

PŘEDNÁŠKA 5.

Tvorba námětových map



Martina Husáková

13. 3. 2014

Obsah přednášky

- Systematický přístup k tvorbě námětové mapy
- Způsoby tvorby námětové mapy
- Syntaxe XTM

2

Systematický přístup k tvorbě námětové mapy

- Návrh a implementace námětové mapy může být dosažena pomocí těchto klíčových kroků:
 - **Analýza**
 - Specifikace domény a stanovení šíře záběru
 - Všeobecné otázky spojené s vývojem námětové mapy
 - Snaha nevyvíjet to, co už bylo jednou vyvinuto
 - Brainstorming
 - **Návrh**
 - Specifikace schématu námětové mapy a instancí
 - **Implementace**
 - Výběr přístupu pro tvorbu kódu námětové mapy
 - **Tvorba dokumentace**
 - **Údržba námětové mapy**

Analýza: specifikace domény a stanovení šíře záběru

- **Omezení zkoumané oblasti**
 - Které předměty má a nemá námětová mapa pokrývat?
 - Nutno určit úroveň detailnosti modelování jednotlivých elementů námětové mapy a množství těchto elementů
 - Toto určujeme pro snadnější management námět. mapy
 - Které informační zdroje mají být zahrnuty a jakou mají mít formu (formát)?
- **Koncoví uživatelé**
 - Jaké uživatelské skupině je námětová mapa určena?
 - Jaká jsou očekávání této uživatelské skupiny?
 - Je zde uživatelská skupina se specifickými požadavky?

Analýza: specifikace domény a stanovení šíře záběru

- Existují zde určitá omezení v rozpočtu nebo času zadavatelů (koncových uživatelů) projektu?
- Jaké jsou dostupné lidské zdroje, resp. jací experti jsou k dispozici?
- V jakém jazyce se bude komunikovat?
- **Hledisko tvorby samotné námětové mapy**
 - Jakým způsobem se bude samotná námětová mapa vytvářet, resp. manuálně, polo-automaticky nebo automaticky?
 - Existují již hotové námětové mapy, které pokrývají požadovanou doménu? Pokud ano, integrujeme z nich znalosti do naší námětové mapy?

Analýza: specifikace domény a stanovení šíře záběru

- Budeme informace a znalosti pro tvorbu námět. mapy čerpat od expertů nebo z knih, časopisů a dalších publikací?
- Požadujeme vícejazyčnou námětovou mapu?
- Předpokládá se budoucí znovupoužití ontologie?
- Předpokládá se propojení s jinou námětovou mapou/aplikací?
- Kolik autorů bude pracovat na vývoji námětové mapy?
- **Hledisko budoucí existence námětové mapy**
 - Jak a kým bude námětová mapa udržována v budoucnu?

Analýza: úvaha nad opakovaným použitím něčeho již hotového

- „Do not re-invent the wheel.“ ☺
- Použít maximum toho, co už bylo dříve vytvořeno
- Uvnitř vlastní organizace
 - Schémata, modely, návrhy vytvořené při jiných příležitostech
 - Databázová schémata
 - Formalizovaná metadata
 - DTD (document type definitions), schémata XML
 - Řízené slovníky, thesauri, klasifikační schémata
- Vně vlastní organizace
 - Existující aplikace nad námětovou mapou
 - Seznamy zveřejněných identifikátorů (PSI)
 - Knihovny ontologií
 - ...

Analýza: identifikace důležitých termínů domény

- **Brainstorming**
 - Cílem je shromáždit soupis všeho, co se zkoumané domény nějak týká
 - Nejprve co nejdelší, neuspořádaný soupis termínů
 - Následně odlišení jedinců, tříd, vztahů, zdrojů apod.
=> základ ontologie, která se bude dále upřesňovat
- **Nakládání se synonymy**
 - Ujistit se, že všechny předměty, které chceme v námětové mapě ve formě námětů reprezentovat, jsou různé (neexistence synonym)
 - Pokud chceme reprezentovat synonyma, tak v podobě alternativních názvů námětů

Návrh námětové mapy

- Brainstorming nám poskytne první vhled na obsah námětové mapy, resp. seznam typů námětů, asociací, rolí, výskytů a námětů
- Seznam zachycen v neformální podobě, např. na papír



Takto NE! ☺

[Zdroj: Brainstormin, obrázek. URL: <http://happybrainstorm.com/index.php/How-to-organise-a-BRAINSTORMING-session.html>]

Pozitivní a negativní příklad brainstormingu

Pozitivní příklad



<http://www.youtube.com/watch?v=7McMFSMzDM&feature=related>

Negativní příklad



<http://www.youtube.com/watch?v=e8kp7Hzcg4&feature=related>

10

Návrh: identifikace typů námětů

- **Kriteria rozhodování, co z domény modelovat jako typy námětů**
 - Jedná se o množinu věcí, které vykazují nějaké společné charakteristiky?
 - Příklad: psi jsou takové „věci“, které: mají čtyři nohy, mají ocas, mají srst, jsou savci, žerou maso, štěkají, dají se ochočít,...
 - Jedná se o množinu jedinců, kteří sdílejí alespoň jednu z hlediska modelované domény důležitou vlastnost?
 - Příklad: Operní skladatel je ten, kdo složil alespoň jednu operu.
 - Jedná se o implicitně nebo explicitně definovanou kategorii, o které můžeme říci, zda do ní jedinec patří nebo nepatří?
 - Příklad: stůl, židle, hrnek... tyto pojmy je sice těžké formálně definovat, ale (většinou) dokážeme v konkrétních případech rozpoznat, zda určitá věc spadá do této kategorie nebo ne

Návrh: odlišení typů námětů, námětů a typů rolí

- **Typ je to, co může mít instance**
 - „geografie“ instance nemá, takže nejde o typ
 - „stát“ instance mít může, takže jej má smysl definovat jako typ
- **Rozlišujeme:**
 - typy námětů s instancemi
 - „osoba“, případně „muž“, „žena“ jsou typy
 - role v asociacích
 - „student“, „učitel“, „skladatel“ jsou role:
 - student = osoba, která studuje
 - skladatel = osoba, která skládá

Návrh: identifikace (typů) asociací



- Začneme souvislostmi mezi těmi náměty, které jsme identifikovali v rámci brainstormingu
- Následně zjišťujeme, jestli je třeba zaznamenat ještě vztahy další
- Součástí definice typu asociace jsou role
 - Vymezují, jakým způsobem se náměty mohou na asociaci podílet
 - Počet rolí není omezen, nejčastěji dvě (binární asociace)
 - Unární asociace definuje podmínku typu ano-ne
 - Příklad: dílo je/není dokončeno; skladatel žije/nežije

Návrh: identifikace (typů) asociací



- Snažme se **zevšeobecňovat**
 - Příklad: role **autor** a **dílo** spolu s asociací **vytvořil** můžeme opakovaně použít pro spisovatele a knihu, skladatele a operu, programátora a program,... namísto tří asociací *napsal, složil, naprogramoval*
- Zavedeme jeden **shrnující** námět pro vymezení účastníků v roli:
 - Příklad: definujeme asociaci “šelma – žere – maso”, nikoli “pes, liška, tygr,... – žere – maso”

Návrh: identifikace typů výskytů



- Pro každý typ externího výskytu zkoumáme:
 - Které náměty mohou mít výskyty takového typu
 - Kolik výskytů stejného typu může námět mít
 - Jaké formáty informačních zdrojů jsou přípustné
- Pro každý typ interního výskytu zkoumáme:
 - Které náměty mohou mít výskyty takového typu
 - Kolik výskytů stejného typu může námět mít
 - Jaká syntaxe řetězcových (stringových) hodnot je přípustná

Návrh: hierarchie typů



- Při tvorbě hierarchie kombinujeme dva typy vztahů:
 - instance-of (is-a): je instancí typu (vztah mezi námětem a typem námětu)
 - is-subclass-of (is-a-kind-of, ako) (vztah mezi typy námětů)
- Příklad:
 - Maxipes Fík is-a Pes
 - Pes ako Šelma

Návrh: hierarchie typů



- Ne každá hierarchie je hierarchií typů!
 - Například nemůže říct: “Paříž AKO Francie”
- Pro hierarchii typů platí:
 - Když typ A je nadtypem typu B, potom
 - (a) A i B jsou typy
 - (b) je-li C instancí typu B, musí být také instancí typu A
 - (c) je-li D podtypem typu B, musí být i podtypem typu A
- Pokud některá z předchozích podmínek neplatí, nelze použít předdefinovanou taxonomickou asociaci superclass-subclass a místo ní je nutné definovat vlastní asociaci

Implementace námětové mapy



- Možnosti pro implementaci:
 - Tvorba námětových map s pomocí editorů
 - Manuální tvorba námětových map
 - Automatická tvorba námětových map
- Ve všech případech lze využít již existující zdroje informací a znalostí:
 - Metadata dokumentů
 - Informační systémy
 - Znalosti expertů explicitně nevýjádřené

Implementace: prakticky

- Editory
 - Ontopia, Wandora, CmapTools, TM4L
- Frameworky
 - TM4J
- Syntaxe standardizované
 - XTM (XML Topic Maps)
 - CTM (Compact Topic Maps Syntax)
- Syntaxe nestandardizované
 - LTM (Linear Topic Maps Syntax)
 - AsTMA=
 - Syntaxe HyTM (HyTime - Hypermedia/time-based)

19

Dokumentace ontologie

- Sama ontologie nestačí (i když je pro námětovou mapu zásadní)
- Dokumentace je potřebná, pokud s ontologií má pracovat více uživatelů
 - Zajišťují konzistentní používání ontologie
 - Ulehčí novým uživatelům porozumět ontologii
- Dokumentace může zahrnovat
 - Návod, vysvětlující ontologii
 - Textový dokument, komentující ontologii
 - UML diagramy, vysvětlující obrázky
 - Tabulky a seznamy typů a jejich popisů
 - Sady zveřejněných identifikátorů předmětů
 - Nejčastěji forma webové stránky
 - Jejich smyslem je zjednodušení vývoje aplikací
 - Schéma ontologie

Syntaxe XTM (XML Topic Maps)

- Část standardu ISO 13250 pod názvem ISO/IEC-3 XML syntaxe
- Syntaxe XTM je založená na značkovacím jazyce XML a jazyce XLink => XTM dokument je XML dokumentem
- Vznik syntaxe na základě požadavku využití technologie Topic Maps i ve webovém prostředí
- Autor syntaxe: nezávislá organizace TopicMaps.org
- První verze syntaxe (1.0) představena r. 2000 na konferenci XML 2000 ve Washingtonu
- Oficiální publikace syntaxe r. 2001 ve formě standardu ISO
- V současné době se jedná o jednu z nejpoužívanějších syntaxí pro tvorbu námětových map
- Možnost využití verze 2.0 (prostředí Ontopia ji má v nabídce)

Proč tvořit námětovou mapu ručně, když máme editor Ontopoly?

- Editor Ontopoly automaticky přiděluje své vlastní identifikátory jednotlivým elementům námětové mapy => problém při dotazování se nad námětovými mapami a s identifikací elementů uvnitř kódu XTM dokumentu

```
<topic id="id101">
  <instanceOf>
    <topicRef xlink:href="#id74"></topicRef>
  </instanceOf>
  <baseName>
    <baseNameString>John C. Lever</baseNameString>
  </baseName>
  <instanceOf>
    <topicRef xlink:href="#id44"></topicRef>
  </instanceOf>
  <resourceRef xlink:href="http://fide.uhk.cz/cleveland"></resourceRef>
  </resourceRef>
  <instanceOf>
    <topicRef xlink:href="#id15"></topicRef>
  </instanceOf>
  <resourceData>
    <resourceData>10:30-11:00</resourceData>
  </resourceData>
</topic>
```

Podoba XTM dokumentu

- Syntaxe pro tvorbu XTM dokumentu je popsána ve specifikaci XTM Topic Maps *

```
<?xml version="1.0" encoding="UTF-8" ?>
<document xmlns="http://www.topicmaps.org/xtm/1.0"
  <topic id="BartSimpson">
    <baseName>
      <baseNameString>Bart Simpson</baseNameString>
    </baseName>
    <instanceOf>
      <topicRef xlink:href="#id101"></topicRef>
    </instanceOf>
    <resourceRef xlink:href="http://www.simpson.com/BartSimpson"></resourceRef>
  </topic>
  <topic id="LisaSimpson">
    <baseName>
      <baseNameString>Lisa Simpson</baseNameString>
    </baseName>
    <instanceOf>
      <topicRef xlink:href="#id101"></topicRef>
    </instanceOf>
    <resourceRef xlink:href="http://www.simpson.com/LisaSimpson"></resourceRef>
  </topic>
  <topic id="MaggieSimpson">
    <baseName>
      <baseNameString>Maggie Simpson</baseNameString>
    </baseName>
    <instanceOf>
      <topicRef xlink:href="#id101"></topicRef>
    </instanceOf>
    <resourceRef xlink:href="http://www.simpson.com/MaggieSimpson"></resourceRef>
  </topic>
</document>
```

Soubor XML tagů

[Zdroj]: TopicMaps.org URL: <http://www.topicmaps.org/xtm/>

Pravidla pro psaní XML dokumentů

- Každý platný dokument XML bez výjimky musí být správně formulován => musí splňovat určitá pravidla:
 - Každá počáteční značka musí mít odpovídající koncovou značku
 - <topic> ... </topic>
 - Elementy se nesmějí překrývat
 - <topic> ... <association> ... </topic> ... </association>
 - Musí existovat právě jeden kořenový element
 - <topicMap> ... </topicMap>
 - Hodnoty atributů musejí být uzavřeny do uvozovek
 - <topic id = "Bart Simpson"> ... </topic>
 - Element nesmí mít dva atributy stejného názvu
 - <topic id = "Lisa Simpson" id = "Maggie Simpson">
 - Komentáře nesmějí být uvnitř značek
 - <!-- můj milý <topic id = „Barnie“ --> -->
 - ...
- XML je case-sensitive

Platné i pro XTM

Základní kostra XTM

Kořenový element Deklarace xml dokumentu

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
```

→ Kódování dokumentu

```
<topicMap xmlns="http://www.topicmaps.org/xtm/1.0/"
  xmlns:xlink="http://www.w3.org/1999/xlink" id="BartSimpsonTM">
```

→ Hodnota atributu

```
<topic id="Bart Simpson">
```

→ Název atributu

```
</topic>
```

→ Element pro deklaraci námětu, typu námětu, typu asociace, typu role i celé námětové mapy

```
<occurrence>
```

→ Element pro deklaraci typu výskytu

```
</occurrence>
```

```
</topic>
```

→ Element pro deklaraci asociace

```
<association>
```

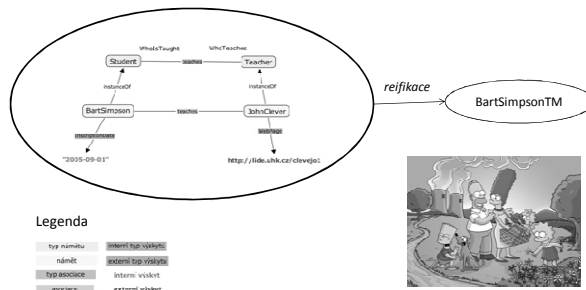
```
</association>
```

```
<mergeMap>
```

```
</mergeMap>
```

```
</topicMap>
```

Tvorba námětové mapy v XTM



(Zdroj: Obrázek URL: <http://simpsonovi.comics.cz/obrazy/obrazky/wallpapers/images/pages/image1.html>)

Založení nové námětové mapy

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
```

```
<topicMap xmlns="http://www.topicmaps.org/xtm/1.0/"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  id="BartSimpsonTM">
```

→ identifikátor námětu

```
</topicMap>
```

→ kořenový element topicMap

celá námětová mapa bude reprezentována jedním námětem

Poznámka: pozor, nezapomeňme, že XML/XTM je case-sensitive!

Tvorba typů námětů

```
<topic id="Teacher">
```

→ námi vytvořený identifikátor typu námětu (bez mezer a diakritiky)

```
<baseName>
```

```
<baseNameString>Teacher</baseNameString>
```

```
</baseName>
```

```
</topic>
```

→ pojmenování typu námětu (co se nám bude zobrazovat v podobě základního názvu v Omnigatoru)

```
<topic id="Student">
```

```
<baseName>
```

```
<baseNameString>Student</baseNameString>
```

```
</baseName>
```

```
</topic>
```

```
...
```

Tvorba námětů

```
<topic id="BartSimpson">
```

→ **námět**

```
<instanceOf>
```

→ je instancí typu námětu Student

```
<topicRef xlink:href="#Student"></topicRef>
```

```
</instanceOf>
```

→ pojmenování námětu

```
<baseName>
```

```
<baseNameString>Bart Simpson</baseNameString>
```

```
</baseName>
```

```
</topic>
```

```
<topic id="JohnClever">
```

→ **námět**

```
<instanceOf>
```

→ je instancí typu námětu Teacher

```
<topicRef xlink:href="#Teacher"></topicRef>
```

```
</instanceOf>
```

→ pojmenování námětu

```
<baseName>
```

```
<baseNameString>John Clever</baseNameString>
```

```
</baseName>
```

```
</topic>
```

Poznámka: Pozor! Nesmíme prohodit pořadí elementů instanceOf a baseName.

Tvorba typu asociací

```
<topic id="teaches">
```

→ **typ asociace**

```
<baseName>
```

```
<baseNameString>teaches</baseNameString>
```

```
</baseName>
```

```
<scope>
```

→ odkaz na typ role WhoTeaches

```
<topicRef xlink:href="#WhoTeaches"></topicRef>
```

```
</scope>
```

```
<baseNameString>teaches</baseNameString>
```

```
</baseName>
```

```
<scope>
```

→ odkaz na typ role WholsTaught

```
<topicRef xlink:href="#WholsTaught"></topicRef>
```

```
</scope>
```

```
<baseNameString>is taught by</baseNameString>
```

```
</baseName>
```

```
</topic>
```

ještě nevytvořené prvky námětové mapy

Poznámka: Pozor! Nesmíme prohodit pořadí elementů scope a baseNameString.

Tvorba typů rolí

```
<topic id="WhoTeaches"> —————> typ role
  <baseName>
    <baseNameString>who teaches</baseNameString>
  </baseName>
</topic>
<topic id="WhosTaught"> —————> typ role
  <baseName>
    <baseNameString>who is taught</baseNameString>
  </baseName>
</topic>
```

Tvorba asociace mezi náměty

```
<association id="teaches"> —————> asociace teaches jako konkrétní výskyt
  <instanceOf> —————> typu asociace teaches
    <topicRef xlink:href="#teaches"></topicRef>
  </instanceOf>
  <member>
    <roleSpec> —————> odkaz na typ asociace teaches
      <topicRef xlink:href="#WhoTeaches"></topicRef> —————> specifikace role
    </roleSpec>
    <roleSpec> —————> odkaz na typ role
      <topicRef xlink:href="#JohnClever"></topicRef>
    </roleSpec>
  </member>
  <member>
    <roleSpec> —————> odkaz na hráče role WhoTeaches
      <topicRef xlink:href="#WhosTaught"></topicRef>
    </roleSpec>
    <roleSpec> —————> odkaz na hráče role WhoTeaches
      <topicRef xlink:href="#BartSimpson"></topicRef>
    </roleSpec>
  </member>
</association>
```

Tvorba typů výskytu

```
<topic id="InscriptionDate"> —————> interní typ výskytu
  <baseName>
    <baseNameString>Inscription date</baseNameString>
  </baseName>
</topic>

<topic id="WebPage"> —————> externí typ výskytu
  <baseName>
    <baseNameString>Web page</baseNameString>
  </baseName>
</topic>
```

Tvorba interního výskytu a jeho přiřazení námětu

```
<topic id="BartSimpson">
  <instanceOf>
    <topicRef xlink:href="#Student"></topicRef>
  </instanceOf>
  <baseName>
    <baseNameString>Bart Simpson</baseNameString>
  </baseName>
  <occurrence> —————> odkaz na interní typ výskytu
    <instanceOf>
      <topicRef xlink:href="#InscriptionDate"></topicRef>
    </instanceOf>
    <resourceData>2005-09-01</resourceData>
  </occurrence>
</topic> —————> interní výskyt
```

Tvorba externího výskytu a jeho přiřazení námětu

```
<topic id="JohnClever">
  <instanceOf>
    <topicRef xlink:href="#Teacher"></topicRef>
  </instanceOf>
  <baseName>
    <baseNameString>John Clever</baseNameString>
  </baseName>
  <occurrence> —————> odkaz na externí typ výskytu
    <instanceOf>
      <topicRef xlink:href="#WebPage"></topicRef>
    </instanceOf>
    <resourceRef xlink:href="http://lde.uhk.cz/clevejo1"></resourceRef>
  </occurrence>
</topic> —————> externí výskyt
```

Vytvoření vztahu mezi námětovou mapou a tématem

- V úvodu tvorby námětové mapy už jsme předem počítali s tím, že vytvoříme vztah mezi námět. mapou a nějakým námětem

```
<topicMap xmlns="http://www.topicmaps.org/xtm/1.0/"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  id="BartSimpsonTM">
```

Teď založíme námět, který nám bude ukazovat na celou námětovou mapu ...

```
<topic id="BartSimpsonTM-topic">
  <subjectIdentity>
    <subjectIndicatorRef xlink:href="#BartSimpsonTM"/> </subjectIdentity>
  <baseName>
    <baseNameString>Bart Simpson topic map</baseNameString>
  </baseName>
</topic>
```

Tímto způsobem jsme schopni o celé námětové mapě něco říci – vytvořit metadata.

Syntaxe XTM 1.0 vs. XTM 2.0



- Změny jsou popsány v samotném standardu ISO 13250-3 (XML syntaxe) * a týkají se zejména:
 - Změny pozice určitého elementu:
 - <mergeMap> element musí být umístěn před všemi náměty a asociacemi
 - Nahrazení určitých elementů:
 - místo <baseName> se používá jen <name>
 - místo <roleSpec> se užívá jen <type>
 - místo <member> se užívá <role>
 - ...
 - Zrušení určitých elementů:
 - <variantName> a <subjectIdentity>
 - Přidání nových elementů:
 - <itemIdentity>, <subjectLocator> a <subjectIdentifier>
 - ...
- Bohužel, ve standardu ver. 2 není specifikováno, proč došlo k těmto změnám.
- Důvody jsou lehce objasněny L. M. Garsholem, viz. blíže *

[Zdroj*: Garshol L. M. Homepage. URL: <http://www.garshol.priv.no/blog/95.html>]

Syntaxe LTM



- Nestandardizovaná syntaxe vyvinutá společností Bouvet (původní myšlenka od S. Peppera rozšířená mj. L. M. Garsholem)
- Oproti XTM syntaxi je LTM syntaxe jednodušší
- Využití tam, kde je požadován rychlý vývoj menších námětových map
- Notace může být zapsána v jakémkoliv textovém editoru a následně zpracována softwarem pro námětové mapy, který LTM podporuje
- Podpora v editoru:
 - Ontopia
 - TM4J (Topic Maps For Java; engine pro tvorbu námět. map)
 - Perl XTM (engine pro tvorbu námět. map)

Příští přednáška dne 20. 3. 2014 na téma

Sémantický web a dotazování