Ontological View-driven Intensional Semantic Integration for Information Systems in a Decentralized Environment

Fateh Mohamed Ali Adhnouss1, Husam M. Ali El-Asfour1, Kenneth McIsaac1, Idris El-Feghia2, Raafat Aburukba3 and Abdulmatalib Wahaishi4

1Dept. of Electrical & Computer Engineering, University of Western, London, Canada 2Faculty of Information Technology, Misurata University, Libya 3Department of Computer Science and Engineering, American University of Sharjah, Sharjah, U.A.E. 4Software Engineering, Rochester Institute of Technology, New York, U.S.A.

Keywords: Ontological View, Extensional Semantics, Intensional Semantics, Semantic Integration.

Abstract: Ontologies are an essential component of semantic integration approaches for information systems. In a decentralized environment, each specification of the domain reflects an Ontological view. However, the semantics characterization of information systems in such decentralized environment poses a significant issue related to their integration. Information systems are viewed as independent intensional entities with their own beliefs, distinct from those held by others. Such autonomy is distorted by traditional extensional semantics. Other entities’ beliefs are introduced into a given entity, thus affecting their beliefs. Additionally, the information that one entity provides to another entity may not be consistent with the information known by the latter. We need an alternative semantics for information integration, which is not dependent on the extension, but rather on the underlying conceptualization. This paper proposes a classification of the environment where the information systems live and a novel modelling paradigm for information integration using intensional logic to model the ontological views. The model comprises a formal modeling approach for the conceptualization as well as for the semantic integration process.

1 INTRODUCTION

Semantic integration based on ontologies is an essential category of solutions to the semantic integration problem. However, the ontology may only be agreed upon in a closed environment. When different information systems designers face the same domain, each will have a specific view of the domain regarding their interest, which results in multiple models (Xue et al., 2012). Since the domain can be viewed in various ways, there is not merely one unique “ontology” for it. Instead, each view can be formally and explicitly specified, and we define the corresponding specification as an ontological view (Ov). Usually, there are no explicit ontologies within information systems. Rather, the associated semantics are implied within the supporting information model (Wang et al., 2009). The information model reflects a specific view of the conceptualization that implicitly defines an Ov: Semantic integration approaches for information systems based on the extensional model are inadequate for a decentralized environment (Majkić and Prasad, 2018; Ali and McIsaac, 2020). This is because they do not account for the dynamic nature of a decentralized environment. The dynamic nature of an environment can be described as a structure using the set of entities present in the environment and the relations between them. The relations between the entities may vary. There is a need for an adequate semantic integration model to account for the decentralized environment to capture the variations of the relations between the entities (dynamic nature) and the changes that may occur in the relations among entities that exist in the domain of interest. This paper discusses conceptualization and modeling languages based on environments classification and proposes a classification of conceptualization and a novel modeling paradigm for information integration using intensional logic to model the ontological views. The model comprises a formal modeling approach for the conceptualization as well as a formal modeling view for the semantic integration process.

Additionally, the work focuses on conceptualization and Ov models to facilitate semantic integration
for information systems in a decentralized environment. The investigation of the extensional, extension reduction and intensional formal models for conceptualization to account for a decentralized environment is the central focus of the work presented here.

The rest of this paper is organized as follows: Section 2 explains the motivation for the research. Section 3 presents a brief background and a literature review relevant to our research. Section 4 proposes a classification of the environment, followed by theoretical foundations for conceptualization and ontological views for semantic integration. The last section of the paper provides conclusions and directions for our future research.

2 MOTIVATION

Within the Ov, there can be very lightweight languages at the representation level that may consist of terms only, with little or no explicit specification of conceptualization. On the other hand, there are rigorously formalized logical theory-based approaches. We provide a formal way to explicitly specify the conceptualization with rich details based on various information models, utilizing logical language with a more explicit specification of conceptualization. This can be done by investigating the nature of the environment, providing an adequate conceptualization structure, and identifying the conceptualization structure’s conversion into a representation. As such, we focus our attention on conceptualization and Ov models to facilitate semantic integration for information systems in a decentralized environment. This can be done by investigating the nature of the environment, providing an adequate conceptualization structure, and identifying the conceptualization structure’s conversion into a representation. More specifically, this research establishes a foundation for semantic integration in a decentralized environment to address the following issues:

- The adequate structure of conceptualization.
- The adequate representation language to provide the manifestation of conceptualization.
- The correctness of the model (sound and complete).

Motivated by the aforementioned issues, the work presents a semantic-based Integration model that is adequate for decentralized information systems, where the Ov is the basis and the following are the underpinning foundation pillars:

- Formal modeling of conceptualization: This involves investigating the nature of the environment, its classifications, the various conceptualization approaches (extensional, extension reduction, and intensional) and their supporting languages (intensional logic, extensional logic). Accordingly, a formal model for conceptualization to account for a decentralized environment is presented.

- Formal modeling of semantic integration: The modeling of the conceptualization should be the basis for the integration of various heterogeneous information systems. This integration will be derived by the mapping between the ontological views of the information systems.

3 BACKGROUND AND LITERATURE REVIEW

A tremendous amount of previous research has been devoted to providing various semantic integration solutions. This section outlines some of the existing approaches which have been applied to semantic integration solutions. Background concepts and views will be provided, and various categories of these existing approaches will be discussed accordingly.

3.1 Semantic Integration

Multiple descriptions of the term “semantic integration” have been developed. Taking human conversation as an example, the heart of the semantic integration problem is how to determine when two statements are about the same subject (Bhatia and Breaux, 2018). In some communities, this is known as the coreferencing problem (Alfrjani et al., 2019). Semantic integration, according to (Vetere and Lenzerini, 2005), is achieved by conceptual mappings that make different datasets and process descriptions equivalent, either pairwise or in relation to some (partial) unified conceptualization. The conceptualization principle is related to the “perception” of a universe within the context of another.

A universe can be categorized with the context of an environment it attempts to describe. An environment includes a set of “entities” and the “relationship” between them.

In the literature, there are two basic semantic integration approaches, structural-based, and semantic-based (Xue et al., 2012). In structural-based approaches, the integration is based on providing or generating a global structure that characterizes the underlying conceptualization of the environment. The global structure can be logically modeled and physi-
cally independent of its implementation. It is mainly a syntactic-based integration, with highly implicit underlying semantics within the structure and the design.

In semantic based approaches, integration is obtained by sharing an equivalent conceptualization among various information systems, which utilize the underlying conceptualization of each system. The semantic integration can be derived by identifying semantic correspondences among a set of specifications of conceptualization, given that multiple system conceptualizations are derivable.

### 3.2 Ontology-driven Integration

Various definitions of "ontology" have been proposed. A basic definition of ontology is "the specification of conceptualizations, used to help programs and humans share knowledge" (Gruber, 1993). This definition then evolves to "a formal, explicit specification of a shared conceptualization" (Guarino et al., 2009). Another view of ontology is a hierarchy of terms corresponding to "concepts" related by subsumption relationships, such as things, events, and a set of relations that are specified in some way in order to create a shared equivalent conceptualization of "reality" in an environment (McGuinness et al., 2000). In the information management and knowledge sharing areas, ontology typically defined informally as a set of logical axioms describing the meanings of terms in a particular community (Guarino et al., 2009).

On the level of ontology representation, research has been conducted on the level of "completeness" and "richness" of the ontology representation, as shown in Figure 1 (Uschold and Gruninger, 2004). At one extreme, there can be very lightweight representation languages that may consist of terms only, with little or no explicit specification of conceptualization. At the other end of the spectrum, there are rigorously formalized logical theory-based approaches. In this work, we treat "ontology" as a formal specification of conceptualization.

In this work, we view "ontology" as a formal specification of conceptualization. In traditional ontology-driven approaches, semantic integration is based on sharing or deriving a common ontology among the information systems. Two main models were discussed in (Sheth and Larson, 1990), namely: local-as-view (LAV) and global-as-view (GAV). In global-as-view approaches, entities and relationships in a global ontology are defined as a view over system ontologies. In local-as-view approaches, entities and relationships in each system local ontology are defined as a view over the global ontology. A major challenge is that in order to enable answering a query expressed over the global ontology, it requires reformulating the query in terms of queries to the local ontology. While in the global-as-view approach such a reformulation is guided by the definitions in the mapping. Yet, the problem in this case is to support a reasoning step in order to infer the ways the local ontology is meaningful to the query.

### 3.3 Related Work

There was excellent work devoted to the definition of conceptualization’s structure. To our knowledge, there are three approaches that have been proposed in the literature, namely, the extensional approach (Gruber, 1993), the extensional reduction approach (Guarino et al., 2009) and PRP approach (Bealer, 1979). Depending on the type of environment, each of these approaches performs slightly different functions. The extensional structure is suitable for describing a snapshot of the environment. Alternatively, the extensional reduction structure (possible worlds structure) is suitable for conceptualizing a dynamic environment in which the relationships between the entities are free to change.

(Xue et al., 2012) attempt to address semantic integration and mapping between ontologies in a decentralized environment. The primary focus was on semantic equivalence relationship discovery. For conceptualization, the author employed possible worlds structure. The author also used the schema-matching approaches to deal with the issue of heterogeneity. However, semantic equivalence relationship discovery based on only syntax and representation language is extensional in nature.

In (Wang et al., 2009), heterogeneity, autonomy, and distribution are addressed. This work proposes a framework for semantic transformation in decentralized systems. Conceptualization structure of the work is based on the Possible Worlds formal model. However, the work does not address the representa-
tion language aspect, instead providing a definition of the ontological commitment through the use of an extensional representation language.

The work in (Ali and McIsaac, 2020), proposed an intensional model for the conceptualization. This model is based on the theory of Properties Relations and Propositions (PRP) proposed by (Bealer, 1979). Intensional based modeling is an adequate and natural choice for modeling in an open environment. However, the focus was on the conceptualization structure aspect, and it does not address the formal specification of a conceptualization as it is the main element for the ontology and ontological views based on the definition of ontology.

4 PROPOSED SOLUTION : A FORMAL MODEL

As the objective is to provide theoretical and practical foundations for developing sound engineering solutions for ontological view-driven semantic integration in decentralized environments, we stress the following assumptions which are considered practical and reasonable:

• For semantic integration, all considered information systems are associated with the Ov that is committed to overlapping intended models.

• For each information system, there is an explicit specification of conceptualization, such as schema that is used to organize the system’s data and convert the data into information.

4.1 Semantic Integration: Foundation and Core Principle

It is noteworthy that, in our work we focus on aspects related to the conceptualization of the “Universe”. A Universe can be thought of in terms of an environment in which the existence of its “entities” is governed by rules and axioms (Stevenson, 2010). Within this context, we classify a particular environment as follows:

• Closed environment (CE): where the existence of the entities is predefined.

• Open environment (OE): where the existence of the entities may vary.

Furthermore, CE, can be further labeled as either being static or dynamic. In a static closed environment (SCE) the entities and their relevant relationships are invariant and predefined in priori. In contrary, in a dynamic closed environment (DCE), the entities are predefined, but the relationships are mutable and may vary. Whereas, in OE, both the existence of the entities and the relationships are mutable and might vary unpredictably (i.e., an entity may join or leave the environment). Figure 2 below illustrates the different categories of the environments and the relevant elements.

In the context of information systems, data is a set of computational symbols that can represent extensional entities and relationships corresponding to a given conceptualization, within an observed universe. Whereas, the specification of a conceptualization, such as a schema, is a semantic level model of a conceptualization which uses a specific language. Typically, in information systems, semantic level modeling is implied in the design and the structure of the model. In this work, the semantic level modeling of information systems is represented in terms of “conceptualization structure”, “language” and “the specification of a conceptualization”. We view conceptualization as an abstraction of the observed universe. The specification of conceptualization using a specific language Ov represents an approximate intended mode as shown in the Figure below. The conceptualization space defines the conceptualization structure, the representation language and the Ov.

Consequently, we classify the semantic level modeling in terms of the type of the associated environment of the observed universe. A universe associated with an environment can be described using a conceptualization structure and supporting language, as depicted in Table1.

In order to support semantic level modelling, it is essential for both the language and the conceptualization structure to be correct in the sense that it is complete and sound. In other words, the language and structure should satisfactorily model all the inherent elements of the environment. The completeness ensures that all intended models can be generated by the Ov. While the soundness ensures that any model generated by the Ov is an intended model.
A particular universe can be described in terms of conceptualization structure which includes a set of entities present in the environment and the set of relations between them.

Therefore, a conceptualization \( (C) \) is defined in terms of the set of the possible worlds \((W)\), the set of entities \((D)\) and intensional relations \((R)\).

In a universe of a static closed environment \((\text{SCE})\) where entities and the relations remain constant, the conceptualization is defined in terms of the following tuple:

\[
C = <D, R>
\]

(Gruber, 1993)

In a universe associated with a dynamic closed environment \((\text{DCE})\), the conceptualization takes into consideration the possible worlds in which the variations of the relations between the entities can be modelled using possible worlds theory and thus is defined per the following tuple:

\[
C = <D, W, R>
\]

(Guarino et al., 2009)

A universe associated with an open environment \((\text{OE})\) can be described utilizing intensional algebraic structure is defined in terms of the following tuple:

\[
C = <D, t, R>
\]

(Bealer, 1979)

As for the modelling language, logic-based languages are complete, as such:

- \text{FOL} as extensional language is complete and sound language for \(\text{SCE}\). The extensional model describes the environment in terms of declarative sentences and ordinary relations. The extensional model is appropriate for describing a closed world.
- \text{FOL} utilizing modal operators (necessity and possibility) is complete and sound for the \(\text{DCE}\). The extensional reduction model is appropriate for describing a closed world in which the relations between the entities can change.

- Intensional-based language can be complete and sound for \(\text{OE}\).

In the following subsections we augment a set of definitions derived from (Guarino et al., 2009) as well as (Wang, 2008)’s work that we deem necessary for formalizing an \(\text{Ov}\) from a traditional ontology perspective.

\[\text{4.2 Conceptualization: A Proposed Formal Conceptualization Model}\]

Given that \(<D, R>\) can relate to a particular Static environment (a particular world), we will refer to this as an extensional structure, \(SwC = <D, R_w>\). A dynamic environment contains many of these world structures, one for each world. Enumerating all the possible world structures is not practical and is in fact impossible. In other words, an extensional specification of the conceptualization would require a listing of all possible extensional structures. However, this is impossible in most cases (e.g., if the universe of discourse \(D\) or the set of possible worlds are infinite) or at least exceedingly impractical.

In light of this, it is more practical to introduce the intensional structure as an extension of the conceptualization to include a dynamic environment. Variations in extensional structures can therefore be expressed in an intensional manner. Conversely, one can extend the intensional structure into all possible extensional structures. Here we present a novel conceptualization based on the fusion of multiple extensional structures, which are represented by \(Sw_1 = <D,R_{w_1}>, Sw_2 = <D,R_{w_2}>\), and the intensional structure that is based on possible world, as shown in Figure 4. An attempt is made to abstract all extensional structures into a generic conceptualization to encompass all possible extensional structures.

\[\text{4.3 Formal Modelling}\]

In the proposed model, the conceptualization is represented by two levels of structures. On the higher level is the intensional structure \(C = <D, W, R>\), which contains all possible worlds.

The lower level can be viewed as a function from possible worlds into extensional structure sets such as: \(Sw_1 = <D,R_{w_1}>, ..., Sw_k = <D,R_{w_k}>\).

Therefore, we formally define the structure as follows:
Table 1: Conceptualization structure and supporting language describes a universe associated with an environment.

<table>
<thead>
<tr>
<th>Environment</th>
<th>Conceptualization structure</th>
<th>Supporting language</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCE</td>
<td>( C = &lt; D, R &gt; )</td>
<td>First order logic FOL</td>
</tr>
<tr>
<td>DCE</td>
<td>( C = &lt; D, W, \mathcal{R} &gt; )</td>
<td>Modal logic</td>
</tr>
<tr>
<td>OE</td>
<td>( C = &lt; D, J, \mathcal{R} &gt; )</td>
<td>Intensional FOL</td>
</tr>
</tbody>
</table>

![Diagram](image)

Figure 4: The conceptualization structure and its two levels.

- \( \rho^p : W \rightarrow 2^D^p \) where \( \rho^p \) is intensional relations are defined on a domain space \( < D, W > \) where \( D \) is a domain and \( W \) is a set of maximal extensional structures of such a domain.
- For a generic extensional relation \( \rho \), \( E\rho = \{ \rho(w) | w \in W \} \) will contain the admissible extensions of \( \rho \).
- A conceptualization for \( D \) can be now defined as triple \( C = < D, W, \mathcal{R} > \), where \( \mathcal{R} \) is a set of extensional structures on the domain space \( < D, W > \).
- \( S_c C = < D, R_c C > \) is the intended extensional structure of \( w \) according to \( C \).
  - \( R_c = \{ \rho(w) | \rho \in \mathcal{R} \} \) is the set of extensions (relative to \( w \)) of the elements of \( \mathcal{R} \).
  - \( SC \) is the set \( \{ S_c C | w \in W \} \) of all the intended world structures of \( C \).
- \( C = < D, R > : S_c C \) is the structure of the universe, in the extensional form. This is a direct model for the structure of \( CE \).

## 4.4 Ontological View: A Proposed Formulation

In order to comply with the conceptualization, we present a formal treatment of ontological views in terms of two distinct semantic levels, namely, intensional (\( \Theta \)) and extensional (\( \Phi \)). Generally, intensional semantics are broader than extensional semantics. Hence, if one knows the intensional of an expression, one can determine its extension with respect to a particular world.(Napoli et al., 2017).

Since the notion of a model is an extensional account of meaning (Guarino et al., 2009), a conceptualization that is intentionally specified would necessitate an ontological commitment to specify the intensional meaning of the vocabulary and to constrain its models.

### In this context, if we consider an intensional structure \( < D, W, \mathcal{R} > \) with an intensional language \( L \), an intensional interpretation and vocabulary \( V \), we can define the intensional semantic level of \( O_v \), which corresponds to ontological commitment in (Guarino et al., 2009), as:

\[
\Theta = < C, \mathcal{I} >
\]

where:
- \( C \) is a conceptualization.
- \( \mathcal{I} \) is an intensional interpretation function assigning elements of \( D \) to constant symbols of \( V \), and elements of \( \mathcal{R} \) to predicate symbols of \( V \).

\[\mathcal{I} : V \rightarrow D \cup \mathcal{R}\]

In order to restrict the intensional semantic level \( \Theta \), utilized by \( \mathcal{I} \), of the intentional logical language \( L \) to be used in a manner intended for a specific domain rather than randomly, an extensional interpretation \( I \) accompanied by a set of axioms is required.

Now, given the intensional semantic level \( \Theta \) and an extensional interpretation \( I \), an intended extension (a model) \( M = < S_w, I > \) of \( L \), is compatible with \( \Theta \) if:
- \( S_w C \in SC \).
- \( \forall c \in V : I(c) = \mathcal{I}(c) \).
Therefore, for a language $L$ and conceptualization $C$, the set of all extensions (models) of $L$ that are compatible with $\Theta$ represents the set of intended extensions $I_{\Theta}(L)$ of $L$ according to $\Theta$.

The extensional semantic level $\Phi$ of $Ov$ can then be expressed as a specification of $C$ formulated by a language $L$, an extensional interpretation $I$ and a set of axioms that it and approximate the intensional interpretation to the intended extensions (models) $I_{\Theta}(L)$.

As a result, we can say:

- $Ov$ commits to $C$ if it has been designed with the purpose of characterizing $C$, and it approximates the reality $D$ through its extensions.
- A language $L$ commits to $\Theta$ if it commits to conceptualization $C$ such that $\Phi$ agrees with $C$.
- $L$ commits to $\Phi$ for a given $\Theta$ such that the $I_{\Theta}(L)$ is captured in the models for $\Phi$.

Figure 6 below illustrates how the intensional semantic level is derived from an intensional interpretation, and the extensional semantic level from an extensional interpretation to form an ontological view.

4.5 Ontological View Completeness and Soundness

An essential aspect of semantic integration is the degree of approximation of the intended extensions (models). The level of the approximation of the intended extensions is the extensional semantic level $\Phi$ of $Ov$ which is a logical theory accounting for intensional semantic level $\Theta$ to a conceptualization of a universe. An $Ov$ indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended extensions. As illustrated in Figure 7, if the $Ov$ is not sufficiently accurate, and the intended extension fails to meet the criteria of being complete and sound we can assume that they are not derived from the same conceptualization.

Therefore, in order to develop an ontological view-driven semantic, $\Phi$ needs to be as accurate as possible, otherwise, the ontological views will intersect even though they are not semantically equivalent.

![Figure 7: Ontological View Completeness and Soundness.](image)

4.6 Semantic Integration: A Proposed Formulation

Utilizing the $Ov$ formally, we view the semantic integration problem within the context of two scenarios. In scenario (a) as in Figure 8, assume there are two ontological views of different information systems $A$ and $B$ for a common observed universe $U$. Assume they are associated with intended extensions $I_{\Theta A}(L)$ and intended extensions $I_{\Theta B}(L)$, respectively, which overlap.

In scenario (b) as in Figure 9, intended extensions $I_{\Theta A}(L)$ and $I_{\Theta B}(L)$ do not overlap.

The semantic integration is feasible if its intended extensions $I_{\Theta A}(L)$ and $I_{\Theta B}(L)$ overlap. Therefore, we can conclude that Information systems $A$ and $B$ can be semantically integrated if, and only if, their intended extensions overlap and therefore we can assume that they are derived from the same conceptualization of the same observed universe.

Here we introduce what we call the semantic integration principle, wherein semantic integration be-
tween ontological views can be guaranteed if, and only if, i) each is sound and complete, and ii) they are associated with overlapped intended extensions.

Within the context of this research, we introduce the following assumptions, which are considered practical and reasonable. The objective is to provide a practical foundation for the research and to help reduce the complexity of the problem without loss of generality.

- For semantic integration, all information systems of consideration are associated with ontological views that are committed to overlapping intended extensions.
- For each information system, there is an explicit specification of conceptualization, such as a schema, that is used to organize the system’s data and convert the data into information.

5 CONCLUSIONS

Semantic integration approaches for information systems based on the extensional model are inadequate for a decentralized environment. This is because they do not account for the dynamic nature of a decentralized environment. The dynamic nature of an environment can be described as a structure using the set of entities present in the environment and the relations between them. The relations between the entities may vary. This work presented in the paper has outlined ongoing research and proposed a new approach of conceptualization classification and a modeling language for information integration using intensional logic to model ontological views. The investigation of the extensional, extension reduction and intensional formal models for conceptualization to account for a decentralized environment is the central focus of the proposed model. The intensional semantic model is an applicable solution that utilizes an Ov suitable for semantic integration that supports the deployment of semantically enabled applications in decentralized environments.

REFERENCES


