Enterprise Information Systems
Vol. 3, No. 1, February 2009, 59–94

Collaborative conceptualisation: towards a conceptual foundation of interoperable electronic product catalogue system design

Jingzhi Guo*

University of Macau, Taipa, Macao

(*Received 17 September 2008; final version received 9 November 2008*)

The maintenance of semantic consistency between numerous heterogeneous electronic product catalogues (EPC) that are distributed, autonomous, interdependent and emergent on the Internet is an unsolved issue for the existing heterogeneous EPC integration approaches. This article attempts to solve this issue by conceptually designing an interoperable EPC (IEPC) system through a proposed novel collaborative conceptualisation approach. This approach introduces collaboration into the heterogeneous EPC integration. It implies much potential for future e-marketplace research. It theoretically answers why real-world EPCs are so complex, how these complex EPCs can be explained and articulated in a PRODUCT MAP theory for heterogeneous EPC integration, how a semantic consistency maintenance model can be created to satisfy the three heterogeneous EPC integration conditions and implemented by adopting a collaborative integration strategy on a collaborative concept exchange model, and how this collaborative integration strategy can be realised on a collaboration mechanism. This approach has been validated through a theoretical justification and its applicability has been demonstrated in two prototypical e-business applications.

Keywords: interoperable electronic product catalogue; collaborative conceptualisation; collaborative concept exchange; collaborative integration; collaboration mechanism; product data integration; PRODUCT MAP; semiotic analysis; product representation; product data management

1. Introduction

In the last decade, the e-marketplace has experienced a drastic transformation from the web presence of products to the back-end business interoperability between firms (Segev *et al*. 1995, Guo and Sun 2004). In this transformation, technologies such as product classification standards (e.g. UNSPSC.org and eclass.de), ontological engineering (Gruber 1993) and collaborative computing (Sun *et al*. 1998) have pushed the development of the e-marketplace, where sellers and buyers can match their supply and demand for business transactions (Bakos 1998).

In the process of e-marketplace construction, an important issue is how to integrate numerous heterogeneous electronic product catalogues (EPCs) of the participating firms. This issue was identified and defined in the research of Linche and Schmid:

> Most systems offered today are proprietary structures that lack interoperability and cross navigation. Despite the growing number of companies that present their products on the Internet, a global search for products and comparative analysis of their features is impeded by semantic differences between the EPCs. Thus, even though buyers enjoy broad access to different vendors' product specifications, integration and evaluation of product information still has to be performed manually. (Linche and Schmid 1998)

---

*Email: jzguo@umac.mo

ISSN 1751-7575 print/ISSN 1751-7583 online

DOI: 10.1080/17517570802610362
http://www.informaworld.com

This quotation precisely points out that the semantic differences of heterogeneous EPCs are a problem that affects interoperability between web-based firms.

1.1. Problem description

To demonstrate the problem consider, for example, two small and medium-sized enterprises (SMEs) that have individually designed their *ad hoc* EPCs for product information exchange, where *ad hoc* refers to no commonly acceptable design rules. The database of each EPC consists of the catalogue schemas that are actively used in all aspects of the individual SME's internal systems. Suppose that the two SMEs are trying to communicate with each other to sell and buy refrigerators. They may encode the refrigerator in two different product representations:

- SME_1: fridge (id: 222, clr: blue, prc: 300)
- SME_2: réfrigérateur (art: x111, couleur: bleu, prix: 300)

The immediate question is: are these two pieces of product information semantically the same or different? It is uncertain and depends on each SME's understanding in its situated context. If these two SMEs have ever co-operated with each other, they may know that they are semantically equal or not equal. If they have never co-operated before, they may have the following understandings:

- If SME_1 knows French, it may assume that SME_2 has a 'fridge' specification where its 'couleur = clr' is 'bleu = blue'. It is not sure whether the value '300' of its 'prix = prc' is the same as its own '300', because SME_1 may implicitly refers to '300' such as US$300 and cannot conclude whether the currency of SME_2 refers to US dollar, Euros or others. SME_1 also does not know what 'art' means.
- If SME_2 knows English, it may infer SME_1 has a 'réfrigérateur' specification if it understands 'fridge' as 'refrigerator'. However, this inference may be wrong if 'fridge' in SME_1 does not refer to 'refrigerator'. SME_2 cannot infer or understand the details of 'fridge' specification of SME_1 because 'id', 'clr' and 'prc' are only understandable in SME_1's own context.
- If SME_1 does not understand French or SME_2 does not understand English, both cannot understand the opposite side's refrigerator specification.

The above example shows that the issue of EPC interoperability is extremely complex. The interpretation of a piece of product information (i.e. a concept) is context dependent. It is impossible for a concept producer to imagine all contexts of concept consumers and a concept consumer has difficulty inferring correctly the contexts of concept producers. By a simple classification, heterogeneous product representations shown in the example have semantic conflicts in syntactic constructs (e.g. conflicts in 'id' and 'art'), semantics encoding (e.g. term conflicts in product, attribute and value naming) and context reference systems (e.g. different interpretation of '300').

Traditionally, the semantic conflict resolution between EPCs for product information interoperability is achieved using the following approaches:

- Mandatory standardisation of all involved EPCs, so all EPCs adopt same standards and thus no semantic conflicts occur when EPC information is exchanged;

- Automated mediation of all involved EPCs by creating shared domain-wide ontologies, so the intelligent mediators can automatically resolve semantic conflicts by applying shared ontologies.

However, some issues remain unsolved because not all SMEs adopt standards and not all semantic conflicts can be resolved by intelligent mediators. To resolve the remaining issues, this article aims at examining the effectiveness of standardisation and mediation approaches and providing an alternative collaborative conceptualisation approach to a conceptual foundation of designing an interoperable EPC (IEPC) for SMEs.

1.2. An overview of the collaborative conceptualisation approach

The collaborative conceptualisation approach thinks that the EPC world is complex. This world comprises heterogeneities of EPC structures, concepts and contexts. The existing approaches of standardisation and mediation cannot effectively cope with the complexity and heterogeneity for semantic interoperability. For the standardisation approach, the mandatory use of standards could reduce the effort of integrating heterogeneous EPCs, but it also increases difficulties of handling changes and contexts in various EPCs. The automated mediation approach could automate the conflict resolution process between heterogeneous EPCs, but intelligent mediators could not correctly infer the meaningful information exchanges if there are no semantic mapping rules found in the rule repository.

The task of the proposed approach is, thus, responsible for simplifying our access to the complex EPC world and resolving the semantic conflicts in communication, arising from EPC heterogeneity.

To keep the above promise, the collaborative conceptualisation approach proposes a novel theory about the structure, concept and context of complex EPCs. It thinks that the complex EPCs can be simplified into a set of generic but flexible EPC representation specifications, as opposed to the mandatory EPC standard forms. It also thinks that the collaborative EPC editing technique is a means of resolving semantic conflicts between heterogeneous EPCs, as opposed to the automated mediation through domain-wide shared ontologies.

Specifically, the collaborative conceptualisation approach to designing an interoperable EPC (IEPC) can be outlined in a deconstruction and reconstruction framework consisting of two processes of EPC deconstruction and EPC reconstruction, shown in Figure 1.

1.2.1. EPC deconstruction

EPC deconstruction is an articulation process of decomposing complex EPCs into simplified EPC constructs for easy handling. It includes a complexity analysis method and a PRODUCT MAP theory.

The complexity analysis method detailed in Section 3 analyses complex EPCs in both spatial and temporal dimensions. By this method, the EPC properties in the notions of distribution, autonomy, emergence, and interdependence could be fairly and accurately captured to understand heterogeneous EPCs.

The PRODUCT MAP, detailed in Section 4, is a theory on structure, concept and context for EPC representation to capture the EPC properties. It applies semiotic theory for the first time (Saussure 1966, Barthes 1968, 1972, Eco 1976) to articulate complex EPC phenomena and abstracts the complex EPCs into a set of simple and manageable representation constructs. With these constructs, it derives the conditions of integrating heterogeneous EPCs.

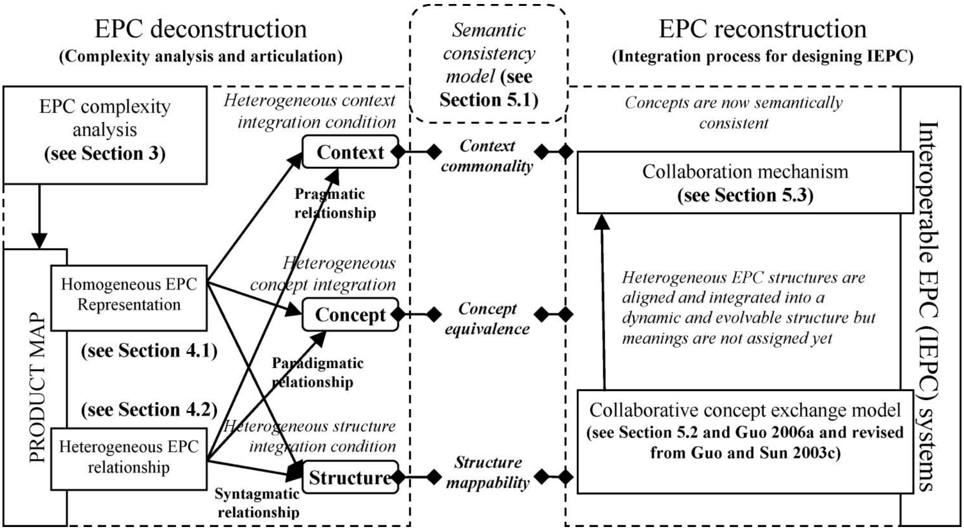


Figure 1. Outline of collaborative conceptualisation approach.

1.2.2. EPC reconstruction

EPC reconstruction is an integration process of re-building heterogeneous EPCs into an integrated yet interoperable EPC (i.e. IEPC) based on the heterogeneous EPC integration conditions. The IEPC is free of semantic conflicts and is the foundation of product information exchange between heterogeneous EPCs. EPC reconstruction consists of a semantic consistency model, a concept exchange model and a collaboration mechanism.

The semantic consistency model, detailed in Section 5.1, formalises the heterogeneous EPC integration conditions on heterogeneous structure, concept and context, shown in Figure 1. This model constrains the way in which heterogeneous EPCs should be semantically integrated by satisfying three semantic consistency properties of structure mappability, concept equivalence and context commonality.

The concept exchange model, detailed in Section 5.2, abstracts various types of real-world EPCs into a model to describe how real-world EPCs could be connected with each other. By this model, the path of product information exchange from one heterogeneous EPC to another could be dynamically determined.

The collaborative mechanism, detailed in Section 5.3, implements the semantic consistency model for heterogeneous EPC integration on the concept exchange model using a collaboration strategy. It introduces for the first time collaborative editing techniques as the key means of semantic consistency maintenance.

The collaborative conceptualisation approach is justified in Section 5.4.1 and its applicability has been demonstrated in several systems prototypes in Section 5.4.2.

1.3. Investigation scope

It should be noted that the semantic interoperability of heterogeneous EPC information involves two aspects of issues – the semantic conflict resolution for maintaining semantic consistency between heterogeneous EPCs, and the exchange of heterogeneous EPC information between EPC sources and EPC destinations. The former is a heterogeneous

EPC information integration issue, where its target is to achieve a semantically consistent state between heterogeneous EPCs. The latter is a heterogeneous EPC information exchange issue, where its goal is to enable that the required EPC information is adequately arrived at the intended destinations from the sources. The latter further relates to business document engineering and business process management.

The investigation of this article is within the scope of the first issue, which is the foundation of the second issue. Thus, how a piece of heterogeneous EPC information is transformed and used between heterogeneous EPC source and destination is out of the discussion scope of this article.

The remainder of this article is organised as follows. In Section 2, a brief survey is made on how the existing approaches of mandatory standardisation and automated mediation work to resolve semantic conflicts between heterogeneous EPCs and why it is still not enough to solve the problem. Section 3 attempts to answer why existing EPCs are complex, so that the properties of complex heterogeneous EPCs can be derived. In Section 4, a new PRODUCT MAP theory is developed to abstract what EPC structure, concept and context are and how they work together to require satisfying heterogeneous EPC integration conditions. In Section 5, the heterogeneous EPC integration approach is discussed through some models and mechanisms. In Section 6, both theoretical justification and some prototypical demonstrations on the applicability of the proposed approach are given. Section 7 compares the collaborative conceptualisation approach with the traditional and contemporary EPC integration approaches. Finally, we summarise the article and describe the particular contribution of this article, and point out the limitation of the proposed approach that leads to future work.

2. Literature survey

In the literature, there are two main approaches, mandatory standardisation and automated mediation, which attempt to solve the problem described in Section 1.1 (Guo and Sun 2003a).

2.1. Mandatory standardisation approach

Mandatory standardisation is an approach that avoids semantic conflicts by enforcing a product standard for all participating EPCs. It is initiated by standard designers or makers in the design stage through defining a shared product vocabulary (i.e. a set of 'commonly acceptable' product terms).

In this approach, resolving semantic conflicts between a product standard and local EPCs is the responsibility of the local users of the standard. Systems adopting this approach include EDI systems (United Nations 1987), barcode systems (e.g. UPC of uc-council.org), and books using the ISBN identifier. The former EDI systems are proprietary and rigid and require substantial amounts of maintenance costs. The latter barcode and ISBN systems are flat identifier systems, where the customised definitions to identifiers are dependent on the users.

More recently, the systems employing this approach often adopt hierarchical classification models (e.g. eCl@ss in eclass.de and UNSPSC in unspsc.org). The contemporary approach of this kind focuses on vocabulary in terms of ontology (Gruber 1993) (e.g. OWL) or thesaurus (Aitchison and Clarke 2004). It believes that semantic interoperability between heterogeneous EPCs could be achieved through standard ontology or thesaurus by means of standard conformance. Similar beliefs are held in industries that develop *de facto* EPC standards to build supply chains and e-marketplaces (e.g. www.martsoft.com/ocp).

In virtually all of the above cases, the technique for semantic conflict resolution implicitly follows an assumption that semantic conflicts in underlying heterogeneous EPCs could be circumvented via a set of commonly acceptable terms. Nevertheless, in spite of the advantages that EPC standards are well formatted in syntax, clear in semantics and highly machine-readable, there are disadvantages for adopting a technique of standardisation. A standardisation process is often a complex socio-economic process (Fomin and Keil 2000). Standards tend to be rigid and are not adaptive when facing users' integration requirements that are constantly changing (Damsgaad and Truex 2000). Hepp *et al.* (2007a) reveal that, though most of industrial standards are actually maintained and updated, many of them are rather inactive and are dead collections. Dogac and Cingil (2001) and Shim *et al.* (2000) pointed out that there are multiple international standards that are not interoperable.

2.2. Automated mediation approach

Automated mediation is an approach that undertakes the semantic conflict detection and resolution using an intelligent mediation system that is independent of heterogeneous EPCs. This mediation system often acts as an information middleware to intelligently transform heterogeneous product information between heterogeneous EPCs.

Traditionally, well-known examples of this approach are Smart Catalogs and Virtual Catalogs (Keller and Genesereth 1996) and NetAcademy (Lincke *et al.* 1998). The former introduces the facilitator concept to perform routing and translation between distributed product EPCs and catalogue web interfaces based on a set of ontologies (Gruber 1993). These ontologies define the various levels of common terms. The latter introduces the mediator concept (Linche and Schmid 1998) as an integrated EPC to mediate the distributed EPCs through a merger called Q-Calculus (i.e. a common product description frame, which is a formal language for description and classification of objects). Similar to the above two examples, Stanoevska-Slabeva and Schmid (2000) favoured the use of mediating ontologies to integrate the external heterogeneous EPCs.

Nevertheless, these examples have exposed a number of research challenges in heterogeneous EPC integration. For example, Baron *et al.* (2000), after analysing the Smart Catalogs and Virtual Catalogs, raised the computer understanding problem between heterogeneous EPCs. Besides, personalisation requirements for EPC construction, discussed in Liu *et al.* (2001) and Yen and Kong (2000), are also concerned. All these pose more challenges of EPC integration in semantic level, including: (1) the semantic accuracy of brokering or translating different product ontologies (Warburton 1999), (2) the reconciliation of semantic differences between heterogeneous EPCs (Linche and Schmid 1998), and (3) the interoperability with the existing international and industrial EPC standards (Schulten *et al.* 2001, Omelayenko and Fensel 2001b).

In response to the above challenges, contemporary researches on EPC integration often adopt a hybrid mediation approach; that is, on the one hand, standardisation emphasised on the development of ontologies or thesauri and, on the other hand, automated mediation is called to mediate heterogeneous EPCs according to either a standardised ontology or an *ad hoc* ontology specification. For example, Fensel *et al.* (2001) proposed the integration of product information by decomposing integration tasks into subtasks and reclassifying heterogeneous EPCs into a core set of marketplace product ontologies to automate the mediation between sellers and buyers. Obrst *et al.* (2001) provided an ontology mapping strategy in the product and service knowledge space for B2B e-commerce. MEMO (Quix *et al.* 2002) proposed a federated database system to store metadata to link the instances of product representations by a generic ontology schema. Omelayenko *et al.* (2001b) proposed a layered mapping approach to integrating *de facto* standards of cxml.org and xcbl.org. MOMIS (Bergamaschi *et al.* 2002, Beneventano *et al.* 2004) suggested the use of a semi-automatic method to define the mapping among product classification standards of UNSPSC and eCl@ss and eBay categorisation through a common thesaurus on a web service infrastructure. Landry (2004) proposed a linking strategy to multilingual subject access between different standard subject heading languages for building a digital library. Kong *et al.* (2005) applied web services to enhance the EPC interoperability.

The above hybrid mediation approaches have many merits. However, they often adopt intelligent mediating methods over static mapping of heterogeneous EPCs and lack the ability to cope with the EPC requirements that change dynamically (Guo and Sun 2003c). The problem is that after the mediating product ontologies/thesauri are designed, the e-marketplace is closed until the next version of their releases. Therefore, the openness for evolving is intermittently static between two versions and not adaptive to continuous changes. In addition, if SMEs are involved, the *ad hoc* EPCs share no public marketplace ontology/thesaurus in EPC design time. Under this circumstance, the commitment of the *ad hoc* EPCs to a certain product ontology/thesaurus introduces issues of incompleteness and inaccuracy (Fensel *et al.* 2001).

In this article, we refer to the above unsolved issues in standardisation and mediation approaches as a semantic consistency problem for heterogeneous EPC integration and interoperability, emphasising the issue of meaning consistency between them.

3. EPC complexity analysis

A semantic consistency problem directly relates to the complexity of real-world EPCs. A complexity analysis on the real-world EPCs is necessary. It is a method of capturing the properties of heterogeneous EPCs. It is helpful to understand these properties in order resolve the semantic consistency problem. In this article, the EPC complexity analysis is made in both spatial and temporal aspects.

3.1. EPC complexity in space

In the real world, EPCs are created in different groups like firms. Centred on interpretation of systems modelling, Robinson and Bannon introduced the term ‘semantic community’ to describe the cause of heterogeneity between different groups.

> Different groups, professions, and subcultures embody different perspectives. They communicate in different ‘jargon’. Much of this cannot be translated in a satisfactory way into terms used by other groups, since it reflects a different way of acting in the world (a different ontology or epistemology). Distinct groups of this sort will be referred to as semantic communities. (Robinson and Bannon 1991)

The term ‘semantic community’ is applicable to describe the numerous SMEs and explains the semantic consistency problem of heterogeneous EPCs in SMEs.

3.1.1. *EPC properties in spatial dimension*

In particular, EPCs of different semantic communities (e.g. SMEs) could have peculiar properties in a spatial dimension – distribution, autonomy and interdependence. Distribution refers to EPCs that reside in firms' computers and are geographically dispersed and connected by the Internet. Autonomy means that EPC designers of firms create EPCs in their own ways. Interdependence refers to the interaction requirement between EPCs where firms intend to work together for electronic business. EPC properties make heterogeneous EPCs very complex.

To illustrate such complexity, an empirical study on real-world refrigerator representations was conducted by the author through a Google search of the existing 54 corporate EPCs in 2003 using three terms of refrigerator, freezer and fridge, shown in Table 1 and Table 2 of the Appendix. The investigation revealed the following problems:

1. *Product meaning representation is highly heterogeneous*. For example, in Table 1, 59% of the sampled terms of 'refrigerator' (48%), 'freezer' (7%) and 'fridge' (4%) are synonyms, i.e. they all mean 'household refrigerator', and 41% of the sampled terms of 'refrigerator', 'freezer' and 'fridge' are homonyms, i.e. a same term in different context does not mean 'household refrigerator'.
2. *Product structure representations are highly heterogeneous*. For example, Table 2 shows that the manners of feature description of 'household refrigerator' (i.e. product structures) in sampled EPCs are highly heterogeneous.
3. *Semantic context reference systems of different EPCs are highly heterogeneous*. For example, in Table 1, 41% of EPCs, using the names of 'refrigerator', 'freezer' and 'fridge', are in fact not referenced to the semantic context of 'household refrigerator' but interpreted as other products.
4. *Different languages have their own terms for semantics*. For example, 'refrigerator' in English and 'réfrigérateur' in French.

The investigation indicated that real-world EPCs are complex due to their properties of distribution, autonomy and interdependence. This is consistent with many theoretical analyses on semantic heterogeneity in multi-databases and EPCs (Goh *et al.* 1994, Kashyap and Sheth 1996, Linche and Schmid 1998, Bergamaschi *et al.* 2002, Guo and Sun 2003b, 2003c). The implication of the investigation is that EPC integration efforts should be toward the reduction of complexity arising from the heterogeneity of semantics, structures and contexts amongst distributed, autonomous and interdependent EPCs.

3.1.2. *Abilities of capturing spatial EPC properties by existing approaches*

'Existing mandatory standardisation approach', in general, does not concern the EPC complexity (e.g. unspsc.org) because it denies the EPC autonomy property. It requires all EPCs to confirm the same standard. It simply pushes the integration task of resolving the semantic consistency issue between a standard and individual EPCs to each participated EPC, and assumes that any individual EPC is able to resolve the issue.

'Existing automated mediation approach' admits the EPC autonomy property, but it neglects the fact that EPCs are created in different semantic communities where interpretation of human-generated representations is needed. With this negligence, the transformation of EPC information between different semantic communities (e.g. xcbl.org and cxml.org) adopts the intelligent mediators using a set of inference rules (e.g. facilitator;

Keller and Genesereth 1996). More specifically, the intelligent mediator infers the semantic meaning of each individual EPC through some established rules using a domain-wide shared ontology and applies the inferred result to semantically communicate with another individual EPC. Since the intelligent mediators are only software fully dependent of the available rules, the circumstances of the unavailability of the rules lead to inaccurate EPC information transformation between heterogeneous EPCs.

3.2. EPC complexity in time

3.2.1. EPC property in temporal dimension

In the real world, EPCs are also dynamic in their states and thus present the emergence property in temporal dimension (Guo and Sun 2003a). Emergence refers to the continuous changing status of something in time. For example:

(1) EPCs for interoperation require the dynamic mapping in real time between buyers and sellers, because individual EPCs may be in changing states.
(2) Sellers and buyers (e.g. SMEs) have their local languages, cultures, preferences and business practices. These may change from time to time.
(3) e-marketplace relationships between sellers, buyers and providers are not stable. Participants may freely enter and leave e-marketplaces. This may change the already formed interdependent relationships between EPCs and make EPCs more complex.

It is important to reduce temporal complexity of EPCs caused by the emergence property, because any statically designed EPC services could lead to dysfunctional behaviour (Maindantchik *et al.* 2002), for example losing buyers' loyalty. This importance is strongly supported by the emergence theory (Giddens 1984, Mogan 1998, Truex *et al.* 1999), which points out that there is no point assuming that stable structures underpin organisations. Social organisations are works-in-process, emergent as their actors respond to adapting to shifting environments, and constantly interacting with each other to re-negotiate the 'rules of the game' for stability while never achieving it (Ngwenyama 1998, Damsgaad and Truex 2000). This theory explicitly states that many available system development means are inadequate, because they are not connected through a coherent framework that focuses on the emergent characteristics of organisations (Truex *et al.* 1999).

3.2.2. Abilities of capturing temporal EPC property by existing approaches

Existing mandatory standardisation approaches (e.g. UNSPSC and ecl@ss) cannot well satisfy the EPC emergence property, because standards are often rigid and time-consuming to change. The way of standard emergence is through versioning, which increases heterogeneous EPC copies that might add more difficulty of integration. Likewise, existing automated mediation approaches face the similar problem of inability in satisfying EPC emergence property. Intelligent mediators themselves cannot create dynamic rules that accurately capture the changing semantics of EPCs such that mediators could not read a human's thought expressed in a word, a phrase or a sentence.

A recent study from Hepp *et al.* (2007b) was interesting. They utilised Wikipedia entries as a vocabulary intended for intelligent mediators. Their testing result showed rather stable semantics for a large percentage of Wikipedia vocabulary over a certain period. This indicates the possibility of adopting Wikipedia as a stable shared vocabulary for search and information retrieval fields, assuming the searched and retrieved terms are also within the Wikipedia vocabulary scope. However, when numerous heterogeneous EPCs are involved (that is, combined with the spatial EPC properties) Wikipedia vocabulary as a shared vocabulary still has the semantic mediation problem between Wikipedia vocabulary and other numerous *ad hoc* EPCs. The reason is simple: *ad hoc* EPCs as a set of different vocabularies may not share the same vocabulary building principles as Wikipedia vocabulary. In addition, when 100% accuracy in mediation is required in exact business information exchange, for example heterogeneous contracts mediation, Wikipedia vocabulary could not be used as a shared vocabulary.

4. PRODUCT MAP theory

To fully satisfy the EPC properties of distribution, autonomy, interdependence and emergence in IEPC design, a method of deconstruction and reconstruction (initially proposed in Guo *et al.* 2004b) is adopted to articulate complex EPCs into manageable and atomic representation constructs and relationships.

Articulation refers to the orderly accomplishment of deconstructing the complex EPCs, which is borrowed from semiotics (Guiraud 1975, p. 32, Eco 1976, p. 231). It is a process of analysing, decomposing, meshing and aligning the complex EPCs to make it easy to capture EPC properties. Integration, on the other hand, is a process of the reconstruction of the articulated EPC constructs and relations into an IEPC system that is fit for semantically exchanging product information.

The articulation of EPC constructs and relations will be discussed in Section 4.1 while the integration of EPCs will be discussed in Section 4.2.

4.1. Homogeneous EPC articulation to understand generic EPC constructs

This subsection will discuss a homogeneous EPC as a representation through describing a theory on structure and concept.

4.1.1. Definition of representation

A representation, applying semiotics, can be defined as follows:

> A sign, or representamen, is something which stands to somebody for something in some aspect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call interpretant of the first sign. The sign stands for something, its object. It stands for that object, not in all aspects, but in reference to a sort of idea, which I have sometimes called the ground of the representamen. (Innis 1985, p. 5)

In this definition, a representation is originated from a semantic (or semiotic) community, shown in Figure 2, where semantic gaps happen between an existence, an observation, an interpretation, and a representation. First, an existence or a real object (e.g. a refrigerator) appears. Second, this real object has been observed, but is not guaranteed to capture the full meaning of the real object. Third, the observed real object is interpreted by human mind in the context of that person. This interpretation again distorts the original semantic meaning of the real object. Fourth, a representation, in which we are interested, is a written form in computable format. It once again semantically deviates from what the real objects refer to.

Thus, a representation has a semantic causal sequence, such as real object → observation → interpretation → representation.

A representation in semiotic theories is called a 'sign' (Saussure 1966, Barthes 1968, 1972, Eco 1976). It takes the form of words, images, sounds, odours, flavours, acts or objects, but such sign has no *a priori* intrinsic meaning and becomes a sign only when people invest it with a meaning; that is, being interpreted to deliver a meaning. 'Nothing is a sign unless it is interpreted as a sign' (Peirce 1931–1958, 2.172). For example, '*xyz*()' is a sign but it has no intrinsic meaning. The '*xyz*()' has a meaning only it is interpreted as something, say a function of *xyz*().

4.1.2. Definition of a homogeneous EPC

A homogeneous EPC as a whole is simply a representation of a set of product notions, residing in computers created by an entity used by all, for example UNSPSC or ecl@ss. It means that all participating EPCs in an interdependent work group follow the same design principles in both syntax and semantics. For example, all participating systems of a community may design their own EPCs based on both OWL and UNSPSC.

To formally define a homogeneous EPC, we employ the dyadic sign model (Saussure 1966, p. 67), which defines a sign as being composed of a 'signifier' and a 'signified'. A signifier is the form that the sign takes, and a signified is the concept (i.e. the meaning) that the sign represents. Applying this model, we refer to any syntactic construct as the signifier called 'structure' (S) to represent the syntax of an EPC representation, and refer to a semantic definition as the signified to express the meaning that an EPC representation carries, called 'concept' (C). By these definitions, any representation in an EPC (e.g. categories and commodities), we refer to it as a PRODUCT MAP (PM), could be notated as:

Definition 1: PRODUCT MAP (PM) = (structure, concept) = (S, C). ☐

This definition states that a representation or PM is a couple of structure and concept that are mutually independent. A structure can convey any concept and a concept can be conveyed in any structure. A structure does not need to imply a concept and a concept does not need to relate to a particular structure. Thus, a structure on its own is meaningless, referring to nothing but merely an existence of a construct. Structure is able to convey something only after a concept is conveyed by this structure. The process of

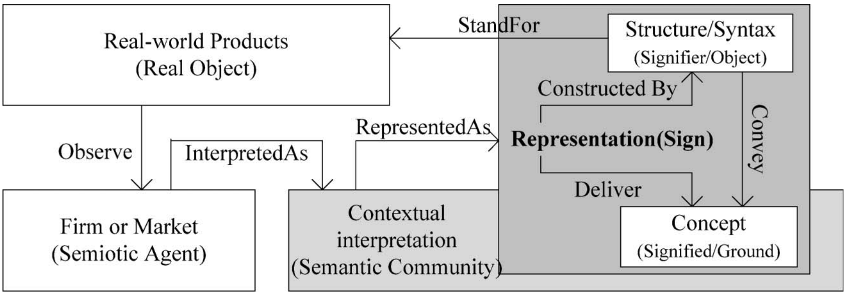


Figure 2. Representation in a semiotic community.

conveyance of concept in structure is referred to as conceptualisation. For example, we draw a circle as structure, which is meaningless. However, we can convey the meaning of either 'pie' or 'moon' to it.

Figure 3 further illustrates the relationship between structure and concept in a PM, where the concept of 'a refrigerator' can be conveyed in the structures of 'concept(123, fridge, prod)' in one context or 'concept(789, item)' in another context or the structure 'concept(789, item)' can convey either a concept 'a refrigerator' in one context or a concept 'a travelling bag' in another context.

Applying the PM, any EPC can be generically modelled as a set of PMs in terms of structures and concepts as Definition 2, because an EPC is a set of representations like categories, commodities, etc.

Definition 2: EPC $= (\mathrm{PM}_1, \mathrm{PM}_1, \ldots, \mathrm{PM}_n)$. □

For example, a barcode system is an EPC where each barcode is a structure and its definition of that barcode is a concept conveyed in the barcode structure.

4.1.3. Hierarchical homogeneous EPC

Real-world objects are often hierarchically placed. For example, our galaxy includes the solar system, the solar system includes Earth, Earth has living things, the living things include human beings and so on and so forth. The hierarchical relationship is applicable to represent a homogeneous EPC in PRODUCT MAP.

The EPC hierarchy in PRODUCT MAP can be explained by Barthes' orders of signification (Barthes 1972, p. 114–115), where a concept (signified) can be described in two types: denotation and connotation (Barthes 1968, p. 89–94, Eco 1976, 54–57). Denotation 'tends to be described as the definitional, 'literal', 'obvious' or 'commonsense' meaning of a sign', while connotation 'is used to refer to the socio-cultural and 'personal' associations ... of the sign', which is more obvious to individual interpreters (Chandler 2003). When a concept C1 (denotation) is conveyed in a structure S1 (signifier), it becomes a representation PM1, called the first system (Barthes 1968, p. 89). This first system (we call it a first level representation) then becomes a part of the structure of a second level representation PM2, where the concept (C2) in the second level representation is the connotation of the first level concept (C1). This second level concept C2 (i.e. C2 is connotation of C1 and C1 is denotation of C2) is again conveyed by the second level structure, which is a combination of first level structure S1 and the current second level structure S2 (i.e. S2 = S1 + S2). This second level representation could be recursively developed as a part of the further lower level structure (i.e. S3 = S2 + S3) to convey

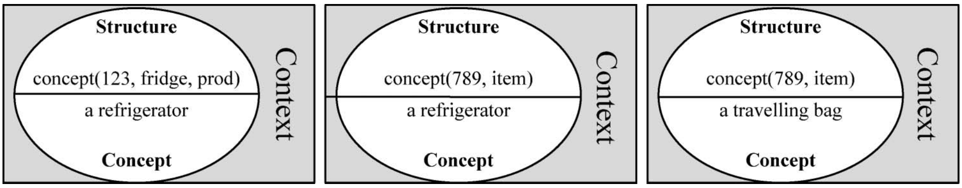


Figure 3. Model of representation or PRODUCT MAP (PM).

concept (C3 as connotation of C2 and C2 as denotation of C3) in conformity with Barthes' statement that the denotation leads to a chain of connotations.

For example, given that UNSPSC is a structure (S1) and a definition of 'UN product catalogue standard' is a concept (C1), the bonding of (UNSPSC → S1, 'UN product standard' → C1) = PM1 is a first level representation. PM1 via S1 then becomes a part of structure of a second level system of (UNSPSC:10000000 → S2, 'Live Plant and Animal Material and Accessories and Supplies' → C2) = PM2. This process can continue until no lower level concept can be found.

By applying Barthes' orders of signification, homogeneous EPCs can further be represented in a set of nested PMs shown in Definition 3.

Definition 3: EPC = PM(PM_i). Here, PM is the root of a homogeneous EPC and PM_i is a set of lower-level representations within the EPC, which could be recursively developed. ☐

4.1.4. Properties of PRODUCT MAP

The PRODUCT MAP exhibits some important properties of the following:

PM Property 1 (independence): S || C. Structure (S) and concept (C) are independent, where '||' notates an independence property, which implies that structure and concept can be created independently on their own. ☐

For example, the concept of 'red in our mind' and the structure 'red' can be independently exists such that 'red in our mind || red'. Here, if we do not associate 'red in our mind' to 'red', we do not know what 'red' means. (Please imagine if you tell computer 'red', what it will respond?)

PM Property 2 (conceptualisation): S\C. Structure can convey any concept if necessary, where '\' notates a conceptualisation property, which implies that concept in structure can be dynamically conveyed. ☐

For example, '123', 'colour' and 'price' are independent symbols. If we use '123' as a structure, it can convey any concept of 'colour', 'price' or anything in our minds in the form of '123\colour' or '123\price'. (Please imagine again if a computer knows this association, it could answer us that '123' refers to colour, or refers to price or refers to anything we give.)

PM Property 3 (causality): S ⇐ C. Structure conveys concept only after the conceptualisation of structure has been executed, where '⇐' notates a causality property based on semantic causal sequence of representation, which implies that conveyed concept in structure determines the meaning expression by structure. ☐

For example, we have a word as 'refrigerator', but we do not know what 'refrigerator' refers to and we simply call it a structure. Now we have a definition about the word 'refrigerator' like 'an appliance, a cabinet, or a room for storing food or other substances at a low temperature'. This definition, we call it as a concept, gives the meaning to 'refrigerator'. Thus, structure and concept has a causality relational property such that 'refrigerator ⇐ an appliance, a cabinet, or a room for storing food or other substances at a low temperature'.

PM Property 4 (hierarchy): $S_i \setminus C_i(S_i \bullet S_\Delta \setminus C_{i+1})$. A lower level concept is framed in the higher level structured concept, where the parentheses '( )' notate hierarchy property, which implies that, among adjacent levels of representations, lower level representation contains the semantics of the higher level representation, and lower level structure conveys higher level concept. ☐

For example, given two PM representations PMa = p00\price and PMb = v00\value. Now we want to express that the 'value' is the 'value' of 'price'. Based on Barthes' orders of signification (Barthes 1972, p. 114–115), 'value' is a connotation of 'price' and should be in the second level of the 'price' concept, such that PMa = p00\price (p00 • v00\value). This is because 'value' is not only conveyed by the concept of 'value' but also conveyed by 'price' to state that the 'value' is the 'price' *s* value.

4.1.5. *Representation of denotation and connotation*

Through the above properties, an EPC can be theoretically developed by capturing the properties of causality, conceptualisation and hierarchy by using independent symbols and re-stated in Definition 4.

Definition 4: $EPC = PM(PM_i) = S_1\backslash C_1(S_1\bullet\ S_2\backslash C_2\ (S_1\bullet S_2\bullet S_3\backslash C_3\ (\ \ldots\ )),\ \ldots\ ,)$ ☐

However, EPC in Definition 4 is awkward for computing and not suitable for computer use. A practical representation is needed.

Luckily, denotation is called only with regard to its lower level concepts, which are reversely called connotations of their higher level concept as their denotation. It gives us the opportunity to convey a denotation in a simple set structure in Definition 5.

Definition 5: Denotation structure (DS) = concept[identifier, annotation, link, option] = C[IID, AN, LK, OP], where the bracket '[ ]' introduces the detailed set of elementary structures for conveying denotative concept C. ☐

In this structure, IID, AN, LK and OP are all elementary structures. Particularly, identifier (IID) is a denotative concept identifier, which is to uniquely identify and convey a denotation, i.e. the annotation (AN) by following PM Property 3. Here, AN is a written definition created at the location of link. Often a link (LK) is a web address defining the web location of which semantic community an AN is created. Options are a set of optional elementary structures reserved for denotation structure. For example, the denotation structure of a refrigerator can be notated as c[iid, an, lk, op], which can convey the denotative concept of a refrigerator as c[1.52.14.15.1, household refrigerator, unspsc:52141501].

While denotation structure describes causal relationship between same level concepts, connotation structure (CS) is to capture PM Property 4 by building hierarchical relationship between concepts. It allows all denotations to be hierarchically evolved within an EPC to satisfy EPC emergence property. The CS notation is defined as:

Definition 6: Connotation structure (CS) = $(PM_1, \ldots, PM_n)$, where $PM_i = C[IID_i,\ AN_i,\ LK_i,\ OP_i]$ ☐

For example, given a denotation 'price' C[1111, price], its connotation is (currency, value, unit), which can be extended as C[1111, price](C[1111.1, currency], C[1111.2, value], C[1111.3, unit]). Particularly for connotation structure, its support on EPC evolvability is by means of a vector concept tree notated as follows:

Definition 7: Vector concept tree (VCT) = $(\mathrm{I}_i^1, \ldots, \mathrm{I}_i^k)$, where IID are indexed such that the level of denotation structures is $k \in \{1, \ldots, n\}$ and the position of connotation structure is $i \in \{1, \ldots, m\}$ (Guo and Sun 2003d). □

Through VCT, an EPC represented as a recursive PM can be freely evolved with IID in VCT form. Take UNSPSC (unspsc.org) for example, we can evolve lower level concepts for the 'household refrigerators' concept to include price concept. The evolution can be illustrated as follows:

(1) UNSPSC ID: 52141501 → PM IID: 1.52.14.15.1, where the first '1' is added as the EPC root. In UNSPSC, the first two digits are segment, the second two digits are family, the third two digits are class, and the fourth two digits are commodity.
(2) Evolve to include price concept under refrigerators: 1.52.14.15.1(1.52.14.15.1.1 \ price). It means the price of domestic refrigerators.
(3) Evolve to include currency concept under price: 1.52.14.15.1.1(1.52.14.15.1.1.1 \ currency). It means the currency of the refrigerator's price.

In the above VCT-based evolution, the real IID is generated from VCT = $(I_i^1, \ldots, I_i^k)$. The '1,*i* … *i*' is the written form of $(I_i^1, \ldots, I_i^k)$, where '1' refers to PM root and '*i*' refers to the sibling PM in the *k* level of a VCT.

4.2. Heterogeneous EPC articulation to understand inter-EPC relations

We have discussed the homogeneous EPCs in a PRODUCT MAP representation. In the real world, as stated in Section 3, EPCs are complex and heterogeneous. A heterogeneous EPC is an EPC that has a different way of constructing a PRODUCT MAP in both structure and concept, compared with others. Heterogeneous EPCs exist because of different semantic communities, which provide different contexts for EPC representations; that is, the different perspectives of building PRODUCT MAPS.

This subsection will represent heterogeneous EPCs by extending the PRODUCT MAP theory with an added concern of context. When contexts are involved, an EPC representation will have the following form:

Definition 8: EPC = $\mathrm{PM}(\mathrm{PM}_i)@\mathrm{X}_i$, where X_i is a set of contexts, $i \in (1, \ldots, n)$. □

This simple change from EPC = $\mathrm{PM}(\mathrm{PM}_i)$ to EPC = $\mathrm{PM}(\mathrm{PM}_i)@\mathrm{X}_i$ is significant, because EPC designers and users will no longer be assumed to work in the same semantic community but in different semantic contexts. Contexts will produce three types of issues: structure heterogeneity, concept heterogeneity and context heterogeneity, which all lead to semantic inconsistency for EPC interoperation.

4.2.1. Representing structure heterogeneity relations

Structure heterogeneity refers to the different constructs of structuring a PRODUCT MAP. It mainly reflects in two aspects of EPC classification and EPC modelling. EPC classification heterogeneity refers to the fact that EPC designers in different contexts classify the terms of an EPC in different classification schemes (Fensel *et al.* 2001, Ng *et al.* 2000). Explained by PRODUCT MAP, PM hierarchies are evolved in different concept chains. For example, UNSPSC and ecl@ss are not the same in their classification schemes. EPC modelling heterogeneity means that EPC designers model denotation structures in their own ways (Kim *et al.* 2002). For example, SME_1 may model a denotation structure as 'concept[classifier, definition, reference]' while SME_2 may model it as 'concept[name, description, relation]'. This leads to structure inconsistency and makes EPC interoperation impossible.

To overcome the structure heterogeneity problem, the term syntagm of semiotics is introduced. Syntagm (Saussure 1966, p. 122–123), relating to structuralising representations, is 'an orderly combination of interacting signifiers which forms a meaningful whole within a text' (Chandler 2003). It concerns term combination and positioning between terms, assuming that terms used as structural symbols are understandable to any involved parties.

To make heterogeneous structures interoperable, an analysis of syntagmatic relations between different structures is helpful. Figure 4 presents syntagms (*A* and *B*) that convey the aggregated concepts in two EPCs. In this example, the EPCs have different sequential syntagms; that is, they have heterogeneous structures to convey concepts – *A* uses (name, description, relation) and *B* uses (classifier, definition, reference), which differ in spatial dimensions and semantically conflict. In capturing PM hierarchy property, *A* implicitly uses sequence numbers (1, 2, 3) to arrange all concept records while *B* explicitly uses hierarchical classifiers (1, 1.1, 1.1.1).

The two EPC structures can be interoperable if they converge on a same isomorphic structure. This is a heterogeneous structure integration condition (HESI condition):

HESI condition: Given two heterogeneous PMs of $C_1[IID_1, AN_1, X_1, OP_1]$ and $C_2[IID_2, AN_2, X_2, OP_2]$ defined in Definition*s* 5 and 6, then C_1 and C_2 are structure-consistent if and only if:

(1) IID is a unique identifier of C such that $C(IID, AN, X, OP) \rightarrow IID$;
(2) X is the context of PM such that PM@X, following Definition 8;
(3) There exists a reference representation Map such that IID_1 and IID_2 are mapped onto $Map(IID_1@X_1, IID_2@X_2)$. □

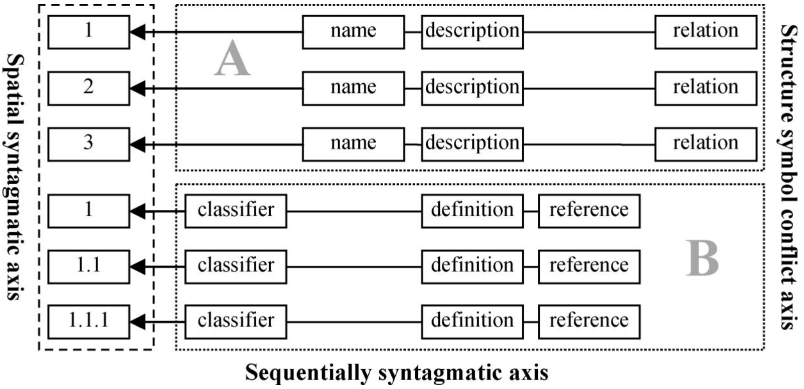


Figure 4. Syntagmatic relations between heterogeneous structures.

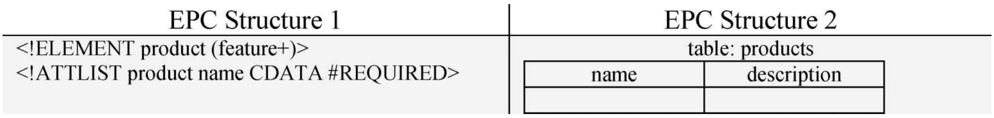
Generically, HESI condition achieves structure consistency by converging all heterogeneous structures onto an isomorphic structure *Map*($\text{IID}_1, \ldots, \text{IID}_n$). For example, given the heterogeneous structures in Figure 5: if IID_{11} = product[@name] and IID_{21} = products.name, and IID_{12} = 'product/feature' and IID_{22} = products.description, then heterogeneous EPC_1 and EPC_2 are structurally consistent on *Map*(IID_{11}@EPC_1, IID_{21}@EPC_2) and *Map*(IID_{12}@EPC_1, IID_{22}@EPC_2).

4.2.2. *Representing concept heterogeneity relations*

Concept heterogeneity refers to the different meanings of EPC terms used in different contexts. It mainly reflects in two aspects of concept naming heterogeneity and concept definition heterogeneity. Concept naming heterogeneity refers to the inconsistent uses of names or identifications for notating a concept. For example, a 'household refrigerator' could be named as 'refrigerator' or 'fridge' and identified as '52141501' by UNSPSC or *ad hoc* identified as '357'. Concept definition heterogeneity refers to the inconsistent definitions of the same presented term. For example, a 'refrigerator' can be defined as 'an appliance, a cabinet, or a room for storing food or other substances at a low temperature' (see *American Heritage Dictionary*), or simply 'a cooling bag for portably carrying drinks in a lower temperature'. The particular reflections of concept heterogeneity are synonymous and homonymous concept expressions (Goh *et al.* 1994). Concept heterogeneity leads to semantic inconsistency, which makes EPCs non-interoperable.

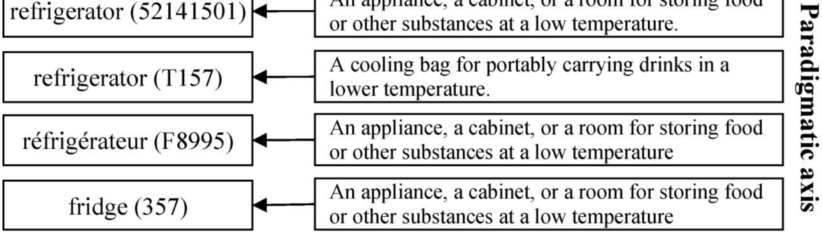
Concept heterogeneity can be overcome by investigating the paradigm of semiotics. Paradigm (Chandler 2003), relating to concept conveyance in structures, is a set of associated terms as members of some defining category. It concerns the autonomy of different EPC paradigms and looks for their possible transformation and transposition or, simply, the substitution of concepts in a paradigmatic relation.

The concern of substitutability of concepts in paradigm can be illustrated in Figure 6. Saussure called substitutability 'association relation' (Saussure 1966, p. 119–125), where each concept in a paradigm is significantly different in expression, for example



```
<!ELEMENT product (feature+)>
<!ATTLIST product name CDATA #REQUIRED>
```

Figure 5. EPC structure mapped onto XML PM structure.



An appliance, a cabinet, or a room for storing food or other substances at a low temperature.

A cooling bag for portably carrying drinks in a lower temperature.

An appliance, a cabinet, or a room for storing food or other substances at a low temperature

An appliance, a cabinet, or a room for storing food or other substances at a low temperature

Figure 6. Paradigmatic relation.

‘refrigerator’ ‘fridge’ and ‘réfrigérateur’ in either their names or definitions. If only concept names are concerned, they will have unreliable perceived similarity or dissimilarity because in Figure 6 we see their different semantic definitions. Saussure uses the term ‘mental association’ (Saussure 1966, p. 121–126) to describe this similarity or dissimilarity. He noted that there is no end (or commonly agreed order) to such association. The unreliable mental association between two concepts requires a way of proving that two associated concepts are also semantically substitutable, for example refrigerator (52141501) semantically equalling fridge (357).

In this article, the heterogeneous concept integration condition (HECI condition) described below is a condition of achieving concept substitutability.

HECI condition: Given two heterogeneous PMs of $C_1[IID_1, AN_1, X_1, OP_1]$ and $C_2[IID_2, AN_2, X_2, OP_2]$ defined in Definitions 5, 6 and 8, then C_1 and C_2 are concept consistent if and only if:

(1) AN is unique concept of C such that $C(IID, AN, X, OP) \rightarrow AN$.
(2) X is the context of PM such that PM@X, following Definition 8.
(3) PM Property 2 of conceptualisation is captured such that $IID \setminus AN$.
(4) PM Property 3 of causality is captured such that $IID \Leftarrow AN$.
(5) AN_1 and AN_2 are semantically substitutable such that $AN_1 \Leftrightarrow AN_2$. □

HECI condition guarantees that two heterogeneous concepts are semantically consistent. For example, the semantic consistency between *concept*[iid = ‘52141501’, annotation = ‘refrigerator’] and *concept*[iid = ‘12345’, annotation = ‘fridge’] could be guaranteed if and only if we could guarantee that ‘refrigerator’ and ‘fridge’ are semantically the same.

4.2.3. Representing context heterogeneity relations

Context heterogeneity refers to the different perspectives of designing EPCs in situated semantic communities. It mainly reflects in three aspects of natural language difference, referencing system heterogeneity and implicit concept involvement. Natural language difference refers to the fact that EPCs are designed in different natural language environments. Referencing system heterogeneity means that EPC designs do not refer to the same specification of modelling, terminology or dictionaries (e.g. refer to UNSPSC or ecl@ss). Implicit concept involvement is the practice of omitting some obvious concepts in EPC designs. For example, many EPC designers omit ‘currency’ and ‘unit’ concepts when designing ‘price’ concept. They simply use, for instance, price = ‘300’ to replace the full representation such as price=(currency = ‘USD’, value = ‘300’, unit = ‘piece’).

Context heterogeneity leads to both structure and concept inconsistency, and is a pragmatic issue. Pragmatics (Dijk 1977) is ‘the relationships between sign and their users’. It describes the semantic interpretation relationship between EPC designers and EPC users about the same representation.

Pragmatic relations are illustrated in Figure 7 and explained by the term of modality judgment (Chandler 2003). Modality judgment states that people of different semantic communities may describe and interpret things using their own living experiences. They could generate heterogeneous structures, concepts and reification (here the reification refers to the way of reifying a concept into a particular concept instance). This fact requires finding a way of semantically transforming the stuff of heterogeneous contexts so that understandings could be reached. In this article, we employ a collaboration mechanism, available to all EPC designers of different contexts, as a common context, shown in Figure 8. Through this common context, heterogeneous structures and contexts could be collaboratively mediated.

The common context achieved by collaboration mechanism is referred to as heterogeneous context integration condition (HEXI condition) as follows:

HEXI condition: Given two heterogeneous PMs of $C_1[IID_1, AN_1, X_1, OP_1]$ and $C_2[IID_2, AN_2, X_2, OP_2]$ defined in Definitions 5, 6 and 8, then C_1 and C_2 are context consistent if and only if:

(1) X is a unique context of C such that C[IID, AN, X, OP]@X as Definition 8.
(2) There exists a collaboration mechanism Γ on which EPC designers of C1 and C2 can work together to negotiate semantic agreements on C1 and C2 such that HESI condition and HECI condition can be independently satisfied following PM Property 1 of independence. ☐

For example, if EPC designer A of context 1 can work with EPC designer B of context 2 on a collaborative EPC editing system to negotiate that 'refrigerator' and 'fridge' are semantically equivalent, the concepts of 'refrigerator' and 'fridge' can be semantically consistent.

HEXI condition is important and sufficient for realising HESI and HECI conditions. It is worth remarking that any machine-based reconciliation between two heterogeneous representations for structure mapping and concept substitution is at most a kind of similarity (e.g. simPro system; Kashyap and Sheth 1996). This is because machines can only infer representation equivalence through pre-encoded rules. They cannot handle uncertainty outside the given rules; that is, the 'mental association' issue (Saussure 1966, pp. 121–126) remains unsolved. For example, machines cannot infer whether 'refrigerator' appearing in two places are semantically equivalent or not if they do not have predefined rules.

HEXI condition is about the establishment of collaboration relationship between EPC designers; it applies the theory of design collaboration (e.g. Geisler and Rogers 2000). By collaboration, EPC designers could make agreements on concept substitution between

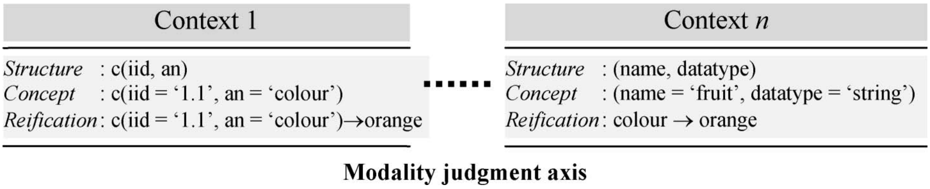


Figure 7. Pragmatic relations.

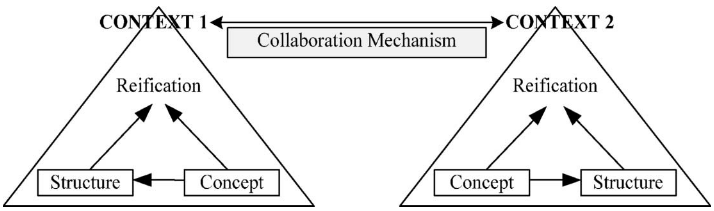


Figure 8. An extended PRODUCT MAP theory.

their heterogeneous representations and convey the agreed concepts in a commonly agreed representation, for example a referenced mapping representation like *Map*(IID_1@X_1, IID_2@X_2).

5. Heterogeneous EPC integration

This section will reconstruct the articulated EPCs into an interoperable EPC on a concept exchange model using a collaboration mechanism by following a semantic consistency model.

5.1. A semantic consistency model

The semantic consistency model aims to maintain semantic consistency in structures, concepts and contexts among numerous heterogeneous EPCs. By this model, the heterogeneous EPC integration conditions discussed in Section 4.2 are satisfied.

Definition 9 (**Semantic consistency model**): Given any two heterogeneous EPC representations in terms of two PM_1 and PM_2, they are said to be semantically consistent if and only if PM_1 and PM_2 ensure the following three properties:

(1) Structure mappability
(2) Concept equivalence
(3) Context commonality ☐

Definitions 10, 11 and 12 specify the details of these three properties.

Definition 10 (structure mappability '$\approx$'): Given two PM_1 and PM_2, then $\mathrm{PM}_1 \approx \mathrm{PM}_2$ if and only if: (1) $\mathrm{IID}_1 \in \mathrm{PM}_1$ and $\mathrm{IID}_2 \in \mathrm{PM}_2$, and (2) there exists a structure map Ξ such that Ξ (IID_1, IID_2), where IID_1 and IID_2 have a mapping relation ($\leftrightarrow$). ☐

Structure mappability satisfies HESI condition to resolve structure isomorphism issue by mapping heterogeneous structures onto an isomorphic PM mapping structure Ξ.

Definition 11 (concept equivalence relation '$\cong$'): Given two PM_1 and PM_2, then $\mathrm{PM}_1 \cong \mathrm{PM}_2$ if and only if: (1) AN_1, $\mathrm{IID}_1 \in \mathrm{PM}_1$ and AN_2, $\mathrm{IID}_2 \in \mathrm{PM}_2$, (2) $\mathrm{IID}_1 \backslash \mathrm{AN}_1$ and $\mathrm{IID}_2 \backslash \mathrm{AN}_2$, (3) $\mathrm{AN}_1 \Rightarrow \mathrm{IID}_1$ and $\mathrm{AN}_2 \Rightarrow \mathrm{IID}_2$, and (4) $\mathrm{AN}_1 \Leftrightarrow \mathrm{AN}_2$, where '$\Leftrightarrow$' is a semantic equivalence relation such that AN_1 and AN_2 refer to the same meaning. ☐

Concept equivalence satisfies HECI condition to achieve concept substitution by ensuring that two concepts are semantically equivalent.

Definition 12 (context commonality relation '$\propto$'): Given three PM_1, PM_2 and PM, then 'PM_1, $\mathrm{PM}_2 \propto \mathrm{PM}$ if and only if PM_1 ⟆ PM and PM_2 ⟆ PM. PM is called the common context of PM_1 and PM_2, where ' ⟆' is a reference relation. ☐

Context commonality satisfies HEXI condition to resolve context reference problems by ensuring that heterogeneous contexts are referenced to a common context. The common context enables EPC designers to work together to satisfy HESI and HECI conditions.

The formal grammar of generic representation of structures, concepts, contexts and maps and its XML implementation in XML PRODUCT MAP (XPM) are given in www.sftw.umac.mo/~jzguo/pages/spec.html.

5.2. *Collaborative concept exchange model*

The complex EPCs could have many types. In this subsection, we design a new collaborative concept exchange model, shown in Figure 9, based on the semantic consistency model and a collaboration integration strategy. This model characterises and integrates the real-world EPC types by improving a previous concept exchange model we proposed (Guo and Sun 2003b).

Real-world e-marketplaces (Guo and Sun 2004) include EPC producers, EPC consumers, EPC mediators and EPC integrators. Each of them is equipped with a type of EPC. These types are linked together with some mechanisms in three layers of LEPCs, CEPCs and SEPCs, which, as a whole, construct a global IEPC system.

Specifically, the LEPC layer consists of many local firms, which are semantic communities that produce their specific legacy systems, cultures, languages and personal preferences. They own and use *ad hoc* EPCs (AEPC) in terms of irregular product data sources such as XML data stores, relational databases or *ad hoc* web pages. AEPCs could be integrated into a local EPC (LEPC) by firms themselves. An LEPC is a firm-wide interoperable vocabulary and is the foundation of ERP system design (Bernus *et al.* 1996).

CEPC layer constitutes many CEPC providers. It provides CEPCs to LEPC designers. CEPC providers are semantic communities that create common concepts in common EPCs (CEPCs). By using CEPCs, LEPC designers could reference common concepts for cross-firm concept exchange. A single CEPC is the foundation of local e-marketplace (Bakos 1998) or SCM system design (Christiaanse and Kumar 2000) in a same natural language. To make CEPCs interoperable in a wider range, many language-different CEPCs (e.g. Chinese or English) can be replicated as a standard EPC (SEPC). SEPC is the foundation of a regional e-marketplace design for cross-country business deals.

SEPC layer comprises SEPC integration providers, which integrate different SEPCs. If all SEPCs could semantically connect together, an interoperable EPC (IEPC) could be established as the foundation of global e-marketplace (Guo and Sun 2004).

To integrate the above heterogeneous EPCs, a collaboration integration strategy is adopted to implement semantic consistency model. It states that collaborative engines (CEs) '●', shown in Figure 9, provide collaboration functions for EPC designers of adjacent semantic communities. These CEs provide common contexts $\chi(PM)$ for EPC

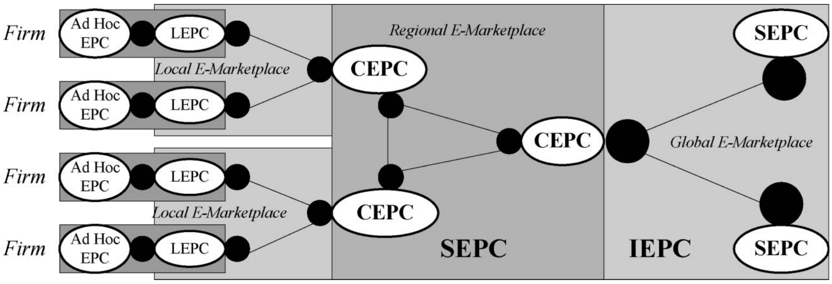


Figure 9. A generic concept exchange model.

designers that have heterogeneous contexts $\chi(\mathrm{PM}_1), \ldots, \chi(\mathrm{PM}_n)$ such that $\chi(\mathrm{PM}_1), \ldots, \chi(\mathrm{PM}_n) \propto \chi(\mathrm{PM})$ to ensure context commonality of semantic consistency model.

Through CEs, EPC designers collaboratively map their heterogeneous EPC structures onto PM mapping structures Ξ (IID_1, IID_2), ..., Ξ (IID_{n-1}, IID_n) during conveying heterogeneous EPC concepts in structures such that $\mathrm{IID}_1(\mathrm{PM}_1) \approx \cdots \approx \mathrm{IID}_n(\mathrm{PM}_n)$ to ensure structure mappability of semantic consistency model.

Through CEs, EPC designers negotiate concept equivalence by collaboration such that heterogeneously expressed annotations $\mathrm{AN}_1(\mathrm{PM}_1), \ldots, \mathrm{AN}_n(\mathrm{PM}_n)$ achieve $(\mathrm{AN}_1 \Rightarrow \mathrm{IID}_1) \Leftrightarrow \cdots \Leftrightarrow (\mathrm{AN}_n \Rightarrow \mathrm{IID}_n)$ for $\mathrm{PM}_1 \cong \cdots \cong \mathrm{PM}_n$ to ensure concept equivalence of semantic consistency model.

Collaborative integration strategy is important because, by this strategy, numerous heterogeneous EPCs can be semantically bridged to accurately transform meaningful information on a concept supply chain through mapping (Guo and Sun 2003d, Guo *et al.* 2004a), such that AEPC1 $\leftrightarrow$ LEPC1 $\leftrightarrow$ CEPC1 $\leftrightarrow$ CEPC2 $\leftrightarrow$ LEPC2 $\leftrightarrow$ AEPC2.

5.3. Collaboration mechanism

In this section, we propose a collaboration mechanism, shown in Figure 10, to physically realise a collaborative concept exchange model. It maintains semantic consistency between heterogeneous LEPCs and CEPCs.

Collaboration mechanism can be described in three types of collaborative engines as follows:

1. A common collaboration engine (CCE) is designed as a common context between natural-language-different CEPCs and used by all CEPC designers to collaboratively design semantically consistent CEPCs such that $\mathrm{comAn}_1 \Rightarrow \mathrm{comIid} \Leftarrow \mathrm{comAn}_2$ (comIid and comAn refer to IID and AN of common concepts in a CEPC).
2. A local collaboration engine (LCE) is designed as a common context between LEPCs and CEPCs in a same natural language and used for LEPC designers to design semantically consistent local concepts of LEPCs with CEPCs under the condition that LEPC designers can find the needed semantically equivalent common concepts in CEPCs for LEPC designs when LEPCs are collaboratively mapped onto the adjacent CEPC (locIid and locAn refer to IID and AN of local concepts in an LEPC).
3. A global collaboration engine (GCE) is designed as a common context between LEPCs and CEPCs in a same natural language and used for all LEPC designers to

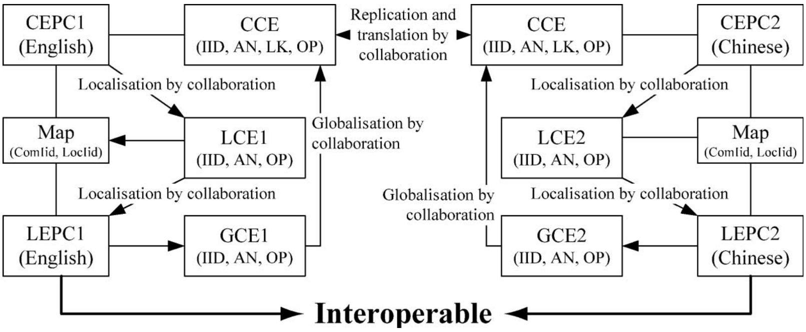


Figure 10. Collaboration mechanism.

design semantically consistent local concepts of LEPCs, such that (new comIid $\Leftarrow$ new comAn) $\leftrightarrow$ (new locIid $\Leftarrow$ new locAn), under the conditions that LEPC designers cannot find the needed semantically equivalent common concepts in CEPCs for LEPC designs when LEPCs are collaboratively mapped onto the adjacent CEPC.

5.3.1. *Common collaboration engine*

A common collaborative engine (CCE) is implemented as a peer-to-peer (P2P) (Guo 2006b) collaborative concept editing system, shown in Figure 11. It takes a partially replicated and centrally managed architecture that fully replicates IID but translates AN of different natural languages between CEPCs.

The challenge of CCE implementation is the semantic consistency maintenance between multiple copies of CEPCs in replication, where the key issue is to ensure PM Property 5 of denotation causality that maintains unique causal order of IID $\Leftarrow$ AN between different CEPCs, such that:

(1) preventing semantically not equivalent concepts $AN_1(PM_1), \ldots, AN_n(PM_n)$ from being conveyed in a same IID such that '$AN_1 \Rightarrow IID_x, \ldots, AN_n \Rightarrow IID_x$' happens;
(2) preventing semantically equivalent concept AN_x from being conveyed in many different $IID_1, \ldots, IID_n$ such that '$AN_x \Rightarrow IID_1, \ldots, AN_x \Rightarrow IID_n$' happens; and
(3) preventing a concept generated in CEPC1 of one natural language from being translated as an unequivalent concept in CEPC2 of another natural language, such that given $AN_1 \Rightarrow IID_x$ in $\chi(PM_1)$ and $AN_2 \Rightarrow IID_x$ in $\chi(PM_2)$, $AN_1 \neg \Leftrightarrow AN_2$.

To solve the above problems, CCE introduces a semantic node locking approach (Guo 2006a) and a collaborative verification approach to respectively resolve the above Issues 1, 2 and 3 in two procedures of replication and translation such that all IID shall be replicated between CEPC1 and CEPC2 and all AN1 shall be correctly translated from CEPC1 to corresponding AN2 in CEPC2.

A semantic node locking approach is an approach of resolving Issue 1 and 2 by issuing semantic locks on some concept nodes of a centrally managed PM hierarchy when a node of the PM is in editing mode. When adding a child concept node, the editing node and its existing children nodes are locked. This prevents Issue 1 such that a newly generated IID will be assigned with many ANs by concurrent EPC designers, and also prevents Issue 2 such that many IID will be concurrently generated following the ADD instructions from many EPC designers. When deleting a concept node, the father node and the children

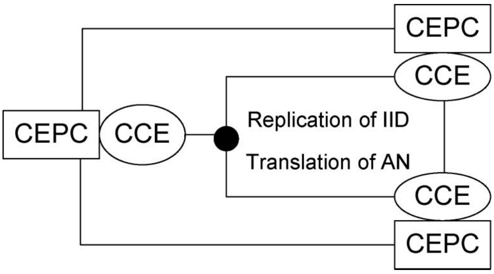


Figure 11. Common collaborative engine (CCE).

nodes of the editing node are locked to prevent the case such that the children nodes of the editing node cannot become the children nodes of the father node after the editing node has been deleted.

A collaborative verification approach is an approach to resolving Issue 3 by providing a verification mechanism for collaboratively supervising the automatic translation results between ANs of the IID-replicated CEPCs during CEPC translation procedure. The mechanism ensures IF $AN_1 \Rightarrow IID_x$ in $\chi(PM_1)$ THEN ($AN_2 \Rightarrow IID_x$ in $\chi(PM_2)$, $AN_1 \Leftrightarrow AN_2$) through a two-stage translation mechanism. In Stage 1, AN_1 is created in $CEPC_1$ and automatically translated into AN_2 in $CEPC_2$ such that $AN_1 \Rightarrow IID_x$ and $AN_2 \Rightarrow IID_x$ with an 'unverified' mark to warn $CEPC_2$ designers that AN_2 may be semantically inconsistent with the original AN_1 of $CEPC_1$ and requires semantic verification. In Stage 2, $CEPC_2$ designers examine all 'unverified' AN_2 to turn them into 'verified' AN_2 manually such that $AN_1 \Leftrightarrow AN_2$.

While automatic translation is necessary for saving labour cost, the collaborative verification is a sufficient condition to ensure concept equivalence between any two concepts of $CEPC_1$ and $CEPC_2$. It is an unavoidable remedy to the inaccurate machine translation under the current IT advancement.

5.3.2. *Local collaboration engine*

Local collaboration engine (LCE) is implemented as a dominant-to-follower (D2F) (Guo 2006b) collaborative concept editing system, shown in Figure 12. It takes a point-to-point and locally managed architecture to link a CEPC and LEPCs and localise needed common concepts of CEPC into local concepts by creating LEPCs.

The LCE implementation assumes that CEPC designers are knowledge workers (Rogoski 1999) and are authentic in designing and publishing common concepts for all to use. Thus, they are dominants like standard makers. On the contrary, LEPC designers are not knowledge workers and are unable to correctly design commonly acceptable concepts. Thus, LEPC designers are followers and must reference a CEPC in LEPCs through a *localisation* procedure, such that LEPC designers browse to reference common concepts, create local concepts, and map local concepts onto common concepts.

The key technical issue in localisation is the impact of dynamic changes of local and common concepts after concept map Ξ (locIid, comIid) has been created. For example, given a map Ξ (1.52.14.15.1, FG255) for (refrigerator → 1.52.14.15.1) ↔ (fridge → FG255). Now due to certain reasons, 1.52.14.15.1 has been changed to 1.52.14.15.20 to refer to the original meaning. This causes the map Ξ (1.52.14.15.1, FG255) obsolete with semantic inconsistency.

To solve this problem, we adopt a buffer mechanism (Guo 2006b) between CEPC, LEPC and their map Ξ such that CEPC → buffer → Ξ (comIid, locIid) ← buffer ← LEPC. With this added mechanism, whenever there is a change, the changes are sent to the buffers. If the maps are available, real-time revisions are made to the maps by emptying the buffer. If the maps are not available (e.g. offline), the changes are queued in the buffers until the maps are available. Applying this buffer mechanism in localisation, we ensure

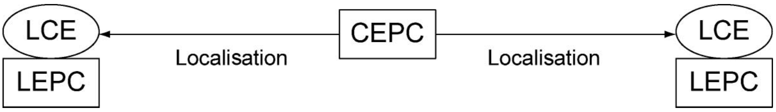


Figure 12. Local collaboration engine (LCE).

semantic consistency between LEPCs and CEPCs such that (locIid $\Leftarrow$ locAn) $\leftrightarrow$ (comIid $\Leftarrow$ comAn).

5.3.3. Global collaboration engine

Global collaboration engine (GCE) is implemented as a requestor-to-answerer (R2A) (Guo 2006b) collaborative concept editing system, shown in Figure 13. It takes a point-to-point and centrally managed architecture to create newly needed local concepts of LEPCs that are currently not mappable onto common concepts in CEPC. It accepts the requests of new concept creation from LEPC designers and propagates them to CEPC designers to real-time answer the requests in a *globalisation* procedure such that (new locIid $\Leftarrow$ new locAn) $\leftrightarrow$ (new comIid $\Leftarrow$ new comAn).

The technical issue in globalisation procedure is how CEPC designers can quickly answer the requests from LEPC designers. For example, given a new request of the needed concept of (locIid = 'LF111-3', locAn = 'size of fridge used in kitchen') for LEPC design, how can CEPC designers immediately receive, design and publish the re-edited equivalent concept of (comIid = '1.52.14.15.1.5', comAn = 'the dimension of household refrigerators') in CEPC for LEPC designers to use.

To improve the response time, an awareness mechanism is designed such that all LEPC requests will be immediately aware by CEPC designers through an awareness interface in GCE, which displays and alerts CEPC designers to re-design the submitted local concepts of LEPCs for publishing and answering.

In summary, the collaboration mechanism implements collaborative integration strategy through three types of collaborative engines. With this mechanism, the requirements of exactness, flexibility and evolvability (Guo and Sun 2003b) for IEPC reconstruction are satisfied.

6. Justification and applicability

This subsection first justifies that LEPCs on a generic concept exchange model (see Figure 9) are semantically connectible for exchanging concepts between any two LEPCs. Thus, the collaborative conceptualisation approach is an effective semantic integration approach for designing interoperable electronic product catalogue (IEPC) systems, where documents derived from LEPCs are semantically exchangeable between heterogeneous contexts. Following this justification, some applications based on this justified approach are demonstrated to illustrate the applicability of the approach.

6.1. Justification of collaborative conceptualisation approach

Proposition (semantic connectivity): Any concept of a local document LD_1 semantically derived from $LEPC_1$ in $Context_1$ is semantically connectible to another corresponding concept of local document LD_2 semantically derived from a remote $LEPC_2$ of $Context_2$ if

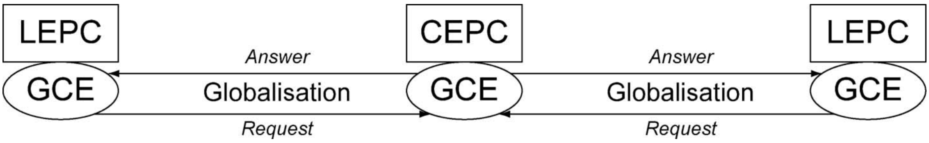


Figure 13. Global collaboration engine (GCE).

and only if $LEPC_1$ and $LEPC_2$ are collaboratively created on collaboration mechanism (CM) (see Figure 10).

Justification:

(1) Given LD_1, $LEPC_1 \in Context_1$; LD_2, $LEPC_2 \in Context_2$, CD_1, $CEPC_1 \in Context_3$; CD_2, $CEPE_2 \in Context_4$, and let all of them be represented as PRODUCT MAP PMs (*Def*. 4, 5, 6) such that $PM_{LD1}(IID_{LD1}, AN_{LD1})$, $PM_{LEPC1}(IID_{LEPC1}, AN_{LEPC1}) \in Context_1$; $PM_{LD2}(IID_{LD2}, AN_{LD2})$, $PM_{LEPC2}(IID_{LEPC2}, AN_{LEPC2}) \in Context_2$; $PM_{CD1}(IID_{CD1}, AN_{CD1})$, $PM_{CEPC1}(IID_{CEPC1}, AN_{CEPC1}) \in Context_3$; $PM_{CD2}(IID_{CD2}, AN_{CD2})$, $PM_{CEPC2}(IID_{CEPC2}, AN_{CEPC2}) \in Context_4$.

(2) Collaborative conceptualisation

2.1. Common concept collaborative conceptualisation (for Section 5.3.1)
Given $\forall IID_{CEPC1} \backslash AN_{CEPC1} \subset PM_{CEPC1}$ and $\forall IID_{CEPC2} \backslash AN_{CEPC2} \subset PM_{CEPC2}$ (following PM Property 1 & 2), then $IID_{CEPC1} \approx IID_{CEPC2}$ (ensuring structure mappability, *Def*. 9, 10) and $AN_{CEPC1} \cong AN_{CEPC2}$ (ensuring concept equivalence, *Def*. 9, 11) on CCE such that PM_{CEPC1}, $PM_{CEPC2} \propto$ CCE (ensuring context commonality, *Def*. 9, 12). CCE is illustrated in Figures 10 and 11.

2.2. Local concept collaborative localisation (for Section 5.3.2)
Given $\forall IID_{LEPC1} \backslash AN_{LEPC1} \subset PM_{LEPC1}$, $\forall IID_{LEPC2} \backslash AN_{LEPC2} \subset PM_{LEPC2}$, $\exists IID_{CEPCx} \backslash AN_{CEPCx} \subset IID_{CEPC1} \backslash AN_{CEPC1} \subset PM_{CEPC1}$ and $\exists IID_{CEPCy} \backslash AN_{CEPCy} \subset IID_{CEPC2} \backslash AN_{CEPC2} \in PM_{CEPC2}$ (following PM Property 1 & 2), then $IID_{LEPC1} \approx IID_{CEPCx}$ (*Def*. 9, 10) and $AN_{LEPC1} \cong AN_{CEPCx}$ (*Def*. 9, 11) on LCE_1 such that PM_{LEPC1}, $PM_{CEPCx} \propto LCE_1$ (*Def*. 9, 12). The same applies to the semantic mapping between $LEPC_2$ and $CEPC_y$. LCE is illustrated in Figures 10 and 12.

2.3. Local concept collaborative globalisation (for Section 5.3.3)
$\forall IID_{LEPC} \backslash AN_{LEPC} \in PM_{LEPC}$ and $\forall IID_{CEPC} \backslash AN_{CEPC} \in PM_{CEPC}$, if $\exists (AN_{LEPCx} \subset_{sem} AN_{LEPC}) \not\subset_{sem} AN_{CEPC}$, then let $AN_{LEPCx} \subset_{sem} AN_{CEPC}$ on GCE such that $\forall PM_{LEPCx}, \forall (PM_{CEPCy} \subset PM_{CEPC}) \propto$ GCE (*Def*. 9, 12), $AN_{LEPCx} \cong AN_{CEPCy}$ (*Def*. 9, 11) and $IID_{LEPCx} \approx IID_{CEPCy}$ (*Def*. 9, 10), where ‘ $\subset_{sem}$’ refers to semantically being included in. GCE is illustrated in Figure 10 and Figure 13.

(3) Semantically consistent concept exchange (for application)

3.1. Given common documents CD_1 and CD_2 are respectively subsets of $CEPC_1$ and $CEPC_2$ such that $PM_{CD1}(IID_{CD1}, AN_{CD1}) = PM_{CEPCx}(IID_{CEPCx}, AN_{CEPCx}) \subset PM_{CEPC1}(IID_{CEPC1}, AN_{CEPC1})$, $PM_{CD2}(IID_{CD2}, AN_{CD2}) = PM_{CEPCy}(IID_{CEPCy}, AN_{CEPCy}) \subset PM_{CEPC2}(IID_{CEPC2}, AN_{CEPC2})$, let $IID_{CD1} \approx IID_{CD2}$ (because of justification 2.1, we can always find $IID_{CD1} \approx IID_{CD2}$), then $IID_{CEPCx} \approx IID_{CD1} \approx IID_{CD2} \approx IID_{CEPCy}$.

3.2. Given local documents LD_1 and LD_2 are respectively subsets of $LEPC_1$ and $LEPC_2$ such that $PM_{LD1}(IID_{LD1}, AN_{LD1}) = PM_{LEPCx}(IID_{LEPCx}, AN_{LEPCx}) \subset PM_{LEPC1}(IID_{LEPC1}, AN_{LEPC1})$ and $PM_{LD2}(IID_{LD2}, AN_{LD2}) = PM_{LEPCy}(IID_{LEPCy}, AN_{LEPCy}) \subset PM_{LEPC2}(IID_{LEPC2}, AN_{LEPC2})$, then $IID_{LD1} = IID_{LEPCx} \subset IID_{LEPC1}$ and $IID_{LD2} = IID_{LEPCy} \subset IID_{LEPC2}$.

3.3. If $IID_{LEPCx} \approx IID_{CEPCx} \subset IID_{CEPC1}$ and $IID_{LEPCy} \approx IID_{CEPCy} \subset IID_{CEPC2}$ do not respectively exist for $LEPC_1$ or $LEPC_2$, GCE is launched to guarantee the existence of such relation. Let $IID_{LEPCx} \approx IID_{CEPCx} \subset IID_{CEPC}$ and $IID_{LEPCy} \approx IID_{CEPCy} \subset IID_{CEPC2}$.

3.4. Thus, $IID_{LD1} = IID_{LEPCx} \approx IID_{CEPCx} \approx IID_{CD1} \approx IID_{CD2} \approx IID_{CEPCy} \approx IID_{LEPCy} = IID_{LD2}$. Because of PM Properties 1, 2, 3 and 4, PM_{LD1} is semantically equivalent to PM_{LD2} and thus LD_1 of $Context_1$ semantically equals LD_2 of $Context_2$.

Thus, the proposition of semantic connectivity between any of two LDs of different contexts is proved and the effectiveness of collaborative conceptualisation for concept exchange is justified.

6.2. Applicability of collaborative conceptualisation approach

The collaborative conceptualisation approach can be applied in many areas of e-commerce and information systems such as interoperable electronic catalogue system design, electronic marketplace construction, inter-organisational cooperation and multilingual translation. In this subsection, two application prototypes are introduced to demonstrate the applicability of the approach.

6.2.1. CAT prototype

CAT (Contextual vAlue Translation) (Guan 2008) is a prototype of multilingual business document translation systems applied to use in different contexts with different natural languages. It aims to maintain semantic consistency between reified terms (i.e. constant values) of any business document when the document is translated from one context of a firm to another. It follows the design principles of accuracy and flexibility (Guo and Sun 2003b). Its system framework can be illustrated in Figure 14, where reified multilingual terms (i.e. constant values) are collaboratively designed between dictionary designers to reach semantic agreements between multilingual terms.

The novelty of this prototype is its context-based term sense disambiguation (TSD) method. The hierarchically arranged PRODUCT MAP concepts are used as the contexts of the reified multilingual terms. This allows that each reified term has its definite context to disambiguate the term sense and thus could be accurately translated. For example, given a term ‘orange’, if its context is ‘colour’, it will not be translated into ‘橙子’ (in the context of ‘fruits’) but ‘桔黄色’.

The evaluation experiment on document translation showed that the translation accuracy between Chinese and English is 92–100% (close to our 100% accuracy expectation) for each test term from evaluation corpse based on the well-known gold standard measure. This is exceptionally good compared with the experiments on Google Translate (42–100%) and Babel Fish (13–96%).

The explanation for why CAT could achieve high accuracy in translating reified multilingual terms is that when the reification of collaborative concepts is controlled within the collaboratively designed reified terms, the accuracy could be guaranteed. The 8% inaccuracy appeared in CAT experiments showed that when reified terms are not taken from the collaboratively designed dictionaries, the inaccuracy rate increases.

6.2.2. HICP prototype

HICP (hotel information collaboration platform) (Luo 2008) is a prototype of distributed online hotel information servicing systems. It aims to resolve the semantic conflicts between the service provider’s CEPC and numerous hotels’ LEPCs at all levels of

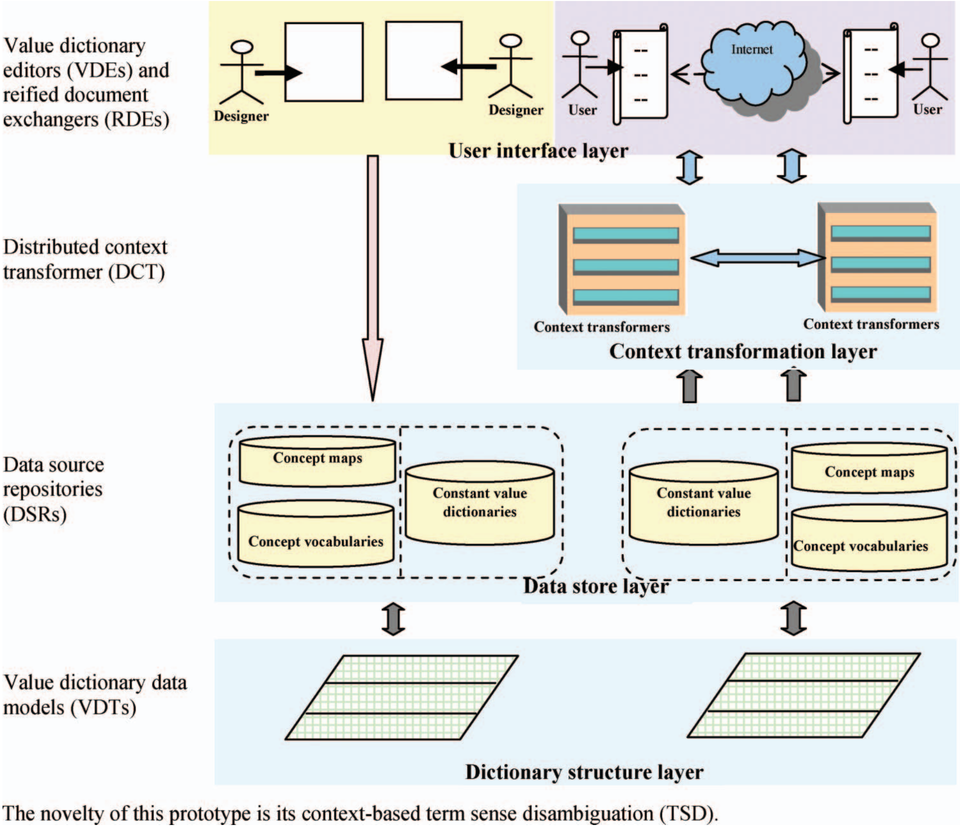


Figure 14. CAT system framework.

structure, concept and context. Its task is to provide the accurate hotel accommodation offers by hotels in real time when tourists lodge their hotel accommodation inquiry from HICP's user interface. It follows the design principles of exactness, flexibility (Guo and Sun 2003b) and richness, where richness refers to as much information as possible mediated between HICP platform and numerous hotel information systems. HICP's system architecture is illustrated in Figure 15, where common concept designers collaboratively design CEPC of HICP, and local common designers localise CEPC to LEPCs, globalise non-included LEPCs to CEPC and semantically map LEPC concepts onto CEPC concepts.

The technical advancement of the HIPC prototype in implementation level is its flexible, rich and document-based inquiry model. Local hotel systems will not just receive several keywords for querying the possible available hotel rooms, but will receive a full complete hotel inquiry document for hotel information systems to make valid offers. The rich yet semantically consistent concept-centred and document-oriented inquiry highly meets tourists' requirements during their placing of hotel room orders.

The evaluation experiments on HICP prototype given below showed that HICP is a big improvement on existing well-known online hotel information systems such as expedia.com, hotels.com and elong.com.

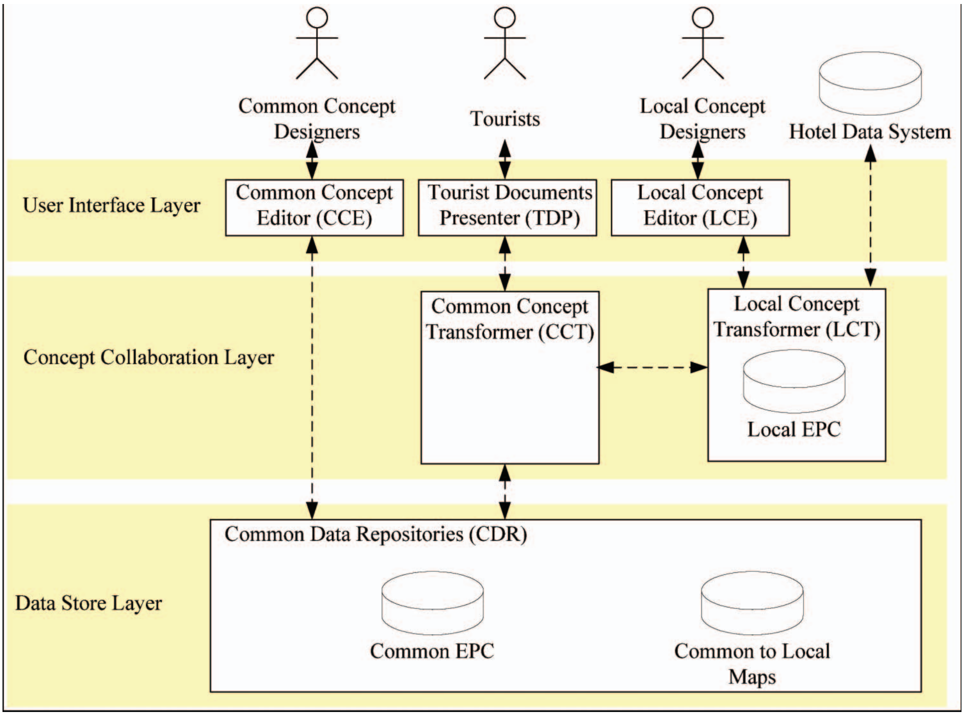


Figure 15. HICP system architecture.

	Quantity of features	Has attributes features	Has attributes ratio	Searchable features	Searchable features ratio
www.expedia.com	165	7	4.24%	18	10.91%
www.hotels.com	151	5	3.31%	18	11.92%
www.elong.com	125	6	4.80%	5	4.00%
HICP	304	259	85.20%	304	100.00%

Both CAT and HICP systems prototypes demonstrated the applicability of collaborative conceptualisation approach. They imply that this approach is promising in many e-commerce application fields.

7. Comparing with related works

The collaborative conceptualisation approach presented in this article is different from the mandatory standardisation approach (e.g. UNSPSC and ecl@ss) and the automated mediation approach (e.g. MOMIS; Bergamaschi *et al.* 2002, Beneventano *et al.* 2004) and Omelayenko *et al.* 2001b) of EPC integration. It is a pioneer approach to adopt collaboration to integrate numerous complex EPCs distributed around the world. Nevertheless, it has more or less absorbed the important integration thoughts related to IEPC design, comparing the contemporary information integration approaches.

First, at a cursory level, the collaborative conceptualisation approach may resemble many contemporary integration approaches. Examples of these commonalities include:

- Specifying semantics in a concept model (e.g. sign model, Peirce 1974/1978/1979) of semiotics; a similar approach is found in conceptual graph (Sowa 2000).
- Classifying product concepts in classification taxonomy (e.g. elc@ss and UNSPSC) and adopting unique identifiers to identify product concepts.
- Isolating heterogeneity in context (e.g. COIN approach; Goh *et al.* 1994) where semantic conflicts are formed.
- Adopting the thought of mapping (e.g. Obrst *et al.* 2001 and MACS, Landry 2004) for context transformation.

It is posited that, despite these superficial similarities, the collaborative conceptualisation approach represents a departure from these contemporary strategies that aim at rejuvenating the mediation and standardisation approaches. These are often characterised by the adoption of a domain-wide common vocabulary construction for more effective semantics mediation, or a standardisation process to limit the effects of inaccurate vocabulary mapping by proposing a homogeneous specification of product representation. To some extent, the systems (e.g. COIN, Goh *et al.* 1994, MOMIS, Bergamaschi *et al.* 2002, Beneventano *et al.* 2004, and MACS, Landry 2004) could be seen as the example of the first approach (by introducing common semantic types or by building mapping ontologies or thesauri on standards) while the systems (e.g. based on UNSPSC and ecl@ss) could be regarded as the latter approach (by promoting the adoption of standards). These strategies differ from the collaborative conceptualisation strategy since they continue to rely on rule-based global vocabularies or standards that mediate heterogeneous EPCs and neglect the *ad hoc* EPCs that are difficult to be inferred using shared vocabularies or standards.

The second difference is the concept defined in the collaborative conceptualisation approach, which only captures the hierarchical semantic relation between concepts (in terms of denotation–connotation relation). It thus differs from conceptual graph (CG; Sowa 2000) where concept relations are complex and nearly every pair of different concepts has a different concept relation. While CG has the advantage of representing richer and more particular concepts, it makes it more difficult for millions of heterogeneous concept representations to build equivalent relations in concept definitions for semantic interoperation. The collaborative conceptualisation also differs from product classification standards such as UNSPSC and ecl@ss, which have fixed hierarchical levels of product concepts. A fixed hierarchy of concept representation is rigid and is not evolvable for meeting emergent requirements of customers (Damsgaad and Truex 2000, Truex *et al.* 1999). The implication is that the concept representation of the collaborative conceptualisation approach is more flexible and simple to manage for concept evolution.

The third difference is that although context is also emphasised in the collaborative conceptualisation approach, the context in this case is a perspective of an individual EPC, which consists of a set of recursive unique concepts that could be decomposed into atomic concepts that are collaboratively designed (see Section 4.1). This differs from the definition that a context is a collection of axioms (e.g. COIN, Goh *et al.* 1994) particular to a source or receiver of heterogeneous systems. The implication is that a context is run-time divisible to cater for the specific need for accuracy in context transformation for product concept exchange. In addition, heterogeneous structures and concepts in different contexts can be collaboratively reconstructed through structure mapping and concept equivalence under a common context. This realises the personalisation of the individual LEPC and allows legacy systems to be connected.

The fourth difference is the way of concept mapping between heterogeneous EPCs. In the collaborative conceptualisation approach, heterogeneous concepts are not simply mapped onto a global vocabulary (e.g. NetAcademy, Lincke *et al.* 1998, and MACS, Landry 2004). They are also not pure structure mapping or similarity mapping between objects (e.g. semPro, Kashyap and Sheth 1996). The map in the collaborative conceptualisation approach is a collaboration result, which is an agreement on two heterogeneous concepts between two contexts. The implication is that a collaborative concept map is dynamic and can be flexibly edited. It expresses exact and equivalent meanings of two concept designers from different contexts. This resolves semantic conflicts arising from inaccurate mapping that bothers mapping systems that adopt technologies of artificial intelligence.

The fifth difference is that collaborative conceptualisation approach emphasises exactness (i.e. 100% full accuracy) in semantic integration research. It asks for semantic equivalence between any two concepts from heterogeneous contexts. It is thus different from those semantic integration systems that aim at achieving higher semantic similarity between two concepts or two data sets (e.g. simPro, Kashyap and Sheth 1996). This is because semantic similarity is not useful for exchanging business documents that are legally binding. A business document should be exchanged such that either it is fully semantically consistent for exchange or it is aborted for not being exchanged.

There are some unique features in the collaborative conceptualisation approach that are not comparable because they are novel and stem from both empirical investigation (see Table 1 and 2 in the Appendix) and theoretical analysis on semiotics (see Section 4). These features are:

(1) The collaborative conceptualisation describes a collection of domain models that are hierarchically distributed in fragmented e-marketplaces (Guo and Sun 2004) based on the nature of uneven distribution of real-world marketplaces (Guo 2007, Wang and Archer 2007). Each domain model has its own vocabulary and context, which does not link to a universal vocabulary but to an adjacent context. This departs from the contemporary approaches that only describe a single domain model with a global vocabulary. For example, COIN mediates a number of databases through a set of common semantic types (Gho *et al.* 1994). MEMO bridges heterogeneous sources with common product ontology in a federated system (Quix *et al.* 2002). It is noted that HERMES (Subrahmania *et al.* 1995) appears similar to the collaborative conceptualisation approach because it also aims to mediate multiple domains. However, it focuses on physical linking of the data sources, where its semantic mediation still relies on a central mediator that includes a conflict handling toolkit.

(2) The EPCs in the collaborative conceptualisation approach are represented by denotation and connotation structures through the analysis of a sign model, which enables the product concepts to be accurately and flexibly denoted and connoted (see Section 4.1). Semantic relations between concepts become simple and manageable. This improves the existing integration approaches that neglect the interdependence property of heterogeneous EPCs, which occurs in the real-world e-business world.

(3) The collaborative integration strategy in the collaborative conceptualisation approach is proposed based on the analysis of a set of semantic consistency conditions on the real-world interdependent heterogeneous EPCs that are

distributed, autonomous and emergent. These conditions are abstracted as EPC structure mappability, EPC concept equivalence and EPC context commonality.

(4) The collaborative creation and reconstruction of product concepts on collaboration mechanism (see Section 5.3) enable the semantic conflicts to be resolved in the stage of IEPC design. The introduction of collaboration departs from traditional and contemporary rule-based mediation and transformation of heterogeneous concepts (e.g. HERMES, Subrahmania *et al.* 1995). It has eliminated semantic conflicts between any participating sources that are heterogeneous, because collaboration between concept designers can effectively make agreement on the semantic equivalence between heterogeneous concepts.

The implication of these distinctive features is that the collaborative conceptualisation approach is able to meet the requirements of flexibility, evolvability and exactness (Guo and Sun 2003b) in a concept exchange model.

8. Conclusion

In this article, we have laid a solid theoretical foundation for designing an interoperable electronic product catalogue (IEPC) system based on our systematic exposition of what are complex EPCs, how these complex EPCs can be deconstructed through an articulation process into a set of manageable structures, concepts and contextual relationships on a theoretical analytical framework of PRODUCT MAP, and how these manageable constructs can be reconstructed into an IEPC system by satisfying a set of heterogeneous EPC integration conditions through a collaborative integration strategy on a collaboration mechanism in terms of several collaborative engines. The systematic exposition of designing IEPC in theory is documented and called the collaborative conceptualisation approach in this article.

The existing heterogeneous EPC integration approaches of standardisation and mediation have their merits. The mandatory standardisation approach could reduce efforts for EPC integration if firms widely adopted a standard. However, standards are often inflexible and slow to react to the emergent requirements. They also push the integration task to users. The automated mediation approach could automate integration tasks, but the accuracy for mediating heterogeneous EPCs is often not satisfactory. When compared with these two approaches, the collaborative conceptualisation approach resolves the unsolved issue of standardisation and mediation approaches in terms of semantic consistency problem between complex heterogeneous EPCs that are autonomously distributed around the emergent yet interdependent world. Compared with our previous researches that were devoted to collaborative EPC integration, this article, for the first time, theoretically and systematically describes how heterogeneous EPC systems can be redesigned into an IEPC system. Besides the contribution of this summarisation, the particular incremental contributions of this article were dedicated to the following points:

- Found and generalised four generic properties attributed to the complex EPCs. These properties are distribution, autonomy, interdependence and emergence.
- Developed a theory of PRODUCT MAP to deconstruct the complex EPCs into a set of manageable constructs of structures, concepts and contexts to ease the integration process towards IEPC.
- Abstracted a semantic consistency model to satisfy heterogeneous EPC integration conditions for IEPC design.

- Proposed a collaborative concept exchange model and rectified the wrong direction of achieving semantic similarity between any two heterogeneous EPCs (Guo and Sun 2003d) by requiring accurate concept exchange between business partners.
- Proposed a collaboration mechanism with several collaborative engines to implement collaborative concept exchange model and finally resolve the semantic consistency problem.

The above contributions are important and have several implications. First, it has confirmed the position of collaboration in business information integration. With collaboration, various semantic conflicts can be resolved to support accurate concept exchange. Second, the complex business information world has a way to be orderly managed through articulation methodology. The heterogeneous information sources can be collaboratively classified to observe and allow their dynamic evolution. This methodology can also be immediately applied to articulate complex business documents and business processes, where they exhibit the similar semantic consistency problem. Third, various tools can be designed based on the collaborative conceptualisation approach to analyse and reconstruct many of existing commercial e-marketplace systems. For instance, our preliminary investigation shows the world-leading Alibaba (www.china.alibaba.com) e-marketplace lacks the ability of semantically integrating Chinese SMEs into overseas buyers (e.g. www.alibaba.com) to enable them to exchange product data in a seamless manner (Guo *et al*. 2006). A tool devised and based on the proposed approach could add this ability to Alibaba.

The limitation of the collaborative conceptualisation approach is that the success of heterogeneous EPC integration depends on the underlying designed collaboration mechanism. If a badly implemented collaboration mechanism is introduced, it may not be able to resolve the semantic consistency issue. In this sense, the future work of the proposed approach should ensure that the implemented collaboration mechanism satisfies the semantic consistency model, so that the collaborative results reflect the truth assignment between all collaborative partners.

Lastly, since the collaborative conceptualisation approach presented in this article is pioneer research in adopting the collaborative integration strategy for integrating complex EPCs, this article wishes to attract more researchers of semantic integration to participate in our discussion.

References

Aitchison, J. and Clarke, S., 2004. The thesaurus: A historical viewpoint, with a look to the future. *Cataloging & Classification Quarterly*, 37 (3/4), 5–21.

Bakos, Y., 1998. The emerging role of electronic marketplaces on the Internet. *Communications of the ACM*, 41 (8), 24–30.

Baron, J., Shaw, M., and Bailey, A., 2000. Web-based E-catalog systems in B2B procurement. *Communications of the ACM*, 43 (5), 93–100.

Barthes, R., 1968. *Elements of semiology*. English version. New York: Hill and Wang.

Barthes, R., 1972. *Mythologies*. English version. New York: Hill and Wang.

Beneventano, D., Guerra, F., Magnani, S., and Vincini, M., 2004. A web service based framework for the semantic mapping amongst product classification schemas. *Journal of Electronic Commerce Research*, 5 (2), 114–127.

Bergamaschi, S., Guerra, F., and Vincini, M., 2002. A data integration framework for e–commerce product classification. *In*: *Proceedings of ISWC 2002*, 9–12 June, Sardinia, Italy. Berlin: Springer, 379–393.

Bernus, P., Nemes, L., and Williams, T., 1996. *Architectures for enterprise integration*. London: Chapman & Hall.

Chandler, D., *Semiotics for Beginners* [online]. http://www.aber.ac.uk/media/documents/S4B/semiotic.html; http://web.archive.org/web/20030210113754/http://www.aber.ac.uk/media/Documents/S4B/semiotic.html [Accessed 18 March 2003].

Christiaanse, E. and Kumar, K., 2000. ICT enabled coordination of dynamic supply web. *International Journal of Physical Distribution & Logistics Management*, 30 (3–4), 268–285.

Damsgaad, J. and Truex, D., 2000. Binary trading relations and the limits of EDI standards: The procrustean bed of standards. *European Journal of Information Systems*, 9 (3), 173–188.

Dijk, T., 1977. *Text and context: Explorations in the semantics and pragmatics of discourse*. London: Longman.

Dogac, A. and Cingil, I., 2001. A survey and comparison of business-to-business e-commerce frameworks. *ACM SIGEcom Exchanges*, 2 (2), 14–25.

Eco, U., 1976. *A theory of semiotics*. Bloomington: Indiana University Press.

Fensel, D., Ding, Y., Omelayenko, B., Schulten, E., Botquin, G., Brown, M., and Flett, A., 2001. Product data integration in B2B e-commerce. *IEEE Intelligent Systems*, 16 (4), 54–59.

Fomin, V. and Keil, T., 2000. Standardisation: Bridging the gap between economic and social theory. *In*: *Proceedings of AIS ICIS 2000*, 1–3 December, Brisbane, Australia [online]. Available from: http://aisel.aisnet.org/icis2000/20/.

Geisler, C. and Rogers, E., 2000. Technological mediation for design collaboration. *In*: *Proceedings of IEEE/ACM IPCC/SIGDOC 2000*, 24–27 September, Cambridge, MA. Los Alamitos: IEEE Educational Activities Department, Article 37.

Giddens, A., 1984. *The construction of society: Outline of the theory of structuration*. Cambridge: Policy Press.

Goh, C., Madnick, S., and Siegel, M., 1994. Context interchange: Overcoming the challenges of large-scale interoperable database systems in a dynamic environment. *In*: *Proceedings of CIKM'94*, 29 November–2 December, Gaithersburg, MD. New York: ACM Press, 337–346.

Gruber, T., 1993. A translation approach to portable ontologies. *Knowledge Acquisition*, 5 (2), 199–220.

Guiraud, P., 1975. *Semiology*. English version, translated by George Gross. London: Routledge & Kegan Paul Ltd.

Guan, X., 2008. *Context-based translation of constant concept values in e-business*. Thesis (Masters), University of Macau.

Guo, J. and Sun, C., 2003a. Collaborative product representation for emergent electronic marketplace. *In*: *Proceedings of Bled 2003*, 9–11 June, Bled, Slovenia. bledconference.org, 847–859.

Guo, J. and Sun, C., 2003b. Concept exchange: Constructing interoperable electronic product catalogues in an emergent environment. *In*: *CEC'03*, 24–27 June, Newport Beach, CA. Los Alamitos: IEEE Computer Society, 165–172.

Guo, J. and Sun, C., 2003c. Context representation of product data. *ACM SIGEcom Exchanges*, 4 (1), 20–28.

Guo, J. and Sun, C., 2003d. Context representation, transformation and comparison for ad hoc product data exchange. *In*: *Proceedings of DocEng'03*, 20–22 November, Grenoble, France. New York: ACM Press, 121–130.

Guo, J. and Sun, C., 2004. Global electronic markets and global traditional markets. *Electronic Markets*, 14 (1), 4–12.

Guo, J., Sun, C., and Chen, D., 2004a. Transforming heterogeneous product concepts through mapping structures. *In*: *Proceedings of 2004 international conference on cyberworlds (CW'04)*, 18–20 November, Tokyo, Japan. Los Alamitos: IEEE Computer Society, 22–29.

Guo, J., Sun, C., and Chen, D., 2004b. Deconstruction and reconstruction of heterogeneous electronic product catalogues for semantic interoperation. *In*: *Proceedings of CEC'04*, 6–9 July, San Diego. Los Alamitos: IEEE Computer Society, 333–336.

Guo, J., 2006a. A transparent collaborative integration approach for ad hoc product data. *In*: *Proceedings of CEC/EEE'06*, 26–29 June, San Francisco. Los Alamitos: IEEE Computer Society, Article 56.

Guo, J., Lam, I., Lei, I., Guan, X., Iong, P., and Ieong, M., 2006. Alibaba International: Building a global electronic marketplace. *In*: *Proceedings of 2006 IEEE international conference on e-business engineering (ICEBE'06)*, 24–26 October, Shanghai. Los Alamitos: IEEE Computer Society, 554–558.

Guo, J., 2006b. Inter-enterprise business document exchange. *In*: *Proceedings of ICEC'06*, 13–16 August, Fredericton, Canada. New York: ACM Press, 427–437.

Guo, J., 2007. Business-to-business electronic market place selection. *Enterprise Information Systems*, 1 (4), 383–419.

Hepp, M., Leukel, J., and Schmitz, V., 2007. A quantitative analysis of product categorisation standards: content, coverage, and maintenance of eCl@ss, UNSPSC, eOTD, and the RosettaNet Technical Dictionary. *Knowledge and Information Systems*, 13, 77–114.

Hepp, M., Siorpaes, K., and Bachlechner, D., 2007. Using Wikipedia entries as vocabulary for knowledge management. *IEEE Internet Computing*, 11 (5), 54–65.

Innis, R. ed. 1985. *Semiotics: An introductory anthology*. London: Hutchinson.

Kashyap, V. and Sheth, A., 1996. Semantic and schematic similarities between database objects: A context-based approach. *The VLDB Journal*, 5, 276–304.

Keller, A. and Genesereth, M., 1996. Multivendor catalogs: Smart catalogs and virtual catalogs. *EDI Forum: Journal of Electronic Commerce*, 9 (3), 87–93.

Kim, W., Kerschberg, L., and Scime, A., 2002. Learning for automatic personalisation in a semantic taxonomy-based meta-search agent. *Electronic Commerce Research and Applications*, 1 (2), 150–173.

Kong, S., Li, H., Liang, Y., Hung, T., Anumay, C., and Chen, Z., 2005. Web services enhanced interoperable construction products catalogue. *Automation in Construction*, 14 (3), 343–352.

Landry, P., 2004. Multilingual subject access: The linking approach of MACS. *Cataloging & Classification Quarterly*, 37 (3/4), 177–191.

Linche, D. and Schmid, B., 1998. Mediating electronic product catalogs. *Communication of the ACM*, 41 (7), 86–88.

Lincke, D., Schmid, B., Schubert, P., and Selz, D., 1998. The NetAcademy: A novel approach to domain-specific scientific knowledge accumulation, dissemination and review. *In*: *Proceedings of HICSS'98*, 6–9 January, Hawaii, USA. Los Alamitos: IEEE Computer Society Press, 131.

Liu, D., Lin, Y., Chen, C., and Huang, Y., 2001. Deployment of personalised e-catalogues: An agent-based framework integrated with XML metadata and user models. *Journal of Network and Computer Applications*, 24, 201–228.

Luo, Y., 2008. *Semantic integration of heterogeneous online hotel information systems*. Thesis (Masters), University of Macau.

Maindantchik, C., Montoni, M., and Santos, G., 2002. Learning organisational knowledge: An evolutionary proposal for requirements engineering. *In*: *Proceedings of SEKE'02*, 15–18 July, Ischia, Italy. New York: ACM Press, 151–157.

Mogan, G., 1998. *Images of organisation*. San Francisco: Berret-Koehler Publishers, Inc. Thousand Oaks: SAGE Publishers, Inc.

Ng, W.K., Yan, G., and Lim, E., 2000. Heterogeneous product description in electronic commerce. *ACM SIGEcom Exchanges*, 1 (1), 7–13.

Ngwenyama, O.K., 1998. Groupware, social action and organisational emergence: On the process dynamics of computer mediated distributed work. *Accounting, Management and Information Technologies*, 8, 127–146.

Obrst, L., Wray, R., and Howard, L., 2001. Ontological engineering for B2B e-commerce. *In*: *Proceedings of ACM FOIS'01*, 17–19 October, Ogunquit, ME. New York: ACM Press, 117–126.

Omelayenko, B. and Fensel, D., 2001a. An analysis of B2B catalogue integration problems: Content and document integration. *In*: *Proceedings of ICEIS'01*, 7–10 July, Setubal, Portugal. iceis.org, 945–952.

Omelayenko, B. and Fensel, D., 2001b. A two-layered integration approach for product information in B2B e-commerce. *In*: *Proceedings of EC-Web 2001*, 4–6 September, Munich, Germany. Berlin: Springer, 226–239.

Peirce, C., 1931–1958. *Collected writing* (8 volumes). C. Hartshome, P. Weiss, and A. Burks, eds. Cambridge, MA: Harvard University Press.

Peirce, C., 1978 (1–2); 1974 (3–6); 1979 (7–8). Phenomenology. *In*: C. Hartshorne, P. Weiss, and A. Burks, eds. *Collected papers of Charles Sanders Peirce*. Cambridge, MA: The Belknap Press of Harvard University Press.

Quix, C., Schoop, M., and Jeusfeld, M., 2002. Business data management for business-to-business electronic commerce. *ACM SIGMOD Record*, 31 (1), 49–54.

Robinson, M. and Bannon, L., 1991. Questioning representations. *In*: *Proceedings of ECSCW'91*, September, Amsterdam. Dordrecht/Boston/London: Kluwer Academic Publishers, 219–233.

Rogoski, R., 1999. Knowledge workers top company assets. *Triangle Business Journal*, 14 (19), 21.

Saussure, F., 1966. *Course in general linguistics*. New York: McGraw-Hill Book Company.

Schulten, E., Akkermans, H., Guarino, N., Botquin, G., Lopes, N., Dörr, M., and Sadeh, N., 2001. The e-commerce product classification challenge. *IEEE Intelligent Systems*, 16 (4), 86–93.

Segev, A., Wan, D., and Beam, C., 1995. Electronic catalogs: A technology overview and survey results. *In*: *Proceedings of CIKM'95*, 28 November–2 December, Baltimore, MD. New York: ACM Press, 11–18.

Shim, S., Pendyala, V., Sundaram, M., and Gao, J., 2000. Business-to-business e-commerce frameworks. *IEEE Computer*, 33 (10), 40–47.

Sowa, J., 2000. Ontology, metadata, and semiotics. *Conceptual Structures: Logical, Linguistic, and Computational Issues,* Lecture Notes in AI #1867. London: Springer, 55–81.

Stanoevska-Slabeva, K. and Schmid, B., 2000. Internet electronic product catalogs: An approach beyond simple keywords and multimedia. *Computer Networks*, 32, 701–715.

Subrahmania, V., Adali, S., Brink, A., Emery, R., Lu, J., Rajput, A., Rogers, T., Ross, R., and Ward, C., 1995. *HERMES: A heterogeneous reasoning and mediator system*. University of Maryland [online]. Available from: http://www.cs.umd.edu//projects/hermes/publications/abstracts/hermes.html.

Sun, C., Jia, X., Zhang, Y., Yang, Y., and Chen, D., 1998. Achieving convergence, causality preservation, and intention preservation in real-time cooperative editing systems. *ACM Transactions on Computer–Human Interaction*, 5 (1), 63–108.

Truex, D., Baskerville, R., and Klein, H., 1999. Growing systems in emergent organisations. *Communication of the ACM*, 42 (8), 117–123.

United Nations, 1987. *EDI for Administration, Commerce, and Transport* (EDIFACT).

Yen, B. and Kong, R., 2002. Personalisation of information access for electronic catalogs on the Web. *Electronic Commerce Research and Applications*, 1 (1), 20–40.

Wang, S. and Archer, N., 2007. Electronic market place definition and classification: Literature review and clarifications. *Enterprise Information Systems*, 1 (1), 89–112.

Warburton, K., 1999. A database of e-commerce terms: Implementation and benefits in producing internationalised software. *In*: *Proceedings of ACM SIGDOC'99*, 12–14 September, New Orleans, LA. New York: ACM Press, 120–126.

Appendix

Table 1. Semantic differences of 'household refrigerator' naming.

Description	Not referring to household refrigerator	Referring to household refrigerator	Named as 'refrigerator'	Named as 'freezer'	Named as 'fridge'
Actual/total	22/54	32/54	26/54	4/54	2/54
Percentage	41%	59%	48%	7%	4%

Note: (1) The research has randomly investigated 54 web sites that provide catalogues including products named as refrigerator, freezer or fridge. The catalogues are randomly selected from the search engines (www.google.com) from March to April 2003. (2) 41% of terms as refrigerator, freezer and fridge do not refer to household refrigerators, but 59% of same terms refer to household refrigerator.

Table 2. Attribute and value structure of 'refrigerator'.

Feature description	Multi-level attribute structure	Single level attribute structure	Attribute number min(*n*), max(*m*)	Simple value structure	Complex value structure	Simple pair-value structure	Complex pair-value structure	Multiple complex value structure
Actual/total	18/32	12/32	average 17	30/32	26/32	12/32	12/32	8/32
Percentage	56%	38%	*n*, 6; *m*, 33	94%	81%	38%	38%	25%

Note: (1) Multiple level attribute structure is a nested attribute structure such as 'dimension (width, height)'. Single level attribute structure is a non-nested attribute structure such as 'colour'. Single value structure is a simple value structure such as 'red' for colour. Complex value structure is a complex value structure such as '2 years' for warranty. Simple pair-value structure is a simple paired value structure such as '2 kg/24 h' for freezing. Complex pair-value structure is the value structure such as 'FOB Rotterdam USD450/piece'. Multiple complex value structure is the value structure such as '600 × 595 × 1850 mm'. (2) The total number is taken from Table 1's 59% of valid refrigerator catalogues. (3) Percentage is the actual number with said structure to total sampled number.