Possibilistic testing of OWL axioms against RDF data

Andrea Tettamanzi, Catherine Faron Zucker, Fabien Gandon

Possibilistic Testing of OWL Axioms Against RDF Data

Andrea G. B. Tettamanzi[a], Catherine Faron-Zucker[a], Fabien Gandon[a]

[a] *Université Côte d'Azur, Inria, CNRS, I3S, 2000, route des Lucioles, Sophia Antipolis, France*

Abstract

We develop the theory of a possibilistic framework for OWL 2 axiom testing against RDF datasets, as an alternative to statistics-based heuristics. The intuition behind it is to evaluate the credibility of OWL 2 axioms based on the *evidence* available in the form of a set of facts contained in a chosen RDF dataset. To achieve it, we first define the notions of development, content, support, confirmation and counterexample of an axiom. Then we use these notions to define the possibility and necessity of an axiom and its acceptance/rejection index combining both of them. Finally, we report a practical application of the proposed framework to test `SubClassOf` axioms against the DBpedia RDF dataset.

Keywords: Possibility Theory, Linked Data, Ontology Learning, OWL 2, Axioms

1. Introduction

Ontology learning [1] is a broad field of research, aiming at overcoming the knowledge acquisition bottleneck through the automatic generation of ontologies, that has started to emerge at the beginning of this century, mainly within the context of the semantic Web. The input for ontology learning can be text in natural language or existing ontologies (typically expressed in OWL) and in-

Email addresses: `andrea.tettamanzi@unice.fr` (Andrea G. B. Tettamanzi), `faron@unice.fr` (Catherine Faron-Zucker), `fabien.gandon@inria.fr` (Fabien Gandon)

stance data (typically represented in RDF) [2]. In the former case, the focus is on the population of ontologies with facts derived from text using natural language processing methods, although the generation of lightweight taxonomies can also be undertaken. In the latter case, which is the one we are interested in, induction-based methods like the ones developed in inductive logic programming and data mining are developed to detect meaningful patterns and learn schema axioms from existing instance data (facts) and their metadata, if available.

On a related note, there exists a need for evaluating and validating ontologies, be they the result of an analysis effort or of a semi-automatic learning method, and/or validating instance data. Indeed, instead of starting from the *a priori* assumption that a given ontology is correct and verify whether the facts contained in an RDF base satisfy it, one may treat ontologies like hypotheses and develop a methodology to verify whether the RDF facts corroborate or falsify them. Ontology learning and validation are thus strictly related. They could even be seen as an agile and test-driven approach to ontology development, where the linked data is used as a giant test case library not only to validate the schema but even to suggest new developments.

Both ontology learning and ontology/data validation rely critically on (candidate) axiom scoring. To see why, let us consider the following example. While constructing an ontology for a given domain (say, politics), based on the description of instances in a given dataset, e.g., DBpedia, we might suspect that a mayor is an elected representative. Before we insert this knowledge into the ontology, we should score the corresponding axiom `SubClassOf(Mayor ElectedRepresentative)` against the statements in the dataset, i.e., measure the extent to which it is compatible with them. Conversely, for validation, imagine that an ontology about politics models the fact that plurality of offices is banned, through a set of axioms like `SubClassOf(Mayor ObjectComplementOf(MP))`. In order to check whether a given country obeys this rule, we might score the above axiom against the linked open government data of that country.

In this paper, we will tackle the problem of testing a single, isolated axiom, which is anyway the first step to solve the problem of validating an entire ontology.

This article is organized as follows: Section 2 discusses related work on ontology learning and validation. Section 3 presents the principles of axiom testing and Section 4 discusses the difficulties and shortcomings of conventional probability-based scoring heuristics, which motivate the search for an alternative. Section 5 presents our proposal of an axiom scoring heuristics based on possibility theory. A computational framework for axiom scoring based on such heuristics is then presented in Section 6 and evaluated on `SubClassOf` axioms. Section 7 draws some conclusions and gives directions for future work.

2. Related Work

Recent contributions towards the automatic creation of OWL 2 ontologies from large repositories of RDF facts include FOIL-like algorithms for learning concept definitions [3], statistical schema induction via association rule mining [4], and light-weight schema enrichment methods based on the DL-Learner framework [5, 6]. All these methods apply and extend techniques developed within inductive logic programming (ILP) [7]. For a recent survey of the wider field of ontology learning, see [2].

The growing need for evaluating and validating ontologies is witnessed by general methodological investigations [8, 9], surveys [10] and tools like OOPS! [11] for detecting pitfalls in ontologies. Ontology engineering methodologies, such as METHONTOLOGY [12], distinguish two validation activities, namely verification (through formal methods, syntax, logics, etc.) and validation through usage. Whilst this latter is usually thought of as user studies, an automatic process of validation based on RDF data would provide a cheap and scalable assistance, whereby the existing linked data may be regarded as usage traces that can be used to test and improve the ontologies, much like log mining can be used to provide test cases for development in the replay approaches. Alternatively, one may regard the ontology as a set of integrity constraints and check if the data satisfy them, using a tool like Pellet integrity constraint validator (ICV), which translates OWL ontologies into SPARQL queries to automatically validate RDF data [13]. The RDF Data Shapes W3C Working Group has been created in 2014 and published in 2017 a working draft, intended to become a W3C recommendation, of the SHACL Shapes Constraint Language, a language for validating RDF graphs against a set of structural conditions.[1] A similar approach also underlies the idea of test-driven evaluation of linked data quality [14]. To this end, OWL ontologies are interpreted under the closed-world assumption and the weak unique name assumption.

The most popular scoring heuristics proposed in the literature are based on statistical inference (see, e.g., [6]). As such a probability-based framework is not always completely satisfactory, we have recently proposed [15, 16] an axiom scoring heuristics based on a formalization in possibility theory of the notions of logical content of a theory and of falsification, loosely inspired by Karl Popper's approach to epistemology, and working with an open-world semantics. In this article, we justify such proposal and we develop a theory of OWL axiom testing against RDF facts based on possibility theory, whose output is a degree of possibility and necessity of an axiom, given the available evidence. Our proposal is coherent with a recently proposed possibilistic extension of description logics [17, 18].

In particular, our first attempts [15] pointed out that a possibilistic approach to test candidate axioms could be beneficial to ontology learning, as well as to ontology and knowledge base validation, although at the cost of a heavier computational cost than the probabilistic scores it aims to complement. Further investigation [16] showed that time capping can alleviate the computation of the proposed possibilistic axiom scoring heuristics without giving up the precision of the scores.

[1] https://www.w3.org/TR/shacl/

3. Principles of OWL 2 Axiom Testing

The problem we study may be stated as follows: given a *hypothesis* about the relations holding among some entities of a domain, syntactically expressed in the form of an OWL 2 axiom, we wish to evaluate its credibility based on the *evidence* available in the form of a set of facts contained in an RDF dataset and, therefore, syntactically expressed in RDF. We call this task *axiom testing*.

If, for a moment, we abstract away from the particular syntax of the hypothesis and of the available evidence, what we have here is a fundamental problem in epistemology, with important ramifications in statistical inference, data mining, inductive reasoning, medical diagnosis, judicial decision making, and even the philosophy of science. Central to this problem is the notion of *confirmation*: see [19] for a general overview of the major approaches to confirmation theory in contemporary philosophy. All the approaches build on logical entailment (from evidence to the hypothesis or from the hypothesis to evidence, to which background knowledge may be added). The approach we follow may be classified as a form of extended hypothetico-deductivism, whereby, roughly speaking, evidence e confirms a hypothesis h if the latter entails it, $h \models e$, and disconfirms it if the former entails the negation of the latter, $e \models \neg h$. As we will see, other considerations will be added to extend this basic idea.

Testing an OWL 2 axiom against an RDF dataset can thus be done by checking whether the formulas entailed by it are confirmed by the facts contained in the RDF dataset.[2] The rest of this section will be devoted to formalizing and developing this intuition.

[2]Note that calling linked data search engines like Sindice could virtually extend the dataset to the whole LOD cloud.

3.1. OWL 2 Direct Model-Theoretic Semantics and Development of OWL 2 Axioms

We refer to the model-theoretic semantics of OWL 2 as defined in [20].[3] An interpretation $\mathcal{I}$ for a datatype map D and a vocabulary V over D is defined by an interpretation domain $\Delta^{\mathcal{I}} = \Delta_I \cup \Delta_D$ (Δ_I is the *object domain* and Δ_D the *data domain*), and a valuation function $\cdot^{\mathcal{I}}$ with seven restrictions: $\cdot^C$ mapping class expressions to subsets of Δ_I, $\cdot^{OP}$ mapping object properties to subsets of $\Delta_I \times \Delta_I$, $\cdot^{DP}$ mapping data properties to subsets of $\Delta_I \times \Delta_D$, $\cdot^I$ mapping individuals to elements of Δ_I, $\cdot^{DT}$ mapping datatypes to subsets of Δ_D, $\cdot^{LT}$ mapping literals to elements of Δ_D and $\cdot^{FT}$ mapping facets to subsets of Δ_D.

Table 1 provides a reference of the model-theoretic semantics of OWL 2 expressions.

Table 2 provides a reference of the semantics of the 32 axiom types of OWL 2.

We aim at operationalizing the model-theoretic semantics of OWL 2 axioms into corresponding first-order logic formulas which will serve as a basis to query an RDF dataset in order to test OWL 2 candidate axioms against it. It was proposed by Hempel [21] that, given some body of evidence, a hypothesis ϕ can be developed into a finite ground formula, which he calls the *development* of the hypothesis. It is useful to recall Hempel's proposal first, which we will then adapt to RDF + OWL 2.

Let $\mathcal{L}$ be a finite first-order language; let $e, h \in \mathcal{L}$ be the available evidence and a hypothesis, respectively; let C be a finite set of individual constants of $\mathcal{L}$ (typically, those occurring non-vacuously in e). The *development* of hypothesis h according to C is the formula $D_C(h)$, such that $h \models D_C(h)$, defined recursively as follows: Let $\phi, \psi \in \mathcal{L}$,

1. if $C = \emptyset$ or ϕ is atomic, then $D_C(\phi) = \phi$;
2. otherwise,
   (a) $D_C(\neg\phi) = \neg D_C(\phi)$;

[3] http://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211/, Section 2.2 Interpretations

Table 1: The model-theoretic semantics of OWL 2 expressions. The first column gives the OWL 2 functional syntax of the expression, the second column its more compact $\mathcal{SHOIQ}$ description logic syntax, and the last column shows its semantics.

OWL 2 Functional Syntax	DL Syntax	Interpretation
$\mathsf{ObjectInverseOf}(R)$	R^-	$(R^-)^\mathcal{I} = \{\langle y, x\rangle \mid \langle x, y\rangle \in R^\mathcal{I}\}$
$\mathsf{DataIntersectionOf}(D_1 \ \ldots \ D_n)$	$D_1 \sqcap \ldots \sqcap D_n$	$D_1^\mathcal{I} \cap \ldots \cap D_n^\mathcal{I}$
$\mathsf{DataUnionOf}(D_1 \ \ldots \ D_n)$	$D_1 \sqcup \ldots \sqcup D_n$	$D_1^\mathcal{I} \cup \ldots \cup D_n^\mathcal{I}$
$\mathsf{DataComplementOf}(D)$	$\neg D$	$\mathtt{D}^{\mathrm{arity}(D)} \setminus D^\mathcal{I}$
$\mathsf{DataOneOf}(d_1 \ \ldots \ d_n)$	$\{d_1, \ldots, d_n\}$	$\{d_1^\mathcal{I}, \ldots, d_n^\mathcal{I}\}$
$\mathsf{DatatypeRestriction}(D \ F_1 \ d_1 \ \ldots \ F_n \ d_n)$		$D^\mathcal{I} \cap \langle F_1, d_1\rangle^\mathcal{I} \cap \ldots \cap \langle F_n, d_n\rangle^\mathcal{I}$
$\mathsf{ObjectIntersectionOf}(C_1 \ \ldots \ C_n)$	$C_1 \sqcap \ldots \sqcap C_n$	$C_1^\mathcal{I} \cap \ldots \cap C_n^\mathcal{I}$
$\mathsf{ObjectUnionOf}(C_1 \ \ldots \ C_n)$	$C_1 \sqcup \ldots \sqcup C_n$	$C_1^\mathcal{I} \cup \ldots \cup C_n^\mathcal{I}$
$\mathsf{ObjectComplementOf}(C)$	$\neg C$	$\Delta^\mathcal{I} \setminus C^\mathcal{I}$
$\mathsf{ObjectOneOf}(a_1 \ \ldots \ a_n)$	$\{a_1, \ldots, a_n\}$	$\{a_1^\mathcal{I}, \ldots, a_n^\mathcal{I}\}$
$\mathsf{ObjectSomeValuesFrom}(R \ C)$	$\exists R.C$	$\{x \mid \exists y.\langle x, y\rangle \in R^\mathcal{I} \wedge y \in C^\mathcal{I}\}$
$\mathsf{ObjectAllValuesFrom}(R \ C)$	$\forall R.C$	$\{x \mid \forall y.\langle x, y\rangle \in R^\mathcal{I} \Rightarrow y \in C^\mathcal{I}\}$
$\mathsf{ObjectHasValue}(R \ a)$	$\exists R.\{a\}$	$\{x \mid \langle x, a^\mathcal{I}\rangle \in R^\mathcal{I}\}$
$\mathsf{ObjectHasSelf}(R)$	$\exists R.\mathsf{Self}$	$\{x \mid \langle x, x\rangle \in R^\mathcal{I}\}$
$\mathsf{ObjectMinCardinality}(n \ R)$	$\geq nR.\top$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I}\}\| \geq n\}$
$\mathsf{ObjectMaxCardinality}(n \ R)$	$\leq nR.\top$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I}\}\| \leq n\}$
$\mathsf{ObjectExactCardinality}(n \ R)$	$= nR.\top$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I}\}\| = n\}$
$\mathsf{ObjectMinCardinality}(n \ R \ C)$	$\geq nR.C$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I} \wedge y \in C^\mathcal{I}\}\| \geq n\}$
$\mathsf{ObjectMaxCardinality}(n \ R \ C)$	$\leq nR.C$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I} \wedge y \in C^\mathcal{I}\}\| \leq n\}$
$\mathsf{ObjectExactCardinality}(n \ R \ C)$	$= nR.C$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I} \wedge y \in C^\mathcal{I}\}\| = n\}$
$\mathsf{DataSomeValuesFrom}(R \ D)$	$\exists R.D$	$\{x \mid \exists y.\langle x, y\rangle \in R^\mathcal{I} \wedge y \in D^\mathcal{I}\}$
$\mathsf{DataAllValuesFrom}(R \ D)$	$\forall R.D$	$\{x \mid \forall y.\langle x, y\rangle \in R^\mathcal{I} \Rightarrow y \in D^\mathcal{I}\}$
$\mathsf{DataHasValue}(R \ d)$	$\exists R.d$	$\{x \mid \langle x, d^\mathcal{I}\rangle \in R^\mathcal{I}\}$
$\mathsf{DataMinCardinality}(n \ R)$	$\geq nR.\top$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I}\}\| \geq n\}$
$\mathsf{DataMaxCardinality}(n \ R)$	$\leq nR.\top$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I}\}\| \leq n\}$
$\mathsf{DataExactCardinality}(n \ R)$	$= nR.\top$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I}\}\| = n\}$
$\mathsf{DataMinCardinality}(n \ R \ D)$	$\geq nR.D$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I} \wedge y \in D^\mathcal{I}\}\| \geq n\}$
$\mathsf{DataMaxCardinality}(n \ R \ D)$	$\leq nR.D$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I} \wedge y \in D^\mathcal{I}\}\| \leq n\}$
$\mathsf{DataExactCardinality}(n \ R \ D)$	$= nR.D$	$\{x \mid \|\{y \mid \langle x, y\rangle \in R^\mathcal{I} \wedge y \in D^\mathcal{I}\}\| = n\}$

Table 2: The model-theoretic semantics of OWL 2 axioms. The first column gives the OWL 2 Functional syntax of the axiom, the second column its more compact $\mathcal{SHOIQ}$ description logic syntax, and the last column shows it semantics.

OWL 2 Functional Syntax	DL Syntax	Semantics
SubClassOf$(C\ D)$	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
EquivalentClasses$(C_1\ \ldots\ C_n)$	$C_i \equiv C_j,\ i,j \in \{1,\ldots n\}$	$C_i^{\mathcal{I}} = C_j^{\mathcal{I}},\ i,j \in \{1,\ldots n\}$
DisjointClasses$(C_1\ \ldots\ C_n)$	$\mathsf{Dis}(C_1,\ldots,C_n)$	$C_i^{\mathcal{I}} \cap C_j^{\mathcal{I}} = \emptyset,\ i,j \in \{1,\ldots n\},\ i \neq j$
DisjointUnion$(C\ C_1\ \ldots\ C_n)$	$C \equiv C_1 \sqcup \ldots \sqcup C_n$, and	$C^{\mathcal{I}} = C_1^{\mathcal{I}} \cup \ldots \cup C_n^{\mathcal{I}}$, and
	$\mathsf{Dis}(C_1,\ldots,C_n)$	$C_i^{\mathcal{I}} \cap C_j^{\mathcal{I}} = \emptyset,\ i,j \in \{1,\ldots n\},\ i \neq j$
SubObjectPropertyOf(S, R)	$S \sqsubseteq R$	$S^{\mathcal{I}} \subseteq R^{\mathcal{I}}$
SubObjectPropertyOf(w, R), with	$S_1 \ldots S_n \sqsubseteq R$	$S_1^{\mathcal{I}} \circ \ldots \circ S_n^{\mathcal{I}} \subseteq R^{\mathcal{I}}$, i.e., $\forall y_0,\ldots,y_n$,
$w =$ ObjectPropertyChain$(S_1\ \ldots\ S_n)$		$\langle y_0, y_1\rangle \in S_1^{\mathcal{I}} \wedge \ldots \wedge \langle y_{n-1}, y_n\rangle \in S_n^{\mathcal{I}}$
		$\Rightarrow \langle y_0, y_n\rangle \in R^{\mathcal{I}}$
EquivalentObjectProperties$(R_1\ \ldots\ R_n)$	$R_i \equiv R_j,\ i,j \in \{1,\ldots n\}$	$R_i^{\mathcal{I}} = R_j^{\mathcal{I}},\ i,j \in \{1,\ldots n\}$
DisjointObjectProperties$(R_1\ \ldots\ R_n)$	$\mathsf{Dis}(R_1,\ldots,R_n)$	$R_i^{\mathcal{I}} \cap R_j^{\mathcal{I}} = \emptyset,\ i,j \in \{1,\ldots n\},\ i \neq j$
ObjectPropertyDomain$(R\ C)$	$\geq 1R \sqsubseteq C$	$\langle x, y\rangle \in R^{\mathcal{I}} \Rightarrow x \in C^{\mathcal{I}}$
ObjectPropertyRange$(R\ C)$	$\top \sqsubseteq \forall R.C$	$\langle x, y\rangle \in R^{\mathcal{I}} \Rightarrow y \in C^{\mathcal{I}}$
InverseObjectProperties$(S\ R)$	$S \equiv R^-$	$S^{\mathcal{I}} = \{\langle y, x\rangle \mid \langle x, y\rangle \in R^{\mathcal{I}}\}$
FunctionalObjectProperty(R)	$\mathsf{Fun}(R)$	$\langle x, y\rangle \in R^{\mathcal{I}} \wedge \langle x, z\rangle \in R^{\mathcal{I}} \Rightarrow y = z$
InverseFunctionalObjectProperty(R)	$\mathsf{Fun}(R^-)$	$\langle x, y\rangle \in R^{\mathcal{I}} \wedge \langle z, y\rangle \in R^{\mathcal{I}} \Rightarrow x = z$
ReflexiveObjectProperty(R)	$\mathsf{Ref}(R)$	$\langle x, x\rangle \in R^{\mathcal{I}}$
IrreflexiveObjectProperty(R)	$\mathsf{Irr}(R)$	$\langle x, x\rangle \notin R^{\mathcal{I}}$
SymmetricObjectProperty(R)	$\mathsf{Sym}(R)$	$\langle x, y\rangle \in R^{\mathcal{I}} \Rightarrow \langle y, x\rangle \in R^{\mathcal{I}}$
AsymmetricObjectProperty(R)	$\mathsf{Asy}(R)$	$\langle x, y\rangle \in R^{\mathcal{I}} \Rightarrow \langle y, x\rangle \notin R^{\mathcal{I}}$
TransitiveObjectProperty(R)	$\mathsf{Tra}(R)$	$\langle x, y\rangle \in R^{\mathcal{I}} \wedge \langle y, z\rangle \in R^{\mathcal{I}} \Rightarrow \langle x, z\rangle \in R^{\mathcal{I}}$
SubDataPropertyOf(S, R)	$S \sqsubseteq R$	$S^{\mathcal{I}} \subseteq R^{\mathcal{I}}$
EquivalentDataProperties$(R_1\ \ldots\ R_n)$	$R_i \equiv R_j,\ i,j \in \{1,\ldots n\}$	$R_i^{\mathcal{I}} = R_j^{\mathcal{I}},\ i,j \in \{1,\ldots n\}$
DisjointDataProperties$(R_1\ \ldots\ R_n)$	$\mathsf{Dis}(R_1,\ldots,R_n)$	$R_i^{\mathcal{I}} \cap R_j^{\mathcal{I}} = \emptyset,\ i,j \in \{1,\ldots n\},\ i \neq j$
DataPropertyDomain$(R\ C)$	$\geq 1R \sqsubseteq C$	$\langle x, y\rangle \in R^{\mathcal{I}} \Rightarrow x \in C^{\mathcal{I}}$
DataPropertyRange$(R\ D)$	$\top \sqsubseteq \forall R.D$	$\langle x, y\rangle \in R^{\mathcal{I}} \Rightarrow y \in D^{\mathcal{I}}$
FunctionalDataProperty(R)	$\mathsf{Fun}(R)$	$\langle x, y\rangle \in R^{\mathcal{I}} \wedge \langle x, z\rangle \in R^{\mathcal{I}} \Rightarrow y = z$
DatatypeDefinition$(T\ D)$	$T \equiv D$	$T^{\mathcal{I}} = D^{\mathcal{I}}$
HasKey$(C\ (R_1\ \ldots\ R_n)\ (S_1\ \ldots\ S_m))$	$\mathsf{Key}(C) =$	$a, b \in C^{\mathcal{I}}$ a, a_i, b, b_i named individuals
with R_i object properties	$\{R_1,\ldots,R_n,S_1,\ldots,S_m\}$	$\wedge\langle a, a_i\rangle \in R_i^{\mathcal{I}} \wedge \langle b, b_i\rangle \in R_i^{\mathcal{I}}$
and S_i data properties		$\wedge\langle a, d_i\rangle \in S_i^{\mathcal{I}} \wedge \langle b, e_i\rangle \in S_i^{\mathcal{I}} \Rightarrow a = b$
SameIndividual$(a_1\ \ldots\ a_n)$	$a_i \doteq a_j,\ i,j \in \{1,\ldots,n\}$	$a_i^{\mathcal{I}} = a_j^{\mathcal{I}},\ i,j \in \{1,\ldots n\}$
DifferentIndividuals$(a_1\ \ldots\ a_n)$	$a_i \not\doteq a_j,\ i,j \in \{1,\ldots,n\},\ i \neq j$	$a_i^{\mathcal{I}} \neq a_j^{\mathcal{I}},\ i,j \in \{1,\ldots n\},\ i \neq j$
ClassAssertion$(C\ a)$	$C(a)$	$a^{\mathcal{I}} \in C^{\mathcal{I}}$
ObjectPropertyAssertion$(R\ a\ b)$	$R(a,b)$	$\langle a^{\mathcal{I}}, b^{\mathcal{I}}\rangle \in R^{\mathcal{I}}$
NegativeObjectPropertyAssertion$(R\ a\ b)$	$\neg R(a,b)$	$\langle a^{\mathcal{I}}, b^{\mathcal{I}}\rangle \notin R^{\mathcal{I}}$
DataPropertyAssertion$(R\ a\ d)$	$R(a,d)$	$\langle a^{\mathcal{I}}, d^{\mathcal{I}}\rangle \in R^{\mathcal{I}}$
NegativeDataPropertyAssertion$(R\ a\ d)$	$\neg R(a,d)$	$\langle a^{\mathcal{I}}, d^{\mathcal{I}}\rangle \notin R^{\mathcal{I}}$

(b) $D_C(\phi \vee \psi) = D_C(\phi) \vee D_C(\psi)$;

(c) $D_C(\phi \wedge \psi) = D_C(\phi) \wedge D_C(\psi)$;

(d) $D_C(\forall x \phi) = \bigwedge_{c \in C} D_C(\phi\{c/x\})$;

(e) $D_C(\exists x \phi) = \bigvee_{c \in C} D_C(\phi\{c/x\})$.

In the above definition, $\phi\{c/x\}$ stands for the formula obtained from ϕ by substituting all free occurrences of variable x with constant c.

We can observe that $D_C(\phi)$, as defined above, can always be transformed either into conjunctive normal form (CNF) or disjunctive normal form (DNF) by repeated application of the De Morgan Laws, i.e.

$$D_C(\phi) = \bigwedge_i \psi_i \quad \text{or} \quad D_C(\phi) = \bigvee_i \psi_i. \tag{1}$$

In either case, the ground formulas ψ_i, which we may call *basic statements*, may be tested directly against the available facts to compute a degree of corroboration of hypothesis ϕ.

We shall now define the notion of *development* of an OWL 2 axiom with respect to an RDF dataset. That notion relies on a transformation, which translates an OWL 2 axiom into a first-order logic formula based on the set-theoretic formulas of the OWL direct semantics.[4]

Definition 1. Let $t(\cdot; x, y)$ be recursively defined as follows, with an OWL 2 entity, expression, or axiom as the first argument and x, y variables:

- Entities:
  - if d is a data value (a literal), $t(d; x, y) = (x = d)$;
  - if a is an individual name (an IRI), $t(a; x, y) = (x = a)$;
  - if C is an atomic concept, $t(C; x, y) = C(x)$;
  - if D is an atomic datatype, $t(D; x, y) = D(x)$;
  - if R is an atomic relation, $t(R; x, y) = R(x, y)$;

[4]This transformation is similar to the two mappings defined in [22] (pages 154–155) to show the equivalence of DL and a two-variable fragment of first-order logic.

- Expressions:
  - $t(R^{-}; x, y) = t(R; y, x)$;
  - $t(D_1 \sqcap \ldots \sqcap D_n; x, y) = t(D_1; x, y) \wedge \ldots \wedge t(D_n; x, y)$;
  - $t(D_1 \sqcup \ldots \sqcup D_n; x, y) = t(D_1; x, y) \vee \ldots \vee t(D_n; x, y)$;
  - $t(\neg D; x, y) = \neg t(D; x, y)$;
  - $t(\{d_1, \ldots, d_n\}; x, y) = t(d_1; x, y) \vee \ldots \vee t(d_n; x, y)$;
  - $t(C_1 \sqcap \ldots \sqcap C_n; x, y) = t(C_1; x, y) \wedge \ldots \wedge t(C_n; x, y)$;
  - $t(C_1 \sqcup \ldots \sqcup C_n; x, y) = t(C_1; x, y) \vee \ldots \vee t(C_n; x, y)$;
  - $t(\neg C; x, y) = \neg t(C; x, y)$;
  - $t(\{a_1, \ldots, a_n\}; x, y) = t(a_1; x, y) \vee \ldots \vee t(a_n; x, y)$;
  - $t(\exists R.C; x, y) = \exists y(t(R; x, y) \wedge t(C; y, z))$;
  - $t(\forall R.C; x, y) = \forall y(\neg t(R; x, y) \vee t(C; y, z))$;
  - $t(\exists R.\{a\}; x, y) = t(R; x, a)$;
  - $t(\exists R.\mathsf{Self}; x, y) = t(R; x, x)$;
  - $t(\geq nR.\top; x, y) = (\|\{y \mid t(R; x, y)\}\| \geq n)$;
  - $t(\leq nR.\top; x, y) = (\|\{y \mid t(R; x, y)\}\| \leq n)$;
  - $t(= nR.\top; x, y) = (\|\{y \mid t(R; x, y)\}\| = n)$;
  - $t(\geq nR.C; x, y) = (\|\{y \mid t(R; x, y) \wedge t(C; y, z)\}\| \geq n)$;
  - $t(\leq nR.C; x, y) = (\|\{y \mid t(R; x, y) \wedge t(C; y, z)\}\| \leq n)$;
  - $t(= nR.C; x, y) = (\|\{y \mid t(R; x, y) \wedge t(C; y, z)\}\| = n)$;
  - $t(\exists R.D; x, y) = \exists y(t(R; x, y) \wedge t(D; y, z))$;
  - $t(\forall R.D; x, y) = \forall y(\neg t(R; x, y) \vee t(D; y, z))$;
  - $t(\exists R.\{d\}; x, y) = t(R; x, d)$;
  - $t(\geq nR.D; x, y) = (\|\{y \mid t(R; x, y) \wedge t(D; y, z)\}\| \geq n)$;
  - $t(\leq nR.D; x, y) = (\|\{y \mid t(R; x, y) \wedge t(D; y, z)\}\| \leq n)$;

- $t(= nR.D; x, y) = (\|\{y \mid t(R; x, y) \wedge t(D; y, z)\}\| = n)$;

- Axioms:
  - $t(C_1 \sqsubseteq C_2; x, y) = \forall x(\neg t(C_1; x, y) \vee t(C_2; x, y))$;
  - $t(C_1 \equiv C_2; x, y) = \forall x((t(C_1; x, y) \wedge t(C_2; x, y)) \vee (\neg t(C_1; x, y) \wedge \neg t(C_2; x, y)))$;
  - $t(\mathsf{Dis}(C_1, \ldots, C_n); x, y) = \bigwedge_{i=1}^{n} \bigwedge_{j=i+1}^{n} (\neg t(C_i; x, y) \vee \neg t(C_j; x, y))$;
  - $t(C \equiv C_1 \sqcup \ldots \sqcup C_n, \mathsf{Dis}(C_1, \ldots, C_n); x, y) = t(C \equiv C_1 \sqcup \ldots \sqcup C_n; x, y) \wedge t(\mathsf{Dis}(C_1, \ldots, C_n); x, y)$;
  - $t(S \sqsubseteq R; x, y) = \forall x \forall y(\neg t(S; x, y) \vee t(R; x, y))$;
  - $t(S_1 \ldots S_n \sqsubseteq R; x, y) = \forall x \forall z_1 \ldots \forall z_{n-1} \forall y(\neg t(S_1; x, z_1) \vee \neg t(S_2; z_1, z_2) \vee \ldots \vee \neg t(S_n; z_{n-1}, y) \vee t(R; x, y))$;
  - $t(R_1 \equiv R_2; x, y) = \forall x \forall y((t(R_1; x, y) \wedge t(R_2; x, y)) \vee (\neg t(R_1; x, y) \wedge \neg t(R_2; x, y)))$;
  - $t(\mathsf{Dis}(R_1, \ldots, R_n); x, y) = \bigwedge_{i=1}^{n} \bigwedge_{j=i+1}^{n} (\neg t(R_i; x, y) \vee \neg t(R_j; x, y))$;
  - $t(\geq 1R \sqsubseteq C; x, y) = \forall x \forall y(\neg t(R; x, y) \vee t(C; x, y))$;
  - $t(\top \sqsubseteq \forall R.C) = \forall x \forall y(\neg t(R; x, y) \vee t(C; y, z))$;
  - $t(S \equiv R^-; x, y) = \forall x \forall y((t(S; x, y) \wedge t(R; y, x)) \vee (\neg t(S; x, y) \wedge \neg t(R; y, x)))$;
  - $t(\mathsf{Fun}(R); x, y) = \forall x \forall y \forall z(\neg t(R; x, y) \vee \neg t(R; x, z) \vee y = z)$;
  - $t(\mathsf{Fun}(R^-); x, y) = \forall x \forall y \forall z(\neg t(R; x, y) \vee \neg t(R; z, y) \vee x = z)$;
  - $t(\mathsf{Ref}(R); x, y) = \forall x(t(R; x, x))$;
  - $t(\mathsf{Irr}(R); x, y) = \forall x(\neg t(R; x, x))$;
  - $t(\mathsf{Sym}(R); x, y) = \forall x \forall y(\neg t(R; x, y) \vee t(R; y, x))$;
  - $t(\mathsf{Asy}(R); x, y) = \forall x \forall y(\neg t(R; x, y) \vee \neg t(R; y, x))$;
  - $t(\mathsf{Tra}(R); x, y) = \forall x \forall y \forall z(\neg t(R; x, y) \vee \neg t(R; y, z) \vee t(R; x, z))$;
  - $t(\top \sqsubseteq \forall R.D) = \forall x \forall y(\neg t(R; x, y) \vee t(D; y, z))$;
  - $t(T \equiv D; x, y) = \forall x((t(T; x, y) \wedge t(D; x, y)) \vee (\neg t(T; x, y) \wedge \neg t(D; x, y)))$;

– $t(\mathsf{Key}(C) = \{R_1, \ldots, R_n\}; x, y) = \forall x \forall z \forall z_1 \ldots \forall a z_n (\neg t(C; x, y) \vee t(C; z, y) \vee \bigvee_{i=1}^{n} (\neg R_i(x, z_i) \vee \neg R_i(z, z_i)) \vee x = z)$;

– $t(a \doteq b; x, y) = (a = b)$;

– $t(a \neq b; x, y) = \neg(a = b)$;

– $t(C(a); x, y) = C(a)$;

– $t(\neg C(a); x, y) = \neg C(a)$;

– $t(R(a, b); x, y) = R(a, b)$;

– $t(\neg R(a, b); x, y) = \neg R(a, b)$;

– $t(R(a, d); x, y) = R(a, d)$;

– $t(\neg R(a, d); x, y) = \neg R(a, d)$;

where z, z_i, denote "fresh" variables, C, C_i denote concepts, D, D_i, T datatypes, R, R_i, S, S_i (object or data) properties, a, b individuals, and d data values.

For instance, let us consider the following OWL 2 axiom:

$$\phi = \texttt{SubClassOf}(\texttt{dbo:LaunchPad}\ \texttt{dbo:Infrastructure}),$$

Its transformation into FOL is:

$$\begin{aligned}
&t(\phi, x, y) = \\
&t(\texttt{SubClassOf}(\texttt{dbo:LaunchPad}\ \texttt{dbo:Infrastructure}), x, y) = \\
&\forall x(\neg t(\texttt{dbo:LaunchPad}, x, y) \vee t(\texttt{dbo:Infrastructure}), x, y)) = \\
&\forall x(\neg \texttt{dbo:LaunchPad}(x) \vee \texttt{dbo:Infrastructure})(x))
\end{aligned}$$

The semantic equivalence of $t(\phi; x, y)$ and ϕ can be readily verified by observing that the definition of $t(\phi; x, y)$ is obtained from the set-theoretic formulas of the OWL direct semantics of ϕ (cf. Tables 1 and 2) by

- substituting all symbols $a^{\mathcal{I}}$ denoting elements of $\Delta^{\mathcal{I}}$ by their names (IRI) a,
- substituting all symbols $C^{\mathcal{I}}$ denoting subsets of $\Delta^{\mathcal{I}}$ by their corresponing class name or datatype name C, and

- substituting all symbols $R^{\mathcal{I}}$ denoting subsets of $\Delta_I \times \Delta_I$ or $\Delta_I \times \Delta_D$ by their corresponding object or data property name R.

Definition 2 (Development of an Axiom)**.** Let ϕ be an OWL 2 axiom and let $\mathcal{K}$ be an RDF dataset. The *development* $D_{\mathcal{K}}(\phi)$ of ϕ with respect to $\mathcal{K}$ is defined as follows:

1. Let $\hat{\phi} = t(\phi; x, y)$ (cf. Definition 1);
2. Let $I(\mathcal{K})$ be the set of (named or blank) individuals occurring in $\mathcal{K}$ (it is reasonable to assume that $I(\mathcal{K}) \neq \emptyset$ and $I(\mathcal{K})$ is finite);
3. $D_{\mathcal{K}}(\phi) = NF(\hat{D}(\hat{\phi}))$, where
   - $\hat{D}(\cdot)$ is recursively defined as follows:
     (a) if $\hat{\phi}$ is atomic, then $\hat{D}(\hat{\phi}) = \hat{\phi}$,
     (b) $\hat{D}(\neg\hat{\phi}) = \neg\hat{D}(\hat{\phi})$,
     (c) $\hat{D}(\hat{\phi} \vee \hat{\psi}) = \hat{D}(\hat{\phi}) \vee \hat{D}(\hat{\psi})$,
     (d) $\hat{D}(\hat{\phi} \wedge \hat{\psi}) = \hat{D}(\hat{\phi}) \wedge \hat{D}(\hat{\psi})$,
     (e) $\hat{D}(\forall x\hat{\phi}) = \bigwedge_{c \in I(\mathcal{K})} \hat{D}(\hat{\phi}\{c/x\})$,
     (f) $\hat{D}(\exists x\hat{\phi}) = \bigvee_{c \in I(\mathcal{K})} \hat{D}(\hat{\phi}\{c/x\})$;
   - and $NF(\cdot)$ is a function transforming a formula either in conjunctive or in disjunctive normal form. We will see in Section 5 that $D_{\mathcal{K}}(\phi)$ being in conjunctive or disjunctive form has some consequences on the way ϕ is scored. We shall call the conjuncts (disjuncts, respectively) of $D_{\mathcal{K}}(\phi)$ if it is in conjunctive (disjunctive) normal form the *basic statements* of $D_{\mathcal{K}}(\phi)$. $NF(\cdot)$ chooses between a conjunctive or disjunctive normal form to produce the formula with the greatest number of basic statements.

3.2. Content, Support, Confirmation, and Counterexample of an OWL 2 Axiom

We are now ready to define the notion of *content* of an axiom, which is at the foundation of axiom testing.

Definition 3 (Content of an Axiom)**.** Let ϕ be an OWL 2 axiom and let $\mathcal{K}$ be an RDF dataset. The *content* of ϕ, given $\mathcal{K}$, $content_{\mathcal{K}}(\phi)$, is defined as the set of all the basic statements of $D_{\mathcal{K}}(\phi)$.

We will omit the subscript $\mathcal{K}$ when there is no ambiguity and write simply $content(\phi)$ to denote the content of axiom ϕ.

For example, let us consider the test of candidate axiom

$$\phi = \texttt{SubClassOf(dbo:LaunchPad dbo:Infrastructure)},$$

against the DBpedia dataset. As we have seen above, this axiom translates into the first-order formula

$$\hat{\phi} = t(\phi; x, y) = \forall x(\neg\texttt{dbo:LaunchPad}(x) \vee \texttt{dbo:Infrastructure}(x)),$$

and is finally developed according to DBpedia into:

$$D_{\mathrm{DBpedia}}(\phi) = \bigwedge_{r \in I(\mathrm{DBpedia})} (\neg\texttt{dbo:LaunchPad}(x) \vee \texttt{dbo:Infrastructure}(x)).$$

We may thus express the content of ϕ as:

$$\begin{aligned}&content(\texttt{dbo:LaunchPad} \sqsubseteq \texttt{dbo:Infrastructure}) =\\ &\{\neg\texttt{dbo:LaunchPad}(r) \vee \texttt{dbo:Infrastructure}(r) :\\ &r \text{ is a resource occurring in DBpedia}\}.\end{aligned}$$

By construction, $\forall\psi \in content(\phi)$, $\phi \models \psi$. Indeed, let $\mathcal{I}$ be a model of ϕ; by definition, $\mathcal{I}$ is also a model of the formula which expresses the semantics of ϕ and *a fortiori*, also of all its groundings; since ψ is a grounding of the formula which expresses the semantics of ϕ, $\mathcal{I}$ is a model of ψ.

Now, given a formula $\psi \in content(\phi)$ and an RDF dataset $\mathcal{K}$, there are three cases:

1. $\mathcal{K} \models \psi$: in this case, we will call ψ a *confirmation* of ϕ;
2. $\mathcal{K} \models \neg\psi$: in this case, we will call ψ a *counterexample* of ϕ;
3. $\mathcal{K} \not\models \psi$ and $\mathcal{K} \not\models \neg\psi$: in this case, ψ is neither a confirmation nor a counterexample of ϕ.

The definition of $content(\phi)$ may be refined by adopting Scheffler and Goodman's principle of *selective confirmation* [23], which characterizes a confirmation as a fact not simply confirming a candidate axiom, but, further, favoring the axiom rather than its contrary. For instance, the occurrence of a black raven *selectively confirms* the axiom $\texttt{Raven} \sqsubseteq \texttt{Black}$ because it both confirms it and fails to confirm its negation, namely that there exist ravens that are not black. On the contrary, the observation of a green apple does not contradict $\texttt{Raven} \sqsubseteq \texttt{Black}$, but it does not disconfirm $\texttt{Raven} \not\sqsubseteq \texttt{Black}$ either, i.e., it does not selectively confirm $\texttt{Raven} \sqsubseteq \texttt{Black}$.

The definition of $content(\phi)$ may thus be further refined, in order to restrict it just to those ψ which can be counterexamples of ϕ, thus leaving out all those ψ which would be trivial confirmations of ϕ. That is like saying that, to test a hypothesis, we have to try, as hard as we can, to refute it.

A formal definition of the content of an axiom taking into account this principle of selective confirmation can hardly be given in the general case, since it depends very closely on the form of the axiom. This should rather be shifted to the computational definition of each type of OWL 2 axioms (see Section 6). For example, in the case of a $\texttt{SubClassOf}(C\ D)$ axiom, all ψ involving the existence of a resource r for which $\mathcal{K} \not\models C(r)$ will either be confirmations (if $\mathcal{K} \models D(r)$) or they will fall into Case 3 otherwise. Therefore, such ψ will not be interesting and should be left out of $content(\texttt{SubClassOf}(C\ D))$.

Applying this principle greatly reduces $content(\phi)$ and, therefore, the number of ψ that will have to be checked.

Definition 4 (Support of an Axiom)**.** Let ϕ be an OWL 2 axiom and let $\mathcal{K}$ be an RDF dataset. We shall denote by u_ϕ the support of ϕ, defined as the cardinality of its content:

$$u_\phi = \|content(\phi)\|.$$

Notice that, since $I(\mathcal{K})$ is finite, $content(\phi)$ is a finite set and, therefore u_ϕ is a natural number.

Definition 5. We denote by u_ϕ^+ the number of formulas $\psi \in content(\phi)$ which are entailed by the RDF dataset (confirmations); and by u_ϕ^- the number of such formulas whose negation $\neg\psi$ is entailed by the RDF dataset (counterexamples).

Notice that it is possible that, for some $\psi \in \mathit{content}(\phi)$, the RDF dataset entails neither ψ nor $\neg\psi$ (Case 3 above). Therefore,

$$u_\phi^+ + u_\phi^- \leq u_\phi. \tag{2}$$

For example, when testing $\phi = \texttt{dbo:LaunchPad} \sqsubseteq \texttt{dbo:Infrastructure}$ against the DBpedia dataset, we found that $u_\phi = 85$, $u_\phi^+ = 83$, i.e., there are 83 confirmations of ϕ in the dataset; and $u_\phi^- = 1$, i.e., there is 1 counterexample in the dataset, namely

$$\texttt{dbo:LaunchPad(:USA)} \Rightarrow \texttt{dbo:Infrastructure(:USA)},$$

since

$$\begin{aligned} \text{DBpedia} &\models \texttt{dbo:LaunchPad(:USA)}, \\ \text{DBpedia} &\models \neg\texttt{dbo:Infrastructure(:USA)}. \end{aligned}$$

and one formula in *content*(ϕ) neither is a confirmation nor a counterexample, namely

$$\texttt{dbo:LaunchPad(:Cape\_Canaveral)} \Rightarrow \texttt{dbo:Infrastructure(:Cape\_Canaveral)},$$

because

$$\begin{aligned} \text{DBpedia} &\models \texttt{dbo:LaunchPad(:Cape\_Canaveral)}, \\ \text{DBpedia} &\not\models \texttt{dbo:Infrastructure(:Cape\_Canaveral)}, \\ \text{DBpedia} &\not\models \neg\texttt{dbo:Infrastructure(:Cape\_Canaveral)}. \end{aligned}$$

The following are further interesting properties of u_ϕ, u_ϕ^+, and u_ϕ^-.

Theorem 1. *Let ϕ be a candidate OWL 2 axiom. Then ϕ and $\neg\phi$ have the same support: $u_\phi = u_{\neg\phi}$.*

Proof. We know that either $D_{\mathcal{K}}(\phi)$ is in CNF or it is in DNF. In the former case,

$$D_{\mathcal{K}}(\phi) = \bigwedge_{i=1}^{u_\phi} \psi_i;$$

by the De Morgan Laws,

$$D_{\mathcal{K}}(\neg\phi) = \neg D_{\mathcal{K}}(\phi) = \neg \bigwedge_{i=1}^{u_\phi} \psi_i = \bigvee_{i=1}^{u_\phi} \neg\psi_i,$$

whence we see that the basic statements of $\neg\phi$ are the negations of the basic statements of ϕ. Therefore, $u_{\neg\phi} = u_\phi$.

Analogously in the case $D_{\mathcal{K}}(\phi)$ is in DNF. □

Theorem 2. *Let ϕ be a candidate OWL 2 axiom. If the RDF dataset $\mathcal{K}$ is consistent, then*

1. $u^+_\phi = u^-_{\neg\phi}$ *(the confirmations of ϕ are counterexamples of $\neg\phi$);*
2. $u^-_\phi = u^+_{\neg\phi}$ *(the counterexamples of ϕ are confirmations of $\neg\phi$).*

Proof. From the proof of Theorem 1, we know that the basic statements of $\neg\phi$ are the negations of the basic statements of ϕ. Therefore, given a basic statement $\psi_i \in content(\phi)$,

- if $\mathcal{K} \models \psi_i$ (ψ_i is a confirmation of ϕ), then $\mathcal{K} \not\models \neg\psi_i$, since $\mathcal{K}$ is consistent; but then $\neg\psi_i$ is a counterexample of $\neg\phi$;
- if $\mathcal{K} \models \neg\psi_i$ (ψ_i is a counterexample of ϕ), then $\mathcal{K} \not\models \psi_i$, since $\mathcal{K}$ is consistent; but then $\neg\psi_i$ is a confirmation of $\neg\phi$;
- if $\mathcal{K} \not\models \psi_i$ and $\mathcal{K} \not\models \neg\psi_i$, then ψ_i is neither a confirmation nor a counterexample for both ϕ and $\neg\phi$.

□

Likewise, we could characterize the support, confirmations, and counterexamples of the conjunction and of the disjunction of OWL axioms. For instance, it would be easy to prove that, if both $D_{\mathcal{K}}(\phi)$ and $D_{\mathcal{K}}(\psi)$ are in CNF, then both

$D_{\mathcal{K}}(\phi \vee \psi)$ and $D_{\mathcal{K}}(\phi \wedge \psi)$ are in CNF too and, furthermore, $u_{\phi\vee\psi} = u_\phi \cdot u_\psi$, $u_{\phi\wedge\psi} = u_\phi + u_\psi$, $u^+_{\phi\vee\psi} = u^+_\phi \cdot u_\psi + u^+_\psi \cdot u_\phi - u^+_\phi \cdot u^+_\psi$, $u^-_{\phi\vee\psi} = u^-_\phi \cdot u^-_\psi$, etc. However, results like these would be of limited interest here, since the conjunction and the disjunction of OWL axioms are not OWL axioms.

4. A Critique of Probabilistic Candidate Axiom Scoring

Before going on to expound our proposal for candidate axiom testing, let us examine what most researchers would consider an obvious first choice for that task, namely an approach based on statistical hypothesis testing, and explain why we believe this is not a suitable choice.

Indeed, all previous work on automatic knowledge base enrichment we are aware of is based on some form of probabilistic axiom scoring. Most work on data mining, too, relies on model performance measures that are essentially probabilistic (of the frequentist type): consider, for example,

- the *confidence* measure used in association rule mining [24], which can be interpreted as an estimate of the conditional probability that the consequent of a rule is satisfied by a transaction, given that the antecedent is satisfied;

- the *accuracy* measure used in binary classification or prediction, defined as the proportion of correct classifications (both true positives and true negatives) over the total number of cases examined;

- *precision* and *recall*, used in information retrieval as well as in classification and prediction.

If we restrict our attention to the scoring heuristics used for the discovery of OWL 2 axioms from RDF datasets, the approach proposed by Bühmann and Lehmann [6] may be regarded essentially as scoring an axiom by an estimate of the probability that one of its logical consequences is confirmed (or, alternatively, falsified) by the facts stored in the RDF repository.

This relies on the assumption of a binomial distribution, which applies when an experiment (here, checking if a logical consequence of a candidate axiom is confirmed by the facts) is repeated a fixed number of times, each trial having two possible outcomes (conventionally labeled *success* and *failure*; here, we might call them *confirmation*, if the observed fact agrees with the candidate axiom, and *counterexample*, if the observed fact contadicts it), the probability of success being the same for each observation, and the observations being statistically independent.

Estimating the probability of confirmation of axiom ϕ just by $\hat{p}_\phi = u^+_\phi / u_\phi$ would be too crude and would not take the cardinality of the content of ϕ in the RDF repository into account. The parameter estimation must be carried out by performing a statistical inference.

One of the most basic analyses in statistical inference is to form a confidence interval for a binomial parameter p_ϕ (probability of confirmation of axiom ϕ), given a binomial variate u^+_ϕ for sample size u_ϕ and a sample proportion $\hat{p}_\phi = u^+_\phi / u_\phi$. Most introductory statistics textbooks use to this end the Wald confidence interval, based on the asymptotic normality of $\hat{p}_\phi$ and estimating the standard error. This $(1 - \alpha)$ confidence interval for p_ϕ would be

$$\hat{p}_\phi \pm z_{\alpha/2} \sqrt{\hat{p}_\phi (1 - \hat{p}_\phi) / u_\phi}, \tag{3}$$

where z_c denotes the $1 - c$ quantile of the standard normal distribution.

However, the central limit theorem applies poorly to this binomial distribution with $u_\phi < 30$ or where $\hat{p}_\phi$ is close to 0 or 1. The normal approximation fails totally when $\hat{p}_\phi = 0$ or $\hat{p}_\phi = 1$. That is why Bühmann and Lehmann [6] base their probabilistic score on Agresti and Coull's binomial proportion confidence interval [25], an adjustment of the Wald confidence interval which goes: "Add two successes and two failures and then use Formula 3." Such adjustment is specific for constructing 95% confidence intervals.

In fact, Agresti and Coull's suggestion is a simplification of the Wilson score interval

$$\left(\hat{p}_\phi + \frac{z^2_{\alpha/2}}{2u_\phi} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_\phi(1-\hat{p}_\phi) + \frac{z^2_{\alpha/2}}{4u_\phi}}{u_\phi}}\right) \Big/ \left(1 + \frac{z^2_{\alpha/2}}{2u_\phi}\right), \tag{4}$$

which is an approximate binomial confidence interval obtained by inverting the approximately normal test that uses the null, rather than the estimated, standard error. When used to compute the 95% score interval, this confidence interval has coverage probabilities close to the nominal confidence level and can be recommended for use with nearly all sample sizes and parameter values.

A remark about Bühmann and Lehmann's approach is in order. Bühmann and Lehmann only look for confirmations of ϕ, and treat the absence of a confirmation as a failure in the calculation of the confidence interval. This is like making an implicit closed-world assumption. In reality, the probability of finding a confirmation and the probability of finding a counterexample do not add to one, because there is a non-zero probability of finding neither a confirmation nor a counterexample for every potential falsifier of an axiom. Their scoring method should thus be corrected in view of the open-world assumption, for example by using $\hat{p}^* = u^+_\phi/(u^+_\phi + u^-_\phi)$ as the sample proportion instead of $\hat{p}$.

However, there is a more fundamental critique to the very idea of computing the likelihood of axioms based on probabilities. In essence, this idea relies on the assumption that it is possible to compute the probability that an axiom ϕ is true given some evidence e, for example e = "$\psi \in \text{content}(\phi)$ is in the RDF repository", or e = "$\psi \notin \text{content}(\phi)$ is in the RDF repository", or e = "$\psi \in \text{content}(\phi)$ is not in the RDF repository", etc., which, by Bayes' formula, may be written as

$$\Pr(\phi \mid e) = \frac{\Pr(e \mid \phi)\Pr(\phi)}{\Pr(e \mid \phi)\Pr(\phi) + \Pr(e \mid \neg\phi)\Pr(\neg\phi)} \tag{5}$$

However, in order to compute (or estimate) such probability, one should at least be able to estimate probabilities such as

- the probability that a fact confirming ϕ is added to the repository given that ϕ holds;

- the probability that a fact contradicting ϕ is added to the repository in error, i.e., given that ϕ holds;
- the probability that a fact confirming ϕ is added to the repository in error, i.e., given that ϕ does not hold;
- the probability that a fact contradicting ϕ is added to the repository given that ϕ does not hold.

Now, it is not hard to argue that the above probabilities may vary as a function of the concepts and properties involved. Let us take a subsumption axiom $C \sqsubseteq D$ as an example. A fact confirming it is a triple "x `a` D", with $x \in C^{\mathcal{I}}$, whereas a fact contradicting it is a triple "x `a` C'", with $x \in C^{\mathcal{I}}$ and $C' \sqcap C = \bot$. Assuming that $C \sqsubseteq D$ holds, we may suspect that a triple "x `a` D" is much likely to be found in the repository if D is either very specific (and thus "closer" to x) or very general (like `owl:Person`), and less likely if it is somewhere in the middle. This supposition is based on our expectations of what people are likely to say about x: for instance, an average person, if asked "what is this?" when pointing to a basset hound, is more likely to answer "a dog" or "an animal" than, say, "a carnivore" or "a mammal", which, on purely logical grounds, would be perfectly valid things to say about it [26], a phenomenon which John Sowa [27] calls *salience* of an ontological or linguistic term. There is thus an inherent difficulty with estimating the above probabilities, one which cannot be solved otherwise than by performing a large number of experiments, whose results, then, would be hard to generalize. By this argument, any axiom scoring method based on probability or statistics is doomed to be largely arbitrary and subjective or, in other words, *qualitative* and therefore hardly more rigorous or objective than our approach based on possibility theory.

5. A Possibilistic Candidate Axiom Scoring Framework

We present now an axiom scoring heuristics which captures the basic intuition behind the process of axiom discovery based on possibility theory: assigning to a candidate axiom a degree of possibility equal to 1 just means that this axiom is possible, plausible, i.e., it is not contradicted by facts in the knowledge base. This is much weaker than assigning a probability equal to 1, meaning that the candidate axiom certainly *is* an axiom.

5.1. Possibility Theory

Possibility theory [28] is a mathematical theory of epistemic uncertainty. Given a finite universe of discourse Ω, whose elements $\omega \in \Omega$ may be regarded as events, values of a variable, possible worlds, or states of affairs, a possibility distribution is a mapping $\pi : \Omega \to [0, 1]$, which assigns to each ω a degree of possibility ranging from 0 (impossible, excluded) to 1 (completely possible, normal). A possibility distribution π for which there exists a completely possible state of affairs ($\exists \omega \in \Omega : \pi(\omega) = 1$) is said to be *normalized*.

There is a similarity between possibility distribution and probability density. However, it must be stressed that $\pi(\omega) = 1$ just means that ω is a plausible (normal) situation and therefore should not be excluded. A degree of possibility can then be viewed as an upper bound of a degree of probability. See [29] for a discussion about the relationships between fuzzy sets, possibility, and probability degrees. A fundamental difference between possibility theory and probability theory is that possibility theory is suitable to represent incomplete knowledge while probability theory is adapted to represent random and observed phenomena.

A possibility distribution π induces a *possibility measure* and its dual *necessity measure*, denoted by Π and N respectively. Both measures apply to a set $A \subseteq \Omega$ (or to a formula ϕ, by way of the set of its models, $A = \{\omega : \omega \models \phi\}$), and are usually defined as follows:

$$\Pi(A) = \max_{\omega \in A} \pi(\omega); \tag{6}$$

$$N(A) = 1 - \Pi(\bar{A}) = \min_{\omega \in \bar{A}} \{1 - \pi(\omega)\}. \tag{7}$$

In other words, the possibility measure of A corresponds to the greatest of the possibilities associated to its elements; conversely, the necessity measure of A is equivalent to the impossibility of its complement $\bar{A}$.

A generalisation of the above definition can be obtained by replacing the min and the max operators with any dual pair of triangular norm and co-norm.

Here are a few properties of possibility and necessity measures induced by a normalized possibility distribution on a finite universe of discourse Ω:

1. $\Pi(\emptyset) = N(\emptyset) = 0, \quad \Pi(\Omega) = N(\Omega) = 1$;
2. $\Pi(A) = 1 - N(\bar{A})$ (duality);
3. $N(A) \leq \Pi(A)$;
4. $N(A) > 0$ implies $\Pi(A) = 1$, and $\Pi(A) < 1$ implies $N(A) = 0$.

In case of complete ignorance on A, $\Pi(A) = \Pi(\bar{A}) = 1$. The above properties are independent of a particular choice of a dual pair of triangular norm and co-norm. Examples of additional properties that are satisfied for $\langle T, S\rangle$ a dual pair of triangular norm and co-norm are the following:

1. $\Pi(A \cup B) = S(\Pi(A), \Pi(B)) \geq \max\{\Pi(A), \Pi(B)\}$;
2. $N(A \cap B) = T(N(A), N(B)) \leq \min\{N(A), N(B)\}$.

5.2. Possibility and Necessity of an Axiom

It was noted by Popper [30] that there is an inherent asymmetry between confirmations and counterexamples of a hypothesis ϕ. When the development of ϕ is conjunctive, a single counterexample is enough to falsify it, even in the face of any number of confirmations. Conversely, when the development of ϕ is disjunctive, a single confirmation is enough to prove ϕ, no matter how many counterexamples are known. Of course, in the presence of noisy data, a single counterexample is hardly a conclusive argument to reject a hypothesis with a conjunctive development and, likewise, a single confirmation is hardly a conclusive argument to accept a hypothesis with a disjunctive development. This is why we turn to the gradual notions of possibility and necessity.

We shall now lay down a number of intuitive postulates the possibility and necessity of a hypothesis (in the form of an OWL 2 axiom) should satisfy and we shall then propose a mathematical definition of these measures that satisfies all the postulates. The basic principle for establishing the possibility of an axiom ϕ should be that the absence of counterexamples to ϕ (if ϕ has a conjunctive development) or the presence of confirmations to ϕ (if ϕ has a disjunctive development) in the RDF repository means that ϕ is completely possible, i.e., $\Pi(\phi) = 1$. A hypothesis should be regarded as all the more *necessary* as it is explicitly supported by facts and, if it has a conjunctive development, not contradicted by any fact; and all the more *possible* as it is not contradicted by facts. We recall that, by Theorem 2, a confirmation of ϕ is a counterexample of $\neg\phi$ and a counterexample of ϕ is a confirmation of $\neg\phi$.

We give here the properties that, based on common sense and the above considerations, necessity and possibility of an axiom should satisfy. These properties may be taken as postulates which will serve as a basis for a formal definition of Π and N:

1. $\Pi(\phi) = 1$ if $u_\phi^- = 0$ or, if $D(\phi)$ is disjunctive, $u_\phi^+ > 0$, i.e., an axiom is fully possible if no counterexample for it is known; furthermore, if its development is disjunctive, which is typical of axioms whose semantics involves an existential quantification, even one confirmation is sufficient to grant its full possibility;
2. $N(\phi) = 0$ if $u_\phi^+ = 0$ or, if $D(\phi)$ is conjunctive, $u_\phi^- > 0$, i.e., for an axiom to have a non-zero degree of necessity, confirmations for it must be known; however, if its development is conjunctive, which is typical of axioms whose semantics involves a universal quantification, a single counterexample is enough to offset any number of known confirmations;
3. let $u_\phi = u_\psi$; then $\Pi(\phi) > \Pi(\psi)$ iff $u_\phi^- < u_\psi^-$ and, if $D(\phi)$ is disjunctive, $u_\psi^+ = 0$, i.e., the possibility of an axiom is inversely proportional to the number of known counterexamples, unless the axiom has a disjunctive development and at least a confirmation, in which case its possibility is 1 and does not depend on the number of counterexamples; as the number of counterexamples increases, $\Pi(\phi) \to 0$ strictly monotonically, if the de-

velopment of ϕ is conjunctive or, if it is disjunctive, if no confirmations are found;

4. let $u_\phi = u_\psi$; then $N(\phi) > N(\psi)$ iff $u_\phi^+ > u_\psi^+$ and, if $D(\phi)$ is conjunctive, $u_\phi^- = 0$, i.e., the necessity of an axiom increases as the number of confirmations for it increases, unless its development is conjunctive and at least a counterexample for it is known, in which case the necessity of the axiom is zero and does not depend on the number of confirmations; $N(\phi) \to 1$ strictly monotonically as the number of confirmations increases and, if the development of ϕ is conjunctive, no counterexamples are found;
5. let $u_\phi = u_\psi = u_\chi$ and $u_\psi^+ = u_\phi^+ = u_\chi^+ = 0$, and let $u_\psi^- < u_\phi^- < u_\chi^-$: then
$$\frac{\Pi(\psi) - \Pi(\phi)}{u_\phi^- - u_\psi^-} > \frac{\Pi(\phi) - \Pi(\chi)}{u_\chi^- - u_\phi^-},$$
i.e., the first counterexamples found to an axiom should determine a sharper decrease of the degree to which we regard the axiom as possible than any further counterexamples, because these latter will only confirm our suspicions and, therefore, will provide less and less information;
6. let $u_\phi = u_\psi = u_\chi$ and $u_\psi^- = u_\phi^- = u_\chi^- = 0$, and let $u_\psi^+ < u_\phi^+ < u_\chi^+$: then
$$\frac{N(\phi) - N(\psi)}{u_\phi^+ - u_\psi^+} > \frac{N(\chi) - N(\phi)}{u_\chi^+ - u_\phi^+},$$
i.e., in the absence of counterexamples, the first confirmations found to an axiom should determine a sharper increase of the degree to which we regard the axiom as necessary than any further confirmations, because these latter will only add up to our acceptance and, therefore, will provide less and less information.

We propose now a definition of Π and N which satisfies the above postulates.

Definition 6. Let ϕ be an OWL 2 axiom and let u_ϕ be the support of ϕ, u_ϕ^+ the number of its confirmations, and u_ϕ^- the number of its counterexamples. The possibility and necessity of ϕ are defined as follows:

- if $u_\phi > 0$ and $D(\phi)$ is in conjunctive normal form,

$$\Pi(\phi) \quad = \quad 1 - \sqrt{1 - \left(\frac{u_\phi - u_\phi^-}{u_\phi}\right)^2}; \tag{8}$$

$$N(\phi) \quad = \quad \begin{cases} \sqrt{1 - \left(\frac{u_\phi - u_\phi^+}{u_\phi}\right)^2}, & \text{if } u_\phi^- = 0, \\ \\ 0, & \text{if } u_\phi^- > 0; \end{cases} \tag{9}$$

- if $u_\phi > 0$ and $D(\phi)$ is in disjunctive normal form,

$$\Pi(\phi) \quad = \quad \begin{cases} 1 - \sqrt{1 - \left(\frac{u_\phi - u_\phi^-}{u_\phi}\right)^2}, & \text{if } u_\phi^+ = 0, \\ \\ 1, & \text{if } u_\phi^+ > 0; \end{cases} \tag{10}$$

$$N(\phi) \quad = \quad \sqrt{1 - \left(\frac{u_\phi - u_\phi^+}{u_\phi}\right)^2}; \tag{11}$$

$$\tag{12}$$

- if $u_\phi = 0$, $\Pi(\phi) = 1$ and $N(\phi) = 0$, i.e., we are in a state of maximum ignorance, given that no evidence is available in the RDF dataset to assess the credibility of ϕ.

Theorem 3. *The measures* Π *and* N *of Definition 6 satisfy all the postulates of axiom possibility and necessity.*

Proof. Postulates 1 and 2 hold trivially.

To prove that postulate 3 holds, we observe that, when the hypotheses of the postulate hold, $\Pi(\cdot)$ can be expressed as a function of the counterexamples of an axiom,

$$\Pi(u_\phi^-) = 1 - \sqrt{1 - \left(\frac{u - u_\phi^-}{u}\right)^2}, \tag{13}$$

where $u = u_\phi = u_\psi$, which is strictly decreasing; therefore, $\Pi(_\phi) > \Pi(_\psi)$ iff $\phi < \psi$. The proof for postulate 4 is analogous.

To prove that postulate 5 holds, we observe that, once again, when the hypotheses of the postulate hold, $\Pi(\cdot)$ can be expressed as in Equation 13 above;

it will thus suffice to observe that $\Pi(\cdot)$ is strictly concave (since $\Pi'' > 0$, see also Figure 1a) and that u^-_ϕ is in the convex hull of u^-_ϕ and u^-_ψ.

The proof for postulate 6 is analogous. $\square$

Notice that this definition, derived from a quadratic equation, is by no means the only possible one, but it is the simplest one, as the following result suggests.

Theorem 4. *Any definition of* Π *and* N *as linear functions of* u^+_ϕ *and* u^-_ϕ *cannot satisfy all the postulates of axiom possibility and necessity.*

Proof. We show that a linear definition of Π and N would not satisfy Postulates 5 and 6. A linear function f satisfies additivity ($f(x+y) = f(x) + f(y)$) and homogeneity of degree 1 ($f(kx) = kf(x)$). Let us assume $\Pi(u^+_\phi, u^-_\phi)$ and $N(u^+_\phi, u^-_\phi)$ are linear. Then, given $u_\phi = u_\psi = u_\chi = u$, let us assume $u^-_\psi < u^-_\phi < u^-_\chi$ (as in Postulate 5) and, furthermore, $u^+_\phi = u^+_\psi = u_\chi = u^+$; then

$$\frac{\Pi(u^+_\psi, u^-_\psi) - \Pi(u^+_\phi, u^-_\phi)}{u^-_\phi - u^-_\psi} = \frac{\Pi(u^+ - u^+, u^-_\psi - u^-_\phi)}{u^-_\phi - u^-_\psi} = \frac{\Pi(0, u^-_\psi - u^-_\phi)}{u^-_\phi - u^-_\psi}$$

and

$$\frac{\Pi(u^+_\phi, u^-_\phi) - \Pi(u^+_\chi, u^-_\chi)}{u^-_\chi - u^-_\phi} = \frac{\Pi(u^+ - u^+, u^-_\phi - u^-_\chi)}{u^-_\chi - u^-_\phi} = \frac{\Pi(0, u^-_\phi - u^-_\chi)}{u^-_\chi - u^-_\phi}.$$

Now, if $u^-_\phi - u^-_\psi = u^-_\chi - u^-_\phi$, we obtain

$$\frac{\Pi(0, u^-_\psi - u^-_\phi)}{u^-_\phi - u^-_\psi} = \frac{\Pi(0, u^-_\phi - u^-_\chi)}{u^-_\chi - u^-_\phi},$$

which violates Postulate 5. A similar reasoning may be applied to N to show that there exist conditions under which Postulate 6 is violated. $\square$

Figure 1 shows $\Pi(\phi)$ and $N(\phi)$ as a function of u^-_ϕ and u^+_ϕ, respectively. The two functions describe an arc of an ellipse between the minor and the major axes. Besides satisfying the postulates of axiom possibility and necessity, Π and N satisfy the general properties of possibility and necessity measures.

Theorem 5. *The measures* Π *and* N *of Definition 6 satisfy the duality property:* $N(\phi) = 1 - \Pi(\neg\phi)$ *and* $\Pi(\phi) = 1 - N(\neg\phi)$.

Proof. The thesis holds mainly because if $D(\phi)$ is conjunctive, $D(\neg\phi)$ is disjunctive and *vice versa*.

Let us assume $D(\phi)$ is in conjunctive normal form; $\Pi(\phi)$ is then given by Equation 8 and $N(\phi)$ by Equation 9; $D(\neg\phi)$ is in disjunctive normal form; thus $\Pi(\neg\phi)$ is given by Equation 10 and $N(\neg\phi)$ by Equation 11. From Theorems 1 and 2, we know that $u_{\neg\phi} = u_\phi$, $u_\phi^+ = u_{\neg\phi}^-$, and $u_\phi^- = u_{\neg\phi}^+$; we can thus write:

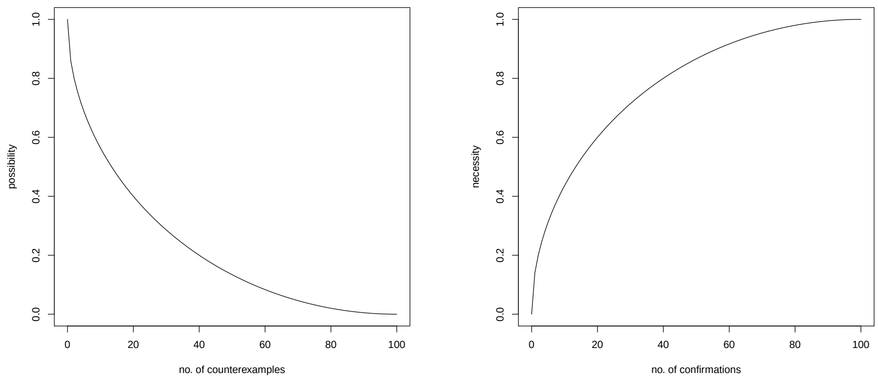
$$\begin{aligned}
N(\phi) &= \begin{cases} \sqrt{1 - \left(\frac{u_\phi - u_\phi^+}{u_\phi}\right)^2}, & \text{if } u_\phi^- = 0, \\ \\ 0, & \text{if } u_\phi^- > 0; \end{cases} \\
&= \begin{cases} \sqrt{1 - \left(\frac{u_{\neg\phi} - u_{\neg\phi}^-}{u_{\neg\phi}}\right)^2}, & \text{if } u_{\neg\phi}^+ = 0, \\ \\ 0, & \text{if } u_{\neg\phi}^+ > 0; \end{cases} \\
&= 1 - \Pi(\neg\phi),
\end{aligned}$$

and

$$\Pi(\phi) = 1 - \sqrt{1 - \left(\frac{u_\phi - u_\phi^-}{u_\phi}\right)^2} = 1 - \sqrt{1 - \left(\frac{u_{\neg\phi} - u_{\neg\phi}^+}{u_{\neg\phi}}\right)^2} = 1 - N(\neg\phi).$$

The same applies when $D(\phi)$ is in disjunctive normal form: just rename ϕ as $\neg\psi$ and $\neg\phi$ as ψ; now $D(\psi)$ is in conjunctive normal form, for which we have just proven the thesis. □

As a matter of fact, we will seldom be interested in computing the necessity and possibility degrees of the negation of OWL 2 axioms, for the simple reason that, in most cases, the latter are not OWL 2 axioms themselves. For instance, while $C \sqsubseteq D$ is an axiom, $\neg(C \sqsubseteq D) \equiv C \not\sqsubseteq D$ is not.



(a) (b)

Figure 1: A plot of $\Pi(\phi)$ as a function of u_ϕ^- (a) and of $N(\phi)$ as a function of u_ϕ^+ (b) when $u_\phi = 100$.

5.3. Axiom Scoring

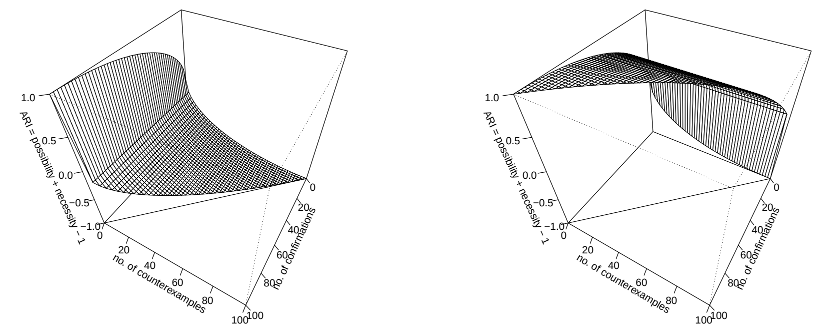
We combine the possibility and necessity of an axiom to define a single handy acceptance/rejection index (ARI) as follows:[5]

$$\mathrm{ARI}(\phi) = N(\phi) - N(\neg\phi) = N(\phi) + \Pi(\phi) - 1 \in [-1, 1]. \tag{14}$$

A negative $\mathrm{ARI}(\phi)$ suggests rejection of ϕ ($\Pi(\phi) < 1$), whilst a positive $\mathrm{ARI}(\phi)$ suggests its acceptance ($N(\phi) > 0$), with a strength proportional to its absolute value. A value close to zero reflects ignorance about the status of ϕ. Figure 2 shows $\mathrm{ARI}(\phi)$ as a function of u_ϕ^- and u_ϕ^+ in the two cases of a ϕ whose development is a conjunction or a disjunction, respectively, of basic statements.

Although this ARI is useful for the purpose of analyzing the results of our experiments and to visualize the distribution of the tested axiom with respect to a single axis, one should always bear in mind that an axiom is scored by the proposed heuristics in terms of two bipolar figures of merit, whose meanings,

[5]The suggestion that this type of representation may simplify the treatment of possibilistic uncertainty in some contexts goes back to [31].



(a) (b)

Figure 2: Two plots of ARI(ϕ) as a function of u_ϕ^- and u_ϕ^+: (a) when ϕ has a conjunctive development and (b) when ϕ has a disjunctive development.

though related, are very different:

- $\Pi(\phi)$ expresses the degree to which ϕ may be considered "normal", in the sense of "not exceptional, not surprising", or not contradicted by actual observations;
- $N(\phi)$, on the other hand, expresses the degree to which ϕ is certain, granted by positive evidence and corroborated by actual observations.

6. Application to `SubClassOf` Axiom Testing

To illustrate how the theory developed in the previous sections can be applied in practice, we summarize here an application to `SubClassOf` axiom testing, developed in [15, 16]. Scoring these axioms with their ARI requires to compute the development of `Class` entities and `ObjectComplementOf` expressions.

We define a mapping $Q(E, \texttt{?x})$ from OWL 2 class expressions to SPARQL graph patterns, where E is an OWL 2 class expression, and `?x` is a variable, such that the query `SELECT DISTINCT ?x WHERE {` $Q(E, \texttt{?x})$ `}` returns all the individuals which are instances of E. We denote this set by $[Q(E, \texttt{?x})]$:

$$[Q(E, \texttt{?x})] = \{v : (?x, v) \in \mathit{ResultSet}(\texttt{SELECT DISTINCT ?x WHERE}\{Q(E, \texttt{?x})\}\}. \tag{15}$$

For a `Class` entity A,

$$Q(A, \texttt{?x}) = \{\texttt{?x a } A\}, \tag{16}$$

where A is a valid IRI.

For an `ObjectComplementOf` expression, things are slightly more complicated, since RDF does not support negation. The model-theoretic semantics of OWL class expressions of the form `ObjectComplementOf(`C`)` ($\neg C$ in DL syntax), where C denotes a class, is $\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$. However, to learn axioms from an RDF dataset, the open-world hypothesis must be made: the absence of supporting evidence does not necessarily contradict an axiom, moreover an axiom might hold even in the face of a few counterexamples. Therefore, as proposed in [15], we define $Q(\neg C, \texttt{?x})$ as follows, to approximate an open-world semantics:

$$\begin{aligned} Q(\neg C, \texttt{?x}) = \{ \ & \texttt{?x a ?dc .} \\ & \texttt{FILTER NOT EXISTS } \{ \texttt{ ?z a ?dc .} \quad Q(C, \texttt{?z}) \ \} \ \}, \end{aligned} \tag{17}$$

where `?z` is a variable that does not occur anywhere else in the query.

For a `Class` entity A, this becomes

$$Q(\neg A, \texttt{?x}) = \{ \texttt{ ?x a ?dc . FILTER NOT EXISTS } \{\texttt{?z a ?dc . ?z a } A\}\}. \tag{18}$$

A computational definition of $u_{C \sqsubseteq D}$ is the following SPARQL query:

$$\begin{aligned} & \texttt{SELECT (count(DISTINCT ?x) AS ?u)} \\ & \texttt{WHERE } \{Q(C, \texttt{?x})\}. \end{aligned} \tag{19}$$

In order to compute the score of `SubClassOf` axioms, $ARI(C \sqsubseteq D)$, we must provide a computational definition of $u^{+}_{C \sqsubseteq D}$ and $u^{-}_{C \sqsubseteq D}$, respectively:

$$\begin{aligned} & \texttt{SELECT (count(DISTINCT ?x) AS ?nConfirm)} \\ & \texttt{WHERE } \{ \ Q(C, \texttt{?x}) \ Q(D, \texttt{?x}) \ \} \end{aligned} \tag{20}$$

and

$$\texttt{SELECT (count(DISTINCT ?x) AS ?nCounter)} \\ \texttt{WHERE } \{ \ Q(C, \texttt{?x}) \ Q(\neg D, \texttt{?x}) \ \}. \tag{21}$$

The results of our first experiments described below showed that an axiom which takes too long to test will likely end up having a very negative score. We defined two heuristics based on this idea.

- We time-cap the SPARQL queries to compute the ARI of a candidate axiom and decide whether to accept or reject it, since above a computation time threshold, the axiom being tested is likely to get a negative ARI and be rejected.
- We construct candidate axioms of the form $C \sqsubseteq D$, by considering the subclasses C in increasing order of the number of classes D sharing at least one instance with C. This enables us to maximize the number of tested and accepted axioms in a given time period, since it appears that the time it takes to test $C \sqsubseteq D$ increases with that number and the lower the time, the higher the ARI.

We evaluated the proposed scoring heuristics by performing tests of `SubClassOf` axioms using DBpedia 3.9 in English as the reference RDF fact repository. In particular, on April 27, 2014, we downloaded the DBpedia dumps of English version 3.9, generated in late March/early April 2013, along with the DBpedia ontology, version 3.9. This local dump of DBpedia, consisting of 812,546,748 RDF triples, with materialized inferences, has been bulk-loaded into Jena TDB and a prototype for performing axiom tests using the proposed method has been coded in Java, using Jena ARQ and TDB to access the RDF repository.

We systematically generated and tested `SubClassOf` axioms involving atomic classes only according to the following protocol: for each of the 442 classes C referred to in the RDF repository, we construct all axioms of the form $C \sqsubseteq D$ such that C and D share at least one instance. Classes D are obtained with the following query:

$$\texttt{SELECT DISTINCT ?D WHERE}\{Q(C, \texttt{?x}) \ \texttt{. ?x a ?D}\}.$$

We experimentally fixed to 20 min the threshold to time-cap the SPARQL queries to compute $u^{+}_{C \sqsubseteq D}$ and $u^{-}_{C \sqsubseteq D}$ in order to decide whether to accept or reject a candidate axiom $C \sqsubseteq D$.

An in-depth quantitative and qualitative analysis of our experimental results is reported in [15] and [16]. Here we summarize the main findings.

We tested 5050 axioms using the time-capping heuristics. Of these, 632 have been also tested without time capping, which is much more expensive in terms of computing time by a factor of 142; the outcome of the test was different on just 25 of them, which represents an error rate of 3.96%, a very reasonable price to pay, in terms of accuracy degradation, in exchange for faster testing. It should be observed that, by construction, these errors are all in the same direction, i.e., some axioms which should be accepted are in fact rejected: our heuristics are conservative, since they do not generate false positives.

Validating the results of our scoring heuristics in absolute term would require having a knowledge engineer tag as true or false every axiom tested and compare her judgment with the test score. Some insights gained from trying to do so are given in [16], but besides being an extremely tedious and error-prone task, manual evaluation is not completely reliable.

In order to obtain a more objective evaluation, we took all `SubClassOf` axioms in the DBpedia ontology and added to them all `SubClassOf` axioms that can be inferred from them, thus obtaining a "gold standard" of axioms that *should* be all considered as valid. This, at least, looks like a reasonable assumption, despite the fact that in [15] a number of potential issues were pointed out with the subsumption axioms of the DBpedia ontology. Of the 5050 tested axioms, 1915 occur in the gold standard; of these, 327 get an ARI below 1/3, which would yield an error rate of about 17%. In fact, in most cases, the ARI of these axioms is around zero, which means that our heuristic gives a suspended judgment. Only 34 axioms have an ARI below $-1/3$. If we took these latter as the real errors, the error rate would fall to just 1.78%.

Finally, a comparison of the proposed scoring heuristic with a probabilistic score, summarized in Figure 3, highlights some remarkable differences in behav-

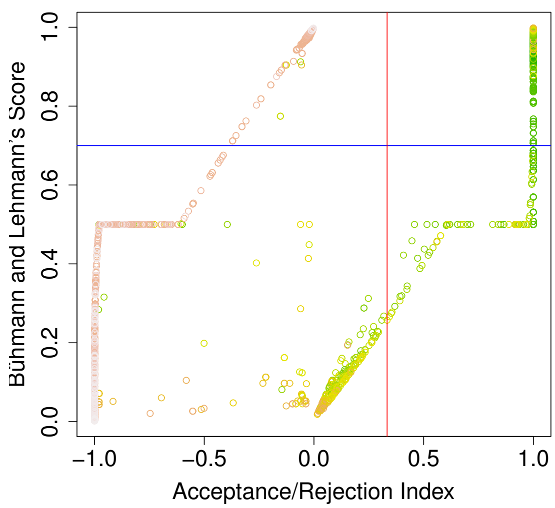


Figure 3: A comparison of the acceptance/rejection index and the probability-based score used in [6] on axioms tested with time capping. The vertical red line shows the acceptance threshold ARI(ϕ) $>$ 1/3; the horizontal blue line the acceptance threshold of 0.7 for the probabilistic score.

ior. In the figure, each axiom is plotted according to its ARI (X-axis) and its probabilistic score computed as in [6] (Y-axis). First of all it is clear that both scores tend to agree in the extremes, with some notable exceptions, but behave quite differently in all other cases. With very few exceptions, all the axioms in the bottom right rectangle are false negatives for the probabilistic score; most axioms in the upper left rectangle are false positives. In addition, the color of the axioms is a function of the time it took to compute their ARI (according to a terrain color scale)

On the 380 axioms tested in [15] without time capping, the probabilistic score with the 0.7 threshold suggested by [6] gave 13 false negatives (7 more than the ARI) and 4 false positives (one more than the ARI). It was observed that most false axiom candidates got an ARI close to -1, whilst their probabilistic scores are almost evenly distributed between 0 and 0.5, which led us to argue that, besides being more accurate, ARI gives clearer indications than the probabilistic score.

The increased accuracy and clarity of the possibilistic score come to a somehow expensive price: we do not have precise figures, but the computational overhead introduced by considering the possibilistic approach instead of a simpler probabilistic score is orders of magnitude higher. The source of such dramatic increase in cost is the execution of the SPARQL query in Equation 17 to approximate the semantics of open-world negation. While it is true that such a query is an integral part of our proposal, one could argue that any probabilistic model wishing to take the open-world assumption into account would have to incur similar costs; furthermore, it is possible that SPARQL query execution engines can be optimized to make the execution of queries of that type significantly faster.

7. Conclusion

We have developed the theory of a possibilistic framework for OWL 2 axiom testing as an alternative to statistics-based heuristics.

The practical application of such a framework has been demonstrated by studying the case of `SubClassOf` axiom testing against the DBpedia database.

A qualitative analysis of the results confirms the interest of using possibilistic axiom scoring heuristics like the one we propose not only to learn axioms from the LOD, but also to drive the validation and debugging of ontologies and RDF datasets.

Future research directions include the systematic computational definition of the content of each kind of OWL 2 axioms, taking into account the principle of selective confirmation. Based on it, we will continue our experiments by testing the possibilistic framework on domain specific datasets and extending it to test other types of OWL axioms, beginning with `SubObjectPropertyOf` and `SubDataPropertyOf` axioms and `SubClassOf` axioms involving `ObjectSomeValuesFrom` class expressions.

References

[1] A. Maedche, S. Staab, Ontology learning for the semantic web, IEEE Intelligent Systems 16 (2) (2001) 72–79. doi:10.1109/5254.920602.

[2] J. Lehmann, J. Völker (Eds.), Perspectives on Ontology Learning, Vol. 18 of Studies on the Semantic Web, IOS Press, Amsterdam, 2014. doi:10.3233/978-1-61499-379-7-i.

[3] N. Fanizzi, C. d'Amato, F. Esposito, DL-FOIL concept learning in description logics, in: F. Zelezný, N. Lavrac (Eds.), Inductive Logic Programming, 18th International Conference, ILP 2008, Prague, Czech Republic, September 10-12, 2008, Proceedings, Vol. 5194 of Lecture Notes in Computer Science, Springer, 2008, pp. 107–121. doi:10.1007/978-3-540-85928-4_12.

[4] D. Fleischhacker, J. Völker, H. Stuckenschmidt, Mining RDF data for property axioms, in: R. Meersman, H. Panetto, T. S. Dillon, S. Rinderle-Ma, P. Dadam, X. Zhou, S. Pearson, A. Ferscha, S. Bergamaschi, I. F. Cruz (Eds.), On the Move to Meaningful Internet Systems: OTM 2012, Confederated International Conferences: CoopIS, DOA-SVI, and ODBASE 2012, Rome, Italy, September 10-14, 2012. Proceedings, Part II, Vol. 7566 of Lecture Notes in Computer Science, Springer, Berlin, 2012, pp. 718–735. doi:10.1007/978-3-642-33615-7_18.

[5] S. Hellmann, J. Lehmann, S. Auer, Learning of OWL class descriptions on very large knowledge bases, Int. J. Semantic Web Inf. Syst. 5 (2) (2009) 25–48. doi:10.4018/jswis.2009040102.

[6] L. Bühmann, J. Lehmann, Universal OWL axiom enrichment for large knowledge bases, in: A. ten Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Aquin, A. Nikolov, N. Aussenac-Gilles, N. Hernandez (Eds.), Knowledge Engineering and Knowledge Management - 18th International Conference, EKAW 2012, Galway City, Ireland, October 8-12, 2012. Proceedings, Vol. 7603 of Lecture Notes in Computer Science, Springer, 2012, pp. 57–71. doi:10.1007/978-3-642-33876-2_8.

[7] S. Muggleton, L. De Raedt, D. Poole, I. Bratko, P. Flach, K. Inoue, A. Srinivasan, ILP turns 20: Biography and future challenges, Machine Learning 86 (2012) 3–23. doi:10.1007/s10994-011-5259-2.

[8] A. Gangemi, C. Catenacci, M. Ciaramita, J. Lehmann, A theoretical framework for ontology evaluation and validation, in: P. Bouquet, G. Tummarello (Eds.), SWAP 2005 - Semantic Web Applications and Perspectives, Proceedings of the 2nd Italian Semantic Web Workshop, University of Trento, Trento, Italy, 14-16 December 2005, Vol. 166 of CEUR Workshop Proceedings, CEUR-WS.org, 2005, p. Article No.: 2.

[9] A. Gangemi, C. Catenacci, M. Ciaramita, J. Lehmann, Modelling ontology evaluation and validation, in: Y. Sure, J. Domingue (Eds.), The Semantic Web: Research and Applications, 3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, June 11-14, 2006, Proceedings, Vol. 4011 of Lecture Notes in Computer Science, Springer, 2006, pp. 140–154. doi:10.1007/11762256_13.

[10] S. Tartir, I. Budak Arpinar, A. P. Sheth, Ontological evaluation and validation, in: R. Poli, M. Healy, A. Kameas (Eds.), Theory and Applications of Ontologies: Computer Applications, Springer, 2010, pp. 115–130. doi:10.1007/978-90-481-8847-5_5.

[11] M. Poveda-Villalón, M. del Carmen Suárez-Figueroa, A. Gómez-Pérez, Validating ontologies with OOPS!, in: A. ten Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Aquin, A. Nikolov, N. Aussenac-Gilles, N. Hernandez (Eds.), Knowledge Engineering and Knowledge Management - 18th International Conference, EKAW 2012, Galway City, Ireland, October 8-12, 2012. Proceedings, Vol. 7603 of Lecture Notes in Computer Science, Springer, 2012, pp. 267–281. doi:10.1007/978-3-642-33876-2_24.

[12] M. Fernández, A. Gómez-Pérez, N. Juristo, METHONTOLOGY: From ontological art towards ontological engineering, Tech. Rep. SS-97-06, AAAI (1997).

[13] E. Sirin, J. Tao, Towards integrity constraints in OWL, in: R. Hoekstra, P. F. Patel-Schneider (Eds.), Proceedings of the 5th International Workshop on OWL: Experiences and Directions (OWLED 2009), Chantilly, VA, United States, October 23–24, 2009, Vol. 529 of CEUR Workshop Proceedings, CEUR-WS.org, 2009, p. Article No.: 9.

[14] D. Kontokostas, P. Westphal, S. Auer, S. Hellmann, J. Lehmann, R. Cornelissen, A. Zaveri, Test-driven evaluation of linked data quality, in: Proceedings of the 23rd International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, Geneva, Switzerland, 2014, pp. 747–758. doi:10.1145/2566486.2568002.

[15] A. G. B. Tettamanzi, C. Faron-Zucker, F. L. Gandon, Testing OWL axioms against RDF facts: A possibilistic approach, in: K. Janowicz, S. Schlobach, P. Lambrix, E. Hyvönen (Eds.), Knowledge Engineering and Knowledge Management - 19th International Conference, EKAW 2014, Linköping, Sweden, November 24-28, 2014. Proceedings, Vol. 8876 of Lecture Notes in Artificial Intelligence, Springer, 2014, pp. 519–530. doi:10.1007/978-3-319-13704-9_39.

[16] A. G. B. Tettamanzi, C. Faron-Zucker, F. Gandon, Dynamically time-capped possibilistic testing of subclassof axioms against rdf data to enrich schemas, in: K. Barker, J. M. Gómez-Pérez (Eds.), K-CAP 2015. Proceedings of the 8th International Conference on Knowledge Capture, Palisades, NY, USA, October 07–10, 2015, ACM, New York, 2015, p. Article No.: 7. doi:10.1145/2815833.2815835.

[17] G. Qi, Q. Ji, J. Z. Pan, J. Du, Extending description logics with uncertainty reasoning in possibilistic logic, International Journal of Intelligent Systems 26 (2011) 353–381. doi:10.1002/int.20470.

[18] G. Qi, Q. Ji, J. Z. Pan, J. Du, PossDL – A possibilistic DL reasoner for uncertainty reasoning and inconsistency handling, in: L. Aroyo, G. Antoniou, E. Hyvönen, A. ten Teije, H. Stuckenschmidt, L. Cabral, T. Tudorache

(Eds.), The Semantic Web: Research and Applications, 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 - June 3, 2010, Proceedings, Part II, Vol. 6089 of Lecture Notes in Computer Science, Springer, 2010, pp. 416–420. doi:10.1007/978-3-642-13489-0_35.

[19] V. Crupi, Confirmation, in: E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, winter 2016 Edition, Metaphysics Research Lab, Stanford University, 2016.
URL https://plato.stanford.edu/archives/win2016/entries/confirmation/

[20] B. Cuenca Grau, B. Motik, P. Patel-Schneider, OWL 2 web ontology language direct semantics (second edition), W3C recommendation, W3C (December 2012).
URL http://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211/

[21] C. G. Hempel, A purely syntactical definition of confirmation, The Journal of Symbolic Logic 8 (4) (December 1943).

[22] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P. Patel-Schneider (Eds.), The Description Logic Handbook: Theory, implementation and applications, Cambridge, 2003.

[23] I. Scheffler, N. Goodman, Selective confirmation and the ravens: A reply to Foster, The Journal of Philosophy 69 (3) (1972) 78–83.

[24] R. Agrawal, T. Imielinski, A. N. Swami, Mining association rules between sets of items in large databases, in: Proc. of the Int. Conf. on Management of Data, ACM Press, 1993, pp. 207–216. doi:10.1145/170035.170072.

[25] A. Agresti, B. A. Coull, Approximate is better than “exact” for interval estimation of binomial proportions, The American Statistician 52 (2) (May 1998). doi:10.1080/00031305.1998.10480550.

[26] G. Lakoff, Women, Fire, and Dangerous Things, University Of Chicago Press, Chicago, 1987.

[27] J. F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks/Cole, Pacific Grove, CA, 2000.

[28] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28.

[29] D. Dubois, H. Prade, Fuzzy sets and probability: Misunderstandings, bridges and gaps, Fuzzy Sets and Systems 40 (1) (1991) 143–202. `doi:10.1109/FUZZY.1993.327367`.

[30] K. Popper, Logik der Forschung, Verlag von Julius Springer, Vienna, 1935.

[31] J. García del Real, R. G. Molina, J. Ríos Carrión, J. Cardeñosa Lera, A simplified technique for using necessity-possibility measures, International Journal of Approximate Reasoning 5 (4) (1991) 399–413. `doi:10.1016/0888-613X(91)90019-I`.