Improving Document Exchanges in the Supply Chain

Jamel Eddine Jridi$^{(\boxtimes)}$ and Guy Lapalme

Department of Computer Science and Operations Research,
Université de Montréal, Montreal, QC, Canada
{jridijam,lapalme}@iro.umontreal.ca

Abstract. In order to help businesses to communicate fruitfully, we present a solution based on ontology alignment for integrating business documents. We focus on detecting and resolving semantic conflicts encountered during the integration process due to different terminologies used in xCBL, cXML and RosettaNet. Our contribution is to benefit from research in the ontology alignment area and considered as empirical study to test if alignment solution can overcome the heterogeneity problems between business systems. As case study, we apply alignment on purchase order ontologies, a common task of the supply chain.

Keywords: Ontology alignment · Semantic heterogeneity · Document exchange · B2B standards · e-business · Interoperability · Semantic web

1 Introduction

Document exchanges in business-to-business has been dominated for a long time by traditional exchange standards such as UN/EDIFACT. xCBL[1], RosettaNet[2] and cXML[3] have been developed based on XML technologies for managing business documents mainly in the supply chain area and overcoming the traditional standards.

Given the XML limits, we proposed, in the 15th International Conference on Enterprise Information Systems (ICEIS), an approach to integrate semantics based on an automatic mapping of existing B2B standards written in DTD or XML Schema to an ontological representation using the OWL language [1]. However, several issues related to heterogeneity, and other issues still need to be addressed. As running example, we consider a company, called HomeSecurity, that offers a remote monitoring configuration and installation of security systems. If it needs to buy monitoring cameras to be installed at a customer's site, it would send electronic purchase order documents with other business partners, such as CameraSystems.

HomeSecurity uses xCBL, but CameraSystems expects documents in RosettaNet [2]. Some terminological heterogeneity exists between these ontologies, e.g. the concepts `ContactInformationType` of RosettaNet and `PartyType` of xCBL are similar even though their names are not. Although these ontologies cover the same area of interest, the definition and representation of concepts differ which hinders communication and interoperability between business applications. Moreover, RosettaNet provides more details in the description of purchase orders compared to xCBL. For example, RosettaNet describes the payment process in details using concepts such as `PartPaymentType` and `PrePayment DetailType` representing payment methods. By cons, xCBL represents it with the single `PaymentMethodType` concept. So, this problem generates a conceptual heterogeneity between xCBL and RosettaNet due to different document models for the same domain of interest, each being designed at various levels of detail. The interoperability between businesses becomes more difficult due to different types of heterogeneity between ontologies.

The example presented above shows many heterogeneity types between studied standards. In this paper, we propose ontology alignment as solution to resolve them. We apply alignment on purchase order which is a common task in the supply chain addressed by cXML, xCBL and RosettaNet standards. Also, we discuss if they help business systems to communicate fruitfully between different formalisms. No work has yet addressed the alignment of business documents, due to a lack of ontologies that models B2B standards. Several matching-based studies have been proposed in the field of e-business specifically on business standards from different points of view: business process [3,4], or web services [5,6] but not on business documents which are the main support of communication between business partners [7].

Our contribution is not to present a new alignment approach but to benefit from research in the ontology alignment area and show if these alignment techniques can help business systems to communicate fruitfully and overcome the heterogeneity problem by finding correspondences between the entities (classes, properties, instances, etc.) of two ontologies [8]. We focus in this paper on detecting and resolving semantic conflicts encountered during the integration process of xCBL, cXML and RosettaNet documents due to differences in terminology.

This paper presents a continuity of our work [1], published in ICEIS 2013, in order to get a complete data integration framework which benefits from ontology advantages and reduce heterogeneity using ontology alignment.

We define, in Sect. 2, the notions studied in this paper. A summary of the related work in ontology alignment in business domain is given in Sect. 3. In Sect. 4, we present the ontology alignment techniques and discuss the results of their application on purchase order document exchanges in B2B standards. Finally, we conclude and suggest future research directions in Sect. 5.

2 Background

In this section, we define important notions studied in this paper.
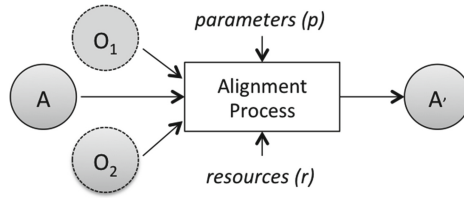
Fig. 1. Ontology alignment process [9].

Ontology provides a shared and common vocabulary for a domain of interest and a specification of the meaning of its terms [9]. It allows users to organize information into a hierarchy of concepts, to describe their relationships, and to make semantics machine processable, not just readable by a human.

Ontology Alignment. Ontology Alignment is a process that takes as input two ontologies and finds semantic links or a mapping between the entities of these ontologies (classes, properties, instances, etc.).

$$\texttt{A'} = \texttt{f}(O_1, O_2, \texttt{A, p, r})$$

According to Euzenat et al. [9], the alignment process, shown in Fig. 1, can be formally presented as a function `f` which takes as input; a pair of ontologies O_1 and O_2 , a starting alignment `A` that must be completed by the alignment process, a set of parameters `p` as weights and thresholds, and external resources `r` (e.g. WordNet [10]) and returns an alignment `A'` between these two ontologies.

RosettaNet is a consortium which provides a global forum for suppliers, customers, and competitors to do business and collaboration in an efficient and profitable manner. To manage business activities, RosettaNet formalizes Partner Interface Processes (PIP) with either Data Type Definition (DTD) format or XML Schema, and defines business processes between trading partners. PIPs are organized into eight groups of core business processes called clusters, themselves further grouped into segments. Each segment includes several PIPs [2]. In this paper, we use as a running example the PIP3A4, named `Request Purchase Order`, that enables a buyer to issue a purchase order and to get a confirmation from a seller.

XCBL is a collection of XML specifications of business documents used in e-business. Technically, xCBL contains 44 documents and uses multiple namespaces, where each namespace represents a functional area [11]. `order management` is one of the functional areas of xCBL containing documents related to manage purchase orders, e.g. `OrderRequestType` and `OrderConfirmation Type`. In xCBL, this process begins when the buyer sends to a seller the `Order RequestType` document. Finally, the document `OrderConfirmationType` will be sent from the seller to the buyer to confirm order details.

CXML provides business documents validated with DTDs [12]. Businesses use cXML to communicate purchase orders. Processing purchase order begins when the customer sends an `OrderRequest` document, an electronic version of an order that contains customer information, billing, delivery information, etc. Once the order is received by the seller, he returns an acknowledgment with an `OrderResponse` document that indicates whether the request was successfully received. The order confirmation will be made by sending a `Confirmation Request` document to the customer.

3   State of the Art

The use of formal semantics for business to business (B2B) communication based on ontologies should help overcome the problem of standards integration by the use of ontology alignment.

In recent years, several studies have focused on ontology alignment but with different perspectives such as dealing with business documents or identifying correspondences between business models or web services sharing the same behaviour.

Schubert et al. [4] studied electronic collaboration between companies in the supply chain. Their project seeks to develop a framework that facilitates collaboration in B2B by building a reference business model combining the shared patterns between companies among a set of business scenarios collected from several german enterprises. This study focuses on the business scenarios at an abstract level in the business process without discussing the causes of heterogeneity between businesses. However, the emergence of business standards, such as RosettaNet, have partially solved this problem. In our case, we focus on the causes for heterogeneity in document exchanges as the principal source of communication and heterogeneity between enterprises.

Zhu et al. [5] and Kim et al. [6] propose alignment approaches for searching similar business process sharing a similar behaviour, especially web services. Zhu et al. [5] define a metric called Business Process Similarity (BPS) based on structure differences and edit distance between two business processes. Kim et al. [6] apply semantic matching on a formal semantics of business processes represented by ontologies. These approaches stay at an abstract level and have not been applied to existing standards to evaluate their effectiveness in a practical environment.

García et al. [13] propose to align the ebXML Business Process (ebBP) ontology to the Business Process Execution Language for Web Services (BPEL-WS) ontology using OWL Ontology Align tool [14] which is unfortunately not available for testing. The choice of the alignment system is not discussed and the performance of the alignment was not evaluated.

Unlike previous approaches, we focus primarily on the data representation in business documents and we try to solve the heterogeneity problem in document exchanges using existing alignment techniques.

4   Ontology Alignment in the Business Domain

Ontology alignment is an interesting tool for integrating multiple knowledge bases, e.g. between business documents, by determining correspondences between concepts (properties and classes) of two ontologies to be aligned. Most alignment systems were evaluated on a benchmark tests during OAEIs or applied to bioinformatic ontologies such as ASMOV [15]. In our study, we depart from these types of tests to deal with other ontologies related to the business domain specifically the supply chain and discuss if they help business systems to communicate fruitfully between different formalisms.

In this section, we review alignment system methods and discuss the highest results presented in the recent Ontology Alignment Evaluation Initiative (OAEI) campaigns[4]. Section 4.2 describes the experimental results of OLA_2 for aligning Purchase Order ontologies in RosettaNet, xCBL and cXML, to reduce the heterogeneity between those documents.

4.1   Alignment Systems Evaluation

Many alignment systems have been proposed in the OAEI campaigns from 2004 to 2014 which differ in methodology and in the type of data used: strings (terminological), structure (structural), data instances (extensional) or models (semantics) [8]. Moreover, systems differ in their alignment strategies (e.g. graph-based approaches) or similarity measures (e.g. lexical, structural, extentional, etc.) such as AOTL [16], OLA_2 [17], Falcon [18] and ASMOV [15]. According to Shvaiko et al. [8], most systems based on terminological and structural techniques seldom use extensional or semantic methods. The quality of each system is measured by their F-measure metric in the OAEI campaigns.

We now review state-of-the-art alignment systems that have been found to be the most effective in the benchmarks tests of the recent Ontology Alignment Evaluation Initiative (OAEI) campaigns.

Shvaiko et al. [8] compared alignment systems participating to the OAEI campaigns found that OLA_2 [17] had the best performance with a of 0.71. Even though other systems got better performance in 2014 [19], we choose OLA_2 because of the availability of its source code and that it remains still an excellent system.

4.2   Experimentation

OLA_2 is an automatic ontology alignment system which takes as input two OWL ontologies, written in OWL Lite or OWL DL, and computes a set of correspondences between their entities. OLA_2 is based on a graph representation that allows similarity to be expressed as graph alignment computed as matrix operations. Ontologies are represented in OLA_2 as an oriented graph with vertices

---

[4] http://oaei.ontologyalignment.org.

corresponding to entities and edges to inter-entity relationships. This graph is used to build a similarity graph by adding weight to each arc representing the similarity between two entities given by their string and lexical distances. OLA_2 computes the similarity measures through an iterative approximation process that first considers the lexical similarity measures using Levenshtein distance[5], and then the structural similarity of the compared ontologies [20]. The similarity of a node at a given iteration is based on similarities of adjacent nodes of the previous iteration until the similarity values of the graph nodes remain the same. Djoufak et al. [20] provides more technical details on the OLA alignment steps.

To evaluate the performance, alignments are compared with a set of reference alignments that we manually developed. We use the precision (P), recall (R), and the F-score (F) metrics calculated as follows:

$$P = \frac{TP}{TP + FP} \qquad R = \frac{TP}{TP + FN} \qquad F = 2 * \frac{P * R}{P + R}$$

where TP denote the number of good correspondences correctly identified, FP is the number of correspondences incorrectly identified, and FN represents the number of correct correspondences not identified by the algorithm. To our knowledge, no work has yet addressed the alignment of business standards, due to the lack of ontologies that define them.

A case study, we apply alignment on purchase order ontologies. The reason in using these ontologies is that the management of purchase orders is the main and common task of the supply chain addressed by cXML, xCBL and RosettaNet standards. Moreover, these ontologies represent the principal support that defines most entities (more than 70 %) used in those standards. Table 1 gives a quantitative study of these ontologies (`Request`) and (`Confirm`) using metrics with `noc` is the number of ontology classes, `nodp` is the number of data properties, and `noop` is the number of object properties.

Table 1. Quantitative study of purchase order ontologies (`Request`) and (`Confirm`) in RosettaNet, xCBL and cXML using metrics.

	RosettaNet		xCBL		cXML	
	Request	Confirm	Request	Confirm	Request	Confirm
`noc`	151	143	66	33	17	27
`nodp`	264	252	174	68	34	36
`noop`	126	130	91	35	18	25
Total	523	525	331	136	69	88

We see from Table 1 that the purchase order ontologies of RosettaNet has more entities than xCBL and cXML, which seems logical as RosettaNet models business document in more details.

---

[5] The Levenshtein distance is the minimum number of substitutions, additions and deletions of characters required to transform one string to another.

Table 2. The alignment results of OLA_2 done on ontologies of purchase order documents in xCBL and cXML with RosettaNet PIP3A4.

	xCBL		cXML	
	Request	Confirm	Request	Confirm
Number of entities	331	136	69	88
TP	193	86	37	50
FP	138	49	21	24
TP+FN	298	122	62	81
Precision (P)	58 %	64 %	64 %	67 %
Recall (R)	65 %	72 %	60 %	62 %
F-score (F)	**61 %**	**68 %**	**62 %**	**64 %**

OLA_2 takes as input two OWL ontologies in RDF/XML. First, a pre-processing step is performed to deal with `owl:import` constructs, which are not supported by OLA_2; it integrates recursively all entities in the imported files with a Java program that uses OWL API[6]. The performance of OLA_2 is illustrated in Table 2.

We obtain an average F-score of 64 % in the alignment of purchase order ontologies of xCBL and cXML with RosettaNet (see Table 2). We consider that the alignment results are promising although we limited ourselves to entity names and their structure, and that the vocabulary used in the studied ontologies is different.

Furthermore, the structure has an impact in the alignment results and helps detect good alignment even different names of entities. To verify this assumption, we aligned using only the Levenshtein distance which resulted in an average performance of 43 % for most of the standards, which is much less than the 64 % obtained with OLA_2.

Many false positives were identified by Levenshtein method based on entities names ignoring their structure. Two strings can be very similar but differ in semantics, e.g. the classes `ProductReferenceType` and `TaxReferenceType`. In addition, there are some strings that are not similar in their Levenshtein distances, even though they are semantically similar. So the use of other sources of information, such as the ontology structure, can help reduce this heterogeneity.

Table 3 shows alignment examples of entities of RosettaNet with xCBL and cXML which have been correctly aligned with OLA_2. For example, The xCBL data property `StreetSupplement2` was misaligned with Levenshtein distance, but correctly aligned with OLA_2.

Although, the heterogeneity, presented in the running example of HomeSecurity and CameraSystems illustrated in Sect. 1, are resolved. The RosettaNet class

---

[6] owlapi.sourceforge.net.

Table 3. Alignment of some entities of RosettaNet with xCBL and cXML.

	xCBL	cXML
PhysicalAddress	NameAddressType	Address
PartnerDescriptionType	OrderRequestPartyType	Request
PurchaseOrderRequestType	OrderRequestType	OrderRequestHeader
AddressLine2	StreetSupplement2	Street

`ContactInformationType` have been correctly aligned with the class `PartyType` of xCBL, also for classes `PaymentMethodType` and `PartPaymentType`.

Discussion. OLA_2 is a generic structure-based method that does depend neither from comments nor from entity names. However, the major limitation of this algorithm is the computation time and the memory space needed relative to the terminological techniques. According to Djoufak et al. [20], these requirements are due to the matrix operations that are costly in time and memory especially in the case of large ontologies. Unfortunately, OLA_2 has heavy memory space requirements when the size of ontologies is growing, so it required a powerful machine.

We note that the use of alignment in business domain, especially for B2B standards, is promising and can reduce the heterogeneity in the document exchanges even if they don't share the same vocabulary. Furthermore, OLA_2 has demonstrated good performance other than testing benchmarks and the alignment results could be further improved by using a thesaurus describing the terms of business domain or using document instances.

On the other side, alignment can be used to reduce heterogeneity between enterprise systems which use the same standard. In the case of RosettaNet ontologies, enterprises can use the same ontologies differently as the messages contain many optional elements that are not implemented by each company. This problem also exists in the case of xCBL and cXML standards. In this case, manual task is required to interpret new information sent by a new partner. When the number of partners increases, the exchange becomes more and more difficult [21]. So, it is assumed that OLA_2 may give better results in this case, because it deals with ontologies of the same standards having the same terminology and structure. But, this assumption remains to be tested, as future work, on real collaboration scenarios between companies.

Furthermore, we would like to design, in future work, a thesaurus to identify, classify and links the different technical terms related to the business domain. This hierarchical structure could reflect the relationships between topics, each entry representing a set of terms with a common semantics. To create the thesaurus, we must develop a vocabulary by parsing the business documents of B2B standards to extract the basic concepts and their relations, e.g. the synonymy, the hyponymy or the hyperonymy.

5   Conclusion

In this paper, we presented using a motivating example different types of heterogeneity (terminological and conceptual) found in business document exchanges. Document models have been developed independently by different organizations to meet the needs of enterprises for collaborating.

We presented ontology alignment as a promising solution to reduce heterogeneity encountered during the integration process of xCBL, cXML and RosettaNet documents. We performed detailed alignments among the generated OWL ontologies in order to get a complete data integration framework. We experimented with a structure-based algorithm to determine the best correspondences between ontologies of purchase order management in cXML and xCBL standards with the RosettaNet documents. This approach provides a promising results and can be applied to other cases as well as supply chain.

With semantic web technologies, industry acquires several benefits such as more efficient business interoperability and information exchange. We demonstrated that the alignment can improve the interoperability between business systems even if using different terminology.

References

1. Jridi, J.E., Lapalme, G.: Adapting RosettaNet B2B standard to semantic Web technologies. In: 15th International Conference on Enterprise Information Systems (ICEIS), Angers, France, pp. 443–450 (2013)
2. RosettaNet Program Office, Overview Clusters, Segments, and PIPs, December 2011
3. Cardoso, J., Bussler, C.: Mapping between heterogeneous XML and OWL transaction representations in B2B integration. Data Knowl. Eng. **70**(12), 1046–1069 (2011)
4. Schubert, P., Legner, C.: B2B integration in global supply chains: an identification of technical integration scenarios. J. Strateg. Inf. Syst. **20**(3), 250–267 (2011)
5. Zhu, J., Pung, H.K.: Process matching: a structural approach for business process search. In: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns, COMPUTATIONWORLD 2009, pp. 227–232, Computation World. IEEE (2009)
6. Kim, G., Suhh, Y.: Ontology-based semantic matching for business process management. ACM SIGMIS Database **41**(4), 98–118 (2010)
7. Glushko, R.J., McGrath, T.: Document Engineering: Analyzing and Designing Documents for Business Informatics and Web Services. MIT Press, Cambridge (2005)
8. Shvaiko, P., Euzenat, J.: Ontology matching: state of the art and future challenges. IEEE Trans. Knowl. Data Eng. **25**(1), 158–176 (2013)
9. Euzenat, J., Shvaiko, P.: Ontology Matching, vol. 18. Springer, Heidelberg (2007)
10. Miller, G.A.: WordNet: a lexical database for English. Commun. ACM **38**(11), 39–41 (1995)
11. xCBL User's Guide Version 4.0 (2003)
12. Ariba, cXML User's Guide Version 1.2.025, April 2014

13. García, R., Gil, R.: Facilitating business interoperability from the semantic web. In: Abramowicz, W. (ed.) BIS 2007. LNCS, vol. 4439, pp. 220–232. Springer, Heidelberg (2007)
14. Zhdanova, A.V.: Towards a community-driven ontology matching. In: Proceedings of the 3rd International Conference on Knowledge Capture, pp. 221–222. ACM (2005)
15. Jean-Mary, Y.R., Shironoshita, E.P., Kabuka, M.R.: Ontology matching with semantic verification. Web semant. Sci. Serv. Agents World Wide Web **7**(3), 235–251 (2009)
16. Khiat, A., Benaissa, M.: Aot/aotl results for oaei (2014)
17. Euzenat, J., Loup, D., Touzani, M., Valtchev, P., et al.: Ontology alignment with OLA. In: Proceedings of the 3rd ISWC2004 Workshop on Evaluation of Ontology-Based Tools (EON), pp. 59–68 (2004)
18. Hu, W., Qu, Y., Cheng, G.: Matching large ontologies: a divide-and-conquer approach. Data Knowl. Eng. **67**(1), 140–160 (2008)
19. Dragisic, Z., Eckert, K., Euzenat, J., Faria, D., Ferrara, A., Granada, R., Ivanova, V., Jiménez-Ruiz, E., Kempf, A.O., Lambrix, P., et al.: Results of the ontology alignment evaluation initiative (2014)
20. Djoufak-Kengue, J.-F., Euzenat, J., Valtchev, P., et al.: Alignement d'ontologies dirigé par la structure. In: Actes 14e journées nationales sur langages et modèles à objets (LMO), pp. 43–57 (2008)
21. Haller, A., Gontarczyk, J., Kotinurmi, P.: Towards a complete SCM ontology: the case of ontologising RosettaNet. In: Proceedings of the 2008 ACM Symposium on Applied computing, pp. 1467–1473. ACM (2008)