Towards Licenses Compatibility and Composition in the Web of Data

Serena Villata* and Fabien Gandon

INRIA Sophia Antipolis, France
{firstname.lastname}@inria.fr

Abstract. We propose a general framework to attach the licensing terms to the data where the compatibility of the licensing terms concerning the data affected by a query is verified, and, if compatible, the licenses are combined into a composite license. The framework returns the composite license as licensing term about the data resulting from the query.

1 Introduction

The absence of clarity concerning the licensing terms does not encourage the reuse of the data in the Web of Data [3]. When consumers query the Web of Data, results from different datasets, and thus released under different licensing terms, are provided. In this paper, we propose first to verify the compatibility among the licensing terms associated to a query result, and second, to compose, if compatible, the distinct licensing terms for creating a composite license. The composite license is retuned together with the query result using the standard SPARQL query results XML format by means of the `<link>`[1] element. We adopt Semantic Web languages only, and reuse the Creative Commons (CC) licenses schema [1] to define the anatomy of our licenses. Licenses are composed by *models*: `cc:Permission`, `cc:Requirement`, and `cc:Prohibition`. Models are composed by *elements* el_i like `ShareAlike`, `Attribution`, and many others. We choose CC because it provides a general schema for licenses specification, even if there are works which should not be released under the CC licenses [5, 3]. For addressing this issue and covering a wider range of machine-readable license specifications, we align the CC vocabulary with the other schemas including licensing terms (Figure 1). We extend and adapt existing proposals for licenses compatibility and composition in the area of service license analysis [2] to the Web of Data scenario. However, the different application scenarios (service composition vs Web of Data) open different problems. The compatibility rules we define are different, and the definition of the composite license mirrors the same differences. Truong et al. [6] address the issue of analyzing data contracts using RDF for the contracts representation. This work concentrates on data contracts and not on data licenses. Krotzsch and Speiser [4] present a semantic framework for evaluating *ShareAlike* recursive statements while we address the problem of licenses composition.

* The author acknowledges the support of the DataLift Project ANR-10-CORD-09.

[1] http://www.w3.org/TR/rdf-sparql-XMLres/#head

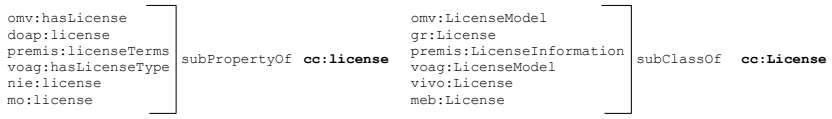

Fig. 1: Alignment between the Creative Commons vocabulary and other vocabularies.

2 Our proposal

Licenses Compatibility. We define a set of compatibility rules assessing the possible compatibility among the elements composing the licenses $\mathcal{L}$, following [2] for service licenses. First, there are certain elements which are broader in scope of permission than other elements, e.g., *Sharing* is more permissive than *Reproduction* [1]. The subsumption rules[2] for `cc:Permission` elements, and the way they can be combined are shown in Table 1. Table 2.a shows whether an element el_1 is compatible with another element el_2, i.e., $el_1 \bigcirc el_2$, concerning `cc:Permission` elements[3]. The rationale is that these elements are compatible if there is a subsumption relation between them.

Subsumption	More permissive	Less permissive
$DerivativeWorks \succ Sharing$	$DerivativeWorks$	$Sharing$
$Distribution \succ Reproduction$	$Distribution$	$Reproduction$
$DerivativeWorks \succ Reproduction$	$DerivativeWorks$	$Reproduction$
$Sharing \succ Reproduction$	$Sharing$	$Reproduction$
$DerivativeWorks \succ Distribution$	$DerivativeWorks$	$Distribution$

Table 1: Compatibility rules for subsumption relation among `cc:Permission` elements.

Second, a possible situation in analyzing license compatibility is that one license $\mathcal{L}_i$ specifies clauses which are not specified by the other license $\mathcal{L}_j$, e.g., $\mathcal{L}_i$ specifies `cc:Prohibition` and $\mathcal{L}_j$ does not specify this clause. Table 2.b shows the compatibility rules for specified elements against *Unspecified* elements[4]. The requirement for specification of *Attribution* does not affect the compatibility with *Unspecified* (the same holds for *Notice*, *SourceCode*, and *CopyLeft*). Concerning prohibitions, these elements are not compatible with *Unspecified*, e.g., commercial use is the default setting of the licenses, thus we cannot assume compatibility if commercial use is denied by *NonCommercial*. For permissions, we follow the "conservative" approach where unspecification means a denial of compatibility[5].

[2] Subsumption $\succ$ means that there is a compatibility if a certain license element el_i is more permissive, i.e., it accepts more, than the other license element el_j.

[3] Rules are expressed under the form of truth tables where elements are evaluated as compatible T, or incompatible F.

[4] We interpret unspecified elements as "do not care", as in [2].

[5] Table 2.c shows three exceptions of compatible elements from distinct models.

el_1	el_2	$el_1 \bigcirc el_2$
$Sharing$	$DerivativeWorks$	T
$Reproduction$	$Distribution$	T
$Reproduction$	$DerivativeWorks$	T
$Reproduction$	$Sharing$	T
$Distribution$	$DerivativeWorks$	T
$Sharing$	$Distribution$	F

(a)

el_1	el_2	$\bigcirc$	$\mathcal{L}_1 \odot \mathcal{L}_2$
$Attribution$	$ShareAlike$	T	$el_1 \wedge el_2$
$Attribution$	$NonCommercial$	T	$el_1 \wedge el_2$
$ShareAlike$	$NonCommercial$	T	$el_1 \wedge el_2$

(c)

el_1	el_2	$el_1 \bigcirc el_2$
$Notice$	$Unspecified$	T
$Attribution$	$Unspecified$	T
$ShareAlike$	$Unspecified$	T
$SourceCode$	$Unspecified$	T
$CopyLeft$	$Unspecified$	T
$NonCommercial$	$Unspecified$	F
$HighIncomeNationUse$	$Unspecified$	F
$Reproduction$	$Unspecified$	F
$Distribution$	$Unspecified$	F
$DerivativeWorks$	$Unspecified$	F
$Sharing$	$Unspecified$	F

(b)

Table 2: (a) Compatibility rules among `cc:Permission` elements, (b) Compatibility rules among `cc:Requirement`, `cc:Prohibition`, `cc:Permission` elements against *Unspecified*, (c) Composition rules among `cc:Requirement` and `cc:Prohibition` elements.

Let $\mathcal{L}(C)$ be the set of licenses associated to the named graphs affected by the consumer's query, we say that two licenses are compatible if the models in both the licenses are compatible [7]. The models are compatible if (i) the models are the same, (ii) the models are composed by elements which satisfy the compatibility rules (Table 2.c), and (iii) their elements are compatible. The elements are compatible if (i) the elements are the same, (ii) the elements satisfy the subsumption rules (Table 1), (iii) the elements satisfy the compatibility rules against *Unspecified* (Table 2.b), and (iv) the elements satisfy the compatibility rules (Table 2.a-c). If the licenses are not compatible, then we leave to the data provider to decide the strategy to deal with this situation, e.g., the data is returned together with the more constraining license among $\mathcal{L}(C)$.

Licenses Composition. If the licenses are compatible then we compose them such that the resulting composite license $\mathcal{L}_c$ ($\odot$ is the composition relation) satisfies the following properties: $\mathcal{L}_c$ can be generated only if all the licenses composing it are compatible, and $\mathcal{L}_c$ is consistent with the set of licenses used to compose it. The definition of $\mathcal{L}_c$ is achieved through the definition of (i) redefinition rules to be applied in case a subsumption relation holds (Table 1), (ii) composition rules necessary to maintain the consistency of $\mathcal{L}_c$ w.r.t. $\mathcal{L}(C)$ (Table 2.c), and (iii) heuristics to compose the elements of each license into $\mathcal{L}_c$. We consider three basic heuristics: *OR-composition*: $\forall l \in \mathcal{L}_i$ then $l \in \mathcal{L}_c$; *AND-composition*: if $\exists l \in \mathcal{L}_1 \wedge \ldots \wedge \mathcal{L}_n$ then $l \in \mathcal{L}_c$; *Constraining-value*: most constraining $l \in \mathcal{L}(C)$ is included in $\mathcal{L}_c$. We leave to the data provider the choice of her best strategy for composing the licenses, e.g., AND-composition typically leads to a shorter and simpler license, while OR-decomposition leads to a more complex license where all the clauses in $\mathcal{L}(C)$ are listed. For example, assume we want to analyze two licenses $\mathcal{L}_1$ and $\mathcal{L}_2$ (Figure 2.a-b). We first compare the two licenses at the model level. Both licenses contains the model `cc:Permission`. Even if the two elements are not the same, there is a subsumption relation between them (Table 1). The first license does not contain the model `cc:Prohibition` and the

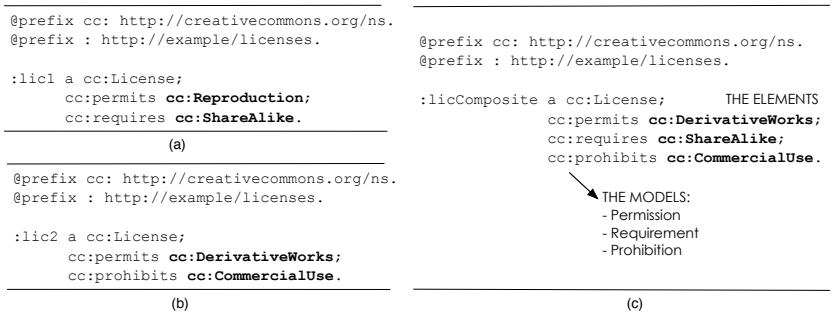

Fig. 2: Example of compatibility evaluation and composition of two licenses.

second license does not contain the model `cc:Requirement`. However, the two elements `cc:ShareAlike` and `cc:CommercialUse` are compatible (Table 2.c). Thus, the two licenses are compatible. $\mathcal{L}_c$ is obtained using the more permissive permission, and the OR-composition heuristic.

3 Future challenges

The first point to be addressed is a legal and social validation of the proposed framework. This means to evaluate, not only quantitatively the performance of the algorithms to retrieve remote licensing statements referenced in the data, but also the legal value of the composite license. We will address an evaluation of the disparate named graphs that might be used in a "typical" query, and their relative proportions that have compatible or incompatible licenses. Finally, we are defining more complex heuristics like the one looking for the minimal composite license.

References

1. H. Abelson, B. Adida, M. Linksvayer, and N. Yergler. ccREL: The creative commons rights expression language. Technical report, 2008.
2. G. R. Gangadharan, M. Weiss, V. D'Andrea, and R. Iannella. Service license composition and compatibility analysis. In *ICSOC, LNCS 4749*, pages 257–269, 2007.
3. T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011.
4. M. Krötzsch and S. Speiser. Sharealike your data: Self-referential usage policies for the semantic web. In *ISWC, LNCS 7031*, pages 354–369, 2011.
5. P. Miller, R. Styles, and T. Heath. Open data commons, a license for open data. In *LDOW*, 2008.
6. H. L. Truong, G. R. Gangadharan, M. Comerio, S. Dustdar, and F. De Paoli. On analyzing and developing data contracts in cloud-based data marketplaces. In *APSCC, IEEE*, pages 174–181, 2011.
7. S. Villata and F. Gandon. Licenses compatibility and composition in the Web of Data. In *COLD*, 2012.