

Articulating Autonomously Distributed Electronic Product Catalogues for Constructing Dynamic CONEX Net

Jingzhi Guo, Chengzheng Sun and David Chen

School of Computing and Information Technology, Griffith University, Australia
{J.Guo, C.Sun, D.Chen}@griffith.edu.au

Abstract

The integration of ad hoc electronic product catalogues (EPCs) is an important issue in constructing dynamic global electronic markets. A particular issue in this integration process is that millions of ad hoc EPCs are autonomously distributed around the world on Internet and fully behave in their own ways. This leads to the unknown sources of their schemas and semantics and cannot adopt traditional product ontologies to mediate heterogeneous product information. To investigate the specific issues, this paper articulates both semantic conflict relations of ad hoc EPCs and their unique uneven distribution properties. By these articulations, it has developed a novel ConexNet on which uneven distributed and autonomous EPCs are aligned and linked in a semantic conflict-free way. Features are provided to examine how ConexNet is made to resolve semantic conflicts. Measuring criteria are proposed to evaluate the quality of ConexNet.

1. Introduction

The integration of ad hoc *electronic product catalogues* (EPCs) is an important issue in constructing dynamic *global electronic markets* (GEM) [2][5]. *Ad hoc EPCs* refer to inconstant, small-scale, irregular and heterogeneous EPCs with unknown numbers [3]. A particular issue in integration process is that ad hoc EPCs autonomously distributed around the world on Internet behave in their own ways, which increases complexity of ad hoc EPC integration and makes difficult in resolving semantic conflicts between them. The complexity is increased when the interdependent ad hoc EPCs request business interoperation, where the semantic conflicts arise in three aspects: *structure, concept* and *context*.

In this paper, we aim to articulate the complexity of autonomous component EPCs from a perspective of semantic conflicts and distribution characteristics through the methodology of *articulation* - a part of deconstruction and reconstruction theory [6] for integrating heterogeneous electronic product catalogues. Articulation makes integration work simple and manipulable. By articulation, we expect two results: (1) semantic conflicts between interdependent component EPCs are orderly classified for diagnosing the problems arisen from integrating distributed component EPCs, and (2) the distributed component EPCs are orderly clustered in global electronic markets in terms of a concept exchange net, called *ConexNet*, that meet the requirements of flexibility, evolvability, exactness and financial cost minimisation [2]. The two results are correlated: the former constrains that the integration of component EPCs should resolve the observed semantic conflicts while the latter requires that the distributed component EPCs integrated together should not only

resolve semantic conflicts but also be clustered and well aligned in a required manner. The purpose of this paper is to investigate what are the problems general to ad hoc EPCs and how they are correlated in terms of EPC distribution.

The rest of the paper will articulate the semantic conflict relations between component EPCs, analyse and cluster distributed component EPCs in a global electronic market, formulate the clustered component EPCs as a ConexNET, and finally provide the measurement criteria of ConexNet.

2. Articulate Semantic Conflict Relations between Component EPCs

Autonomy leads to heterogeneity between component EPCs, which further causes product information conflicts. Roughly, information conflicts can be distinguished between the two types of *schematic conflicts* and *semantic conflicts*. The former does not care the confusion of human's interpretation on the exchanged data. The latter, on the other hand, arise from the incongruent semantic assignments to the data representations and the different interpretations on the exchanged data.

In integrating ad hoc component EPCs, we assume that the EPC information sources are ad hoc. It means that they have no available (or obvious) schemas, and have no inferable semantics on terms that are traditionally called *product ontologies*. Besides, ad hoc sources are uncertain in number, which may be in millions and their semantics of schemas (may even not exist) are unknown. By this clarification, there is a need to articulate what semantic conflicts exist in ad hoc EPC integration. In our investigation, we articulate semantic conflicts related to ad hoc EPCs into three types: *structure conflicts*, *concept conflicts* and *context conflicts*.

Structure conflicts, which refer to the different effects of semantic inconsistency between component EPCs. It could be categorised in two types. *Vocabulary classification conflicts* refer to the different taxonomy methodologies for classifying the concepts of EPC contents. The differences may stem from the different selections of product content domains, or from the different perspectives of the semantic relationships between the concepts of product contents. *Product modelling conflicts* refer to the different abstraction methodologies for modelling products. Given a same category of products, different EPCs have different abstraction models for product features.

Concept conflicts, which refer to the semantic inconsistency effects from inconsistent EPC items. Given a classified EPC, an EPC item denotes a concept or a semantic unit, which can be a category, a product or an attribute. Concept conflicts can be categorised into three types. *Naming conflicts* arise from the unmatched concept names or concept identifiers in the contexts

where concept names or identifiers are used to identify EPC items. The recognised conflicts are synonyms and homonyms. In some proposed solutions, mapping tables are provided to bridge synonyms and homonyms. In real world, component EPCs are emergent, meaning that a mapping will not be effective unless it provides the dynamic mapping of new synonyms and homonyms. *Annotation conflicts* refer to the different semantics assignment to a concept or an item. It is similar to homonyms but arises from the different semantic definitions to an EPC item. This type of conflicts affects the interpretation of product metadata, especially for many intelligent systems that infer meanings through keywords. *Value structure conflicts* refer to the adoption of different measurement systems for a value structure. In complex value structures, this type of conflicts also arises from the different definitions of an attribute value. Limited solutions are ontology mapping and data registry [8][10]. **Context conflicts**, which refer to the different semantic inconsistency effects from the different semantic contexts where EPC representations are modelled in different *semantic communities* [4]. We identify three types of context conflicts between component EPCs. *Natural language conflicts* refer to the differences in using natural languages for defining the semantics of product information. This type of conflicts can roughly be classified into three levels: natural language conflicts, dialect conflicts and proprietary language conflicts. *Concept referencing conflicts* refer to the different contextual references of EPC items, which mean that different component EPCs may reference to different schemas, definitions and product standards for their EPC items. Concept referencing conflicts will result in misunderstanding between component EPCs. *Implicit meaning conflicts* refer to the partial representation of the EPC features that are important to other component EPCs. This type of conflicts arises from the implicit assumption that some EPC features are commonly known to all and can be omitted in the EPC representations.

3. Articulate Distributed EPC Representation

The characteristic of high distribution of component EPCs increases complexity between component EPCs, which leads to semantic conflicts more severe. Schmidt et al [11] explain that interdependent systems tend to be more complex than independent systems because of the involvement of coordination work arrangement for supporting indirect and distributed cooperative work relationships. Hence, articulation work is required to reduce complexity so that the involved work arrangement could be orderly managed. By articulating the unevenly distributed EPC representations, we align common, local and source representations in terms of structures, concepts and contexts in a novel

ConexNet as a concept exchange network topology where CONEX framework [6] can be situated for resolving semantic conflicts.

3.1. Uneven Distribution of Component EPCs

The component EPCs are unevenly distributed between electronic markets. More specifically, the interdependent component EPCs, which are carriers and transmitters of product information between electronic markets, are “fragmented by geography, which creates inefficient markets and uninformed buyers and sellers” [9]. The *geography* here refers to the differences of network locations that component EPCs are created on Internet and related to various fragmented markets. Geography leads to the heterogeneous designs and uses of component EPCs,

Geographic discrepancies often arise from three aspects. *Natural causes* refer to geographic region, country border, natural language, customs and culture. *Marketing strategies* refer to the deliberate designs of certain product features, unnatural languages and descriptions that could cater for the need of a small portion of customers (e.g. product differentiation [1]). *Resource constraints* refer to the legacy systems, technical difficulties and financial limitations of firms [5]. These causes lead to an *uneven distribution* of component EPCs in both contents and representations over Internet. The uneven distribution of component EPCs raises an issue for EPC representation integration, that is, in which way component EPCs should be connected to resolve semantic conflicts?

3.2. Cluster Unevenly Distributed EPCs

Cluster analysis is “a convenient method commonly used in many disciplines to categorise entities ... into groups that are homogeneous along a range of observed characteristics” ([12]:39). It can provide the underlying systems with transparency in architecture and communication, and can accommodate heterogeneity in various aspects.

We cluster component EPCs based on five criteria: languages, geographical locations, industrial classification schemes, geo-demographic factors and resource constraints. This clustering base provides an angle of observing real-world electronic markets that could help build cost-effective EPC integration system accommodating heterogeneous component EPCs.

Language and *geographical locations* as our first two criteria lie in the fact that component EPCs are represented by natural languages for business communication between different geographical locations through global languages (e.g. English), national language (e.g. Chinese), or by special community lan-

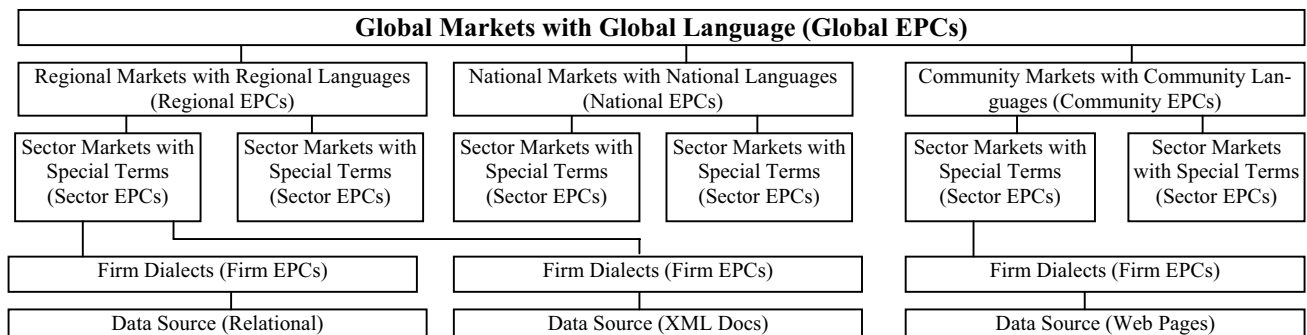


Fig. 1: Illustration of hierarchical distribution of component EPCs

guages (e.g. supply chain specifications or vocabularies). *Industrial classification schemes* and *geo-demographic factors* divide component EPCs into industrial sectors towards different producers and consumers. These two criteria are the two facets of representing products: supplier-based versus consumer-based sectors. They lead to finer clusters of component EPCs over languages and geographical locations. *Resource constraints* are criteria related to internal situations of a firm. They reveal the ability of firms what component EPCs they are able to provide and accept for product information exchange. These constraints are data sources of component EPC, connections to legacy systems and technical and financial capabilities of EPC reengineering for adapting to external component EPCs.

The clustering of component EPCs provides an insight into the uneven distribution of component EPCs: there is a *hierarchical relationship* between the uneven distributed component EPCs. This relationship can be depicted in Fig. 1 where heterogeneous component EPCs are hierarchically arranged and every two of adjacent component EPCs are linked by their respective *context line*. The hierarchical distribution of component EPCs is consistent with the natural market distribution: global markets, regional markets, national markets, sector markets and firms.

3.3. Align Distributed EPCs into ConexNet

The *hierarchical distribution of component EPCs* characterises the levels of GEM where product information is exchanged. In this part, we abstract the Fig. 1 into multi-sets of hierarchical contexts shown as in Fig. 2. This abstraction characterises the architecture of product information exchange along the layers of

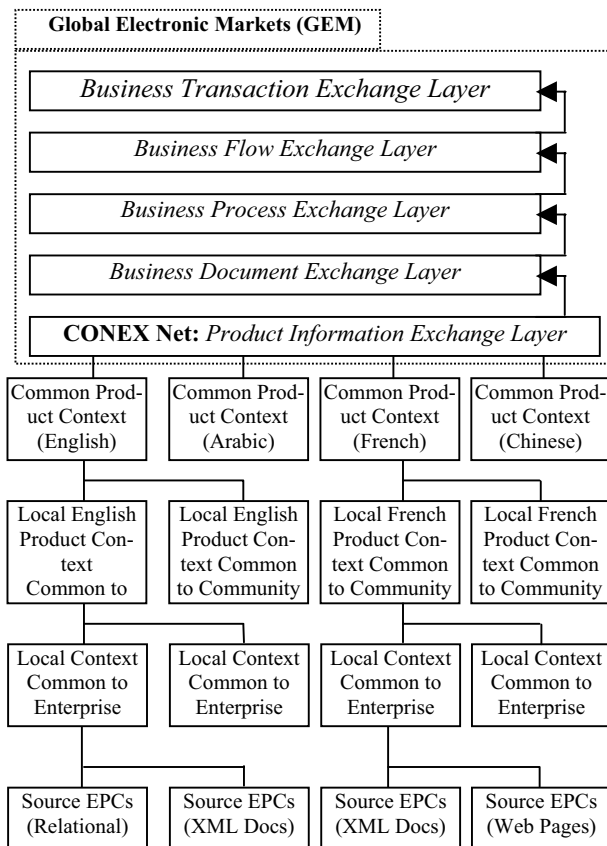


Fig. 2: Relation between GEM and ConexNet

GEM. The abstraction divides product information exchanged on GEM into common context (a set of common catalogue), local context (multi-sets of local catalogues) and source context (multi-sets of source EPCs). This abstraction presents an integrated global piece of product information representation.

A *ConexNet* (Fig. 2) is defined as a set of integrated component EPC representations that are expressed as a set of contexts, which are hierarchically arranged over Internet. It characterises the properties of component EPCs that are unevenly distributed in different market segments according to their commonality and locality. These common, local and source contexts together with the operations on them present a global concept exchange network for transporting product information between heterogeneous component EPCs. This network is the foundation layer of GEM, over which the higher-layer business representations are enabled for further investigations (e.g. business documents, business processes, flows and transactions). A ConexNet consists of two types of components: *context representation components* and *context connection components*. The former includes *common contexts*, *local contexts* and *source contexts* (expressed as squares), while the latter includes mapping maps between contexts in terms of source-local context mapping, local-common context mapping and common-common context mapping (expressed as connection lines between squares) [3][6]. In ConexNet, source, local and common contexts are three types of heterogeneous component EPCs that are created from different semantic communities. A specification for each of them could not automatically resolve the semantic conflicts between them because of *modality judgment* [7]. For this reasons, context maps between them must be collaboratively designed for resolving semantic conflicts. In CONEX approach [6], three types context maps are collaboratively created to resolve semantic conflicts between different contexts. They are *source-local context map*, *local-common context map* and *common-common context map*. Context maps as connection components between heterogeneous contexts are critical in constructing ConexNet for resolving semantic conflicts and exchanging product concept.

4. Features of ConexNet

We abstract ConexNet as a network topology shown in Fig. 3, where nodes of source, local and common contexts are renamed as two types of distinctive nodes: *source node* (S) and *mediation node* (M). By joining in ConexNet as a node, participants can subscribe or provide services for integrating heterogeneous product concepts.

First, ConexNet is a generalisation hierarchy, where directly linked higher-level nodes are parents. Concepts that are created on parent nodes could be *localised* for reconstructing local concepts of child nodes through collaboration mechanisms [7]. The relationship between a parent and a child is a context inclusion relation, expressed as $ChildNode \subseteq ParentNode$, which means local concepts are *semantically included* in common concepts.

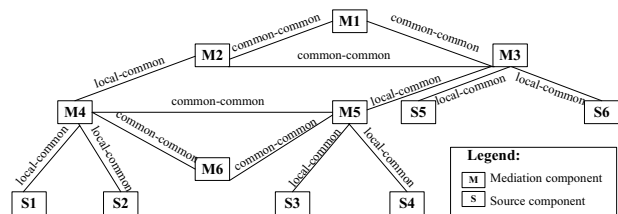


Fig. 3: A Generalisation of ConexNet

Second, nodes of ConexNet are crossly linked to form multiple product *concept supply chains* [3] where services of collaboration and transformation are provided. Multiple routes of product concept exchange provide redundancy and increase *delivery rate* between two sources. Third, a mediation node M can provide services of concept *creation* and *transformation* to either source or mediation child nodes S and M. It acts as a *collaborator* to create common concepts in its peer network with other peer nodes M. It can also take the role of a child node of other mediation nodes M. Fourth, a source node S is a terminal node of ConexNet, which is a product information source or receiver and cannot mediate product concepts for other component nodes. A source node S only connects to a local mediation node M such that $S \cap M \subseteq$ (a Same semantic community). This feature can be interpreted as a condition that local concept designers near sources should understand both schemas and semantics of the sources. Fifth, ConexNet is open, which means any number of components can join in for service subscription and/or provision. The openness allows any product concept mediation systems to be plugged into ConexNet.

The features of ConexNet have characterised the properties of autonomy and distribution of component EPCs. By ConexNet, semantic conflicts are resolved through both *concept collaboration functions* and *concept transformation functions* provided by mediation component nodes M.

5. Measurements of CONEX Net

Four measurement criteria of ConexNet are: flexibility, evolvability, exactness and financial cost minimisation for overall ConexNet.

Flexibility is measured against the indicators of the linkability of heterogeneous EPC structures, concepts and contexts. *Structure flexibility* refers to *concept combination ability*, which defines that different concepts classified and modelled on relational databases, XML documents, and ad hoc sources can be aligned in context maps as parts of mediation nodes, which connect source nodes. Given structure flexibility, heterogenous data structures can be interoperable. *Concept flexibility* refers to *concept substitutability*, which defines that concepts developed in different nodes of ConexNet can be substituted if any two of them are deemed semantically same via collaboration mechanisms. *Context flexibility* refers to *concept referencing ability*, which defines the ability of a given context of a certain node to establish the context reference with other contexts in other nodes.

Evolvability is measured against the indicators of new concept creation ability for whole ConexNet. It is defined as the ability that a new concept created in one node can also be recognised in other nodes. In our approach, evolvability is achieved through collaboration mechanisms of replication, translation and globalisation discussed in [7].

Exactness is measured against indicator of the accuracy of concept mapping between two linked nodes. The exactness assurance is provided by collaboration mechanisms [7].

Financial cost minimisation for overall ConexNet refers to that the overall systems developed on ConexNet should maintain minimal financial cost. This criterion is difficult to physically measure and is proved by a *specialisation theorem*: the sum of all individual works has larger cost than a single work for all individuals if there exist some individual works intersected or exactly same.

6. Conclusion

This paper has articulated the semantic conflict relations of autonomously distributed component EPCs into three sets corresponding to structure, concept and contexts. By this articulation, the unevenly distributed component EPCs are placed into a set hierarchically arranged nodes. A novel ConexNet is contributed to align the articulated component EPCs, where semantic conflicts could be resolved through context mappings. ConexNet includes a series of features: generalisation hierarchy, cross-linked concept supply chain, concept mediation and context transformation, terminal source nodes, and open service subscription and provision. These features enable product concepts are created and exchanged in a flexible, evolvable, exact and cost minimal way. ConexNet is the higher-level topology of CONEX approach [6] that integrates heterogeneous, distributed yet interdependent ad hoc EPCs. For this reason, ConexNet emphasises on the aspects of linkability of its nodes and encapsulates the details of conflict resolution in specific CONEX approaches developed in [2][3][4][6][7]. By increasing the linkability, semantic conflicts are resolved between linked contexts.

ConexNet has signified a research direction of what should be a suitable topology for future global electronic markets. With this perspective, one of our future researches will be the investigation of applying ConexNet to fragmented electronic markets.

7. Reference

- [1] Dickson, P. and J. Ginter, "Market Segmentation, Product Differentiation, and Marketing Strategy", *Journal of Marketing*, Vol. 51, 1987, pp. 1-10.
- [2] Guo, J. and Sun, C., "Concept Exchange: Constructing Interoperable Electronic Product Catalogues in an Emergent Environment", in: *CEC'03: Proc. of IEEE Int'l Conf. on E-Commerce*, IEEE Computer Society Press, 2004, pp.165-172.
- [3] Guo, J. and Sun, C., "Context Representation, Transformation and Comparison for Ad Hoc Product Data Exchange", in *DocEng'03: Proc. of the 2003 ACM Symposium on Document Engineering*, ACM Press, 2003, pp. 121-130.
- [4] Guo, J. and Sun, C., "Context Representation of Product Data", *ACM SIGEcom Exchanges*, Vol. 4. 1, ACM Press, 2003, 20-28.
- [5] Guo, J. and Sun, C., "Global Electronic Markets and Global Traditional Markets", *Electronic Markets*, Vol. 14, No. 1, 2004, 4-12.
- [6] Guo, J., Sun, C. and Chen, D., "Deconstruction and Reconstruction of Heterogeneous Electronic Product Catalogues for Semantic Interoperation", *the 2004 IEEE Conference on E-Commerce Technology (CEC'04)*, July 6-9, 2004, San Diego, CA, USA.
- [7] Guo, J., Sun, C. and Chen, D., "Collaborative Integration of Product Data: Analysis, Strategy and Protocol", *International Conference on Electronic Commerce and Research (ICECR 2004)*, Dallas, Texas, USA, June 10-13, 2004 (Accepted).
- [8] Ide, N. and L. Romary, "A Common Framework for Syntactic Annotation", *Proceedings of ACL'2001*, Toulouse, 2001, 298-305.
- [9] Phillips, C. and Meeker, M. (chief authors) "*Collaborative Commerce, The B2B Internet Report*", April, Morgan Stanley Dean Witter, 2000.
- [10] Romary, L. (ed), "Implementing a Data Category Registry with ISO TC37", *Language Resource Management*, ISO TC 37/SC4 N055, ISO TC 37/SC 4/WG1 N26, 2003-06-23.
- [11] Schmidt, K. and Simmon, C., "Coordination Mechanisms: Towards a Conceptual Foundation of CSCW Systems Design", *Computer Supported Cooperative Work: The Journal of Collaborative Computing* 5, 1996, pp. 155-200.
- [12] *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, 1998.