

Mapování schémat v prostředí Sémantického webu*

Zdeňka Linková

3rd ročník PGS, email: linkova@cs.cas.cz

Katedra matematiky, Fakulta jaderná a fyzikálně inženýrská, ČVUT
školicel: Július Štuller, Ústav informatiky Akademie Věd ČR

Abstract. The paper deals with steps of the non-materialized data integration, and focuses on schema matching and schema mapping issues. The proposal is for data sources on the Semantic Web; the crucial assumption for the considered task is the availability of the ontologies describing data to integrate. These ontologies are used to find correspondences between source schemas elements, and also for found mapping expression.

Abstrakt. Článek se zabývá úlohami, které je třeba řešit při nematerializované integraci dat. Zaměřuje se na hledání korespondencí mezi schémata a mapování schémat. Návrh přístupu řešení těchto úloh na Sémantickém webu těží z dostupných ontologiích popisujících integrované zdroje. Ontologie jsou využity jak k hledání mapování, tak i při jejich popisu.

1 Úvod

Integrace dat [1] je úloha, která se zabývá sloučením dat. Jejím cílem je prezentovat data pocházející z různých datových zdrojů jako jediný celek a umožnit je zpracovávat, jako by z jediného datového zdroje opravdu pocházela. V případě tzv. nematerializovaného přístupu [25] bývá řešením úlohy poskytnutí unifikovaného pohledu na zdroje dat. Tento pohled je využíván jako nový zdroj obsahující všechna data. Ve skutečnosti jde o pohled virtuální a data zůstávají fyzicky uložena v původních zdrojích.

Aby bylo možné integraci založit na využití virtuálního pohledu, neboli aby bylo možné k datům přes tento pohled přistupovat, je nezbytné definovat jeho vazby na fyzická data. Proto je třeba se v tomto přístupu zabývat schémata dat. Vazby mezi pohledem a daty se pak zajistí definováním vztahů mezi jednotlivými částmi schématu pohledu a částmi schémat původních zdrojů. Ty jsou pak dále využity při zpracování dat, například při dotazování.

Proces integrace je možné nahlížet jako kolekci několika úloh, které společně přinášejí požadovaný výsledek. Základní kroky při řešení integrace dat pomocí virtuálního pohledu jsou:

- **Hledání korespondencí mezi schémata** (schema matching) - Za předpokladu, že datové zdroje, které mají být integrovány, byly vytvořeny nezávisle, různými designéry a pro různé účely, jsou jejich schémata obecně heterogenní. Proto je důležitou úlohou nalezení jejich vzájemných korespondencí. Problém hledání korespondencí mezi schémata bývá označován jako schema matching [20], [21].

*Práce byla podpořena projektem 1ET100300419 programu Informační společnost (Tématického programu II Národního programu výzkumu v ČR: “Inteligentní modely, algoritmy, metody a nástroje pro vytváření sémantického webu”) a výzkumným záměrem AV0Z10300504 “Informatika pro informační společnost: Modely, algoritmy, aplikace”.

- **Mapování schémat** (schema mapping) - Obvyklým způsobem jak vyjádřit nalezené souvislosti mezi schématy zdrojů je použití tzv. *mapování*. Mapování je struktura, např. množina tvrzení, která popisuje vazbu mezi elementy schématu pohledu (obvykle označovaného jako globální schéma) a schémat datových zdrojů (označovaných jako lokální schémata).

Při tvorbě mapování jsou využívány dva základní přístupy [14], [3]: *Global As View* (GAV) přístup, který spočívá v definování globálního schématu jako množiny pohledů nad lokálními schématy, a *Local As View* (LAV) přístup, který definuje lokální schéma zdroje pomocí pohledů nad globálním schématem. Je samozřejmě možné oba přístupy kombinovat.

- **Zpracování dotazů** (query processing) - Vytvoření mapování je stěžejní úloha, jejíž výsledek má důležitou roli při přístupu k datům pomocí dotazů.

Při použití systému, který integruje data, klade uživatel dotazy tvořené nad poskytnutým pohledem, tj. využívá jeho jazyk, schéma atd. Pro vyhodnocení dotazu nad daty je třeba původní (globální) dotaz nějakým způsobem zpracovat [18].

Tím se zabývají dva základní přístupy. Prvním je *query rewriting* - dotaz je dekomponován na části odpovídající lokálním zdrojům. Ty jsou dále přepsány tak, aby byly vyjádřeny v prostředí příslušného lokálního zdroje. Nad lokálními zdroji jsou pak vzniklé lokální dotazy vyhodnoceny a ze získaných lokálních odpovědí je opět sestavena globální odpověď, která je vrácena jako výsledek na původní dotaz.

Druhou možností je *query answering*, která nijak nespécifikuje, jak má být daný dotaz zpracován. Jejím cílem je využít všechny dostupné informace k získání odpovědi na dotaz. Příkladem může být hledání takových dat, u nichž lze dle dostupných znalostí usuzovat, že jsou hledaným výsledkem.

Tento článek se dále zabývá prvními dvěma kroky integrace, tedy hledání korespondencí mezi schématy a jejich popisem pomocí vhodné struktury. Zaměřuje se na datové zdroje Sémantického webu.

Sémantický web [2], [13], [8] je zamýšlen jako sémantické rozšíření webu současného. V současné době jsou hlavními technikami při popisu dat Sémantického webu především:

- jazyk XML [33] pro strukturování dat
- RDF(S) [30], [31] pro popis metadat
- OWL [28] pro specifikaci ontologií.

Omezení na data Sémantického webu spočívá v požadavku vyjádření dat pomocí RDF/XML a dostupných ontologií [6] popisující jednotlivé zdroje.

Článek je členěn následovně: Kapitola 2 představuje obecně úlohu hledání korespondencí spolu s přístupy, které se touto úlohou zabývaly. Kapitola 3 se zaměřuje na ontologický přístup hledání korespondence mezi schématy na Sémantickém webu. Vyjádření mapování se věnuje Kapitola 4.

2 Úloha hledání korespondencí mezi schémata

Vstupem v úloze hledání korespondencí jsou dvě schémata, mezi nimiž je třeba nalézt vzájemné vztahy. Tato úloha je předmětem mnoha výzkumů. Bohužel však je v různých projektech či implementacích řešena především manuálně [15], tzn. je založena na lidském zásahu, uživatel - člověk je ten, kdo vztahy nalezne. To s sebou přináší mnohá omezení, je to například časově náročné, náchylné k chybám, drahé. Přírozená snaha o zautomatizování prováděné operace má ovšem většinou za následek pouze tzv. *kandidáty* možných korespondencí a je to opět člověk, kdo musí rozhodnout, zda nalezené možné korespondence skutečně platí.

Způsoby, kterými bývají korespondence hledány, lze rozdělit na základě úrovně informací, které jsou při porovnávání schémat využívány:

- **Na úrovni instancí** - Srovnávací přístupy pracují s vlastními daty ze zdrojů, aby našly korespondence mezi jejich schémata.
- **Na úrovni použitých pojmů** - Přístupy pracující na této úrovni bývají lingvisticky založené (např. jsou založené na jménech a textových popisech elementů schémat). Mohou pracovat se známými vztahy mezi použitými pojmy (synonyma, homonyma, apod.) nebo mohou pojmy zpracovávat jako řetězec znaků (a využívat vztahů jako je prefix, sufix, kořen apod.)
- **Na úrovni struktury** - Při hledání korespondencí (především mezi schémata, které mají složitější strukturu) bývá brána v úvahu i vlastní struktura zdroje. K porovnávání struktur mohou být využity například metody z oblasti teorie grafů.

Tyto techniky mohou být samozřejmě kombinovány. Například při porovnávání jednotlivých elementů schémat je možné brát v úvahu jak jejich jména, datové typy, aktivní domény, ale i jejich strukturu.

Možnost existence mapování, ke kterému se nalezený kandidát vztahuje, bývá často vyjádřena pomocí nějaké funkce, která podobnost porovnávaných elementů vyjadřuje. Je možné ji založit na pravděpodobnosti [16], kosinové míře příznakových vektorů [23], nebo míře vyjadřující počet shodných zkoumaných aspektů [27]. Použitá míra může být využita při výběru skutečných korespondencí z kandidátů, čímž je možné lidský zásah více eliminovat. Někdy jsou navíc použity i další techniky, jako například zpřesňování kandidátů [7] či machine learning [26].

3 Hledání korespondencí mezi schémata na Sémantickém webu

V prezentovaném přístupu se předpokládá, že spolu s integrovanými zdroji jsou k dispozici také ontologie, které popisují data uložená ve zdrojích. Pomocí nich jsou vyvozovány požadované korespondence mezi jednotlivými elementy schémat. Jelikož je přístup orientován na Sémantický web, předpokládá, že ontologie zdrojů jsou vyjádřeny v jazyce OWL [28].

V obecném případě může jeden element korespondovat s jedním nebo více jinými elementy, může korespondovat s kombinací elementů, nebo nemusí korespondovat s žádným jiným elementem. V této souvislosti se obvykle při hledání korespondencí používá pojem *kardinalita*, která pro určitou korespondenci vyjadřuje, kolik elementů mapovaných schémat do vztahu vstupuje. Kardinalita korespondence může být 1:1, 1:N, N:1, N:M. Ovšem většina existujících přístupů využívá kardinalit 1:1 or 1:N.

Prezentovaný přístup uvažuje vztahy následujících kardinalit:

- **1:1** - při vzájemném porovnávání dvou schémat. Tento případ vyjadřuje, že element jednoho schématu je ve vztahu s jedním elementem druhého schématu.
- **1:N** - při porovnávání jednoho schématu s více dalšími schématy. Tento případ je možné vidět jako množinu korespondencí kardinalit 1:1. Kardinality 1:N se často využívá v integraci dat pro vyjádření korespondencí mezi schématem globálního virtuálního pohledu a schématy lokálních zdrojů.

Pojetí korespondence při porovnávání schémat je formalizováno následovně:

- **Korespondence** kardinality **1:1** je tvrzení:

$$\varepsilon_1 \rho \varepsilon_2$$

kde

ε_1 je element jednoho schématu

ε_2 je element druhého schématu

ρ je vztah mezi ε_1 a ε_2 , který vyjadřuje jejich vzájemnou korespondenci.

- **Korespondence** kardinality **1:N** je množina tvrzení kardinalit 1:1:

$$\{\varepsilon_1 \rho_i \varepsilon_i\}$$

kde

ε_1 je element jednoho schématu

ε_i je element druhého schématu

ρ_i je vztah mezi ε_1 a ε_i , který vyjadřuje jejich vzájemnou korespondenci.

Vztahem ρ mohou být následující druhy korespondencí:

- **Is-a** hierarchický vztah (tj. jeden element je obecnější než druhý, nebo naopak). Tento druh je označen jako \supseteq , resp. \subseteq .
- **Ekvivalence** mezi elementy. Tento druh je označen jako $=$.
- **Disjunktnost**, tj. mezi elementy není žádná souvislost.

3.1 Hledání korespondencí v případě sdílené ontologie

V nejjednodušším případě je popis všech zdrojů dostupný v jediné ontologii. Tato ontologie je lokálními zdroji sdílena a pokrývá popis všech lokálních dat. Vztahy mezi elementy jednotlivých schémat mohou být nalezeny přímo v této ontologii.

Pro to je použito pravidlo:

Sémantický vztah mezi pojmy definovaný v ontologii implikuje stejný vztah mezi elementy schémat, které jsou těmito pojmy označené.

Uvažujeme-li dříve zmíněné typy korespondencí, je možné přístup založit na *is-a hierarchii* definované sdílenou ontologií. Jsou-li porovnávána dvě schémata, pro každý element jednoho schématu a každý element druhého schématu je jejich vztah hledán v této ontologii. Je-li mezi nimi vztah nalezen, je příslušná korespondence i mezi uvažovanými elementy.

Některé vztahy nemusí být v ontologii vyjádřeny přímo, ale je možné je z ontologie získat využitím tranzitivity *is-a* vztahu. Je-li například použit přístup k ontologii jako grafu s třídami popisujícími jednotlivé pojmy jako uzly a s orientovanými hranami vyjařujícími existenci *is-a* vztahu, nalezenou korespondenci neznamena pouze existující hrana, ale také příslušně značená cesta.

V případě, že jsou elementy disjunktí, znamená to, že by v *is-a* hierarchii neměla být žádná cesta a není tedy nutné nějaký vztah hledat. V praxi vede tato situace ke stejnému efektu, jako když je vztah hledán, ale žádný není nalezen. Ovšem je vhodné tuto informaci o disjunktnosti dále uchovávat, protože může být dále využita při rozšiřování přístupu například o další usuzování apod.

Všechny korespondence, které jsou ze sdílené ontologie získány, jsou přijaty. Není na ně nahlíženo nejprve jako na kandidáty, neboť zde není žádný odhad korespondencí - všechny z nich jsou v dané ontologii definovány. Tento krok tedy nevyžaduje žádný zásah (lidského) uživatele.

3.2 Obecný případ hledání korespondencí založený na ontologiích

Obecně nemusí být ontologie, která by popisovala všechna zpracovávaná data, dostupná. Některé zdroje mohou sdílet některé pojmy, avšak sdílení všech pojmů všemi zdroji nelze předpokládat. Je třeba pracovat obecně s více ontologiemi.

Sloučením všech ontologií, které popisují integrované datové zdroje, získáme “novou” sdílenou ontologii, a tak je tento obecný případ převeden na předchozí. Slučováním ontologií se zabývá řada výzkumů z oblastí ontology alignment a ontology merging a je tedy pro toto možné využít některou ze známých metod.

V souvislosti s ontologiemi, pojmy alignment a merging spolu úzce souvisí [10]. Pro oba jsou také relevantní úlohy hledání korespondencí (matching) a mapování (mapping). *Ontology alignment* obvykle označuje stanovení binárních vztahů mezi dvěma ontologiemi. To umožňuje definovat způsob, jak tyto ontologie sloučit. Výsledkem *ontology merging* je nová integrovaná ontologie.

Metody ontology alignment a ontology merging jsou, podobně jako metody při porovnávání schémat, provozovány na několika úrovních: *instance* (např. srovnání množiny instancí popisovaného pojmu), *element* (např. lexikální techniky) a *struktura* (např. grafové techniky), a také využívají nejen sémantické, ale i syntaktické přístupy.

V obou oblastech lze najít i podobnost s používáním kandidátů. Metody vyžadují lidskou interakci nebo jsou založeny na heuristikách z předchozích rozhodnutí. Ačkoliv při odvozování vztahů schémat ze sdílené ontologie žádní kandidáti nevznikají a korespondence jsou přímo určeny, v obecném případě mohou vznikat právě při využívání existujících metod při řešení podúlohy jak sdílenou ontologii najít.

Je patrné, že metody pro hledání korespondencí v ontology merging a ontology alignment jsou založeny na podobných principech jako metody pro hledání korespondencí mezi schémata. Důvodem toho je, že ontologie a datová schémata spolu úzce souvisí. Hlavním důvodem je účel, ke kterému jsou použity. Ontologie jsou vytvářeny, aby popisovaly pojmy používané v nějaké oblasti, zatímco schémata jsou vytvářena, aby modelovala nějaká konkrétní data. Speciálně pro schémata využívající sémantický model není často patrný rozdíl a není zřejmý způsob, jak identifikovat, která reprezentace je schéma a která je ontologie. V praxi mají často schémata i ontologie dobře definované použité pojmy. Protože schémata obecně neposkytují explicitní sémantiku pro data, používají se při hledání korespondencí techniky, pomocí nichž se odhaduje význam užívaných pojmů. Předpokládáme-li, že datové zdroje jsou popsány v dostupných ontologiích, použití takových technik není nutné, neboť potřebnou informaci máme.

Metodami pro ontology merging, jež je například možné při hledání sdílené ontologie použít, se zabývá mnoho výzkumných projektů:

- **Chimaera** [12] - Systém Chimaera poskytuje nástroj pro slučování ontologií. Je založen na ontologickém editoru Ontolingua [9]. Uvažuje pouze hierarchický is-a vztah.
Chimaera je interaktivní nástroj, který vyžaduje interakci uživatele: generuje seznam pojmů (kandidátů pro vztah), což pomáhá uživateli při určování pojmů ke sloučení. Chimaera ponechává rozhodnutí plně na uživateli, sám nenabízí žádné návrhy.
- **PROMPT** [17] - PROMPT je algoritmus pro semiautomatické sloučení ontologií. Provádí některé akce automaticky. Také determinuje možné nekonzistence plynoucí z uživatelských rozhodnutí a nabízí, jak je vyřešit.
PROMPT nejprve vytvoří iniciální seznam pro korespondence založený na pojmech. Následuje cyklus výběru kandidátů uživatelem a automaticky prováděných akcí - algoritmus využívá datové typy, lingvistické techniky a is-a hierarchii.
Algoritmus PROMPT byl implementován jako rozšíření ontologického editoru Protégé-2000 [29].
- **FCA-MERGE** [22] - FCA-MERGE je metoda pro slučování ontologií, která nabízí strukturální popis. Pro zdrojové ontologie extrahuje instance z relevantních textových dokumentů dané domény a aplikuje techniky zpracování přirozeného jazyka. Po extrakci instancí, jsou použity techniky FCA (Formal Concept Analysis) [19] a je získán strukturální výsledek FCA-MERGE. Extrakce instancí a FCA-MERGE algoritmus jsou plně automatické. Vygenerovaný výsledek je transformován do sloučené ontologie se zásahem uživatele.
- **HCONE** [11] - Přístup HCONE využívá WordNet [32], externí informační zdroj. HCONE z WordNetu získává lexikální informace. Lingvistické a strukturální infor-

mace o ontologiích jsou získány pomocí metody LSI - Latent Semantics Indexing [5]. Jednotlivé koncepty jsou asociovány s jejich neformálními, lidsky orientovnými interpretacemi z WordNetu.

Metoda překládá formální definice pojmů do běžného slovníku, které pak vyšetřuje s využitím deskripční logiky. Cílem je ověřit mapování mezi ontologiemi a najít minimální množinu axiomů pro výslednou sloučenou ontologii. Není plně automatický, lidský zásah je nutný v počátečních fázích procesu.

4 Mapování na Sémantickém webu

Výsledek úlohy hledání vzájemných vztahů mezi schématy, tedy nalezené korespondence, se často označuje jako mapování. Obecně může mapování představovat libovolná struktura. K vyjádření mapování lze použít od jednoduchých 1-1 mapovacích pravidel vyjadřujících přímou korespondenci mezi elementy, přes mapování konceptu na dotaz nebo pohled [4], až po pomocné mapovací struktury (například referenční model v [24]). Různé projekty obvykle používají vlastní pojetí mapování.

Kromě například používání mapovacích pravidel jako tvrzení pro elementy globálních a lokálních schémat, které jsou orientovány na konkrétní řešenou úlohu, je možné využít složitější a dokonce standardizovanou strukturu, jenž by pokrývala všechna mapování. K popisu mapování mezi elementy schématu globálního pohledu a schémat lokálních zdrojů bude sloužit *ontologie OWL*.

Užití ontologie pro mapování přináší možnost znovupoužití také v jiných úlohách či situacích. Je také možné při odvozování dalších korespondencí, například při integrování dalšího zdroje, využít mapování v ontologii jako další ontologii, která integrované zdroje popisuje. Tak je možné dále využívat již jednou zjištěné skutečnosti. Navíc, bude-li v budoucnu třeba zachytit i další typy vztahů mezi elementy, může být ontologie dále využita, neboť je schopna zachytit různé typy vztahů.

K popisu mapování bude v závislosti na typu vztahu využít odpovídající [28] konstrukt. Abstraktním mechanismem pro seskupování popisovaných zdrojů v OWL je třída (class). Zdrojem je na webu jakákoli identifikovatelná entita. Proto bude pojetí `owl:Class` použito pro korespondenci elementů:

- **Is-a** hierarchický vztah, tj. $\varepsilon_1 \subseteq \varepsilon_2$, lze vyjádřit pomocí podtříd. Příslušným rysem OWL je `rdfs:subClassOf`, který umožňuje vyjádřit, že extenze popisu jedné třídy je podmnožinou extenze popisu jiné třídy.
- Vztah **ekvivalence**, tj. $\varepsilon_1 = \varepsilon_2$, lze v OWL vyjádřit s `owl:equivalentClass`. `owl:equivalentClass` umožňuje vyjádřit, že dvě třídy mají stejnou extenzi. V tomto případě může být také použit `rdfs:subClassOf` tak, že definujeme ε_1 jako podtřídou třídy ε_2 a současně ε_2 jako podtřídou třídy ε_1 , říkáme, že ε_1 a ε_2 jsou ekvivalentní třídy.
- **Disjunktnost** (neboli tvrzení, že extenze popisu jedné třídy nemá žádné společné prvky s extenzí popisu jiné třídy) lze vyjádřit pomocí `owl:disjointWith`.

K vyjádření mapování slouží ontologie OWL. Zdrojem, ze kterého je mapování získáváno je ontologie sdílená zdroji, také ontologie OWL. Daná sdílená ontologie je “nadontologií” hledané ontologie v tom významu, že popisuje všechny třídy a jejich vztahy obsažené v mapování.

5 Shrnutí a závěr

Hledání korespondencí mezi schémata (schema matching) je stěžejní částí integračního procesu. Jeho výsledkem je mapování, které je dále využíváno při zpracovávání integrovaných dat. Při hledání vztahů je možné využít různých technik založených na různých informacích o datech. Jsou-li dostupné ontologie zdrojů, je možné odvodit hledané korespondence také z nich.

Důležitou otázkou je také způsob, jak nalezené mapování zaznamenat. V popsaném přístupu je k tomuto využita ontologie OWL. To přináší možnost mapování sdílet či znovu používat. Navíc mapování, které je vyjádřeno pomocí standardizované struktury může být dále využíváno i v jiných situacích a lze jej zpracovávat různými nástroji. Pro toto mapování je například možné používat metody vyvinuté pro zpracovávání ontologií.

Je-li k dispozici jediná ontologie, která popisuje data v integrovaných datových zdrojích, lze mapování v podstatě snadno získat přímo z této ontologie. V obecném případě, kdy je pro popis dat využito více ontologií, jsou tyto ontologie integrovány. Výsledkem integrace ontologií je sdílená ontologie a úloha je převedena na předchozí případ. Tímto způsobem je úloha hledání korespondencí mezi schémata převedena na úlohu slučování ontologií, pro kterou je možné využít některou z dostupných metod.

Mapování schémat založené na ontologiích je podúlohou celého procesu integrace. V budoucnu je proto plánováno zaměřit se také na následující fázi, tj. využití mapování pro zpracování dotazů.

Literatura

- [1] Z. Bellahsene, “Data integration over the Web”, *Data & Knowledge Engineering*, 44 (2003), pp. 265-266.
- [2] T. Berners-Lee, J. Hendler and O. Lassila, “The Semantic Web”, *Scientific American*, vol. 284, 5, pp. 35-43, 2001.
- [3] A. Cali, D. Calvanese, G. De Giacomo, and M. Lenzerini, “On the Expressive Power of Data Integration Systems”, In *Proceedings of the 21st Int. Conf. On Conceptual Modeling (ER 2002)*, LNCS 2503, Springer, pp. 338-350, 2002.
- [4] D. Calvanese, G. De Giacomo, and M. Lenzerini, “Ontology of integration and integration of ontologies”, In *Proceedings of the 2001 Description Logic Workshop (DL 2001)*, 2001.
- [5] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman, “Indexing by Latent Semantic Analysis”, *Journal of the American Society of Information Science* 41(6), pp. 391-407, 1990.

-
- [6] Y. Ding, D. Fensel, M. Klein, and B. Omelayenko, “The semantic web: yet another hip?”, *Data & Knowledge Engineering*, 41 (2002), pp. 205-227.
- [7] H.-H. Doa and E. Rahmb, “Matching large schemas: Approaches and evaluation”, *Information Systems*, (in print), 2007.
- [8] J. Euzenat, “Research challenges and perspectives of the Semantic Web”. Report of the EU-NSF Strategic Research Workshop, Sophia-Antipolis, France, October, 2001.
- [9] A. Farquhar, R. Fikes, and J. Rice, “The Ontolingua Server: a Tool for Collaborative Ontology Construction”, Technical report, Stanford KSL 96-26, 1996.
- [10] Y. Kalfoglou and M. Schorlemmer, “Ontology mapping: the state of the art”, *The Knowledge Engineering Review* 18(1), pp. 1-31, 2003.
- [11] K. Kotis and G. A. Vouros, “The HCONE Approach to Ontology Merging”, In *ESWS*, LNCS 3053, Springer, pp. 137-151, 2004.
- [12] D. L. McGuinness, R. Fikes, J. Rice, and S. Wilder, “An Environment for Merging and Testing Large Ontologies”, In Proceedings of the *Seventh International Conference*, 2000.
- [13] M.-R. Koivunen and E. Miller, “W3C Semantic Web Activity”, in the proceedings of the *Semantic Web Kick/off Seminar*, Finland, 2001.
- [14] M. Lenzerini, “Data Integration: A Theoretical Perspective”, In Proceedings of the *21st ACM SIGMOD - SIGACT - SIGART symposium on Principles of database systems*, ACM Press, pp. 233-246, 2002.
- [15] P. Mitra, G. Wiederhold, and J. Jannink, “Semi-automatic integration of knowledge sources”, In Proceeding of the *2nd Int. Conf. On Information FUSION'99*, 1999.
- [16] H. Nottelmann and U. Straccia, “Information retrieval and machine learning for probabilistic schema matching”, *Inf. Process. Manage.*, 43(3), pp. 552-576, 2007.
- [17] N. F. Noy and M. A. Musen, “PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment”, In *AAAI/IAAI*, pp. 450-455, 2000.
- [18] R. Pottinger and A. Levy, “A Scalable Algorithm for Answering Queries Using Views”, In the Proceedings of the *26th VLDB Conference*, Cairo, Egypt (2000).
- [19] U. Priss, “Formal Concept Analysis in Information Science (draft)”, <http://www.upriss.org.uk/papers/arist.pdf>.
- [20] E. Rahm and P. A. Bernstein, “A survey of approaches to automatic schema matching”, In *VLDB Journal: Very Large Data Bases*, 10(4), pp. 334-350, 2001.
- [21] P. Shvaiko and J. Euzenat, “A survey of schema-based matching approaches”, 3730, pp. 146-171, 2005.

- [22] G. Stumme and A. Maedche “FCA-MERGE: Bottom-Up Merging of Ontologies”, In *IJCAI*, pp. 225-234, 2001.
- [23] X. Su and J. A. Gulla, “An information retrieval approach to ontology mapping”, *Data & Knowledge Engineering* 58(1), pp. 47-69, 2006.
- [24] H. T. Uitermark, P. J. M. van Oosterom, N. J. I. Mars, and M. Molenaar, “Ontology-based integration of topographic data sets”, *International Journal of Applied Earth Observation and Geoinformation* 7 (2005), pp. 97-106.
- [25] J. D. Ullman, “Information integration using logical views”, *Theoretical Computer Science* 239 (2000), pp. 189-210.
- [26] L. Xu and D. W. Embley, “A composite approach to automating direct and indirect schema mappings”, *Inf. Syst.*, 31(8), pp. 697-732, 2006.
- [27] S. Yi, B. Huang, and W. T. Chan, “Xml application schema matching using similarity measure and relaxation labeling”, *Inf. Sci.*, 169(1-2), pp. 27-46, 2005.
- [28] Web Ontology Language (OWL),
<http://www.w3.org/2004/OWL>.
- [29] The Protégé Ontology Editor and Knowledge Acquisition System,
<http://protege.stanford.edu/>.
- [30] Resource Description Framework (RDF),
<http://www.w3.org/RDF/>.
- [31] RDF Vocabulary Description Language 1.0: RDF Schema, *W3C Recommendation*,
<http://www.w3.org/TR/2004/REC-rdf-schema-20040210>.
- [32] WordNet, a lexical database for the English language,
<http://wordnet.princeton.edu/>.
- [33] Extensible Markup Language (XML),
<http://www.w3.org/XML/>.