

# ZKUŠENOSTI S AUTOMATIZACÍ CITAČNÍ ANALÝZY NA ZAKONECHPROLIDI.CZ

21.9.2017

ČPIT 2017

Pavel Gardavský / AION CS, s.r.o

# ○ čem budu mluvit?

2

- ZPL a citační analýza, typy, čísla, příklady
- Principy trochu podrobněji
- Výhled dalšího využití citační analýzy do budoucna
- Drobné překvapení na závěr 😊

# Zákony pro lidi a citační analýza

3

- strojově prováděná citační analýza je pro ZPL jedním ze základních stavebních kamenů
  - ▣ prostřednictvím citační analýzy lokalizujeme sémantické entity podle předem daných pravidel
  - ▣ pro lokalizované sémantické entity nalézáme cíle
    - používáme při vyhledávání
    - tvoříme dvojice zdroj  $\leftrightarrow$  cíl, propojujeme je obousměrnými vazbami, které navenek vypadají jako odkazy
- postupujeme (až na výjimky) automaticky

## 2 detaily ...

4

### 1. Fragmentální datový model

- Princip 1. – rozklad předpisů na fragmenty
- Princip 2. – skládání fragmentů do hierarchií předpisů
- Princip 3. – asociační vazby
- Princip 4. – konsolidační vazby

### 2. Vše je zapsáno v databázi

- jednotlivé fragmenty
- hierarchie fragmentů
- různé typy vazeb

# Pár čísel

## □ externí odkazy

- ▣ odkazy na celé předpisy
  - 338.291 unikátních
  - 1.629.423 přes všechna znění
- ▣ odkazy na celé předpisy „ve znění“
  - 149.223 unikátních
  - 800.802 přes všechna znění
- ▣ odkazy na části jiných předpisů
  - 128.785 unikátních
  - 608.972 přes všechna znění
- ▣ odkazy prováděcích předpisů
  - 15.093 unikátních
  - 44.952 přes všechna znění

## □ interní odkazy

- ▣ 339.261 unikátních
- ▣ 2.329.362 přes všechna znění

## □ odkazy na EUR-Lex

- ▣ 52.928 unikátních
- ▣ 212.684 přes všechna znění

# Pár příkladů

## □ Externí odkazy:

- [Poznámky pod čarou Zákoníku práce 262/2006 Sb.](#)
- [Záložka Souvislosti ...](#)

## □ Interní odkazy:

- [zakonyprolidi.cz/cs/2006-262#p363](#)

## □ Využití citační analýzy při hledání

|                                                               |                                  |   |
|---------------------------------------------------------------|----------------------------------|---|
| 262/2006                                                      | paragraf 25 odstavec 2 písmeno a | Q |
| lokálně na § 25 odst. 2 písm. a) - a) zaměstnavatele, kteří j |                                  |   |

## □ Kuriozita – [skript analýzy citací nad EUR-Lexem](#)

# Podrobněji ...

## Dvoufázový systém propojování sémantických entit

7

1. Extrakce entit (= lokalizace a označkování)
  - ▣ předpis (89/2012 Sb., daňový řád,..)
  - ▣ ustanovení (§1 bod 2, §1 až §5,..)
    - paragraf, článek, hlava, odstavec, písmeno, bod, věta, část, kapitola,..
  - ▣ instituce (vláda, ministerstvo spravedlnosti, Česká národní rada,...)
  - ▣ typ (zákon, vyhláška, opatření,..)
  - ▣ datумы
2. Identifikace vztahů zdroj  $\leftrightarrow$  cíl (= nalezení cíle)
  - ▣ výsledkem je
    - změna textu na prokliknutelný odkaz
    - zápis oboustranné vazby
  - ▣ určení časového řezu zákona
    - “ve znění”  
 (“zákon 89/2012 ve znění novely 298/2016 Sb.”  $\rightarrow$  “ve znění od 1.1.2012”)
      - je třeba zjistit účinnost zdrojového ustanovení
    - vyřešení mnohoznačnosti textových citací zákonů  
 (“Občanský zákoník” - 40/1964 Sb. vs. 89/2012 Sb.)
  - ▣ identifikace ustanovení v časovém řezu cílového zákona

# A ještě podrobněji ...

## 1. fáze: Extrakce entit

8

schopnost identifikovat

- posloupnosti citovaných ustanovení
  - ▣ “§1, §3 písm. a), §5 odst. 1 a 2, §10”
- rozsahy citovaných ustanovení
  - ▣ “§1-§5”, “Hlava X až XV”, “§5 a násl.”
- dynamicky řízené pravidla
  - ▣ editory kontrolované zkratky zákonů “TrZ”, „NOZ“
- kombinace předcházejících
  - ▣ “§1 písm. b) až c), §2, § 4, §6-9 a §15 TrZ”
- posloupnosti předpisů
  - ▣ obsluha sekvence “ve znění”



# A ještě podrobněji ...

## 2. fáze Identifikace předpisů a ustanovení

9

context aware system

- analýza určení vztahů bere v potaz stromovou sktrukturu “zdrojového” textu
  - ▣ odkazy na předcházející odstavce (chybí kompletní jednoznačná sémantická adresa citovaného ustanovení: “§1 odstavec 3” vs. “odstavec 3”)
- rozsahy typu “§1 a násl.” (stromová struktura cíle)
- identifikace citací “tohoto zákona”,
- identifikace definovaných zkratk v textu “89/2012 (dále jen ‘zákon’)”

# A úplně nejpodrobněji:

## Podstata věci – extraktor entit

10

- řízený vlastním regex engine
- možnost definice kombinací (vnořování) předešle pojmenovaných výrazů
  - ▣ vnořené části
    - “§2 odstavec 3 písmeno a) až odstavec 4 písmeno d)”
    - sada výrazů pro písmena je znovupoužitelná pro jiné výrazy (odstavce, paragraf)
- možnost vlastních maker pro detekci
- schopnost extraktoru správně interpretovat
  - ▣ numerické, slovní a římská čísla
  - ▣ datумы v libovolném formátu (číselné a slovní vyjádření jednotlivých složek)

# A kudy dál s citační analýzou na ZPL?

11

## □ Především analyzovat další dokumenty

### ▣ soudní rozhodnutí

- podle „stále zelených principů“ popsanych panem Mgr. Jaromírem Šavelkou před několika lety [na JinémPrávu](#)
- jak to bude vypadat:
  - [§ 52 písm. c\) zákona č. 262/2006 Sb. zákoník práce](#)

# Překvapení ...

12

- ❑ Dnes zprovozněna na počest ČPIT nová funkčnost ZPL
- ❑ Monitor změn
  - ▣ sledování práce našich zákonodárců
  - ▣ informace „na stříbrném podnose“ 😊

