

Mapování ontologií (OM)

Ondřej Šváb-Zamazal
KIZI

Agenda

- Vysvětlení pojmu
- Zdůvodnění potřeby mapování
- Příklady užití
- Metody mapování
- Příklady systémů

<http://nb.vse.cz/~svabo/ukazkyOM.zip>

Schéma mapování ontologií

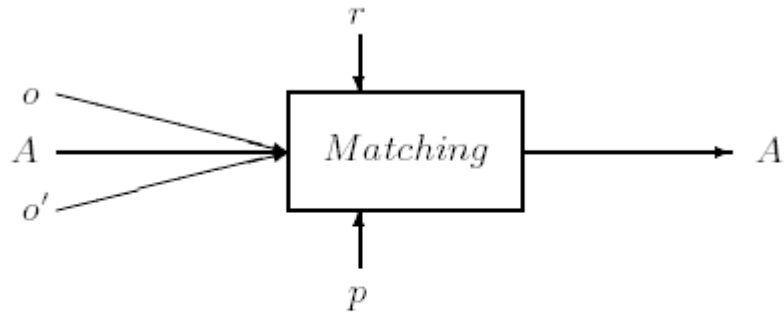


Fig. 2. The matching process

Matching

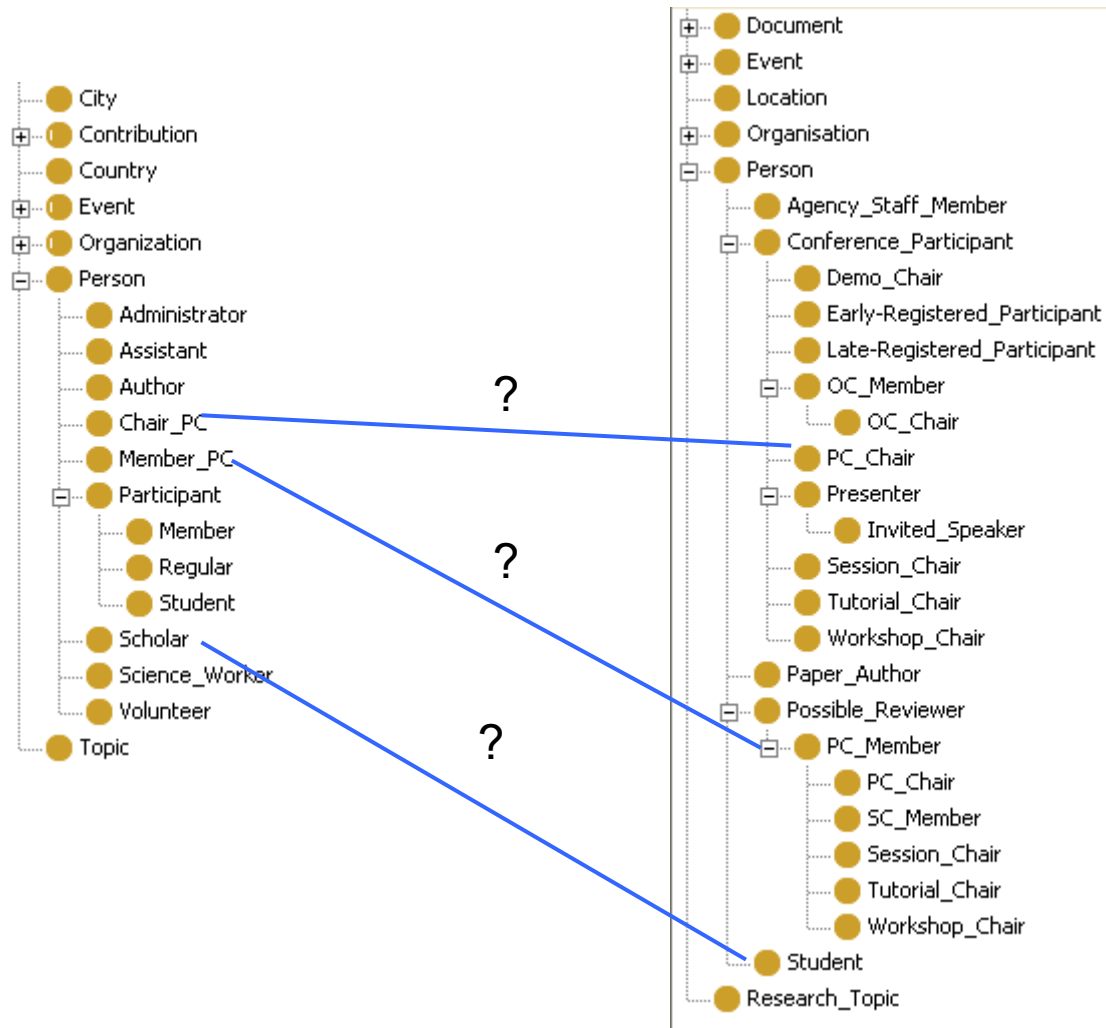
Mapping

Alignment

Potřeba mapování ontologií

- Existence řady překrývajících ontologií - pokrývající jednu oblast různým způsobem (různé definice,...)
- Příčiny heterogenity:
 - Různé pohledy na danou oblast
 - Ontologie jsou vytvářeny za nějakým účelem
 - Promítají se odborné znalosti ale i různé názory <- ontologie jsou subjektivní
 - Různé konvence při návrhu ontologií, které na druhou stranu mohou být v rámci jedné komunity „široce“ sdíleny – návrhové vzory, dobré zásady
- Možné záporné důsledky:
 - Nástroje, které používají různé ontologie se nemohou „domluvit“ -> potřeba sémantické integrace
 - př. Komunikace agentů (2 ontologie)
- Řešení:
 - 1 závazná ontologie - NE
 - Nelze; distribuovaný (charakter) charakter webu (znalostí), žádná jedna správná ontologie
 - Hledání vztahů mezi elementy ontologií - snad ANO

Příklad OM



For Project: ● ekaw

Asserted Hierarchy

- owl:Thing
 - ▼ ● Document
 - ▼ ● Abstract
 - Invited_Talk_Abstract
 - Tutorial_Abstract
 - Flyer
 - ▼ ● Multi-author_Volume
 - ▼ ● Proceedings
 - Conference_Proceedings
 - ▼ ● Paper
 - Camera_Ready_Paper
 - Conference_Paper
 - Demo_Paper
 - Industrial_Paper
 - Poster_Paper
 - Regular_Paper
 - ▼ ● Submitted_Paper
 - ▼ ● Assigned_Paper
 - ▼ ● Evaluated_Paper
 - Accepted_Paper
 - Rejected_Paper
 - Workshop_Paper
 - Programme_Brochure
 - ▼ ● Review
 - Negative_Review
 - Neutral_Review
 - Positive_Review
 - Web_Site

For Project: ● PCS

Asserted Hierarchy

- owl:Thing
 - ▼ ● Conference
 - Programme
 - ▼ ● DOCUMENT
 - ▼ ● Paper
 - Accepted_paper
 - Rejected_paper
 - Review
 - ▼ ● Electronic_proceeding
 - CD
 - Web_site
 - ▼ ● Email
 - Information
 - Invitation
 - ▼ ● PERSON
 - Administrator
 - ▼ ● ConferenceMember
 - Author
 - ▼ ● CommitteeMember
 - Primary_Reviewer
 - Secondary_Reviewer
 - External_Reviewer
 - Report
 - System

Use-cases (scénáře)

- portál porovnávání výrobků
 - Porovnávání výrobků z různých katalogů
 - Katalogy jsou tezaurové struktury, ontologie
 - Aplikace typu „Catalog integration“ („Data translation“)
- Při fúzování firem
 - Odlišná databázová schémata, nutnost uvést je do souladu
 - Aplikace typu „schema integration“ („merging“)

Mapování ontologií

- 1. mapování ontologií (alignment)
- 2. použití mapování (merging, transforming, ...)

Použití mapování - sémantická integrace

- Datová transformace pomocí integrace schémat, ontologií
 - Uvedení do souladu řady databázových schémat (mapování schémat), ontologií
 - Aplikační oblasti:
 - Datové sklady
 - Elektronické obchodování (ontologie produktů)
 - komunikační agenti
- Odvozování nad mapováními
- Zprostředkování dotazů – např. v peer-to-peer systémech (přeformulování dotazů s použitím mapování mezi peers)

Kategorie aplikací

Application	instances	run time	automatic	correct	complete	operation
Ontology evolution	✓			✓	✓	transformation
Schema integration	✓			✓	✓	merging
Catalog integration	✓			✓	✓	data translation
Data integration	✓			✓	✓	query answering
P2P information sharing		✓				query answering
Web service composition		✓	✓	✓		data mediation
Multi agent communication		✓	✓	✓	✓	data translation
Context matching in ambient computing		✓	✓	✓		data translation
Query answering	✓	✓				query reformulation
Semantic web browsing	✓	✓		✓		navigation

Table 2.1: Summary of applications requirements.

Ontologie podle složitosti (komplexita)

- Teseurus-like
- Schema-like (SQL data types)
- Bohaté axiomy (dle DL)

- Tyto všechny mohou být vyjádřeny jako ontologie.
- Všechno jsou to modely.
- Ontologie je „znalostní“ model.
- + data (instance)

Charakter účelu mapování

- Pro kontrolu konzistence
 - Nutno uvažovat i logické axiomy
- Pro lepší vyhledávání
 - Menší důraz na axiomy
- Pro integraci dat podle integrovaných modelů
 - Podle datových typů a struktury

Vztahy mezi elementy ontologií

- Ekvivalence
- Subsumpce

- Překrytí (overlapping)
- Rozdílnost (disjunction)

Elementy ontologií:

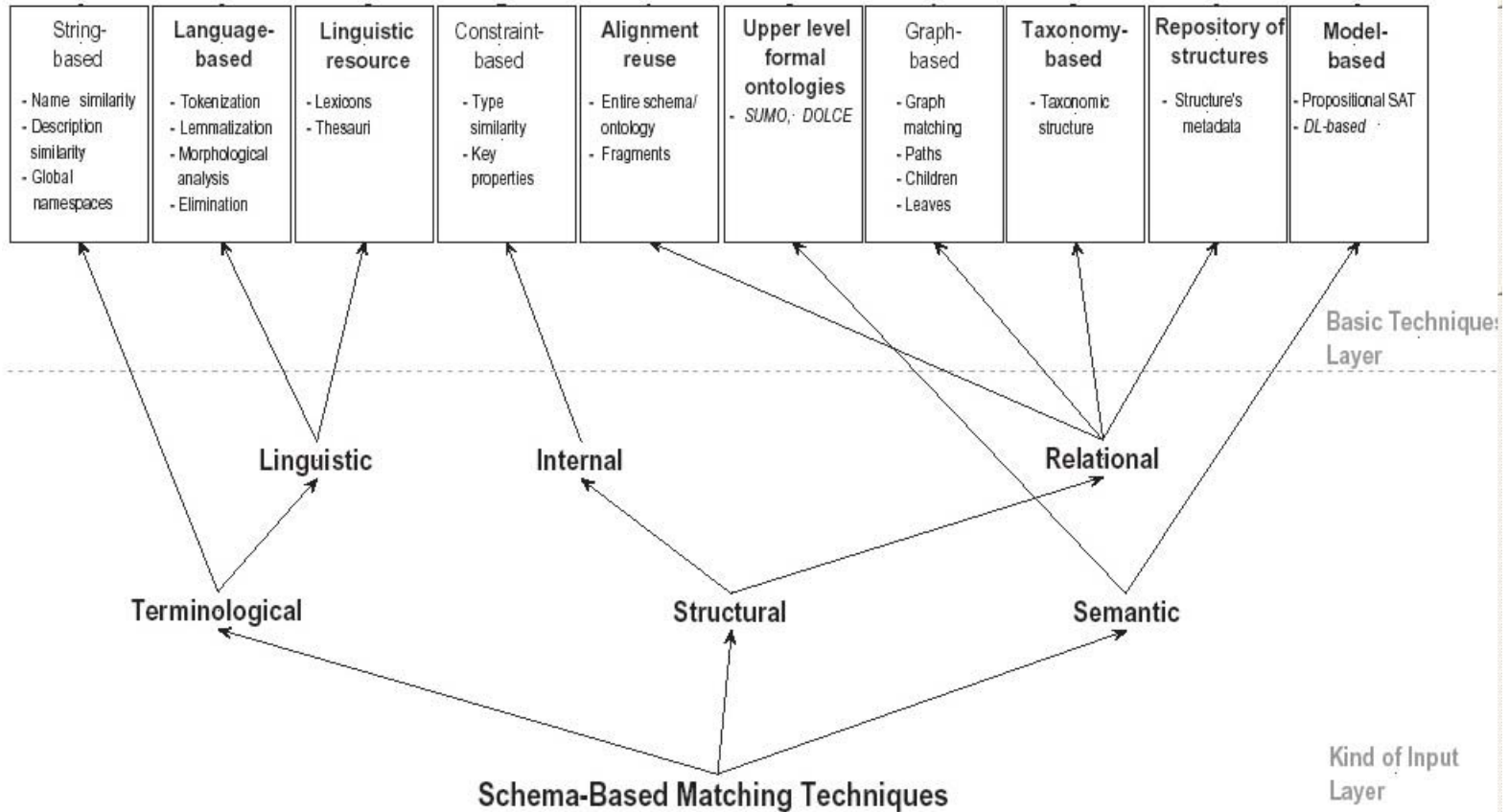
- Třídy, vlastnosti, instance

Metody mapování ontologií – mapování jejich elementů

■ Pomůcky při mapování

- Referenční ontologie (SUMO, Dolce, Cyc)
- Terminologické nástroje
 - Řetězcová normalizace (malá písmena, stop-slova)
 - Vzdálenost slov – editační vzdálenost, ...
 - Tezaury – např. WordNet
- Vstup uživatele (úvodní namapování, zpětná vazba)
- Předcházející mapování
- Struktura ontologií

Techniky mapování - klasifikace



Metody mapování ontologií

- Pravidlové
- Řetězcové
- Grafové – Similarity flooding
- Pravděpodobnostní
- Kombinování výsledků jiných metod pomocí bayesovských sítí

Argumenty namapování

- Interní struktura – př. Vlastnosti třídy C jsou namapovány na třídu C'
- Externí struktura – př. Třídy C a C' mají namapované sousedy
- Terminologické – řetězcová podobnost
- Extensionální – př. instance tříd C a C' jsou namapovány
- Sémantická interpretace – deskripční logiky

Různé přístupy OM

- Intenzionální
- Extenzionální
- Interní
- Externí

Řetězcové metody

- Porovnávání jmen konceptů a jejich popisků (vlastnosti comment a description)
- Vzdálenost založená na míře překrytí slov
 - CharJaccard, Jaccard, Jaro
- Editační vzdálenost slov
 - Levenshtein, Smith-Waterman

Cupid

- Generický algoritmus pro mapování schémat (lingvistický a strukturální přístup)
- Pravidla:
 - Elementy (listy) ve stromech jsou ekvivalentní, jestliže se shodují lexikálně a datotypově a jestliže elementy v jejich sousedství jsou rovněž ekvivalentní
 - Dva elementy (nelistové) jsou ekvivalentní, jestliže jsou lexikálně shodné a jestliže jejich podstromy jsou shodné
 - ...

Similarity Flooding

- 1. ontologie jsou převedeny na orientované pojmenované grafy
- Principy:
 - A) Elementy jsou podobné, pokud jsou podobní jejich sousedi
 - B) Podobnost je částečně propagována k sousedům (fixpoint computation)
 - C) filtrují se podobnosti->alignment

Similarity Flooding

■ Algorithm

1. $G1 = \text{Graph}(S1); G2 = \text{Graph}(S2)$
2. $\text{initialMap} = \text{StringMatch}(G1, G2)$
3. $\text{product} = \text{SFJoin}(G1, G2, \text{initialMap})$
4. $\text{result} = \text{selectThreshold}(\text{product})$

HCONE algoritmus

- Používají WordNet and LSI pro mapování ontologií
- C_1 je koncept z ontologie O_1
- 1. získej všechny synsety z WordNetu (WordNet senses) – m
- 2. získej všechny hypernoma (nadtržidy) and hyponoma (podtržidy) m synsetů z kroku 1 - n 'termů'
 - Všechny se nepoužijí. Jen ty, které se vyskytly často v sousedství m synsetů z 1. kroku
- 3. Sestavení $n \times m$ asociační matice termů-synsetů
- 4. dotaz jako pseudo-dokument – termy ze sousedství konceptu C_1 (vektor v k -dimenzionálním prostoru)
- 5. spočítat podobnosti 'dokumentů' (synsetů) s dotazem pseudo-dokumentem
- 6. dokument (synset), který má nejvyšší podobnost je vybrán jako ideální vyjádření smyslu konceptu C_1

To samé pro koncept C_2 z ontologie O_2 . Namapování pak mezi přiřazenými „smysly“.

Vyhodnocení (evaluation)

- Porovnání nalezených mapování vzhledem k referenčnímu mapování
- Tradiční měřítka
 - Přesnost (precision)
 - Úplnost (recall)

$$\begin{array}{ll} \textit{precision} & \pi = \frac{tp}{tp+fp}, \\ \textit{accuracy} & \eta = \frac{tp+tn}{tp+tn+fn+fp}, \end{array}$$

$$\begin{array}{ll} \textit{recall} & \rho = \frac{tp}{tp+fn}, \\ \textit{F}_1\text{-measure} & F_1 = \frac{2\pi\rho}{\pi+\rho}. \end{array}$$



Vybrané ukázky nástrojů

Falcon-AO

Drago

<http://nb.vse.cz/~svabo/ukazkyOM.zip>

Vybraná ukázka – Falcon-AO

- Kombinuje tři mapovací metody:
 - V-Doc – vytvářejí virtuální dokumenty pro elementy ontologií – zahrnují slova z popisků, názvy a další „okolní“ informace (názvy sousedů, předků, potomků, vlastností apod.), podobnost elementů je pak počítána ve vektorovém prostoru
 - I-Sub – terminologické porovnávání
 - GMO – měření podobnosti struktur mezi RDF grafy
 - Modul pro kombinaci výsledků

Demo – Falcon-AO

■ Nástroj:

<http://xobjects.seu.edu.cn/project/falcon/matching/resources.html>

■ Ontologie:

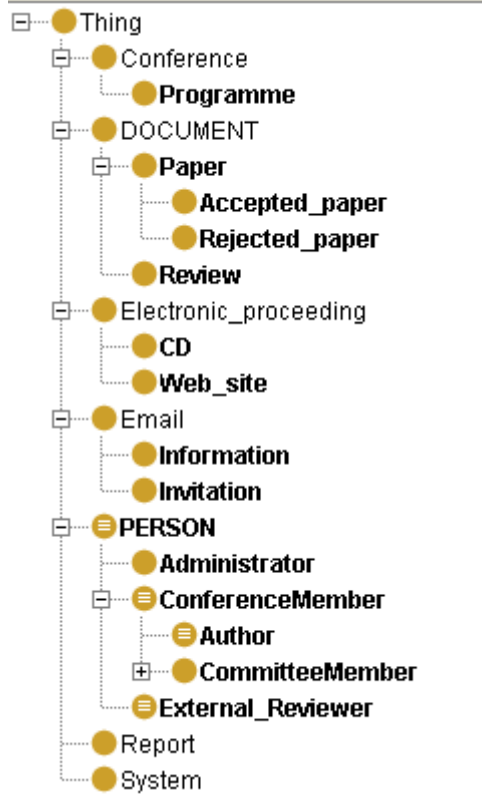
□ #1

■ <http://nb.vse.cz/~svabo/oaei2006/data/PCS.owl>

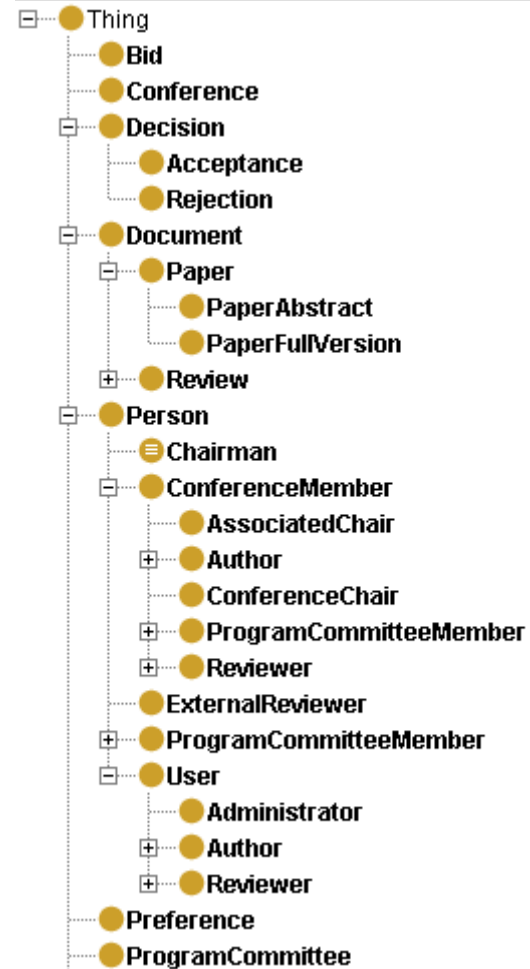
□ #2

■ <http://nb.vse.cz/~svabo/oaei2006/data/cmt.owl>

PCS.owl



cmt.owl



Drago – Distributed Reasoning Architecture for Galaxy of Ontologies

■ Nástroj pro distribuované odvozování

- Pracuje s DDL (Distributed Description Logics)

■ Služby

- Kontrola konzistence ontologie,
- odvození taxonomie (klasifikace),
- ověření splnitelnosti konceptů,
- kontrola odvození (entailment)

- Resource: [<http://drago.itc.it/>]

DDL

- Umožňuje reprezentovat sémanticky propojené (namapované) ontologie
 - Rozšíření deskripční logiky (lokální interpretace, distribuovaná interpretace,...)
 - Distribuovaný T-box
- Stručně formální aparát:
 - Sémantické relace (mapování) reprezentované tzv. bridge rules:

$$i : X \xrightarrow{\sqsubseteq} j : Y \text{ - an into-bridge rule}$$
$$i : X \xrightarrow{\sqsupseteq} j : Y \text{ - an onto-bridge rule}$$

- Závisí na pořadí ontologií, z pohledu ontologie j

DDL inferenční mechanismus

- Rozšíření tablového algoritmu
- Odvození „nové“ subsumpce pomocí ‘subsumption propagation mechanism’

$$\frac{i : A \sqsubseteq B, \quad i : A \xrightarrow{\exists} j : G, \quad i : B \xrightarrow{\sqsubseteq} j : H}{j : G \sqsubseteq H}$$

- A jeho zobecnění s disjunkcemi,...

Drago - architektura

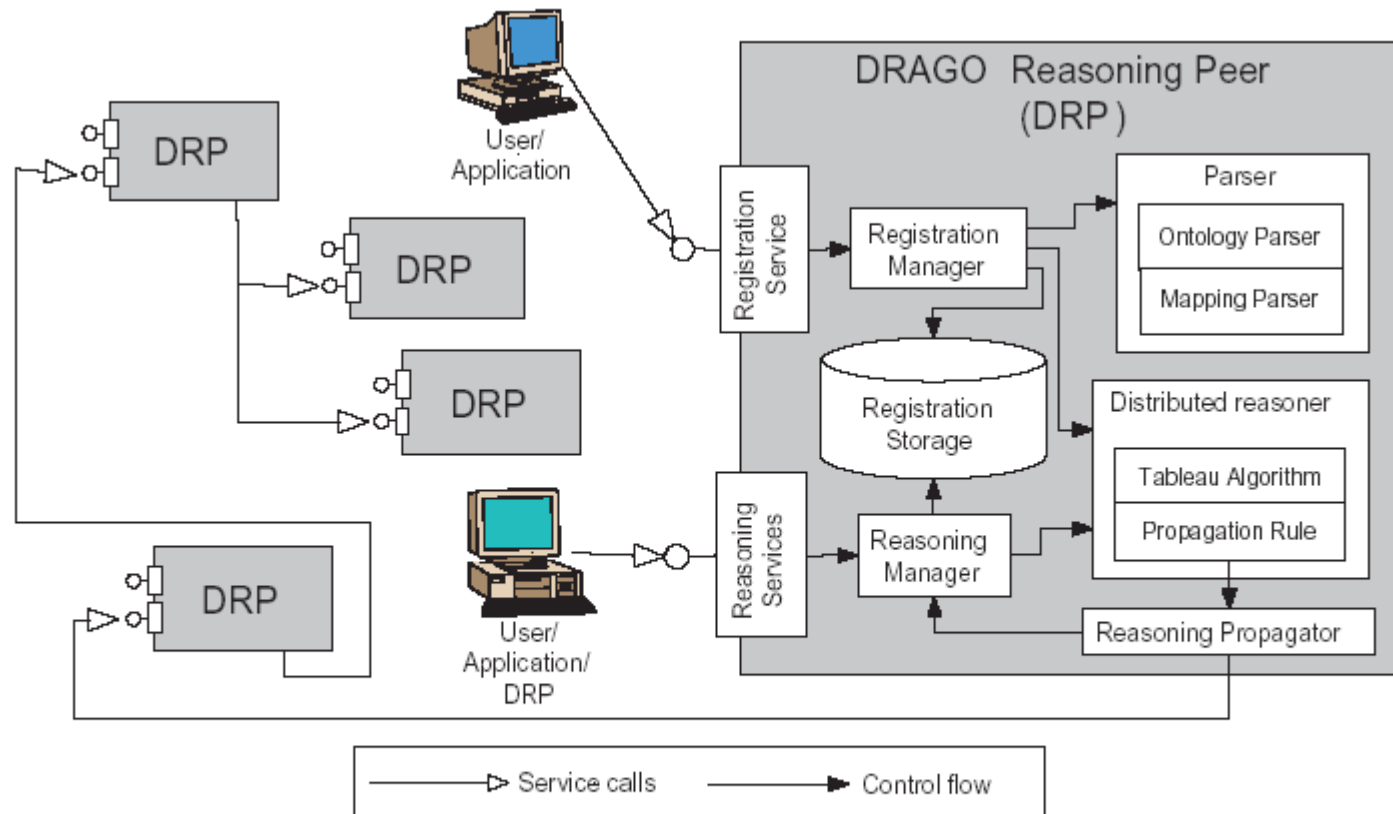


Fig. 5. DRAGO architecture.

DRP=Drago Reasoning Peer

peer-peer network of DRPs

Drago - implementace

- Ontologický jazyk OWL
- Mapování mezi ontologiemi v jazyce C-OWL
- Distributed Reasoner – rozšíření OWL reasoner Pellet (<http://www.mindswap.org/2003/pellet/>)
- Komunikace DRP protokolem HTTP

Drago - ukázky

- 1. odvození nového vztahu subsumpce, obohacení distribuované taxonomie cílové ontologie
- 2. množina mapování vedoucí k nesplnitelným konceptům v cílové (target) ont

Unsatisfiable classes

CONFTOOL:Conference
CONFTOOL:Paper
CONFTOOL:Poster
CONFTOOL:Short_paper
CONFTOOL:Tutorial

Removed bridge rules

PCS:Conference $\xrightarrow{\text{E}}$ CONFTOOL:Organization,
PCS:Report $\xrightarrow{\text{E}}$ CONFTOOL:Event,
PCS:Report $\xrightarrow{\text{E}}$ CONFTOOL:Organization,
PCS:PERSON $\xrightarrow{\exists}$ CONFTOOL:Poster,
PCS:Accepted_paper $\xrightarrow{\text{E}}$ CONFTOOL:Event,

Reference

- Euzenat J. et al. D2.2.3: State of the art on ontology alignment. KnowledgeWeb. 2004.
- Euzenat J. et al. D1.2.6: Case-based recommendation of matching tools and techniques. 2006. Draft.
- Shvaiko P., Euzenat J. A Survey of Schema-based Matching Approaches. Journal on Data Semantics, 2005.
- Serafini L., Tamilin A. DRAGO: Distributed Reasoning Architecture for the Semantic Web. ESWC'05. Springer-Verlag. 2005.
- Melnik S., Garcia-Molina H., Rahm E. Similarity flooding: A versatile graph matching algorithm. ICDE 2002.
- OAEI-2007. <http://nb.vse.cz/~svabo/oaei2007/>