Detecting and Correcting Conservativity Principle Violations in Ontology-to-Ontology Mappings

Alessandro Solimando[1], Ernesto Jiménez-Ruiz[2], Giovanna Guerrini[1]

[1] DIBRIS, University of Genova, Italy
[2] Department of Computer Science, University of Oxford, UK

October 23th 2014, Riva del Garda, ISWC 2014

Acknowledgements: student grant by Semantic Web Science Association

Context

General Context

- Ontologies are key in the development of the Semantic Web
- An application domain can be modeled with different points of view and purposes
- Therefore often many ontologies with different naming and modeling conventions
- Matching these ontologies to enable interoperability between ontology-based information systems

Ontology Matching (OM)

- The problem of (semi-)automatically computing mappings between independent, but semantically overlapping, ontologies
- OM systems rely on lexical/structural heuristics, the integration of the input ontologies and the mappings may lead to undesired logical consequences

Motivations

OM Problems

- Ontology alignments may suffer from logical problems (*i.e.*, violations)
- These violations may hinder their usefulness in practical use (*e.g.*, query answering)
- Three principles in literature:
  - consistency principle (no new incoherencies in the aligned ontology)
  - conservativity principle (no new entailments in the aligned ontology for classes of the same input ontology),
  - locality principle (mappings should link entities with similar *neighborhood*)

Conservativity Principle

Deductive Difference

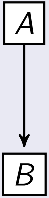
- Conservativity principle tightly linked with the notion of conservative extension/deductive difference
- Deductive difference between $\mathcal{O}$ and $\mathcal{O}'$ w.r.t. a signature Σ is the set of entailments constructed over Σ that do not hold in $\mathcal{O}$', but do hold in $\mathcal{O}$
- Its computation is in *NEXPTIME-Hard* for acyclic $\mathcal{ALC}$ terminologies, tractable algorithm only for $\mathcal{EL}$ family [Konev et al., 2008]

(Definition) Conservativity Principle Violation

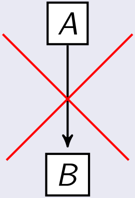
Let A, B be atomic concepts from $\Sigma \cup \{\top, \bot\}$, $\mathcal{O}$ and $\mathcal{O}'$ be two OWL 2 ontologies, Σ a signature. A conservativity violation is any axiom of the form $A \sqsubseteq B$ satisfying: $\mathcal{O} \not\models A \sqsubseteq B$, and $\mathcal{O}' \models A \sqsubseteq B$

Conservativity Principle Violations

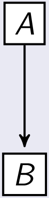
Conservativity Principle Violation ($diff^{\approx}$)



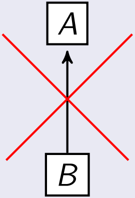
such that



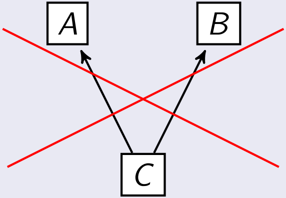
Conservativity Principle Violation Approximation ($consViol$)



such that



and



Key Ideas of our Approach

- Approximation of the deductive difference, targeting named classes [Jiménez-Ruiz et al., 2009]
- Scalability:
  - use of modularization techniques
  - projection of the ontologies to Horn Propositional fragment
  - reuse of semantic indexing techniques of LogMap matcher (based on interval indexing) [Jiménez-Ruiz and Cuenca Grau, 2011]
- basic violations: partly reduced to consistency problem by adding enough disjoint axioms, following the *assumption of disjointness* [Schlobach, 2005]

Application Scenario

Optique

FP7 Project Optique (www.optique-project.eu)

- Scalable Ontology-based Access to BigData in gas and oil industry (Siemens AG and Statoil)
- Ontology $\mathcal{O}_1$ bootstrapped from the relational database containing the data and directly linked to it using ontology-to-schema mappings
- Visual query formulation interface using ontology $\mathcal{O}_2$ (Norwegian Petroleum Directorate ontology)
- Ontology-to-Ontology mappings between $\mathcal{O}_1$ and $\mathcal{O}_2$
- None of the two ontologies can be modified
- Query answering therefore depending on both kinds of mappings

Algorithm Sketch

1. Classification step (input and aligned ontologies)
2. Interval indexing to encode subsumption/disjointness relations
3. Conservativity principle violations detection
4. Projection of the ontologies to Horn Propositional fragment
5. Disjointness clauses addition to projections, two variants:
    - basic: add a disjointness clause for each pair of classes not in a subsumption relationship
    - optimized: add a disjointness clause for each basic violation
6. Repair based on Dowling and Gallier algorithm (over the projections)

Example (input ontologies and mappings)

	Ontology $\mathcal{O}_1$		Ontology $\mathcal{O}_2$
α_1	WellBore $\sqsubseteq$ $\exists$belongsTo.Well	β_1	Exploration_well $\sqsubseteq$ Well
α_2	WellBore $\sqsubseteq$ $\exists$hasOperator.Operator	β_2	Explor_borehole $\sqsubseteq$ Borehole
α_3	WellBore $\sqsubseteq$ $\exists$locatedIn.Field	β_3	Appraisal_exp_borehole $\sqsubseteq$ Explor_borehole
α_4	AppraisalWellBore $\sqsubseteq$ WellBore	β_4	Appraisal_well $\sqsubseteq$ Well
α_5	ExplorationWellBore $\sqsubseteq$ WellBore	β_5	Field $\sqsubseteq$ $\exists$hasFieldOperator.Field_operator
α_6	Operator $\sqsubseteq$ Owner	β_6	Field_operator $\sqcap$ Owner $\sqsubseteq$ Field_owner
α_7	Operator $\sqsubseteq$ Company	β_7	Company $\sqsubseteq$ Field_operator
α_8	Field $\sqsubseteq$ $\exists$hasOperator.Company	β_8	Field_owner $\sqsubseteq$ Owner
α_9	Field $\sqsubseteq$ $\exists$hasOwner.Owner	β_9	Borehole $\sqsubseteq$ Continuant $\sqcup$ Occurrent

Mappings $\mathcal{M}$

id	e_1	e_2	n	ρ
m_1	$\mathcal{O}_1$:Well	$\mathcal{O}_2$:Well	0.9	$\equiv$
m_2	$\mathcal{O}_1$:WellBore	$\mathcal{O}_2$:Borehole	0.7	$\equiv$
m_3	$\mathcal{O}_1$:ExplorationWellBore	$\mathcal{O}_2$:Exploration_well	0.6	$\sqsubseteq$
m_4	$\mathcal{O}_1$:ExplorationWellBore	$\mathcal{O}_2$:Explor_borehole	0.8	$\equiv$
m_5	$\mathcal{O}_1$:AppraisalWellBore	$\mathcal{O}_2$:Appraisal_exp_borehole	0.7	$\equiv$
m_6	$\mathcal{O}_1$:Field	$\mathcal{O}_2$:Field	0.9	$\equiv$
m_7	$\mathcal{O}_1$:Operator	$\mathcal{O}_2$:Field_operator	0.7	$\sqsupseteq$
m_8	$\mathcal{O}_1$:Company	$\mathcal{O}_2$:Company	0.9	$\equiv$
m_9	$\mathcal{O}_1$:hasOperator	$\mathcal{O}_2$:hasFieldOperator	0.6	$\equiv$
m_{10}	$\mathcal{O}_1$:Owner	$\mathcal{O}_2$:Owner	0.9	$\equiv$

Example (conservativity violations)

Conservativity violations affecting the aligned ontology:

σ	**Entailment:**	**follows from:**	**Violation?**
σ_1	$\mathcal{O}_2$:Explor_borehole $\sqsubseteq$ $\mathcal{O}_2$:Exploration_well	m_3, m_4	YES
σ_2	$\mathcal{O}_1$:AppraisalWellBore $\sqsubseteq$ $\mathcal{O}_1$:ExplorationWellBore	β_3, m_4, m_5	YES
σ_3	$\mathcal{O}_2$:Field_operator $\sqsubseteq$ $\mathcal{O}_2$:Field_owner	$\alpha_6, \beta_6, m_7, m_{10}$	YES
σ_4 σ_5	$\mathcal{O}_1$:Company $\equiv$ $\mathcal{O}_1$:Operator $\mathcal{O}_2$:Field_operator $\equiv$ $\mathcal{O}_2$:Company	$\alpha_7, \beta_7, m_7, m_8$	NO (*)
σ_6	$\mathcal{O}_1$:Company $\sqsubseteq$ $\mathcal{O}_1$:Owner	σ_4, α_6	YES
σ_7	$\mathcal{O}_2$:Company $\sqsubseteq$ $\mathcal{O}_2$:Field_owner	σ_3, σ_5	YES

Example (basic violations)

σ	**Entailment:**	**follows from:**	**Violation?**
σ_1	$\mathcal{O}_2$:Explor_borehole $\sqsubseteq$ $\mathcal{O}_2$:Exploration_well	m_3, m_4	YES
σ_2	$\mathcal{O}_1$:AppraisalWellBore $\sqsubseteq$ $\mathcal{O}_1$:ExplorationWellBore	β_3, m_4, m_5	YES
σ_3	$\mathcal{O}_2$:Field_operator $\sqsubseteq$ $\mathcal{O}_2$:Field_owner	$\alpha_6, \beta_6, m_7, m_{10}$	YES
σ_4 σ_5	$\mathcal{O}_1$:Company $\equiv$ $\mathcal{O}_1$:Operator $\mathcal{O}_2$:Field_operator $\equiv$ $\mathcal{O}_2$:Company	$\alpha_7, \beta_7, m_7, m_8$	NO
σ_6	$\mathcal{O}_1$:Company $\sqsubseteq$ $\mathcal{O}_1$:Owner	σ_4, α_6	YES
σ_7	$\mathcal{O}_2$:Company $\sqsubseteq$ $\mathcal{O}_2$:Field_owner	σ_3, σ_5	YES

$$m_3 = \langle \mathcal{O}_1\text{:ExplorationWellBore}, \mathcal{O}_2\text{:Exploration\_well}, 0.6, \sqsubseteq \rangle$$
$$m_4 = \langle \mathcal{O}_1\text{:ExplorationWellBore}, \mathcal{O}_2\text{:Explor\_borehole}, 0.8, \equiv \rangle$$

Conservativity Violations in a Query Answering Task

- Conjunctive query $\mathrm{CQ(x)} \leftarrow \mathcal{O}_2$:Well(x) asks for wells
- Query formulated using the Optique's query formulation interface (using the vocabulary of $\mathcal{O}_2$)
- The query is rewritten, according to the ontology axioms and mappings $\beta_1, \beta_4, m_1, m_3, m_4$ in $\mathcal{O}_\mathcal{U} = \mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{M}$, into the following union of conjunctive queries
  $\mathrm{UCQ(x)} \leftarrow \mathcal{O}_2$:Well(x) $\cup$ $\mathcal{O}_1$:Well(x) $\cup$
  $\mathcal{O}_2$:Exploration_well(x) $\cup$ $\mathcal{O}_2$:Appraisal_well(x) $\cup$
  $\mathcal{O}_1$:ExplorationWellBore(x) $\cup$ $\mathcal{O}_2$:Explor_borehole(x).
- Since only $\mathcal{O}_1$ is linked to the data, the query is simplified as
  $\mathrm{UCQ(x)} \leftarrow$ Well(x) $\cup$ ExplorationWellBore(x)
- Original query asking for wells, while the rewritten query return also data about exploration wellbores

Example (basic violations repair)

σ	Entailment:	follows from:	Violation?
σ_1	$\mathcal{O}_2$:Explor_borehole $\sqsubseteq$ $\mathcal{O}_2$:Exploration_well	m_3, m_4	YES
σ_2	$\mathcal{O}_1$:AppraisalWellBore $\sqsubseteq$ $\mathcal{O}_1$:ExplorationWellBore	β_3, m_4, m_5	YES
σ_3	$\mathcal{O}_2$:Field_operator $\sqsubseteq$ $\mathcal{O}_2$:Field_owner	$\alpha_6, \beta_6, m_7, m_{10}$	YES
σ_4 σ_5	$\mathcal{O}_1$:Company $\equiv$ $\mathcal{O}_1$:Operator $\mathcal{O}_2$:Field_operator $\equiv$ $\mathcal{O}_2$:Company	$\alpha_7, \beta_7, m_7, m_8$	NO
σ_6	$\mathcal{O}_1$:Company $\sqsubseteq$ $\mathcal{O}_1$:Owner	σ_4, α_6	YES
σ_7	$\mathcal{O}_2$:Company $\sqsubseteq$ $\mathcal{O}_2$:Field_owner	σ_3, σ_5	YES

- $\mathcal{O}_1, \mathcal{O}_2 \xrightarrow[\textit{horn enconding}]{} \mathcal{P}_1, \mathcal{P}_2 \xrightarrow[\textit{disjointness addition}]{} \mathcal{P}_1^d, \mathcal{P}_2^d$
- $\mathcal{P}_1^d \cup \mathcal{P}_2^d \cup \mathcal{M} \cup \{true \rightarrow \mathcal{O}_1\text{:ExplorationWellbore}\}$ = UNSAT
- Justification:
  - $m_3 = \langle \mathcal{O}_1\text{:ExplorationWellBore}, \mathcal{O}_2\text{:Exploration\_well}, 0.6, \sqsubseteq \rangle$
  - $m_4 = \langle \mathcal{O}_1\text{:ExplorationWellBore}, \mathcal{O}_2\text{:Explor\_borehole}, 0.8, \equiv \rangle$
  - $\psi_1 = \mathcal{O}_2\text{:Explor\_borehole} \wedge \mathcal{O}_2\text{:Exploration\_well} \rightarrow \textit{false}$
- between involved mappings (m_3 and m_4) the one with lower confidence is discarded (m_3)

Example (equivalence violations)

σ	**Entailment:**	**follows from:**	**Violation?**
σ_1	$\mathcal{O}_2$:Explor_borehole $\sqsubseteq$ $\mathcal{O}_2$:Exploration_well	m_3, m_4	YES
σ_2	$\mathcal{O}_1$:AppraisalWellBore $\sqsubseteq$ $\mathcal{O}_1$:ExplorationWellBore	β_3, m_4, m_5	YES
σ_3	$\mathcal{O}_2$:Field_operator $\sqsubseteq$ $\mathcal{O}_2$:Field_owner	$\alpha_6, \beta_6, m_7, m_{10}$	YES
σ_4 σ_5	$\mathcal{O}_1$:Company $\equiv$ $\mathcal{O}_1$:Operator $\mathcal{O}_2$:Field_operator $\equiv$ $\mathcal{O}_2$:Company	$\alpha_7, \beta_7, m_7, m_8$	NO
σ_6	$\mathcal{O}_1$:Company $\sqsubseteq$ $\mathcal{O}_1$:Owner	σ_4, α_6	YES
σ_7	$\mathcal{O}_2$:Company $\sqsubseteq$ $\mathcal{O}_2$:Field_owner	σ_3, σ_5	YES

$$\alpha_7 = \mathcal{O}_1\text{:Operator} \sqsubseteq \mathcal{O}_1\text{:Company}$$
$$\beta_7 = \mathcal{O}_2\text{:Company} \sqsubseteq \mathcal{O}_2\text{:Field\_operator}$$
$$m_7 = \langle \mathcal{O}_1\text{:Operator}, \mathcal{O}_2\text{:Field\_operator}, 0.7, \sqsupseteq \rangle$$
$$m_8 = \langle \mathcal{O}_1\text{:Company}, \mathcal{O}_2\text{:Company}, 0.9, \equiv \rangle$$

- σ_4 and σ_5 are not included in our variant of conservativity violation, due to axiom α_7 and β_7, respectively
- however, they lead to violations σ_6 and σ_7

Experiments

- Violation rate: high for alignments computed by the top-level ontology matchers, but also reference alignments, such as that of *Ontology Alignment Evaluation Initiative* (OAEI) dataset and UMLS-Metathesaurus [Bodenreider, 2004] (UMLS), manually curated alignments for integrating biomedical knowledge bases (*e.g.*, SNOMED-NCI $>$ 500K violations)
- Evaluation using the reference alignments of OAEI 2013:
  - Efficiency: 256 seconds to detect violations in aligned ontology SNOMED-NCI (the biggest OAEI's test case)
  - Effectivity: repair algorithm solved $>$ 99% violations, effective in practice even if repair algorithm is approximated
- Total repair time ranges from an average of 10 ms for the smallest testcase to an average of 22 min for the biggest one ($>$ 75K classes)
- Optimized disjointness addition is faster and does not reduce the rate of solved violations

Related Work

- Assumption of disjointness for debugging unsatisfiable ontologies with few disjointness constraints [Schlobach, 2005]
- Several algorithms for enriching ontologies with negative constraints [Völker et al., 2007, Meilicke et al., 2008, Fleischhacker and Völker, 2011, Ferré and Rudolph, 2012]
- Conservativity principle use case study on biomedical ontologies [Jiménez-Ruiz et al., 2011, Beisswanger et al., 2012]
- Enriching incomplete input ontologies [Dragisic et al., 2014, Ivanova and Lambrix, 2013, Lambrix and Liu, 2013, Lambrix et al., 2013a, Lambrix et al., 2013b]
- In OM only patterns and heuristics to limit the violations, lack of theoretical foundation [Jean-Mary et al., 2009, Wang and Xu, 2012, Ngo and Bellahsene, 2012]
- Proposals never analyzed/compared for conservativity violations, only precision/recall w.r.t. a (often not violations-free) reference alignment

Conclusions and Future Work

- Completely automatic detection/repair algorithms, following a "better safe than sorry approach"
- Practical efficiency and effectivity confirmed by conducted experiments, $> 99\%$ of the violations are removed
- Repair can be aggressive (from 16% to 48% of the mappings)
- Given the presence of false positives (violations caused by incompleteness of the input ontologies), a suitable GUI seems the missing ingredient
- Repair algorithms are very precise (can selectively repair a subset of the violations), but visualizing them for user labeling is really challenging (in some cases > 30 millions violations), and information/context to be visualized to help the user is not clear yet
- Short-term plans for deployment in the Optique industry partners Statoil and Siemens (techniques already integrated within the "ontology and mapping management module" ([Kharlamov et al., 2013])

Thanks for your attention!

References I

Beisswanger, E., Hahn, U., et al. (2012).
Towards valid and reusable reference alignments—ten basic quality checks for ontology alignments and their application to three different reference data sets.
J. Biomed. Semant., 3(Suppl 1):S4.

Bodenreider, O. (2004).
The Unified Medical Language System (UMLS): integrating biomedical terminology.
Nucleic Acids Research, 32:267–270.

Dragisic, Z., Lambrix, P., and Wei-Kleiner, F. (2014).
A system for debugging missing is-a structure in el ontologies.
In *Int'l Workshop on Debugging Ontologies and Ontology Mappings (WoDOOM)*.

References II

Ferré, S. and Rudolph, S. (2012).
Advocatus Diaboli - Exploratory Enrichment of Ontologies with Negative Constraints.
In *Int'l Conf. on Knowl. Eng. (EKAW)*, pages 42–56.

Fleischhacker, D. and Völker, J. (2011).
Inductive Learning of Disjointness Axioms.
In *OTM Conferences*, pages 680–697.

Ivanova, V. and Lambrix, P. (2013).
A Unified Approach for Aligning Taxonomies and Debugging Taxonomies and their Alignments.
In *Eur. Sem. Web Conf. (ESWC)*, pages 1–15. Springer.

References III

Jean-Mary, Y. R., Shironoshita, E. P., and Kabuka, M. R. (2009).
Ontology Matching With Semantic Verification.
J. Web Sem., 7(3):235–251.

Jiménez-Ruiz, E. and Cuenca Grau, B. (2011).
LogMap: Logic-based and Scalable Ontology Matching.
In *Int'l Sem. Web Conf. (ISWC)*, pages 273–288.

Jiménez-Ruiz, E., Cuenca Grau, B., Horrocks, I., and Berlanga, R. (2009).
Ontology integration using mappings: Towards getting the right logical consequences.
In *Eur. Sem. Web Conf.*

References IV

Jiménez-Ruiz, E., Cuenca Grau, B., Horrocks, I., and Berlanga, R. (2011).
Logic-based Assessment of the Compatibility of UMLS Ontology Sources.
J. Biomed. Semant., 2(Suppl 1):S2.

Jiménez-Ruiz, E., Meilicke, C., Grau, B. C., and Horrocks, I. (2013).
Evaluating Mapping Repair Systems with Large Biomedical Ontologies.
In *Description Logics*, pages 246–257.

Kharlamov, E., Jiménez-Ruiz, E., Zheleznyakov, D., et al. (2013).
Optique: Towards OBDA Systems for Industry.
In *Eur. Sem. Web Conf. (ESWC) Satellite Events*, pages 125–140.

Konev, B., Walther, D., and Wolter, F. (2008).
The Logical Difference Problem for Description Logic Terminologies.
In *Int'l Joint Conf. on Automated Reasoning (IJCAR)*, pages 259–274.

Lambrix, P., Dragisic, Z., and Ivanova, V. (2013a).
Get My Pizza Right: Repairing Missing Is-a Relations in $\mathcal{ALC}$ Ontologies.
In *Semantic Technology*, pages 17–32. Springer.

Lambrix, P. and Liu, Q. (2013).
Debugging the Missing Is-a Structure Within Taxonomies Networked by Partial Reference Alignments.
Data Knowl. Eng. (DKE), 86:179–205.

References VI

Lambrix, P., Wei-Kleiner, F., Dragisic, Z., and Ivanova, V. (2013b).
Repairing Missing Is-a Structure in Ontologies is an Abductive Reasoning Problem.
In *Int'l Workshop on Debugging Ontologies and Ontology Mappings (WoDOOM)*, page 33.

Meilicke, C., Völker, J., and Stuckenschmidt, H. (2008).
Learning Disjointness for Debugging Mappings between Lightweight Ontologies.
In *Int'l Conf. on Knowl. Eng. (EKAW)*, pages 93–108.

Ngo, D. and Bellahsene, Z. (2012).
YAM++ : A Multi-strategy Based Approach for Ontology Matching Task.
In *Int'l Conf. on Knowl. Eng. (EKAW)*, pages 421–425.

References VII

Schlobach, S. (2005).
Debugging and Semantic Clarification by Pinpointing.
In *Eur. Sem. Web Conf. (ESWC)*, pages 226–240. Springer.

Völker, J., Vrandecic, D., Sure, Y., and Hotho, A. (2007).
Learning Disjointness.
In *Eur. Sem. Web Conf. (ESWC)*, pages 175–189.

Wang, P. and Xu, B. (2012).
Debugging Ontology Mappings: A Static Approach.
Computing and Informatics, 27(1):21–36.

Algorithm 1 Conducted evaluation over the Optique and OAEI data sets

Input: $\mathcal{O}_1$, $\mathcal{O}_2$: input ontologies $\mathcal{M}$: reference mappings for $\mathcal{O}_1$ and $\mathcal{O}_2$

1: $\mathcal{O}_\mathcal{U} := \mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{M}$

2: Store size of $\mathsf{Sig}(\mathcal{O}_1)$ (**I**), $\mathsf{Sig}(\mathcal{O}_2)$ (**II**) and $\mathcal{M}$ (**III**)

3: Compute number of conservativity principle violations: $\mathsf{consViol} := |\mathsf{consViol}(\mathcal{O}_1, \mathcal{O}_\mathcal{U})| + |\mathsf{consViol}(\mathcal{O}_2, \mathcal{O}_\mathcal{U})|$ (**IV**)

4: Compute number of conservativity principle violations: $\mathsf{diff}^{\approx} := |\mathsf{diff}^{\approx}_{\mathsf{Sig}(\mathcal{O}_1)}(\mathcal{O}_1, \mathcal{O}_\mathcal{U})| + |\mathsf{diff}^{\approx}_{\mathsf{Sig}(\mathcal{O}_2)}(\mathcal{O}_2, \mathcal{O}_\mathcal{U})|$ (**V**)

5: Compute two repairs $\mathcal{R}^{\approx}$ using Algorithm 2 for $\mathcal{O}_1$, $\mathcal{O}_2$, $\mathcal{M}$, with the *Optimization* set to false and true

6: Store number of added disjointness *disj* (**VI** and **XII**), size of repair $|\mathcal{R}^{\approx}|$ (**VII** and **XIII**), time to compute disjointness rules t_d (**VIII** and **XIV**), and time to compute the mapping repair t_r (**IX** and **XV**)

7: $\mathcal{O}_\mathcal{U} := \mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{M} \setminus \mathcal{R}^{\approx}$

8: Compute number of remaining conservativity principle violations (our variant): $\mathsf{consViol} := |\mathsf{consViol}(\mathcal{O}_1, \mathcal{O}_\mathcal{U})| + |\mathsf{consViol}(\mathcal{O}_2, \mathcal{O}_\mathcal{U})|$ (**X** and **XVI**)

9: Compute number of remaining conservativity principle violations (general notion): $\mathsf{diff}^{\approx} := |\mathsf{diff}^{\approx}_{\mathsf{Sig}(\mathcal{O}_1)}(\mathcal{O}_1, \mathcal{O}_\mathcal{U})| + |\mathsf{diff}^{\approx}_{\mathsf{Sig}(\mathcal{O}_2)}(\mathcal{O}_2, \mathcal{O}_\mathcal{U})|$ (**XI** and **XVII**)

Original Problem Metrics

Dataset	$\mathcal{O}_1 \sim \mathcal{O}_2$	Problem size			Original Violations	
		I	II	III	IV	V
		$\lvert\text{Sig}(\mathcal{O}_1)\rvert$	$\lvert\text{Sig}(\mathcal{O}_2)\rvert$	$\lvert\mathcal{M}\rvert$	consViol	$\text{diff}^{\approx}$
Optique	NPD$\sim$BootsOnto	757	40,671	102	214	220
LargeBio	SNOMED$\sim$NCI	122,519	66,914	36,405	>525,515	>546,181
	FMA$\sim$SNOMED	79,042	122,519	17,212	125,232	127,668
	FMA$\sim$NCI	79,042	66,914	5,821	19,740	19,799
Anatomy	MO$\sim$NCI$_{\text{Anat}}$	2,747	3,306	3,032	1,321	1,335
Library	STW$\sim$TheSoz	6,575	8,376	6,322	42,045	42,872
Conference	cmt$\sim$confof	89	75	32	11	11
	conference$\sim$edas	124	154	34	8	8
	conference$\sim$iasted	124	182	28	9	9
	confof$\sim$ekaw	75	107	40	6	6
	edas$\sim$iasted	154	182	38	7	7

Basic Method Results

Dataset	$\mathcal{O}_1 \sim \mathcal{O}_2$	Solution size		Times		Remaining Violations	
		VI	VII	VIII	IX	X	XI
		#disj	$\lvert\mathcal{R}^{\approx}\rvert$	$t_d(s)$	$t_r(s)$	consViol	diff$^{\approx}$
Optique	NPD∼BootsOnto	4,716,685	49	9,840	121	0	0
LargeBio	SNOMED∼NCI	–	–	–	–	–	–
	FMA∼SNOMED	1,106,259	8,234	35,817	1,127	0	121
	FMA∼NCI	347,801	2,176	2,471	38	103	112
Anatomy	$\text{MO}\sim\text{NCI}_{\text{Anat}}$	1,331,374	461	397	56	0	3
Library	STW∼TheSoz	591,115	2,969	4,126	416	0	24
Conference	cmt∼confof	50	6	0.01	0.01	0	0
	conference∼edas	774	6	0.03	0.01	0	0
	conference∼iasted	2,189	4	0.06	0.02	0	0
	confof∼ekaw	296	6	0.02	0.01	0	0
	edas∼iasted	1,210	4	0.06	0.02	1	1

Optimized Method Results

Dataset	$\mathcal{O}_1 \sim \mathcal{O}_2$	Solution size		Times		Remaining Violations	
		XII	XIII	XIV	XV	XVI	XVII
		#disj	$\lvert\mathcal{R}^{\approx}\rvert$	t_d(s)	t_r(s)	consViol	diff$^{\approx}$
Optique	NPD∼BootsOnto	214	41	2.54	0.17	0	0
LargeBio	SNOMED∼NCI	525,515	15,957	275	3,755	>411	>1,624
	FMA∼SNOMED	125,232	8,342	30	251	0	131
	FMA∼NCI	19,740	2,175	34	6.18	103	112
Anatomy	$\text{MO}\sim\text{NCI}_{\text{Anat}}$	1,321	491	1.39	0.53	0	3
Library	STW∼TheSoz	42,045	3,058	4.93	41	0	40
Conference	cmt∼confof	11	6	0.05	0.01	0	0
	conference∼edas	8	6	0.07	0.01	0	0
	conference∼iasted	9	1	0.22	0.01	0	0
	confof∼ekaw	6	5	0.04	0.01	0	0
	edas∼iasted	7	4	0.21	0.02	1	1

Algorithm 2 Algorithm to detect and solve conservativity principle violations

Input: $\mathcal{O}_1$, $\mathcal{O}_2$: input ontologies; $\mathcal{M}$: (coherent) input mappings; *Optimization*: Boolean value

Output: $\mathcal{M}'$: output mappings; $\mathcal{R}^{\approx}$: approximate repair; *disj*: number of disjointness rules

$\langle \mathcal{O}'_1, \mathcal{O}'_2 \rangle :=$ ModuleExtractor$(\mathcal{O}_1, \mathcal{O}_2, \mathcal{M})$
$\langle \mathcal{P}_1, \mathcal{P}_2 \rangle :=$ PropositionalEncoding$(\mathcal{O}'_1, \mathcal{O}'_2)$
$SI_1 :=$ StructuralIndex$(\mathcal{O}'_1)$
$SI_2 :=$ StructuralIndex$(\mathcal{O}'_2)$
if (*Optimization* = true) **then**
    $SI_{\mathcal{U}} :=$ StructuralIndex$(\mathcal{O}'_1 \cup \mathcal{O}'_2 \cup \mathcal{M})$
    $\langle \mathcal{P}^d_1, disj_1 \rangle :=$ DisjointAxiomsExtensionOptimized$(\mathcal{P}_1, SI_1, SI_{\mathcal{U}})$ ▷ See Algorithm 4
    $\langle \mathcal{P}^d_2, disj_2 \rangle :=$ DisjointAxiomsExtensionOptimized$(\mathcal{P}_2, SI_2, SI_{\mathcal{U}})$
else
    $\langle \mathcal{P}^d_1, disj_1 \rangle :=$ DisjointAxiomsExtensionBasic$(\mathcal{P}_1, SI_1)$ ▷ See Algorithm 3
    $\langle \mathcal{P}^d_2, disj_2 \rangle :=$ DisjointAxiomsExtensionBasic$(\mathcal{P}_2, SI_2)$
end if
$\langle \mathcal{M}', \mathcal{R}^{\approx} \rangle :=$ MappingRepair$(\mathcal{P}^d_1, \mathcal{P}^d_2, \mathcal{M})$ ▷ See Algorithm 2 in [Jiménez-Ruiz et al., 2013]
$disj := disj_1 + disj_2$
return $\langle \mathcal{M}', \mathcal{R}^{\approx}, disj \rangle$

Algorithm 3 Basic disjointness axioms extension

Input: $\mathcal{P}$: propositional theory; SI: structural index
Output: $\mathcal{P}^d$: extended propositional theory; $disj$: number of disjointness rules

$disj := 0$
$\mathcal{P}^d := \mathcal{P}$
for each pair $\langle A, B\rangle \in$ OrderedVariablePairs($\mathcal{P}$) **do**
  if not (areDisj(SI, A, B) **or** inSubSupRel(SI, A, B) **or** shareDesc(SI, A, B)) **then**
    $\mathcal{P}^d := \mathcal{P}^d \cup \{A \wedge B \rightarrow \mathit{false}\}$
    $SI :=$ updateIndex($SI, A \sqcap B \rightarrow \bot$)
    $disj := disj + 1$
  end if
end for
return $\langle \mathcal{P}^d, disj\rangle$

Algorithm 4 Optimised disjointness axioms extension

Input: $\mathcal{P}$: propositional theory; SI: structural index $SI_{\mathcal{U}}$: structural index of the union ontology

Output: $\mathcal{P}^d$: extended propositional theory;*disj*: number of disjointness rules

$disj := 0$
$\mathcal{P}^d := \mathcal{P}$
for $A \rightarrow B \in \text{ConservativityViolations}(SI, SI_{\mathcal{U}})$ **do**
    if not $(\text{areDisj}(SI, A, B))$ **then**
        $\mathcal{P}^d := \mathcal{P}^d \cup \{A \wedge B \rightarrow false\}$
        $SI := \text{updateIndex}(SI, A \sqcap B \rightarrow \bot)$
        $disj := disj + 1$
    end if
end for
return $\langle \mathcal{P}^d, disj \rangle$