

INTLIB

a

Alfa

Projekt (TA02010182/Inteligentní knihovna)
je řešen s finanční podporou TA ČR.

Osnova

- O projektu
- Postupy prací podle oblastí
 - ▣ Legislativní doména
 - Judikatura
 - Účetní poddoména
 - ▣ Environmentální doména
 - ▣ Uplatnění výsledků
- Nejblíže plánované činnosti
- Postup prací - souhrn
- Vynaložené náklady na projekt

○ projektu – Představení

- Účel
 - ▣ Zefektivnění prohledávání a vytěžování znalostí z odborných textů v českém jazyce
- Cílem projektu je vybudovat systém použitelný v přesně vymezených doménách:
 - ▣ legislativa
 - ▣ životní prostředí – dokumentace EIA
- Přizpůsobíme a propojíme vědecké výsledky z oblastí lingvistiky, dolování dat, databází, Linked Data apod.
- Certifikovaná metodika a v praxi využitelný SW

○ projektu – Představení

- Systém bude částem dokumentů přiřazovat význam dle definovaného slovníku (modelu, ontologie).
 - ▣ inteligentní indexování dokumentů
- Výstupem budou extrahovaná data, pojmy a vazby z jednotlivých dokumentů s jednoznačně určeným sémantickým přiřazením.
- Oindexovaná a vzájemně propojená data usnadní prohledávání a procházení dokumentů uživateli (oproti existujícím nástrojům dostupným na trhu).
- Reprezentace dat v podobě tzv. Linked Data.

○ projektu – Cíle

- Nová certifikovaná metodika
 - ▣ dolování sémantiky odborných textů učícím se systémem a zpracování, prohledávání a prezentace získaných sémantických dat dle principů Linked Data
- Softwarový nástroj
 - ▣ pro dolování, zpracování, prohledávání a prezentaci dat. SW je přizpůsoben dvěma daným konkrétním doménám:
 - legislativní doména - legislativní dokumentace, kde sémantikou rozumíme legislativní vztahy v ní obsažené
 - environmentální doména – aplikovaný zejména na dokumentaci EIA a související posudky a stanoviska

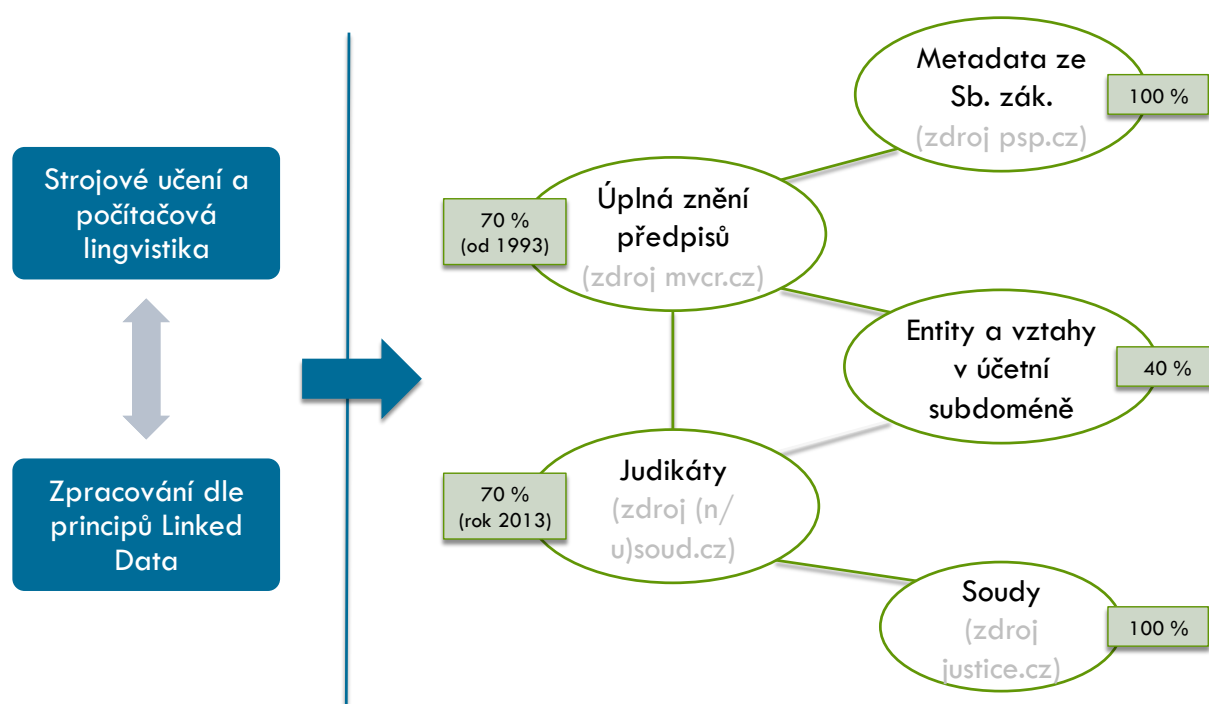
Postup prací – legislativní doména

- Po konzultacích s odborníky z oblasti práva a účetnictví, vyplynul požadavek na řešení specifických oblastí:
 - ▣ judikáty
 - ▣ účetnictví
- Znalosti z uvedených oblastí jsou nyní minimálně strojově vytěžovány, tudíž pro uživatele obtížně prohledatelné.
- Úprava algoritmů pro větný rozbor pro potřeby legislativní domény.
- Ruční anotace legislativních dokumentů – trénovací množina
- Probíhá strojové zpracování relevantních dokumentů na základě ručně oannotovaných předpisů.
 - ▣ Automatizované extrahování pojmů a vztahů z legislativních dokumentů

Postup prací – legislativní doména

- Účetní poddoména – specifické činnosti
 - ▣ Sondáž u potenciálních klientů – specifikace požadavků na funkcionalitu a vzhled
 - Ideálně výstup jako modul do existujícího systému
 - Návrh prezentační vrstvy nástroje
 - ▣ Proběhla jednání s členy Komory auditorů a Komory daňových poradců ohledně využití jejich interpretací pro vytěžování
 - ▣ Tvorba návrhu prezentační vrstvy na základě požadavků uživatelů
 - slouží i pro snadnější demonstraci vyvíjených funkcionalit

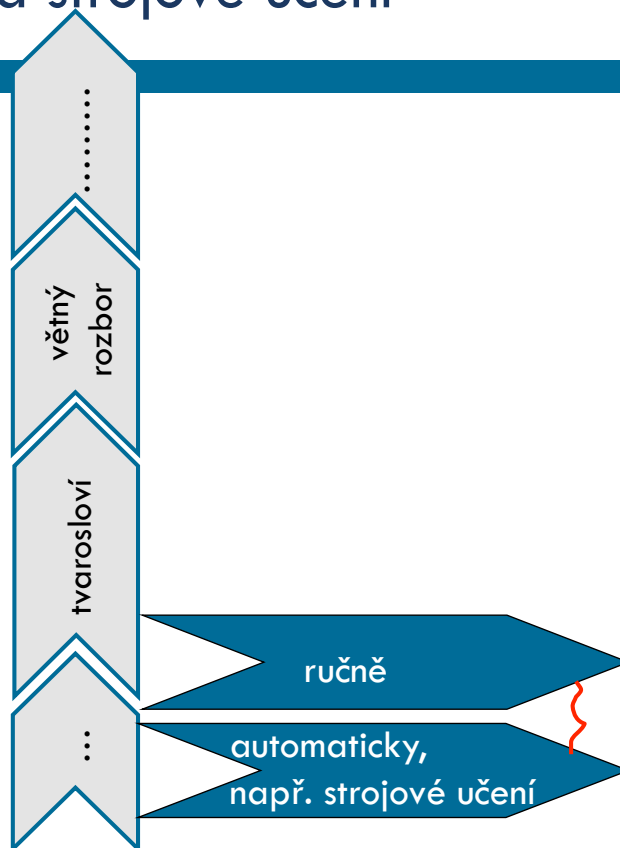
Postup prací – legislativní doména



Počítačová lingvistika a strojové učení

- (3) Účetní jednotky, které vedou účetnictví ve zjednodušeném rozsahu
- a) tvoří opravné položky a rezervy pouze dle zvláštních právních předpisů, a proto použijí ustanovení § 16, 26, 31, 55 a 57 přiměřeně,
- b) neoceňují majetek a závazky reálnou hodnotou podle § 27 zákona, s výjimkou přeměny bytových družstev, a nepoužijí ustanovení § 14, 39, § 51 až 55, § 58, 60 a 69, v rozsahu v jakém upravují oceňování majetku a závazků reálnou hodnotou,
- c) sestavují účetní závěrku ve zjednodušeném rozsahu.

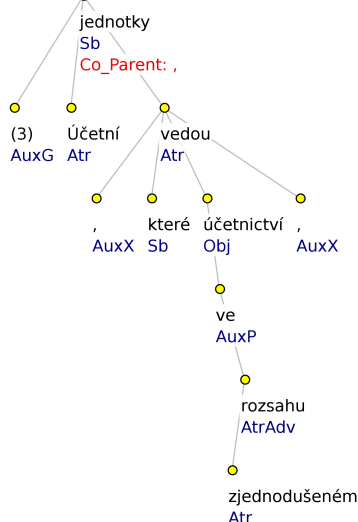
Doplnění
lingvistických znalostí
od znaků k významu



Větný rozbor legislativních textů

větný
rozbor

a-vyhlasaka.iso-001-p3s5
AuxS

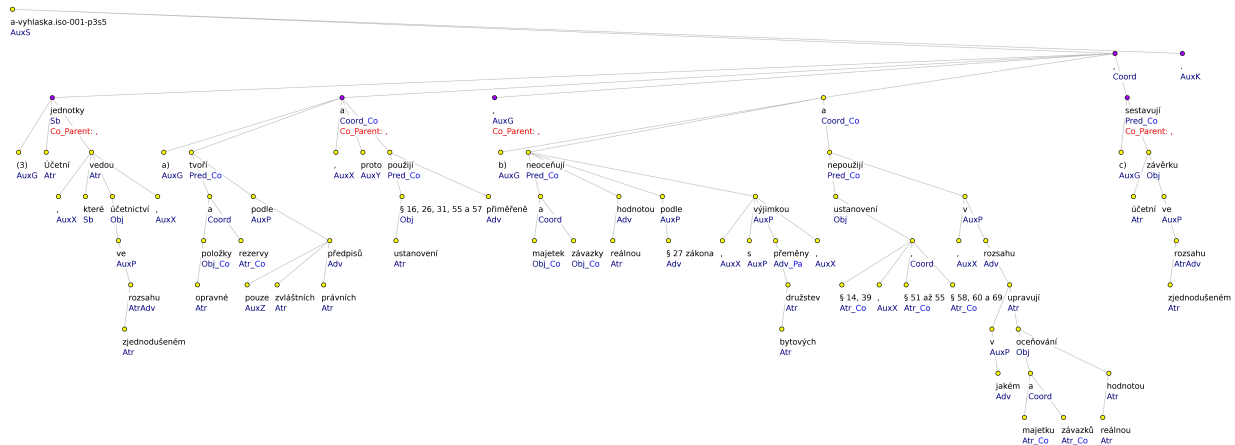


strojové učení
na publicistických textech

Lze použít na legislativní texty?

- (3) Účetní jednotky, které vedou účetnictví ve zjednodušeném rozsahu
- a) tvoří opravné položky a rezervy pouze dle zvláštních právních předpisů, a proto použijí ustanovení § 16, 26, 31, 55 a 57 přiměřeně,
- b) neoceňují majetek a závazky reálnou hodnotou podle § 27 zákona, s výjimkou přeměny bytových družstev, a nepoužijí ustanovení § 14, 39, § 51 až 55, § 58, 60 a 69, v rozsahu v jakém upravují oceňování majetku a závazků reálnou hodnotou,
- c) sestavují účetní závěrku ve zjednodušeném rozsahu.

Větný rozbor legislativních textů



Pojmenované entity v textech

§ 2

(1) Vyhláška se vztahuje na účetní jednotky podle § 1 odst. 2 písm. a) a b) zákona, s výjimkou účetních jednotek uvedených v odstavci 2, a na účetní jednotky podle § 1 odst. 2 písm. d) až h) zákona.

(2) Z účetních jednotek uvedených v odstavci 1 se tato vyhláška nevztahuje na účetní jednotky podle § 19a zákona, pokud zvláštní právní předpis 1c) nestanoví jinak, a na účetní jednotky, jejichž účetnictví upravuje zvláštní právní předpis 1d). Dále se tato vyhláška, s výjimkou § 62 odst. 2 až 5, nevztahuje na účetní jednotky podle § 23a zákona.

(3) Účetní jednotky, které vedou účetnictví ve zjednodušeném rozsahu,

a) tvoří opravné položky a rezervy pouze podle zvláštních právních předpisů, 1e) a proto použijí ustanovení § 16, 26, 31, 55 a 57 přiměřeně,

b) neocetňují majetek a závazky reálnou hodnotou podle § 27 zákona, s výjimkou přeměny bytových družstev, a nepoužijí ustanovení § 14, 39, § 51 až 55, § 58, 60 a 69, v rozsahu v jakém upravují oceňování majetku a závazků reálnou hodnotou,

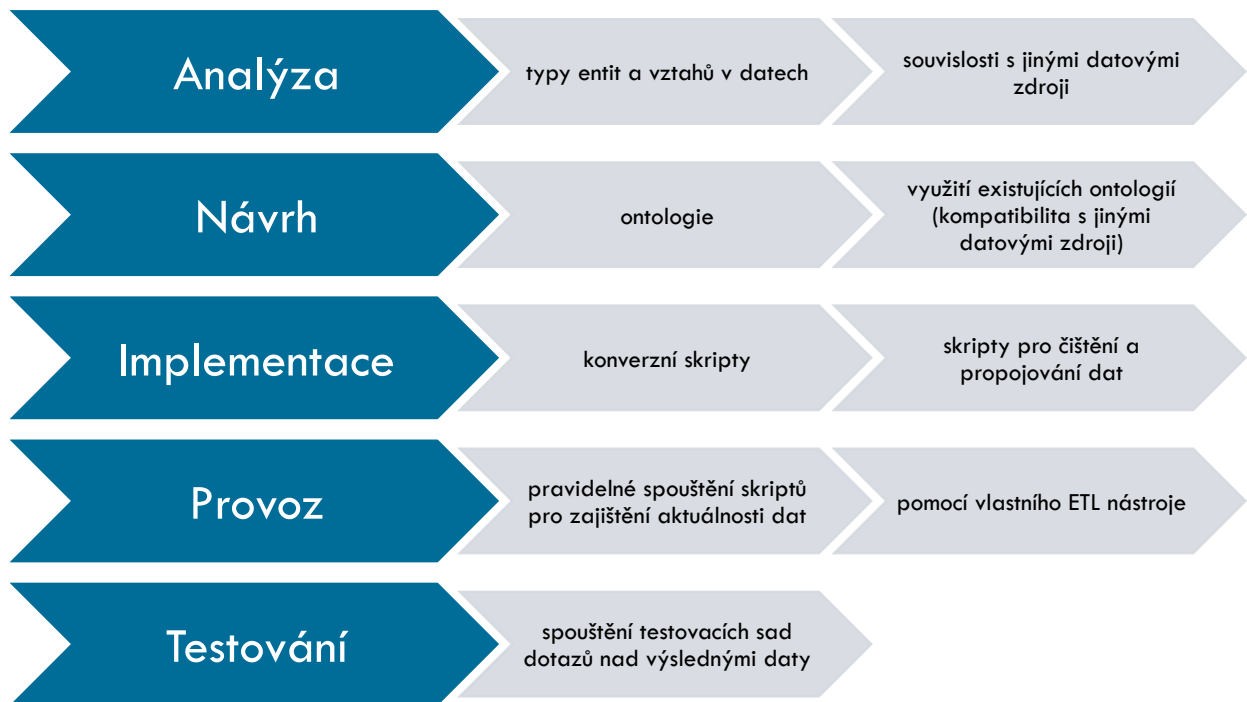
c) sestavují účetní závěrku ve zjednodušeném rozsahu.

[illegible]

Response	Percentage
Yes	85%
No	15%

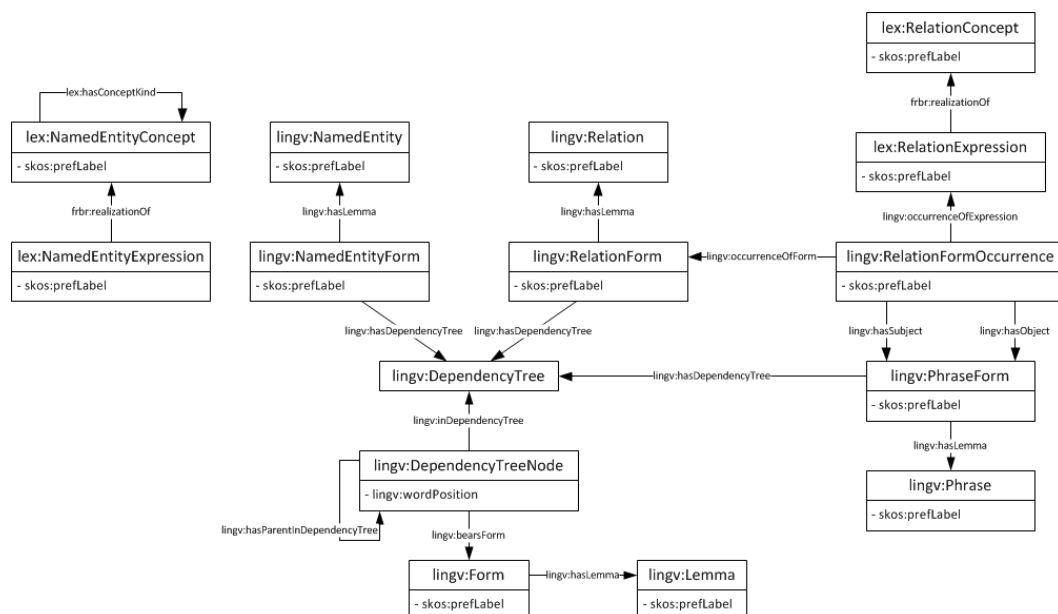


Konverze do Linked Data podoby obecně



Linked Data reprezentace pojmů a vazeb

- ontologie popisující Linked Data reprezentaci lingvistického výstupu



Linked Data reprezentace pojmů a vazeb

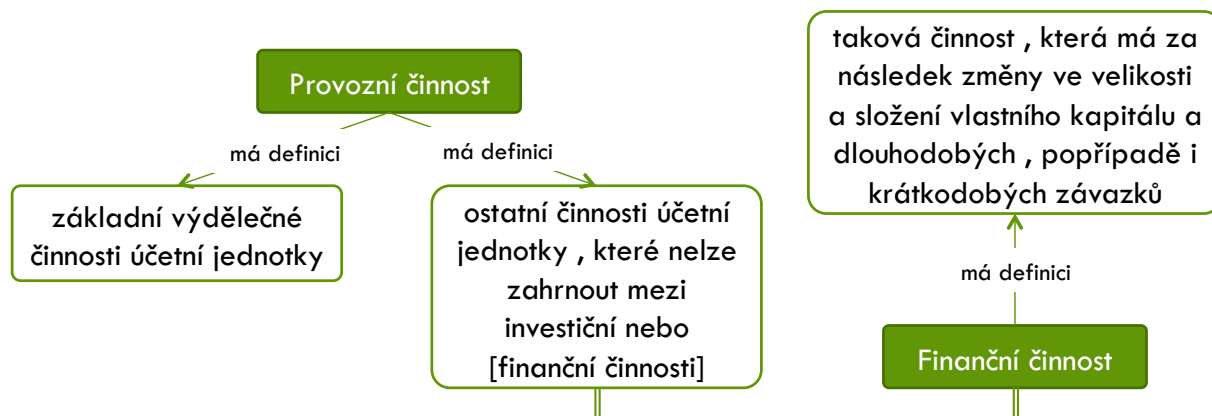
- lingvistický výstup v Linked Data reprezentaci připravený k dotazování

http://purl.org/lex/ucetnictvi#UcetnipojemConcept	656
http://purl.org/lex/ucetnictvi#PredpisConcept	156
http://purl.org/lex/ucetnictvi#PolozkaConcept	136
http://purl.org/lex/ucetnictvi#PojemConcept	94
http://purl.org/lex/ucetnictvi#UcetnipojemAktivumConcept	76
http://purl.org/lex/ucetnictvi#UcetnipojemPasivumConcept	59
http://purl.org/lex/ucetnictvi#MetodaConcept	54
http://purl.org/lex/ucetnictvi#OkamzikConcept	54
http://purl.org/lex/ucetnictvi#UcetnipojemNakladConcept	45
http://purl.org/lex/ucetnictvi#SubjektConcept	41
http://purl.org/lex/ucetnictvi#UcetConcept	33
http://purl.org/lex/ucetnictvi#ObdobiConcept	30
http://purl.org/lex/ucetnictvi#SubjektPravnickaosobaConcept	30
http://purl.org/lex/ucetnictvi#DokumentUcetnivyzkazConcept	27
http://purl.org/lex/ucetnictvi#SubjektObecnysubjektConcept	26

být	143
obsahovat	111
moci	80
použít	61
uvést	51
rozumět	48
účtovat	45
muset	33
považovat	27
sestavovat	26
odpisovat	23
lze	22
stanovit	18
uvádět	17
zahrnovat	12

Linked Data reprezentace pojmů a vazeb

- pomocí jazyka SPARQL extrahujeme např. definice pojmů
 - (všechny dvojice podstromů lingvistického rozboru vět propojené vazbou "rozumět", "považovat", atd.)



Postup prací – legislativa – účetnictví

Ukázka prezentační vrstvy nástroje

Hledaný výraz: Seznam pojmů ze slovníku LEX

Volba vztahu: Vysvětlení vztahů mezi pojmy

☒ Související pojmy ☒ Podřízené pojmy

Výsledky vyhledávání

Hledaný výraz	Vztah	Souvislost	Legislativa
Účetní výkaz	rozumí se	dokumenty zobrazující finanční situaci podniku	Zákon č. 563/1991, o účetnictví §4, odst. 2, písm. a
	je definován jako	soubor finančních výkazů, které účetní jednotka sestavuje za účetní období k rozvahovému dni (tj. k poslednímu dni účetního období)	Vyhláška č. 500/2002, prováděcí vyhláška k podvojnému účetnictví §15, odst. 1

Preference legislativa

K výrazu „účetní výkaz“ existuje zkratka „Výkaz“. Je pravděpodobné, že existují další relevantní výsledky. →

Související pojmy

- [Přehled o peněžních tocích](#)
- [Přehled o změnách vlastního kapitálu](#)
- [Příloha k účetní závěrce](#)
- [Rozvaha](#)
- [Účetní závěrka](#)
- [Výkaz zisku a ztráty](#)

Podřízené pojmy

- [Hlavní účetní výkaz](#)
- [Finanční účetní výkaz](#)

Postup prací – legislativa – účetnictví

Ukázka prezentační vrstvy nástroje

Hledaný výraz: Seznam pojmů ze slovníku LEX

Volba vztahu: Vysvětlení vztahů mezi pojmy

☒ Související pojmy ☒ Podřízené pojmy

Mapa souvisejících pojmů Mapa zobrazuje pojmy, které jsou přímo propojeny s hledaným výrazem.

The concept map shows 'Účetní výkaz' at the center, connected to various related terms:

- Účetní jednotka** (connected by 'sestavování')
- Účetní závěrka** (connected by 'je speciálním případem')
- Rozvaha** (connected by 'je speciálním případem')
- Výkaz zisku a ztráty** (connected by 'je speciálním případem')
- Přehled o peněžních tocích** (connected by 'je speciálním případem')
- Přehled o změnách vlastního kapitálu** (connected by 'je speciálním případem')
- Příloha k účetní závěrce** (connected by 'je speciálním případem')
- Definice** (connected by 'definice')
- Zákon č. 563/1991, o účetnictví §4, odst. 2, písm. a** (connected by 'definice')
- Vyhláška č. 500/2002, prováděcí vyhláška k podvojnému účetnictví §15, odst. 1** (connected by 'definice')
- Výkaz** (connected by 'zkratka')
- Účetní období** (connected by 'sestavování')

Kliknutím na související pojem bude zobrazena obdobná mapa zachycující nejbližší vztahy ke zvolenému pojmu.

Legislativa - nejbližší plánované činnosti

- Zdokonalení vyhledávacího nástroje nad oanotovanými daty – na základě uskutečněných experimentů
- Kontaktování komerčních společností, které provozují:
 - ▣ právníkové systémy (ASPI, Beck, případně další), do nichž by bylo možné zintegrovat nástroj na prohledávání judikátů,
 - ▣ účetní systémy (Pohoda, Helios a další) – nabídka SW jako modulu
- Plánování alternativního uplatnění – např.:
 - ▣ tvorba samostatného nástroje přístupného pomocí webových služeb (zpoplatněný),
 - ▣ modul pro edukativní systémy
 - návrhy prezentace zákazníkům, volba vhodné formy reklamy
- Implementace navržené prezentační vrstvy

Environmentální doména – úvod do problematiky

- ELA – posuzování vlivů plánovaných záměrů na životní prostředí
 - ▣ Záměr – např. stavby, těžba nerostných surovin, výrobní zařízení
 - ▣ V procesu posuzování vznikají dokumenty – oznámení, posudek, stanovisko k záměru
 - ▣ Dokumenty obsahují cenné informace, které nejsou běžnými metodami vyhledatelné
 - ▣ Záměry vč. průběhu procesu zveřejněny na http://portal.cenia.cz/eiasea/view/eia100_cr
- Další možné využití v oblasti ŽP
 - ▣ Registr CITES – regulace obchodu s ohroženými druhy zvířat a rostlin

Postup prací – environmentální doména

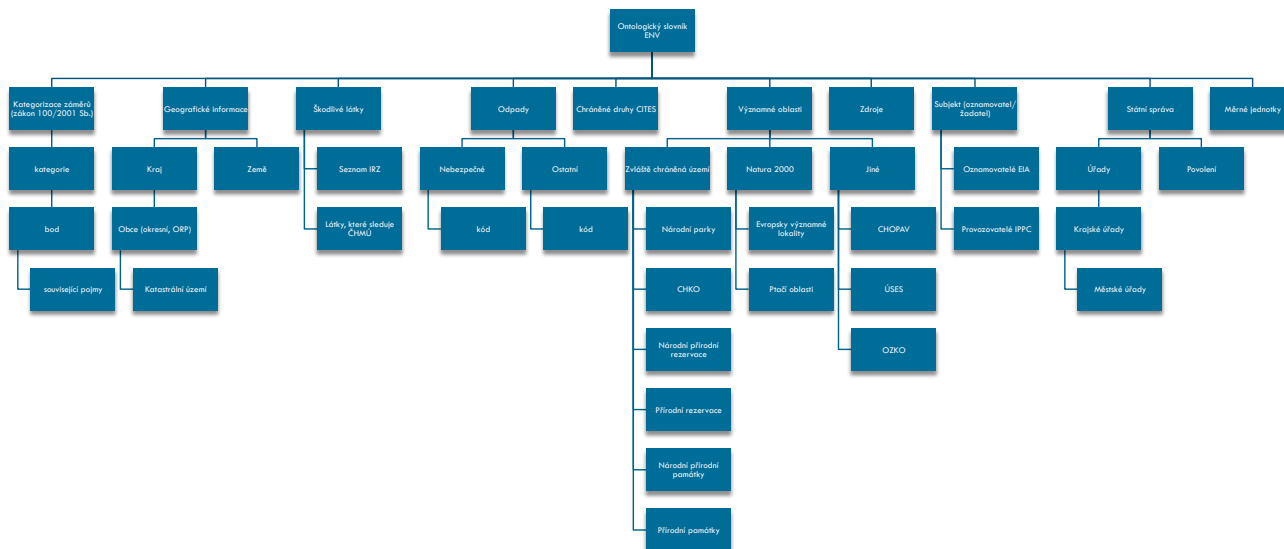
- Analýza environmentální dokumentace – EIA, CITES
 - ▣ Oznámení
 - Základní údaje (oznamovatel, umístění, termín)
 - Podle typu záměru např. plocha, kapacita zařízení, kusy dobytka... (hodnota, jednotka)
 - Navazující rozhodnutí (typicky např. stavební povolení)
 - Dopad na ŽP – zabor půdy, spotřeba zdrojů, vypouštění znečišťujících látek, odpady a další
 - ...
 - ▣ Posudek – doporučení pro vydání stanoviska
 - ▣ Stanovisko – průběh posuzování, rozhodnutí
 - ▣ CITES – rozhodnutí AOPK o povolení obchodní transakce

Postup prací – environmentální doména

- Sondáž u potenciálních klientů – specifikace požadavků na nástroj u prohledávání dokumentace EIA a CITES
- Probíhá jednání s MŽP a CENIA ohledně ideálního způsobu využití nástroje – jeho integrace do vyvíjeného systému nebo samostatně
- Tvorba ontologického slovníku pro potřeby nástroje – pojmy používané v oblasti EIA a CITES
 - ▣ Identifikovány oblasti, pro něž je třeba vytvořit či použít relevantní seznam
 - ▣ Seznamy pojmů pro jednotlivé dílčí oblasti – např. znečišťující látky, odpady, měrné jednotky, chráněná území, obce...
- Tvorba návrhu prezentační vrstvy pro potřeby prezentace možností nástroje potenciálním klientům

Postup prací – environmentální doména

□ Struktura ontologického slovníku



Postup prací – environmentální doména

□ Ukázky návrhu prezentační vrstvy pro ENV

Název záměru

Dálnice D1 - Mirošovice - Kývalka, zkapacitnění

Najít záměr

IC oznamovatele

Kód záměru

Parametrické vyhledávání

Název záměru	Oznamovatel	Kategorie	Kraj
Dálnice D1 - Mirošovice - Kývalka, zkapacitnění	Ředitelství silnic a dálnic ČR	I/9.3 Novostavby, rozšiřování a přeložky dálnic a rychlostních silnic.	Jihomoravský Středočeský Vysočina

Kapacita a rozsah

Lokalita

Termín realizace

Dotčené územně samosprávné celky

Navazující rozhodnutí

Vstupy

Výstupy

Dotčené významné oblasti

Vlivy na obyvatelstvo a životní prostředí

Přeshraniční vlivy - dotčené země

Dokumenty

System

zobrazuje pouze data, která bezpečně rozpozná v dokumentech. Pro přesnou informaci nahleďte, prosím, do originálního dokumentu.

Postup prací – environmentální doména

Název záměru Najít záměr

IC oznamovatele Kód záměru

► Parametrické vyhledávání

Název záměru	Oznamovatel	Kategorie	Kraj
Dálnice D1 - Mirošovice - Kývalka, zkapacitnění	Ředitelství silnic a dálnic ČR IČ: 65993390	I/9.3 Novostavby, rozšiřování a přeložky dálnic a rychlostních silnic.	Jihomoravský Středočeský Vysocina

► Kapacita a rozsah

► Lokalita

▼ Termín realizace

Zahájení 2015
Dokončení 2020

► Dotčené územně samosprávné celky

► Navazující rozhodnutí

► Vstupy

▼ Výstupy

Ovzduší	Odpadní vody	Odpady	Ostatní
Oxidy dusíku ≥ 17 000 t/rok Oxid uhelnatý ≥ 9 000 t/rok PM ₁₀ ≥ 8 000 t/rok benzen ≥ 55 t/rok	Splaskové odpadní vody Dešťové odpadní vody 31 770 m³/rok Podzemní vody	Nebezpečné uniklé (rozlité) ropné látky ostatní rozpouštědla a směsi rozpouštědel	Pneumatiky biologicky rozložitelný odpad

► Dotčené významné oblasti

► Vlivy na obyvatelstvo a životní prostředí

► Přeshraniční vlivy - dotčené země

► Dokumenty

⚠️ Systém zobrazuje pouze data, která bezpečně odpovídají v dokumentech. Pro přesnou informaci nahleďte, prosím, do originálního dokumentu.

Postup prací – environmentální doména

Název záměru Najít záměr

IC oznamovatele Kód záměru

▼ Parametrické vyhledávání

Geografická lokalizace záměru

Kraj

Obec

Produkované škodlivé látky

Produkované odpady

Klíčová slova

Dotčené oblasti

☒ Zvláště chráněná území
☐ Natura 2000 ☐ CHOPAV ☐ ÚSES ☐ OZKO

Zasažené ohrožené druhy CITES

Provozovatelé s povolením IPPC

Plánovaná doba realizace
Od Do

Volba zobrazení seznamu výsledků

☒ Název záměru
☒ Oznamovatel
☒ Kategorie záměru
☒ Kraj lokalizace záměru
☐ Stanovisko (souhlasné/nesouhlasné)
☐ Plánovaná doba realizace
☐ Datum zveřejnění informace o záměru
☐ Dotčené oblasti

Hledat záměry dle parametrů

Postup prací – environmentální doména

Parametrické vyhledávání

Geografická lokalizace záměru

Kraj: Jihočeský
Obec: [všechny]
Produkovávané škodlivé látky: [všechny]
Produkovávané odpady: [všechny]
Klíčová slova: [všechny]

Dotčené oblasti

☒ Zvláště chráněná území
☐ Natura 2000 ☐ CHOPAV ☐ ÚSES ☐ OZKO
CHKO Šumava

Zasažené ohrožené druhy CITES

[všechny]

Provozovatel s povolením IPPC

[všechny]

Plánovaná doba realizace

Od: [všechny] Do: [všechny]

Volba zobrazení seznamu výsledků

☒ Název záměru
☒ Oznamovatel
☒ Kategorie záměru
☒ Kraj lokalizace záměru
☐ Stanovisko (soulasně/nesoulasně)
☐ Plánovaná doba realizace
☐ Datum zveřejnění informace o záměru
☐ Dotčené oblasti

Hledat záměry dle parametrů

5 výsledků vyhledávání

#	Název záměru	Oznamovatel	Kategorie	Kraj
1	Rekonstrukce a rozšíření kapacity kravína, farma Plav, okr. Č. Budějovice	Zemědělský podnik Malše a.s.	II/1.5 Chov hospodářských zvířat s kapacitou od 50 do 180 dobytčích jednotek (1 dobytčí jednotka = 500 kg živé hmotnosti).	Jihočeský
2	Rozšíření těžby mimo dobývací prostor Ševětín I.	Kámen a písk spol. s r.o., Český Krumlov	II/2.5 Těžba nerostných surovin 10 000 až 1 000 000 tun/rok; těžba rašeliny na ploše do 150 ha.	Jihočeský
3	Silnice I/22 Strakonice	Ředitelství silnic a dálnic ČR, České Budějovice		Jihočeský
4	Zásobník kyseliny chlorovodíkové	SIGRUM spol. s r.o.	II/10.4 Skládování vybraných nebezpečných chemických látek a chemických přípravků (vysoko toxických, toxických, zdraví škodlivých, žravých, dráždivých, senzibilizujících, karcinogenních, mutagenních, toxických pro reprodukci, nebezpečných pro životní prostředí) (11) a pesticidů v množství nad 1 t kapalných hnojiv, farmaceutických výrobků, barev a laků v množství nad 100 t.	Jihočeský
5	Filtrace užitkových olejů v areálu firmy ZITA Kájov	ZITA Kájov s.r.o.	II/10.1 Zařízení ke skladování, úpravě nebo využívání nebezpečných odpadů zařízení k fyzikálně-chemické úpravě, energetickému využívání nebo odstraňování ostatních odpadů.	Jihočeský

Společná tabulka pro zveřejnění údajů o záměrech. Pro zveřejnění údajů o záměrech, pro zveřejnění údajů o záměrech, pro zveřejnění údajů o záměrech.

Environmentální doména nejbližší plánované činnosti

- Dokončení ontologického slovníku
- Anotace trénovací množiny pomocí pojmů z vytvářeného slovníku
- Tvorba pravidel pro prohledávání relevantní environmentální dokumentace
- Extrakce pojmů z environmentální dokumentace na základě ontologického slovníku a lingvistických pravidel
- Implementace navržené prezentační vrstvy

Postup prací – uplatnění výsledků

- **Judikatura**
 - ▣ Budou uskutečněna jednání o uplatnění nástroje jako dílčího modulu pro právníkový systém ASPI
 - ▣ Pokud nebude uskutečněna dohoda, bude navržen samostatný nástroj, který bude zpřístupněn přes web
- **Účetní poddoména**
 - ▣ Budou kontaktovány komerčních společností, které provozují účetní systémy (Pohoda, Helios a další) ohledně možné integrace nástroje jako modulu do jejich systémů
 - ▣ V případě špatné odezvy bude navrženo alternativní uplatnění – např. tvorba samostatného nástroje, který bude zpřístupněn přes web
- **Environmentální doména**
 - ▣ Probíhají jednání s MŽP a CENIA o možném uplatnění nástroje jako dílčího modulu do nově vyvíjeného IS na ministerstvu, případně o jeho prezentaci na samostatném portálu.

Postup prací - souhrn

Dílčí cíl 2012	Splněno	Dílčí cíl 2013	Splněno
DC1.1	100%	DC2.1	75%
DC1.2	100%	DC2.2	75%
DC1.3	100%	DC2.3	60%
DC1.4	100%	DC2.4	10%
DC1.5	100%	DC2.5	75%
DC1.6	100%	DC2.6	70%
		DC2.7	75%
		DC2.8	50%*

* 2 články popisující dílčí výsledky z roku 2013 již vyšly. Pracujeme na souhrnné výzkumné zprávě za celý rok 2013.

Vynaložené náklady na projekt

□ 2012

Položka	Uznané náklady	Skutečné náklady
Osobní náklady	3 052 000,-	2 752 665,-
Služby	165 000,-	450 561,55
Ostatní náklady	591 000,-	602 214,07
Náklady CELKEM	3 808 000,-	3 805 440,62
Výše podpory	2 436 000,-	2 433 251,62
Neveřejné zdroje	1 372 000,-	1 372 189,-
Zdroje CELKEM	3 808 000,-	3 805 440,62

Vynaložené náklady na projekt

□ 2013 (čerpaní do 31.8.2013)

Položka	Uznané náklady	Vyčerpané náklady	Vyčerpáno (%)
Osobní náklady	3 052 000,-	1 774 703,-	58%
Služby	105 000,-	177 935,-	169%
Ostatní náklady	705 000,-	403 626,-	68%
Náklady CELKEM	3 862 000,-	2 434 268,-	63%
Výše podpory	2 492 000,-	1 133 818,-	46%
Neveřejné zdroje	1 370 000,-	1 300 450,-	95%
Zdroje CELKEM	3 862 000,-	2 434 268,-	63%

INTLIB – TA02010182



Děkujeme za pozornost

Dotazy?

□ Stránky projektu: <http://ufal.mff.cuni.cz/intlib>