Automatic Partitioning of OWL Ontologies Using $\mathcal{E}$-Connections

Bernardo Cuenca Grau, Bijan Parsia, Evren Sirin, Aditya Kalyanpur
University of Maryland, College Park, MD, USA
`bernardo@minswap.org;bparsia@isr.umd.edu;{evren,aditya}@cs.umd.edu`

1 Motivation

On the Semantic Web, the ability to combine, integrate and reuse ontologies is crucial. The Web Ontology Language (OWL) defines the *owl:imports* construct, which allows to include by reference all the axioms contained in another knowledge base (KB) on the Web. This certainly provides some *syntactic* modularity, but not a *logical* modularity. We have proposed [3] $\mathcal{E}$-Connections as a suitable formalism for combining KBs and for achieving modular ontology development on the Web. $\mathcal{E}$-Connections are KR languages defined as a combination of other logical formalisms. They were originally introduced in [4] mostly as a way to go beyond the expressivity of each of the component logics, while preserving the decidability of the reasoning services in the combination. We have found that $\mathcal{E}$-Connections can help process, evolve, reuse, and understand OWL ontologies.

In this paper, we address the problem of automatically transforming an OWL KB $\mathcal{O}$ into a $\mathcal{E}$-Connection Σ in such a way that each of the relevant sub-domains modeled in $\mathcal{O}$ is represented in a different component of Σ. We present a formal definition and investigation of different variants of the problem, a polynomial solution for some of them, an optimised implementation and some promising empirical results.

We have found that in some large KBs, partitioning to an $\mathcal{E}$-Connection provides modularity benefits. In particular, if a KB can be partitioned, it typically contains several "free standing" components, that is, sub-KBs which do not "use" information from any other components. These KBs can be easily reused and evolved without reference to the rest of the $\mathcal{E}$-Connection. We believe that the factoring out of such independent parts of the original KB alone justifies partitioning for many applications

2 The Partitioning Problem

In this section, we introduce $\mathcal{E}$-Connections in a Semantic Web context, i.e. as a language for combining $\mathcal{SHOIN}(\mathcal{D})$ ontologies. We define different $\mathcal{E}$-Connections based partitioning problems depending on the relationship existing between the input $\mathcal{SHOIN}$ KB and the output $\mathcal{E}$-Connection. For brevity, we give here a slightly simplified version of the language that does not capture datatypes. However, all the results presented in this paper also extend for combinations of ontologies with datatypes [1]

Definition 1 ***(Syntax and Semantics of*** $\mathcal{C}^{\epsilon}_{\mathcal{IHN}}(\mathcal{SHOIN})$***))***

Let, for $i, j = 1, ..., n$ V_{C_i}, $V_{\mathcal{I}_i}, \epsilon_{ij}$ *be countable and pair-wise disjoint sets of atomic concepts, individuals and property names respectively. The set of ij-properties is* $\epsilon_{ij} \cup \{P^{-} | P \in \epsilon_{ji}\}$*. When* $i = j$*, ij-properties are called "roles", whereas if* $j \neq i$ *they are called "links". The sets of i-concepts are built by simultaneous induction as follows:*

$$C := A|\top|\neg D|D \sqcap E|D \sqcup E|\{a\}|\exists P.Z|\forall P.Z| \geq nS| \leq nS$$

Where $A \in V_{C_i}$*,* D, E *are i-concepts,* $a \in V_{\mathcal{I}_i}$*,* Z *is a j-concept and* P, S *ij-properties with* S *simple* [2]*. An i-axiom* $\mathcal{A}$ *is an expression of either of the following forms:*

$$\mathcal{A} := C \sqsubseteq D|P \sqsubseteq Q|Trans(R)|C(a)|P(a,b)$$

With P, Q *ij-properties,* R *an ij-property with* $i = j$*,* C, D *i-concepts,* $a \in V_{\mathcal{I}_i}, b \in V_{\mathcal{I}_j}$*. An* $\mathcal{E}$*-Connection* Σ *with vocabulary* $V_n = \{\{V_{C_i}\}, \{V_{\mathcal{I}_i}\}, \{\epsilon_{ij}\}\}$ *is a collection* $\Sigma = \{\Sigma_1, ..., \Sigma_n\}$*, where* Σ_i *is a finite set of i-axioms.*

An interpretation is a tuple of the form:

$$\mathcal{M} = (\{\mathcal{M}_i\}, \{\mathcal{M}_{ij}\}, i, j = 1..., n)$$

Where $\mathcal{M}_i = (W_i, .^{\mathcal{M}_i})$ *with* $W_i \cap W_j = \emptyset, \forall i \neq j$*. The interpretation functions are applied to ij-properties as follows, with* $P \in \epsilon_{ij}, Q \in \epsilon_{ji}$*:*

$$P^{\mathcal{M}_{ij}} \subseteq W_i \times W_j \mid (Q^{-})^{\mathcal{M}_{ij}} = \{(x,y)|(y,x) \in Q^{\mathcal{M}_{ji}}\}$$

For individuals, $a^{\mathcal{M}_i} \in W_i$ *with* $a \in V_{\mathcal{I}_i}$*. The interpretation functions are applied to i-concepts and i-axioms as shown in Table 1.* $\mathcal{M}$ *is a model of* Σ $(\mathcal{M} \models \Sigma)$ *if* $\forall i = 1, ..., n$*,* $\mathcal{M} \models \mathcal{A}$ *for each i-axiom* $\mathcal{A} \in \Sigma_i$*.*

The language $\mathcal{C}^{\epsilon}_{\mathcal{IHN}}(\mathcal{SHOIN})$ allows for combinations of $\mathcal{SHOIN}$ ontologies in which inverses, number restrictions and hierarchies are allowed on link properties. This is a very expressive language for which we do not have a practical decision procedure yet. However, our goal here has been to provide maximum flexibility for the partitioning. Nontheless, there already exist practical (and implemented) tableau-based algorithms for some expressive fragments

Semantics of i-concepts	Semantics of i-axioms
$A^{\mathcal{M}_i} \subseteq W_i$; $\top^{\mathcal{M}_i} = W_i$ $\{a\}^{\mathcal{M}_i} \subseteq W_i, \#\{a\}^{\mathcal{M}_i} = 1$	$\mathcal{M} \models (C \sqsubseteq D) \leftrightarrow C^{\mathcal{M}_i} \subseteq D^{\mathcal{M}_j}$
$(\neg C)^{\mathcal{M}_i} = W_i - C^{\mathcal{M}_i}$	$\mathcal{M} \models (P \sqsubseteq Q) \leftrightarrow P^{\mathcal{M}_{ij}} \subseteq Q^{\mathcal{M}_{ij}}$
$(C \sqcap D)^{\mathcal{M}_i} = C^{\mathcal{M}_i} \cap D^{\mathcal{M}_i}$	$\mathcal{M} \models Trans(R)$, iff $R^{\mathcal{M}_{ii}} = (R^{\mathcal{M}_{ii}})^+$
$(C \sqcup D)^{\mathcal{M}_i} = C^{\mathcal{M}_i} \cup D^{\mathcal{M}_i}$	$\mathcal{M} \models C(a) \leftrightarrow a^{\mathcal{M}_i} \in C^{\mathcal{M}_i}$
$(\exists P.Z)^{\mathcal{M}_i} = \{x \in W_i \mid \exists y \in W_j, (x,y) \in P^{\mathcal{M}_{ij}}, y \in Z^{\mathcal{M}_j}\}$	$\mathcal{M} \models P(a,b) \leftrightarrow (a^{\mathcal{M}_i}, b^{\mathcal{M}_j}) \in P^{\mathcal{M}_{ij}}$
$(\forall P.Z)^{\mathcal{M}_i} = \{x \in W_i \mid \forall y \in W_j, if\ (x,y) \in P^{\mathcal{M}_{ij}}, \rightarrow y \in Z^{\mathcal{M}_j}\}$	
$(\leq nS)^{\mathcal{M}_{ij}} = \{x \in W_i \mid \#\{t \in W_j \mid (x,t) \in P^{\mathcal{M}_{ij}}\} \leq n\}$	
$(\geq nS)^{\mathcal{M}_{ij}} = \{x \in W_i \mid \#\{t \in W_j \mid (x,t) \in P^{\mathcal{M}_{ij}}\} \geq n\}$	

Table 1: Semantics of i-Concepts and i-Axioms

of this $\mathcal{E}$-Connection language [1], namely $\mathcal{C}^{\epsilon}_{\mathcal{HN}}(\mathcal{SHIQ}, \mathcal{SHOQ}, \mathcal{SHIO})$ and $\mathcal{C}^{\epsilon}_{\mathcal{HI}}(\mathcal{SHIQ}, \mathcal{SHOQ}, \mathcal{SHIO})$.

Definition 2 ***(Partitioned Vocabulary)***
Let $\mathcal{O}$ be a $\mathcal{SHOIN}$ KB with vocabulary $V = (V_C, V_R, V_I)$, the collection $V_n = \{\{V_{C_i}\}, \{V_{I_i}\}, \{\epsilon_{ij}\}\}; i, j = 1, ..., n;.$ is a ***partitioned vocabulary*** *of V iff:*

- $V_C = \bigcup_i V_{C_i}$; $V_I = \bigcup_i V_{I_i}$; $V_R = \bigcup_{ij} \epsilon_{ij}$
- $V_{C_i} \cap V_{C_j} = \emptyset$, $V_{I_i} \cap V_{I_j} = \emptyset$, $\epsilon_{ij} \cap \epsilon_{kl} = \emptyset$ *if either* $i \neq k$ *or* $j \neq l$

The first relationship that one could think of between an OWL KB $\mathcal{O}$ and an $\mathcal{E}$-Connection Σ is a "structural" one, i.e. one in which Σ contains exactly the same entities (concepts, properties and individuals) and axioms as $\mathcal{O}$, but divided into its different components.

Definition 3 ***(Structural Compatibility)*** *Let $\mathcal{O}$ and Σ have respective vocabularies V, V_n. Σ is* ***structurally compatible with $\mathcal{O}$*** *($\Sigma \sim \mathcal{O}$) iff:*

1. *V_n is a partitioned vocabulary of $\mathcal{O}$*
2. $\mathcal{A} \in \Sigma \leftrightarrow \mathcal{A} \in \mathcal{O}$

Structurally compatible $\mathcal{E}$-Connections reveal as a plausible output for partitioning, since they preserve the modeling choices in the original ontology, since no entities or axioms are added, removed or changed during the partitioning process.

However, the semantics of $\Sigma \sim \mathcal{O}$ differs from the semantics of $\mathcal{O}$. For example, suppose that the negation $(\neg C)$ is present in $\mathcal{O}$ and also in Σ_i, for $\Sigma = (\Sigma_1, ..., \Sigma_n)$ with $\Sigma \sim \mathcal{O}$. If $I = (W, .^{\mathcal{I}})$ is an interpretation of $\mathcal{O}$, and $\mathcal{M} = (\{\mathcal{M}_i\}, \{\mathcal{M}_{ij}\}, i, j = 1..., n)$ an interpretation of Σ, then:

$$(\neg C)^{\mathcal{I}} = W - C^{\mathcal{I}} \quad ; \quad (\neg C)^{\mathcal{M}} = W_i - C^{\mathcal{M}_i}$$

---

[1] Obviously, in the restricted way OWL-DL handles datatypes

[2] S is *simple* if it is not transitive and none of its sub-properties is transitive

Observe that $\mathcal{I}$ computes the set difference w.r.t. the whole interpretation domain, whereas $\mathcal{M}$ does it w.r.t. the restricted "local" domain W_i. Similar variations in the semantics are revealed in other constructors.

In order to dillucidate how these differences affect relationship between $\mathcal{O}$ and Σ, we will compare "corresponding" interpretations $\mathcal{I}$ and $\mathcal{M}$, i.e., interpretations that "agree" on atomic concepts, properties and individuals, but "disagree" in general in the evaluation of complex constructs.

Definition 4 ***(Partitioned Interpretation)***

Let $\mathcal{O}$ be an KB with vocabulary V and let V_n be a partitioned vocabulary for V. An interpretation $\mathcal{I} = (W, .^{\mathcal{I}})$ for $\mathcal{O}$ of the following form:

- $W = \bigcup_{i=1,\ldots,n} W_i$ *with* $W_i \cap W_j = \emptyset$ *for* $i \neq j$, *and* $W_i \neq \emptyset$;
- *For each* $A \in V_{C_i}$, $A^{\mathcal{I}} \subseteq W_i$
- *For each* $P \in \epsilon_{ij}$, $P^{\mathcal{I}} \subseteq W_i \times W_j$
- *For each* $a \in V_{I_i}$, $a^{\mathcal{I}} \in W_i$

Is a ***partitioned interpretation*** *of $\mathcal{O}$ with vocabulary V_n, denoted by $\mathcal{I}(V_n)$. Let $\mathcal{O}$ be consistent. We say that $\mathcal{O}$ is* ***partitionable for*** *V_n if there exists an interpretation $\mathcal{I}(V_n)$ s.t. $\mathcal{I}(V_n) \models \mathcal{O}$.*

Definition 5 *Let $\mathcal{O}$ and Σ have vocabularies V and V_n respectively. Let V_n be a partitioned vocabulary for V, then we can establish a relation '$\leftrightarrow$' between the interpretations of Σ and the partitioned interpretations of $\mathcal{O}$ with vocabulary V_n s.t. $\mathcal{I} \leftrightarrow \mathcal{M}$ are related as follows*

- $W_i' = W_i$; $\top^{\mathcal{I}} = \bigcup_i (\top_i)^{\mathcal{M}}$
- $A^{\mathcal{I}} = A^{\mathcal{M}_i}$, *for each* $A \in V_{C_i}$
- $P^{\mathcal{I}} = P^{\mathcal{M}_{ij}}$, *for each* $P \in \epsilon_{ij}$
- $a^{\mathcal{I}} = a^{\mathcal{M}_i}$, *for each* $a \in V_{I_i}$

Clearly, the relation '$\leftrightarrow$' is a bijection between partitioned interpretations of $\mathcal{O}$ and interpretations of Σ.

We are now ready to formulate the notion of semantic compatibility between a DL KB and an $\mathcal{E}$-Connection:

Definition 6 ***(Semantic Compatibility)***

Let $\mathcal{O}$ with vocabulary V be consistent. Let Σ with vocabulary V_n, with V_n a partitioned vocabulary for V, then Σ is ***semantically compatible*** *with $\mathcal{O}$ ($\Sigma \approx \mathcal{O}$) if:*

1. *$\mathcal{O}$ is partitionable for V_n*
2. *If $\mathcal{I} \leftrightarrow \mathcal{M}$, then $\mathcal{M} \models \Sigma$ iff $\mathcal{I} \models \mathcal{O}$*

Semantic compatibility is a desirable relation between the input and the output of a partitioning process. It ensures that equivalent KBs have exactly the same set of compatible $\mathcal{E}$-Connections. It preserves consistency and ensures that existing subsumptions in the class tree and property tree will hold in the $\mathcal{E}$-Connection, as shown in the following Lemma:

Lemma 1 ***(Properties of Semantic Compatibility)***

1. *Let* $\Sigma = \{\mathcal{O}'\}$ *be an* $\mathcal{E}$*-Connection with a single component, then* $\Sigma \approx \mathcal{O}$ *iff* $\mathcal{O} \equiv \mathcal{O}'$
2. *Let* $\Sigma \equiv \Sigma'$*, then* $\Sigma \approx \mathcal{O} \leftrightarrow \Sigma' \approx \mathcal{O}$
3. *Let* $\Sigma \approx \mathcal{O}$*, then* $\mathcal{O}$ *is consistent iff* Σ *is consistent*
4. *Let* $\Pi(\mathcal{O})$ *be the set of* $\mathcal{E}$*-Connections that are semantically compatible with* $\mathcal{O}$*. Then two ontologies* $\mathcal{O}_1$ *and* $\mathcal{O}_2$ *are equivalent (have the same models) iff* $\Pi(\mathcal{O}_1) = \Pi(\mathcal{O}_2)$
5. *Let* $A, B \in V_C$*,* $P, Q \in V_R$*,* $a \in V_i$*. If* $\Sigma \approx \mathcal{O}$*, then:*

$$\begin{gathered}
\mathcal{O} \models (A \sqsubseteq B) \rightarrow \Sigma \models (A \sqsubseteq B) \\
\mathcal{O} \models (P \sqsubseteq Q) \rightarrow \Sigma \models (P \sqsubseteq Q) \\
\mathcal{O} \models A(a) \rightarrow \Sigma \models A(a) \\
A \textit{ unsatisfiable in } \mathcal{O} \rightarrow A \textit{ unsatisfiable in } \Sigma
\end{gathered}$$

We believe that the implications in property 5) also hold in the converse way, i.e. that semantically compatible $\mathcal{E}$-Connections do not introduce new entailments concerning atomic entities.

As an example of structural and semantic compatibility, suppose the following KB:

$$\mathcal{O} = \{(C \sqsubseteq D \sqcup E); (C \sqsubseteq \neg D); (B \sqsubseteq \exists P.A); (A \sqsubseteq F)\}$$

and the $\mathcal{E}$-Connections $\Sigma = (\Sigma_1, \Sigma_2, \Sigma_3)$ and $\Upsilon = (\Upsilon_1, \Upsilon_2, \Upsilon_3)$:

$$\begin{gathered}
\Sigma_1 = \{C \sqsubseteq D \sqcup E,\, C \sqsubseteq \neg D\};\ \Sigma_2 = \{B \sqsubseteq \exists P.A\};\ \Sigma_3 = \{A \sqsubseteq F\} \\
\Upsilon_1 = \{C \sqsubseteq E\}\ ;\ \Upsilon_2 = \{B \sqsubseteq \exists P.A\};\ \Upsilon_3 = \{A \sqsubseteq F\};\ \Upsilon_4 = \{D \sqsubseteq \bot\}
\end{gathered}$$

Observe that $\Sigma \sim \mathcal{O}$ and $\Upsilon \approx \mathcal{O}$. We can formulate at least three partitioning problems depending on what is the desired relationship between the input $\mathcal{O}$ and the output Σ

Definition 7 *(Partitioning Problems) The partitioning problem* ***P1)*** *is the problem of finding, for an input* $\mathcal{O}$*, the* $\mathcal{E}$*-Connection* Σ *with the largest number of components s.t.* $\Sigma \sim \mathcal{O}$*. In the problem* ***P2)*** *we require* $\Sigma \approx \mathcal{O}$ ***in addition to*** $\Sigma \sim \mathcal{O}$*. In* ***P3)*** *we require* $\Sigma \approx \mathcal{O}$ ***instead of*** $\Sigma \sim \mathcal{O}$*.*

In other words, we can enforce structural compatibility only, both structural and semantic compatibility, or semantic compatibility only.

In this paper, we show that **P1)** and **P2)** are solvable in polynomial time, without the intervention of a reasoner in any stage of the process. We leave the solution of **P3)** as an open problem. The key step towards a solution for **P2)** is to devise under what conditions $\Sigma \sim \mathcal{O}$ is also semantically compatible with $\mathcal{O}$. For such analysis, the notion of **$\mathcal{E}$-safety** reveals crucial.

Definition 8 *(**$\mathcal{E}$-safety**) Let: $g : C \in \mathcal{O} \rightarrow \{T, F\}$ be a function mapping every concept $C \in \mathcal{O}$ to a boolean value and recursively defined as follows:*

- *If C is $\top$, then $g(C) = F$*
- *Let $C \in V_C$, then $g(C) = T$*
- *Let C be $\{a\}$, for $a \in V_I$, then $g(C) = T$*
- *Let C be $D \sqcap E$.If $g(D)$=F and $g(E)$=F, then $g(C)$=F. Otherwise, $g(C) = T$*
- *Let C be $D \sqcup E$. If $g(D) = T$ and $g(E) = T$, then $g(C) = T$. Otherwise, $g(C) = F$*
- *Let C be $\neg D$. If $g(D) = T$, then $g(C) = F$ and if $g(D) = F$, then $g(C) = T$.*
- *Let C be $\exists P.D$ or $\geq nP$, then $g(C) = T$*
- *Let C be $\forall P.D$ or $\leq nP$, then $g(C) = F$*

*The KB $\mathcal{O}$ is **$\mathcal{E}$-safe** iff it contains no axiom of the form $C \sqsubseteq D$ s.t. $g(C) = F$ and $g(D) = T$.*

The rationale under the definition of the g function is provided by the following result:

Lemma 2 *Let $\Sigma \sim \mathcal{O}$ and $\mathcal{I} \leftrightarrow \mathcal{M}$. Let C be a concept in $\mathcal{O}$ and an i-concept in Σ, then:*

$$C^{\mathcal{I}} = C^{\mathcal{M}_i} \text{ iff } g(C) = T$$
$$C^{\mathcal{I}} = C^{\mathcal{M}_i} \bigcup_{j \neq i} W_j \text{ iff } g(Z) = F$$

Also $C^{\mathcal{M}_i} = C^{\mathcal{I}} \cap W_i$ for all values of $g(C)$.

Observe that the function g determines which concepts are interpreted in the same way by corresponding interpretations ($g(C) = T$) or differently ($g(C) = F$), due to the differences between DL and $\mathcal{E}$-Connections semantics. $\mathcal{E}$-safety is a property of the input ontology that indicates which axioms are "dangerous" for the preservation of semantic compatibility in an $\mathcal{E}$-Connections based decomposition.

Note that $\mathcal{E}$-safety is a property of $\mathcal{SHOIN}$ ontologies. It is possible to provide a rationale for its definition independently from $\mathcal{E}$-Connections. Roughly, $\mathcal{E}$-safety ensures that certain kinds of General Concept Inclusion Axioms (GCIs) do not appear explicitly in the KB. The following theorem shows that, if an ontology is $\mathcal{E}$-safe, then those kinds of GCIs cannot be *entailed* by the KB.

Theorem 1 *Let $\mathcal{O}$ with vocabulary V and consistent be $\mathcal{E}$-safe, then there are no $\mathcal{SHOIN}$ concepts C,D in the vocabulary V s.t. $g(C) = F$, $g(D) = T$ and $\mathcal{O} \models (C \sqsubseteq D)$*

As a consequence of the theorem the following properties of $\mathcal{E}$-safety can be shown:

Corollary 1 *Let $\mathcal{O}$ with vocabulary V be consistent and $\mathcal{E}$-safe, then:*

1. *There is no concept C in the vocabulary of $\mathcal{O}$ with $g(C) = T$ s.t. $\mathcal{O} \models (\top \sqsubseteq C)$*
2. *There is no atomic concept $A \in V_C$ s.t. $\mathcal{O} \models (\top \sqsubseteq A)$*
3. *All the models of $\mathcal{O}$ do not share the same finite interpretation domain.*
4. *$\mathcal{O} \equiv \mathcal{O}'$, then $\mathcal{O}$ is $\mathcal{E}$-safe iff $\mathcal{O}'$ is $\mathcal{E}$-safe.*

The exact connection between structural and semantical compatibility is given by the following theorem:

Theorem 2 *If $\mathcal{O}$ is consistent and $\mathcal{E}$-safe and $\Sigma \sim \mathcal{O}$, then $\Sigma \approx \mathcal{O}$. If $\mathcal{O}$ is not $\mathcal{E}$-safe, then the only Σ s.t. $\Sigma \sim \mathcal{O}$ and $\Sigma \approx \mathcal{O}$ is $\Sigma = \{\mathcal{O}\}$*

The theorem shows how **P2)** can be reduced to **P1)**. In order to solve **P2)**, first decide $\mathcal{E}$-safety and then solve **P1)**.Obviously, $\mathcal{E}$-safety can be computed efficiently and therefore the reduction from **P2)** to **P1)** is polynomial.

Definition 9 ***(Merge)***
*Let Σ have n components. We say that the $\mathcal{E}$-Connection Υ with $(n-k+1)$ $(k \leq n)$ components is a **k-merge** of Σ if it is obtained by replacing any k components $\{\Sigma_1, ..., \Sigma_k\}$ of Σ by a single component containing the union of all the axioms in $\Sigma_1, ..., \Sigma_k$.*

It is easy to see that, given Σ with n-components, there are $C(n,k) = \binom{n}{k}$ different k-merges of Σ, where$\binom{n}{k}$ is the binomial coefficient with parameters n,k.

Lemma 3 *Let $\Sigma \sim \mathcal{O}$ and Υ a k-merge of Σ, then $\Upsilon \sim \mathcal{O}$. Moreover, if additionally $\Sigma \approx \mathcal{O}$, then also $\Upsilon \approx \mathcal{O}$.*

3 The Partitioning Algorithm

In this section, we describe our proposed algorithm for solving **P1)** and **P2)**, specified in Figure 1.

The algorithm accepts $\mathcal{O}$ as input and returns an $\mathcal{E}$-Connection $\Sigma = \{\Sigma_1, ..., \Sigma_n\}$. The algorithm consists of a succession of n *partitioning steps*. Each step involves *a pair* of KBs: the *original* KB, $\mathcal{O}$, from which entities and axioms are removed, and a *target* KB, Σ_i, generated from scratch in the ith step, to which these are added. In the process some roles in $\mathcal{O}$ will eventually become link properties in Σ.

In a given partitioning step, each concept, individual and link property (generated in a previous step) can be in one of two possible "states": either as entities in $\mathcal{O}$, or in Σ_i. A role, however, can be in one of four possible states: as a role in $\mathcal{O}$(State 1), as a link property from $\mathcal{O}$ to Σ_i (State 2), as a link property from Σ_i to $\mathcal{O}$ (State 3), and finally as an object property in Σ_i (State 4). Only the transitions $1 \rightarrow 2$, $1 \rightarrow 3$, $1 \rightarrow 4$ $2 \rightarrow 4$ and $3 \rightarrow 4$ are allowed.

The algorithm initially checks if $\mathcal{O}$ is $\mathcal{E}$-safe. If the check is negative, then it returns $\mathcal{O}$ as a result. Otherwise, the algorithm starts a partitioning step by creating a new component Σ_i and by forcing an initial state transition on an arbitrary entitiy (concept, role or individual) in $\mathcal{O}$

The initial transition will trigger new ones, due to the structural constraints imposed by $\mathcal{E}$-Connections. For example, if $(C \sqsubseteq D) \in \mathcal{O}$, and we move C, then D *must* be exported as well to Σ_i, since an axiom cannot relate complex classes in different components in an $\mathcal{E}$-Connection However, there is still a choice to make in certain situations that involve roles. For example, if $\exists R.C \in \mathcal{O}$ and we move C, two possible actions would be allowed: first, to make R a Link Property from $\mathcal{O}$ to Σ_i; second, to make R a role in Σ_i. Analogously, suppose that $P(a, b) \in \mathcal{O}$, and we move a; again, two choices would be admissible: either we make P a Link Property from Σ_i to $\mathcal{O}$, or we make it a role in Σ_i. In both examples, each choice would result in a syntactically valid $\mathcal{E}$-Connection. In order to obtain a maximal partitioning of $\mathcal{O}$ we will transform roles into link properties *whenever possible.*

The set *bounded*(P) represents the set of entities that are "forced" to end up in the same partition due to the semantics of the link property P. For example, imagine we have 3 components:$\Sigma_0, \Sigma_1, \Sigma_2$. Suppose we first create Σ_1 and then Σ_2. Suppose that there is a Link Property P, generated in the first step, from Σ_1 to Σ_0 and $\exists P.C, \exists P.B \in \Sigma_0$. Then, assume that C is moved from Σ_0 to Σ_2 in the second step. Obviously, P should now link Σ_1 and Σ_2, instead of Σ_1 and Σ_0, and hence B must be moved as well, since it is bounded to C by P.

Once all the state transitions in a partitioning step have been completed, the relevant axioms are moved. Each axiom in the input ontology is moved *only once*, i.e. whenever it is exported from $\mathcal{O}$ to a newly created component, it will never be put back into $\mathcal{O}$, nor moved to a different component.

As an example, let us consider again the KB:

$$\mathcal{O} = \{(C \sqsubseteq D \sqcup E); (C \sqsubseteq \neg D); \ (B \sqsubseteq \exists P.A); \ (A \sqsubseteq F)\}$$

The KB is clearly $\mathcal{E}$-safe. Suppose the algorithm first selects E, which changes its state ($\mathit{State}(E) \leftarrow 2$). A new (empty) component Σ_1 is created and added to the $\mathcal{E}$-Connection. During the execution of the state machine it is detected that $(D \sqcup E) \in \Sigma_0$ with $\mathit{State}(E) = 2$, then both $(D \sqcup E)$ and D update its state to 2. Then, $\mathit{State}(C) \leftarrow 2$ since it is a subclass of $D \sqcup E$ and $\mathit{State}(\neg D) \leftarrow 2$, since $\neg D$ is a subclass of C. Since no other transition occurs, the axioms $(C \sqsubseteq D \sqcup E)$ and $(C \sqsubseteq \neg D)$ are removed from Σ_0 and added to Σ_1. At the end of this first step, the state of the $\mathcal{E}$-Connection is as follows:

$$\begin{gathered} \Sigma_0 = \{(B \sqsubseteq \exists P.A); \ (A \sqsubseteq F)\} \\ \Sigma_1 = \{(C \sqsubseteq D \sqcup E); (C \sqsubseteq \neg D)\} \end{gathered}$$

The component Σ_1 won't be examinated anymore. In the next partitioning step, the component Σ_2 is created. Suppose that the atomic concept A is selected ($\mathit{State}(A) \leftarrow 2$. The algorithm detects that $\exists P.A \in \Sigma_0$. Since $\mathit{State}(P) = 1$ and $\mathit{State}(A) = 2$, the state of P is updated ($\mathit{State}(P) \leftarrow 2$). Also, since $(A \sqsubseteq F)$, $\mathit{State}(F) \leftarrow 2$. Since no more state transitions occur, the axiom $(A \sqsubseteq F)$ is exported to Σ_2 and P becomes a link property pointing to Σ_2.

$$\begin{gathered} \Sigma_0 = \{(B \sqsubseteq \exists P.A)\} \\ \Sigma_1 = \{(C \sqsubseteq D \sqcup E); (C \sqsubseteq \neg D)\} \\ \Sigma_2 = \{(A \sqsubseteq F)\} \end{gathered}$$

In the next step, Σ_3 is created. At this point, B is an atomic concept with $\mathit{State}(B) = 1$, $\exists P.A$ a concept with $\mathit{State}(\exists P.A) = 1$ and P is a link property with $\mathit{State}(P) = 1$ in Σ_0. Suppose that the algorithm selects P and updates its state ($\mathit{State}(P) \leftarrow 2$). Then, the state of $\exists P.A$ changes to $\mathit{State}(\exists P.A) \leftarrow 2$, and also $\mathit{State}(B) \leftarrow 2$. At the end of the step, the axiom $(B \sqsubseteq \exists P.A)$ is removed from Σ_0 and added to Σ_3. Since Σ_0 is empty, the algorithm removes Σ_0 from the $\mathcal{E}$-Connection and returns $\Sigma = \{\Sigma_1, \Sigma_2, \Sigma_3\}$, with:

$$\begin{gathered} \Sigma_1 = \{(C \sqsubseteq D \sqcup E); (C \sqsubseteq \neg D)\} \\ \Sigma_2 = \{(A \sqsubseteq F)\} \\ \Sigma_3 = \{(B \sqsubseteq \exists P.A)\} \end{gathered}$$

Theorem 3 *The algorithm Partition(O) is worst-case quadratic in the size of the KB. The output Σ is a solution for **P2)** with input $\mathcal{O}$ and a solution for **P1)** if the safety check is omitted.*

KB	Atomic Concepts	Complex Descriptions	Roles	Indiv.	Number Components/ Leaf Comp.	Atomic Concepts Largest	Atomic Concepts Smallest	Links in Σ	Time(s)
OWL-S	51	49	54	9	17/7	21	1	37	0.291
NASA	1537	232	102	194	43/36	1100	1	22	2.8
GALEN	2749	2011	413	0	2/1	2748	1	0	11
NCI	27652	4000	71	0	17/9	7663	34	55	45

Table 2: Some Partitioned Ontologies

4 Implementation and Evaluation

We implemented partitioning algorithm on top of Manchester's OWL-API, which we have extended to provide support for $\mathcal{E}$-Connections. The UI in SWOOP [3] for browsing $\mathcal{E}$-Connections has been extended to support automated partitioning. We have applied our algorithm to a set of OWL ontologies available on the Web and stored the partitions in an online repository [4]. Table 2 summarizes the results obtained for some relevant cases.

GALEN and NCI are both large, carefully designed ontologies dealing with the biomedical domain, but which follow very different modeling paradigms. In GALEN, most of the knowledge is ultimately dependending on a common *top* class (*TopCategory*) and *top* role. Hence, although it is possible to identify intuitively several disjoint sub-domains in GALEN, the ontology follows a very "monolithic" design pattern, which prevents a good partitioning. However, NCI follows a more modular design pattern, since the knowledge has been structured around separate top entities, each of which delimites a different sub-domain.

The partitioning of NCI reproduces each of the intuitive sub-domains in a different component. The link properties in the resulting $\mathcal{E}$-Connection provide useful information on the original ontology. In the partitioning of NCI the component dealing with genes is the one that contains the largest number of "outgoing" link properties and also the one that "uses" information from the largest number of components. This implies that genes are central to the ontology. Other components, like the one dealing with anatomical structures, are "leaf components" in the sense that they have only "incoming" link properties. These components do not use information from any other components in the $\mathcal{E}$-Connection, are written in plain OWL and can be directly reused.

The OWL-S ontologies describe Web services, whereas NASA's SWEET-JPL ontologies model several interrelated domains of interest for the space industry. Within both sets of KBs, each ontology seems to model a well-defined sub-domain. The domain as a whole is modeled in both cases by using *owl:imports*. We have collapsed each set of ontologies applied the partitioning algorithm to the result. In the case of OWL-S the partitions correspond closely to the original

---

[3] http://www.mindswap.org/2004/SWOOP

[4] http://www.mindswap.org/2004/multipleOnt/FactoredOntologies

sub-domains: no information from different KBs in the original set comes together, while some of the original sub-domains appear further decomposed in a reasonable way. However, in the case of NASA-JPL, the partitioning shows important flaws in the way the knowledge was originally structured. The physically distinct ontologies do not correspond to semantically distinct ones.

5 Related Work and Conclusion

In this paper, we have analyzed the problem of partitioning expressive DL KBs using $\mathcal{E}$-Connections. We have argued that partitioning can be a useful tool for Semantic Web applications. We have provided an efficient solution for some interesting partitioning problems and shown some empirical results.

Partitioning OWL ontologies for modeling purposes has been recently addressed in [6]. We consider that the main limitation of that approach is the lack of a suitable formalism for representing the output of the problem. The results are represented as a visualization of the different kinds of information contained in the input ontology. No shareable partitions are ever obtained.

[5] explores partitioning FOL theories to improve theorem prover performance. We believe that reasoning $\mathcal{E}$-Connections not only does not affect existing optimizations in DL reasoners, but also suggest new ones. In particular it may help to detect obvious non-subsumptions, to alleviate the effect of non-absorbable GCIs (since they would be added only to certain nodes in the tableau expansion and not to all of them), and to enhance ABox reasoning, since the individuals in the original KB are split into different components. We plan to confirm that experimentally using our reasoner Pellet, which already provides $\mathcal{E}$-Connections support and compare the results with [5] for the DL case.

6 Proofs of Lemmas and Theorems

Proof for Lemma 1

1. *It's straightforward from the definition of semantic compatibility*

2. *Suppose that $\Sigma' \equiv \Sigma$, then they have the same models. From the definition of semantic compatibility, it is easy to see that the property holds.*

3. *We prove each of the directions:*

   - *($\rightarrow$) Let $\mathcal{O}$ be consistent. Since $\Sigma \approx \mathcal{O}$, $\mathcal{O}$ is partitionable for V_n (the vocabulary of Σ). Therefore, there exists a partitioned interpretation $\mathcal{I}(V_n)$ s.t. $\mathcal{I}_(V_n) \models \mathcal{O}$. Let $\mathcal{M}$ be the interpretation for Σ s.t. $\mathcal{I}(V_n) \leftrightarrow \mathcal{M}$. Since $\Sigma \approx \mathcal{O}$, we have that $\mathcal{M} \models \Sigma$ and hence Σ is consistent*

- *($\leftarrow$) Let Σ be consistent, then there exists an interpretation $\mathcal{M} \models \Sigma$. Let $\mathcal{I}(V_n)$ be the partitioned interpretation s.t. $\mathcal{I}(V_n) \leftrightarrow \mathcal{M}$. Since $\Sigma \approx \mathcal{O}$, $\mathcal{I}(V_n) \models \mathcal{O}$ and hence $\mathcal{O}$ is consistent.*

4.
    - ($\rightarrow$) *Let $\mathcal{O}_1 \equiv \mathcal{O}_2$ and suppose that there is a Σ that belongs to one of $\Pi(\mathcal{O}_1), \Pi(\mathcal{O}_2)$, but not to the other (say $\Sigma \in \Pi(\mathcal{O}_1)$ and $\Sigma \notin \Pi(\mathcal{O}_2)$). Since $\Sigma \approx \mathcal{O}_1$, then $\mathcal{O}_1$ is partitionable for V_n (the vocabulary of Σ), and therefore there is a partitioned interpretation $\mathcal{I}(V_n)$ s.t. $\mathcal{I}(V_n) \models \mathcal{O}_1$. However, since $\mathcal{O}_1$ and $\mathcal{O}_2$ have the same models, then $\mathcal{I}(V_n) \models \mathcal{O}_2$, and therefore $\mathcal{O}_2$ is partitionable for V_n. Finally, since $\mathcal{O}_1 \equiv \mathcal{O}_2$, the set of partitioned models for V_n is the same for both ontologies, we can establish exactly the same bijection '$\leftrightarrow$' between partitioned models for V_n and models of Σ. Since $\Sigma \approx \mathcal{O}_1$ if $\mathcal{I}(V_n) \leftrightarrow \mathcal{M}$, then $\mathcal{I}(V_n) \models \mathcal{O}_1 \leftrightarrow \mathcal{M} \models \Sigma$. Therefore this property also holds for $\mathcal{O}_2$. Hence, $\Sigma \approx \mathcal{O}_2$*
    - *Let $\Pi(\mathcal{O}_1) = \Pi(\mathcal{O}_2)$. Using 1), we have that $\Sigma = \{\mathcal{O}_2\} \in \Pi(\mathcal{O}_2)$. By the assumption, $\Sigma \in \Pi(\mathcal{O}_1)$ and hence $\Sigma \approx \mathcal{O}_1$. Therefore, $\mathcal{O}_1$ and Σ have the same models and thus $\mathcal{O}_1 \equiv \mathcal{O}_2$*

5.
    - ($\rightarrow$) *Suppose $\mathcal{O}$ entails the axiom $\mathcal{A}$ of the form $A \sqsubseteq B$, then every model of $\mathcal{O}$ satisfies $\mathcal{A}$. In particular all the partitioned models with vocabulary V_n satisfy it. Let $\mathcal{I}(V_n) \models \mathcal{A}$, then $A^{\mathcal{I}_n} \subseteq B^{\mathcal{I}_n}$. Let $\mathcal{M} \leftrightarrow \mathcal{I}(V_n)$, then by definition $\mathcal{M}$ and $\mathcal{I}(V_n)$ interpret the atomic concepts in the same way and hence $A^{\mathcal{I}} = A^{\mathcal{I}_n}; B^{\mathcal{I}} = B^{\mathcal{I}_n}$ and hence $\mathcal{M} \models \mathcal{A}$. By compatibility, $\mathcal{M} \models \Sigma$ and since '$\leftrightarrow$' is a bijection between partitioned models of $\mathcal{O}$ with vocabulary V_n and (all) the models of Σ, we have that $\Sigma \models \mathcal{A}$. An analogous argument can be used for showing the remaining cases*

Proof for Lemma 2 *First, note that if P is an ij-property (belonging to ϵ_{ij} or the inverse of a property in ϵ_{ji}), then it is easy to see from the definition of corresponding interpretations that $P^{\mathcal{I}} = P^{\mathcal{M}_{ij}}$.*

Then, the proof goes by induction on the structure of the concept.

- *If C of the form $A \in V_{C_i}$, then by definition of $\mathcal{I}$, $A^{\mathcal{I}} = A^{\mathcal{M}_i}$. By definition of g, $g(C) = T$*
- *If C is $\top$, then $(\top)^{\mathcal{I}} = W = \bigcup_{k=1,\ldots,n} W_k = W_i \bigcup_{j \neq i} W_j = (\top)^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$. In this case $g(C) = F$*
- *If C of the form $\{a\}$, for $a \in V_{I_i}$, then $C^{\mathcal{I}} = \{a^{\mathcal{I}}\}$. By definition, $a^{\mathcal{I}} = a^{\mathcal{M}_i}$, and hence $C^{\mathcal{I}} = C^{\mathcal{M}_i}$. By definition of g, $g(C) = T$*

- *If C is of the form $\neg D$, then, $C^{\mathcal{I}} = W - D^{\mathcal{I}}$*
  - *If $D^{\mathcal{I}} = D^{\mathcal{M}_i}$, then $C^{\mathcal{I}} = W - D^{\mathcal{M}_i}$, which implies that $C^{\mathcal{I}} = (W_i \bigcup_{j \neq i} W_j) - D^{\mathcal{M}_i}$. By induction and def. of g, $g(D) = T$ in this case. Since $D^{\mathcal{M}_i} \subseteq W_i$ and $W_i \cap W_j = \emptyset$ for $j \neq i$, then $C^{\mathcal{I}} = (W_i - D^{\mathcal{M}_i}) \bigcup_{j \neq i} W_j = C^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$. Therefore, $g(C) = F$*
  - *If $D^{\mathcal{I}} = D^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$, then $C^{\mathcal{I}} = W - (D^{\mathcal{M}_i} \bigcup_{j \neq i} W_j) = W_i - D^{\mathcal{M}_i}$ and $g(D) = F$. Therefore, $C^{\mathcal{I}} = C^{\mathcal{M}_i}$ and hence $g(C) = T$.*

- *If C of the form $(D \sqcap E)$, then $(D \sqcap E)^{\mathcal{I}} = D^{\mathcal{I}} \cap E^{\mathcal{I}}$. By induction, $D^{\mathcal{I}}$ is either $D^{\mathcal{M}_i}$ in which case $g(D) = T$ or $D^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$, in which case $g(D) = F$ and anagolously, $E^{\mathcal{I}}$ is either $E^{\mathcal{M}_i}$ ($g(E) = T$) or $E^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$ ($g(E) = F$). Since $C^{\mathcal{M}_i}$ and $E^{\mathcal{M}_i}$ are disjoint with $\bigcup_{i \neq j} W_j$, we have for every possible combination that $D^{\mathcal{I}} \cap E^{\mathcal{I}}$ is either equal to $D^{\mathcal{M}_i} \cap E^{\mathcal{M}_i}$ or to $(D^{\mathcal{M}_i} \cap E^{\mathcal{M}_i}) \bigcup_{j \neq i} W_j$. Note that the latter case only appears if $D^{\mathcal{I}} = D^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$, and also $E^{\mathcal{I}} = E^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$. Therefore, if $g(D) = F$ and $g(E) = F$, then $g(C) = F$ and otherwise $g(C) = T$*

- *Let $C = D \sqcup E$. Then $(D \sqcup E)^{\mathcal{I}} = D^{\mathcal{I}} \cup E^{\mathcal{I}}$. Again, by induction, $D^{\mathcal{I}}$ is either $D^{\mathcal{M}_i}$ or $D^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$ and anagolously, $E^{\mathcal{I}}$ is either $E^{\mathcal{M}_i}$ or $E^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$. Again, for every possible combination $D^{\mathcal{I}} \cup E^{\mathcal{I}}$ is either equal to $D^{\mathcal{M}_i} \cup E^{\mathcal{M}_i}$ or to $(D^{\mathcal{M}_i} \cup E^{\mathcal{M}_i}) \bigcup_{j \neq i} W_j$. Note that the former case only appears if $D^{\mathcal{I}} = D^{\mathcal{M}_i}$ and also $E^{\mathcal{I}} = E^{\mathcal{M}_i}$. Therefore, if $g(D) = T$ and $g(E) = T$, then $g(C) = T$ and otherwise $g(C) = F$*

- *If C of the form $\exists P.Z$, with P an ij-property, then $(\exists P.Z)^{\mathcal{I}} = \{x \in W | \exists y \in W, (x, y) \in P^{\mathcal{I}}, y \in Z^{\mathcal{I}}\}$. Note that $P^{\mathcal{I}} = P^{\mathcal{M}_{ij}}$, by definition of $\mathcal{I}$:*
  - *Suppose $Z^{\mathcal{I}} = Z^{\mathcal{M}_j}$. In this case, it is easy to see that $(\exists P.Z)^{\mathcal{I}} = (\exists P.Z)^{\mathcal{M}_j}$*
  - *Suppose $Z^{\mathcal{I}} = Z^{\mathcal{M}_i} \bigcup_{i \neq j} W_j$. Since $P^{\mathcal{I}} = P^{\mathcal{M}_{ij}}$, there is no element $y \in W$ s.t. $(x, y) \in P^{\mathcal{I}}$ and $y \in \bigcup_{i \neq j} W_j$, since y must be in W_i. Therefore $(\exists P.Z)^{\mathcal{I}} = (\exists P.Z)^{\mathcal{M}_i}$. Since in both cases $(\exists P.Z)^{\mathcal{I}} = (\exists P.Z)^{\mathcal{M}_i}$, by induction we have that $g(C) = T$*

- *If C of the form $\forall P.Z$, with P an ij-property, then $(\forall P.Z)^{\mathcal{I}} = \{x \in W | (x, y) \in P^{\mathcal{I}} \rightarrow y \in Z^{\mathcal{I}}\}$. Again, using that $P^{\mathcal{I}} = P^{\mathcal{M}_{ij}}$ and that $W_i \cap W_j = \emptyset$ for $i \neq j$ it can be seen that $(\forall P.Z)^{\mathcal{I}} = (\forall P.Z)^{\mathcal{M}_i} \cup_{j \neq i} W_j$ always holds. Since in both cases $(\exists P.Z)^{\mathcal{I}} = (\exists P.Z)^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$, by induction we have that $g(C) = F$*

- *If C of the form $\geq nP$, then $(\geq nP)^{\mathcal{I}} = \{x \in W | card(\{y | (x, y) \in P^{\mathcal{I}}\}) \geq n$. Since $P^{\mathcal{I}} = P^{\mathcal{M}_{ij}}$, $(\geq nP)^{\mathcal{I}} = (\geq nP)^{\mathcal{M}_{ij}}$. By induction, $g(C) = T$*

- *If* C *of the form* $\leq nP$, P *an ij-property, then* $(\leq nP)^{\mathcal{I}} = \{x \in W | card(\{y | (x, y) \in P^{\mathcal{I}}\}) \leq n$. *Since* $P^{\mathcal{I}} = P^{\mathcal{M}_{ij}}$, *the set of elements in* W *that have at most* n *elements in* W *connected to it by* $P^{\mathcal{I}}$ *is the union of those that are in* W_i *and those in* $\bigcup_{j \neq i} W_j$. *The former ones are precisely those in* $(\leq nP)^{\mathcal{M}_i}$ *and the latter ones are precisely* $\bigcup_{j \neq i} W_j$, *since no element in* W_j *is related through* P *to an element of the domain. Hence,* $(\leq nP)^{\mathcal{I}} = (\leq nP)^{\mathcal{M}_i} \bigcup_{j \neq i} W_j$. *By induction,* $g(C) = F$

Proof for Theorem 1 *We need to prove that if* $\mathcal{O}$ *is* $\mathcal{E}$*-safe, there is no axiom of the form* $C \sqsubseteq D$ *with* $g(C) = F$ *and* $g(D) = T$ *s.t.* $\mathcal{O} \models (C \sqsubseteq D)$. *Let* $\mathcal{I} = (W, .^{\mathcal{I}})$ *be an interpretatain s.t.* $\mathcal{I} \models \mathcal{O}$ *and* $\mathcal{I} \models (C \sqsubseteq D)$. *We show that it is possible to obtain from* $\mathcal{I}$ *an interpretation* $\mathcal{J} = (W', .^{\mathcal{J}})$ *s.t.* $\mathcal{J} \models \mathcal{O}$, *but s.t. it does not satisfy the axiom* $C \sqsubseteq D$.

Let us define $\mathcal{J}$ *as follows:*

- $W' = W \cup \{x\}$, *where* $x \notin W$
- $A^{\mathcal{J}} = A^{\mathcal{I}}$, *for every atomic concept* A *in* $\mathcal{O}$
- $R^{\mathcal{J}} = R^{\mathcal{I}}$, *for every atomic role* R *in* $\mathcal{O}$
- $a^{\mathcal{J}} = a^{\mathcal{I}}$, *for every individual* a *in* $\mathcal{O}$

First note that for every $\mathcal{SHOIQ}$ *role* R *(an atomic role or its inverse) we have that* $R^{\mathcal{J}} = R^{\mathcal{I}}$.

We show that $\mathcal{J} \models \mathcal{O}$. *For such a purpose we will prove that for every concept* C *in the vocabulary of* $\mathcal{O}$ *s.t.* $g(C) = T$ *it holds that* $C^{\mathcal{J}} = C^{\mathcal{I}}$, *whereas if* $g(C) = F$, *then* $C^{\mathcal{J}} = C^{\mathcal{I}} \cup \{x\}$. *We proceed by induction:*

- *If* C *is an atomic concept, then* $g(C) = T$ *and by definition* $C^{\mathcal{J}} = C^{\mathcal{I}}$
- *If* C *is of the form* $\{a\}$, *then* $g(C) = T$ *and, since* $a^{\mathcal{J}} = a^{\mathcal{I}}$, *it follows that* $C^{\mathcal{J}} = C^{\mathcal{I}}$
- *If* C *is the top concept* $\top$, *then* $g(C) = F$ *and, since* $(\top)^{\mathcal{J}} = W'$, $(\top)^{\mathcal{I}} = W$ *and* $W' = W \cup \{x\}$, *we have that* $C^{\mathcal{J}} = C^{\mathcal{I}} \cup \{x\}$
- *Let* C *be of the form* $\neg D$*:*
    - *If* $g(D) = T$, *then* $g(C) = F$. *By induction,* $D^{\mathcal{J}} = D^{\mathcal{I}}$. *Since* $C^{\mathcal{J}} = W' - D^{\mathcal{J}}$, $W' = W \cup \{x\}$ *and* $D^{\mathcal{J}} = D^{\mathcal{I}}$, *we have that* $C^{\mathcal{J}} = C^{\mathcal{I}} \cup \{x\}$

  - If $g(D) = F$, then $g(C) = T$. By induction, $D^{\mathcal{J}} = D^{\mathcal{I}} \cup \{x\}$. Since $C^{\mathcal{J}} = W' - D^{\mathcal{J}}$, $W' = W \cup \{x\}$ and $D^{\mathcal{J}} = D^{\mathcal{I}} \cup \{x\}$, we have that $C^{\mathcal{J}} = C^{\mathcal{I}}$

- Let C be of the form $D \sqcap E$. If $g(D) = T$, then $D^{\mathcal{J}} = D^{\mathcal{I}}$, whereas if $g(D) = F$, then $D^{\mathcal{J}} = D^{\mathcal{I}} \cup \{x\}$. Analogously, if $g(E) = T$, $E^{\mathcal{J}} = E^{\mathcal{I}}$ and if $g(E) = F$, then $E^{\mathcal{J}} = E^{\mathcal{I}} \cup \{x\}$. We have the following cases:
  - If $g(D) = F$ and $g(E) = F$, then $g(C) = F$. We have that $C^{\mathcal{J}} = D^{\mathcal{J}} \cap E^{\mathcal{J}} = (D^{\mathcal{I}} \cup \{x\}) \cap (E^{\mathcal{I}} \cup \{x\}) = (D^{\mathcal{I}} \cap E^{\mathcal{I}}) \cup \{x\} = C^{\mathcal{I}} \cup \{x\}$
  - If any of $g(D), g(E)$ is True, then $g(C) = T$. It is easy to see that in any of those cases $C^{\mathcal{J}} = D^{\mathcal{I}} \cap E^{\mathcal{I}} = C^{\mathcal{I}}$, since $x \notin D^{\mathcal{I}}$ and $x \notin E^{\mathcal{I}}$.

- Let C be of the form $D \sqcup E$. The argument is analogous to the above case.

- Let C be of the form $\exists R.D$ In this case $g(C) = T$. Note that, since $R^{\mathcal{J}} = R^{\mathcal{I}}$, both if $g(D) = T$ (in which case $D^{\mathcal{J}} = D^{\mathcal{I}}$), or if $g(D) = F$ (in which case $D^{\mathcal{J}} = D^{\mathcal{I}} \cup \{x\}$), $C^{\mathcal{J}} = C^{\mathcal{I}}$.

- Let C be of the form $\geq nR$. In this case $g(C) = T$. It is easy to see that, since $R^{\mathcal{J}} = R^{\mathcal{I}}$, also $C^{\mathcal{J}} = C^{\mathcal{I}}$

- Let C be of the form $\leq nR$. In this case $g(C) = F$. We have that $R^{\mathcal{J}} = R^{\mathcal{I}}$ and $W' = W \cup \{x\}$. $(\leq nR)^{\mathcal{J}}$ contains the elements of W' that are related by $R^{\mathcal{J}}$ with at least 1 element of W' and at most with n and by those that are not connected by $R^{\mathcal{J}}$ with any element of W'. It is easy to see that all the elements of $(\leq nR)^{\mathcal{I}}$ are contained in $(\leq nR)^{\mathcal{J}}$ and that the only element that belongs to $(\leq nR)^{\mathcal{J}}$ and does not belong to $(\leq nR)^{\mathcal{J}}$ is precisely $\{x\}$

- Let C be of the form $\forall R.D$. The argument is analogous to the above case.

Now we prove that $\mathcal{J}$ is a model of $\mathcal{O}$.

Since $\mathcal{O}$ is $\mathcal{E}$-safe, each of the concept inclusion axioms $C \sqsubseteq D$ in $\mathcal{O}$ are of either of the following forms:

1. $g(C) = T$ and $g(D) = T$. In this case $C^{\mathcal{J}} = C^{\mathcal{I}}$ and $D^{\mathcal{J}} = D^{\mathcal{I}}$. Since $\mathcal{I} \models \mathcal{O}$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, and hence $C^{\mathcal{J}} \subseteq D^{\mathcal{J}}$. Therefore, $\mathcal{J}$ satisfies the axiom

2. $g(C) = T$ and $g(D) = F$. In this case $C^{\mathcal{J}} = C^{\mathcal{I}}$ and $D^{\mathcal{J}} = D^{\mathcal{I}} \cup \{x\}$. Since $\mathcal{I} \models \mathcal{O}$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, and hence $C^{\mathcal{J}} \subseteq D^{\mathcal{J}}$. Therefore, again $\mathcal{J}$ satisfies the axiom

3. $g(C) = F$ and $g(D) = F$. In this case $C^{\mathcal{J}} = C^{\mathcal{I}} \cup \{x\}$ and $D^{\mathcal{J}} = D^{\mathcal{I}} \cup \{x\}$. Since $\mathcal{I} \models \mathcal{O}$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, and hence $C^{\mathcal{I}} \cup \{x\} \subseteq D^{\mathcal{I}} \cup \{x\}$. Therefore, $C^{\mathcal{J}} \subseteq D^{\mathcal{J}}$ and again $\mathcal{J}$ satisfies the axiom

For axioms of the form $R \sqsubseteq S$, for R, S roles, since $R^{\mathcal{I}} \sqsubseteq S^{\mathcal{I}}$ and $R^{\mathcal{I}} = R^{\mathcal{J}}$ and $S^{\mathcal{I}} = S^{\mathcal{J}}$ for any $\mathcal{SHOIN}$ role (atomic role or inverse of an atomic role) , we have that $\mathcal{J}$ satisfies the axiom. For axioms of the form $C(a)$, since $a^{\mathcal{I}} \in C^{\mathcal{I}}$, and $a^{\mathcal{I}} = a^{\mathcal{J}}$ and $C^{\mathcal{J}}$ is either $C^{\mathcal{I}}$ (if $g(C) = T$) or $C^{\mathcal{I}} \cup \{x\}$, if $g(C) = F$, we have that in both cases $a^{\mathcal{J}} \in C^{\mathcal{J}}$. Similar arguments can be used to verify that if $\mathcal{I}$ satisfies an axiom of the form $R(a, b)$, then $\mathcal{J}$ satisfies it.

Finally, we show that if $\mathcal{I}$ satisfies an axiom of the form $(C \sqsubseteq D)$ with $g(C) = F$ and $g(D) = T$ then $\mathcal{J}$ does not satisfy it. Since $g(C) = F$, then $C^{\mathcal{J}} = C^{\mathcal{I}} \cup \{x\}$ and since $g(D) = T$ we have that $D^{\mathcal{J}} = D^{\mathcal{I}}$. Now we have that $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. However $\{x\} \notin D^{\mathcal{I}}$ by construction and hence $C^{\mathcal{J}}$ is not a subset of $D^{\mathcal{J}}$. This implies that $\mathcal{O}$ cannot entail such an axiom, since there are models of $\mathcal{O}$ that do not satisfy it.

Proof for Corollary 1

1. *Direct consequence of the theorem*
2. *Direct consequence of the theorem*
3. *All the models of $\mathcal{O}$ share the same finite interpretation domain iff $\mathcal{O} \models \top \sqsubseteq \{a_1, ..., a_n\}$ for some n, with $a_i \in V_I$. But this cannot happen by the theorem, since $g(\top) = F$ and $g(\{a_1, ..., a_n\} = T$*
4. *Suppose $\mathcal{O} \equiv \mathcal{O}'$, $\mathcal{O}$ $\mathcal{E}$-safe and $\mathcal{O}'$ not $\mathcal{E}$-safe. Since $\mathcal{O}'$ is not $\mathcal{E}$-safe, there is a pair of concepts C, D in the vocabulary of $\mathcal{O}'$ (and $\mathcal{O}'$) s.t. $g(C) = F$, $g(D) = T$ and $\mathcal{O}' \models (C \sqsubseteq D)$. But, since $\mathcal{O} \equiv \mathcal{O}'$ then it also holds that $\mathcal{O} \models (C \sqsubseteq D)$ and hence $\mathcal{O}$ cannot be $\mathcal{E}$-safe, which is a contradiction*

Proof for Theorem 2 *1)*

1.1) To Prove: *Σ with vocabulary V_n, $\Sigma \sim \mathcal{O}$ and $\mathcal{O}$ consistent and $\mathcal{E}$-safe, then $\mathcal{O}$ partitionable for V_n*

Since $\mathcal{O}$ is consistent, there exists an interpretation $\mathcal{I} = (W, .^{\mathcal{I}})$ s.t. $\mathcal{I} \models \mathcal{O}$. We show that we can obtain from $\mathcal{I}$ a partitioned interpretation $\mathcal{I}(V_n) = (W', .^{\mathcal{I}'})$ for V_n s.t. $\mathcal{I}(V_n) \models \mathcal{O}$. First, we redefine the domain W of $\mathcal{I}$ by adding n elements $\{x_1, ..., x_n\}$ that we do not add to the interpretation of any atomic concept. Note that this is always possible, since $\mathcal{E}$-safety ensures that the interpretation domain for every model of $\mathcal{O}$ needs not be finite and also that there is no atomic concept that is equivalent to $\top$. With this trivial domain expansion, $\mathcal{I}$ is still a model of $\mathcal{O}$.

We define the domain W' of $\mathcal{I}(V_n)$ using the following steps:

1. *$W' \leftarrow W$*
2. *For each $a \in W$, if a appears at the same time in any two of the following:*
   - *in $A^{\mathcal{I}}$ for $A \in V_{C_i}$, $B^{\mathcal{I}}$ for $B \in V_{C_j}$, $(a, x) \in P^{\mathcal{I}}$ with $P \in \epsilon_{ji}$, $(x, a) \in P^{\mathcal{I}}$ with $P \in \epsilon_{ij}$*

then do $W' \leftrightarrow W - \{a\}$ *and remove from* $P^{\mathcal{I}}$ *(for any* $P \in V_R$*) any pair that involves* a*. Clearly, at the end of this process* $W' \subseteq W$*. Note that this process will never suppress from the domain any element* $o^{\mathcal{I}}$ *for* $o \in V_{\mathcal{I}}$*, because of the compatibility relation* $\Sigma \sim \mathcal{O}$ *and the syntax of* $\mathcal{E}$*-Connections.*

3. *Define, for* $i, j = 1, ..., n$ *the set* W_i *as the* ***union*** *of the following sets:*
    (a) *The sets* $A^{\mathcal{I}} \cap W'$ *for every* $A \in V_{C_i}$
    (b) *The sets* $\{x \in W' | \exists y \in W', (x, y) \in P^{\mathcal{I}}\}$ *for* $P \in \epsilon_{ij}$
    (c) *The sets* $\{a^{\mathcal{I}}\}$ *for each* $a \in V_{I_i}$
    (d) *The sets* $\{x \in W' | \exists y \in W', (y, x) \in P^{\mathcal{I}}\}$ *for* $P \in \epsilon_{ji}$

    If W_i *is empty after adding those sets, do* $W_i \leftarrow \{x_i\}$

By construction of W' *and its sub-domains, the following properties hold:*

1. $W_i \cap W_j = \emptyset$ *for* $j \neq i$ *and* $W' = \bigcup_i W_i$
2. $A^{\mathcal{I}} \cap W' = A^{\mathcal{I}} \cap W_i$ *for each* $A \in V_{C_i}$
3. $P^{\mathcal{I}} \cap (W' \times W') = P^{\mathcal{I}} \cap (W_i \times W_j)$ *for each* $P \in \epsilon_{ij}$
4. $a^{\mathcal{I}'} = a^{\mathcal{I}} \in W_i$ *for each* $a \in V_{I_i}$

Therefore, if the define the interpretation function $.^{\mathcal{I}'}$ *as follows:*

- $A^{\mathcal{I}'} = A^{\mathcal{I}} \cap W_i$ *for each* $A \in V_{C_i}$
- $P^{\mathcal{I}'} = P^{\mathcal{I}} \cap (W_i \times W_j)$ *for each* $P \in \epsilon_{ij}$
- $a^{\mathcal{I}'} = a^{\mathcal{I}}$ *for each* $a \in V_{I_i}$

The interpretation $\mathcal{I}(V_n)$ *is a partitioned interpretation for* $\mathcal{O}$ *with vocabulary* V_n*. We need to show that* $\mathcal{I}(V_n) \models \mathcal{O}$*. The first step is to find the relationship between* $C^{\mathcal{I}}$ *and* $C^{\mathcal{I}'}$ *for every concept* $C \in \mathcal{O}$*. First, note that since* $\Sigma \sim \mathcal{O}$*, every* $C \in \mathcal{O}$ *must be an i-concept in* Σ *for a (single)* $i \in \{1, ..., n\}$*. We prove the following:*

$$C^{\mathcal{I}'} = C^{\mathcal{I}} \cap W' \text{ for every concept}$$

We proceed by induction:

- *If* C *of the form* $\top$*, then* $\top^{\mathcal{I}'} = W' = (W \cap W')$
- *If* C *of the form* $A \in V_{C_i}$*, then* $A^{\mathcal{I}'} = A^{\mathcal{I}} \cap W_i = A^{\mathcal{I}} \cap W'$

- *If* C *of the form* $\{a\}$ *for* $a \in V_{I_i}$*, then* $a^{\mathcal{I}'} = a^{\mathcal{I}} \in W_i$ *and hence* $C^{\mathcal{I}'} = C^{\mathcal{I}} = C^{\mathcal{I}} \cap W'$

- *If* C *of the form* $D \sqcap E$*, then* $C^{\mathcal{I}'} = D^{\mathcal{I}'} \cap E^{\mathcal{I}'}$*. By induction,* $D^{\mathcal{I}'} = D^{\mathcal{I}} \cap W'$ *and* $E^{\mathcal{I}'} = E^{\mathcal{I}} \cap W'$*, hence* $C^{\mathcal{I}} = D^{\mathcal{I}} \cap E^{\mathcal{I}} \cap W' = (D \sqcap E)^{\mathcal{I}} \cap W'$

- *If* C *of the form* $D \sqcup E$*, then* $C^{\mathcal{I}'} = D^{\mathcal{I}'} \cup E^{\mathcal{I}'}$*. By induction,* $D^{\mathcal{I}'} = D^{\mathcal{I}} \cup W'$ *and* $E^{\mathcal{I}'} = E^{\mathcal{I}} \cup W'$*, hence* $C^{\mathcal{I}} = D^{\mathcal{I}} \cup E^{\mathcal{I}} \cup W' = (D \sqcup E)^{\mathcal{I}} \cup W'$

- *If* C *of the form* $\neg D$*, then* $C^{\mathcal{I}'} = W' - D^{\mathcal{I}'}$*. By induction,* $D^{\mathcal{I}'} = D^{\mathcal{I}} \cap W'$ *and hence* $C^{\mathcal{I}'} = W' - (D^{\mathcal{I}} \cap W') = (W - D^{\mathcal{I}}) \cap W' = C^{\mathcal{I}} \cap W'$

- *If* C *of the form* $\exists P.D$*, then* $(\exists P.D)^{\mathcal{I}'}$ *is the set of elements in* W' *that are related through* $P^{\mathcal{I}'}$ *with at least an element of* $D^{\mathcal{I}'}$*. Since* $P^{\mathcal{I}}$ *and* $P^{\mathcal{I}'}$ *coincide in* $W' \times W'$*, and by induction we have* $D^{\mathcal{I}'} = D^{\mathcal{I}} \cap W'$*, then* $C^{\mathcal{I}'} = C^{\mathcal{I}} \cap W'$*. Similar arguments allow to prove the cases* $\forall P.D$*,* $\leq nP$ *and* $\geq nP$*.*

Now, for every axiom $\mathcal{A} \in \mathcal{O}$*, we have:*

- *If* $\mathcal{A}$ *of the form* $C \sqsubseteq D$*, since* $\mathcal{I} \models \mathcal{A}$*, we have that* $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$*. Now, since* $C^{\mathcal{I}'} = C^{\mathcal{I}} \cap W'$ *and* $D^{\mathcal{I}'} = D^{\mathcal{I}} \cap W'$*, then also* $C^{\mathcal{I}'} \subseteq D^{\mathcal{I}'}$*, and therefore* $\mathcal{I}(V_n) \models \mathcal{A}$

- *If* $\mathcal{A}$ *of the form* $P \sqsubseteq Q$*, since* $\mathcal{I} \models \mathcal{A}$*, we have that* $P^{\mathcal{I}} \subseteq Q^{\mathcal{I}}$*. Now, since* $P^{\mathcal{I}'} = P^{\mathcal{I}} \cap W' \times W'$ *and* $Q^{\mathcal{I}'} = Q^{\mathcal{I}} \cap W' \times W'$*, then also* $P^{\mathcal{I}'} \subseteq Q^{\mathcal{I}'}$*, and therefore* $\mathcal{I}(V_n) \models \mathcal{A}$

- *If* $\mathcal{A}$ *of the form* $C(a)$*, since* $\mathcal{I} \models \mathcal{A}$*, we have that* $a^{\mathcal{I}} \in C^{\mathcal{I}}$*. Since* $a^{\mathcal{I}'} = a^{\mathcal{I}} \in W'$ *and* $C^{\mathcal{I}'} = C^{\mathcal{I}} \cap W'$*, then* $\mathcal{I}(V_n) \models \mathcal{A}$

- *Analogous arguments can be applied to* $P(a, b)$

- *If* $\mathcal{I} \models Trans(P)$*, then if* $(a, b) \in P^{\mathcal{I}}$ *and* $(b, c) \in P^{\mathcal{I}}$*, then* $(a, c) \in P^{\mathcal{I}}$*. Since* $P^{\mathcal{I}'} = P^{\mathcal{I}} \cap W' \times W'$*, then if* $(a, b) \in P^{\mathcal{I}'}$ *and* $(b, c) \in P^{\mathcal{I}'}$*, then* $(a, c) \in P^{\mathcal{I}'}$*, whereas the only other case is when all of* a, b, c *are outside* W'*. Note that it cannot happen that any of* a, b, c *is in* W' *whereas the rest is outside, by the way we have constructed* W'*.*

1.2) To Prove *We prove that if* $\mathcal{I} \leftrightarrow \mathcal{M}$*, then* $\mathcal{M} \models \Sigma$ *iff* $\mathcal{M} \models \mathcal{O}$*.*
$(\rightarrow)$

Let $\mathcal{A}$ *be an axiom in* Σ*. Since* $\Sigma \sim \mathcal{O}$*,* $\mathcal{A} \in \mathcal{O}$*. We show that if* $\mathcal{M} \models A$*, then* $\mathcal{I} \models \mathcal{A}$

- *Let* $\mathcal{A}$ *be of the form* $C \sqsubseteq D$*. We have 4 cases:*

   1. $C^{\mathcal{I}} = C^{\mathcal{M}_i}$ and $D^{\mathcal{I}} = D^{\mathcal{M}_i}$. Since $\mathcal{M} \models \mathcal{A}$, $C^{\mathcal{M}_i} \subseteq D^{\mathcal{M}_i}$, and therefore $\mathcal{I} \models \mathcal{A}$.
   2. $C^{\mathcal{I}} = C^{\mathcal{M}_i}$ and $D^{\mathcal{I}} = D^{\mathcal{M}_i} \bigcup_{j\neq i} W_j$. Since $\mathcal{M} \models \mathcal{A}$, $C^{\mathcal{M}_i} \subseteq D^{\mathcal{M}_i}$, and $C^{\mathcal{M}_i}$ and $D^{\mathcal{M}_i}$ are disjoint with $\bigcup_{j\neq i} W_j$ it follows that $\mathcal{I} \models \mathcal{A}$.
   3. $C^{\mathcal{I}} = C^{\mathcal{M}_i} \bigcup_{j\neq i} W_j$ and $D^{\mathcal{I}} = D^{\mathcal{M}_i}$. This case is not possible by $\mathcal{E}$-safety of $\mathcal{O}$, since $g(C) = F$ and $g(D) = T$
   4. $C^{\mathcal{I}} = C^{\mathcal{M}_i} \bigcup_{j\neq i} W_j$ and $D^{\mathcal{I}} = D^{\mathcal{M}_i} \bigcup_{j\neq i} W_j$. Again, since Since $\mathcal{M} \models \mathcal{A}$, $C^{\mathcal{M}_i} \subseteq D^{\mathcal{M}_i}$, and $C^{\mathcal{M}_i}$ and $D^{\mathcal{M}_i}$ are disjoint with $\bigcup_{j\neq i} W_j$ it follows that $\mathcal{I} \models \mathcal{A}$.

- Let $\mathcal{A}$ of the form $P \sqsubseteq Q$. Since $\mathcal{A} \in \Sigma$, both P, Q are ij-properties. Since $\mathcal{M} \models \mathcal{A}$, then $P^{\mathcal{M}_{ij}} \subseteq Q^{\mathcal{M}_{ij}}$ and since $P^{\mathcal{M}_{ij}} = P^{\mathcal{I}}$ and $Q^{\mathcal{M}_{ij}} = Q^{\mathcal{I}}$, it follows that $P^{\mathcal{I}} \subseteq Q^{\mathcal{I}}$. Therefore $\mathcal{I} \models \mathcal{A}$

- Let $\mathcal{A}$ of the form $C(a)$. We have two cases:

   1. $C^{\mathcal{I}} = C^{\mathcal{M}_i}$. Since $\mathcal{M} \models \mathcal{A}$, then $a^{\mathcal{M}_i} \in C^{\mathcal{M}_i}$. Since $a^{\mathcal{I}} = a^{\mathcal{M}_i}$ and $C^{\mathcal{I}} = C^{\mathcal{M}_i}$, then $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and therefore $\mathcal{I} \models \mathcal{A}$
   2. $C^{\mathcal{I}} = C^{\mathcal{M}_i} \bigcup_{j\neq i}$. Since $\mathcal{M} \models \mathcal{A}$, then $a^{\mathcal{M}_i} \in C^{\mathcal{M}_i}$. Since $a^{\mathcal{I}} = a^{\mathcal{M}_i}$ and $C^{\mathcal{I}} = C^{\mathcal{M}_i} \bigcup_{j\neq i} W_j$, then $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and again $\mathcal{I} \models \mathcal{A}$

- Let $\mathcal{A}$ of the form $P(a, b)$ with $P \in \epsilon ij$. Since $\mathcal{M} \models \mathcal{A}$, $(a^{\mathcal{M}_i}, b^{\mathcal{M}_j}) \in P^{\mathcal{M}_{ij}}$. Since $a^{\mathcal{M}} = a^{\mathcal{I}}$, $b^{\mathcal{M}} = b^{\mathcal{I}}$ and $P^{\mathcal{M}} = P^{\mathcal{I}}$, it follows that $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in P^{\mathcal{I}}$ and therefore $\mathcal{I} \models \mathcal{A}$

- Let $\mathcal{A}$ be of the form $Trans(P)$. Since $\mathcal{M} \models \mathcal{A}$, $P^{\mathcal{M}_{ij}}$ is transitive. Since $P^{\mathcal{M}_{ij}} = P^{\mathcal{I}}$, $P^{\mathcal{I}}$ is transitive and therefore $\mathcal{I} \models \mathcal{A}$

Therefore $\mathcal{I} \models \mathcal{O}$.
($\leftarrow$)
Now we prove that, if $\mathcal{I} \models \mathcal{O}$, then $\mathcal{M} \models \Sigma$. From Lemma (...) we have that:

$$Z^{\mathcal{M}_i} = Z^{\mathcal{I}} \cap W_i \text{ (1)}$$

We prove that $\mathcal{M}$ is a model of every axiom $\mathcal{A} \in \Sigma$. The only interesting case is when $\mathcal{A}$ is of the form $C \sqsubseteq D$. Since $(C \sqsubseteq D)^{\mathcal{M}_i}$ implies that $C^{\mathcal{M}_i} \subseteq D^{\mathcal{M}_i}$ and (1) holds, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ and therefore $\mathcal{I} \models \mathcal{A}$

2)

Suppose that $\mathcal{O}$ is not $\mathcal{E}$-safe and $\Sigma \sim \mathcal{O}$, then there is an axiom $\mathcal{A}$ of the form $C \sqsubseteq D$ s.t. $C^{\mathcal{I}} = C^{\mathcal{M}_i} \bigcup_{j\neq i} W_j$ and $D^{\mathcal{I}} = D^{\mathcal{M}_i}$. Since $D^{\mathcal{M}_i}$ is disjoint with $\bigcup_{j\neq i} W_j$, then $C^{\mathcal{I}}$ is not a subset of $D^{\mathcal{I}}$, unless $\bigcup_{j\neq i} W_j \subseteq W_i$ which implies that $W = W_i$ and therefore $\Sigma = \{\mathcal{O}\}$. If $C^{\mathcal{I}}$ is not contained in $D^{\mathcal{I}}$, $\mathcal{I}$ is not a model of $\mathcal{A}$ and since $\mathcal{A} \in \mathcal{O}$, $\mathcal{I}$ is not a model of $\mathcal{O}$.

Proof for Lemma 3 *The first part is obvious. The second is a direct consequence of Theorem 1.*

Proof for Theorem 3 *Termination stems from the fact that the the number of partitioning steps is bounded by the number of entities in O (including class descriptions). The complexity is justified by the fact that each element (axiom or entity) in O is moved only once and each partitioning step involves checks that are linear in the size of the input.*

We need to show that $\Sigma \sim \mathcal{O}$*. Property 1) follows from the fact that the components of* Σ *do not share entities. On the other hand, no new entities are created and no existing entities are removed. Finally, no entities other than roles are transformed into link properties. Property 2) holds since no new explicit axioms are generated during the partitioning process and no existing axioms are removed.*

The proof for maximality is obtained by showing that the axioms and entities that are grouped together by $Partition(\mathcal{O})$ ***must*** *be grouped together; otherwise either compatibility with* $\mathcal{O}$ *would be violated or we would end up with a syntactically invalid* $\mathcal{E}$*-Connection. This shows that the partitions are "minimal" and hence the number of partitions is maximal.*

Suppose $(C \sqsubseteq D) \in \mathcal{O}$ *and* C, D *are in different components; then we must either remove the axiom, in which case we are violating compatibility, or we would obtain a syntactically invalid* $\mathcal{E}$*-Connection, since an axiom* $(C \sqsubseteq D)$ *cannot involve two concepts in different components. A similar argument yields to a contradiction in the following cases:* $C(a) \in \mathcal{O}$ *and* a *and* C *in different components (separated);* $P(a, b) \in \mathcal{O}$ *and* P, a *separated;* $P \sqsubseteq Q$ *and* P, Q *separated;* $(C \sqcap D) \in \mathcal{O}$ *and* $C, D, (C \sqcap D)$ *separated,* $(C \sqcup D) \in \mathcal{O}$ *and* $C, D, (C \sqcup D)$ *separated,* $\{a\} \in \mathcal{O}$ *and* $a, \{a\}$ *separated,* $\exists R.C \in \mathcal{O}$ *and* $R, \exists R.C$ *separated,* $\forall R.C \in \mathcal{O}$ *and* $R, \forall R.C$ *separated,* $(\geq nR) \in \mathcal{O}$ *and* $R, \geq nR$ *separated,* $(\leq nR) \in \mathcal{O}$ *and* $R, \leq nR$ *separated,* $\neg C \in \mathcal{O}$ *and* $C, (\neg C)$ *separated. Finally,* $QualifiedRes(R, C), QualifiedRes(R, D) \in \mathcal{O}$ *and* C, D *separated, where* $QualifiedRes(R, C)$ *is either* $\exists R.C$ *or* $\forall R.C$*. This last condition holds because a link property cannot point to two different domains nor be a role and a link property simultaneously.*

Finally, note that, whenever there is a choice, the algorithm always chooses to separate elements into different components, and that always involves the creation of a new link property.

The safety check ensures that $\Sigma \approx \mathcal{O}$ *by Theorem 1. Therefore, the algorithm is a solution for* ***P2)*** *and a solution for* ***P1)*** *if the safety check is not performed.*

References

[1] B. Cuenca-Grau, B. Parsia, and E.Sirin. Tableau algorithms for e-connections of description logics. Technical report, UMIACS, 2004. Available at http://www.mindswap.org/2004/multipleOnt/papers/EconnTableau.ps.

[2] B. Cuenca-Grau, B. Parsia, E.Sirin, and A.Kalyanpur. Automatic partitioning of owl ontologies using e-connections. Technical report, UMIACS, 2005. Available at http://www.mindswap.org/2004/multipleOnt/papers/Partition.pdf.

[3] B. Cuenca-Grau, B. Parsia, and E. Sirin. Working with multiple ontologies on the semantic web. In *Proc. of ISWC 2004*, 2004.

[4] O. Kutz, C. Lutz, F. Wolter, and M. Zakharyaschev. E-connections of abstract description systems. *Artificial Intelligence 156(1):1-73*, 2004.

[5] B. MacCartney, S. A. McIlraith, E. Amir, and T. Uribe. Practical partition-based theorem proving for large knowledge bases. In *Proc. of IJCAI 2003*, 2003.

[6] H. StuckenSchmidt and M. Klein. Structure-based partitioning of large class hierarchies. In *Proc. of ISWC 2004*, 2004.

```
-Algorithm Partition($\mathcal{O}$)
-Input: An OWL KB $\mathcal{O}$
-Output: A maximal partition $\Sigma = \{\Sigma_1, ..., \Sigma_n\}$

$\Sigma \leftarrow \{\Sigma_0\}$ with $\Sigma_0 = \mathcal{O}$
$\mathcal{S} \leftarrow$ Set of all concepts, roles and individuals in $\mathcal{O}$
if $\mathcal{O}$ not $\mathcal{E}$-safe, return $\Sigma = \{\Sigma_0\}$
for each $X \in S$, $State(X) \leftarrow 1$
Repeat     Select $X \in S$ with $State(X) = 1$
  if $X$ a concept, individual or link property, then $State(X) \leftarrow 2$
  if $X$ a role then $State(X) \leftarrow 3$
  Create new KB $\Sigma_i$. $\Sigma \leftarrow \Sigma \cup \{\Sigma_i\}$
  Repeat
    for all concept $C \neq \top$ with $State(C) = 1$
      if any of the following conditions holds:
        1)$(C \sqsubseteq D)$ or $(D \sqsubseteq C) \in \mathcal{O}$, and $State(D) = 2$
        4)$C(a) \in \mathcal{O}$ and $State(a) = 2$
        5)$\exists P.C$ or $\forall P.C \in \mathcal{O}$, for $P$ a role, and $State(P) \in \{2, 4\}$
        6)$C \sqcap D$ $(C \sqcup D) \in \mathcal{O}$ and $State(D) = 2$
        7)$C = D \sqcap E$, or $C = D \sqcup E$ and $State(E) = 2$ or $State(D) = 2$
        8)$C$ a restriction $\exists P.D, \forall P.D, \geq nP$ or $\leq nP$, with $P$ a role and $State(P) \in \{3, 4\}$, or a link with $State(P) = 2$
        9)$C = \{a\}$ and $State(a) = 2$
        10)$(\neg C) \in \mathcal{O}$ and $State(\neg C) = 2$
        10)$C, E \in Bound(P)$ and $State(E) = 2$
      then $State(C) \leftarrow 2$
    for all individual $a$ with $State(a) = 1$
      if any of the following conditions holds:
        1)$C(a) \in \mathcal{O}$ and $State(C) = 2$
        2) $\{a\} \in \mathcal{O}$ and $State(\{a\}) = 2$
        3)$P(a, b) \in \mathcal{O}$ and either of the following holds:
          3.1) $P$ a role with $State(P) = 3$ or $State(P) = 4$
          3.2) $P$ a Link Prop. with $State(P) = 2$
        4)$P(b, a) \in \mathcal{O}$, $P$ a role, and $State(P) = 2$ or $State(P) = 4$
      then $State(a) \leftarrow 2$
    for all Link Prop. $P$ with $State(P) = 1$
      if any of the following conditions holds:
        1)$(P \sqsubseteq Q)$ or $(Q \sqsubseteq P) \in \mathcal{O}$ , and $State(Q) = 2$
        3)There is a restriction $D$ with property $P$ and $State(D) = 2$
        4)$P(a, d) \in \mathcal{O}$ and $State(a) = 2$
      then $State(P) \leftarrow 2$
    for all role $P$ with $State(P) \in \{1, 2, 3\}$
      if $(P \sqsubseteq Q)$ or $(Q \sqsubseteq P) \in \mathcal{O}$, and $State(Q) \in \{2, 3, 4\}$ then $State(P) \leftarrow State(Q)$
      if $State(P) \in \{2, 3\}$ and $P$ transitive, then $State(P) \leftarrow 4$
      if $State(P) = 1$, $P(a, b) \in \mathcal{O}$ and $State(a) = 2$, then $State(P) \leftarrow 3$
      if $State(P) = 2$, $P(a, b) \in \mathcal{O}$ and $State(a) = 2$, then $State(P) \leftarrow 4$
      if $State(P) = 1$, $P(a, b) \in \mathcal{O}$ and $State(b) = 2$, then $State(P) \leftarrow 2$
      if $State(P) = 3$, $P(a, b) \in \mathcal{O}$ and $State(b) = 2$, then $State(P) \leftarrow 4$
      if $State(P) = 1$, $D$ a restr. on $P$, $State(D) = 2$, then $State(P) \leftarrow 3$
      if $State(P) = 2$, $D$ a restr. on $P$, $St(D) = 2$, then $St(P) \leftarrow 4$
      if $State(P) = 1$, $\exists P.C$ or $\forall P.C \in \mathcal{O}$, $State(C) = 2$, then $State(P) \leftarrow 2$
      if $State(P) = 3$ and $\exists P.C$ or $\forall P.C \in \mathcal{O}$ with $State(C) = 2$, then $State(P) \leftarrow 4$
      if $State(P) = 1$ and $State(Inv(P)) = 2$, then $State(P) \leftarrow 3$
      if $State(P) = 1$ and $State(Inv(P)) = 3$, then $State(P) \leftarrow 2$
      if $State(P) = 1, 2,$ $or$ $3$ and $State(Inv(P)) = 4$, then $State(P) \leftarrow 4$
  until no state transition has occurred
  for each Axiom $\mathcal{A} \in \mathcal{O}$
    if any of the following conditions holds
      1)$\mathcal{A}$ of the form $C \sqsubseteq D$, $State(C) = State(D) = 2$
      2)$\mathcal{A}$ of the form $C(a)$, $State(a) = State(C) = 2$
      3)$\mathcal{A}$ of the form $R_1 \sqsubseteq R_2$, and either $State(R_1) = 3, State(R_2) = 3$, or $State(R_1) = 4, State(R_2) = 4$
      4)$\mathcal{A}$ of the form $Trans(R)$, $State(R) \in \{3, 4\}$
      5)$\mathcal{A}$ of the form $R(a, b)$, $State(R) = \in \{3, 4\}$
      6)$\mathcal{A}$ of the form $G_1 \sqsubseteq G_2$,$State(G_1) = State(G_2) = 2$
      7)$\mathcal{A}$ of the form $G(a, b)$, $State(G) = 2$
    then $\mathcal{O} \leftarrow \mathcal{O} - \{\mathcal{A}\}$ and $\Sigma_i \leftarrow \Sigma_i \cup \{\mathcal{A}\}$

   for each P, s.t. $State(P) = 2$, make $P$ a link prop. with $State(P) \leftarrow 1$
  for each link prop. $P \in \Sigma_j$, $\Sigma_j \in \Sigma$, and $\Sigma_j \neq \Sigma_i$
    if $P$ points to $\Sigma_0$ and $\forall X \in bounded(P)$, $State(X) = 2$, make $P$ point to $\Sigma_i$
until $\Sigma_0 = \emptyset$
$\Sigma \leftarrow \Sigma - \Sigma_0$
return $\Sigma$
```

Figure 1: The Automatic Partitioning Algorithm