



VIKMA05

Organizace znalostí

7. 4. 2017: Formální struktura dokumentu, reprezentace dokumentu (P7+K3)

FF MU, jaro 2017

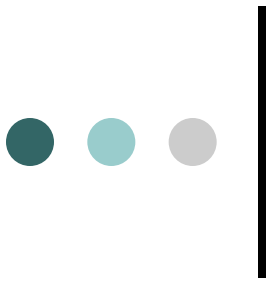
Mgr. Josef Schwarz

126172@mail.muni.cz



Osnova

- Přehled (minulých a budoucích) témat
- Praktická část
- Teoretická část
 - SGML
 - právní texty – struktura
- Zadání 7. dílčího kroku DA



PŘEHLED TÉMAT



Přehled (minulých) témat

Témata

- Úvod
- Přístupy ke KO, doménová analýza
- Ontologie, epistemologie, teleologie (pojem, metafora)
- Kategorie, kategorizace, klasifikace (tradiční vs kognitivistický přístup: prototypy, bázové kategorie)
- Termíny a terminologie
- Notace a notační systém

Úkoly

- Volba domény, zdroje
- Metafory
- Prototypy a bázové kategorie
- Reprezentace pojmu
- Návrh notačního systému



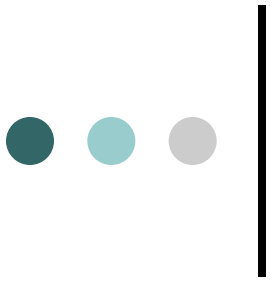
Přehled (budoucích) témat

- Formální struktura dokumentu, reprezentace dokumentu
- Procesy organizace znalostí
- Ontologie
- Univerzální klasifikační systémy
- Systémy založené na přirozeném jazyce
- Reprezentace a vizualizace znalostí
- Závěrečná hodina - test



Praktická část

DA krok č. 6



STRUKTURA DOKUMENTU



Formální struktura dokumentu

- Tradiční

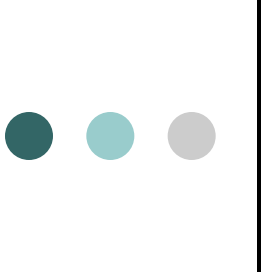
- Obsah

- [Příklad1](#), [příklad2](#)

- Elektronická

- SGML (ISO 8879:1986)

- Obecná struktura dokumentu a další atributy
 - Odvozené jazyky: např. HTML, XML
 - Pro základní informace o SGML a jeho struktuře viz web W3.org

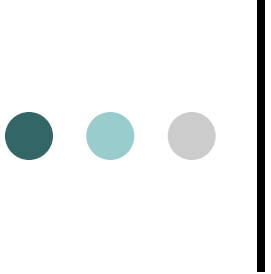


SGML – příklad 1

PhDr. Rostislav Kaprál

Dělnická 356/5

280 02 Kolín II



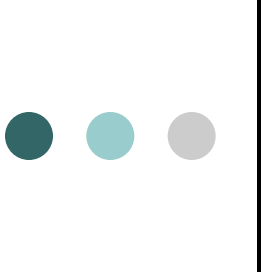
SGML – příklad 1

- <DOKUMENT>
 - <adresa stát="cs">
 - <osoba>
 - <titul>PhDr. </titul><křestní jméno>Rostislav<křestní jméno><příjmení>Kaprál</příjmení>
 - </osoba>
 - <ulice>Dělnická</ulice>
<číslo_popisné>356</číslo_popisné>/<číslo_orientační>5</číslo_orientační>
 - <PSČ>280 02</PSČ> <město>Kolín </město>
<městská část> II </městská část>
 - </adresa>
- </DOKUMENT>



SGML – příklad 1

- <DOKUMENT>
 - <adresa stát="cs">
 - <subjekt typ="osoba">
 - <titul>PhDr. </titul><křestní jméno>Rostislav<křestní jméno><příjmení>Kaprál</příjmení>
 - </subjekt>
 - <ulice>Dělnická</ulice>
<číslo_popisné>356</číslo_popisné>/<číslo_orientační>5</číslo_orientační>
 - <PSČ>280 02</PSČ> <město>Kolín </město>
<městská část> II </městská část>
 - </adresa>
- </DOKUMENT>



SGML – příklad 2

- Není známo, kdo přesně radil voličům Úsvitu přímé demokracie Tomia Okamury, nicméně stalo se, že Úsvit získal dost hlasů na to, aby se dostal do parlamentu, a tudíž aby dostal nárok na příspěvek ze státní kasy. To už nezjistíme, kdo radil voličům. Nicméně některé záhady objasněny byly. Ten, kdo radí Úsvitu, je Tomio Okamura a bere za to mega měsíčně. Jeho hnutí má členů míň, než měl Ježíš učedníků. Jakpak se jim asi četl Reflex, kde se o tom píše? Asi si řekli to, co si asi řeknou voliči Tomia Okamury. Reflex je hnusný štvavý plátek, který pomlouvá a lže. Vždyť přece Tomio Okamura je ztělesnění té pravé demokracie, má to i v názvu svého podniku. Jak by mohl udělat něco tak hnusného? Možná, že se v příští etapě Okamura zachová podobně jako Zeman. Ten nechal své jméno odstranit z názvu své politické strany. Okamura by mohl nechat z názvu odstranit slovo demokracie. V obou případech to budí smích.
- Zdroj: [Neviditelný pes](#), 4. 4. 2014

SGML – příklad 2

- `<dokument typ=„politický komentář“ země=„ČR“ čas=„2014-04-04“> <stav typ=„výsledek voleb“ podtyp=„parlamentní volby ČR 2013“> Není známo, kdo přesně radil voličům <subjekt typ=„instituce“ podtyp=„politické strany“> Úsvitu přímé demokracie </subjekt> <subjekt typ=„osoba“ podtyp=„politik“> Tomia Okamury </subjekt>, nicméně stalo se, že Úsvit získal dost hlasů na to, aby se dostal do parlamentu, a tudíž aby dostal <finance typ=„finanční subvence“ podtyp=„příspěvek politickým stranám“> nárok na příspěvek ze státní kasy </finance>. To už nezjistíme, kdo radil voličům. </stav> Nicméně některé záhady objasněny byly. <funkce typ=„politický poradce“>Ten, kdo radí Úsvitu, je Tomio Okamura a <finance typ=„finanční odměna“> bere za to mega měsíčně</finance>. <stav typ=„politické strany“ podtyp=„počet členů“> Jeho hnutí má členů míň, než měl Ježíš učedníků. </stav> <reakce typ=„politický komentář“> Jakpak se jim asi četl Reflex, kde se o tom píše? Asi si řekli to, co si asi řeknou voliči Tomia Okamury. Reflex je hnusný štvavý plátek, který pomlouvá a lže. </reakce> Vždyť přece Tomio Okamura je ztělesnění té pravé demokracie, má to i v názvu svého podniku. Jak by mohl udělat něco tak hnusného?</funkce> Možná, že se v příští etapě Okamura zachová podobně jako <subjekt typ=„osoba“ podtyp=„politik“> Zeman </subjekt>. Ten nechal své jméno odstranit z názvu své politické strany. Okamura by mohl nechat z názvu odstranit slovo demokracie. V obou případech to budí smích. </dokument>`



Analýza právních textů

Kořenský – Cvrček – Novák: Juristická a lingvistická analýza právních textů. Praha: Academia, 1999

○ A. Typologie dokumentů

● 1. podle druhu

- **právní předpisy**
- **soudní rozhodnutí**
- **individuální rozhodnutí a normativní instrukce státní správy**
- **důvodové zprávy**
- **návrhy a projekty zákonů a dalších předpisů**
- **záznamy parlamentních debat**
- **výsledky připomínkových řízení**
- **komentáře**
- **učebnice, skripta a odborné monografie**
- **odborné články**
- **expertizy**
- **přehledy platných právních předpis**



Analýza právních textů

- **A. Typologie dokumentů**
 - **2. Podle formy publikace**
 - oficiální sbírky
 - neoficiální sbírky
 - knihy
 - odborné časopisy
 - jednotlivé dokumenty uložené v archivech institucí
 - **3. Podle vztahu k původnímu znění**
 - původní
 - upravené
 - obohacené
 - metadokumenty nebo adresy



Analýza právních textů

- **B. Struktura položek v základním slovníku (LEMMA)**
 - **1. Základní slovní tvar**
 - **2. Typ**
 - Speciální znaky
 - Číslice, římské číslice
 - Zkratky s tečkou (*Sb.*), zkratky bez tečky (*ČR*)
 - Pomocné znaky (§)
 - Slovenská slova
 - Cizí slova (ne chemické vzorce, ne názvy nemocí apod.)
 - **3. Slovní druh**
 - **4. Rod**
 - **5. Životnost**
 - **6. Příslušnost ke kmenové sémantické skupině (*daně, zdanění, danit*)**
 - **7. Typy právních významů**
 - **8. Četnost**
 - **9. Číslo základního slovního tvaru**



Analýza právních textů

○ C. Sémantika morfologicko-syntagmatických typů

● Příklad – škoda

- *A – S:* *případná / věcná / hrozící*
- *S – S:* *velkého rozsahu / značného rozsahu*
- *S – praep S:* *na zdraví / na majetku / na pozemku*
- *S – VF:* *vznikla / přesahuje / hrozí*
- *VF – S:* *zavinil / způsobil / zabránil (škodu/škodě)*
- *VF – praep S:* *dochází (ke škodám) / odpovídá (za škodu)*
- *S – (praep) S:* *zjišťování / náhrada / úhrada (škody) / odpovědnost (za škodu)*
- *INF – S:* *zabránit (škodě) / nahradit (škodu)*
- *INF – praep S:* *odpovídat (za škodu)*

● Sémantické hodnocení vazeb:

- Ad A - S
- Věcná povaha škody
- Vyjádření původu a příčiny škody
- Kvantitativní hodnocení škody
- Společenské hodnocení škody

● Model morfologicko-syntagmatické struktury



Analýza právních textů

- **D. Výrazová segmentace textu**
- **E. Vztah mezi formální větou a textovou strukturací**
- **F. Výskyt vyhledávaných syntagmat v segmentech textu**
- **G. Distinktivní vlastnosti spojovacích prostředků (spojky a obdobné výrazy)**
- **H. Morfologie, gramatika, sémantika právních textů**
- **I. Struktura právních textů a jejich formální vztahy**
- **J. Využití:**
 - **1. Tvorba právních předpisů**
 - Formulace koncepčního záměru
 - Koncepce konkrétní legislativní úpravy
 - Zjišťování kontextových souvislostí
 - Právnělingvistická podpora legislativní činnosti
 - Vytváření úplných znění právních předpisů
 - Zpracování důvodových zpráv
 - Příprava legislativních výzkumů
 - **2. Tvorba právní terminologie**
 - **3. Tvorba oborových klasifikací a tezurů**
 - **4. Vyhledávání v oborových databázích**