

# PONK

## Asistent přístupné úřední komunikace

Tereza Novotná

novotna@ufal.mff.cuni.cz

České právo a informační technologie

Právnická fakulta MUNI Brno

13. září 2024



Univerzita Karlova  
Matematicko-fyzikální fakulta  
Ústav formální a aplikované lingvistiky



unless otherwise stated

- PONK
- projekt č.
- poskytovatel
- doba řešení
- řešitel
- aplikační garant
- web

Psaní Orientované Na Klienta  
TQ01000526

TAČR

9/2023-12/2025, **jsme v polovině**

MFF UK

Frank Bold Society, z.s.

<https://ufal.mff.cuni.cz/grants/ponk>

Vytvořit webovou aplikaci pro automatické hodnocení srozumitelnosti českých právních textů s využitím metod umělé inteligence a ručních pravidel.

- **Srozumitelnost** posuzujeme v těchto aspektech
  - předvídatelnost - míra překvapení při čtení textu
  - stylistika - stylistická správnost textu
  - právní argumentace - relevance a koherence textu
- Metody umělé inteligence budeme trénovat na korpusu **KUK**
- Jako vedlejší výsledek projektu vzniká nástroj **MasKIT** pro pseudonymizaci a anonymizaci českých právních textů



# Lexikální překvapení

## Definice

- Situace, kdy čtenář narazí na slovo či slovní spojení, které je **nečekané** nebo **neobvyklé** v daném kontextu.

## Příčiny

- Neznámé slovo
- Neobvyklé použití slova
- Výraz netypický pro danou situaci či text

## Dopad na porozumění

- Může zpomalit porozumění
- Vyžaduje větší kognitivní úsilí
- Může vést k nepochopení textu a chybnému jednání

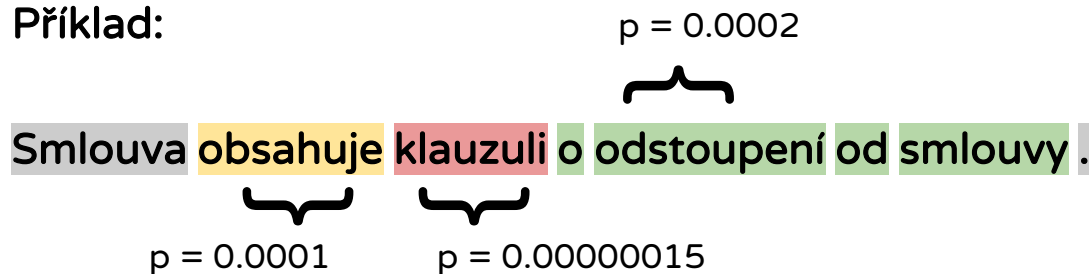
# Lexikální překvapení

## PONK-GLITTER

- Modul aplikace PONK
- Použití jazykových modelů pro výpočet pravděpodobnosti/entropie textu

## Pravděpodobnost následujícího slova ~ lexikální překvapení

Příklad:



# Gramatická pravidla & stylometrické indikátory

## Gramatická pravidla

- Detekujeme **výskyty jevů**, které by mohly **zhoršovat srozumitelnost**
- Inspirujeme se příručkami srozumitelného právního a akademického psaní
- Např. trpný rod, pozice přísudku, podstatná jména slovesná, přebytek negací, ...

## Stylometrické indikátory

- **Číselné** “zhodnocení”
- Vztahují se na **text jako celek**
- Např. type-token ratio, aktivita, Gunning fog index, ...

Používáme nástroj **UDPipe** pro větný rozbor textů (<https://lindat.cz/services/udpipe/>).

# Gramatická pravidla & stylometrické indikátory

## Příklad

Navrhovatelé dále vyjádřili názor, že zásady územního rozvoje mohou **být** soudem **přezkoumány** již v okamžiku jejich **vydání**, **neboť** od tohoto okamžiku „právně **existují**“, a není tak potřeba čekat na **nabytí** účinnosti.

## Detekovali jsme

- Trpný rod opisný
- Podstatná jména slovesná
- Přísudek daleko ve větě

## Naměřené indikátory

- Activity: 0.29
- Entropy: 8.71
- TTR: 0.16
- Gunning Fog: 11.86

# Anotace právní argumentace

## Relevance



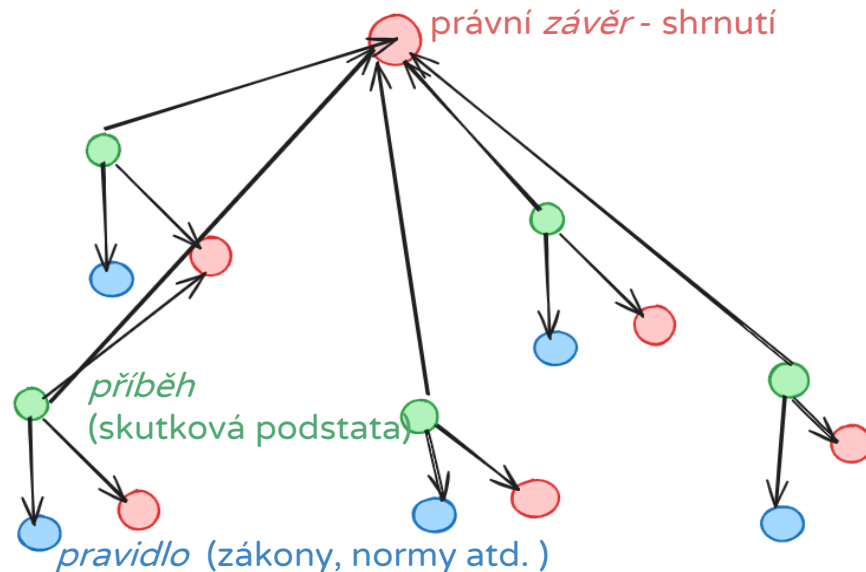
nesrozumitelné

zbytečné

relevantní

## Koherence

Každý **právní závěr** má být podložen **příběhem** a **pravidlem**, nebo trojicí závěru, příběhu a pravidla.





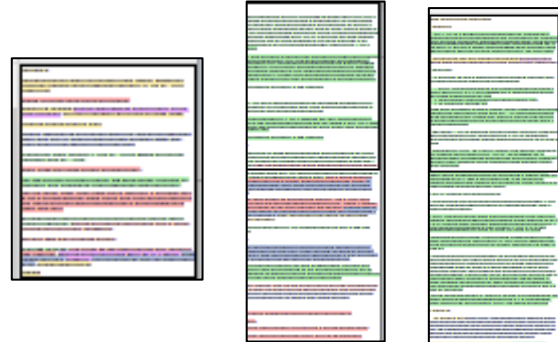
# Anotace právní argumentace

## Anotační nástroj Gloss (Jaromír Šavelka)

### Relevance

Types 4 Annotations 41		
1_Nesrozumitelné	Komise předně podotýká, že p... neznámá, že žadatelé o dlou... musí mimo realizaci formálních... zejména k naplnění obligatorn... dostatečně vysvětlit svoji moti... mají žadatelé o dlouhodobé vi... nezbytných k individuálnímu p...	šech náležitost... um bude toto... apř. předlože... sti žádosti o d... ž k ověření tě... st zodpověděl... jich konkrétní... h požadavků... onných důvo...
2_Zbytečné		
3_Relevantní	víza nepostačí pouhé splnění... vnitřní musí rovněž vyloučit pří... uvedených v § 56 zákona č. 3...	
komentář anotátora	Komise upřesňuje, že pohovo... v souladu se zákonem č. 326/... zjištění objektivních skutečnos...	žadatelem ná... dná se o star... i je při pohovc...

### Koherence



Data	Text
Types 8 Annotations 19	ArgMOmbu032024_odBz
Komentář o čemkoli	Již v době objednání díla žalobcem měl dlužník peněžité závazky 30 dní po lhůtě splatnosti, jak vyplývá z přihlášek pohledávek věř... specifikovaného insolvenčního řízení vedeného proti dlužníku (n... vedené pod čísly přihlášek P4, P5, P7, P8, P9, P11, P12 atd.), p... pochybností, že tyto své závazky nebyl schopen splnit. Žalobce m... žalobcovy objednávky díla byl dlužník předlužen, soudě dle pomě... likvidátora dlužníka ze dne 30. 7. 2013 vyplývá, že aktiva dlužníka... pohledávkami za třetími osobami, jejichž vymahatelnost označil likvidátor... pasiva společnosti byla dle účetní závěrky přiložené ke sdělení likvidátora v obd... než dospět k závěru předlužení, a to vzhledem k nutnosti posuzování aktiv dle je... účetní hodnoty, neboť pokud jsou pohledávky dlužníka za třetími osobami obližn... jejich tržní hodnota značně nižší než uvedená hodnota účetní.
Metatext	
Pravidlo	
Právní otázka	
Příběh	
Rada	Žalobce je názoru, že žalovaný, toho času jednatel dlužníka (ve funkci od 1.1.20... blíže viz obchodní rejstřík), porušil svou povinnost stanovenou v ustanoveních §... zákona č. 182/2006 Sb., o úpadku a způsobech jeho řešení (dále jen „InsZ“), ve... předpisů, tedy povinnost statutárního orgánu podat insolvenční návrh bez zbyte... poté, co se dozvěděl nebo při náležitě pečlivosti měl dozvědět o úpadku dlužníka... povinnosti žalovaného podat insolvenční návrh žalobce vyzvojuje z následujících
Výzva	
Závěr	

# Anotace právní argumentace

## Koherenční vztahy

- Anotátor vychází z příběhu.
- Segmentuje příběh tak, aby ho co nejlépe propojil s pravidly a závěry.
- Pole bez vztahů: metatext, rady, výzvy

[Barbora Kubíková] Porušení povinnosti žalova...

[Barbora Kubíková] Důkazy: sdělení likvidátor...

[Barbora Kubíková] V situaci, kdy pohledávka ...

[Barbora Kubíková] Pokud by žalovaný danou ...

author: Barbora Kubíková  
type: Příběh  
start: 4776  
end: 5181

pravidla: V situaci, kdy pohledávka věřitele vznikla až při p... (Barbora Kubíková)  
Ustanovení § 99 InsZ pak stanoví, že osoba, která p... (Barbora Kubíková)

závěry: přičemž je bez jakýchkoliv pochybností že tyto své ... (Barbora Kubíková), jelikož byl ale žalovaný v prodlení se splněním své ... (Barbora Kubíková), \*\*V.\* \*\*Vzhledem k vyše uvedenému žalobce navrhl ... (Barbora Kubíková)

zbytečného odkladu poté žalovaného podat insolvenční návrh.

Důkazy: sdělení likvidátora.

\*\*IV.\*\*

V situaci, kdy pohledávka vznikla až při prodlení se splněním své povinnosti Nejvyššího soudu (viz nálezení Nejvyššího soudu, kterou věřitel posléze obdržel, dlužníkem představuje dlužník).

Pokud by žalovaný danou povinností uzavřel za zcela odlišných okolností (jakýchkoliv vad), případně v rámci specifickovanou povinností.

Jelikož byl ale žalovaný v prodlení se splněním své povinnosti, pak stanoví, že osoba, která v prodlení se splněním své povinnosti způsobila jinou újmu, kterou způsobila.

# Anotace právní argumentace



Originál:  
mísení segmentů  
v rychlém sledu, často uvnitř  
jedné věty



Přepis:  
často stručnější, méně  
střídání segmentů



MasKIT je webový nástroj a služba REST API pro pseudonymizaci a anonymizaci českých právních textů:

1. Zpracuje vstupní text (např. dopis zaslaný právním orgánem občanovi) pomocí nástrojů UDPipe a NameTag
2. V textu identifikuje a nahradí osobní informace

K dispozici na adrese <https://quest.ms.mff.cuni.cz/maskit/run.php>

**NameTag** rozpoznává jména osob, institucí, geografické názvy a jiné jmenné entity v textu (<https://lindat.cz/services/nametag/>).

## MaskIT

 [O programu](#)

 [Spustit](#)

 [Dokumentace REST API](#)



▼ MaskIT je webový nástroj a služba REST API pro pseudonymizaci a anonymizaci českých právních textů.

▼ Informace o serveru MaskIT: verze: 0.64 20240820, stav: *online*

**Vstup:**  Prostý text  Předrozdělený (každá věta na samostatném řádku)

**Výstup:**  TXT (s označením pomocí speciálních znaků)  HTML (s barevným zvýrazněním)

**Další volby:**  Náhodná volba při nahrazování  Použití tříd při nahrazování


 Vstupní text

[Smazat vstupní text](#)

Paní Marie Nováková z Myslíkovy ulice č. 25 dostala dopis od firmy Škoda.

↓ Zpracovat vstup ↓

původní  
 barvy

 Výstup

[Uložit](#)

 Další informace

[Uložit](#)

Paní **Anna**<sub>[Marie]</sub> **Svobodová**<sub>[Nováková]</sub> z **Polní**<sub>[Myslíkovy]</sub> ulice č. **29**<sub>[25]</sub> dostala dopis od firmy **ProfiServis**<sub>[Škoda]</sub>.

# MaskIT - podporované vlastnosti

▼ Informace o serveru MaskIT: verze: 0.64 20240820, stav: *online*

## Informace o serveru MaskIT

- verze: 0.64 20240820, stav: *online*
- podporované vlastnosti: *first names (not judges) • surnames (not judges, male and female tied) • phone numbers (non-emergency) • e-mails • www addresses • street names (incl. multiword) • street numbers • town/town part names (incl. multiword; not places of court) • ZIP codes • company names (incl. multiword) • cult./educ./scient. institutions (incl. multiword) • commercial register number (IČO) • tax register number (DIČ) • land registration numbers (registrační čísla pozemků) • birth registration numbers (rodná čísla) • dates of birth/death • agenda reference numbers (čísla jednací) • vehicle registration numbers (SPZ)*

Pro testovací účely je služba poskytována zdarma, za [těchto podmínek užívání](#). Pro komerční využití softwaru je požadován výslovný písemný souhlas autorů MaskITu a volaných externích služeb ([UDPipe](#) a [NameTag](#)).

## Anonymita

V zájmu zachování anonymity na tomto serveru MaskIT ani externí služby UDPipe a NameTag neukládají zpracovávané texty. Jediné informace, které mohou být ukládány, jsou: čas použití, velikost zpracovávaných dat, konfigurace systému a IP adresa, ze které bylo k systému přistupováno; pokud službu používáte prostřednictvím rozhraní REST API, ujistěte se, že používáte metodu POST, protože adresy URL rozhraní REST API mohou být také ukládány.

Všechny připomínky a reakce jsou vítány.

- Korpus Úřední Komunikace
- České právní a administrativní texty pro trénování metod umělé inteligence
- Verze 0.0
  - 6 051 dokumentů, 19 196 037 slov
    - právní analýzy a zodpovězené dotazy z databáze Frank Bold
    - stanoviska a letáky vydané Kanceláří veřejného ochránce práv
    - soudní rozhodnutí tří nejvyšších českých soudů
    - právní texty v několika parafrázích ze studie srozumitelnosti českých administrativních textů
  - ke stažení v repozitáři LINDAT/CLARIAH-CZ [url](#)
  - k prohlížení v nástroji TEITOK [url](#)

# KUK 0.0 - metadata

**Legend**  
 PK - primary key  
 PFK - primary foreign key

DocumentFileFormat	
KUK_ID (PFK)	Text
FileName	Text
FileFormat	Text
FolderPath	Text

ContentLinks	
KUK_ID (PFK)	Text
RefersTo	Text

DocumentVersion	
KUK_ID (PFK)	Text
Version	Original Translational Partial Redesign Redesign
CreationDate	Text
SourceOriginalID	Text

DocumentIdentificationGenreProperties	
KUK_ID (PK)	Text
SourceDB	FrBo_analyses FrBo_counseling Ombu_findings Ombu_flyers CzCDC
SourceID	Integer
DocumentTitle	Text
ClarityPursuit	Boolean
Anonymized	No Anonymized by source On-site anonymization
RecipientType	natural person legal person combined
RecipientIndividuation	individual bulk public
AuthorType	authority individual
Objectivity	quasiobjective persuasive
LegalActType	individual normative
Bindingness	Boolean

AnalysesSpecificColumns	
KUK_ID (PFK)	

ArticlesSpecificColumns	
KUK_ID (PFK)	

srozumitelné ano/ne