Empirical Analysis of Impacts of Instance-Driven Changes in Ontologies

Yalemisew M. Abgaz[1], Muhammad Javed[2], Claus Pahl[3]

Centre for Next Generation Localization (CNGL),
School of Computing, Dublin City University, Dublin 9, Ireland
{yabgaz[1]|mjaved[2]|cpahl[3]}@computing.dcu.ie

Abstract. Changes in the characterization of instances in digital content are one of the rationales to evolve ontologies that support a domain. These changes can have impacts on one or more of interrelated ontologies. Before implementing changes, their impact on the target ontology, other dependent ontologies or dependent systems should be analysed. We investigate three concerns for the determination of impacts of changes in ontologies: representation of changes to ensure minimum impact, impact determination and integrity determination. Key elements of our solution are the operationalization of changes to minimize impacts, a parameterization approach for the determination of impacts, a categorization scheme for identified impacts, and prioritization technique for change operations based on the severity of impacts.
Keywords: ontology evolution, impact determination, instance-driven change.

1 Introduction

Ontology evolution is a continuous process. Whenever there is a change in the domain, its conceptualization or specification, the ontology needs to be changed [1–3]. Ontologies, built to give support for specific content within a domain, change as content and embedded ontology instances change and need to be updated synchronously with changes in the domain [4–6].

When new concepts are added or existing ones are deleted or modified in the content, the respective ontology needs to be updated. Implementing the changes requires understanding them correctly and representing them accurately using ontology change operations. However, this only solves few of the problems associated. These changes can trigger further cascaded changes and affect one or more interrelated ontologies. The effects of the change may propagate back to the domain instances in the content leaving the process in a circle. An ontology engineer who detects a change of an instance in a content document and trying to maintain the ontology accordingly may end up with so many unseen impacts. In large and interrelated ontologies, the process of determining impacts of change operators will become time consuming and error prone. Thus, the determination of change impact is crucial.

In this research, some of the key features we investigate are:

– a case-based real-world requirement analysis.
– an analysis to determine the impacts of instance-driven changes in ontologies:
  • Operationalization: how to operationalize changes to ensure minimum impact? We measure minimum impact in terms of consistency, validity, number of operations, cascaded effect, ontologies affected etc
  • Parameterization and Categorization: how to determine different impact categories and parameters to determine impact?
  • Integrity: how to determine the integrity (consistency within and among dependent ontologies and validity within instances and ontologies) of the ontology due to the changes?
  • Prioritization: how to choose the best options with minimum impacts in different situations?

While a significant number of approaches [7–10] focus on addressing consistency and validity of the ontology at the time of change, we focus on analysis and determination of impacts of change operations to minimize their impact on not yet evolved ontologies. We operationalize changes to ensure minimum impact, define parameters to identify and determine impacts, categorize impacts to deal with them at different levels of expertise and prioritize impacts to enable us to choose the options with minimum impacts in different situations.

This paper is organized as follows: Section 2 describes the empirical study and Section 3 focuses on the selection of schemes for impact analysis and identification of parameters. Section 4 presents our proposed framework for the impact determination process. The actual categorization of impacts using different criteria is discussed in Section 5. Evaluation of the results and related work is given in Section 6. We give conclusions in Section 7.

2 Empirical Study

We conducted an empirical study on a repository of help files of a content management software system with the aim of supporting our proposed theoretical solution with an empirical experiment. The case study is selected because it has a wide coverage of domains from the application domain to software systems, which are interdependent on another. Moreover concepts and instances are distributed throughout the content of the help files and create a strong link between the instances in the content and the concepts in the ontologies. This makes it of great interest to investigate instance-driven change impacts because the changes made in the content of the help files will have a direct impact on the ontology and vice versa. We use the term elements to refer to ontology elements such as classes, individuals, attributes, relations, axioms etc [11].

There are four primary ontologies identified for supporting software systems help files. A high-level description of these ontologies and their dependencies are depicted in Fig. 1. The DocBook ontology gives structure to and defines how elements in the help ontology are organized. The help ontology guides the software ontology in a way that explains how the software ontology makes use of the topics, procedures etc. The software ontology allows us implement the

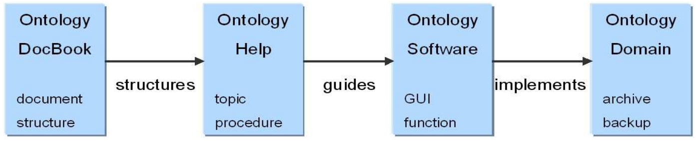

Fig. 1. Ontology Interrelation

ontology domain which is specific to the components in each application. Using the help files, we identified representative scenarios that explain possible changes that may occur when there is a new software release, technology change and/or software structure adjustment. The scenarios further represent frequent changes that occur on the instances of the help files. The scenarios are extracted from the real world changes in the software industry - specifically changes between an old and a new version of the software help content. Our focus is on content changes that trigger a change in an ontology and the scenarios are selected based on the impact of changes in instances. Initially, 15 scenarios were identified that cover all the four ontologies. Based on the frequency of the change, their cascaded impacts, the operations involved and the number of ontologies affected, we selected scenarios that are most representative of the evolution process. Two scenarios are discussed below.

Scenario 1: The new version of the software resulted in a change of component "X" which contains other two sub components "Y" and "Z". The component "X" and its subcomponent "Y " are removed but the subcomponent "Z" is moved up. Here, all the previous instances of "X" and "Y" are preserved as instances of "Z". The desired output is an updated ontology which reflects the change requested. The primary ontology affected is the domain ontology(Fig. 1). The change operations are:

- Move up ("Z")
  - Add instance of ("instance of X", "Z")...
  - Add instance of ("instance of Y", "Z")...
- Delete concept ("Y")
- Delete concept ("X")

Scenario 2: The software engineers introduced a new software component "NC". The new component has new associated help files. The desired output is a software application ontology that has a description of the new component and its properties. The primary ontology affected are the domain and help ontologies (Fig. 1). The change operations are:

- Add Concept ("NC")
- Add sub concept ("NC", "software application")
- Add instance ("help file")...
- Add instance of ("help file", "NC")

3 Schemes for Impact Analysis

In situations where interrelated ontologies are used, the change of one element in one ontology may have an impact on other elements within the same ontology or elements among the interrelated ontologies. The dependency between ontologies, especially when they are used in a specialized domain is often high. Thus, the impact determination process focuses on identifying impacts of change operations on one or more interrelated ontologies.

3.1 Types of Ontology Change Impact

The term impact refers to a consequential change of elements in the ontology due to the application of a change operation on one or more of the elements in the ontology [8, 1, 12]. The impact can be structural or semantic. Structural impact is an impact that occurs on the structural relationship between the elements of the ontology. Semantic impact is an impact that occurs on the interpretation of the ontology and its elements.

Structural impacts are possible consequences on the structure of the ontology due to a structural change.

- Broken Structure:
  - Orphan concept: the change operator may introduce an orphan concept in the ontology
  - Orphan Properties: the change operator may introduce an orphan property in the ontology (properties without domain or without parents)
  - Orphan instance: the change operator may introduce an orphan instance in the ontology
- Cyclic structure: the change operator may introduce a cyclic structure

Semantic impacts are possible inconsistencies and invalidities that arise from the interpretation of the ontology due to structural changes [8].

- Generalization/ specialization: elements (concepts, instances, domains/ranges of properties) move up or down in the hierarchy
- More/less description: a data type property or instance level object property is added to or deleted from a concept
- More/less restrictive: a change to the restriction further restricts or extends its semantics
- More/less extended: a change to its axioms further extends or restricts its semantics

To determine the impact of change operations, we identified the following parameters that determine the nature of impact of a change operation on the elements of an ontology, see Table 1. An example shall explain the approach. All instances of the concept "Assign Role", from the help ontology, in one version of a help file have been changed to either "Assign AdministrativeRole" or "Assign UserRole" in the newer software version. This change can be represented by a

Table 1. Sample parameters for impact determination

General Parameters	**Concept Parameters**	**Property Parameters**	**Axiom Parameters**
Operation	Target concept	Target property	Target axiom
Ontology element type	Has sub/super class	Has sub/super property	Has domain
Ontology target element	Has domain/range	Has data/object property	Has range

composite operation SplitConcept (“Role”, “AdministrativeRole”, “UserRole”). To determine the impact of the change on the ontology, we need to know what the target entity is (the concept “Role”), whether it has subclasses and/or super class, whether it has a data property or object property, and if it is a domain or a range of a property. These parameters about the target concept provide us with information about what potential impacts are associated with a change.

3.2 Change Operations

The following is a list of possible changes that may occur on the structure and the semantics of ontology. Changes can be atomic, composite or domain specific. Higher levels of changes (composite and domain-specific) are created by combining atomic changes in a certain order [13]. Renaming can be done by a series of addition and deletion operations; thus, is not discussed here. Addition and deletion are applied to concepts, properties, restrictions, axioms and instances as target elements. Thus, we have change operations like Add concept, Delete concept, Add property, and Delete axiom.

4 Framework for Change Impact and Integrity Analysis

The empirical study further clarifies that impact determination is a step-by-step process, see Fig. 2. These steps fit into the semantics of change phases of the general ontology evolution process [1].

4.1 Change Request Capturing and Representation

In this first step, the objective is to represent detected changes using suitable change operators that ensure the efficient implementation of the required change. The execution depends on how the change is represented [14, 15] and relies on two factors. First, the selection of the appropriate operator and, second, the order of execution of the operations focusing on efficient ordering of atomic change operations into composite and higher-level granularity to minimize impact [12, 13].

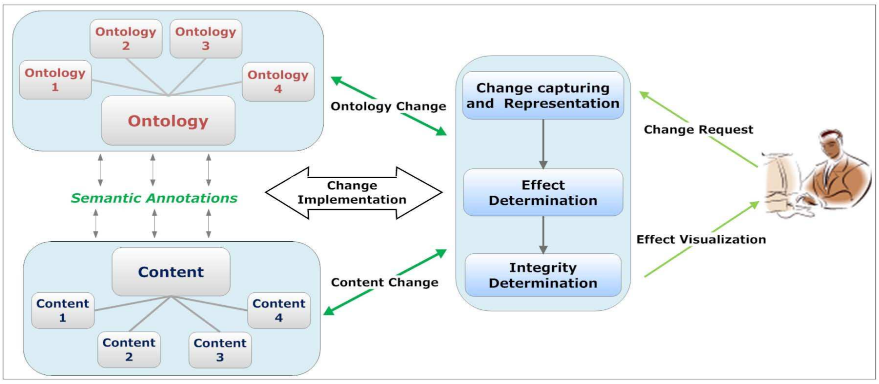

Fig. 2. Instance-driven change and impact determination process

4.2 Impact Determination

This step mainly focuses on determining the impacts of the captured change operations on the elements of the ontology. Impact determination process focuses on analyzing the nature of the operations and the target ontology elements using different parameters. Based on these parameters, this phase enables the categorization of change operations into different category of impacts. This phase further identifies the part of the ontology that is affected by the change. This process is crucial to deal with those identified parts.

4.3 Integrity Determination

Consistency. Once the scope of the impact of the change operations is identified, the next step is to analyze how the consistency of the ontology is affected by these changes. Consistency is analyzed based on consistency rules that are defined for the ontology. Since these rules are defined prior to the implementation of the change operations, it is possible to determine which consistency rules will be violated if a change is made on the ontology [16]. Defining consistency rules is important for consistency analysis. We focus here only on the following widely used rules for an ontology [14]:

1. *Identity invariant:* no two elements should have the same id (URI)
2. *Rootedness invariant:* there should be a single root in the ontology
3. *Concept hierarchy invariant:* no element should have a cyclic graph
4. *Closure invariant:* every concept should have at least one parent concept except the root concept
5. *Cardinality invariant:* the cardinality of a constraint should be a non-negative integer greater than or equal to the minimum cardinality and less than or equal to the maximum cardinality
6. *User-defined constraints:* these constraints are user-defined and needs to be stated in the way they can be implemented like the other invariants.

Validity. Instances in content centric systems are linked to the ontology using semantic annotation [5]. Thus determining the impact of change operations with regard to the instances is crucial. The determination of the validity of instances and instance properties is also based on the validity rules. Theses rules determine how instances/ instance properties should exist in the ontology structurally and how they should be interpreted:

1. *Invalid instance:* given a consistent ontology, if there is an instance that does not correspond to any of the concepts, then that instance is invalid.
2. *Invalid interpretation:* given a consistent ontology, if there is an instance whose interpretation contradicts any interpretation denoted by the consistent ontology, that instance has an invalid interpretation.

5 Categorization of Impact

The categorization of impact is important to systematically handle changes and prioritise crucial impacts to save much time.

5.1 Impact based on Severity

Severity is the degree of impact of a change operation on ontology. The impact is measured qualitatively using consistency and validity, or quantitatively using number of change operations required, ontology elements affected and cascaded effect on dependent ontologies. The impact can be on the structure or semantic, consistency or validity of the existing ontology. Using the selected scenarios to determine the severity of impact of change operations, we analyze how different change operations impact the ontologies. This gives us a better understanding of which operations under what condition have a more severe impact than the others. Defined categories and identified properties are needed as input for automatic categorization at a later stage.

1. *Less or no impact:* changes with no effect on the consistency or the validity of the ontology, e.g. addition of a concept at the bottom of a hierarchy.
2. *Medium impact:* changes with medium impact that can be solved using predefined operations, e.g. addition of concepts in the middle of a hierarchy.
3. *High impact:* changes that create structural inconsistency and require little or no human involvement, usually restricted to a single ontology, e.g. deletion of concepts with subclasses and annotation links.
4. *Crucial impact:* changes that significantly affect the consistency of the ontology, affecting dependent ontologies and instances and their interpretations. They require expert involvement, e.g. deletion of concepts or addition of axioms which create invalid or inconsistent interpretation of elements.

5.2 Impacts based on Type of Operation and Target Elements

Based on the parameters identified in Table 1; Table 2 summarizes the severity of impacts of atomic change operations. The table indicates the severity of the change operations when they are applied to the target elements in the ontology. The severity of impacts of composite change operations can be determined using the atomic change operations involved. The severity within the operations is given in decreasing order. However the severity among the operations is highly tied to the element types and other parameters, for example, addition of an axiom has a more severe effect than deletion of an instance.

Table 2. Severity of change operations and type of elements

Operation Type	**Element Type**
Deletion	Concept, Property [object property, then data property]
	Axioms, Restrictions
	Instances, then instance properties
Addition	Axioms, Restrictions
	Properties, concepts
	Instances, instance properties

5.3 Impacts based on Constraints Violated

The constraints that are violated by the change operation have different levels of impact on the ontology. This idea is backed by the empirical study and is described below.

1. the strength of the consistency and validity rules being violated, e.g. invariant constraints, soft constraints and user defined constraints.
2. the level of human involvement required, e.g. can the system carry out the operations autonomously or does it need a human intervention?

The severity of the constraints violated is listed below in descending order.
- Invalid interpretation, Closure invariant
- Concept hierarchy invariant, Invalid instances, Cardinality invariant
- Identity invariant, Rootedness invariant, Soft constraints
- User defined invariants, maybe severe based on the requirements of the user

6 Discussion and Related Work

Our approach has been evaluated based on its practical and operational applicability in the real world. From the case study, we found out that the proposed solution is effective in allowing to reduce the number of change operations, and consequently the number of cascaded impacts, significantly. We found that it enables us classify impacts into the appropriate categories.

To put our findings into context, we give a brief summary of current practice in the area of ontology evolution, specifically in handling instance-driven change. An interesting research [8] looks at determining the validity of instances in evolving ontologies. The authors evaluated the validity of data instances against changing ontologies and came up with a formal model. They presented a formal notion of structural and semantic validity of data instances. Compared to our work, their work focuses on determination of validity of data instances after a change takes place, but do not address the problem of determining impact. In [3], the effects of domain changes on the performance and validity of the knowledge-based systems are discussed. They analyzed the problems using non-evolved ontologies and present a solution for enabling consistent description of knowledge sources. However, their work emphasizes problems related with metadata evolution and annotation and does not focus on impact determination. The work in[16] discusses consistent evolution of OWL ontologies with the aim of guaranteeing consistency whenever the ontology evolves. Their focus is on structural, logical and user-defined consistency, but does not formally focus on analysis, parametrization and categorization of change impact.

7 Conclusion and Future Work

In this paper, we empirically analyzed and determined the impact of instance-driven change in ontologies. Based on results of the empirical study, we can operationalize, categorize, parameterize and prioritize changes and can analyze their impacts. Based on the severity of the impact, the changes are further analyzed. The case study has highlighted details of problems associated with instance-driven ontology changes and the difficulty of the problem solutions.

The major contribution of our work is the determination of the impact of change operations that are carried out on content-oriented ontologies. The research further contributes to identifying and categorizing change operations based on their impacts, identifying parameters that play significant role in determination of impacts and categorization of the change operations based on the severity of the impacts. We identified parameters for determining severity of the impacts like the cascaded effect, the time required (number of operations) and the human involvement to resolve complex choices.

Our future work will focus on content-oriented ontology change impact determination in a web-based multilingual environment. Specifically, we will focus on sliced Web content annotated using the domain ontologies[17]. Another complexity to be investigated is multilingual content and ontology infrastructures. We will also investigate how we can translate the empirical result into ontology languages such as Web Ontology Language (OWL).

Acknowledgment. This material is based upon works supported by the Science Foundation Ireland under Grant No. 07/CE/I1142 as part of the Centre for Next Generation Localisation (www.cngl.ie) at Dublin City University (DCU).

References

1. Stojanovic, L.: Methods and tools for ontology evolution. PhD thesis, University of Karlsruhe (2004)
2. Maedche, A., Motik, B., Stojanovic, L., Studer, R., Volz, R.: Ontologies for enterprise knowledge management. Intelligent Systems, IEEE. **18**(2) (2003) 22–33
3. Stojanovic, N., stojanovic, L., Handschuh, S.: Evolution in the ontology-based knowledge management systems. Proceedings of the European Conference on Information Systems - ECIS 2002, Gdask, Poland. (2002)
4. Benjamins, V., Contreras, J., Corcho, O., Gomez-perez, A.: 'six challenges for the semantic web'. In Cristani, M(ED): KR2002 Workshop on the Semantic Web, Toulouse, France. (2002)
5. Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna, F.: Semantic annotation for knowledge management:requirements and survey of the state of the art. Web Semantics: Science, Services and Agents on World Wide Web. **4**(1) (2006) 14–28
6. Boyce, S., Pahl, C.: The development of subject domain ontologies for educational technology systems. Journal of Educational Technology and Society (ETS) IEEE **10**(3) (2007) 275–288
7. Stojanovic, L., Maedche, A., Motik, B., Stojanovic, N.: User-driven ontology evolution management. Lecture Notes in Computer Science. **6**(4) (2002) 285–300
8. Qin, L., Atluri, V.: Evaluating the validity of data instances against ontology evolution over the semantic web. Information and Software Technology. **51**(1) (2009) 83–97
9. Liang, Y., Alani, H., Shadbolt, N.: Ontology change management in protégé. In Proceedings AKT DTA Colloquium, Milton Keynes, UK. (2005)
10. Gruhn, V., Pahl, C., Wever, M.: Data model evolution as a basis of business process management, Springer (1995) 270–281
11. W3C: Owl web ontology language guide, w3c recommendation 10 february 2004 http://www.w3.org/TR/owl-guide.
12. Plessers, P., De Troyer, O., Casteleyn, S.: Understanding ontology evolution: A change detection approach. Web Semantics: Science, Services and Agents on the World Wide Web. **5**(1) (2007) 39–49
13. Javed, M., Abgaz, Y., Pahl, C.: A pattern-based framework of change operators for ontology evolution. The 4th International Workshop on Ontology content OnToContent 2009. Springer-Verlag. LNCS Series. (2009)
14. Stojanovic, L., Maedche, A., Stojanovic, N., Studer, R.: Ontology evolution as reconfiguration-design problem solving. Proceedings of the 2nd international conference on Knowledge capture (2003)
15. Oliver, D.E., Shahar, Y., Shortliffe, E.H., Musen, M.A.: Representation of change in controlled medical terminologies. Artificial Intelligence in Medicine. **15**(1) (1999) 53–76
16. Haase, P., Stojanovic, L.: Consistent evolution of owl ontologies. The Semantic Web:Research and Application **3532** (2005) 182–197
17. Holohan, E., McMullen, D., Melia, M., Pahl, C.: Adaptive courseware generation based on semantic web technologies. In: Proceeding of the International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL2005) at the Twelveth International Conference on Artificial Intelligence in Education (AIED2005), IOS Press (2005)