Automatizovaný návrh pravidel pro integraci dat a sémantický web

Zdeňka Linková, Martin Řimnáč

Ústav informatiky AV ČR, v.v.i.

Znalosti 2008
Bratislava 13.-15.02. 2008

1. Motivace
   - Integrace dat pomocí pohledů
   - Automatická integrace dat
   - Formalismus pro datový zdroj
2. Návrhy mapování, míry
   - Globální matice úložiště
   - Mapování
   - Pravidlo ekvivalence
   - Pravidlo hierarchie
3. Výsledky a shrnutí
   - Využití měr
   - Závěr

Integrace dat

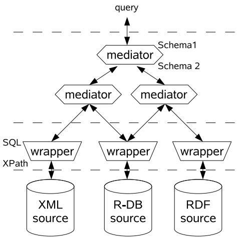
- Web jako studnice dat
  - mnoho zdrojů (stále progresivně rostoucí)
  - (částečně) redundantní popis
- Integrace dat
  - umožňuje data z lokálních zdrojů prezentovat pomocí jednoho zdroje
    - matrializovaně
    - virtuálně - pomocí pohledů

      Mediační systémy

Integrace dat pomocí pohledů

Úlohy v procesu integrace dat:

- schema matching
- schema mapping:
  - LAV (Local As View)
  - GAV (Global As View)
  - GLAV (Global Local As View)
- zpracování dotazů:
  - query rewriting



Automatická integrace dat

1. Manuální návrh integrace dat
   - návrh pravidel - človek na základě své intrepretace schématů
   - práce s mnoha daty
   - možnost omylu, důvěryhodnost třetích stran
2. (Semi)automatický návrh pravidel
   - na základě dat návrh kandidátů
     - jako (ohodnocené) doporučení pro návrháře
     - jako nejlepší možný odhad
   - ohodnocení kandidátů - kosinové fuzzy míry
     - lexikální analýza
     - strukturální analýza
     - ....

Formalismus - datový zdroj

- Formalismus binárních matic
  1. Matice úložiště

$$\Phi = [\phi_{ij}], \quad \phi_{ij} = \begin{cases} 1 & \text{pokud } e_i \rightarrow e_j \in \mathscr{I} \\ 0 & \text{jinak} \end{cases}$$

  2. Matice funkčních závislostí

$$\Omega = [\omega_{ij}], \quad \omega_{ij} = \begin{cases} 1 & \text{pokud } A_i \rightarrow A_j \in \mathscr{F} \\ 0 & \text{jinak} \end{cases}$$

- Vztah (transformace)

$$\Omega = \Delta^T \Phi \Delta \qquad \Phi' = \Phi \odot \Delta \Omega \Delta^T$$

pomocí matice aktivních domén atributů

$$\Delta = [\delta_{ij}], \quad \delta_{ij} = \begin{cases} 1 & \text{pokud } e_i = (A_j, v_\star) \in \mathscr{E} \\ 0 & \text{jinak} \end{cases}$$

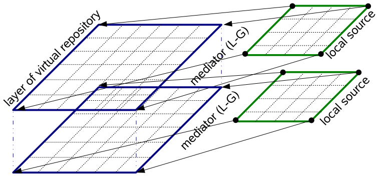
Příklad

$$\Phi = \begin{array}{c|l} \left[\begin{array}{ccc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ \hline 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right] & \begin{array}{l} \text{Město, Praha} \\ \text{Město, Košice} \\ \text{Město, Bratislava} \\ \hline \text{Stát, ČR} \\ \text{Stát, Slovensko} \\ \hline \text{Měna, CZK} \\ \text{Měna, SSK} \end{array} \end{array}$$

$$\Omega = \begin{array}{cc} \left[\begin{array}{c|c|c} 1 & 0 & 0 \\ \hline 1 & 1 & 0 \\ \hline 1 & 1 & 1 \end{array}\right] & \begin{array}{c} \text{Město} \\ \hline \text{Stát} \\ \hline \text{Měna} \end{array} \end{array} \qquad \vec{y} = \Phi \vec{x}$$

Virtuální globální matice úložiště

Centralizovaně

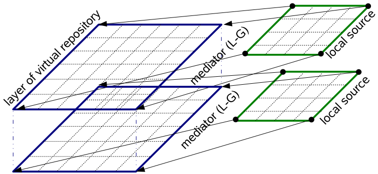


$$\Phi_{\mathscr{S}} = \sum_{\forall S_l \in \mathscr{S}} \Gamma_{S_l} \Phi_{S_l} \Gamma_{S_l}^T$$

- Γ_{S_l} .. mediátor elementů mezi Φ_{S_l} a $\Phi_{\mathscr{S}}$

Virtuální globální matice úložiště

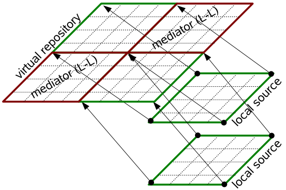
Centralizovaně



$$\Phi_{\mathscr{S}} = \sum_{\forall S_l \in \mathscr{S}} \Gamma_{S_l} \Phi_{S_l} \Gamma_{S_l}^T$$

- Γ_{S_l} .. mediátor elementů mezi Φ_{S_l} a $\Phi_{\mathscr{S}}$

Decentralizovaně

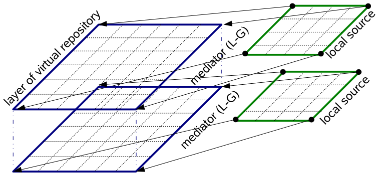


$$\Phi_{\mathscr{S}} = \begin{bmatrix} \Phi_1 & \Psi_{12} & \cdots & \Psi_{1|\mathscr{S}|} \\ \Psi_{21} & \Phi_2 & \cdots & \Psi_{2|\mathscr{S}|} \\ \vdots & & \ddots & \\ \Psi_{|\mathscr{S}|1} & \cdots & \cdots & \Phi_{|\mathscr{S}|} \end{bmatrix}$$

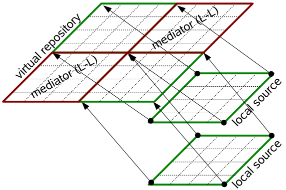
- Ψ_{ij} .. mediátor elementů mezi Φ_{S_i} a Φ_{S_j}

Virtuální globální matice úložiště

Centralizovaně



Decentralizovaně



$$\Phi_{\mathscr{S}} = \sum_{\forall S_l \in \mathscr{S}} \Gamma_{S_l} \Phi_{S_l} \Gamma_{S_l}^T$$

- Γ_{S_l} .. mediátor elementů mezi Φ_{S_l} a $\Phi_{\mathscr{S}}$

$$\Phi_{\mathscr{S}} = \begin{bmatrix} \Phi_1 & \Psi_{12} & \cdots & \Psi_{1|\mathscr{S}|} \\ \Psi_{21} & \Phi_2 & \cdots & \Psi_{2|\mathscr{S}|} \\ \vdots & & \ddots & \\ \Psi_{|\mathscr{S}|1} & \cdots & \cdots & \Phi_{|\mathscr{S}|} \end{bmatrix}$$

- Ψ_{ij} .. mediátor elementů mezi Φ_{S_i} a Φ_{S_j}

$$\Psi_{ij} = \Gamma_{S_i}^T \Gamma_{S_j}$$

Mapování na úrovni elementů

1. Binární

$$\Psi_{kl} = [\psi_{ij}^{kl}]; \psi_{ij}^{kl} = \begin{cases} 1 & \text{pokud } e_i \text{ z } S_l \text{ odpovídá } e_j \text{ z } S_k \\ 0 & \text{jinak} \end{cases}$$

Mapování na úrovni elementů

1. Binární

$$\Psi_{kl} = [\psi_{ij}^{kl}]; \psi_{ij}^{kl} = \begin{cases} 1 & \text{pokud } e_i \text{ z } S_l \text{ odpovídá } e_j \text{ z } S_k \\ 0 & \text{jinak} \end{cases}$$

2. Vážené

$$\Psi_{kl} = [\psi_{ij}^{kl}]; \psi_{ij}^{kl} = \begin{cases} \mu^{\mathcal{E}}(e_i, e_j) & \text{pokud } e_i \text{ z } S_l \text{ odpovídá } e_j \text{ z } S_k \\ & \text{s jistotou } \mu^{\mathcal{E}}(e_i, e_j) \\ 0 & \text{jinak} \end{cases}$$

Mapování na úrovni elementů

1. Binární

$$\Psi_{kl} = [\psi_{ij}^{kl}]; \psi_{ij}^{kl} = \begin{cases} 1 & \text{pokud } e_i \text{ z } S_l \text{ odpovídá } e_j \text{ z } S_k \\ 0 & \text{jinak} \end{cases}$$

2. Vážené

$$\Psi_{kl} = [\psi_{ij}^{kl}]; \psi_{ij}^{kl} = \begin{cases} \mu^{\mathcal{E}}(e_i, e_j) & \text{pokud } e_i \text{ z } S_l \text{ odpovídá } e_j \text{ z } S_k \\ & \text{s jistotou } \mu^{\mathcal{E}}(e_i, e_j) \\ 0 & \text{jinak} \end{cases}$$

3. Triviální návrh mapování

$$\Psi'_{kl} = [\psi_{ij}^{kl}];\ \psi_{ij}^{kl} = \begin{cases} 1 & \text{pokud } e_i = (A_I, v) \wedge e_j = (A_J, v) \\ 0 & \text{jinak} \end{cases}$$

Mapování na úrovni atributů

1. Binární

$$\Pi_{kl} = [\pi_{ij}^{kl}]; \pi_{ij}^{kl} = \begin{cases} 1 & \text{pokud } A_i \text{ z } S_l \text{ odpovídá } A_j \text{ z } S_k \\ 0 & \text{jinak} \end{cases}$$

Mapování na úrovni atributů

1. Binární

$$\Pi_{kl} = [\pi_{ij}^{kl}]; \pi_{ij}^{kl} = \begin{cases} 1 & \text{pokud } A_i \text{ z } S_l \text{ odpovídá } A_j \text{ z } S_k \\ 0 & \text{jinak} \end{cases}$$

2. Vážené

$$\Pi_{kl} = [\pi_{ij}^{kl}]; \pi_{ij}^{kl} = \begin{cases} \mu^{\mathscr{A}}(e_i, e_j) & \text{pokud } A_i \text{ z } S_l \text{ odpovídá } A_j \text{ z } S_k \\ & \text{s jistotou } \mu^{\mathscr{A}}(e_i, e_j) \\ 0 & \text{jinak} \end{cases}$$

Mapování na úrovni atributů

1. Binární

$$\Pi_{kl} = [\pi_{ij}^{kl}]; \pi_{ij}^{kl} = \begin{cases} 1 & \text{pokud } A_i \text{ z } S_l \text{ odpovídá } A_j \text{ z } S_k \\ 0 & \text{jinak} \end{cases}$$

2. Vážené

$$\Pi_{kl} = [\pi_{ij}^{kl}]; \pi_{ij}^{kl} = \begin{cases} \mu^{\mathscr{A}}(e_i, e_j) & \text{pokud } A_i \text{ z } S_l \text{ odpovídá } A_j \text{ z } S_k \\ & \text{s jistotou } \mu^{\mathscr{A}}(e_i, e_j) \\ 0 & \text{jinak} \end{cases}$$

- Mapování atributů versus elementů

$$\Psi_{kl} = \Psi'_{kl} \odot \Delta_{S_l} \Pi_{kl} \Delta_{S_k}^T \qquad \Pi_{kl} = \Delta_{S_l}^T \Psi_{kl} \Delta_{S_k}$$

Cosinové míry - Pravidlo ekvivalence

- Předpoklad:
  - podobné atributy budou mít podobné (aktivní) domény
- Cosinová míra:

$$\pi_{ij} = \frac{|\mathscr{D}_\alpha^{S_k}(A_i) \cap \mathscr{D}_\alpha^{S_l}(A_j)|}{|\mathscr{D}_\alpha^{S_k}(A_i) \cup \mathscr{D}_\alpha^{S_l}(A_j)|}$$

Cosinové míry - Pravidlo ekvivalence

- Předpoklad:
  - podobné atributy budou mít podobné (aktivní) domény
- Cosinová míra:

$$\pi_{ij} = \frac{|\mathscr{D}_{\alpha}^{S_k}(A_i) \cap \mathscr{D}_{\alpha}^{S_l}(A_j)|}{|\mathscr{D}_{\alpha}^{S_k}(A_i) \cup \mathscr{D}_{\alpha}^{S_l}(A_j)|}$$

- Symetrie ekvivalence

$$\Pi_{kl} = \Pi_{lk}^{T}$$

Výběr kadidátů

$$\Phi_1 = \left[\begin{array}{ccc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ \hline 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right] \begin{array}{l} \text{město, Praha} \\ \text{město, Košice} \\ \text{město, Bratislava} \\ \hline \text{stát, ČR} \\ \text{stát, Slovensko} \\ \hline \text{měna, CZK} \\ \text{měna, SSK} \end{array}$$

- Mediátor:

$$\hat{\Pi}_{12} = \begin{bmatrix} 0 & \frac{2}{3} \\ \frac{1}{3} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\Phi_2 = \left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right] \begin{array}{l} \text{země , Česká republika} \\ \text{země , Slovensko} \\ \hline \text{hlavní\_město , Praha} \\ \text{hlavní\_město , Bratislava} \end{array}$$

- preference pravidel s maximální podporou

$A_i \sim A_j$	$\mu^{\mathscr{A}}(A_i, A_j)$
hlavní_město $\sim$ město	$\frac{2}{3}$
země $\sim$ stát	$\frac{1}{3}$

Cosinové míry - Pravidlo hierarchie

- Cosinová míra:

$$\pi_{ij} = \frac{|\mathscr{D}_{\alpha}^{S_k}(A_i) \cap \mathscr{D}_{\alpha}^{S_l}(A_j)|}{|\mathscr{D}_{\alpha}^{S_k}(A_i)|}$$

Cosinové míry - Pravidlo hierarchie

- Cosinová míra:

$$\pi_{ij} = \frac{|\mathscr{D}_\alpha^{S_k}(A_i) \cap \mathscr{D}_\alpha^{S_l}(A_j)|}{|\mathscr{D}_\alpha^{S_k}(A_i)|}$$

- není symetrické

Cosinové míry - Pravidlo hierarchie

- Cosinová míra:

$$\pi_{ij} = \frac{|\mathscr{D}_{\alpha}^{S_k}(A_i) \cap \mathscr{D}_{\alpha}^{S_l}(A_j)|}{|\mathscr{D}_{\alpha}^{S_k}(A_i)|}$$

- není symetrické
- vyjma kombinací pravidel vedoucích na cykly:

$$A_j \sqsubset A_k \sqsubset A_i : A_i, A_i \in \mathscr{A}_l, A_k \in \mathscr{A}_k$$

Cosinové míry - Pravidlo hierarchie

- Cosinová míra:

$$\pi_{ij} = \frac{|\mathscr{D}_{\alpha}^{S_k}(A_i) \cap \mathscr{D}_{\alpha}^{S_l}(A_j)|}{|\mathscr{D}_{\alpha}^{S_k}(A_i)|}$$

- není symetrické
- vyjma kombinací pravidel vedoucích na cykly:

$$A_j \sqsubset A_k \sqsubset A_i : A_i, A_i \in \mathscr{A}_l, A_k \in \mathscr{A}_k$$

- Pro kombinaci ekvivalence a hierarchie - preference

$$\sigma_{ij}^{\sim} = \pi_{ij}^{kl} \cdot \pi_{ji}^{lk}$$
$$\sigma_{ij}^{\sqsubset} = \pi_{ij}^{kl} \cdot (1 - \pi_{ji}^{lk})$$

Výběr kadidátů

$$\Phi_1 = \left[\begin{array}{ccc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ \hline 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right] \begin{array}{l} \text{město, Praha} \\ \text{město, Košice} \\ \text{město, Bratislava} \\ \hline \text{stát, ČR} \\ \text{stát, Slovensko} \\ \hline \text{měna, CZK} \\ \text{měna, SSK} \end{array}$$

$$\Phi_2 = \left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right] \begin{array}{l} \text{země , Česká republika} \\ \text{země , Slovensko} \\ \hline \text{hlavní\_město , Praha} \\ \text{hlavní\_město , Bratislava} \end{array}$$

- Hierarchie:

$$\Pi_{12}^{\sqsubset} = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{2}{2} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\Pi_{21}^{\sqsubset} = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{2}{3} & 0 & 0 \end{bmatrix}$$

- preference pravidel s maximální podporou

z_2 : hlavní_město $\sqsubset$ z_1 : město	$\frac{2}{3}$	$\oplus$
z_2 : hlavní_město $\sim$ z_1 : město	$\frac{2}{9}$	$\ominus$
z_2 : stát $\sim$ z_1 : země	$\frac{1}{4}$	$\oplus$
z_2 : stát $\sqsubset$ z_1 : země	$\frac{1}{4}$	$\ominus$
z_1 : země $\sqsubset$ z_2 : stát	$\frac{1}{4}$	$\ominus$
z_1 : město $\sqsubset$ z_2 : hlavní_město	$\frac{1}{9}$	$\ominus$

Použití měr - dotaz - ekvivalence

Dotaz na Prahu

$$
\Phi_{\mathscr{S}} = \left[\begin{array}{ccc|cc|cc||cc|cc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{3} & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{3} \\
\hline
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\
\hline
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
\hline\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 1 & 0 & 1 \\
\hline
\frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{array}\right]
\begin{array}{l}
\text{město, Praha} \\
\text{město, Košice} \\
\text{město, Bratislava} \\
\hline
\text{stát, ČR} \\
\text{stát, Slovensko} \\
\hline
\text{měna, CZK} \\
\text{měna, SSK} \\
\hline\hline
\text{země , Česká Republika} \\
\text{země , Slovensko} \\
\hline
\text{hlavní\_město , Praha} \\
\text{hlavní\_město , Bratislava}
\end{array}
$$

$$
\begin{array}{rl}
\vec{x}_0 = & \left[\begin{array}{ccc|cc|cc||cc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right] \\
\vec{x}_1 = & \left[\begin{array}{ccc|cc|cc||cc|cc} 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{array}\right]
\end{array}
$$

z_1:město, z_2:hlavní_město	$\frac{2}{3}$	Praha	$1+1=2$
z_1:stát, z_2:země	$\frac{1}{3}$	ČR	1
	$\frac{1}{3}$	Česká Republika	1
z_1:měna	1	CZK	1

Použití měr - dotaz - ekvivalence

Dotaz na hlavní město Praha

$$
\Phi_{\mathscr{S}} = \left[\begin{array}{ccc|cc|cc||cc|cc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{3} & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{3} \\
\hline
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\
\hline
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
\hline\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 1 & 0 & 1 \\
\hline
\frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{array}\right]
\begin{array}{l}
\text{město, Praha} \\
\text{město, Košice} \\
\text{město, Bratislava} \\
\hline
\text{stát, ČR} \\
\text{stát, Slovensko} \\
\hline
\text{měna, CZK} \\
\text{měna, SSK} \\
\hline\hline
\text{země} = \text{Česká Republika} \\
\text{země} = \text{Slovensko} \\
\hline
\text{hlavní\_město} = \text{Praha} \\
\text{hlavní\_město} = \text{Bratislava}
\end{array}
$$

$$
\begin{array}{ll}
\vec{x}_0 = & [\begin{array}{ccc|cc|cc||cc|cc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}] \\
\vec{x}_1 = & [\begin{array}{ccc|cc|cc||cc|cc} \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{array}] \\
\vec{x}_2 = & [\begin{array}{ccc|cc|cc||cc|cc} \frac{2}{3} & 0 & 0 & \frac{2}{3} & 0 & \frac{2}{3} & 0 & 1 & 0 & 1 & 0 \end{array}]
\end{array}
$$

z_1:město, z_2:hlavní_město	$\frac{2}{3}$	Praha	$1 + 0.66 = 1.66$
z_1:stát, z_2:země	$\frac{1}{3}$	ČR	0.66
	$\frac{1}{3}$	Česká Republika	1
z_1:měna	1	CZK	0.66

Použití měr - dotaz - ekvivalence

Dotaz na město Košice

$$
\Phi_{\mathscr{S}} = \begin{array}{c}
\left[\begin{array}{ccc|cc|cc||cc|cc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{3} & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{3} \\
\hline
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\
\hline
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
\hline\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 1 & 0 & 1 \\
\hline
\frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{array}\right]
\end{array}
\begin{array}{l}
\text{město, Praha} \\
\text{město, Košice} \\
\text{město, Bratislava} \\
\hline
\text{stát, ČR} \\
\text{stát, Slovensko} \\
\hline
\text{měna, CZK} \\
\text{měna, SSK} \\
\hline\hline
\text{země} = \text{Česká Republika} \\
\text{země} = \text{Slovensko} \\
\hline
\text{hlavní\_město} = \text{Praha} \\
\text{hlavní\_město} = \text{Bratislava}
\end{array}
$$

$$
\begin{array}{l}
\vec{x}_0 = \left[\begin{array}{ccc|cc|cc||cc|cc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \\
\vec{x}_1 = \left[\begin{array}{ccc|cc|cc||cc|cc} 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & \frac{1}{3} & 0 & 0 \end{array}\right] \\
\vec{x}_2 = \left[\begin{array}{ccc|cc|cc||cc|cc} 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \end{array}\right]
\end{array}
$$

z_1:město, z_2:hlavní_město	$\frac{1}{3}$	Košice	1
	$\frac{1}{3}$	Bratislava	0.33
z_1:stát, z_2:země	$\frac{1}{3}$	Slovensko	1.33
z_1:měna	1	SKK	1

Použití měr - dotaz - Hierarchie

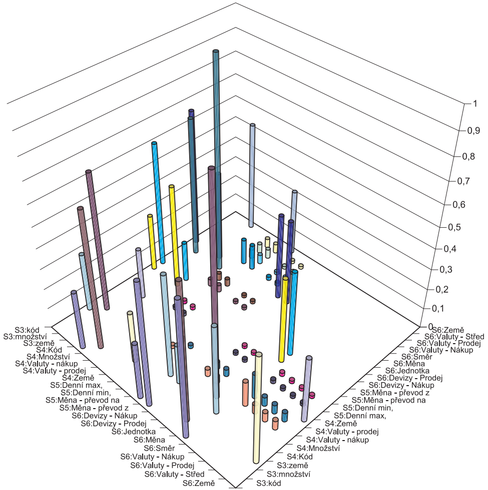
Dotaz na město Košice

$$
\Phi_{\mathscr{S}} = \left[\begin{array}{ccc|cc|cc||cc|cc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\hline
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\
\hline
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
\hline\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 1 & 0 & 1 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{array}\right]
\begin{array}{l}
\text{město, Praha} \\
\text{město, Košice} \\
\text{město, Bratislava} \\
\hline
\text{stát, ČR} \\
\text{stát, Slovensko} \\
\hline
\text{měna, CZK} \\
\text{měna, SSK} \\
\hline\hline
\text{země , Česká Republika} \\
\text{země , Slovensko} \\
\hline
\text{hlavní\_město , Praha} \\
\text{hlavní\_město , Bratislava}
\end{array}
$$

$$
\begin{array}{ll}
\vec{x}_0 = & \left[\begin{array}{ccc|cc|cc||cc|cc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \\
\vec{x}_1 = & \left[\begin{array}{ccc|cc|cc||cc|cc} 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & \frac{1}{2} & 0 & 0 \end{array}\right] \\
\vec{x}_2 = & \left[\begin{array}{ccc|cc|cc||cc|cc} 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right]
\end{array}
$$

z_1:město		Košice	1
z_2:hlavní_město		Bratislava	0.50
z_1:stát, z_2:země	$\frac{1}{2}$	Slovensko	1.50
z_1:měna	1	SKK	1

Reálná data


1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
S3:kód
S3:množství
S3:země
S4:Kód
S4:Množství
S4:Valuty - nákup
S4:Valuty - prodej
S4:Země
S5:Denní max,
S5:Denní min,
S5:Měna - převod na
S5:Měna - převod z
S6:Devizy - Nákup
S6:Devizy - Prodej
S6:Jednotka
S6:Měna
S6:Směr
S6:Valuty - Nákup
S6:Valuty - Prodej
S6:Valuty - Střed
S6:Země
S6:Země
S6:Valuty - Střed
S6:Valuty - Prodej
S6:Valuty - Nákup
S6:Směr
S6:Měna
S6:Jednotka
S6:Devizy - Prodej
S6:Devizy - Nákup
S5:Měna - převod z
S5:Měna - převod na
S5:Denní min,
S5:Denní max,
S4:Země
S4:Valuty - prodej
S4:Valuty - nákup
S4:Množství
S4:Kód
S3:země
S3:množství
S3:kód

Závěr

- Možnost (semi)automatického návrhu integračních pravidel
- Využití měr
  - při návrhu
    - ověření navržených pravidel na datech
  - při dotazování
    1. rozhodnutí při nekonzistenci (preference)
    2. ochrana lokálního zdroje před ostatními (reputace)
- Pravidla
  - vhodnost uvažovat hierarchii - nutnost dalších testů
  - polynomiální složitost
  - jednoznačné přiřazení při disjunktnosti globálních domén atributů