



**FAKULTA
INFORMAČNÍCH
TECHNOLOGIÍ
ČVUT V PRAZE**

Zadání diplomové práce

Název:	Optimalizace vyhledávání v judikatuře vrcholných soudů České republiky
Student:	Bc. David Záleský
Vedoucí:	Ing. Jaroslav Kuchař, Ph.D.
Studijní program:	Informatika
Obor / specializace:	Znalostní inženýrství
Katedra:	Katedra aplikované matematiky
Platnost zadání:	do konce letního semestru 2021/2022

Pokyny pro vypracování

- 1) Seznamte se s problémem vyhledávání v judikatuře vrcholných soudů se zaměřením na identifikaci důležitých judikátů a jejich částí.
- 2) Seznamte se se způsobem, jakým vrcholné soudy zveřejňují svá rozhodnutí na internetu. Navrhněte způsob strojového přístupu a transformace do formátu umožňujícím další datovou analýzu.
- 3) Proveďte analýzu těchto dat, zejména ze statistického hlediska, a identifikujte případné chybějící či nekonzistentní záznamy.
- 4) Na základě znalosti problému, získaných dat a provedených analýz navrhněte postup a dílčí metody, jak klasifikovat a s pomocí grafových algoritmů aází provázat soudní rozhodnutí za účelem optimalizace vyhledávání nejrelevantnějších z nich.
- 5) Navržené metody implementujte ve vhodném jazyku.
- 6) Implementované metody ověřte, vyhodnoťte kvalitu a navrhněte možná zlepšení.

Elektronicky schválil/a Ing. Karel Klouda, Ph.D. dne 1. října 2020 v Praze.



**FAKULTA
INFORMAČNÍCH
TECHNOLGIÍ
ČVUT V PRAZE**

Diplomová práce

Optimalizace vyhledávání v judikatuře vrcholných soudů České republiky

Mgr. Bc. David Záleský

Katedra teoretické informatiky

Vedoucí práce: Ing. Jaroslav Kuchař, Ph.D.

27. června 2021

Poděkování

Děkuji vedoucímu práce, Ing. Jaroslavu Kuchařovi, Ph.D., za jeho nezměrnou trpělivost a opakovaně projevenou shovívavost k neplnění dohodnutých termínů z mé strany a k nepřiměřeným průtahům při psaní této práce.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů. V souladu s ust. § 2373 odst. 2 zákona č. 89/2012 Sb., občanský zákoník, ve znění pozdějších předpisů, tímto uděluji nevýhradní oprávnění (licenci) k užití této mojí práce, a to včetně všech počítačových programů, jež jsou její součástí či přílohou a veškeré jejich dokumentace (dále souhrnně jen „Dílo“), a to všem osobám, které si přejí Dílo užít. Tyto osoby jsou oprávněny Dílo užít jakýmkoli způsobem, který nesnižuje hodnotu Díla a za jakýmkoli účelem (včetně užití k výdělečným účelům). Toto oprávnění je časově, teritoriálně i množstevně neomezené. Každá osoba, která využije výše uvedenou licenci, se však zavazuje udělit ke každému dílu, které vznikne (buť jen zčásti) na základě Díla, úpravou Díla, spojením Díla s jiným dílem, zařazením Díla do díla souborného či zpracováním Díla (včetně překladu) licenci alespoň ve výše uvedeném rozsahu a zároveň zpřístupnit zdrojový kód takového díla alespoň srovnatelným způsobem a ve srovnatelném rozsahu, jako je zpřístupněn zdrojový kód Díla.

V Brně dne 27. června 2021

.....

České vysoké učení technické v Praze
Fakulta informačních technologií

© 2021 David Záleský. Všechna práva vyhrazena.

Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí a nad rámec oprávnění uvedených v Prohlášení na předchozí straně, je nezbytný souhlas autora.

Odkaz na tuto práci

Záleský, David. *Optimalizace vyhledávání v judikatuře vrcholných soudů České republiky*. Diplomová práce. Praha: České vysoké učení technické v Praze, Fakulta informačních technologií, 2021.

Abstrakt

Tato práce se zaměřuje na efektivitu vyhledávání v databázích rozhodnutí vrcholných soudů České republiky. Popisuje současná existující řešení a jejich nedostatky a předestírá způsoby, jak zajistit, aby významná rozhodnutí v jednotlivých právních otázkách byla snadněji dostupná. A to jak samotným soudům, tak i široké veřejnosti. Vybraná řešení následně implementuje v programovacím jazyce Python.

Klíčová slova Judikatura, Vrcholné soudy, Efektivní vyhledávání, Významná rozhodnutí, Rozsudky

Abstract

This thesis focuses on the effectiveness of searching the databases of decisions of the supreme courts of the Czech Republic. It describes existing solutions and their shortcomings, and outlines the ways how to make the significant decisions in individual legal matters more easily accessible to both the courts themselves and the general public. The selected solutions are then implemented in the Python programming language.

Keywords Case law, Supreme courts, Effective search, Important decisions, Judgments

Obsah

Úvod	1
1 Základní informace	3
1.1 Základní pojmy	3
1.2 Soudní soustava	4
1.2.1 Trestní soudnictví	4
1.2.1.1 Dovolání	4
1.2.1.2 Významná rozhodnutí	6
1.2.2 Civilní soudnictví	6
1.2.2.1 Přípustnost dovolání	7
1.2.2.2 Významná rozhodnutí	8
1.2.3 Správní soudnictví	8
1.2.4 Ústavní soudnictví	9
1.2.5 Mezinárodní soudnictví	11
1.3 Druhy soudních rozhodnutí podle jejich významu	11
1.3.1 Nálezy Ústavního soudu	11
1.3.2 Rozhodnutí rozšířeného senátu a velkého senátu	12
1.3.3 Rozhodnutí konfliktního senátu	13
1.3.4 Rozhodnutí s právní větou	13
1.3.5 Rozsudky Nejvyššího soudu	14
1.3.6 Rozsudky Nejvyššího správního soudu	14
1.3.7 Rozsudky okresních, krajských a vrchních soudů	14
2 Dostupnost rozhodnutí v elektronické podobě	17
2.1 Ústavní soud	17
2.1.1 Struktura dat	17
2.1.2 Vyhledávač	18
2.2 Nejvyšší soud	19
2.2.1 Struktura dat	19

2.2.2	Vyhledávač	20
2.3	Nejvyšší správní soud	21
2.3.1	Struktura dat	21
2.3.2	Vyhledávač	22
2.3.3	Nový vyhledávač	22
2.4	Nižší soudy	23
2.5	Databáze judikatury	24
2.5.1	Salvia Kraken	24
2.5.2	Placené databáze	25
3	Získávání a analýza dat	27
3.1	Ústavní soud	27
3.1.1	Implementace	28
3.2	Nejvyšší soud	29
3.2.1	Implementace	30
3.3	Nejvyšší správní soud	31
3.3.1	Implementace	32
3.4	Problémy v datech	32
3.4.1	Stará rozhodnutí	32
3.4.2	Chybějící rozhodnutí	34
4	Klasifikace rozhodnutí	35
4.1	Statická klasifikace	35
4.2	Klasifikace na základě vztahů	36
4.2.1	Počet citací	36
4.2.2	PageRank	38
4.2.2.1	Vážený PageRank	39
4.2.2.2	Užití PageRanku ke klasifikaci judikatury	40
4.2.3	Lokální klasifikace	41
5	Implementace	43
5.1	Prostředí	43
5.1.1	Django / Python	43
5.1.2	Databáze	44
5.1.3	Plnění databáze	46
5.1.3.1	Zpracování dat Ústavního soudu	47
5.1.3.2	Zpracování dat Nejvyššího soudu	48
5.1.3.3	Zpracování dat Nejvyššího správního soudu	48
5.1.3.4	Obecné zpracování dat	49
5.2	Několik poznámek k datům	51
5.2.1	Rozsah dat	51
5.2.2	Problémové atributy	51
5.2.3	Zajímavé atributy	52
5.2.3.1	Právní věta	52

5.2.3.2	Soudem posouzený význam	53
5.2.3.3	Další atributy	53
5.3	Klasifikační metody	54
5.3.1	Statická váha	54
5.3.1.1	Volba parametrů	54
5.3.1.2	Možné nedostatky zvoleného postupu	56
5.3.1.3	Podrobnosti o implementaci	56
5.4	Počet citací	56
5.5	PageRank	57
6	Hodnocení kvality	59
6.1	Metodika hodnocení	59
6.2	Problémy s vyhledáváním	60
6.3	Praktický postup hodnocení	62
6.4	Statické metody	63
6.5	PageRank	64
6.6	Vliv jednotlivých případů	66
	Závěr	69
	Literatura	71
	A Seznam použitých zkratk	75
	B Obsah příloženého CD	77
	C Návod k použití zdrojových kódů	79
C.1	Požadavky	79
C.2	Postup instalace	79
C.3	Získávání rozhodnutí	80
C.4	Práce s databází	80

Seznam obrázků

4.1 PageRank vzorec	39
5.1 Databáze	45
6.1 Validací sada	60
6.2 Výsledky měření 1	63
6.3 Výsledky měření 2	65
6.4 Jednotlivé případy	67

Úvod

Nezávislá a efektivní soudní moc je jedním ze základních pilířů každého demokratického právního státu. Řádná dělba moci totiž vyžaduje nejen to, aby zákony byly přijímány v souladu s demokratickými pravidly a základními lidskými právy, nýbrž i to, aby byly orgány moci výkonné uplatňovány a vymáhány férově, bez známek svévole. A vzhledem k tomu, že moc lze vyvažovat jedině mocí, je nezbytné, aby existoval kontrolní orgán příslušný k přezkumu zákonnosti postupů veřejné moci (a případně i ústavnosti samotných zákonů), který bude současně nadán pravomocí zjednat nápravu v případech, kdy shledá, že ze strany ostatních složek veřejné moci došlo k pochybení. Právě tuto roli plní soudní moc.

Aby však soudní moc mohla být považována za efektivní, musí být mimo jiné předvídatelná. Ať již řadový občan, či státní instituce, musí mít možnost s rozumnou mírou pravděpodobnosti předvídat, jak o jejich věci soudy rozhodnou v případech, kdy se již obdobnými otázkami zabývaly v minulosti. Soudci by tedy měli mít povědomí o tom, jak obdobné věci rozhodovaly jiné soudy, přičemž v případě, že by se od těchto dřívějších rozhodnutí chtěli odchýlit, měli by to přesvědčivě odůvodnit. V opačném případě by se soudní řízení stalo zcela nepředvídatelným, a soudy by jen stěží mohly plnit svoji roli ochránce spravedlnosti a základních lidských práv.

Vzhledem k obrovskému počtu rozhodnutí, které soudy vydávají, je však pro řadového soudce objektivně nemožné znát všechna rozhodnutí jiných soudů. Při rozhodování každé věci je proto soudce značně limitován existujícími nástroji, které umožňují práci s judikaturou. Pokud totiž soudce nedokáže významné rozhodnutí týkající se jeho věci snadno najít, nebude je moci při svém rozhodování zohlednit. A vzhledem k tomu, že zájem na předvídatelnosti rozhodování není v činnosti soudů jediný (v tomto kontextu lze především zmínit zájem na rychlosti rozhodování), nemůže soudce strávit

vyhledáváním těchto rozhodnutí příliš mnoho času.

Kvalita nástrojů umožňujících efektivní vyhledávání v existující judikatuře tak přímo ovlivňuje kvalitu (a jednotnost) soudního rozhodování. V této práci se proto zaměřuji právě na možnosti přístupu k existující judikatuře a otázku efektivity vyhledání v ní.

V následujících kapitolách popíši současný stav, co se přístupu k rozhodnutím týče, se zaměřením na to, nakolik jsou současné nástroje při vyhledávání schopny zohlednit význam jednotlivých rozhodnutí. U každého z popisovaných nástrojů rozeberu jejich přednosti i slabé stránky.

V praktické části práce následně navrhnu a implementuji technická řešení, která by mohla přispět ke zefektivnění vyhledávání klíčových rozhodnutí z jednotlivých právních oblastí. Na základě vyhodnocení jejich úspěšnosti bych tak měl být schopný posoudit, zda v oblasti vyhledávání v judikatuře existují nástroje existujícími vyhledávači opomíjené, které by však byly schopny významněji přispět k rozlišování rozhodnutí podle jejich významu.

Smyslem této práce je tedy přispět ke zlepšení přístupu k významným rozhodnutím vrcholných soudů České republiky, a v důsledku toho rovněž přispět ke zvýšení předvídatelnosti rozhodování české justice.

Základní informace

Dříve než přejdu k samotnému zkoumání efektivity vyhledávání, je nutné si z teoretického hlediska vymezit oblast tohoto zkoumání a její základní pojmy. Pro jejich lepší pochopení je však rovněž nutný obecnější úvod zaměřený na organizaci a funkce soudní soustavy v České republice.

1.1 Základní pojmy

Věcí je v této práci myšlen konkrétní případ či spor projednávaný před českými soudy.

Soudním **rozhodnutím** je v této práci myšleno jakékoliv písemné rozhodnutí vydané soudem bez ohledu na to, zda se jedná o rozhodnutí ve věci samé, rozhodnutí dílčí či konečné, či rozhodnutí, kterým jsou řešeny jen některé procesní otázky. Tento pojem je nadřazeným pojmem zejména pro pojmy **rozsudek**, **usnesení** a **nález**, které označují jednotlivé formy rozhodnutí (stejně jako například pojmy **trestní příkaz** či **platební rozkaz**, které však v této práci s ohledem na její zaměření nebudou příliš používány).

Pojem **meritorní** rozhodnutí označuje takové rozhodnutí, kterým se věc u daného soudu končí a kterému předcházelo plnohodnotné posouzení věci. (Tento pojem tedy zejména nezahrnuje rozhodnutí, jimiž bylo řízení zastaveno z procesních důvodů, či jimiž byl opravný prostředek odmítnut pro nepřípustnost, nepřijatelnost nebo zjevnou neopodstatněnost.)

Judikátem je v této práci myšleno takové soudní rozhodnutí, které obsahuje zobecněný (či alespoň snadno zobecnitelný) právní názor na určitou právní otázku, a zpravidla je citováno v následujících rozhodnutích soudů rozhodujících ve věcech týkajících se obdobných otázek.

Judikaturou je v této práci myšlen soubor soudních rozhodnutí.

1.2 Soudní soustava

Od roku 1994, kdy došlo ke zrušení vojenských soudů¹, se česká soudní soustava skládá ze dvou větví – z větve obecného soudnictví a z větve specializovaného soudnictví, kterou v českém prostředí představuje již jen soudnictví ústavní.

Podstatou obecného soudnictví je, že tyto soudy nejsou specializovány na nějakou zvláštní agendu. Rozhodují tedy ve všech oblastech svěřených do působnosti soudní moci s výjimkou těch, které byly svěřeny do působnosti specializovaným soudům. V českém právním prostředí tvoří hlavní náplň činnosti obecných soudů rozhodování ve třech základních právních oblastech – občanskoprávní (též nazývané civilní), trestněprávní a správněprávní.

To, do jaké právní oblasti daná věc spadá, přitom ovlivňuje nejen procesní předpisy a postup v řízení, ale i míru přípustných opravných prostředků a vrcholný soud, který je v dané věci příslušný. Tyto aspekty přitom výrazným způsobem ovlivňují to, jaký soud bude v dané věci nejvyšší autoritou – a z pohledu zaměření této práce tedy i rozhodnutí jakého soudu budou v dané oblasti zpravidla nejvýznamnějšími.

1.2.1 Trestní soudnictví

Řízení ve věcech trestních je v České republice upraveno především v trestním řádu². Jedná se o řízení v zásadě trojinstanční, přičemž nejvyšší instancí je zpravidla Nejvyšší soud České republiky³. V trestních věcech tedy budou nejvýznamnější zdroj informací tvořit rozsudky Nejvyššího soudu, jimiž tento soud rozhodoval o dovoláních. Povaha prvních dvou instancí se liší v závislosti na závažnosti souzeného trestného činu. Závažné trestné činy (tedy ty, u nichž dolní hranice trestní sazby činí nejméně 5 let⁴) soudí v prvním stupni krajské soudy a o odvolání rozhodují vrchní soudy. U méně závažných trestných činů rozhodují v prvním stupni okresní soudy a o odvolání rozhodují soudy krajské.

1.2.1.1 Dovolání

Pro správné zhodnocení významu rozhodnutí Nejvyššího soudu je však nutné si přiblížit i samotnou podstatu dovolání. Dovolání představuje mimořádný opravný prostředek, jímž „*lze napadnout pravomocné rozhodnutí soudu ve*

¹Zákon č. 284/1993, kterým se mění a doplňuje zákon České národní rady č. 436/1991 Sb., o některých opatřeních v soudnictví, o volbách přísedících, jejich zproštění a odvolání z funkce a o státní správě soudů České republiky, ve znění zákona České národní rady č. 580/1991 Sb., zákona České národní rady č. 23/1993 Sb. a zákona č. 171/1993 Sb.

²Zákon č. 141/1961 Sb., o trestním řízení soudním (trestní řád)

³§ 265a a následující trestního řádu

⁴§ 17 trestního řádu

*věci samé, jestliže soud rozhodl ve druhém stupni a zákon to připouští.*⁵ Již z této definice jsou zřejmá dvě úskalí dovolacího přezkumu – dovolání totiž přichází v úvahu pouze tzv. *ve věci samé* a pouze v některých (zákonem připuštěných) případech.

Co *ve věci samé* znamená, upravuje § 265a odst. 2 trestního řádu. Z tohoto ustanovení přitom vyplývá, že rozhodnutí *ve věci samé* je zejména takové rozhodnutí, jímž se řízení končí (tedy např. rozsudek, jímž je obviněný uznán vinným či zproštěn obvinění, či zastavení trestního stíhání). Naproti tomu v dílčích procesních otázkách a v otázkách týkajících se výkonu trestu dovolání přípustné není – nejvýznamnějšími jsou v této oblasti rozhodnutí o vazbě⁶ či o podmíněném propuštění z výkonu trestu⁷.

V trestněprávních otázkách, v nichž dovolání není přípustné, tak nelze vycházet z rozhodnutí Nejvyššího soudu. A to i přesto, že některá taková rozhodnutí existovat budou. (Skutečnost, že dovolání není přípustné, totiž lidem sama o sobě nebrání dovolání podat. Brání jim pouze s ním úspěš.) Taková rozhodnutí však zpravidla nebudou zachycovat žádné významné právní myšlenky, neboť v nich Nejvyšší soud bude pouze konstatovat, že není příslušný k projednání dovolání.

Trestní řád rovněž Nejvyššímu soudu umožňuje odmítnout dovolání, která jsou *zjevně neopodstatněná*⁸. U takto odmítnutých dovolání přitom postačí, když soud „*jen stručně uvede důvod odmítnutí poukazem na okolnosti vztahující se k zákonnému důvodu odmítnutí*“⁹. Ani u těchto rozhodnutí o odmítnutí dovolání tedy nelze očekávat zásadní význam z hlediska v nich obsažené právní argumentace. Nejvyšší soud přitom právě možnosti odmítnout dovolání bez podrobného odůvodnění hojně využívá. Ve své rozhodovací činnosti přitom např. již dovedl, že „*opakuje-li obviněný v dovolání v podstatě jen námitky uplatněné již v řízení před soudem prvního stupně a v odvolacím řízení, s kterými se soudy obou stupňů dostatečně a správně vypořádaly, jde zpravidla o dovolání zjevně neopodstatněné*“¹⁰. S odkazem na tento závěr pak Nejvyšší soud dovolání zpravidla odmítá, pokud se v nich neobjevují nějaké zásadní námitky.

⁵ § 265a odst. 1 trestního řádu

⁶ § 67 a následující trestního řádu

⁷ § 331 a následující trestního řádu

⁸ § 265i odst. 1 písm. e) trestního řádu

⁹ § 265o odst. 2 trestního řádu

¹⁰ Usnesení Nejvyššího soudu ze dne 29. 5. 2002, sp. zn. 5 Tdo 86/2002

1.2.1.2 Významná rozhodnutí

Z výše uvedeného je zřejmé, že pouze malá část rozhodnutí Nejvyššího soudu o dovoláních představuje významná rozhodnutí z hlediska zaměření této práce. Bude se jednat především o ta rozhodnutí o dovolání, v nichž se Nejvyšší soud dovoláním věcně zabýval a meritorně o něm rozhodl včetně úplného odůvodnění. Půjde tedy o taková rozhodnutí, jimiž bylo dovolání vyhověno či jimiž bylo dovolání zamítnuto (nikoliv odmítnuto). Vhodným klíčem k rozlišení těchto rozhodnutí může být jejich forma. Rozhodnutí, jimiž Nejvyšší soud dovolání odmítá, jsou totiž vydávána formou usnesení¹¹, zatímco rozhodnutí vydávaná po plnohodnotném meritorním posouzení věci jsou vydávána formou rozsudků.

Nejvýznamnějšími rozhodnutími v trestněprávní oblasti tedy budou zpravidla rozsudky Nejvyššího soudu v těch právních otázkách, v nichž je přípustné dovolání a rozsudky odvolacích soudů (krajských či vrchních v závislosti na závažnosti souzeného trestného činu) v těch otázkách, v nichž dovolání přípustné není. Vzhledem k organizaci soudní soustavy a systému opravných prostředků přitom nelze nadřazovat význam rozhodnutí vrchních soudů (u závažnějších zločinů) nad rozhodnutí krajských soudů (u méně závažných trestných činů). V těchto případech totiž vrchní soudy nejsou nadřazené soudům krajským a nemají pravomoc přezkoumávat a hodnotit jejich rozhodnutí o odvolání.

1.2.2 Civilní soudnictví

Řízení ve věcech civilních (slovy zákona a starší literatury též občanskoprávních) je v České republice upraveno především v občanském soudním řádu¹² a zákoně o zvláštních řízeních soudních¹³. I civilní řízení je zpravidla trojinstanční v čele s Nejvyšším soudem¹⁴, a tedy i v civilních věcech budou nejvýznamnější zdroj informací tvořit rozsudky Nejvyššího soudu, jimiž tento soud rozhodoval o dovoláních. Oproti trestnímu soudnictví je však mnohem výraznější rozsah věcí, v nichž není dovolání přípustné, a existuje dokonce nezanedbatelné množství věcí, v nichž není přípustné ani odvolání.

Strukturou soudní soustavy, procesními postupy i skladbou opravných prostředků se z hlediska zaměření této práce civilní soudnictví značně podobá soudnictví trestnímu, a není tedy nutné vše výše uvedené znovu opakovat. V následující části tedy budou uvedeny zejména ty nejvýznamnější odlišnosti.

¹¹§ 265i odst. 2 trestního řádu

¹²Zákon č. 99/1963 Sb., občanský soudní řád

¹³Zákon č. 292/2013 Sb., o zvláštních řízeních soudních

¹⁴§ 236 a následující občanského soudního řádu

1.2.2.1 Přípustnost dovolání

Obdobně jako v trestních věcech, tak i v těch civilních platí, že dovolání představuje mimořádný opravný prostředek, jímž „lze napadnout pravomocná rozhodnutí odvolacího soudu, pokud to zákon připouští.“¹⁵ Ustanovení § 238 však významným způsobem omezuje přípustnost dovolání, když mimo jiné stanoví, že dovolání není přípustné „proti rozsudkům a usnesením vydaným v řízeních, jejichž předmětem bylo v době vydání rozhodnutí obsahujícího napadený výrok peněžité plnění nepřevyšující 50 000 Kč, včetně řízení o výkon rozhodnutí a exekučního řízení, ledaže jde o vztahy ze spotřebitelských smluv a o pracovněprávní vztahy“. Obdobné omezení přitom občanský soudní řád zakotvuje i ve vztahu k odvolání, když stanoví, že „odvolání není přípustné proti rozsudku vydanému v řízení, jehož předmětem bylo v době vydání rozsudku peněžité plnění nepřevyšující 10 000 Kč“¹⁶. Tato omezení přitom vedou k tomu, že o některých v praxi se často vyskytujících otázkách (jako např. jízda načerno ve veřejné hromadné dopravě či rychlé spotřebitelské půjčky částek nižší hodnoty) probíhá pouze jednoinstanční řízení před okresními soudy bez možnosti uplatnění opravných prostředků, a tedy i bez možnosti sjednocovat rozhodovací činnost ze strany Nejvyššího soudu.

Obdobná omezení přitom zavádí i zákon o zvláštních řízeních soudních ve vztahu k některým typům řízení. Podle tohoto zákona je např. nepřípustné dovolání ve věcech týkajících se rozvodu manželství či péče o nezletilé děti.¹⁷ Ani v těchto otázkách tedy nelze spoléhat na sjednocovací roli Nejvyššího soudu.

I v civilním soudnictví občanský soudní řád umožňuje Nejvyššímu soudu odmítnout dovolání bez provedení podrobného přezkumu a odůvodnění. Dovolání je totiž přípustné pouze proti takovým rozhodnutím, která „závisí na vyřešení otázky hmotného nebo procesního práva, při jejímž řešení se odvolací soud odchýlil od ustálené rozhodovací praxe odvolacího soudu nebo která v rozhodování odvolacího soudu dosud nebyla vyřešena nebo je dovolacím soudem rozhodována rozdílně anebo má-li být dovolacím soudem vyřešena právní otázka posouzena jinak.“¹⁸ To v podstatě znamená, že Nejvyšší soud může odmítat dovolání založená na právních otázkách, které již dříve posoudil, a nemá v úmyslu se od tohoto posouzení odchýlovat. (Samozřejmě pouze za předpokladu, že soudy nižší instance postupovaly v souladu s těmito dříve vyslovenými závěry.) Ani u těchto rozhodnutí o odmítnutí dovolání tedy nelze očekávat zásadní význam z hlediska v nich obsažené právní argumentace.

¹⁵ § 236 odst. 1 občanského soudního řádu

¹⁶ § 202 odst. 2 občanského soudního řádu

¹⁷ § 30 zákona o zvláštních řízeních soudních

¹⁸ § 237 občanského soudního řádu

1.2.2.2 Významná rozhodnutí

Z výše uvedeného je zřejmé, že i v civilní oblasti pouze malá část rozhodnutí Nejvyššího soudu o dovoláních představuje významná rozhodnutí z hlediska zaměření této práce. Bude se jednat především o ta rozhodnutí o dovolání, v nichž se Nejvyšší soud dovoláním meritorně zabýval a řádně o něm rozhodl včetně úplného odůvodnění. Půjde tedy o taková rozhodnutí, jimiž bylo dovolání vyhověno či jimiž bylo dovolání zamítnuto (nikoliv odmítnuto). Stejně jako v případě trestního soudnictví půjde o rozhodnutí vydaná ve formě rozsudku, nikoliv usnesení.

Usnesení o odmítnutí dovolání jsou však v civilním právu pro účely této práce mnohem významnější než ta trestněprávní. Ústavní soud totiž ve vztahu k odmítání dovolání dovedl, že „rozhodnutí Nejvyššího soudu, z jehož odůvodnění se nelze dovědět, proč otázku předloženou mu v dovolání nepovažoval za otázku zásadního právního významu (např. citací judikátů tohoto soudu, které věc řeší a pro jejichž změnu či odchýlení se od nich neshledal Nejvyšší soud důvod), je třeba považovat za nepřezkoumatelné, a tudíž stojící v rozporu s principy právního státu.“¹⁹ Nejvyšší soud pak v reakci na toto rozhodnutí Ústavního soudu začal ve svých odmítavých usneseních poctivě uvádět odkazy na svá dřívější rozhodnutí, v nichž předmětné otázky věcně posuzoval. A i když tedy tato usnesení primárně nebude možné považovat za zásadní z hlediska zaměření této práce, lze předpokládat, že v nich uvedené odkazy budou směřovat k rozhodnutím svým významem podstatnějším.

1.2.3 Správní soudnictví

Řízení ve věcech správněprávních prošlo zásadní změnou v roce 2003, kdy vstoupil v účinnost soudní řád správní²⁰, který významným způsobem zreformoval organizaci soudní soustavy ve správním soudnictví. Do té doby roztržštěná soudní příslušnost – významně se lišící podle jednotlivých právních oblastí – byla sjednocena. V první instanci nyní ve správněprávních věcech zásadně rozhodují krajské soudy. Rozhodování o opravném prostředku proti jejich rozhodnutí pak přísluší nově zřízenému Nejvyššímu správnímu soudu. (Tento opravný prostředek se ve správním soudnictví označuje jako kasační stížnost.)

Vzhledem k tomu, že kasační stížnost je přípustná v zásadě proti každému rozhodnutí krajského soudu, je situace ve správním soudnictví mnohem přehlednější. Rozhodnutími zásadní povahy tak zde v naprosté většině případů budou rozsudky Nejvyššího správního soudu. Možnost odmítat kasační stížnosti bez plnohodnotného přezkumu byla navíc dosud značně

¹⁹Nález Ústavního soudu ze dne 15. 7. 2004, sp. zn. II. ÚS 594/02

²⁰Zákon č. 150/2002 Sb., soudní řád správní

omezena (týkala se v podstatě pouze přezkumu rozhodnutí o neudělení azylu²¹), takže naprostá většina rozhodnutí Nejvyššího správního soudu obsahuje plnohodnotné věcné posouzení.

Novelou účinnou od 1. 4. 2021 došlo k rozšíření tzv. institutu nepřijatelnosti tím, že bylo nově zakotveno, že „*jestliže kasační stížnost ve věcech, v nichž před krajským soudem rozhodoval specializovaný samosoudce, svým významem podstatně nepřesahuje vlastní zájmy stěžovatele, odmítne ji Nejvyšší správní soud pro nepřijatelnost.*“²² Otázkou přesahu vlastních zájmů stěžovatele se přitom Nejvyšší správní soud zabýval již za účinnosti předchozí právní úpravy, přičemž dospěl k závěru, že kasační stížnost přesahuje vlastní zájmy stěžovatele „*v následujících typových případech: 1) kasační stížnost se dotýká právních otázek, které dosud nebyly vůbec či nebyly plně řešeny judikaturou Nejvyššího správního soudu; 2) kasační stížnost se týká právních otázek, které jsou dosavadní judikaturou řešeny rozdílně; 3) kasační stížnost bude přijatelná pro potřebu učinit judikaturní odklon, kdy Nejvyšší správní soud ve výjimečných a odůvodněných případech sezná, že je namístě změnit výklad určité právní otázky řešené dosud správními soudy jednotně; 4) kasační stížnost je dále přijatelná, pokud by bylo v napadeném rozhodnutí krajského soudu shledáno zásadní pochybení, které mohlo mít dopad do hmotněprávního postavení stěžovatele.*“²³

Ze srovnání výše uvedeného výkladu provedeného Nejvyšším správním soudem s podmínkami dovolání v civilních věcech přitom na první pohled vyplývá jejich značná podobnost. Z toho lze dovozovat, že i u kasačních stížností odmítnutých pro nepřijatelnost lze rovněž s vysokou mírou pravděpodobnosti očekávat, že budou obsahovat odkazy na rozsudky, které se posuzovanými právními otázkami již podrobně zabývaly.

1.2.4 Ústavní soudnictví

Jedinou oblastí výkonu soudní moci, která se může pyšnit existencí vlastního specializovaného soudu, je v České republice ústavní soudnictví. Ústavní soud tedy není nadřízeným soudem žádného obecného soudu a není oprávněn přezkoumávat jejich rozhodnutí z hlediska jejich prosté správnosti. Tuto skutečnost připomíná téměř v každém ze svých rozhodnutí zhruba tímto způsobem: „*Ústavní soud je soudním orgánem ochrany ústavnosti (čl. 83 Ústavy), který stojí mimo soustavu soudů (čl. 91 odst. 1 Ústavy). Vzhledem k tomu jej nelze, vykonává-li svoji pravomoc tak, že podle čl. 87 odst. 1 písm. d) Ústavy rozhoduje o ústavní stížnosti proti pravomocnému soudnímu rozhodnutí, považovat za další, „superrevizní“ instanci v systému obecné justice,*

²¹§ 104a soudního řádu správního ve znění účinném do 31. 3. 2021

²²§ 104a soudního řádu správního

²³Usnesení Nejvyššího správního soudu ze dne 26. 4. 2006, č. j. 1 Azs 13/2006 - 39

oprávněnou svým vlastním rozhodováním (nepřímo) nahrazovat rozhodování obecných soudů; jeho úkolem je „toliko“ přezkoumat ústavnost soudních rozhodnutí, jakož i řízení, které jejich vydání předcházelo. Nutno proto vycházet z pravidla, že vedení řízení, zjišťování a hodnocení skutkového stavu, výklad podústavního práva a jeho aplikace na jednotlivý případ je v zásadě věcí obecných soudů, a o zásahu Ústavního soudu do jejich rozhodovací činnosti lze uvažovat za situace, kdy je jejich rozhodování stíženo vadami, které mají za následek porušení ústavnosti.“²⁴

Ústavní soud je tedy oprávněn zrušit jakékoliv rozhodnutí obecného soudu, avšak může tak učinit pouze tehdy, pokud daným rozhodnutím (či postupem řízení, v jehož rámci bylo rozhodnutí vydáno) došlo k porušení některého z ústavních principů. Vzhledem k tomu, že součástí ústavního pořádku České republiky je i Listina základních práv a svobod²⁵, přistupuje Ústavní soud k rušení rozhodnutí obecných soudů zejména tehdy, došlo-li jimi k porušení některého ze základních lidských práv účastníků řízení. (Nejčastěji se jedná o právo na spravedlivý proces zaručené článkem 36 Listiny, nicméně není výjimečné, že Ústavní soud shledává i porušení jiných práv.)

Výše uvedené má pro tuto práci zásadní význam, a to v tom, že to umožňuje nahlížet na význam rozhodnutí Ústavního soudu ze správné perspektivy. K těmto rozhodnutím je totiž nutné přistupovat s určitou rezervou. Pokud Ústavní soud ústavní stížnosti nevyhoví, neznamená to automaticky, že rozhodnutí obecných soudů byla správná. Znamená to pouze tolik, že nebyla protiústavní. Naopak, pokud Ústavní soud rozhodnutí zruší pro jeho protiústavnost, neznamená to, že tímto rozhodnutím podává závazný výklad všech otázek, které před ním řešily obecné soudy. Ústavní soud se ve svém rozhodování vyjadřuje pouze k otázkám s ústavněprávním přesahem, přičemž řešení otázek výkladu podústavního práva nechává na obecných soudech. Zcela běžně se tedy stává, že Ústavní soud ruší rozhodnutí obecných soudů například pro jejich nepřezkoumatelnost²⁶ nebo skutečnost, že některý z účastníků řízení neměl prostor k uplatnění svých procesních práv²⁷. A přestože takováto rozhodnutí jsou sice nepochybně významná z hlediska výkladu práva na spravedlivý proces, zpravidla nejsou významná z hlediska otázek, které posuzovaly obecné soudy v předcházejících řízeních.

Rozhodnutí Ústavního soudu tedy představují významná rozhodnutí zejména z hlediska výkladu základních lidských práv. Vzhledem k Ústavním soudem zastávané doktríně „prozařování“ základních lidských práv napříč

²⁴Nález Ústavního soudu ze dne 2. 9. 2020, sp. zn. IV. ÚS 4156/19

²⁵článek 3 Ústavy České republiky

²⁶srov. např. nález Ústavního soudu ze dne 18. 2. 2020, sp. zn. I. ÚS 2002/19

²⁷srov. např. nález Ústavního soudu ze dne 19. 12. 2019, sp. zn. I. ÚS 3073/19

celým právním řádem²⁸ sice není úplně neobvyklé, že se Ústavní soud vyjadřuje i k podústavní úpravě, nicméně i v těchto případech tak činí pouze v rozsahu, jaký považuje za nezbytný pro naplnění základních lidských práv, a nesnaží se tedy o ucelený výklad dané problematiky.

1.2.5 Mezinárodní soudnictví

Významným zdrojem důležitých judikátů jsou rovněž některé mezinárodní soudy. Zejména se jedná o Soudní dvůr Evropské unie (který je oprávněn autoritativně posuzovat právní otázky vycházející z předpisů založených na právu Evropské Unie – tedy zejména v oblasti ochrany spotřebitele a některých daní) a o Evropský soud pro lidská práva (který plní roli obdobnou Ústavnímu soudu, avšak místo české Listiny základních práv a svobod se opírá o Evropskou úmluvu o ochraně lidských práv). Některá rozhodnutí těchto soudů představují zásadní prameny vnitrostátní judikatury a jsou hojně citována v rozhodnutích českých soudů.

Vzhledem k zaměření této práce (zabývající se judikaturou českých vrcholných soudů) však tato rozhodnutí (navzdory významu některých z nich) nebudou detailněji zkoumána.

1.3 Druhy soudních rozhodnutí podle jejich významu

V předchozí části byl představen nástin fungování české soudní soustavy a její rozdělení do jednotlivých relativně samostatných částí. Rovněž v ní byly stručně představeny některé formy rozhodovací činnosti a jejich předpokládaný význam. Jelikož cílem této práce je usnadnit orientaci ve významných soudních rozhodnutích, pokusím se v této kapitole formulovat některá základní pravidla, podle nichž lze předběžně odhadnout význam jednotlivých soudních rozhodnutí, a to zejména s ohledem na jejich formu a vydávající soud.

1.3.1 Nálezy Ústavního soudu

Přestože v předchozí části byla uvedena mnohá omezení, která mohou redukovat význam rozhodnutí Ústavního soudu, i tak se v mnoha právních oblastech jedná o nejvýznamnější zdroj judikatury. Bude tomu tak zejména v případě tzv. nálezů. (Výraz „náleží“ v tomto kontextu představuje označení formy rozhodnutí Ústavního soudu²⁹). V zásadě se jedná o formu, která více méně odpovídá rozsudkům obecných soudů. Užití odlišného názvu

²⁸srov. např. náleží Ústavního soudu ze dne 24. 4. 2014, sp. zn. IV. ÚS 2265/13

²⁹§ 54 Zákona č. 182/1993 Sb., o Ústavním soudu

pro obdobnou formu rozhodnutí v případě Ústavního soudu však podle právních teoretiků „lépe vyjadřuje skutečnost, že soud nalézá právo – zejména v přirozenoprávním smyslu – a nikoliv že by právo tvořil nebo naopak jej pouze jako stroj aplikoval doslova“³⁰⁾

Právě v nálezech totiž Ústavní soud plnohodnotně věcně posuzuje soulad procesního postupu, výkladu práva, či samotné zákonné úpravy s ústavními principy a základními lidskými právy. Vzhledem k dlouhodobosti fungování Ústavního soudu, rozsahu jeho činnosti a ustálenosti jeho judikatury přitom platí, že naprostou většinu ústavních stížností Ústavní soud usnesením odmítá pro zjevnou neopodstatněnost³¹⁾, jelikož se již obdobnými otázkami opakovaně zabýval a má na ně ustálený právní názor. Přistoupí-li tak Ústavní soud k plnohodnotnému meritornímu přezkumu ústavní stížnosti ústíci mu ve vydání nálezu, zpravidla to znamená, že ve věci shledal závažné porušení lidských práv. To platí zejména pro tzv. „senátní“ nálezy – tedy nálezy, které vydal některý z tříčlenných senátů Ústavního soudu v řízení o ústavních stížnostech. V případě, že v senátní věci Ústavní soud přistoupí k rozhodnutí nálezem, téměř vždy to znamená, že alespoň jeden ze členů senátu shledal v předchozím postupu obecných soudů něco protiústavního.

Kromě tříčlenných senátů rozhoduje Ústavní soud též v plénu – tedy ve sboru všech patnácti soudců. V plenárních věcech neplatí, že by vydání nálezu téměř vždy implikovalo shledání protiústavnosti (neboť plénum rozhoduje nálezem prakticky vždy). Význam plenárních nálezů je však zpravidla dán významem posuzovaných otázek. V plénu totiž Ústavní soud rozhoduje zejména o posouzení souladu platných zákonů s Ústavou. Výsledek tohoto posouzení (resp. důvody, o něž Ústavní soud své rozhodnutí opřel) je tedy pro interpretaci těchto zákonů významný bez ohledu na to, zda je Ústavní soud shledal souladnými s Ústavou, či nikoliv.

1.3.2 Rozhodnutí rozšířeného senátu a velkého senátu

V oblasti správního práva mezi nejvýznamnější judikáty patří zejména rozhodnutí tzv. rozšířeného senátu Nejvyššího správního soudu. Nejvyšší správní soud zpravidla rozhoduje ve tříčlenných senátech. Dospěje-li však při svém rozhodování tříčlenný senát „k právnímu názoru, který je odlišný od právního názoru již vyjádřeného v rozhodnutí Nejvyššího správního soudu“, je povinen postoupit věc k rozhodnutí rozšířenému senátu³²⁾. Cílem tohoto postupu je zajistit jednotnost v rozhodování Nejvyššího správního soudu a obtížnost změnitelnosti dříve přijatých právních názorů. Pokud tedy

³⁰⁾DOSTÁL, M., LANGÁŠEK, T., POSPÍŠIL, I., WAGNEROVÁ, E. *Zákon o Ústavním soudu: Komentář*. Wolters Kluwer [cit. 2021-6-22]. ISSN 2336-517X.

³¹⁾§ 43 odst. 2 písm. a) zákona o Ústavním soudu

³²⁾§ 17 odst. 1 soudního řádu správního

1.3. Druhy soudních rozhodnutí podle jejich významu

o některé právní otázky rozhodoval rozšířený senát Nejvyššího správního soudu, lze to považovat za zpravidla nejvýznamnější posouzení této otázky. A vzhledem k tomu, že rozšířený senát běžně neposuzuje tytéž otázky opakovaně (s výjimkou případů, kdy některý ze tříčlenných senátů dospěje k závěru, že by se předchozí právní názor měl změnit), lze mít oprávněně za to, že každé rozhodnutí rozšířeného senátu (v němž se tento senát danou otázkou věcně zabýval) představuje významné rozhodnutí o unikátní právní otázce.

Obdobou rozšířeného senátu je u Nejvyššího soudu tzv. *velký senát kolegia* (či jen *velký senát*)³³. I ten funguje jako nástroj pro sjednocování judikatury, neboť stejně jako v případě rozšířeného senátu je možné změnit dříve vyslovený právní názor pouze předložením věci velkému senátu. I rozhodnutí tohoto senátu proto lze zpravidla považovat za významné judikáty.

1.3.3 Rozhodnutí konfliktního senátu

V právních otázkách vztahujících se k rozhraní působnosti mezi civilní a správní justicí, případně mezi civilní justicí a správními orgány, jsou klíčovými rozhodnutí tzv. zvláštního senátu zřízeného podle zákon č. 131/2002 Sb.³⁴ (též označovaného jako „*konfliktní senát*“). Tento senát byl zřízen za účelem rozhodování kompetenčních sporů mezi vybranými větvemi veřejné správy a je tvořen třemi soudci Nejvyššího soudu a třemi soudci Nejvyššího správního soudu.

Přestože se tento senát zaměřuje na velmi úzkou oblast právních otázek, velká část jeho rozhodnutí představuje v této oblasti zásadní judikáty. To je dáno i tím, že jakmile tento senát nějakou věc jednou rozhodne, jen zřídkakdy je mu předložena stejná či obdobná věc k rozhodování, neboť soudy i správní orgány se jeho rozhodnutími důsledně řídí.

1.3.4 Rozhodnutí s právní větou

U všech tří vrcholných soudů České republiky mají soudci možnost při vkládání rozhodnutí do databáze opatřit jej tzv. právní větou. Právní věta představuje stručné a výstižné vyjádření hlavní myšlenky, na níž je vydané rozhodnutí založeno. Právními větami jsou opatřována převážně taková rozhodnutí, která řeší dosud neřešenou problematiku, či v nichž soud zaujal nový právní názor. Tato rozhodnutí tedy rovněž mají vysoký potenciál představovat významné judikáty ve své oblasti.

³³§ 19 odst. 1 Zákona č. 6/2002 Sb., o soudech, soudcích, přísedících a státní správě soudů a o změně některých dalších zákonů (zákon o soudech a soudcích)

³⁴Zákon č. 131/2002 Sb., o rozhodování některých kompetenčních sporů

Vzhledem k tomu, že opatřit rozhodnutím právní větou není povinnost (naopak, soudci jsou spíše vedeni k tomu, aby jejich užíváním šetřili), představuje již sama existence právní věty významnou informaci. Skutečnost, že rozhodnutí je opatřené právní větou, totiž svědčí o tom, že soudce, který jej vydal, jej považoval za důležité.

1.3.5 Rozsudky Nejvyššího soudu

I u rozsudků Nejvyššího soudu lze zpravidla mít za to, že se jedná o významné judikáty, a to zejména s ohledem na převládající praxi Nejvyššího soudu spočívající v masivním odmítání dovolání ve věcech, v nichž již Nejvyšší soud dříve rozhodoval. Vzhledem k tomu, že odmítnutí dovolání se provádí usnesením, lze tedy mít důvodně za to, že přistoupí-li Nejvyšší soud k vydání rozsudku, činí tak buď proto, aby posoudil dosud neřešenou právní otázku, změnil svůj dosavadní názor, nebo zrušil rozhodnutí soudů nižšího stupně pro nerespektování dříve vysloveného právního názoru Nejvyššího soudu.

1.3.6 Rozsudky Nejvyššího správního soudu

U rozsudků Nejvyššího správního soudu je již situace podstatně horší. Nejvyšší správní soud totiž dosud neměl příliš širokou možnost odmítat kasační stížnosti, takže se všemi musel věcně zabývat a rozhodnout o nich rozsudkem. Existuje tedy velké množství rozsudků pojednávajících o téže právní otázce (zpravidla stejným způsobem), přičemž není možné snadno určit, které z nich lze považovat za významné z hlediska kvality odůvodnění.

1.3.7 Rozsudky okresních, krajských a vrchních soudů

Jak již bylo výše uvedeno, v některých věcech není přípustné dovolání (a někdy ani odvolání) a danou věcí se tak nemůže zabývat Nejvyšší soud. Pokud je přitom nějaká právní oblast charakteristická právě tím, že u věcí do ní spadajících nejsou některé opravné prostředky přípustné, nemůže Nejvyšší soud v těchto otázkách sjednocovat judikaturu. Pokud pak daná právní otázka nedosahuje ústavněprávního významu, neexistuje žádný soud s celostátní působností, který by se k ní mohl závazně vyjádřit.

V těchto případech je proto za významná nutné považovat rozhodnutí těch soudů, které rozhodují v poslední instanci (zpravidla tedy soudy odvolací). U těch je však situace ještě horší než v případě rozsudků Nejvyššího správního soudu. Bez podrobnější analýzy textu rozhodnutí totiž nelze snadno určit, zda v dané věci bylo přípustné dovolání, či nikoliv. A i u skupiny rozsudků, v nichž dovolání přípustné není, narážíme na tentýž problém jako u Nejvyššího správního soudu – tedy na existenci velkého množství

1.3. Druhy soudních rozhodnutí podle jejich významu

obdobných rozsudků řešících tutéž otázku.

V tomto případě je však situace ještě horší v tom, že tyto rozsudky nevydává pouze jeden soud, nýbrž v případě vrchních soudů 2, a v případě krajských soudů dokonce 8. Není přitom výjimkou, že se napříč jednotlivými soudy přístup ke stejným právním otázkám značně liší. Za významná by v takovém případě bylo nutné považovat všechna rozhodnutí, která ve svém souhrnu dávají ucelený obraz o rozhodování těchto právních otázek napříč republikou.

Dostupnost rozhodnutí v elektronické podobě

Znalost typů rozhodnutí a jejich předpokládaného významu je pro tuto práci sice důležitá, nicméně rozhodně není sama o sobě dostatečná. Aby totiž bylo možné v rozhodnutích efektivně vyhledávat pomocí nástrojů výpočetní techniky a znalostního inženýrství, je nejprve nutné mít tato rozhodnutí k dispozici ve strojově zpracovatelném formátu. V této kapitole se proto zaměřím na to, jakým způsobem jsou soudní rozhodnutí přístupná veřejnosti a jak je v nich možné vyhledávat.

2.1 Ústavní soud

Ústavní soud zveřejňuje všechna svá rozhodnutí v aplikaci zvané NALUS dostupné na webu nalus.usoud.cz. Tato databáze obsahuje v zásadě všechna rozhodnutí Ústavního soudu od doby jeho vzniku (v roce 1993). A přestože je na webu uvedeno, že „*rozhodnutí Ústavního soudu publikovaná v elektronické podobě v aplikaci NALUS slouží pouze pro informaci o rozhodovací činnosti Ústavního soudu*“, odpovídají uvedená rozhodnutí v drtivě většině případů jejich přesnému znění, přičemž jediným zásadním rozdílem oproti autentickému znění je jejich anonymizace. Pro účely vyhledávání významných judikátů je proto takto pojatá databáze zcela dostatečná, neboť identita konkrétních účastníků jednotlivých řízení je pro zobecněné posouzení dotčených právních otázek zcela irelevantní.

2.1.1 Struktura dat

Co se formy týče, jsou rozhodnutí v NALUSu dostupná ve formátu html a rtf. Html je přitom v podstatě srovnatelné s plaintextem, neboť jediným rozdílem je občasné ztučnění písma či kurzíva. Získání obsahu jednotlivých rozhodnutí ve strojově zpracovatelné formě tedy není příliš náročné. Aplikace

NALUS však neumožňuje hromadné stahování obsahu rozhodnutí (respektive umožňuje jich najednou stáhnout pouze 20). Na druhou stranu však ani nijak významným způsobem neomezuje přístup ke svému vyhledávači, takže je při vyvinutí jisté snahy možné všechna rozhodnutí jedno po druhém stáhnout.

Aplikace NALUS rovněž kromě samotného textu rozhodnutí obsahuje rozsáhlou sadu metadat. Z hlediska této práce jsou přitom zajímavými především *spisová značka* (umožňující identifikaci rozhodnutí), *forma rozhodnutí* (zda jde o nález či usnesení), *předmět řízení* a *věcný rejstřík* (oba obsahující informace o právní oblasti, jíž se rozhodnutí věnuje) a *dotčené předpisy* (obsahující seznam relevantních zákonných ustanovení, která byla v dané věci významná). Jednoznačně nejzajímavější je však položka *význam*, jejímž prostřednictvím sám Ústavní soud hodnotí judikaturní význam daného rozhodnutí na pětistupňové škále. Zatímco stupně 4 a 5 jsou vyhrazeny odmítavým a procesním usnesením, stupně 1, 2 a 3 jsou schopny ještě lépe rozlišit význam nálezů – tedy zejména vytrdit ty, které nově vyslovují významné právní myšlenky (zejména stupeň 1, částečně i 2) a ty, které tyto již dříve formulované závěry pouze aplikují v následných věcech (stupeň 3).

2.1.2 Vyhledávač

Jedním z hlavních cílů aplikace NALUS je rovněž umožnění efektivního vyhledávání v judikatuře Ústavního soudu. Tohoto úkolu se přitom zhostila celkem dobře. Vzhledem k rozsáhlému množství metadat, jimiž jsou rozhodnutí opatřena, totiž umožňuje velmi vysokou úroveň precizace dotazu. Kromě základních filtrů jako jsou datum vydání, typ rozhodnutí či soudce, obsahuje i mnohé filtry obsahové. Vyhledávání judikatury vztahující se ke konkrétní právní oblasti tak podporuje možnost filtrace podle konkrétních ustanovení právních předpisů, která jsou rozhodnutím dotčena, tak rovněž možnost filtrace podle předmětu řízení a věcného rejstříku. Předmět řízení a věcný rejstřík jsou rozsáhlé číselníky, přičemž u každého rozhodnutí platí, že soudce (či spíše asistent soudce) mu při jeho vkládání do databáze přiřadí některou kategorii (nebo i více kategorií) z uvedených číselníků. V případě, že se tedy povahou svého dotazu uživatel střelí do existující položky některého z číselníků, poskytuje NALUS velmi přesné výsledky.

NALUS rovněž podporuje fulltextové vyhledávání včetně možnosti použití některých logických či jiných operátorů, jimiž lze precizovat vztah mezi hledanými pojmy. V základním režimu přitom vyhledává výsledky bez ohledu na gramatický tvar hledaných slov. Kromě běžných operátorů jako AND, OR, či "", obsahuje i pokročilejší operátory jako SENTENCE, PARAGRAHPH či NEAR(n), které umožňují vyhledat pouze ta rozhodnutí, v nichž se hledaná slova objevují v menší vzdálenosti (např. ve stejné větě či odstavci).

Mezi nevýhody NALUSu pak patří skutečnost, že neumožňuje filtraci podle přítomnosti právní věty a že význam rozhodnutí je omezen pouze na 5 kategorií, přičemž v rámci těchto kategorií nelze určit, která rozhodnutí jsou významnější. Aplikace přitom nijak nehodnotí význam rozhodnutí ve vztahu k zadaným vyhledávacím parametrům. Všechna rozhodnutí, která tyto parametry splní, jsou považována za rovnocenné výsledky vyhledávání – aplikace hodnotí relevanci výsledků pouze binárně (vyhovuje/nevyhovuje), aniž by zohledňovala, že některá rozhodnutí mohou být pro dané vyhledávání relevantnější.

2.2 Nejvyšší soud

I Nejvyšší soud zveřejňuje svá rozhodnutí online, a to na svém webu nsoud.cz/Judikatura/ns_web.nsf/WebSpreadSearch. Tato databáze obsahuje v zásadě všechna rozhodnutí Nejvyššího soudu od 1. 6. 2000. Stejně jako u NALUSu lze přitom i u Nejvyššího soudu konstatovat, že zveřejněná rozhodnutí v zásadě odpovídají anonymizovaným verzím autentických znění, a pro účely vyhledávání významných judikátů jsou tedy zcela dostatečná.

2.2.1 Struktura dat

Co se formy týče, jsou rozhodnutí v databázi Nejvyššího soudu dostupná ve formátu html a doc. Html je přitom rovněž v podstatě srovnatelné s plaintextem, neboť jediným rozdílem je občasné ztučnění písma či kurzíva. Získání obsahu jednotlivých rozhodnutí ve strojově zpracovatelné formě tedy není příliš náročné. Ani web Nejvyššího soudu neumožňuje hromadné stahování obsahu rozhodnutí. Rovněž neumožňuje prázdný dotaz, a omezuje maximální počet výsledků na 900. Pro stažení celého obsahu databáze tedy bude nutné položit takovou sadu dotazů, která pokryje všechna rozhodnutí a zároveň žádný z jednotlivých dotazů nevrátí víc než 900 výsledků. Jako ideální se proto nabízí omezení podle data vydání rozhodnutí (třeba po týdnech či měsících v závislosti na počtu rozhodnutí).

I web Nejvyššího soudu k rozhodnutím obsahuje sadu metadat, byť ani z daleka ne tak rozsáhlou jako NALUS. Z hlediska této práce jsou přitom zajímavými *spisová značka* (umožňující identifikaci rozhodnutí), *typ rozhodnutí* (zda jde o rozsudek či usnesení), *heslo* (obsahující informace o právní oblasti, jíž se rozhodnutí věnuje), *právní věta a dotčené předpisy* (obsahující seznam relevantních zákonných ustanovení, která byla pro rozhodnutí v dané věci podstatná). Stejně jako Ústavní soud, tak i Nejvyšší soud připojuje k rozhodnutím informaci o jejich významu, a to v položce *kategorie rozhodnutí*.

2. DOSTUPNOST ROZHODNUTÍ V ELEKTRONICKÉ PODOBĚ

Kategorizace probíhá v rámci následujících skupin³⁵:

- A – rozhodnutí navržená k publikaci ve Sbírce soudních rozhodnutí a stanovisek. Tato rozhodnutí jsou rozesílána k připomínkám vybraným institucím (např. ústřední orgány státní správy, soudy, Nejvyšší státní zastupitelství, právnické fakulty, profesní komory) a poté projednána na zasedání příslušného kolegia Nejvyššího soudu. Není-li rozhodnutí k publikaci schváleno, přeřadí se do nižší kategorie.
- B – významná rozhodnutí nezařazená do skupiny A.
- C – ostatní rozhodnutí řešící hmotněprávní otázku, pokud nepatří do skupin A nebo B
- D – věci výkonu rozhodnutí a ostatní rozhodnutí řešící procesní problematiku, nepatří-li do skupin A nebo B
- E – u Občanskoprávního a obchodního kolegia jednoduchá procesní rozhodnutí (např. zastavení dovolacího řízení, odmítnutí dovolání, odklad vykonatelnosti). U Trestního kolegia rozhodnutí, jimiž se nerozhoduje ve věci samé (např. rozhodnutí o vazbě, opravné usnesení).
- EU – rozhodnutí, ve kterých Nejvyšší soud aplikoval právo Evropské unie.

Z výše uvedeného je zřejmé, že především rozhodnutí z kategorie A a B budou zpravidla představovat významná rozhodnutí z hlediska této práce.

2.2.2 Vyhledávač

I na webu Nejvyššího soudu je k dispozici vyhledávač, který umožňuje filtraci dostupných rozhodnutí. Tento vyhledávač rovněž obsahuje obsahovou filtraci podle právní oblasti a umožňuje omezit vyhledávání pouze na právní věty rozhodnutí či na rozhodnutí určitých významových kategorií. Na rozdíl od NALUSu však web Nejvyššího soudu neumožňuje vyhledávat podle jednotlivých ustanovení aplikovaných právních předpisů (resp. umožňuje to pouze v rámci fulltextového vyhledávání).

Tento vyhledávač rovněž podporuje fulltextové vyhledávání včetně možnosti použití některých logických či jiných operátorů, jimiž lze precizovat vztah mezi hledanými pojmy. K jejich rozsahu uvádí, že „*můžete například vytvořit hledání dvou slov, která se vyskytují ve stejném dokumentu, a určit jejich*

³⁵Nejvyšší soud. *Nápověda k vyhledávači rozhodnutí Nejvyššího soudu* [online] [cit. 21.6.2021]. Dostupné z: https://www.nsoud.cz/Judikatura/ns_web.nsf/web/Napoveda-Vyhledavanirozhodnuti~Obsah_databaze_rozhodnuti_a_stanovisek_Nejvyssiho_soudu~?openDocument&lng=CZ.

*maximální vzájemnou vzdálenost, pole, ve kterých se musí nacházet, a psaní malých a velkých písmen. Navíc můžete jedno ze slov označit jako důležitější a dokumenty obsahující první slovo budou v seznamu výsledků hledání zobrazeny výše. Pomocí zástupných znaků lze hledat také části slov.*³⁶ V rámci konkrétních příkladů syntaxe se však nápověda omezuje na základní operátory AND, OR a "", které doplňují pouze operátory ? a *, které představují zástupné symboly.

Významným nedostatkem vyhledávače Nejvyššího soudu je, že neumožňuje vyhledávání všech tvarů zadaného slova. V případě, že chce uživatel i takové výsledky najít, je nutné k jejich vyhledání použít zástupné znaky, což však zvyšuje šanci, že budou nalezena i slova podobná, avšak nesouvisející. Význam rozhodnutí je rovněž i zde omezen pouze na 5 kategorií, přičemž v rámci těchto kategorií nelze určit, která rozhodnutí jsou významnější a vyhledávač nijak nehodnotí význam rozhodnutí ve vztahu k zadaným vyhledávacím parametrům.

2.3 Nejvyšší správní soud

Nejvyšší správní soud v současné době rovněž svá rozhodnutí zveřejňuje online, a to na webu www.nssoud.cz/main0Col.aspx?cls=JudikaturaExtendedSearch. Tato databáze obsahuje všechna rozhodnutí Nejvyššího správního vydaná od jeho vzniku (od roku 2003). Rovněž obsahuje velkou část rozhodnutí krajských soudů vydaných ve správněprávních věcech. Stejně jako u NALUSu a webu Nejvyššího soudu lze přitom i u webu Nejvyššího správního soudu konstatovat, že zveřejněná rozhodnutí v zásadě odpovídají anonymizovaným verzím autentického znění, a pro účely vyhledávání významných judikátů jsou tedy zcela dostatečná.

Nejvyšší správní soud však v nejbližší době plánuje přechod na nový vyhledávač, který je v posledních letech usilovně vyvíjen. Přesný okamžik přechodu na nový systém sice byl již několikrát odsunut, nicméně v současné době se jeví vysoce pravděpodobným, že k přechodu dojde nejpozději v průběhu léta roku 2021.

2.3.1 Struktura dat

Co se formy týče, jsou rozhodnutí v databázi Nejvyššího správního soudu dostupná pouze ve formátu pdf. Struktura těchto souborů je přitom natolik netradiční, že ani osvědčené nástroje pro extrakci textu z pdf dokumentů

³⁶Nejvyšší soud. *Vyhledávání rozhodnutí: Zadávání dotazů pro pokročilé* [online] [cit. 21.6.2021]. Dostupné z: https://www.nssoud.cz/Judikatura/ns_web.nsf/web/Napoveda~Vyhledavanirozhodnuti~Vyhledavani_rozhodnuti__Zadavani_dotazu_pro_pokrocile~?openDocument&lng=CZ.

(jako např. PDF Miner³⁷) si s ní nebyly schopny poradit. Získávání textu rozhodnutí ve strojově zpracovatelné podobě tak není zdaleka jednoduchým úkolem. A s ohledem na skutečnost, že připravovaný nový vyhledávač bude umožňovat i konverzi do formátu html a do plaintextu, lze úspěšně pochybovat o smysluplnosti této snahy v současné době.

Web Nejvyššího správního soudu přitom k rozhodnutím neposkytuje ani téměř žádná zajímavá metadata. Kromě klasické *spisové značky a typu rozhodnutí* obsahuje už jen pouze případnou *právní větu*. Z hlediska této práce nejzajímavější informací, kterou lze ve vyhledávači Nejvyššího správního soudu k jednotlivým rozhodnutím dohledat, je tzv. *prejudikatura* – tedy seznam odkazů na dřívější rozhodnutí, z nichž dané rozhodnutí vychází. Způsob vkládání nových rozhodnutí do databáze přitom vyžaduje, aby každé nově vkládané rozhodnutí ve věci samé bylo buď opatřeno právní větou, nebo aby odkazovalo alespoň na jedno rozhodnutí opatřené právní větou, z něž vychází. Lze tedy rozumně předpokládat, že v seznamu prejudikatury budou obsažena spíše významnější rozhodnutí.

2.3.2 Vyhledávač

I Nejvyšší správní soud na svém webu nabízí vyhledávač, který umožňuje filtraci dostupných rozhodnutí. I tento vyhledávač obsahuje obsahovou filtraci podle právní oblasti a dotčených právních předpisů, rovněž umožňuje omezit vyhledávání pouze na právní věty rozhodnutí. Co se významu rozhodnutí týče, umí však tento vyhledávač rozlišit pouze mezi formami rozhodnutí a mezi rozhodnutími rozšířeného senátu.

Tento vyhledávač rovněž podporuje fulltextové vyhledávání včetně možnosti použití některých logických či jiných operátorů, jimiž lze precizovat vztah mezi hledanými pojmy. Činí tak však pouze ve velmi omezeném rozsahu. Konkrétně umožňuje pouze užití operátorů AND, OR a "", přičemž ani neumožňuje jejich kombinování či řetězení. Vyhledávač rovněž neumožňuje vyhledávání všech tvarů zadaného slova a na rozdíl od vyhledávače na webu Nejvyššího soudu nenabízí ani možnost, jak tento nedostatek obejít (s výjimkou spojení všech mluvnických tvarů operátorem OR). Vyhledávač rovněž nijak neposuzuje význam nalezených rozhodnutí, natož aby jej hodnotil ve vztahu k zadaným vyhledávacím parametrům.

2.3.3 Nový vyhledávač

Jak již bylo uvedeno, Nejvyšší správní soud připravuje v nejbližší době spuštění nového vyhledávače. Na jeho vývoji se podílím jako jeden z prostředníků

³⁷SHINYAMA Y. *PDFMiner: Python PDF parser and analyzer dokumentace* [online] [cit. 21.6.2021]. Dostupné z: <https://www.unixuser.org/~euske/python/pdfminer/index.html>.

mezi zadavatelem (Nejvyšším správním soudem) a zpracovatelem, a jsem tedy informován o jeho plánované funkcionalitě.

Ani u nového vyhledávače se sice neplánuje hodnocení významu nalezených rozhodnutí, nicméně jeho spuštěním by mělo dojít k odstranění mnoha výše uvedených nedostatků. Zejména se jedná o zpřístupnění rozhodnutí ve strojově snadno zpracovatelných formátech a o rozšíření sady metadat, která budou spolu s rozhodnutími zobrazována – konkrétně bude například přidán seznam dotčených právních předpisů a u řízení o opravných prostředcích budou prodány odkazy na přezkoumávaná rozhodnutí krajských soudů. Z pohledu této práce by tedy nový vyhledávač měl být mnohem lepším a snadněji přístupným zdrojem dat než ten současný.

Rovněž je v novém vyhledávací plánováno zlepšení kvality fulltextového vyhledávání – a to výrazným rozšířením podporovaných operátorů i umožněním vyhledávání všech gramatických forem zadaných slov. Dojde též ke značnému rozšíření rozsahu položek, v nichž bude možné vyhledávat. Uživatelé budou mít také možnost vytvořit si vlastní vyhledávací profily, v jejichž rámci si mohou své vyhledávání personalizovat.

2.4 Nižší soudy

Rozhodnutí nižších soudů dosud nebyla nikdy na internetu uceleně zveřejňována. Tuto praxi má změnit nedávno přijatá novela zákona o soudech a soudcích³⁸, která s účinností od 1. 7. 2022 zakotvuje povinnost okresních, krajských a vrchních soudů zveřejňovat v databázi spravované Ministerstvem spravedlnosti pravomocná rozhodnutí.

I v rámci této nové povinnosti se však počítá s tím, že soudy nebudou muset zveřejňovat všechna svá rozhodnutí, když stanoví, že povinnost zveřejňovat se bude vztahovat pouze na ta rozhodnutí, „*jejichž kategorie, postup při jejich zveřejnění a rozsah, ve kterém se tato rozhodnutí zveřejní, stanoví ministerstvo vyhláškou.*“ Za předpokladu, že však tato vyhláška nebude nastavena úplně nesmyslně, lze však pravděpodobně očekávat, že budou zveřejňována přinejmenším ta rozhodnutí, která lze považovat za významná ve smyslu zaměření této práce.

I v současné době jsou některá rozhodnutí nižších soudů zveřejňována, a to prostřednictvím webu Nejvyššího soudu. Rozsah těchto rozhodnutí je

³⁸Zákon č. 218/2021 Sb., kterým se mění zákon č. 6/2002 Sb., o soudech, soudcích, přísedících a státní správě soudů a o změně některých dalších zákonů (zákon o soudech a soudcích), ve znění pozdějších předpisů, a další související zákony

však zjevně nedostatečný, neboť u krajských soudů jsou jich zveřejněny zpravidla jen jednotky a obou vrchních soudů dohromady jen necelých 200.

2.5 Databáze judikatury

Kromě databází a vyhledávačů provozovaných samotnými vrcholnými soudy existují na českém internetu i soukromé projekty, které veřejnosti poskytují přístup k rozhodnutím soudů v elektronické podobě. Většina z nich agreguje rozhodnutí z různých soudů a některé z nich umožňují i kvalitní vyhledávání ve svých databázích. V následující sekci proto představím některé z těchto projektů.

2.5.1 Salvia Kraken

Projekt Salvia Kraken – dostupný na webu kraken.slv.cz – je soukromou iniciativou Mgr. Milana Kvasnici. Vyhledávač je volně přístupný a funguje na celkem jednoduchém principu – rozhodnutí vrcholných soudů získaná z jejich webů zobrazuje na statických URL adresách a umožňuje jejich indexaci Googlem. Hlavní strana vyhledávače se pak sestává pouze z jednoho vyhledávacího pole, které iniciuje prohledání souboru všech těchto rozhodnutí Googlem.

Výhodou tohoto postupu nepochybně je, že vyhledávač tak může těžit ze všech kvalit, kterými Google jako vyhledávač oplývá. Zřejmou nevýhodou pak je skutečnost, že Google neumí posoudit významnost rozhodnutí z právního hlediska, a tento aspekt tedy při řazení výsledků není schopen řádně zohlednit. Samotný text rozhodnutí je přitom na tomto webu doplněn o seznam rozhodnutí, na která dané rozhodnutí odkazuje, seznam (některých) rozhodnutí, ze kterých na něj naopak odkaz vede a o seznam dotčených právních předpisů.

Kromě způsobu realizace vyhledávání je však projekt Salvia Kraken zajímavý i dostupností prohledávaných judikátů. Z hlavní stránky webu totiž přímo vede odkaz do adresářové struktury, v níž jsou uloženy texty jednotlivých rozhodnutí. Rozhodnutí jsou přitom uložena ve složkách podle jednotlivých soudů s omezením, že v jedné složce se nalézá nanejvýš zhruba 10 000 dokumentů. Z těchto složek lze texty rozhodnutí snadno stáhnout a zpracovat. Projekt Salvia Kraken tak může představovat alternativu pro získávání textů rozhodnutí, a to zejména k těm soudům, které neumožňují snadné získání rozhodnutí ve strojově dobře zpracovatelném formátu.

2.5.2 Placené databáze

Kromě výše popsané veřejně přístupné databáze (a databází jednotlivých soudů) existuje rovněž několik placených databází judikatury – např. zakonyprolidi.cz/judikaty, judikaty.info či iudictum.cz. Své vlastní databáze judikatury mají též hlavní placené právní informační systémy ASPI.cz a codexis.cz. Tyto placené databáze zpravidla poskytují pokročilejší vyhledávací nástroje než ty zdarma přístupné. A byť se ne vždy vyrovnají databázím provozovaným vrcholnými soudy (zejména NALUSu), jejich nespornou výhodou oproti těmto databázím je, že obsahují naprostou většinu významných rozhodnutí bez ohledu na to, který soud je vydal.

Zejména databáze právnických informačních systémů rovněž doplňují k významnějším rozhodnutím další informace, např. shrnutí obsahu rozhodnutí či vlastní právní věty vytvořené správci těchto informačních systémů. A přestože kvalita a přesnost těchto dodatečných informací není vždy na nejvyšší úrovni, umožňují rychlejší a snadnější orientaci v daných rozhodnutích.

Co se vyhledávání týče, zaměřují se především na umožnění co nejširší možnosti přizpůsobení si vyhledávacích filtrů tak, aby si uživatel mohl co nejlépe specifikovat oblast právní úpravy, k níž chce vyhledávat relevantní rozhodnutí. Hodnocení významu rozhodnutí však tyto systémy provádějí především staticky na základě jeho základních charakteristik (tedy např. jaký soud jej vydal, v jaké formě, zda je opatřeno právní větou, etc.), tedy bez přihlídnutí ke specifikaci dotazu. Některé systémy se pokoušejí i o hodnocení relevance výsledků ve vztahu k dotazu, nicméně toto hodnocení se vztahuje zejména k fulltextovému vyhledávání.

O pokročilejší metodu hodnocení významu rozhodnutí se pokouší systém ASPI. Ten řadí nalezené výsledky na základě „13 kritérií, např. počet citací, datum vydání, autor a další“³⁹. Podrobnosti fungování tohoto hodnocení však nejsou známy, resp. jsou obchodním tajemstvím společnosti Wolters Kluwer, jež systém ASPI provozuje.

³⁹Wolters Kluwer. *Tipy pro práci s ASPI* [online] [cit. 21.6.2021]. Dostupné z: https://static.wolterskluwer.cz/content/mediagallery/wkshop_system/file/product/file/18028.pdf.

Získávání a analýza dat

Na základě informací z předchozích kapitol tedy již víme, jaká data jsou pro tuto práci potřeba a kde je lze najít. Tato kapitola bude zaměřena na samotný postup získávání dat a jejich převod do formátu použitelného pro další zpracování.

3.1 Ústavní soud

Rozhodnutí Ústavního soudu jsou dostupná na webu nalus.usoud.cz. Tento web však neumožňuje jednorázově stáhnout všechna na něm obsažená rozhodnutí. Vzhledem k tomu, že však web obsahuje použitelný vyhledávač, je možné z něj rozhodnutí získat pomocí *scrapingu*⁴⁰ (tedy strojového zpracování obsahu webových stránek). Při zkoumání funkcionality vyhledávače jsem zjistil, že neumožňuje zadání prázdného dotazu, který by vrátil všechna rozhodnutí. Nicméně jelikož systém nelimituje počet výsledků, lze prázdný dotaz snadno nahradit požadavkem na vyhledání všech rozhodnutí vydaných po 1. 1. 1993.

I přes existenci této možnosti jsem se však rozhodl místo použití jednoho dotazu rozdělit *scraping* na dotazy po jednotlivých měsících. Vedla mě k tomu zejména snaha o jisté zpřehlednění tohoto procesu a možnost jeho udržitelnosti do budoucna. V případě spuštění vyhledávače vycházejícího z myšlenek obsažených v této práci totiž bude nutné pravidelně doplňovat databázi judikatury o nová rozhodnutí. Měsíční frekvence se přitom jeví jako rozumný kompromis mezi požadavkem na aktuálnost databáze a zatížením výpočetních prostředků v důsledku aktualizace dat.

Veden těmito úvahami jsem začal podrobněji zkoumat proces pokládání

⁴⁰Zyte team. *What is web scraping?* [online] [cit. 24. 6. 2021]. Dostupné z: <https://www.zyte.com/learn/what-is-web-scraping/>.

dotazů přes vyhledávač NALUSu. Zjistil jsem přitom, že parametry dotazu je nutné posílat `http POST` requestem. Přepínání na další stránky výsledků je pak možné pouze v rámci téže `http session`. Karta výsledku každého jednotlivého rozhodnutí pak obsahuje text rozhodnutí a většinu metadat. Případná právní věta a abstrakt rozhodnutí však nejsou součástí karty výsledku, nýbrž je nutné je získat samostatným `GET` requestem na adrese `nalus.usoud.cz/Search/Abstrakt.aspx?id=[id_rozhodnuti]`, kde `id_rozhodnuti` je unikátní číselný identifikátor rozhodnutí zjištělý z karty rozhodnutí.

3.1.1 Implementace

Získávání dat z databáze NALUS zajišťuje modul `src/website/daemons/usoud.py`, a to prostřednictvím třídy `UstavniSoudGetter` a její hlavní metody `get_all_verdicts`. Konstruktor této třídy mimo jiné definuje všechny parametry, které je nutné na různých místech tohoto procesu uvádět v requestech, aby je systém NALUS řádně zpracoval. Jedná se zejména o následující:

```
self.base_data = {"nalezy":"on", "usneseni":"on",
"stanoviska_plena":"on", "naveti":"on", "vyrok":"on",
"oduvodneni":"on", "odlisne_stanovisko":"on", "text":"",
"ecli":"", "citace":"", "popularni_nazev":"", "typ_rizeni":"",
"decidedFrom":"","decidedTo":"","publicationFrom":"","
"publicationTo":"","submissionFrom":"","submissionTo":"","
"availableFrom":"","availableTo":"","soudce_zpravodaj":"","
"soudce_stanovisko":"","navrhovatel":"","
"affected_organ_type":"","affected_organ_spec":"","
"actkind":"","actkindnumber_txt":"","actkindname_txt":"","
"actkindclause_txt":"","vyrok_multi":"","
"vztah_k_predpisum":"","predmet_rizeni":"","
"klicove_slovo":"","poznamka":"","but_search": "Vyhledat",
"razeni":"2", "resultsPageSize":"80", "resultsFontSize":"10"}
self.page_tags = ["__EVENTVALIDATION", "__VIEWSTATEGENERATOR",
 "__VIEWSTATE", "__EVENTARGUMENT", "__EVENTTARGET"]
self.verdict_tags = ["__VIEWSTATE", "__VIEWSTATEGENERATOR",
 "__EVENTVALIDATION", "positionHidden", "docIdHidden",
 "sourceLinkTypeHidden", "rcHeightHidden", "contentHeightHidden"]}
```

Funkce `get_all_verdicts` je tvořena pouze cyklem, který opakovaně volá funkci `get_monthly_verdicts`, a to za období od zvoleného data do současnosti. Smyslem této funkce je tedy pouze jednorázové získání historické judikatury. Pro budoucí aktualizaci obsahu bude postačovat pravidelné měsíční volání funkce `get_monthly_verdicts`.

Funkce `get_monthly_verdicts` pak řeší samotný *scraping*. Ten

probíhá tak, že funkce nejprve přistoupí k úvodní stránce vyhledávače nalus.usoud.cz/Search/Search.aspx, z níž získá hodnoty čtyř parametrů uvedených v atributu `self.page_tags` a tyto hodnoty překopíruje do připravovaného POST requestu. Tyto hodnoty představují jakési session ID a bez nich by request nevrátil požadované výsledky. Následně funkce provede tento POST request, čímž si na serveru vyžádá vygenerování seznamu výsledků. Následným GET requestem (v rámci stejné session) na stránku nalus.usoud.cz/Search/Results.aspx pak k těmto výsledkům získá přístup.

Po získání seznamu výsledků si funkce zjistí počet výsledků a následně je začne zpracovávat. Zpracovávání jednotlivých výsledků zajišťuje funkce `process_single_verdict`, která postupuje tak, že nejprve zavolá POST request, kterým si vyžádá další výsledek v pořadí. Získaný výsledek (jako html soubor) nejprve uloží do zadané složky, následně k němu prostřednictvím funkce `get_abstract` zajistí případný abstrakt a právní větu a všechny takto získané informace následně rovněž zapíše do CSV souboru, který agreguje data o rozhodnutí za jednotlivé měsíce. Ke každému rozhodnutí mohou být v tomto CSV souboru obsaženy následující údaje: *Identifikátor evropské judikatury, Název soudu, Spisová značka, Paralelní citace (Sbírka zákonů), Paralelní citace (Sbírka nálezů a usnesení), Populární název, Datum rozhodnutí, Datum vyhlášení, Datum podání, Datum zpřístupnění, Forma rozhodnutí, Typ řízení, Význam, Navrhovatel, Dotčený orgán, Soudce zpravodaj, Napadený akt, Typ výroku, Dotčené ústavní zákony a mezinárodní smlouvy, Ostatní dotčené předpisy, Odlišné stanovisko, Předmět řízení, Věcný rejstřík, Jazyk rozhodnutí, Poznámka, URL adresa, Text Rozhodnutí, Právní věta a Abstrakt.*

Po zpracování výsledku se přesune na další v pořadí. Samotnou extrakci dat z html souborů zajišťuje modul `BeautifulSoup`⁴¹.

Všechny html soubory obsahující jednotlivé judikáty i csv soubory obsahující agregovaná data za jednotlivé měsíce jsou průběžně ukládány do složky, která byla uvedena jako parametr konstruktoru třídy `UstavniSoudGetter`. Smyslem samostatného ukládání html souborů je zajistit, aby v případě zjištění, že při extrakci dat z html došlo k nějakému pochybení, nebylo nutné znovu provádět *scraping* textů rozhodnutí, a zbytečně tak zatěžovat web Ústavního soudu.

3.2 Nejvyšší soud

Rozhodnutí Nejvyššího soudu jsou dostupná na webu nsoud.cz/Judikatura/ns_web.nsf/webSpreadSearch. Ani tento web však neumožňuje jednorázově

⁴¹Crummy.com. *Beautiful Soup Documentation*. Dostupné z: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.

stáhnout všechna na něm obsažená rozhodnutí, a rozhodnutí je tedy nutné získat pomocí *scrapingu* prostřednictvím vyhledávače. Při zkoumání funkcionality vyhledávače jsem zjistil, že nejen že neumožňuje zadání prázdného dotazu, který by vrátil všechna rozhodnutí, nýbrž rovněž omezuje maximální počet výsledků na 900 – tedy, pokud by dotaz měl vrátit více než 900 výsledků, vrátí jich pouze 900.

Toto omezení tak rovněž svědčí pro rozdělení dotazů na jednotlivé měsíce. Počet rozhodnutí vydaných v jednom měsíci se pohybuje kolem 500, takže šance, že některý měsíc bude překročena hranice 900, zatím není příliš pravděpodobná. I tak je však nutné být si této možnosti vědom a po dokončení *scrapingu* ověřit, zda náhodou v nějakém měsíci opravdu počet vydaných rozhodnutí nedosáhl této hranice.

Veden těmito úvahami jsem začal podrobněji zkoumat proces pokládání dotazů přes vyhledávač Nejvyššího soudu. Zjistil jsem přitom, že tento vyhledávač umožňuje zadávat dotazy prostřednictvím `http GET` requestu, a to bez nutnosti předchozího získávání session ID či jiné identifikace. URL adresa dotazu týkajícího se ledna 2004 vypadá následovně:

```
https://www.nsoud.cz/Judikatura/judikatura\_ns.nsf/\$\$WebSearch2?SearchView&Query=%5Bdatum\_rozhodnuti%5D%3E%3D01%2F01%2F2004%20AND%20%5Bdatum\_rozhodnuti%5D%3C%3D01%2F02%2F2004%20AND%20%5BSoudCreate%5D%3D%22Nejvy%C5%A1%C5%A1%C3%AD%20soud%22&Start=0&Count=50&SearchOrder=4&SearchMax=0&lng=
```

Jak je vidět, časové rozmezí je v dotazu vyjádřeno parametrem `datum_rozhodnuti`, které je v adrese vždy následováno datem ve formátu DD/MM/RRRR, přičemž jednou je tomu tak s operátorem `>=` a jednou `<=`. Pro automatizaci dotazů tedy stačí v adrese vždy pouze změnit tyto dva parametry.

Kliknutím na další stránku jsem rovněž zjistil, že adresa následujících stránek se od první stránky výsledků liší pouze hodnotou parametru `Start`. Pro *scrapování* dalších stránek tedy stačí zjistit počet výsledků a následně v cyklu postupně zvyšovat parametr `Start` po krocích o velikosti odpovídající počtu záznamů na stránce – tedy o 50.

3.2.1 Implementace

Získávání dat z databáze Nejvyššího soudu zajišťuje modul `src/website/daemons/nsoud.py`, a to prostřednictvím třídy `NSGetter` a její hlavní metody `get_all_verdicts`. Funkce `get_all_verdicts` je opět tvořena pouze cyklem, který opakovaně volá funkci `get_monthly_verdicts`, a to za období od zvoleného data do současnosti. Smyslem této funkce

je tedy opět pouze jednorázové získání historické judikatury. Pro budoucí aktualizaci obsahu bude postačovat pravidelné měsíční volání funkce `get_monthly_verdicts` (za předpokladu, že měsíční přírůstky nepřekročí hranici 900 – v takovém případě bude nutné zkrátit tento interval).

Funkce `get_monthly_verdicts` pak řeší samotný *scraping*. Ten probíhá tak, že funkce nejprve výše popsaným způsobem upraví URL adresu a načte ji. Z takto získané stránky zjistí počet výsledků a následně začne prostřednictvím funkce `process_page` zpracovávat jednotlivé stránky. Tato funkce přitom funguje tak, že na stránce najde (jedinou) tabulku obsahující odkazy na výsledky, a z ní získá všechny hypertextové odkazy. Z těchto odkazů následně vybere jen ty, které obsahují řetězec `WebSearch`, který svědčí o tom, že se jedná o dokument, který je výsledkem vyhledávání (a nikoliv o odkaz na tisk dokumentu, či na jeho verzi ve formátu doc).

Po získání seznamu výsledků je funkce začne po jednom zpracovávat prostřednictvím volání funkce `process_single_verdict`. Vzhledem k tomu, že v případě Nejvyššího soudu jsou všechny informace o rozhodnutí k dispozici na téže stránce, není nutné získávat jinde další metadata a ta následně skládat do CSVčka. Funkce `process_single_verdict` tedy pouze jednoduše načte stránku obsahující informace o rozhodnutí a uloží ji do složky, která byla uvedena jako parametr konstruktoru třídy `NSGetter`.

Poznámka: Po stažení všech rozhodnutí bylo zjištěno, že v žádném měsíci počet rozhodnutí hranice 900 nedosáhl. Nejblíže jí byl v říjnu roku 2017, kdy bylo vydáno 844 rozhodnutí. V ostatních měsících se však počet rozhodnutí pohybuje řádově o stovky pod touto hranicí.

3.3 Nejvyšší správní soud

Rozhodnutí Nejvyššího správního soudu jsou sice dostupná na webu nssoud.cz/main0Col.aspx?cls=JudikaturaExtendedSearch, nicméně narozdíl od předchozích soudů trpí významnými nedostatky. Zásadním problémem je především to, že rozhodnutí jsou dostupná pouze ve formátu pdf, který není z hlediska stojového zpracování zrovna přívětivý. Případnou snahu o překonání této překážky přitom rovněž hatí skutečnost, že současný vyhledávač na stránkách Nejvyššího správního soudu bude v nejbližší době nahrazen vyhledávačem novým, který již bude umožňovat zobrazení rozhodnutí v plaintextu, či ve formátu html.

S ohledem na tyto skutečnosti, jsem se rozhodl jako hlavní zdroj rozhodnutí Nejvyššího správního soudu využít nikoliv přímo web tohoto soudu, nýbrž databázi veřejně přístupného online vyhledávače Salvia Kraken – kraken.slv.cz.

Databáze na těchto stránkách je organizována tak, že k získání rozhodnutí není nutné vůbec používat vyhledávač. Jednotlivá rozhodnutí jsou totiž uložena ve složkách dostupných ze stránky <http://kraken.slv.cz/sitemap>. Manuálním prozkoumáním těchto složek jsem zjistil, že rozhodnutí Nejvyššího správního soudu se nacházejí ve složkách 13 až 20.

Jednotlivá rozhodnutí jsou v těchto složkách dostupná ve formátu html, který je v části obsahující samotný text rozhodnutí opět spíše plaintextem. Jeho převedení do využitelné podoby tedy nebude náročné.

3.3.1 Implementace

Získávání dat z databáze Salvia Kraken zajišťuje modul [website/daemons/nssoud.py](#), a to prostřednictvím třídy `NSSGetter` a její hlavní metody `process_kraken`. Samotný *scraping* je v podstatě triviální. V rámci funkce `process_kraken` dochází k načtení odkazů na rozhodnutí z nadřazených složek. Jednotlivá rozhodnutí jsou následně zpracovávána funkcí `get_from_kraken`, která jednotlivé rozhodnutí načte a uloží jej ve formátu html do složky, která byla uvedena jako parametr konstruktoru třídy `NSSGetter`.

3.4 Problémy v datech

Vzhledem k tomu, že sada získaných rozhodnutí je velmi rozsáhlá a zahrnuje judikaturu za období téměř 30 let, není příliš velkým překvapením, že obsahuje mnohé chyby či nekonzistentnosti. Některé z nich jsou s ohledem na zaměření této práce bezvýznamné, některé je však třeba řádně zohlednit a případně i ošetřit. V této části některé z těchto problémů popíši, a to včetně zhodnocení jejich významu a návrhu případného řešení.

3.4.1 Stará rozhodnutí

Mnohá rozhodnutí Nejvyššího a Ústavního soudu pocházejí z doby, kdy český internet byl ještě v plenkách a nikdo nepočítal s tím, že jednou budou všechna tato rozhodnutí zveřejněna. Nástroje pro podporu psaní rozhodnutí (na počítači) se proto v té době v podstatě nijak nezabývaly případnou strojovou zpracovatelností rozhodnutí, nýbrž se zaměřovaly primárně na to, jak budou rozhodnutí vypadat po vytištění. U takovýchto rozhodnutí se lze tedy zcela běžně setkat např. s tím, že každý řádek je ukončen odřádkováním, zatímco u současných textů je naopak běžné, že odřádkováním jsou ukončovány odstavce, nikoliv jednotlivé řádky.

Takovéto formátování přitom může výrazně narušit vazby mezi jednotlivými slovy a zkomplikovat tak fulltextové vyhledávání (např. při použití

operátoru požadujícího přítomnost hledaných slov ve stejném odstavci). Významné problémy působí i při extrakci konkrétních významových entit z textu rozhodnutí – např. odkazů na jiná rozhodnutí či právní předpisy. Stejně tak komplikuje i analýzu členění rozhodnutí (tedy identifikaci hranic mezi větám, výrokem, odůvodněním a poučením).

Řešením některých z výše uvedených nedostatků by mohlo být sloučení veškerého textu rozhodnutí do jednoho odstavce (a nahrazení odřádkování mezerami) – tím by se odstranil problém rozdělených významových entit. Ani takovýto postup by se však nijak pozitivně nepodepsal na omezení fulltextového vyhledávání na stejný odstavec (spíše naopak – vzhledem k tomu, že celý text by byl jedním odstavcem, byl by tento operátor zbytečný), a analýzu členění rozhodnutí na samostatné významové celky by ještě více zkomplikoval.

Vzhledem k tomu, že pro tuto práci je však identifikace odkazů na právní předpisy a zejména jiná rozhodnutí významnější než kvalita fulltextového vyhledávání či členění rozhodnutí, lze mít za to, že u těchto rozhodnutí by bylo i přesto vhodnější sloučení provést.

Dalším problémem je však identifikace těchto rozhodnutí. Bez podrobnější analýzy jejich obsahu totiž nelze určit, zda jsou odstavce ukončovány předčasně. U rozhodnutí, u nichž je každý řádek ukončen odřádkováním by sice bylo možné toto odhalit – ať již prostřednictvím analýzy délky odstavce či gramatické správnosti a větné skladby (z nichž by vyplynulo, že mnoho odřádkování je vloženo uprostřed vět), ale problémová rozhodnutí nejsou vždy takto šablonovitá. U spousty z nich je chybně vloženo odřádkování pouze na několika málo místech (nikoliv na každém řádku), a snaha o jeho nápravu sloučením všech odstavců by tak působila více problémů než užitku.

S ohledem na komplikace spojené s nápravou těchto problémů je možné rovněž přistoupit k radikálnějšímu řešení – a to k vynechání všech rozhodnutí vydaných v době před zahájením jejich zveřejňování ve vhodné vnitřní struktuře (tedy zhruba od roku 2004). Tento postup by však měl za následek ztrátu mnoha potenciálně významných rozhodnutí. Vzhledem k tomu, že však smyslem této práce není vytvořit plně funkční úplný vyhledávač judikatury (nýbrž jen zhodnotit, navrhnout a otestovat dílčí metody, které by mohly přispět ke zpřesnění klasifikace významu jednotlivých rozhodnutí), lze pro účely této práce považovat sadu rozhodnutí začínající rokem 2004 za dostatečnou.

Takovýmto omezením množiny dat navíc odpadne (či bude alespoň výrazně omezena) řada dalších problémů vyskytujících se zejména u starších rozhodnutí – jako například absence (či jiná struktura) metadat. Po důkladném zvážení jsem tedy nakonec přistoupil k tomu, že jsem množinu rozhodnutí

omezil pouze na ta, která byla vydána nejdříve v roce 2004.

3.4.2 Chybějící rozhodnutí

Omezením množiny rozhodnutí se rovněž výrazně snížila závažnost problému chybějících rozhodnutí. Pokud totiž nějaká rozhodnutí ve vyhledávacích vrcholných soudů chybí, jedná se především o rozhodnutí z dob před spuštěním těchto vyhledávačů. Po přijetí závazku zveřejňovat svá rozhodnutí totiž vrcholné soudy upravily své vnitřní procesy tak, že zveřejnění rozhodnutí je běžnou součástí agendy prováděné po jeho vydání. Tyto vnitřní procesy tak zaručují, že (až na případná ojedinělá pochybení) dochází ke zveřejnění všech rozhodnutí.

Naproti tomu rozhodnutí z doby před zahájením procesu jejich zveřejňování byla do databází rozhodnutí nainportována až dodatečně. U mnohých z těchto rozhodnutí se přitom s jejich zveřejněním nikdy nepočítalo, a tak u spousty z nich neexistovaly anonymizované verze (ty se totiž v minulosti tvořily pouze *ad hoc* u těch rozhodnutí, o jejichž poskytnutí někdo konkrétně požádal např. postupem podle zákona o svobodném přístupu k informacím⁴²). A vzhledem k tomu, že zveřejňování neanonymizovaných rozhodnutí není s ohledem na ochranu osobních údajů možné, mnoho starších rozhodnutí nikdy zveřejněno nebylo.

⁴²Zákon č. 106/1999 Sb., o svobodném přístupu k informacím.

Klasifikace rozhodnutí

Hlavním smyslem této práce je navrhnout metody, které by umožnily kvalitně rozlišovat mezi významnými a méně významnými rozhodnutími, a tím přispět k zefektivnění vyhledávání v judikatuře. V této kapitole tedy představím některé z možných přístupů ke klasifikaci rozhodnutí a vhodné metody.

Na tomto místě je přitom vhodné upozornit, že výraz „klasifikace“ není v této práci užíván ve smyslu, v jakém je v oblasti znalostního inženýrství obecně chápán – tedy jako úloha či proces, jejichž podstatou je predikování diskrétně rozdělené veličiny⁴³. Primární úlohou sice v této práci je rozdělit soudní rozhodnutí do několika kategorií podle významu, nicméně hranice mezi těmito kategoriemi nejsou jasně dané. V některých případech by tedy daný výraz odpovídal své obecné definici (například při dělení rozhodnutí na významná a méně významná), někdy by však poněkud pokulhával – například při řazení rozhodnutí podle jejich významu.

Vzhledem k tomu, že mezi popsány příklady není z hlediska této práce zásadnější rozdíl, zdá se jako zcela zbytečné v zájmu zachování formální čistoty užívaných pojmů zavádět nové výrazy pro části úkolů, jimž se tato práce věnuje. Tím spíše, že hranice mezi nimi není úplně zřejmá. Pro účely této práce se tedy klasifikací (zpravidla podle významu) myslí jak třídění do větších celků s obdobnými vlastnostmi, tak vzájemné srovnávání rozhodnutí, tak i přidělování skóre (vyjadřujícího jejich význam) rozhodnutím.

4.1 Statická klasifikace

Nejzákladnější metodou hodnocení významu je statické hodnocení vycházející z některých aspektů rozhodnutí samotného. Tuto metodu používají Nejvyšší

⁴³Sidath Asiri. *Machine Learning Classifiers* [online] [cit. 26.6.2021]. Dostupné z: <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623>.

soud, Ústavní soud, i některé soukromé databáze. Tato metoda v podstatě přiřadí každému rozhodnutí číselnou váhu, která umožňuje jeho porovnání s ostatními rozhodnutími. Nejvyšší soud i Ústavní soud používají pětistupňovou škálu, přičemž váhu jednotlivých rozhodnutí stanovuje ten, kdo je do databáze vkládá (resp. ten, kdo k tomu dal pokyn), což je zpravidla soudce, který v dané věci rozhodoval (či jeho asistent).

Vzhledem k rozsahu dostupných metadat, jimiž jsou rozhodnutí opatřena, však není nutné omezovat se při klasifikaci na jediný parametr. Rozhodně totiž neplatí, že rozhodnutí lze rozdělit do pěti skupin tak, aby v rámci jedné skupiny bylo možno rozhodnutí považovat za rovnocenná. Jemnější rozdělení přitom mohou být schopny zajistit další parametry – například údaj o tom, který soud rozhodnutí vydal, zda je opatřeno právní větou, jakou formou bylo vydáno, etc. Výsledná váha rozhodnutí (reprezentující jeho význam) by tedy byla váženým průměrem několika relevantních příznaků. Klíčovým tak bude nalezení nejhodnějších vah pro parametry.

4.2 Klasifikace na základě vztahů

Přestože staticky vypočtená váha umožňuje posoudit význam rozhodnutí, činí tak pouze na základě velmi omezených informací. Zatímco u některých rozhodnutí je již při jejich vydání jasné, že se bude jednat o velmi významná či přelomová rozhodnutí, některá rozhodnutí na významu nabývají až časem. Význam rozhodnutí totiž není dán jen tím, kdo jej vydal, ale rovněž tím, zda jej ostatní soudy ve své následné rozhodovací činnosti akceptovaly a ve svých rozhodnutích v obdobných věcech z něj vycházely. Pro posouzení, zda tomu tak bylo, je však nutné znát i obsah jiných rozhodnutí. Pro význam rozhodnutí je tedy zásadní, kolik následných rozhodnutí jej citovalo, a jak významná tato rozhodnutí byla.

4.2.1 Počet citací

Jednou z jednodušších možností, jak při posuzování významu rozhodnutí zohlednit jeho následnou popularitu, je hodnocení na základě prostého počtu citací v následných rozhodnutích, obdobně jako je tomu například i v akademické či výzkumné sféře. Tento přístup nepochybně bude schopen poskytnout některé zajímavé informace – např. díky němu bude možné vytřídit ta rozhodnutí, která nebyla nikdy citována, a (pokud nepocházejí ze zcela nedávné doby) je tedy velmi nepravděpodobné, že se z hlediska jejich významu jedná o rozhodnutí důležitá.

Takový postup však vykazuje stejné nedostatky jako onen pověstný

„*kafemlejnek*“⁴⁴. Tím hlavním samozřejmě je, že nijak nezohledňuje fakt, že ne všechny citace mají stejnou váhu. Vzhledem k tomu, že s problematikou zkoumání významu citací se (v rámci snahy o hodnocení kvality vědců, vědeckých institucí či časopisů) světová věda již delší dobu potýká, existuje v této oblasti řada algoritmů.

Tak například vydavatelství odborné literatury Elsevier v roce 2021 nasadilo novou verzi své metriky k hodnocení významu vědeckých časopisů zvané *CiteScore*⁴⁵, která je obdobou v České republice populárního *Impakt faktoru*⁴⁶. Obě tyto metody hodnotí význam vědeckých časopisů tím, že porovnávají počet citací, které články vydané v daném časopisu zaznamenaly napříč ostatními posuzovanými časopisy, s počtem článků v daném časopisu publikovaných. Zároveň obě omezují období, za které se hodnocení počítá, pouze na několik málo let (v případě *CiteScore* na 4, v případě *Impakt faktoru* na 2).

Z výše uvedeného plyne, že metriky založené na počtu citací, obvykle zohledňují stáří zdroje. Dává tedy smysl takový postup aplikovat i u hodnocení významu soudních rozhodnutí. Pokud totiž významné rozhodnutí zůstane relevantní po mnoho let, je schopné nasbírat rozsáhlé množství citací. To však automaticky neznamená, že je významnější než rozhodnutí novější, které dosud tolik citací nenasbíralo. Naopak, pokud nově vydané rozhodnutí v obdobné právní otázce začne být rozsáhle citováno, bude tomu zpravidla proto, že závěry obsažené ve starším rozhodnutí již nejsou aktuální, nebo přinejmenším nejsou úplné.

Z hlediska významu rozhodnutí by tedy nemusel být relevantní pouze počet citací, nýbrž například i vývoj těchto citací v čase – tedy, zda je dané rozhodnutí v praxi citováno pravidelně a dlouhodobě, či zda svůj objem citací zaznamenalo v rámci krátkého období „popularity“ a následně upadlo v zapomnění. Na základě informací o tom, kým je rozhodnutí citováno, pak lze například snížit význam rozhodnutí, která za významná považují jen jejich autoři (či maximálně úzká skupina soudců v jejich okolí), kteří je proto hojně citují a tím uměle zvyšují jejich význam v klasifikaci na základě metrik založených na citacích.

⁴⁴KONVALINKA, Jan. „*Kafemlejnek*“ na českou vědu aneb k financování vědy. Živa. 2009, 2009(4), XLIX - LXVIII.

⁴⁵Elsevier. *CiteScore Journal Metric - FAQs* [online] [cit. 24.6.2021]. dostupné z: https://service.elsevier.com/app/answers/detail/a_id/30562/supporthub/scopus/

⁴⁶Clarivate. *The Clarivate Analytics Impact Factor* [online] [cit. 24.6.2021]. Dostupné z: <https://clarivate.com/webofsciencelgroup/essays/impact-factor/>.

4.2.2 PageRank

Významným nedostatkem algoritmů založených na počtu citací je, že zpravidla nerozlišují váhu citací podle jejich zdroje. Existují však úspěšné a v praxi hojně používané algoritmy, které tuto schopnost mají – jedním z nich je například *PageRank*⁴⁷ – algoritmus používaný Googlem pro hodnocení relevantnosti výsledků vyhledávání. Funguje tak, že „*spočítá počet a kvalitu odkazů na stránku, aby určil hrubý odhad důležitosti webu. Základním předpokladem je, že důležitější webové stránky pravděpodobně obdrží více odkazů z jiných webových stránek.*“⁴⁸ Přestože algoritmus byl původně navržen k hodnocení významu webových stránek na základě počtu a kvality na ně směřujících odkazů, měl by být jeho princip bez problémů přenositelný i na hodnocení významu soudních rozhodnutí. I ta jsou totiž provázána vzájemnými odkazy.

PageRank samozřejmě není jediným nástrojem, který Google používá pro řazení výsledků vyhledávání (dokonce ani není jisté, zda jej dosud vůbec používá, resp. v jaké podobě jej používá). Pro účely hodnocení relativního významu vzájemně provázaných entit je však značně efektivní, neboť je schopný celkem jednoduše zohlednit nejen počet, ale i kvalitu odkazů, kterými jsou dané entity propojeny.

Základem algoritmu *PageRank* je iterativní přístup. Jeho jednotlivé kroky lze popsat následovně⁴⁹:

1. Nastavení hodnot všech uzlů na 1.
2. Aktualizace hodnot všech uzlů podle *PageRank* vzorce uvedeného na obrázku 4.1.
3. Opakování kroků 2 dokud nedojde k ustálení hodnot uzlů.

V každém kroku algoritmu je tedy hodnota každého uzlu determinována hodnotou uzlů, z nichž do daného uzlu vedou hrany (ve vzorci je to vyjádřeno čitatelem sumy). Každý uzel však svoji váhu *rozděluje* mezi všechny uzly, do kterých z něj vedou hrany (jmenovatel sumy). Nejvyšších hodnot tak dosáhnou uzly, do kterých vedou hrany z významných uzlů. Významným uzlem je přitom takový uzel, do nějž vedou hrany z mnoha uzlů, ale z něhož

⁴⁷Page, Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry. 1999. *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report. Stanford InfoLab.

⁴⁸Google. *Facts about Google and Competition* [online] [cit. 24.6.2021]. Dostupné z: <https://web.archive.org/web/20111104131332/https://www.google.com/competition/howgooglresearchworks.html>.

⁴⁹Chonny. *PageRank: Link Analysis Explanation and Python Implementation from Scratch* [online] [cit. 24.6.2021]. Dostupné z: <https://towardsdatascience.com/pagerank-3c568a7d2332>.

$$PR(P_i) = \frac{(d)}{n} + (1-d) \times \sum_{I_j \in E} PR(P_j) / \text{Outdegree}(P_j)$$

$$D(\text{damping factor}) = 0.1 \sim 0.15$$

$$n = |\text{page set}|$$

Obrázek 4.1: Vzorec používaný pro aktualizaci vah uzlů.

vedou hrany pouze do několika málo uzlů. V případě užití *PageRanku* pro hodnocení významu webových stránek představují uzly samotné stránky, zatímco hrana mezi dvojicí stránek existuje tehdy, pokud z jedné stránky vede hypertextový odkaz na stránku druhou.

Damping factor d ve vzorci pak simuluje chování uživatele. V praxi je totiž zcela běžné, že uživatel ne vždy pokračuje v klikání na odkazy na stránkách dokud tyto existují, nýbrž že po nějaké době přejde na jinou stránku (která s tou předchozí nemusí být nijak propojena). V kontextu tohoto algoritmu jde o variaci náhodné procházky⁵⁰, respektive „teleportaci“. Přejed na novou stránku bez ohledu na existenci odkazů v podstatě odpovídá zadání náhodné URL adresy do internetového prohlížeče uživatelem. Všeobecně je přijímáno, že k tomu dochází zhruba s 15% pravděpodobností⁵¹.

Damping factor tedy redukuje váhu sousedních (rodičovských) uzlů při výpočtu hodnoty zkoumaného uzlu tím, že část váhy nepočítá pouze ze sousedních uzlů, nýbrž ze všech uzlů. Tím také rovněž přispívá k urychlení konvergence skóre jednotlivých uzlů a rychlejšímu ukončení algoritmičského cyklu. V rámci výpočtu algoritmu PageRank totiž často dochází k opakovanému *přelévání* hodnot tam a zpět mezi sousedními uzly. *Damping factor* zajišťuje, že takovéto tendence budou mít s přibývajícím časem stále menší a menší význam.

4.2.2.1 Vážený PageRank

Přestože je *PageRank* schopný zohlednit při výpočtu váhy význam sousedních uzlů, nelze to pro účely této práce považovat za dostatečné. Význam uzlů je totiž opět dán výlučně množstvím hran, které do něj (a jeho předků) směřují. Všechny hrany přitom mají stejnou váhu. Jak však bylo v předchozích kapitolách uvedeno, ne všechny citace v judikátech mají stejnou váhu.

⁵⁰Khan Academy. *Random walks* [online] [cit. 24.6.2021]. Dostupné z: <https://www.khanacademy.org/computing/computer-programming/programming-natural-simulations/programming-randomness/a/random-walks>

⁵¹BRIN, S.; PAGE, L. *The anatomy of a large-scale hypertextual Web search engine*. Computer Networks and ISDN Systems. 30 (1–7): 107–117. doi:10.1016/S0169-7552(98)00110-X. ISSN 0169-7552.

Existují však varianty algoritmu *PageRank*, které tento problém řeší – jedná je o tzv. vážený *PageRank*⁵².

Tato verze představuje rozšíření klasického *PageRanku* v tom, že umožňuje nastavit vlastní váhu pro všechny hrany. Při výpočtu tak není zohledňován pouze počet hran, nýbrž i jejich význam. Varianty tohoto algoritmu rovněž umožňují nastavení výchozích hodnot jednotlivých uzlů, což je schopné výrazně přispět ke zkrácení doby výpočtu.

4.2.2.2 Užití *PageRanku* ke klasifikaci judikatury

PageRank je sice na první pohled efektivní způsob identifikace významných rozhodnutí, je však nutné si uvědomit, že byl vytvářen pro jiné prostředí – tedy pro internetové vyhledávače. Pro dosažení optimálních výsledků při jeho použití je proto nutné vhodně posoudit hlavní rozdíly mezi těmito případy použití, a adekvátně na ně reagovat.

Zásadním rozdílem je zde především absence nutnosti zohledňovat jakéhosi imaginárního webového uživatele, který se může při klikání na navazující odkazy „unavit“ a přejít na náhodnou stránku. Při hodnocení významu judikatury je totiž důležité pouze to, jak jsou mezi sebou jednotlivá rozhodnutí provázána, nikoliv to, zda bude někdo ochotný všechny tyto odkazy následovat. V implementaci zaměřené na hodnocení významu judikatury by tedy nejspíš neměl hrát tak významnou roli damping factor.

Při užití váženého *PageRanku* je dále nutné zamyslet se nad tím, jak ohodnotit jednotlivé hrany (a případně i jaké použít výchozí hodnoty pro uzly). Jak přitom bylo již uvedeno v předchozích kapitolách, existují rozhodnutí, u nichž je pravděpodobné, že odkazují na významná rozhodnutí – zejména se jedná o odmítavá usnesení Nejvyššího správního soudu Nejvyššího soudu v civilních věcech. Odkazům vedoucím z těchto rozhodnutí je tedy vhodné přidělit vyšší váhu. Stejně tak přitom platí, že pokud důležité rozhodnutí odkazuje na jiné rozhodnutí, existuje rozumný předpoklad, že i toto rozhodnutí je důležité. Lze tedy shrnout, že parametry ovlivňující váhu hran se do značné míry podobají parametrům statického hodnocení významu rozhodnutí. Z tohoto ukazatele je proto pro zjednodušení možné vyjít.

Kvalitu klasifikace lze rovněž výrazně vylepšit analýzou samotného textu rozhodnutí. Ne všechny odkazy totiž mají v rámci rozhodnutí stejný význam. Pokud by tedy bylo možné odlišit ty části rozhodnutí, v nichž soud předkládá svoji právní argumentaci, již opírá o odkazy na dřívější rozhodnutí, od těch

⁵²Justin. *Personalized PageRank with Edge Weights* [online] [cit. 24.6.2021]. Dostupné z: <https://www.briggsby.com/personalized-pagerank>

částí, v nichž například jen shrnuje předchozí průběh řízení či argumentaci účastníků, rovněž by to přispělo ke zvýšení kvality klasifikace.

4.2.3 Lokální klasifikace

Všechny výše popsané metody lze označit za globální – tedy takové metody, které pracují s celým souborem rozhodnutí najednou, a jejich výsledek představuje obecnou váhu rozhodnutí. Takový postup však nemusí nutně přinášet nejlepší výsledky. Při vyhledávání není nutné vždy pracovat s celou množinou dat. Zadáním vyhledávacích parametrů totiž zpravidla dojde k výraznému omezení množiny relevantních výsledků. Je pak otázkou, nakolik jsou pro váhu takto vyselektovaných rozhodnutí relevantní jejich vazby na rozhodnutí jiná.

Lze přitom mít obecně za to, že rozhodnutí, které bude hojně citováno napříč množinou nalezených rozhodnutí bude vzhledem k zadanému dotazu významnější, než obecně významné rozhodnutí, které sice splňuje podmínky dotazu, ale jeho význam pochází především z citací mimo rozsah vyhledávání. Při implementaci *PageRanku* je proto potřeba zohlednit i tuto skutečnost. A to buď tak, že citace z rozhodnutí, která nebudou splňovat podmínky dotazu budou mít značně omezenou váhu, či budou zcela ignorována.

Implementace

V předchozích kapitolách bylo vysvětleno jaká data jsou potřeba, kde je získat a jaké metody aplikovat při jejich zpracování. V této kapitole se zaměřím na to, jak výše nastíněné teoretické úvahy realizovat v praxi.

5.1 Prostředí

Než však přistoupím k samotné implementaci klasifikačních metod, považuji za nutné předestřít širší kontext implementace – zejména představit nástroje, které při své implementaci používám a způsob, jakým tak činím.

5.1.1 Django / Python

Jako programovací jazyk jsem zvolil Python⁵³. A to jednak proto, že se jedná o programovací jazyk s významnou popularitou a rozsáhlou podporou pro úlohy z oblasti strojového učení⁵⁴, jednak proto, že s tímto jazykem již mám jisté zkušenosti. Konkrétně používám Python 3, verzi 3.6.1 (byť se domnívám, že aplikace není závislá na žádných knihovnách specifických pro tuto verzi, a měla by být spustitelná i v novějších verzích Pythonu).

Vzhledem k tomu, že původním záměrem projektu (jehož je tato práce součástí) bylo vytvoření vyhledávače judikatury, využívám při implementaci nástroj Django⁵⁵ (konkrétně ve verzi 2.0.2), který podporuje tvorbu webových serverů. Tento nástroj sice není nezbytný pro realizaci úkolů vytyčených v této práci, jeho užití mi však usnadní následnou inkorporaci výsledků této

⁵³Python Software Foundation. *Python documentation* [online] [cit. 24. 6. 2021]. Dostupné z: <https://www.python.org/doc/>

⁵⁴Developer Economics. *What is the best programming language for Machine Learning?* [online] [cit. 24.6.2021]. Dostupné z: <https://towardsdatascience.com/what-is-the-best-programming-language-for-machine-learning-a745c156d6b7?>

⁵⁵Django Software Foundation. *Django documentation* [online] [cit. 24. 6. 2021]. Dostupné z: <https://docs.djangoproject.com/en/3.2/>

práce do širšího projektu.

Vzhledem ke způsobu fungování Django jsou veškeré parametry definující nastavení prostředí obsaženy v souboru `src/website/website/settings.py`. Spuštění samotného serveru se pak provádí voláním scriptu `src/website/manage.py`, a to s parametry `runserver 0.0.0.0:8001` (port 8001 je možné nahradit za libovolný port, na němž má server běžet).

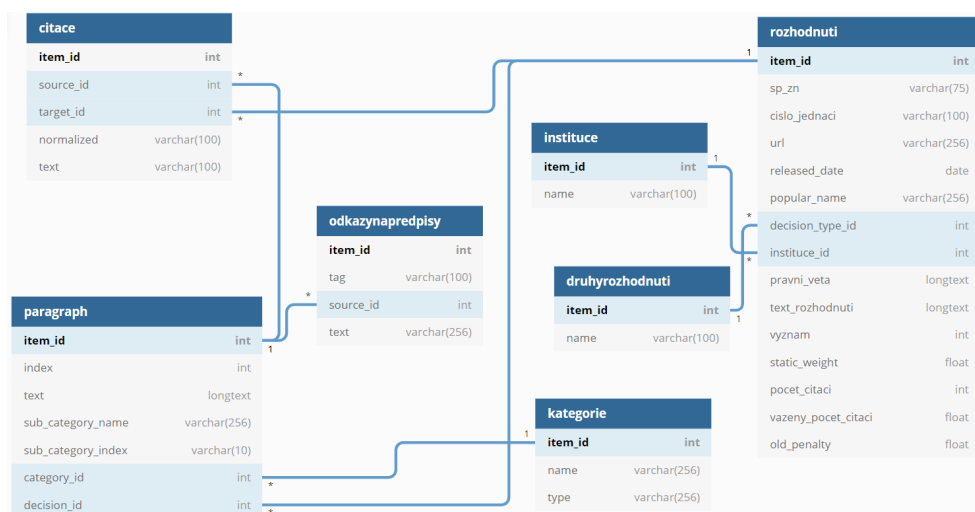
5.1.2 Databáze

Jako databázový server používám MySQL⁵⁶ ve verzi 8.0.22. V rámci této práce však s databází přímo pracuji jen velmi omezeně, neboť její strukturu i komunikaci s ní obstarává Django. Jediné, co jsem tedy musel v databázovém stroji manuálně udělat, bylo vytvořit příslušnou databázi a uživatele, který k ní má všechna práva. Tyto údaje jsem následně zanesl do výše uvedeného souboru `settings.py` do položky `DATABASES`. Jak je z tohoto záznamu zřejmé, databázový server typu MySQL poslouchá na standardním portu 3306, hlavní databáze se jmenuje `judicial_map`, uživatel je `database_commander` a jeho přístupové heslo je `0wbT:~d!nm?X&fJ^~n~e0&8aFKf+j8K&l`. (Všechny tyto parametry je však možné změnit.)

Samotnou strukturu databáze definuje soubor `src/website/main/models.py+`. K vytvoření databáze na základě tohoto modelu lze opět použít script `manage.py`, a to tak, že je nejprve spuštěn s parametrem `makemigrations` a následně s parametrem `migrate`⁵⁷. Po provedení migrací bude v databázovém stroji vytvořena databáze se strukturou popsanou na obrázku^{5.1}.

⁵⁶Oracle. *MySQL Documentation* [online] [cit. 24. 6. 2021]. Dostupné z: <https://dev.mysql.com/doc/>

⁵⁷Django Software Foundation. *Migrations* [online] [cit. 24. 6. 2021]. Dostupné z: <https://docs.djangoproject.com/en/3.2/topics/migrations/>.



Obrázek 5.1: Struktura databáze

Tato struktura zachycuje pouze tabulky definované ve výše uvedeném modulu – jejich názvy jsou v samotné databázi doplněny o předponu `main`. Django kromě těchto tabulek vytváří ještě další tabulky s předponami `django` a `auth`, které jsou však z hlediska této práce irelevantní.

Klíčovou tabulkou z tohoto modelu je tabulka `rozhodnuti`, jež obsahuje informace o jednotlivých rozhodnutích – a to konkrétně jejich úplný text (`text_rozhodnuti`), právní větu (`pravni_veta`), spisovou značku (`sp_zn`), číslo jednací (`cislo_jednaci`), odkaz na web, kde se dané rozhodnutí nachází (`url`), datum vydání rozhodnutí (`released_date`), soudem přidělený význam (`vyznam`), jeho populární název (`popular_name`) a odkaz na soud, který rozhodnutí vydal, (`instituce_id`) a na typ rozhodnutí (`decision_type_id`). Většina těchto polí je přitom vyplněna pouze u některých rozhodnutí v závislosti na tom, jaká metadata o nich byla k dispozici na webu, z něžž byla získána.

Pole `static_weight`, `pocet_citaci`, `vazeny_pocet_citaci` a `old_penalty` jsou zpočátku prázdná. Do těchto polí budou následně jednotlivě implementované ukládat své poznatky o jednotlivých rozhodnutích (např. počet do nich směřujících citací či vlastní odhad jejich váhy).

Tabulka `rozhodnuti` pak přímo odkazuje na dvě další tabulky, které blíže popisují některé aspekty rozhodnutí. Jedná se o tabulky `druhyrozhodnuti` a `instituce`, který obsahují doplňující informace o formě rozhodnutí a instituci (zpravidla soudu), která rozhodnutí vydala. V současné podobě každá z těchto tabulek obsahuje pouze jedno textové pole, nicméně existuje

reálný předpoklad, že v budoucnu může být užitečné používat více informací o těchto položkách. Forma samostatných tabulek tedy byla zvolena zejména s ohledem na předpokládaný budoucí vývoj projektu.

S tabulkou rozhodnutí je následně provázána tabulka `paragraph`, popisující jednotlivé odstavce rozhodnutí. (V současné době je v názvu této tabulky překlep – její název je tedy `paraghraph`. Tento překlep nemá žádný vliv na funkci, nicméně jeho odstranění by si vyžádalo rozsáhlejší zásah do kódu. Vzhledem k tomu, že takový (a ještě o něco širší) zásah stejně bude nutný při inkorporaci výsledků této práce do zbytku projektu, bude tento překlep odstraněn společně s tím.) Tato tabulka obsahuje jednotlivé odstavce ze všech rozhodnutí. Podporuje rovněž možnost klasifikace odstavců do různých kategorií (prostřednictvím propojení s tabulkou `kategorie`), což může nadále zpřesnit hodnocení významu rozhodnutí.

Pro tuto práci klíčové jsou pak tabulka `citace`, která obsahuje seznam odstavců, které odkazují na jiná rozhodnutí, a tabulka `odkazynapredpisy`, která obdobným způsobem uvádí ty odstavce, v nichž se objevuje citace ustanovení některého právního předpisu.

5.1.3 Plnění databáze

Přestože v předchozích kapitolách byl popsán způsob, jakým lze relevantní data získat, pro realizaci práce s nimi je nutné nejprve tato data dostat v příslušné podobě do databáze. K tomu slouží modul `src/website/dameons/databaser.py` ve spojení s moduly zpracovávajícími datové soubory pro jednotlivé soudy. Těmito jsou obecný modul `src/website/dameons/soud.py` a moduly `src/website/dameons/usoud.py` (pro Ústavní soud), `src/website/dameons/nsoud.py` (pro Nejvyšší soud) a `src/website/dameons/nssoud.py` (pro Nejvyšší správní soud).

Samotný modul `databaser` je zkonstruován poměrně zvláštním, neintuitivním a nepraktickým způsobem. Jeho spuštění je totiž možné pouze odesláním požadavku na webový server (ať již přes webový prohlížeč či příkazovou řádku) – konkrétně na adresu (při respektování konfigurace uvedené v předchozí sekci) `http://localhost:8001/daemon/databaser`. Je tomu tak proto, že aby modul mohl využívat nástrojů, které Django poskytuje pro práci s databází, musí běžet v rámci stejného webového serveru. A vzhledem k tomu, že hlavním (a pro účely této práce jediným) účelem tohoto modulu je jednorázově naplnit databázi získanými daty, není uživatelská přívětivost postupu nijak významnou hodnotou.

Co se samotného fungování tohoto modulu týče, tak hlavní úlohu v něm plní třída `Databaser` a její metoda `add_verdicts_to_database`. Ta přijímá jako

svůj parametr seznam rozhodnutí (a označení soudu), která následně vkládá do databáze prostřednictvím rozhraní, které poskytuje Django. Stejným způsobem k těmto rozhodnutím následně vkládá jejich jednotlivé odstavce, citace a odkazy na právní předpisy. Aby však tato funkce byla schopná tato rozhodnutí zpracovat, musejí být ve specifickém formátu – a to ve formátu třídy `verdict` definované v modulu `soud.py`.

O převod neupravených zdrojových dat (tedy stažených html či csv souborů) do tohoto formátu se stará trojice metod `process_usoud`, `process_nsoud` a `process_nssoud`. Každá z těchto metod funguje tak, že načte názvy všech relevantních souborů, a tyto pak následně po částech předává třídám zajišťujícím předzpracování dat pro jednotlivé soudy.

5.1.3.1 Zpracování dat Ústavního soudu

U Ústavního soudu jde o třídu `UsoudPreprocessor`. Vzhledem k tomu, že v případě Ústavního soudu dochází (vzhledem ke způsobu získávání dat) k jisté separaci a čištění dat již při jejich získávání, není tato třída příliš objemná. Její role v podstatě spočívá jen v tom, že otevře příslušný csv soubor a následně v cyklu načítá jednotlivé řádky do proměnné `row`, přičemž do jejich obsahu – s jedinou výjimkou – nijak nezasahuje. Zmíněnou výjimku pak představuje text rozhodnutí, který je před dalším zpracováním rozdělen na odstavce (přičemž odstavce nulové délky jsou vyřazeny). Pole odstavců a řádek načtený z csv souboru jsou následně předávány konstruktoru třídy `UsoudVerdict`.

V rámci třídy `UsoudVerdict` jsou informacemi z řádku csv souboru naplněny některé atributy. Většina z nich je přímo použitelná ve formátu, v jakém byla získána z csv souboru. Pouze datum rozhodnutí je konvertováno z textové podoby do datového typu reprezentujícího datum a význam rozhodnutí (soudem přidělený) je převeden z textu na číslo.

Třída rovněž upravuje standardní postup, jímž jsou z textu rozhodnutí extrahovány citace (jenž bude podrobněji popsán v následující části práce), když jej rozšiřuje o metodu `_extract_usoud_links`. Tato metoda je založena na znalosti fungování systému NALUS. Při zobrazení rozhodnutí v tomto systému jsou totiž odkazy na jiná rozhodnutí Ústavního soudu ve formě hypertextových odkazů s pevně danou strukturou. Uvedená funkce této znalosti využívá, když uvedenou strukturu v textu vyhledává speciálně vytvořeným regulárním výrazem `(<\s*a[^\>]*?sz=([^\>]*?) "[^\>]*?>(.*?)<\s*/\s*a>)`. Z nalezených řetězců je následně schopna získat relevantní spisovou značku odkazovaného rozhodnutí i jeho webovou adresu.

5.1.3.2 Zpracování dat Nejvyššího soudu

O zpracování dat získaných *scrapingem* vyhledávače Nejvyššího soudu se stará třída `NsoudPreprocessor`. Tato funkce zpracovává jednotlivé html soubory, z nichž extrahuje metadata a samotný text rozhodnutí. Metadata jsou obsažena v jediné tabulce na stránce, přičemž na každém řádku se nachází jedno z nich. Řádky se skládají ze dvou buněk – v té první je název atributu, v té druhé jeho hodnota. V rámci zpracovávání tabulky probíhá iterace po jejích jednotlivých řádcích, přičemž názvy atributů jsou porovnávány s klíči proměnné `header_names`, která definuje relevantní atributy, jež je záhodno extrahovat. Pokud narazí na shodu, uloží hodnotu do proměnné `headers`.

Vzhledem k tomu, že tabulka s metadaty je spolu s textem rozhodnutí jediným obsahem `<div>`u třídy `main_detail`, není po extrakci metadat problém získat i samotný text rozhodnutí. To probíhá tak, že z `divu` je nejprve odstraněna tabulka. Zbytek je následně rozdělen na jednotlivé odstavce. Seznam odstavců je pak spolu s proměnnou `headers` předán konstruktoru třídy `NsoudVerdict`.

V rámci třídy `NsoudVerdict` jsou takto postoupenými informacemi plněny některé atributy. Tomuto plnění však předchází ještě několik úprav. Nachází-li se v některém z úvodních odstavců věta začínající slovy „*Citace rozhodnutí Nejvyššího soudu by měla obsahovat formu rozhodnutí*“, je celý tento odstavec vyřazen. Tento odstavec totiž není skutečnou součástí rozhodnutí. Jedná se o odstavec, který je přidáván ke každému rozhodnutí a uvádí návod, jak citovat rozhodnutí Nejvyššího správního soudu. Vzhledem k tomu, že součástí návodu je i příklad citace, stal by se (v případě neodstranění tohoto odstavce) z tohoto příkladu nejcitovanější judikát ze všech rozhodnutí Nejvyššího soudu.

Datum rozhodnutí je rovněž konvertováno z textové podoby do datového typu reprezentujícího datum, nicméně z důvodu míchání používaných formátů (DD.MM.YYYY a MM/DD/YYYY) je nutné tak učinit nadvakrát. Nakonec tato třída provádí konverzi soudem uvedeného významu rozhodnutí (na škále A až E) do číselného vyjádření (v rozsahu 1 až 5).

5.1.3.3 Zpracování dat Nejvyššího správního soudu

Zpracování dat o rozhodnutích Nejvyššího správního soudu (získaných z webu *Salvia Kraken*) se stará třída `NSsoudPreprocessor`. Tato funkce zpracovává jednotlivé html soubory, z nichž extrahuje texty rozhodnutí. Vzhledem k tomu, že soubory neobsahují žádná metadata, pokouší se některá z nich (konkrétně význam, číslo jednací a spisovou značku) vyčíst z textu rozhodnutí.

Získání textu rozhodnutí z html souboru je triviální úkol, jelikož celý

text rozhodnutí (a nic jiného) je součástí `<div>` s id `samotnytext`. Poté, co je text zbaven html elementů, je rozdělen na jednotlivé odstavce, přičemž ty prázdné (respektive obsahující pouze netisknutelné znaky) jsou odstraněny. Následně je z prvního odstavce extrahováno číslo jednací a spisová značka.

Předběžné hodnocení významu rozhodnutí probíhá na základě formy rozhodnutí (rozsudek vs. usnesení). Forma je zjišťována analýzou textu prvních čtyř odstavců. Pokud některý z nich obsahuje slovo „rozsudek“, má se za to, že rozhodnutí je vydáno ve formě rozsudku, a je mu přidělena hodnota významu 3 (v opačném případě bude mít nejnižší hodnotu – 5). Pokud se navíc v některém z prvních odstavců objeví sousloví „rozšířeném senátu“, svědčí to o tom, že se jedná o rozhodnutí rozšířeného senátu, a jako takovému je mu přidělena nejvyšší hodnota významu – 1.

Seznam odstavců je následně spolu se získanými metadaty předán konstruktoru třídy `NSsoudVerdict`. V rámci této třídy jsou v podstatě již jen zkopírována získaná metadata do třídních atributů.

S ohledem na očekávané zprovoznění nového vyhledávače Nejvyššího správního soudu jsem se při analýze a vytěžování (meta)dat ze získaných rozhodnutí omezil pouze na nezbytné minimum nutné k řádnému splnění úkolu vytyčeného v této práci. Přestože by bylo nepochybně možné z rozhodnutí získat mnohem rozsáhlejší sadu informací, jeví se úsilí, jež by k tomu bylo nutné vynaložit, tváří v tvář uvedeným okolnostem jako bezúčelné. Po vydání nového vyhledávače totiž stejně bude nutné systém upravit tak, aby bylo možné data o rozhodnutích Nejvyššího správního soud aktualizovat, přičemž v tomto vyhledávací bude k dispozici mnohem rozsáhlejší sada metadat, která bude i snadněji přístupná.

5.1.3.4 Obecné zpracování dat

Výše uvedené části se zabývaly specifickou přípravou dat podle toho, od kterého soudu byla získána. Uvedenými postupy však proces zpracovávání dat nekončí. Zejména samotný text rozhodnutí je totiž následně ještě zpracováván třídou `Verdict` z modulu `soud.py`. V rámci tohoto zpracovávání nejprve dojde k odstranění případných dosud neodstraněných html elementů z textu rozhodnutí. Následně jsou jednotlivé odstavce podrobeny různým analýzám za účelem získání dodatečných informací.

Metoda `_process_paragraphs` realizuje první z nich. Jejím cílem je identifikovat, do kterého významového celku v rámci rozhodnutí daný odstavec patří. Relevantními významovými celky jsou pro účely této funkce *úvod* (též někdy označovaný jako *návěti*), *výrok*, *odůvodnění*, *závěr* a *odlišné stanovisko*. Činí tak ve spolupráci s metodou `_get_number_and_text`. Tato

metoda analyzuje znaky na začátku odstavce a snaží se z nich zjistit, zda odstavec není uvozen číslem (římským či arabským). Pokud by tomu tak bylo, jde o významnou skutečnost z hlediska zařazení odstavce do významových celků. Arabské číslice se totiž užívají téměř výhradně v odůvodnění a odlišných stanoviscích. Římské číslice se pak používají ve výroku a někdy rovněž jako součást názvů kapitol v odůvodnění. Návěti a závěr jsou pak zpravidla nečíslované.

Pro posouzení hranic jednotlivých částí textu rovněž uvedená funkce využívá znalosti některých klíčových výrazů, které se zpravidla neobjevují v jiných částech rozhodnutí. Jde zejména o první výskyt slova „odůvodnění“ (jež běžně zahajuje blok *odůvodnění*), a o situaci, kdy je odstavec uvozen výrazem „poučení“ (jež zpravidla zahajuje závěrečnou část rozhodnutí) či „odlišné stanovisko“ (zahajující část obsahující odlišná stanoviska). Poté, co funkce tuto analýzu dokončí, budou odstavce rozděleny do jednotlivých skupin podle obsahu. Na každou z těchto skupin bude následně zavolána metoda `_process_paragraph_block`.

Metoda `_process_paragraph_block` prochází jednotlivé odstavce z každé skupiny a snaží se v nich identifikovat případné podskupiny (např. kapitoly v rámci odůvodnění). Za hranici mezi podskupinami jsou považovány krátké odstavce očíslované jinou sadou čísel než okolní odstavce (tedy zpravidla nadpisy kapitol). Pokud funkce žádné takové podskupiny nenajde, pokusí se alespoň najít číslování jednotlivých odstavců. A pokud nenajde ani to, očíslovuje je vzestupnou posloupností přirozených čísel. Všechna takto získaná data o obsahu odstavců se následně použijí pro tvorbu objektů třídy `Paragraph`.

Po prozkoumání obsahu odstavců pokračuje zpracovávání textu rozhodnutí metodou `_extract_links`, které hledá v rozhodnutí odkazy na jiná rozhodnutí. Odkazy na rozhodnutí identifikuje pomocí dvojice vlastních regulárních výrazů, z nichž jeden by měl popisovat všechny relevantní spisové značky Nejvyššího a Nejvyššího správního soudu a druhý by měl detekovat odkazy na rozhodnutí Ústavního soudu. Pokud odkaz najde, uloží jej do atributu `links` a v textu původní spisovou značku uzavře do dvojitých hranatých závorek.

Smyslem uzavření odkazů do závorek je označit již identifikované odkazy a zabránit tak jejich opakované detekci v případě, že by v budoucnu došlo ke zdokonalení (či rozšíření) metody jejich detekce, a bylo by tedy potřeba znovu posoudit texty všech rozhodnutí. Uvedené označení rovněž zjednodušuje identifikaci odkazů v textu, což bude relevantní při zobrazování výsledků v případě, že projekt postoupí do fáze tvorby vyhledávače.

Poslední metodou, která je součástí procesu zpracování dat, je

`_extract_laws`. Tato metoda v textu vyhledává odkazy na zákonná ustanovení, která následně ukládá do atributu `laws`. Funguje tak, že nejprve za pomoci regulárního výrazu `(\s*\.[1-9][0-9a-zA-Z]*\s[,a]?)` najde označení právního předpisu, a následně prozkoumává řetězec směrem před tento text, v němž hledá případnou podrobnější specifikaci právního ustanovení (např. článek, paragraf, odstavec či písmeno).

5.2 Několik poznámek k datům

Před samotným uvedením metod používaných ke klasifikaci dat, považuji za vhodné se alespoň stručně zmínit o obsahu a některých zvláštnostech získaných dat. Tyto informace je totiž nutno vést v patrnosti při hodnocení efektivity použitých metod a nejspíš i při jejich designu.

5.2.1 Rozsah dat

Navzdory omezení rozsahu rozhodnutí pouze na ta vydaná po roce 2004 dosahují zajištěná data i tak značného objemu. Html a csv soubory z jednotlivých soudů zabírají 10,3 GB . Databáze pak rovněž přesahuje 8 GB. Tento objem představuje 68 405 rozhodnutí Ústavního soudu, 118 259 rozhodnutí Nejvyššího soudu a 52 228 rozhodnutí Nejvyššího správního soudu.

Vzhledem k nepochybné časové náročnosti výpočtů nad takto velkou množinou dat je tedy vhodné zamyslet se nad tím, zda lze navrhnout takové metody, které budou schopny si své metriky předpočítat a nebudou tak vyžadovat provádění rozsáhlých výpočtů nad celou databází při zpracovávání dotazu. A i pokud nebude z hlediska efektivity takový postup dostatečný, je možné tyto metody použít přinejmenším pro omezení rozsahu výpočtu v době zpracovávání dotazu na základě kvalifikovaného odhadu.

5.2.2 Problémové atributy

Atributy, které se společně s rozhodnutím přenášejí do databáze se liší nejen ve svém významu pro klasifikaci rozhodnutí, nýbrž i ve vztahu k jejich spolehlivosti. Často se v attributech objevují drobné i závažnější nekonzistentnosti. Nejvýznamnější je tento problém u rozhodnutí Nejvyššího správního soudu, kde v důsledku absence metadat zcela chybějí významné informace, které jsou pak následně obtížně rekonstruovány s nezanedbatelnou mírou chybovosti. Tento problém snad bude již brzy vyřešen zpřístupněním nového vyhledávače.

Chyby se však vyskytují i na dalších místech. Například ve spisové značce rozhodnutí se občas objeví více rozhodnutí spojených středníkem. Nejde přitom o pochybení soudu – k uvedené situaci dochází ve věcech, kdy řízení byla vedena samostatně, nýbrž v jejich průběhu soud dospěl k závěru,

že jsou si natolik podobná, že je z hlediska procesní hospodárnosti vhodné řízení spojit. V takovém řízení je následně vydáno pouze jedno rozhodnutí, které je označeno spisovými značkami všech spojených řízení⁵⁸.

Spojené spisové značky se u rozhodnutí obsažených v databázi objevují u 294 rozhodnutí. To sice není nijak závratné číslo (a pro účely této práce nemá necelých 300 rozhodnutí zásadnější význam), nýbrž daný problém by měl být vyřešen před případným spuštěním řádného vyhledávacího nástroje. V úvahu přitom přicházejí dvě možnosti, jak tak učinit – tedy zalistováním téhož rozhodnutí pod všemi spisovými značkami, či jeho omezení na pouze jednu spisovou značku. První řešení přitom zvyšuje význam daného rozhodnutí (bylo-li by někdy v budoucnu citováno), ten druhý ne.

V této otázce považuji za správné takové řešení, které umožňuje ponechat rozhodnutí v databázi pod všemi spisovými značkami, ale nenásobit kvůli tomu obdržené citace (a to i v případě, kdy je rozhodnutí citováno pod všemi jeho spisovými značkami). Málom který soudce totiž při své právní argumentaci běžně cituje více totožných rozhodnutí. Nelze tedy předpokládat, že byla-li by tato (identická) rozhodnutí vydána samostatně, byla by (všechna najednou) běžně citována ve stejné míře.

5.2.3 Zajímavé atributy

I některé z atributů, které v této práci nehrají zásadní roli, si však rovněž zaslouží přinejmenším drobnou zmínku. Při případné tvorbě vyhledávače totiž mohou být v zájmu co nejvyšší efektivity důležité i na první pohled ne příliš významné atributy, pokud budou schopny přispět ke zvýšení přesnosti klasifikace.

5.2.3.1 Právní věta

U obou soudů, které ve snadno zpracovatelné podobě poskytují spolu s rozhodnutími i právní věty (tedy Ústavní soud a Nejvyšší soud), je zřejmé, že velká část jejich rozhodnutí není opatřena právní větou. V případě Ústavního soudu je to pouze 4 472 rozhodnutí (tedy zhruba 6,5 %). U Nejvyššího soudu se pak jedná o 2 603 rozhodnutí (tedy o pouhých 2,2 %).

Tato zjištění jen podtrhují význam právních vět jakožto silných indicií svědčících o významnosti rozhodnutí (tento předpoklad vychází ze znalosti role právních vět v rámci rozhodovací činnosti soudů). Svědčí totiž o tom, že soudy právními větami neplýtvají a že tedy rozhodnutí jimi opatřená zpravidla významná budou.

⁵⁸např. podle § 112 občanského soudního řádu

Z hlediska implementace je vhodné na tomto místě upozornit na skutečnost, že Ústavní soud v položce *právní věta* u rozhodnutí, která nejsou opatřena právní větou, uvádí: „*Právní věta není k dispozici.*“ Tuto hodnotu pole *právní věta* je tedy nutné považovat za srovnatelnou s absencí právní věty a buď ji z databáze odstranit, nebo při implementaci zajistit, že s ní bude správně nakládáno.

5.2.3.2 Soudem posouzený význam

Obdobně jako u právní věty, i u položky *význam* (která představuje hodnocení významu rozhodnutí na pětistupňové škále provedené soudcem, který ve věci rozhodoval) je zřejmá střídmost v užívání nejvyšších významových kategorií. Ústavní soud totiž do kategorie 1 zařadil pouze 154 rozhodnutí; do kategorie 2 jich zařadil 693 a do kategorie 3 to bylo 3 854. Nejvyšší soud přiznal status rozhodnutí 1. kategorie 2804 rozhodnutím; v kategorii 2 pak bylo zastoupeno 3 722 rozhodnutí. Ostatní kategorie již byly mírou svého užívání výrazně četnější.

Výše uvedené podporuje přesvědčení, že 1. kategorie jsou u obou soudů vyhrazeny pouze těm nejpřelomovějším rozhodnutím, která lze téměř vždy považovat za důležitá. Ani rozhodnutí 2. (a u Ústavního soudu i 3.) kategorie přitom nejsou nijak výrazně nadreprezentována. Tento atribut by tedy při klasifikaci rozhodnutí rovněž měl hrát významnou roli.

5.2.3.3 Další atributy

Výše uvedeným se seznam případných užitečných atributů nevyčerpává. V následujících odstavcích tedy poskytují alespoň stručné posouzení potenciálu některých z nich. Mezi další potenciálně využitelné atributy patří např.:

paralelní citace ve sbírce zákonů – ta udává, zda rozhodnutí bylo vyhlášeno ve sbírce zákonů. Ve sbírce zákonů se přitom vyhlášují pouze taková rozhodnutí Ústavního soudu, jimiž bylo rozhodnuto o (ne)souladu zákona s Ústavou (a jeho případném zrušení)⁵⁹. Rovněž přitom platí, že „*má-li právní názor Ústavního soudu vyslovený v nález, který se nevyhlašuje ve Sbírce zákonů, všeobecný význam, může Ústavní soud rozhodnout, že tento právní názor bude uveřejněn ve Sbírce zákonů*“. Zejména takováto rozhodnutí (jež Ústavní soud z vlastní vůle rozhodl publikovat ve sbírce zákonů) by tedy mohlo být vhodné sledovat.

paralelní citace ve sbírce nálezů a usnesení udává, zda rozhodnutí bylo publikováno ve sbírce nálezů a usnesení Ústavního soudu. V této sbírce

⁵⁹§ 57 zákona o Ústavním soudu

se zveřejňují nálezy a vybraná usnesení Ústavního soudu. Zejména ve vztahu k usnesením se tak jedná o cenný zdroj informací ohledně jejich významu.

odkazy na právní předpisy a precedenci představují významnou informaci o tom, které právní předpisy a dřívější rozhodnutí považoval soudce za důležité v konkrétní věci.

5.3 Klasifikační metody

Jelikož databáze je již na základě postupů z předchozí části naplněna, je již vše připraveno pro implementaci samotných metod umožňujících efektivní klasifikaci judikatury.

5.3.1 Statická váha

Přestože metoda statické váhy je dosti triviální, její implementace je nutná pro jednu z variant metody *PageRank*. Její existence bude rovněž velmi užitečná při hodnocení kvality ostatních metod. Pokud totiž náročnější (jak implementačně, tak zejména z hlediska výpočetního času a požadovaných prostředků) metody nedosáhnou ani takových výsledků jako metoda základní, nelze je považovat za úspěšné.

5.3.1.1 Volba parametrů

Na základě výše uvedených zjištění, rozsáhlého studia získaných dat a mých zkušeností nabytých při výkonu právního povolání jsem se při výpočtu skóre pro jednotlivá rozhodnutí rozhodl použít následující parametry:

Hodnocení významu ze strany soudu – Tento parametr je nepochybně nejvýznamnější, neboť je založen na odborném posouzení osoby zodpovědné za vydání rozhodnutí.

Rozhodovací formace – Pokud bylo rozhodnutí vydáno rozšířeným senátem Nejvyššího správního soudu, velkým senátem Nejvyššího soudu, nebo konfliktním senátem, lze s velkou pravděpodobností konstatovat, že jde o rozhodnutí významné. (U rozhodnutí ostatních formací naopak nelze jen na základě této informace činit zásadnější závěry.)

Právní věta – Rozhodnutí opatřené právní větou je zpravidla významné. O tom svědčí skutečnost, že rozhodující soudce měl potřebu jej touto větou opatřit.

Forma rozhodnutí – Rozhodnutí vydané ve formě rozsudku je téměř vždy významnější než rozhodnutí vydané formou usnesení (nejedná-li se

o usnesení některé z výše uvedených zvláštních formací). Rozhodnutí vydané ve formě nálezu je zpravidla rozhodnutím významným.

Datum vydání – U novějšího rozhodnutí je jistá pravděpodobnost, že je významnější než starší rozhodnutí obdobných charakteristik. Pokud by totiž starší rozhodnutí všechny relevantní právní otázky uspokojivě vyřešilo, byla by novější rozhodnutí vydávány v méně významných formách, neboť by starší rozhodnutí pouze citovala.

Veden těmito úvahami jsem vytvořil algoritmus, který uvedené charakteristiky zohledňuje. Skládá se z následujících kroků:

1. Pokud je rozhodnutí opatřeno hodnocením významu ze strany vydávajícího soudu, použije se za základ dvojnásobek tohoto rozhodnutí (tedy 2 až 10 bodů). Pokud jím opatřeno není, budou mu přiděleny 3 body (tedy význam na pomezí dvou nejnižších tříd). K zdvojnásobení vstupního hodnocení jsem přistoupil pro to, abych v dalších krocích nemusel příliš používat desetinná čísla.
2. Pokud je rozhodnutí vydáno rozšířeným senátem či velkým senátem, zvýší se jeho hodnocení o 20 %. Dosáhlo-li by rozhodnutí po tomto zvýšení méně než sedmi bodů, zvýší se jeho skóre na 7 bodů.
3. Pokud je rozhodnutí opatřeno právní větou, zvýší se jeho hodnocení o 15 %. Dosáhlo-li by rozhodnutí po tomto zvýšení méně než osmi bodů, zvýší se jeho skóre na 8 bodů.
4. Splňuje-li rozhodnutí dva předchozí body zároveň, udělí se mu tolik bodů, aby jich mělo alespoň 10.
5. Pokud je rozhodnutí vydáno ve formě rozsudku či nálezu, zvýší se jeho hodnocení o 2 %. Dosáhlo-li by rozhodnutí po tomto zvýšení méně než šesti bodů, zvýší se jeho skóre na 6 bodů.
6. Rozhodnutí se přičte 1 bod, bylo-li vydáno v roce 2021. Bylo-li vydáno dříve, sníží se tato bonifikace o 0,05 bodu za každý rok.

Na první pohled se může jevit nelogické, proč u vydání rozhodnutí vybranou rozhodovací formací dochází ke 20% bonifikaci s minimem 7 bodů, zatímco v případě rozhodnutí opatřeného právní větou je bonifikace nižší, ale minimum vyšší. Je tomu tak proto, že přestože rozhodnutí uvedených formací jsou zpravidla významná, objevují se mezi nimi i rozhodnutí zcela bezvýznamná (např. rozhodnutí, v nichž soud pouze konstatuje, že věc nespadá do jeho pravomoci). Naproti tomu právní větou bývají běžně opatřena i méně významná rozhodnutí (proto nižší bonifikace), ale nestává se, že by právní větou byla opatřena rozhodnutí bezvýznamná (proto vyšší minimum).

5.3.1.2 Možné nedostatky zvoleného postupu

Volba parametrů a jejich vah tak, jak byla popsána v předcházejících odstavcích, samozřejmě není jediná možná, a rozhodně ani není nejlepší. V předchozí části práce byla předestřena řada užitečných parametrů, které se do závěrečného výběru nedostaly, přestože by nepochybně mohly přispět ke zvýšení úspěšnosti metody. Jak selekce, tak stanovení vah jednotlivých parametrů přitom mohly být provedeny za základě mnohem rigoróznějšího procesu, který by např. zahrnoval i rozsáhlé testování, optimalizaci a vyhodnocování pozitiv a negativ každé konfigurace.

Takový postup by však s ohledem na zaměření této práce byl zbytečně časově náročný – zejména v důsledku dlouhých časových intervalů, které by si vyžadovalo důkladné ladění parametrů na takto rozsáhlé databázi. A přestože pokročí-li projekt do fáze přípravy samotného vyhledávače, bude nejspíš nutné uvedený proces absolvovat, pro naplnění cílů této práce není tato otázka příliš podstatná. Metoda statických vah totiž není středobodem této práce, nýbrž především představuje jednoduché řešení, s nímž je možné následně srovnávat řešení náročnější. Množství informací, které je z takového srovnání možné získat přitom závisí především na nastavení podmínek tohoto srovnávání, nikoliv na přesných parametrech metody statických vah.

5.3.1.3 Podrobnosti o implementaci

Výsledný algoritmus je implementován v souboru `src/website/daemons/static_weights.py`. Funguje tak, že iteruje nad celou množinou rozhodnutí, přičemž pro každé z nich vypočítá skóre a uloží ho do položky `static_weight` v tabulce rozhodnutí. Při samotném výpočtu vychází především z metadat, nicméně pokud některé informace nejsou dostupné, pokouší se je zrekonstruovat z textu rozhodnutí a dalších údajů. To se týká informace o typu rozhodnutí a zvláštní rozhodovací formaci, kterou rekonstruuje z prvních odstavců obdobně jako při zpracovávání dat z databáze Salvia Kraken. Informaci o tom, že ve věci rozhodoval konfliktní senát, dovozuje ze spisové značky – neboť pouze konfliktní senát má ve svých spisových značkách výraz `konf`. Pokud chybí informace o datu vydání rozhodnutí, pokouší se ji metoda najít v posledních čtyřech odstavcích (jelikož datum vydání rozhodnutí je zpravidla na jednom z posledních řádků jeho textu).

5.4 Počet citací

Metody založené na počtu citací jsou implementovány v souboru `src/website/daemons/citations.py` v rámci třídy `CitationCalculator`. Tato třída podporuje současně 4 metriky založené na počtu citací:

Prostý počet citací – metoda, která každému rozhodnutí přiřadí skóre odpovídající jeho počtu citací napříč ostatními rozhodnutími.

Vážený počet citací – metoda, která se od té předchozí liší tím, že každá citace se na výsledku podílí s vahou odpovídající statické váze daného rozhodnutí. (Pokud je tedy zkoumané rozhodnutí citováno ve dvou rozhodnutích, z nichž jedno má váhu 5 a druhé 7,5, bude zkoumanému rozhodnutí přiřazeno skóre 12,5.)

Prostý počet citací se zohledněním stáří – obdoba první metody, která však váhu počtu citací snižuje v závislosti na stáří rozhodnutí. Zatímco u rozhodnutí z roku 2021 má každá citace plnou váhu, u rozhodnutí z roku 2004 má váhu poloviční. Mezi těmito obdobími váha klesá lineárně.

Vážený počet citací se zohledněním stáří – obdoba druhé metody využívající rovněž stejný princip jako metoda třetí.

Samotný proces výpočtu vah zajišťuje metoda `process`. Ta nejprve iteruje nad celou množinou citací (tedy položek tabulky `citace`), z níž opakovaným voláním metody `calculate_citations` sestavuje slovník citačních vah (v atributu `citations`). Po dokončení cyklu by tak v této proměnné měly být obsaženy spisové značky všech rozhodnutí, která byla alespoň jednou citována, společně s jejich celkovým počtem citací a váženým počtem citací. Metoda `save_citation_counts` následně iteruje nad tabulkou rozhodnutí, přičemž ke každému z nich uloží jeho počet citací (do sloupce `pocet_citaci`) a vážený počet citací (do sloupce `vazeny_pocet_citaci`). Rovněž vypočte a uloží koeficient, kterým bude ponižována váha citací v závislosti na jeho stáří (do sloupce `old_penalty`).

5.5 PageRank

K implementaci variant algoritmu *PageRank* jsem využil nástroj `NetworkX`⁶⁰. Ten pro svoji činnost vyžaduje matici hran. Její tvorbu zajišťuje metoda `create_graph` ze třídy `Acquirer` (soubor `src/website/daemons/acquirer.py`). Tato metoda iteruje nad seznamem všech citací, který postupně přepisuje do csv souboru.

Samotný výpočet pak realizuje modul `src/website/daemons/page_rank.py` a funkce `page_rank`. Po zajištění matice hran, seznamu relevantních rozhodnutí pro každý dotaz a seznamu všech výsledků nalezených pro jednotlivé dotazy funkce z těchto dat vytvoří grafy odpovídající požadavkům pro

⁶⁰NetworkX Developers. *pagerank* [online] [cit. 24. 6. 2021]. Dostupné z: https://networkx.org/documentation/networkx-1.10/reference/generated/networkx.algorithms.link_analysis.pagerank_alg.pagerank.html

5. IMPLEMENTACE

zpracování NetworkX *PageRankem*.

Proces eliminace hran probíhá tak, že z matice hran jsou vyřazeny všechny hrany, které spojují pouze rozhodnutí, jež nejsou obsažena ve výsledcích vyhledávání, čímž vznikne *širší* hranová matice. Kromě *širší* matice vzniká eliminací i matice *uzší*. Do té jsou zařazeny pouze hrany, které spojují dvojici uzlů, jež oba náležejí mezi výsledky vyhledávání.

Z těchto dvou matic je následně vytvořena čtveřice grafů (jednou vážených a jednou nevážených), u nichž je následně vypočten jejich *PageRank*. K tomuto výpočtu dochází rovněž nadvakrát – a to jednou s *damping factorem* 0.85 a podruhé 0.9.

Hodnocení kvality

Data jsme získali, metody objevili a naimplementovali. Po jejich spuštění vyšlo najevo, že dokonce vracejí zhruba taková data, jaká jsme očekávali. Pro řádné posouzení úspěšnosti této práce jsou však tyto skutečnosti nedostatečné. Abychom mohli vyslovit alespoň trochu přesvědčivý závěr o užitečnosti (či naopak neúspěchu) implementovaných metod, je nutné provést nějaké objektivizovatelné testy.

6.1 Metodika hodnocení

Problémem mnoha optimalizačních problémů však je, že „správný výsledek“ neznáme, a nemůžeme tedy jednoduše posoudit, zda použité řešení jej dosahuje, či nikoliv. To samozřejmě neznamená, že je to nemožné. V oblasti získávání informací (*information retrieval*) existuje mnoho osvědčených metrik umožňujících zhodnotit kvalitu implementace. Mezi ty bezpochyby nejznámější a nejpoužívanější lze jistě zařadit *precision* (přesnost) a *recall* (výťažnost)⁶¹. Přesnost udává, jaká část z vyhledávačem nalezených dokumentů (vzhledem ke konkrétnímu dotazu) je relevantní. Výťažnost pak udává poměr mezi počtem relevantních dokumentů, které pro zadaný dotaz vyhledávač vrátil, a počtem všech relevantních dokumentů, které vrátil mohl (a které by vrátil dokonalý vyhledávač).

Obě metody však trpí stejným zásadním problémem – pro jejich řádnou aplikaci je totiž nutné vědět, jaké dokumenty jsou relevantní pro jednotlivé dotazy. A vzhledem k tomu, že poskytnout univerzální odpověď na tuto otázku je prakticky nemožné, omezují se reálné aplikace těchto metod zpravidla na hodnocení předem připravených dotazů, k nimž je reálné získat vcelku přesnou informaci o tom, jaké výsledky jsou pro ně relevantní.

⁶¹TING K.M. (2011) Precision and Recall. In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA. ISBN 978-0-387-30768-8

6. HODNOCENÍ KVALITY

Obdobně jsem proto postupoval i v této práci. Vytvořil jsem si seznam 50 dotazů (spočívajících v ustanovení právního předpisu), k nimž jsem v komentářové literatuře byl schopný dohledat rozhodnutí, která právní experti označili za nejvýznamnější z hlediska daných ustanovení. Takto získané výsledky jsem ještě doplnil o některá rozhodnutí, která se (při zadání stejného dotazu) objevovala na prvních místech ve vyhledávači ASPI. Tyto informace jsem zapsal do souboru [src/website/daemons/test.csv](#). Jejich struktura je zachycena na obrázku č. **6.1**.

Právní předpis	Relevantní judikáty					Text právního předpisu				
	Pl. ÚS 8/06	I. ÚS 2930/09	III. ÚS 455/08	II. ÚS 1540/08	II. ÚS 233/09	plní-li povinný ve lhůtě 30 dnů ode dne doručení	anikne-li oprávnění exekutora k vedení exekuce	okud exekutor před tím, než z důvodů podle odstavce	nebyl-li exekutor pověřen vedením exekuce a zamítl-li	okud exekutor nebyl pověřen vedením exekuce
LAW 330/2001 sb. §11										
LAW 40/2009 sb. §22	5 Tdo 741/2013	8 Tdo 797/2010	5 Tdo 875/2009	1 To 38/70		achatelem trestného činu je, kdo svým jednáním naplnil znaky	achatelem trestného činu je i ten, kdo k provedení činu užil jiné osoby			
LAW 40/2009 sb. §28	II. ÚS 3383/14	5 Tdo 1121/2019	11 Tdo 1233/2016	6 Tdo 871/2013		in jinak trestný, kterým někdo odvrací nebezpečí přímo	ejde o krajní nouzi, jestliže bylo možno toto nebezpečí			

Obrázek 6.1: Sada dat určená k ověření efektivity implementace

V prvním sloupci je uvedeno příslušné ustanovení právního předpisu ve formě LAW| [číslo předpisu] / [rok vydání] sb. | § [dotčený odstavec]. V následujících pěti sloupcích jsou pak uvedena významná rozhodnutí vztahující se k tomuto ustanovení. Závěrečná pětice sloupců pak obsahuje citace významných částí dotčeného ustanovení.

6.2 Problémy s vyhledáváním

Tato práce je především zaměřena na hodnocení kvality nalezených výsledků – tedy posouzení toho, která z rozhodnutí (jež odpovídají danému dotazu) jsou nejvýznamnější, a uživatel by tak měla být zobrazena nejdříve. Dosažené výsledky jsou však významně ovlivněny samotnou kvalitou vyhledávače – tedy zda vůbec relevantní výsledky budou nalezeny (v opačném případě totiž nebude možné posoudit správnost jejich seřazení, neboť nebude co řadit). Při hodnocení kvality klasifikačních metod je tedy nutné rozumně odstínit nedostatky při samotném vyhledávání.

S tímto problémem jsem se vypořádal tak, že jsem postupně provedl vyhledání výsledků pro všechny dotazy a vyhodnotil jsem jejich úspěšnost. Vyhledávání (realizované metodami `find_matches` a `find_more_matches` z modulu `acquirer`) kombinovalo dva zdroje dat. Jedním z nich byla tabulka popisující vztahy rozhodnutí k právním předpisům (`OdkazyNaPredpisy`). Každý rozsudek, který měl v této tabulce odpovídající záznam (ve vztahu k zadanému dotazu), byl funkcí `find_more_matches` přidán do seznamu výsledků. Stejně tak byly do seznamu výsledků (funkcí `find_matches`) zařazena ta rozhodnutí, která sice neměla v příslušné tabulce výslovně uvedený odkaz na dané rozhodnutí, nicméně v jeho textu se nacházelo číslo příslušného předpisu, číslo relevantního ustanovení, a alespoň jedna z významných částí textu daného ustanovení (tedy alespoň jedna položka z posledních pěti sloupců ze souboru `test.csv`).

Soubory obsahující výsledky pro jednotlivé dotazy jsou obsaženy ve složce `src/website/daemons/matches/`. Samotné posouzení úspěšnosti vyhledávání následně proběhlo prostřednictvím funkce `test_testing_set` z modulu `src/website/daemons/page_rank.py`. Jeho výsledky jsou obsaženy v souboru `test_results/search_results`. Z těchto výsledků vyplývá, že pouze ke třem dotazům se podařilo najít všechna relevantní rozhodnutí. Naproti tomu u celých 20 rozhodnutí se nepodařilo najít žádný relevantní výsledek. Průměrná úspěšnost vyhledávání se pohybuje kolem 31 %. Při vynechání dotazů, které nenašly žádné výsledky, dosahuje úspěšnost 50 %. Průměrný počet výsledků na dotaz je 129, medián je 67.

Uvedené výsledky pochopitelně nejsou zrovna uspokojivé. Redukci validační sady na 30 dotazů, u nichž bylo relevantní řešení nalezeno (zbytek je totiž pro hodnocení kvality řazení výsledků nepoužitelný), je však ještě akceptovatelné, a 50% úspěšnost pořád celkem obstojně umožňuje hodnotit rozdíly ve výsledcích jednotlivých metod.

Co se příčiny těchto ne zrovna oslnivých výsledků týče, to nelze v tento okamžik určit přesně. Mezi potenciální důvody však může patřit především:

- složitost identifikace vztahu rozhodnutí ke konkrétnímu ustanovení právního předpisu. Právní předpisy totiž nejsou vždy citovány svým plným názvem a často jsou nahrazovány zkratkami, které v textu nejsou ani definovány. Pro jejich přesnější identifikaci je tedy potřeba rozsáhlejší právní znalosti a důkladnější analýzy textu rozhodnutí jakožto přirozeného jazyka. Nabízí se např. možnost supervizovaného učení na základě velkého množství rozhodnutí, u nichž jsou odkazy řádně vyznačeny.
- neexistence oficiálních seznamů významných rozhodnutí ve vztahu k jednotlivým předpisům. Přestože komentářová literatura je z hlediska její

erudice zpravidla na výši, výběr v ní uvedených rozhodnutí je do značné míry na vůli samotného autora. Občas se tak stane, že autor nezmíní některá významná rozhodnutí (např. proto, že v nich obsažené právní názory mu přijdou natolik samozřejmé, že není potřeba je v komentáři podpořit odkazem na příslušnou judikaturu), či jejich význam subjektivně vnímá jinak.

- nedokonalá extrakce metadat a tvorba vyhledávače. Vzhledem k tomu, že smyslem této práce bylo analyzovat, zda je vůbec reálné zefektivnit dosud existující nástroje určené pro vyhledávání v judikatuře, je pochopitelné, že samotný vyhledávač není na špičkové úrovni. K jeho vývoji a zdokonalování bude totiž namísto přistoupit pouze tehdy, pokud by výsledky této práce svědčily o tom, že je možné dosáhnout výrazného zlepšení oproti současnému stavu.

6.3 Praktický postup hodnocení

Pro samotné hodnocení jsem se rozhodl zvolit následující postup kombinující *precision* a *recall*: Pro výsledek každého vyhledávání budu uvažovat polovinu celkového počtu nalezených výsledků, nanejvýš však 10. Pokud některé z nalezených významných rozhodnutí nebude v této skupině, bude považováno za nenalezené. Jednotlivé metody budou hodnoceny ve vztahu ke konkrétním dotazům na škále od nuly do sta procent. 100% hodnocení dosáhne taková metoda, která výsledky seřadí tak, že všechna významná rozhodnutí budou na začátku seznamu, 0% pak taková, u které budou všechna významná rozhodnutí považována za nenalezená. Každé nenalezené rozhodnutí pak znamená penalizaci ve výši nepřímo úměrné počtu významných rozhodnutí (při 5 rozhodnutích to bude 20 %, při dvou 50 %, etc.)

Hodnocení za nalezená významná rozhodnutí bude nadále sníženo v případě, že tato budou zařazena až za některými nevýznamnými rozhodnutími. Výše této penalizace se určí na základě vzorce $(1/N * (jumps / (maxJumps + 1)))$, kde N je počet všech významných rozhodnutí, $jumps$ udává, kolikrát bylo některé z významných rozhodnutí „přeskočeno“ rozhodnutím nevýznamným, přičemž $maxJumps$ udává maximální počet těchto přeskočení při daném počtu nalezených významných rozhodnutí – tato situace nastane, jsou-li všechna nalezená relevantní řešení zařazena až na konci relevantního seznamu rozhodnutí (tedy těsně před hranicí, kdy už by byla považována za nenalezená).

Samotný počet skoků je pak suma přes všechna nalezená nevýznamná rozhodnutí. Každé z nich přitom do této sumy přispěje hodnotou odpovídající počtu významných rozhodnutí, která přeskočilo. Takto nastavená metoda tak bude schopna zajistit, že řešení, které najde více relevantních výsledků,

bude vždy hodnoceno jako lepší. Zároveň je schopna vhodně rozlišovat míru úspěšnosti mezi řešeními, kdy byl nalezen stejný počet relevantních výsledků.

6.4 Statické metody

Nejprve jsem vyhodnotil statické metody – tedy metodu statických vah a metody založené na statickém počtu citací. Implementace testu je obsažena v modulu `src/website/daemons/validator.py`. Výsledky provedených testů jsou pak k dispozici v souborech `test_results/static_methods_aggregated_results.csv` (ve stručné podobě) a `test_results/static_methods_results.csv` (podrobné výsledky rozepsané podle jednotlivých metod a dotazů). Výsledky, kterých jednotlivé metody dosáhly, jsou uvedeny na obrázku č. 6.2).

Metoda	Skóre
Statická váha	19,69 %
Prostý počet citací	25,85 %
Počet citací se zohledněním stáří (maximálně 50% penalizace)	26,87 %
Počet citací se zohledněním stáří (maximálně 90% penalizace)	26,71 %
Vážený počet citací	27,13 %
Vážený počet citací se zohledněním stáří (maximálně 50% penalizace)	26,25 %
Vážený počet citací se zohledněním stáří (maximálně 90% penalizace)	26,40 %

Obrázek 6.2: Výsledky měření efektivity statických metod

Jak je vidět, metody založené na počtu citací vykazují znatelně lepší výsledky než metoda založená pouze na informacích o rozhodnutí samotném. Tento výsledek tedy potvrzuje předpoklad, že vzájemné provázání rozhodnutí odkazy je významným ukazatelem jejich významu. Využití tohoto druhu metod se tedy v případné implementaci pokročilejší podoby vyhledávače jeví jako rozumné.

I samotná metoda statické váhy však nevykazuje nijak tristní výsledky. A přestože sama o sobě nepochybně není nejvhodnější pro hodnocení kvality rozhodnutí, svědčí naměřené výsledky o tom, že atributy ani jejich váha nebyly vybrány zcela nesprávně. Po případném doladění rozsahu používaných informací a jejich vah pak tato metoda může úspěšně sloužit jako vhodný doplněk metod pracujících s analýzou textu rozhodnutí či jejich vzájemných vazeb.

Jednotlivé varianty metod založených na počtu citací se pak od sebe již tak výrazně neliší, nicméně i tak lze pozorovat, že metoda založená na váženém počtu citací dosáhla mírně lepších výsledků, než metoda opírající se o prostý počet citací. I to je v souladu s původními očekáváními založenými na tom, že pokud je rozhodnutí citováno významnými rozhodnutími, vzrůstá u něj šance, že je také významné. Při případné implementaci této metody do finálního vyhledávače bude však namísto věnovat výraznější pozornost způsobu, jakým je váha počítána. Dalšímu zvýšení úspěšnosti by pak mohlo nejspíš pomoci důkladnější třídění obsahu rozhodnutí na významové celky a následná diferenciací vah v závislosti na tom, v jaké části rozhodnutí se odkaz nachází.

Zatímco dvě výše uvedené hypotézy při testování (v rámci možností) obstály, o třetí z nich to již neplatí. Přestože totiž vliv stáří rozhodnutí na jeho význam byl zkoumán hned ve čtyřech variantách použitých metod, jeho dopad je zcela zanedbatelný. To svědčí o tom, že úvaha, která mě vedla k aplikaci tohoto principu, zřejmě nebyla správná. Možná jsem podcenil konzervatismus justice a s ní spojené upřednostňování starších textů. Možná, že právní texty nezastarávají tak rychle, jak jsem předpokládal. Ať je již důvod jakýkoliv, je zjevné, že stáří rozhodnutí nemá na jeho význam takový vliv, jaký jsem předpokládal. Pokud tedy vůbec bude tento parametr ve finálním vyhledávači hrát nějakou roli, bude spíše podpůrná.

6.5 PageRank

Implementace testu PageRanku je obsažena v modulu `src/website/daemons/page_rank.py` (společně s dalšími funkcemi zaměřenými na vyhodnocování testů a jejich zpracování do prezentovatelné podoby). V rámci testování jeho variant jsem se zaměřil především na následující 3 parametry:

rozsah grafu – tedy, zda mají být v modelu zohledněny pouze vazby mezi rozhodnutími, která obě splňují podmínky vyhledávacího dotazu (*užší* varianta), nebo zda stačí, tyto podmínky splňovalo alespoň jedno rozhodnutí z dané vazby (a jsou tedy zohledňovány i externí odkazy).

váhy vazeb mezi rozhodnutími – tedy, zda mají mít všechny vazby (odkazy) stejnou váhu, či zda vazby z významnějších rozhodnutí (posuzováno na základě jejich statické váhy), mají být důležitější.

damping factor (též označován jako *alfa*) – jeden z významných parametrů samotného PageRanku mimo jiné ovlivňující rychlost konvergence výsledků.

Rozsah	Vážený	Alfa	Skóre
Široký	Ne	0,85	11,87 %
Široký	Ano	0,85	11,87 %
Úzký	Ne	0,85	19,56 %
Úzký	Ano	0,85	19,56 %
Široký	Ne	0,95	11,53 %
Široký	Ano	0,95	11,53 %
Úzký	Ne	0,95	19,50 %
Úzký	Ano	0,95	19,50 %

Obrázek 6.3: Výsledky měření efektivity variant algoritmu PageRank

Výsledky provedených testů jsou pak k dispozici v souborech `test_results/page_rank_aggregated_results.csv` (ve stručné podobě) a `test_results/page_rank_results.csv` (podrobné výsledky rozepsané podle jednotlivých metod a dotazů). Výsledky, kterých jednotlivé metody dosáhly, jsou uvedeny na obrázku č. 6.3).

Tyto výsledky naznačují, že *PageRank* poskytuje výsledky nanejvýš srovnatelné s metodami založenými výhradně na informacích o rozhodnutí samotném (statická váha), a výrazně zaostává za metodami vycházejícími ze statického počtu citací. Rovněž se ukázalo, že ze tří zkoumaných atributů je výrazněji schopen ovlivnit chování metody pouze rozsah grafu. Varianty, které posuzování relativního významu rozhodnutí prováděly pouze nad množinou rozhodnutí splňující podmínky zadaného dotazu, opakovaně dosahovaly výrazně lepších výsledků, než varianty založené na širším přístupu.

Lze tedy konstatovat, že význam potenciálních dodatečných informací, jež mohou pro daný dotaz nerelevantní rozhodnutí přinést, zpravidla nepřeváží nad problémy s tím spojenými. Teoreticky by snad šlo uvažovat o tom, že vazby z nerelevantních rozhodnutí budou zohledňovány s výrazně nižší vahou. Vzhledem k tomu, jakých výsledků dosáhly varianty užívající vážený *PageRank*, však ani na toto případné řešení nelze příliš spoléhat.

Ze získaných dat však nelze přímo zjistit, jaké jsou důvody neúspěchu váženého *PageRanku*. Zčásti to může být důsledek použití vah v rozporu s jejich smyslem, jímž bylo odlišit od sebe význam odkazů v rámci jednoho dokumentu (což je zjevné z toho, že se jedná o váhy hran, nikoliv uzlů). Vzhledem k tomu, že pro posouzení významu jednotlivých odkazů v rámci dokumentů jsem však neměl dostatečné nástroje, nastavil jsem váhy podle

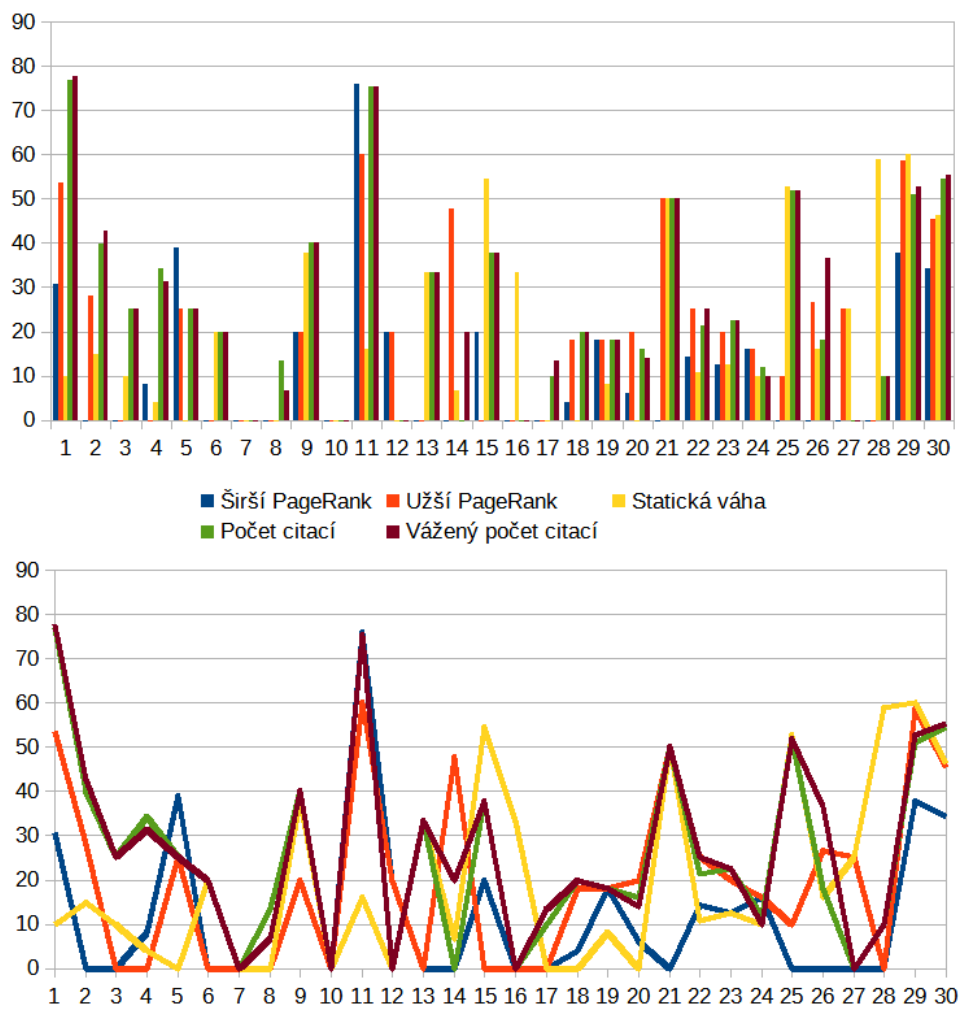
významu rozhodnutí jako celku. A byť jsem věděl, že výsledky velmi pravděpodobně nebudou na takové úrovni, jako kdybych měl k dispozici informace u významu jednotlivých odkazů, je tento výsledek značně překvapivý.

Totéž přitom platí i pro minimální dopad změny *damping factoru*. Kromě dvou v tabulce uvedených hodnot jsem vyzkoušel desítky dalších, a nikdy to nevedlo k větší změně výsledků než v řádech desetin procentního bodu. Hypotéza o tom, že použitím vysokých hodnot tohoto faktoru bude dosahováno lepších výsledků, tedy rovněž nebyla potvrzena.

6.6 Vliv jednotlivých případů

Na obrázku č. [6.3](#) je možné vidět výsledky testů pěti vybraných metod (těch, které vykazovaly nejvýznamnější rozdíly oproti ostatním) na jednotlivých případech vyhledávání. Na ose y je skóre v procentech, na ose x pak jednotlivá vyhledávání.

Z těchto výsledků je zřejmé, že žádná z uvedených metod nevykazuje chování, z něžž by bylo možné usuzovat na výjimečnost některých vyhledávacích podmínek. Rovněž není patrné, že by některé metody pravidelně vykazovaly lepší výsledky v případech, kdy jiné metody selhávají. Nelze tedy reálně uvažovat o tom, že by rozdíly ve formě či obsahu soudních rozhodnutí měly z hlediska použitých metod nějaký zásadní dopad (například, že jedna metoda by vykazovala výraznější úspěchy v oblasti trestněprávní či správněprávní a jiná zase v civilní).



Obrázek 6.4: Výsledky měření rozdělené podle jednotlivých vyhledávání

Závěr

Cílem této práce bylo přispět ke snaze o zefektivnění vyhledávání v judikatuře vrcholných soudů České republiky. Toho mělo být dosaženo identifikováním metod, které by byly schopny rozpoznat významná rozhodnutí v záplavě těch bezvýznamných lépe, než do dokáží v současné době existující nástroje.

Za tímto účelem byla provedena podrobná analýza právních aspektů soudního rozhodování se zaměřením na to, jak se soudní postupy a procedury promítají do způsobu rozhodování a do samotného textu jednotlivých rozhodnutí. Na základě těchto zjištění byly identifikovány prvky rozhodnutí, které svědčí o jeho významu, a byl posouzen význam těchto dílčích indicií pro závěr o míře důležitosti jednotlivých rozhodnutí.

Následně byla zhodnocena současná situace z hlediska dostupnosti soudních rozhodnutí v elektronické podobě. Byly představeny způsoby, jakými vrcholné soudy České republiky svá rozhodnutí zveřejňují. U těchto soudních systémů byly rovněž stručně popsány přednosti a nedostatky jejich vyhledávacích nástrojů.

Vycházející ze zjištění ohledně dostupnosti soudních rozhodnutí v elektronické podobě byly navrženy metody, jak obsah uvedených rozhodnutí automaticky stáhnout a převést do strojově dobře zpracovatelného formátu. Tyto metody byly následně úspěšně implementovány. Data získaná tímto postupem byla následně prozkoumána, ošetřena, vzájemně propojena na základě identifikovaných vazeb a uložena do databáze pro další použití.

Při znalosti rozsahu dostupných dat byly následně navrženy metody, které posoudí relativní význam získaných rozhodnutí a identifikují mezi nimi ta důležitá. Každá z navržených metod upřednostňovala jako klíčovou jinou oblast dostupných informací – to jest získaná metadata, obsahy rozhodnutí či jejich vzájemné vazby – s cílem umožnit hodnocení přínosnosti těchto

kategorií pro posouzení významu rozhodnutí.

Navržené metody byly implementovány v programovacím jazyce Python a otestovány. Na základě takto získaných informací byla potvrzena úzká vazba mezi počtem citací rozhodnutí a jeho významem. Jistých úspěchů dosáhla i metoda opírající se pouze o metadata rozhodnutí samotných. Další zkoušená metoda – *PageRank* – naopak vykazala neuspokojivé výsledky.

Na základě učiněných zjištění tak lze konstatovat, že vzájemné vztahy mezi soudními rozhodnutími představují perspektivní směr při snaze o zlepšování efektivity vyhledávání v judikatuře. Plné využití jejich potenciálu však vyžaduje mnohem podrobnější zpracování a přípravu dostupných dat, které dalece přesahuje náplň této práce.

V rámci této práce byly rovněž identifikovány některé parametry rozhodnutí, které mohou hrát při posuzování jeho významu zásadní roli, avšak s jejich vytěžením jsou spojeny značné obtíže. Pro zefektivnění vyhledávání důležitých rozhodnutí však bude nejspíš nezbytné i tyto znalosti využít. Jejich zpracování a použití tedy představuje přirozený další krok na cestě ke zvýšení efektivity české justice a předvídatelnosti soudního rozhodování.

Literatura

BRIN, S.; PAGE, L. *The anatomy of a large-scale hypertextual Web search engine*. Computer Networks and ISDN Systems. 30 (1–7): 107–117. doi:10.1016/S0169-7552(98)00110-X. ISSN 0169-7552.

DOSTÁL, M.; LANGÁŠEK, T.; POSPÍŠIL, I., WAGNEROVÁ, E. *Zákon o Ústavním soudu: Komentář*. Wolters Kluwer [cit. 2021-6-22]. ISSN 2336-517X.

KONVALINKA, Jan. „Kafemlejnek“ na českou vědu aneb k financování vědy. *Živa*. 2009, 2009(4), XLIX - LXVIII.

PAGE, L.; BRIN, S.; MOTWANI, R.; WINOGRAD, T. 1999. *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report. Stanford InfoLab.

TING, K.M. 2011 Precision and Recall. In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA. ISBN 978-0-387-30768-8.

Asiri, Sidath. *Machine Learning Classifiers* [online] [cit. 26.6.2021]. Dostupné z: <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623>.

Clarivate. *The Clarivate Analytics Impact Factor* [online] [cit. 24.6.2021]. Dostupné z: <https://clarivate.com/webofsciencegroup/essays/impact-factor/>

Crummy.com. *Beautiful Soup Documentation*. Dostupné z: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

ZÁVĚR

[//www.crummy.com/software/BeautifulSoup/bs4/doc/](https://www.crummy.com/software/BeautifulSoup/bs4/doc/).

Developer Economics. *What is the best programming language for Machine Learning?* [online] [cit. 24. 6. 2021]. Dostupné z: <https://towardsdatascience.com/what-is-the-best-programming-language-for-machine-learning-a745c156d6b7?>.

Django Software Foundation. *Django documentation* [online] [cit. 24.6.2021]. Dostupné z: <https://docs.djangoproject.com/en/3.2/>.

Elsevier. *CiteScore Journal Metric - FAQs* [online] [cit. 24.6.2021]. dostupné z: https://service.elsevier.com/app/answers/detail/a_id/30562/supporthub/scopus/

Google. *Facts about Google and Competition* [online] [cit. 24.6.2021]. Dostupné z: <https://web.archive.org/web/20111104131332/https://www.google.com/competition/howgooglesearchworks.html>

Chonyy. *PageRank: Link Analysis Explanation and Python Implementation from Scratch* [online] [cit. 24.6.2021]. Dostupné z: <https://towardsdatascience.com/pagerank-3c568a7d2332>.

Justin. *Personalized PageRank with Edge Weights* [online] [cit. 24.6.2021]. Dostupné z: <https://www.briggsby.com/personalized-pagerank>

Khan Academy. *Random walks* [online] [cit. 24.6.2021]. Dostupné z: <https://www.khanacademy.org/computing/computer-programming/programming-natural-simulations/programming-randomness/a/random-walks>

Nejvyšší soud. *Nápověda k vyhledávací rozhodnutí Nejvyššího soudu* [online] [cit. 21.6.2021]. Dostupné z: https://www.nsoud.cz/Judikatura/ns_web.nsf/web/Napoveda~Vyhledavanirozhodnuti~Obsah_databaze_rozhodnuti_a_stanovisek_Nejvyssiho_soudu~?openDocument&lng=CZ.

Nejvyšší soud. *Vyhledávání rozhodnutí: Zadávání dotazů pro pokročilé* [online] [cit. 21.6.2021]. Dostupné z: https://www.nsoud.cz/Judikatura/ns_web.nsf/web/Napoveda~Vyhledavanirozhodnuti~Vyhledavani_rozhodnuti__Zadavani_dotazu_pro_pokrocile~?openDocument&lng=CZ.

NetworkX Developers. *pagerank* [online] [cit. 24. 6. 2021]. Dostupné z: https://networkx.org/documentation/networkx-1.10/reference/generated/networkx.algorithms.link_analysis.pagerank_alg.pagerank.html

Oracle. *MySQL Documentation* [online] [cit. 24. 6. 2021]. Dostupné z:

<https://dev.mysql.com/doc/>.

Python Software Foundation. *Python documentation* [online] [cit. 24. 6. 2021]. Dostupné z: <https://www.python.org/doc/>.

Shinyama, Yusuke. *PDFMiner: Python PDF parser and analyzer dokumentace* [online] [cit. 21.6.2021]. Dostupné z: <https://www.unixuser.org/~euske/python/pdfminer/index.html>.

Wolters Kluwer. *Tipy pro práci s ASPI* [online] [cit. 21.6.2021]. Dostupné z: https://static.wolterskluwer.cz/content/mediagallery/wkshop_system/file/product/file/18028.pdf.

Zyte team. *What is web scraping?* [online] [cit. 24. 6. 2021]. Dostupné z: <https://www.zyte.com/learn/what-is-web-scraping/>.

Ústavní zákon č. 1/1993 Sb., Ústava České republiky

Ústavní zákon č. 2/1993 Sb., Listina základních práv a svobod

Zákon č. 141/1961 Sb., o trestním řízení soudním (trestní řád)

Zákon č. 99/1963 Sb., občanský soudní řád

Zákon č. 182/1993 Sb., o Ústavním soudu.

Zákon č. 284/1993, kterým se mění a doplňuje zákon České národní rady č. 436/1991 Sb., o některých opatřeních v soudnictví, o volbách přisedících, jejich zproštění a odvolání z funkce a o státní správě soudů České republiky, ve znění zákona České národní rady č. 580/1991 Sb., zákona České národní rady č. 23/1993 Sb. a zákona č. 171/1993 Sb.

Zákon č. 106/1999 Sb., o svobodném přístupu k informacím.

Zákon č. 6/2002 Sb., o soudech, soudcích, přisedících a státní správě soudů a o změně některých dalších zákonů (zákon o soudech a soudcích)

Zákon č. 131/2002 Sb., o rozhodování některých kompetenčních sporů

Zákon č. 150/2002 Sb., soudní řád správní

Zákon č. 292/2013 Sb., o zvláštních řízeních soudních

Zákon č. 218/2021 Sb., kterým se mění zákon č. 6/2002 Sb., o soudech, soudcích, přísedících a státní správě soudů a o změně některých dalších zákonů (zákon o soudech a soudcích), ve znění pozdějších předpisů, a další související zákony.

Nález Ústavního soudu ze dne 18. 2. 2020, sp. zn. I. ÚS 2002/19. Dostupné na: nalus.usoud.cz

Nález Ústavního soudu ze dne 15. 7. 2004, sp. zn. II. ÚS 594/02. Dostupné na: nalus.usoud.cz

Nález Ústavního soudu ze dne 24. 4. 2014, sp. zn. IV. ÚS 2265/13. Dostupné na: nalus.usoud.cz

Nález Ústavního soudu ze dne 19. 12. 2019, sp. zn. I. ÚS 3073/19. Dostupné na: nalus.usoud.cz

Nález Ústavního soudu ze dne 2. 9. 2020, sp. zn. IV. ÚS 4156/19. Dostupné na: nalus.usoud.cz

Usnesení Nejvyššího soudu ze dne 29. 5. 2002, sp. zn. 5 Tdo 86/2002. Dostupné na nsoud.cz

Usnesení Nejvyššího správního soudu ze dne 26. 4. 2006, č. j. 1 Azs 13/2006 - 39. Dostupné na: nssoud.cz

Seznam použitých zkratk

html Hypertext markup language

rtf Rich text format

doc Document

pdf Portable documentformat

url Uniform Resource Locator

ID Identifikátor

Obsah přiloženého CD

```
/
├── readme.txt ... stručné shrnutí obsahu CD
├── src ... primárně zdrojové kódy
│   ├── judikaty ... ukázky souborů stažených se soudních webů
│   ├── website ... soubory spojené s během webového serveru
│   │   ├── daemons ... klíčové soubory projektu
│   │   │   └── matches ... výsledky vyhledávání pro zadané dotazy
│   │   ├── main ... adresář obsahující část konfigurace webového
│   │   │   serveru a jeho (v této práci nevýznamnou) vnější část
│   │   ├── website ... hlavní definice pro fungování webserveru
│   │   └── manage.py ... Djangoem poskytnutý script umožňující
│   │       správu webserveru
│   └── settings.ini ... soubor definující některé parametry
│       užívaných metod
├── test_results ... adresář obsahující výsledky provedených
│   testů
├── text práce ... adresář obsahující soubory textu práce
│   └── prace.pdf ... text práce ve formátu pdf
```

Návod k použití zdrojových kódů

Tato příloha obsahuje instrukce ke zprovoznění přiloženého kódu.

C.1 Požadavky

Pro řádný provoz programu je potřeba následující:

- Python 3 (ideálně ve verzi 3.6.1)
- Django (ideálně ve verzi 2.0.2)
- Mysql datbázový sevrver

C.2 Postup instalace

Poté, co se ujistíte, že disponujete výše uvedenými nástroji, překopírujte obsah CD na takovou část disku, k níž máte právo zápisu.

Poté může být nutné upravit konfiguraci serveru. Ta se nachází v souboru `src/website/website/settings.py`. Zde se na řádce 81 nachází položka `DATABASES`, definující připojení k MySQL serveru. Upravte tedy v tomto souboru data tak, aby odpovídala vašemu nastavení serveru. Pokud jste dosud ještě nevytvořili databázi pro tento projekt a uživatele (s plnými právy k dané databázi), přes kterého k ní budete přistupovat, učiňte tak. Rovněž v souboru `src/settings.py` můžete upravit cestu ke složkám, do nichž budou ukládány soubory stažené z webů jednotlivých soudů.

Nyní je načase vytvořit samotnou databázovou strukturu. Otevřte tedy příkazový řádek a přejděte v něm do složky `src/website`.

Nyní zavolejte příkaz `python manage.py makemigrations` a po jeho dokončení příkaz `python manage.py migrate`. Po dokončení obou příkazů můžete spustit webový server. ten je však nutné spouštět z nadřazené složky. Přejděte tedy do nadřazené složky `src` a příkazem `python website/manage.py runserver 0.0.0.0:8001` (port záleží na vašem uvážení) spusťte server. Nyní můžete přistoupit na domovskou stránku. V současném stavu toho na serveru příliš dělat nemůžete, protože v databázi zatím nejsou žádná rozhodnutí.

C.3 Získávání rozhodnutí

Ke stažení rozhodnutí slouží následující trojice příkazů (které je nutné rovněž spouštět z příkazové řádky ve složce `src`):

- `python website/daemons/nsoud.py` pro získání rozhodnutí Ústavního soudu.
- `python website/daemons/nsoud.py` pro získání rozhodnutí Nejvyššího soudu.
- `python website/daemons/nssoud.py` pro získání rozhodnutí Nejvyššího správního soudu.

Před spuštěním těchto příkazů se ujistěte, že máte na disku dost místa (alespoň 11 GB). Rovněž počítejte s tím, že pokud budete chtít stáhnout všechna rozhodnutí od roku 2004, poběží uvedené programy v případě Nejvyššího a Ústavního soudu zhruba 10 či více hodin; v případě Nejvyššího správního soudu pak program nejspíš poběží několik dní. Příkazy je možné nechat běžet paralelně.

C.4 Práce s databází

Poté, co budete mít stažena rozhodnutí, s nimiž chcete pracovat, můžete se vrátit k webovému serveru. Pokud jste jej v mezidobí vypnuli (což bych vám vzhledem k době stahování rozhodnutí doporučil), můžete jej znovu zapnout opětovným voláním příkazu `python website/manage.py runserver 0.0.0.0:8001` ze složky `src`.

Domovská stránka víceméně obsahuje všechny nástroje, které server v současné době umožňuje, a to v pořadí, které respektuje jejich vzájemnou provázanost (tedy zejména skutečnost, že některé z funkcí ke své řádné činnosti vyžadují výsledky funkcí předchozích). Volání každé funkce je realizováno kliknutím na příslušné tlačítko. Stránka se bude po dobu zpracovávání funkce „načítat“. T, že funkce byla dokončena, tedy poznáte tak, že stránky

se donačte.

DŮLEŽITÉ UPOZORNĚNÍ: Buďte trpěliví a nespouštějte více funkcí najednou, ani funkce nespouštějte opakovaně před jejich dokončením! Taková akce totiž jen přeruší výpočet předchozí funkce a začne provádět výpočet nový. Pokud budete pracovat se stejným rozsahem dat jako já (tedy se všemi rozhodnutími od roku 2004), bude zcela běžné, že mnohé z funkcí budou probíhat desítky minut či jednotky hodin.

Hlavní stránka webu víceméně uvádí, co funkce dělají, a kde najít jejich případné výsledky. Podrobnější detaily přitom byly uvedeny v předchozích částech této práce.