
Collaboration role in semantic integration for electronic marketplace

Jingzhi Guo

Faculty of Science and Technology,
Department of Computer and Information Science,
University of Macau,
Av. Padre Tomas, Pereira, S.J., Taipa, Macau
Fax: +853-2883-8314
E-mail: jzguo@umac.mo

Abstract: Semantic integration is important in constructing electronic marketplaces (e-marketplaces). Although many previous researches have been done, the task of semantic integration and the mechanism of maintaining semantic consistency are still not well known. This paper proposes a Collaboration-Role-Based Semantic Integration Mechanism (CRB-SIM). It promotes collaboration as a key means of resolving semantic conflicts in concepts creation and editing, via a set of service-oriented collaboration roles. This proposal has contributed to the semantic integration research a novel approach with the features of the role-based and service-oriented semantic consistency maintenance of heterogeneous concepts. The applicability of proposed mechanism is exemplified by two application prototypes.

Keywords: electronic marketplace; semantic integration; semantic consistency maintenance; concept exchange; document exchange; role-based system; service-oriented design.

Reference to this paper should be made as follows: Guo, J. (2010) 'Collaboration role in semantic integration for electronic marketplace', *Int. J. Electronic Business*, Vol. 8, No. 6, pp.528–549.

Biographical notes: Jingzhi Guo is an Assistant Professor of the Department of Computer and Information Science at the University of Macau in Macao. He received his PhD in Electronic Commerce and Internet Computing from Griffith University, Australia. His current research interests are in the electronic marketplace, business semantic integration, e-payment systems, virtual world and virtual marketplaces. He is an editorial board member of *Enterprise Information Systems*.

1 Introduction

Semantic integration plays a very important role in constructing various types of e-marketplaces for firms to do business online (Wang et al., 2008). More and more researchers have realised its importance and applied this concept to integrate different disconnected systems into newly integrated systems. Historically, three general approaches are devised, which are mandatory standardisation approach, intelligent mediation approach and collaborative conceptualisation approach (Guo, 2009).

For example, the work of Guo (2008) investigated collaborative concepts to integrate numerous ad hoc electronic product catalogues. The paper by Omelayenko and Fensel (2001) researched the standard integration approach to integrate business documents between xCBL (xcbl.org) and cXML (cxml.org), and the paper by Antoniou et al. (2007) developed brokering systems to intelligently mediate heterogeneous systems through OWL-based ontology. Nevertheless, while the importance of semantic integration is confirmed, the specific tasks and methods of semantic integration are not well known and elaborated.

For example, when a firm sends a product offer to another firm, should this offer be fully comprehensible and understandable to the receiving firm? Traditional researches in artificial intelligence often believed that the full semantic match between the sender and receiver is often not possible. Thus, they promoted the semantic similarity (e.g., Gabrilovich and Markovitch, 2007) to match the understanding between senders and receivers. Semantic similarity is very important. It can be widely used in the search of products and customers, and various types of information retrieval. However, specific to electronic commerce (e-commerce), when buyers and sellers communicate with each other over internet for doing business, the inaccurate information exchange, which is semantically similar, is prohibited. This is because if the receiver wrongly interprets the received purchase order from the sender, legal consequences between buyers and sellers will happen, and they have to resolve the misunderstanding between product offer and order execution.

This paper aims to specify the task of semantic integration in e-marketplaces and provide a promising role-based collaboration approach as a way to the 100% accurate semantic integration. With these discussions, this paper argues that using collaboration to semantically integrate contextual e-commerce information is a laudatory approach for maintaining semantic consistency. Following this argument, it proposes a generic CRB-SIM, which describes the working mechanism of collaboration in a role-based system to maintain semantic consistency during business concept creation and edition on e-marketplaces. To explain the feasibility of the proposed mechanism in a more understandable way, this paper presents two application protocols (Guan, 2008; Luo, 2008), designed and implemented based on CRB-SIM model as an illustration to the proposal.

The remainder of this paper will be organised as follows. The task of semantic integration will be first presented in Section 2, followed by the discussion of the collaboration roles in Section 3. The generic role-based collaborative integration mechanism, described in Section 4, outlines collaboration relationships between various integration parties, and is illustrated in Section 5 through Contextual vAlue Translation (CAT) and Hotel Information Collaboration Platform (HICP) application prototypes. The previous and related works regarding semantic integration and role-based collaboration are presented in Section 6. Finally, a brief conclusion, contribution and future work are given.

2 Task of semantic integration

Confucius very early in 250 BC in China said that if title (i.e., truth concept of terminology for understanding) is not given correctly, then what is said cannot be followed; if what is said cannot be followed, then work cannot be accomplished.

In current e-marketplace designs, the situation is extremely similar: the predefined vocabularies in designers' contexts cannot be followed by the users' understanding in users' contexts (Guo and Sun, 2003a, 2003b), and thus the constructed e-marketplaces cannot correctly exchange business information between e-marketplace participants. Semantic inconsistency between business systems thus becomes an issue for e-marketplace participants to exchange meaningful business information effectively.

Semantic integration technology can be defined as a technology towards the improvement of the meaning understanding of business information distributed amongst the disparate messaging systems of known and unknown business partners. In this technology, the term *meaning understanding* here is essential. It refers to reaching a certain level of semantic agreement between the communicating partners of human through underlying disparate business messaging systems over various networks or internet. Particularly speaking, semantic consistency between terms of heterogeneous vocabularies will enable business documents to be consistent, and consistent documents will allow business processes unambiguously created (Guo et al., 2008). Thus, consistency maintenance between heterogeneous vocabularies is the foundation of all levels of semantic conflict resolution. The quality indicators of the meaning understanding are the accuracy, reach and richness. *Accuracy* reflects the reliability of the semantic agreement between human partners. It can be measured by mutual acceptability, where the term acceptability refers to the "satisfactoriness by virtue of conforming to approved standards" (WordNet.org). For example, when both a seller and a buyer use UNSPSC codes (unspsc.org) as their product identifiers, we say that both the buyer and the seller have a very accurate meaning understanding in their communication about the referred products. This is because both of them agree and are satisfied at conforming to UNSPSC as their product communication media. Thus, the accuracy in this case has 100% mutual acceptability in measurement. On the contrary, if a buyer and a seller communicate based on a set of keywords created in their individual minds, the reliability of the mutual agreement is doubtful. This is because the mutual acceptability of the keyword set is low. *Reach* "is about access and connection" (Evans and Wurster, 1999). It reflects the coverage of the semantic agreement between human partners. It can be measured by the number of connected messaging systems and the population size of business partners reaching semantic agreements. For example, we can compare the reach between UNSPSC (unspsc.org) and ecl@ss (eclass.de) by checking their coverage according to the number of people who adopt these two standards and their numbers of adoption in existing messaging systems. *Richness* "is the depth of the information that a business gives to or collects about its customers" (Evans and Wurster, 1999). It reflects the extent of the functionality support on the covered systems and can be measured by the number of functions where semantic agreements exist. For example, comparing firm A and B that both adopt UNSPSC, if firm A uses UNSPSC codes in all product-related functions and firm B only uses it in marketing and sales functions, we then say that firm A has better semantic integration.

The task of semantic integration is to eliminate the semantic inconsistency and to enable e-marketplace participants to exchange semantically consistent business information, i.e., to achieve mutual understanding in a manner of higher accuracy, wider reach and more richness. Only achieving this task, the e-marketplace activities can then be done in a healthy and smooth manner. For example, an enquiry of buyer can be exactly interpreted by a potential seller, or a procurement order can be accepted by

a buyer exactly as what the seller means without misunderstanding or unfavourable legal consequences.

To fulfil the above-mentioned task, a collaborative integration system is promising if it can satisfy four heterogeneous e-marketplace properties of distribution, autonomy, interdependence and emergence (Guo, 2007). To investigate how to present such a system that could achieve mutual understanding in exchanging business information, the next section first describes the role of collaboration in semantic integration.

3 Role of collaboration in semantic integration

A role is often functional (Strijbos et al., 2005) and can also include goal (Pitts and Edelson, 2004) and knowledge (Gerdt and Sajaniemi, 2006). It is the part that one has to play and undertake in a particular social context with the particular behaviours in adapting to the one's environment for interaction. In e-marketplace design, the function of the semantic integration role is to maintain semantic consistency in business concept creation and edition by cooperating between their semantic interactions initiated from their heterogeneous contexts. In this paper, we call this type of roles as the *role of collaboration*.

Collaboration as a role for semantic integration in e-marketplace is a mandatory requirement for constructing an interoperable e-marketplace. It is necessary because of the following reasons:

- E-marketplace is a common business information space (Guo, 2007), where different heterogeneous e-business systems are required to work together and to resolve their semantic consistency issue during business concept exchange.
- Business concepts are created in different distributed contexts of heterogeneous e-business systems (Yu et al., 2008).
- Each context is a semantic community (Robinson and Bannon, 1991), which is the cause of heterogeneity.
- People in different semantic communities have different modality judgements (Chandler, 2009) for the truth value assignment to a representation and thus experience ontological drifts (Robinson and Bannon, 1991) in representing semantically consistent business concepts.
- Non-negotiated mental association could only offer perceived similarities in form (e.g., homophones) or meaning (e.g., synonyms) of words (Saussure, 1966, pp.123–126).
- The contents of the representation of organisational context or the classification schemes of the represented contents are contingent and negotiable. Contents and schemes should be constructed and maintained cooperatively (Schmidt, 1994).
- Collaboration can be used effectively to resolve conflict or advance shared vision (Gary, 1989).

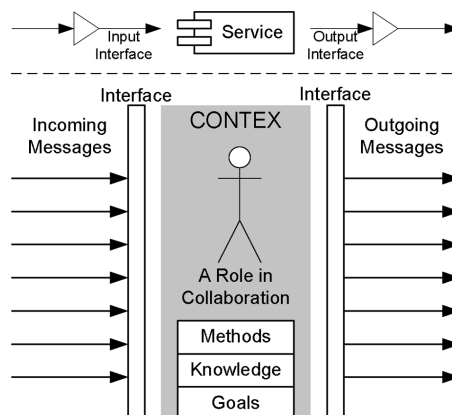
3.1 Definition of collaboration role

While collaboration is one of the important roles of e-marketplace, a collaboration role can be defined as a part such that it has its context that surrounds the part for its behaviours that coordinate with other parts to fulfil common tasks. In this context, a collaboration role has its memory (i.e., contextual knowledge), its belief (i.e., goal), its behavioural habits (i.e., methods), its own perception (i.e., incoming messages) and its own reaction to the external world (i.e., outgoing messages). For example, when you play the role of Hamlet in the play of Shakespeare with other roles in the play, you are in the context of Hamlet and are one of the collaboration roles, where you remembered Hamlet's father (i.e., his knowledge), kept thinking about his revenge (his belief), tried different means (i.e., his methods), and perceived and behaved as he did (i.e., his input and output). But, since you are in the play, you also need to collaborate with other roles in case that they do not exactly follow the script of the play. A collaboration role is, thus, a contextual, autonomous, functional yet interdependent part of the play as what script frames. *Contextual* means that the role is in a distributed location with its own perspective; *autonomous* refers to that the role has its self-regulation of actions; *functional* means that the role has particular tasks; *interdependent* implies that the role has relations to other roles. The total actions and behaviours of the role can be regarded as a service to other interacting roles in collaboration.

With this definition, a collaboration role always executes an interaction function to not only understand others with its own knowledge and perspective but also agree with others to achieve semantic synchronisation in meaning understanding and finally to fulfil its tasks. Between collaboration roles, this mutual understanding can be perceived as a type of mutual services, which constitute a collaborative service system.

The *service of a collaboration role* can be described in Figure 1, where a role collaborates with other roles by receiving incoming messages to respond with outgoing messages via the internal processes by examining its *goals*, *knowledge* and *methods* in its own context. Speaking in a system design way, all these of a collaboration role can be wrapped in a generic *service component* consisted of multiple *objects*, as shown in the upper part of Figure 1, open to receiving *input* from some collaboration roles through *input interface* and sending *output* for collaboration with other collaboration roles through *output interface*.

Figure 1 A collaboration role for semantic integration



More accurately, in Figure 1, a collaboration role as a collaborative service component has its particular collaborative service context in heterogeneous e-business systems, its goals to control its collaboration behaviours, utilise its knowledge to understand the collaborative inputs, apply its methods to handle external collaborative inputs, and use its interfaces to receive and respond to other collaborative services.

3.2 Collaboration roles in semantic integration

Semantic integration is to achieve mutual understanding during business information exchange between heterogeneous e-business systems in e-marketplace. This is the goal of collaboration role of e-marketplace. In this paper, we achieve this goal through collaborative work between collaboration roles by means of collaborative concept creation, mapping and transformation, as exemplified in Guo (2008). Particularly in e-marketplace construction, which requires exact semantic consistency for business communication, two types of collaboration roles can be detected and classified. They are human and agent. This classification is not new but its further discussion is helpful, as it treats human as the core services of semantic consistency maintenance for achieving mutual understanding, which is different in artificial intelligence study where computer agents can also be responsible for semantic conflict resolution.

3.2.1 Human

Human as a collaboration role is the primary means of resolving semantic conflicts in business history (e.g., contract negotiation and dispute arbitration). In e-marketplace, human takes the collaboration role to collaborate with other human to design and edit mutually understandable business vocabularies, documents and processes (Guo, 2008). Human as a collaboration role is inevitable. This is because, in distributed and heterogeneous contexts, human creates ambiguities. These ambiguities have to be resolved by human itself through collaborative negotiation. In current technology conditions, we still cannot find any other means that can resolve human ambiguities. For example, an order with quantity of '1,000,000' gloves is difficult for any computer agent to disambiguate without any explicit rules. Only human can eliminate such ambiguity by collaborative negotiation through their intrinsic knowledge relevant to their contexts. Thus, human is the first collaboration role to be responsible for semantic integration.

3.2.2 Agent

Agent acting as a collaboration role is the second means of resolving semantic conflicts. Unlike human, an agent cannot make correct decision on behalf of its human master when it cannot tell true or false. It can only react correctly if and only if there are rule sets that stipulate its behaviours and these rule sets are explicit and complete. For instance, given the example of the above-mentioned paragraph, if the rule set includes an explicit rule that the quantity of a product must accompany a 'unit for quantity', the agent could then disambiguate the semantic conflict by sending a confirmation message to the glove order originator, asking for what the glove unit is. Thus, agents are the secondary collaboration role in semantic integration.

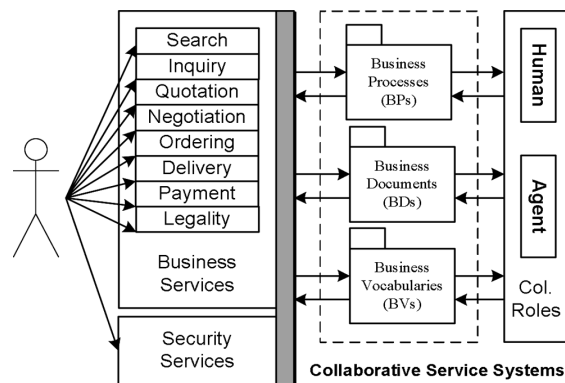
Of course, for a particular collaboration role, it can be played by many e-marketplace participants or service-based agents, which are grouped by groups such as common concept designers, local concept designers and business concept users.

3.3 Collaboration roles in e-marketplace

In this paper, roles provide all sorts of services in an e-marketplace. An e-marketplace is a common business information space that aggregates numerous service functions or simply services. These services can be classified into three categories: *business services* such as matching buyers and sellers, facilitation of transaction, and institutional infrastructure building (Bakos, 1998), *security services* such as e-marketplace confidentiality, integrity and availability, and *information exchange services* such as semantic messaging for business documents and business processes. Particularly, collaboration role provides semantically consistent information exchange services by collaboratively creating and editing Business Vocabularies (BVs), Business Documents (BDs) and Business Processes (BPs) for e-marketplace facilitator to provide other business and security services.

Responsible for the above-mentioned service functions, collaboration roles of human and agents in e-marketplace can be depicted in Figure 2, where each collaborative role provides its own service and collaborates with others to aggregate a collaborative service system.

Figure 2 Collaboration roles in e-marketplace



In Figure 2, service systems such as design systems of BVs, BDs and BPs could be the result of the joint services of multiple collaboration roles of human or agents. These service systems constitute the parts of the technical infrastructure of a semantically integrated and interoperable e-marketplace with regard to the creation and the edition of semantic consistent business concepts throughout the provision of other e-marketplace services such as search, enquiry, quotation, negotiation, ordering, shipment, payment, legality and security used by all e-marketplace participants.

For instance, many human collaborators can work together through agent collaborators to collaboratively edit numerous semantically interoperable common vocabularies. These common vocabularies then are further used by other levels of collaborative roles to create semantically consistent business documents and business processes, and finally build many business and security services.

4 Collaboration role-based semantic integration mechanism

In this section, we will apply the collaboration role discussed earlier to design a novel CRB-SIM, and use it as a collaborative *semantic integration* system to describe how multiple collaborative roles can work together to fulfil the task of semantic integration, i.e., how to achieve mutual understanding in the exchange of business concepts, documents and processes between heterogeneous e-business systems during constructing a semantically interoperable e-marketplace. To make it more intuitive, we compare the CRB-SIM to a play. As we know, a play is specified by a play script that frames many role stereotypes. Each role stereotype can be played by many role actors. Each role actor has its own living context that formulates its own understanding of how to act (i.e., methods on its knowledge).

4.1 Model of CRB-SIM

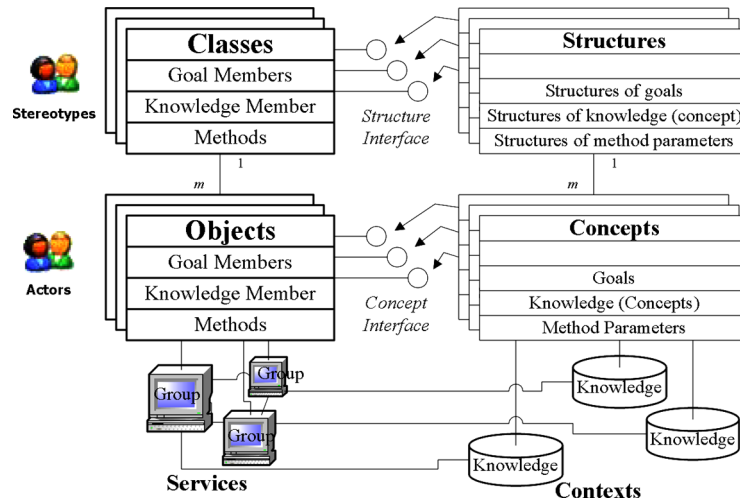
4.1.1 Model description

The metaphor of CRB-SIM is a play, which can be modelled as follows:

- *CRB-SIM as a play*: A CRB-SIM is a *collaborative semantic integration system* behaved like a play, in which there are many role stereotypes fixed by the *play script* to specify how each stereotype should play. A collaboration role *stereotype* is a *class* that abstracts the stereotype with *goal members*, *knowledge member* and *methods* for behaviours. For each stereotype, it can be played by many role *actors* in the form of *objects*. Each object is instantiated from class, conceptualises goal members, knowledge members, and implements methods. Multiple objects work together collaboratively through *interfaces* to pass *messages* of semantic integration.
- *Structures as a play script*: A play script is fixed during the play. All stereotypes cannot change the script. For this reason, we regard a script as *structures* (Guo, 2008, 2009) (similar to schemas) of all participated stereotypes such that an abstract class cannot be changed over its life (i.e., within a CRB-SIM version). It consists of *structural schemas* referenced by the goal members for rule setting, the knowledge members for conceptualisation, and the functions for behaviours.
- *Concepts as actors' understanding*: Since a stereotype is played by many actors, the understanding of the play script by actors is different. This causes the *concept inconsistency* [0, 0] in acting in the play such that each actor (in terms of an object) has its own *concepts* for its knowledge on same structural schemas.
- *Interfaces as channels for collaboration services*: In a real play, misunderstanding is resolved by face-to-face interaction. In CRB-SIM, an *interface* between two actors serves collaboration for passing concepts as collaboration messages to resolve semantic conflicts for mutual agreements of meaning understanding.
- *Knowledge as a group of concepts*: Different actors of a same context share a set of commonly understandable concepts, which are knowledge of a *context*. The actors sharing same knowledge are a *group*. Actors in a same group provide the same *collaboration services*.

With the above-mentioned description, a CRB-SIM can be generally described in Figure 3, where a stereotype is played by actors sharing a common collaboration class referenced to a set of schemas (i.e., structures) for conveying concepts. Each actor as an object has its own concepts of vocabulary, documents and processes referenced to its context.

Figure 3 CRB-SIM model (see online version for colours)



With this model, stereotypes eliminate the *syntactic inconsistency* by means of fixed mapping of predefined structures in different contexts through structure interfaces, for example, the same definitions for class interface structures and parameter lists. Actors eliminate the *semantic inconsistency* of heterogeneous concepts between different contexts through dynamic collaborative concept passing from their particular vocabularies through concept interfaces, for example, the collaborative negotiation of parameter list instantiation for predefined interface parameters.

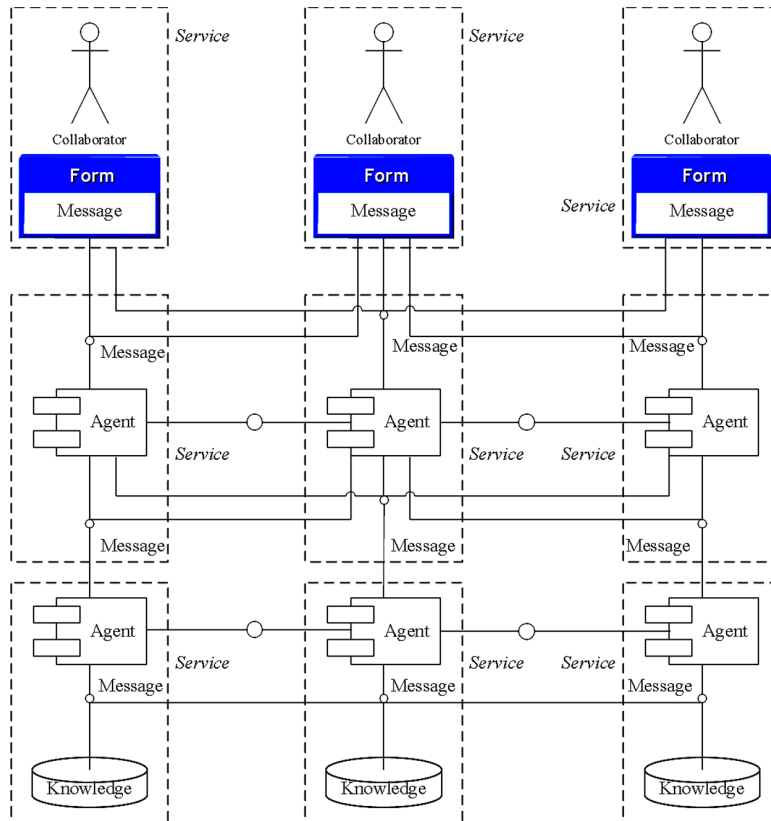
4.1.2 Encapsulation mechanism

The CBR-SIM model satisfies the e-marketplace properties of distribution, autonomy, interdependence and emergence (Guo, 2007) by designing an encapsulation mechanism. Particularly, this mechanism is designed as a CBR-SIM interface system, shown in Figure 4, which is used to separate the semantic heterogeneities in individual contexts.

In this interface system, there are two types of interfaces. *Human-to-agent interface* describes the ability of collaboration between a human and an agent through service provision and consumption. This type of interface is designed in the form of *user interfaces*, such that a form as a *structure* can be understood by both human and agent, and a human knows how to fill concepts in the form as messages for an agent to pass to another agent or human. This type of interaction is the primary approach for resolving semantic consistency issue during business concept creation and edition in CRB-SIM. *Agent-to-agent interface* describes the abilities of collaboration between any two agents in service provision and consumption. It is abstracted as a structure in the form of a parameter list that both interaction agents agree. Concepts are mapped and passed as

messages through the structure from the sending agent to receiving agent for interpretation based on the receiving agent’s goal, knowledge and methods.

Figure 4 CBR-SIM interface system (see online version for colours)



By this interface system, the heterogeneous contexts of individual human environment are encapsulated and separated, satisfying the properties of autonomy, distribution and emergence of e-marketplace. What’s more, the task of achieving mutual meaning understanding for semantic consistency maintenance between human is collaboratively fulfilled through combing human-to-agent interfaces and agent-to-agent interfaces, where the interdependence property of e-marketplace is further satisfied.

4.2 Generalisation of CRB-SIM model

On the basis of the above-mentioned idea description, the CRB-SIM can be generalised as a tuple $\Sigma ::= \langle C, O, H, A, S, K, F, M, R, G, I, X, E \rangle$, where:

- C* A set of classes
- O* A set of objects
- H* A set of human collaborators
- A* A set of agents

- S* A set of schemas as fixed structures
- K* A set of knowledge in different contexts
- B* A set of methods that actors behave
- M* A set of messages as conveyed concepts
- R* A set of roles with different goals
- G* A set of groups where roles are grouped
- I* A set of interfaces for message passing
- X* The initial state of the CRB-SIM system
- E* A set of editorial arbitrators for making final decision on semantic consistency.

In this system, human collaborators H and agents A are actors who play their collaborative roles R specified by the role groups G where their goals are set. Actors behave in a specific way set by behavioural methods B using their contextual knowledge K . Their knowledge is structurally framed in the predefined schemas S . They interact with each other for semantic consistency maintenance through passing messages M (i.e., concepts) via their own interfaces I of services that are components made of a set of objects O (including knowledge objects) and a set of interfaces I . The objects are instantiated from classes C . Since actors may concurrently edit a same concept that is represented differently, a special role of arbitrators E is required to make the final decision among the semantic conflicts.

4.3 Semantic consistency maintenance approach in CRB-SIM model

In CRB-SIM model, the most important task is the semantic consistency maintenance for semantic integration of various heterogeneous messages between contextual e-business systems in e-marketplace. This task is fulfilled in the following procedures:

Step 1: Actors negotiate to join their collaboration role groups. Given an initial state X of the CRB-SIM system, actors negotiate to specify the roles relevant to their goals of semantic consistency maintenance to form role groups such that $G ::= \langle H_0, A_0, G_s, R, K, B \rangle$ where H_0 and A_0 are empty sets, G_s is a set of subgroups, R is the role of the group, K is the existing knowledge for the group and B is a set of methods for group behaviours. If compromises or agreements are made, role groups are created and then the collaboration goes to Step 2 otherwise it is aborted.

Step 2: Actors assign their collaboration roles for semantic consistency maintenance. Each actor is assigned one or more group roles to play such that A or $H ::= \in (G_1, \dots, G_n)$. If agreements are made, the collaboration goes to Step 3 otherwise abort.

Step 3: Actors play their roles. Actors of human collaborators and agents work according to their roles goals until the task of semantic consistency maintenance is accomplished with success. In this step, two sub-procedures are experienced, which are:

Step 3.1: Actors check incoming messages M from interfaces I to consume collaboration services and to find out whether they are semantically consistent with their own concept understanding. If they are consistent, then collaboration

goes to Step 3.2. If conflicts occur and cannot be resolved within the actors' abilities, then collaboration goes to Step 4.

Step 3.2: Actors issue outgoing messages M through interfaces I to provide collaboration services or ask for services from other actors until their tasks are accomplished.

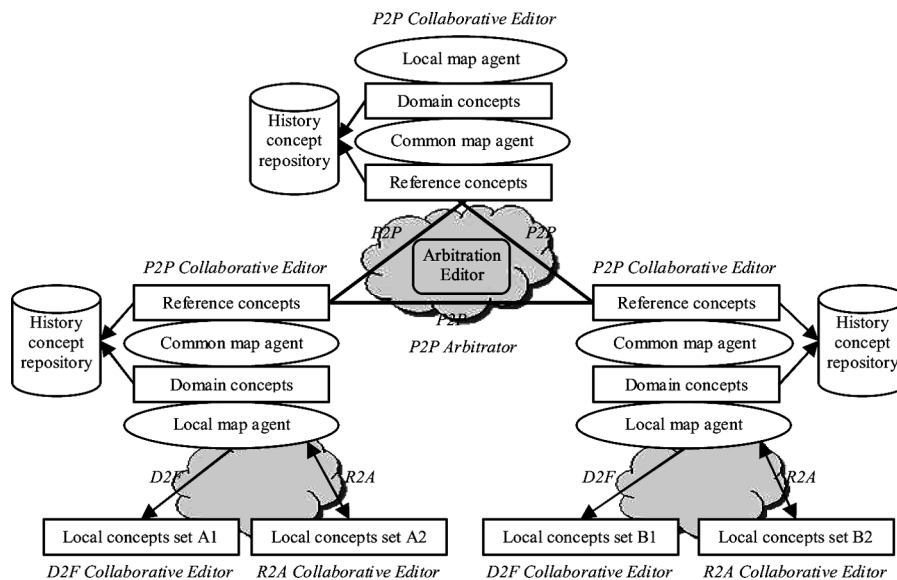
Step 4: Editorial arbitrators E receive the unsolvable conflicts of concepts between any two actors such that $K_1 \neq K_2$ and make final decision such that $K_1 \rightarrow K_2$ or $K_1 \leftarrow K_2$. An editorial arbitrator can be either a human or an agent or a combination such that $E:: \in \langle H, A, (H, A) \rangle$.

In the above-mentioned procedures, the implication of Step 1 is that a group of actors can flexibly include subgroups, contextual knowledge and any number of functions. Thus, CRB-SIM is flexible. Step 4 is unique by introducing the roles of editorial arbitrators to maintain the policy of final arbitration. This guarantees that CRB-SIM system is semantically consistent by its own and avoids the intrinsic semantic errors that may be gradually developed to destroy the whole system.

4.4 Framework of implementing CRB-SIM model

Particularly, the CRB-SIM model can be implemented in a hierarchical collaborative framework, shown in Figure 5, which fulfils the task of achieving semantic consistency between local concept sets A and local concept sets B . In this framework, there are three types of human collaboration namely P2P, D2F and R2F, and three types of agent collaboration namely L2C, C2C and C2L.

Figure 5 Hierarchical collaborative framework



$P2P$ is a peer-to-peer collaboration relationship between domain concept designers of heterogeneous contexts to maintain semantic consistency between heterogeneous

domain concepts (i.e., concepts shared within a domain) through *reference concepts* (i.e., concepts in a single natural language as system-wide standard concepts). Since P2P collaboration may incur some important semantic editing problems such as circular editing or multiple copies of concept definitions, an *arbitration editor* must be provided to finalise a usable copy. An *arbitrator* is often an experienced knowledge engineer whose better quality can improve the time-to-market of newly designed concepts and slow down the speed of needing to re-edit those badly edited concepts or to obsolesce those badly edited concepts. In addition, owing to the emergence property of e-marketplace, some concepts are gradually changing and become obsolete. Direct deleting of these concepts will severely cause global semantic consistency problem. Thus, all obsolete concepts must be maintained in a *history concept repository* so that the old concepts in reference will not be lost.

D2F is a dominator-to-follower collaboration relationship between domain concept designers and local concept designers to maintain semantic consistency between domain concepts and *local concepts* (i.e., personalised concepts in different forms but semantically the same as a subset of domain concepts). In this relationship, local concept designers have no right to modify or add new concepts to domain concepts but can only localise the existing domain concepts to local concepts. Similarly, *R2A* is a requestor-to-answerer collaboration relationship between domain concept designers and local concept designers like *D2F*. The difference is that, in *R2A*, local concept designers cannot find the needed concepts in domain concept set. Thus, they request the domain concept designers to create the needed new domain concepts, so they can find their needed domain concepts for local concept localisation. In general, *P2P*, *D2F* and *R2A* manage the human collaboration relationships for creating and editing reference concepts, domain concepts and local concepts through three types of editors: *P2P collaboration editor*, *D2F collaboration editor* and *R2A collaboration editor*, which are all human user interfaces.

Differently, *L2C* is a local-to-common agent collaboration relationship between local concepts and domain concepts. It maps and transforms local concepts to domain concepts. *C2C* is a common-to-common agent collaboration relationship between domain concepts. It maps and transforms domain concepts to reference concepts to ensure that heterogeneous domain concepts are exactly connected and transformed. *C2L* is a common-to-local agent collaboration relationship between domain concepts and local concepts. It is a reverse of *L2C* agent collaboration relationship. *L2C*, *C2C* and *C2L* are implemented by two types of agents, i.e., *common map agent* and *local map agent*.

4.5 Features of CRB-SIM

The CRB-SIM model is used to maintain semantic consistency for collaborative integration of heterogeneous concepts in the forms of vocabulary, documents and business processes. It has the following features:

- *Accurate*: In CRB-SIM model, the semantic consistency of concepts that are created and edited in heterogeneous contexts is accurately maintained. A concept can accurately reflect the exact understanding between collaborative concept designers. For example, in the creation of a concept ‘refrigerators’, if actor 1 uses concept ‘fridge’ while actor 2 uses concept ‘réfrigérateur’, it is difficult to confirm that these two concepts all refer to ‘refrigerators’. However, if the two actors share a

collaborative role, they may communicate and discuss with each other to resolve the semantic conflict between ‘fridge’ and ‘réfrigérateur’.

- *Flexible*: CRB-SIM model is modularly designed through encapsulation such that a set of objects builds a service component, and multiple service components build a service system. This makes it flexible to add or delete functions in a service component and plug in or remove services in a system. This service-oriented design further increases the system reusability and scalability.
- *Separation of structure from semantics*: CRB-SIM model clearly separates knowledge schemas from the concepts conveyed in the schemas. This separation detaches the design of system structures (which is a responsibility of system designers) from the design of business concepts (which is a responsibility of concept designers), and makes the semantic integration task easier to accomplish.

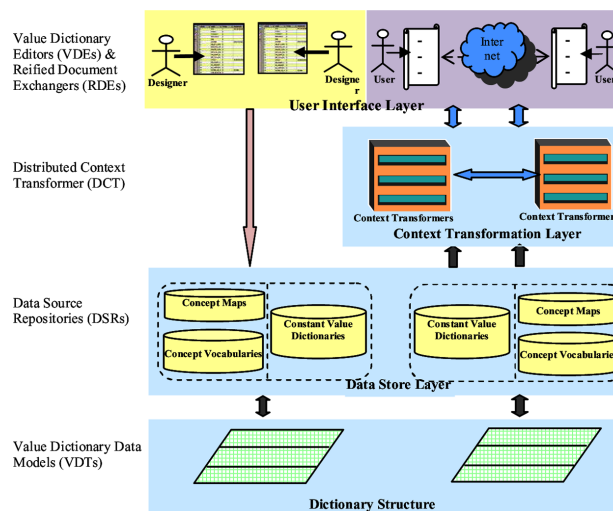
5 Applicability of CRB-SIM model

The CRB-SIM model can be applied in many areas of e-commerce and information systems such as interoperable electronic catalogue system design, e-marketplace construction, e-marketplace activity inference, inter-organisational cooperation and multilingual translation. In this subsection, two application prototypes (see demo screen shots¹) are introduced to demonstrate the applicability of the approach.

5.1 CAT prototype

CAT (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. Its system framework can be illustrated in Figure 6.

Figure 6 CAT system framework (see online version for colours)

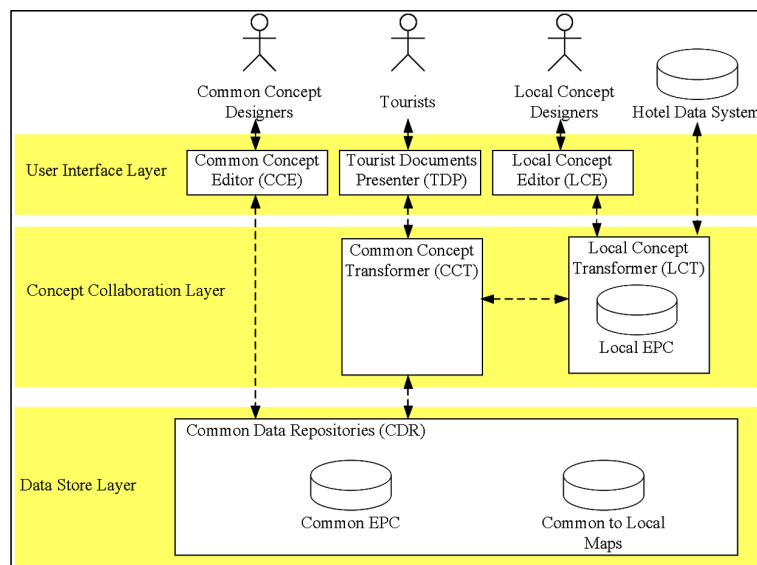


In this system, the collaborative role is dictionary designers, who collaboratively design reified multilingual terms (i.e., constant values) to reach semantic agreements between multilingual terms. The evaluation experiment on document translation showed that the translation accuracy between Chinese 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 when compared with the experiments on Google Translate (42–100%) and Babel Fish (13–96%). The evaluation is based on the golden standard measure. The comparison evaluation result is obtained by sampling through questionnaire. The 8% of accuracy gap comes from the lack of enough mapped terms for translation in multilingual vocabularies. The gap indicates that the future research must be made by improving term mapping mechanism.

5.2 HICP prototype

HICP (Luo, 2008) is a prototype of distributed online hotel information servicing systems, which aim to resolve semantic conflicts between the service provider's CEPC and numerous hotels' LEPCs at all levels of structure, concept and context, so accurate hotel accommodation can be offered. Its system architecture is illustrated in Figure 7.

Figure 7 HICP system architecture (see online version for colours)



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

The evaluation experiments on HICP prototype given in Table 1 showed that HICP has a big improvement on existing well-known online hotel information systems such as expedia.com, hotels.com and elong.com.

Table 1 Experiment result of HICP

	<i>Quantity of features</i>	<i>Has attributes features</i>	<i>Has attributes ratio (%)</i>	<i>Searchable features</i>	<i>Searchable features ratio (%)</i>
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

If a feature has details or attributes to describe it, then the feature 'has attributes features' group, otherwise, the feature has 'no attributes features' group.

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.

6 Previous and related works

6.1 Previous works

Collaborative integration concept can be traced back to Anderson's work (Anderson, 1993, 1995), which was used to understand the collaboration relationship between small and large firms and their collaborative linkages that needed to be developed. The early work directly related to CRB-SIM model regarding semantic integration by collaboration was a COPE collaboration framework (Guo and Sun, 2003a), which proposed to create product concepts through a collaborative editor. With the more understanding of the impacts of heterogeneous business contexts on product representation (Guo and Sun, 2003b), semantic similarity is known as an insufficient means to business document exchange and thus semantic equivalence for accurate exchange of business concepts such as enquiries, offers and orders is required.

When the importance of ensuring the semantic equivalence between heterogeneous business concepts was found, the research work of CONEX (Guo, 2008, 2009) (which was published four years later) formally introduced collaboration as an important mean of semantic integration for maintaining semantic consistency between heterogeneous contexts of distributed ad hoc product catalogues. In this work, the Product Map theory was developed as a foundation of collaborative concept design and the prototype of CONEX design system was created to support the semantic integration by collaboration. However, semantic heterogeneities exist in all levels of business information, business documents and business processes. To control the complexity of semantic consistency and simplify the possible solution, the paper (Guo, 2006a) proposed a hierarchical TRANSCODE representation of business information, documents and processes. This representation enables any business documents to be created by vocabularies and any business processes to be created by documents and vocabularies. This idea makes it possible to resolve all types of semantic inconsistency if semantic consistency of

heterogeneous business vocabularies can be maintained. For example, based on this representation, the CODEX research (Guo, 2006b) provided a collaboration solution to business document template design and reification. The COPEX research (Guo et al., 2008) proposed a document-oriented business process representation method, where a business process can be assembled and reassembled in run-time. By this approach, heterogeneous business process patterns can be semantically interoperable in any time.

6.2 Related works

The CRB-SIM model, evolved from the above-mentioned previous works, relates to two aspects of works: semantic integration and role-based collaboration system.

6.2.1 Semantic integration approaches

Semantic integration approaches can generally be classified into three categories of mandatory standardisation approach, automated mediation approach and collaborative conceptualisation approach (Guo, 2009). *Mandatory standardisation* is an approach that avoids semantic conflicts by enforcing a standard for all participating systems. It is initiated by standard designers or makers in the design stage through defining a shared vocabulary (i.e., a set of ‘commonly acceptable’ product terms). *Automated mediation* is an approach that undertakes the semantic conflict detection and resolution using an intelligent mediation system that is independent of heterogeneous systems. This mediation system often acts as an information middleware to intelligently transform heterogeneous information between heterogeneous systems. *Collaborative conceptualisation* is an approach that avoids semantic conflicts and creates semantic consistent concepts through a collaboration mechanism. This mechanism is presented as a type of collaborative editing systems, targeting at resolving semantic conflicts between business systems.

Standardisation approach is widely used. Most industrial solutions to semantic integration adopt standardisation. These include UNSPSC (unspsc.com) and ecl@ss (eclass.de) for product classification, EDI (United Nations, 1987) for proprietary digital information exchange (particularly document messaging) and BPEL² for business process design and execution on web service (w3.org/TR/wsdl). Standards are important in providing common senses between heterogeneous semantic communities, yet, on the other hand, they are rigid and time-consuming to renegotiate the new common senses confronting with the emergent changes of new business requirements. In addition, the current situation is not that we lack of standards but we have too many mutually conflicting standards. The issue of conflicting standards was explicitly discussed by some researches with attempts to integrate multiple standards. For example, Omelayenko and Fensel (2001) attempted to integrate xCBL and cXML through a layered approach on a given ontology. Bergamaschi et al. (2002) tried to integrate UNSPSC, ecl@ss and other standards by providing a common thesaurus. However, an open issue for these standard integration approaches is that some concepts in two standards are larger than their concept overlapping and are not mappable.

Automated mediation approach is also widely used in e-marketplace construction. Classically, two well-known examples of this approach are Smart Catalogues and Virtual Catalogues (Keller and Genesereth, 1996) and NetAcademy (Lincke et al., 1998).

The former introduces the facilitator concept to perform routing and translation between distributed product catalogues and catalogue web interfaces based on a set of ontologies (Gruber, 1993). The latter introduces the mediator concept (Linche and Schmid, 1998) as an integrated catalogue to mediate the distributed catalogues through a merger called Q-Calculus (i.e., a common product description frame, which is a formal language for description and classification of objects). Modern automated mediation approach often applies technologies of multi-agent systems (e.g., Dani et al., 2007; Gates and Nissen, 2001) and ontology systems (e.g., w3.org/TR/owl-ref/) to construct an e-marketplace as a matchmaking system (e.g., Noia et al., 2003; Veit, 2004) or a brokering system (e.g., Antoniou et al., 2007; Du et al., 2004; Hämäläinen et al., 1996; Segev and Beam, 1999) between various e-business systems. However, mediation systems based on agent and ontology technologies have a serious problem such that the new system is still domain-wide and the activity inference must have fully supported rules, otherwise the semantic consistency cannot be maintained between heterogeneous systems.

To solve the problem that standardisation and mediation cannot solve, the research group of Guo et al. (Guo, 2006b, 2008, 2009; Guo et al., 2008) proposed the collaborative conceptualisation approach, where a semantic consistency maintenance model was developed based on collaborative mapping of all heterogeneous concepts through both real-time and non-real-time collaboration between business concept designers. This approach is promising because it has separated the structure semantics from concept semantics. With this separation, not only concepts can be messaged between multiple parties in semantically consistent structures but also the semantically consistent concepts can be created and edited through the given collaboration mechanism.

6.2.2 Collaboration role-based design approaches

Collaboration role-based system design methodology has been proposed in many different application contexts such as collaborative heterogeneous resource management and sharing (Jin and Ahn, 2006; Park and Hwang, 2003), view sharing (Berry et al., 2005) and collaborative design (Seland, 2006).

Initially, the role concept was used to model behaviours of an object in Pernici (1990), where roles allow one to describe different perspectives for object evolution. This provided the foundation of observing an object in different contexts where each contextual perspective relates to a role with a partition of messages. It is rather similar to an interface of an object for providing and consuming services.

In role-based access control researches (e.g., Nynachama and Osborn, 1999; Park et al., 2001), the terms of privilege, role, user and groups are often used to build authorisation systems. These provide the policy framework of role collaboration, for example, the role of editorial arbitrators designed in CRB-SIM model of this paper.

In Park and Hwang (2003), web service was adopted to P2P collaboration to support role-based system for managing heterogeneous resources. This assists us to improve the design of reusable collaborative concept editing systems through the introduction of service provision and consumption.

Strijbos et al. (2005) defined roles as more or less stated functions/duties or responsibilities that guide individual behaviour and regulate intra-group interaction. They thought that “roles appear to be most relevant when a group pursues a shared goal requiring a certain level of task division, coordination and integration of individual

activities” (Strijbos et al., 2005). This helps explain why CRB-SIM model must adopt collaboration as a role for semantic integration.

Seland (2006) described role play in IT design through a qualitative study. The research showed that role play is useful for collaborators in development process, creating a focus on user needs, fast idea creation, and enhancing the developers’ understanding of the future context-of-use.

The CRB-SIM model designed in this paper is influenced by both thoughts of collaboration as a means of discrepancy resolution between roles, and roles as service components for collaborative message passing to achieve modularity and reuse of system development.

7 Conclusion

This paper has proposed a CRB-SIM for semantic consistency maintenance in semantic integration during heterogeneous concept creation and edition. It first elaborates the task of semantic integration on e-marketplace, and then the collaboration roles are defined with regard to semantic integration and e-marketplaces. Applying this definition, a CRB-SIM model has been described and designed as a collaborative semantic integration system with the features of accuracy, flexibility and the separation of structures from semantics. These features outline CRB-SIM system as a set of role-based service components, where on the one hand semantic consistency during concept creation and edition has been maintained, and on the other hand CRB-SIM has achieved modularity and reusability in system design for easy plugging in and removing service components. On the basis of the general CRB-SIM model, two system prototypes of CAT and HICP have been described to illustrate the applicability of CRB-SIM model.

The CRB-SIM model has the following contributions:

- the collaboration role-based semantic consistency maintenance method for eliminating semantic conflicts during business concept creation and edition
- the separation of concept structures (or schemas) from concept semantics created and edited by concept designers
- the service-oriented design in collaboration role construction for modularity and reusability, which is especially important for any growing e-marketplaces.

The CRB-SIM model is important. In business semantic integration theory, the CRB-SIM model has provided a principle such that human is always more important than agent in semantic conflict resolution. Human is always the masters and agent is always the slaves because of the understanding gap between human and agent. In management thought, the CRB-SIM model has recognised that collaboration between business roles/parties will create better business services that might bring more revenues.

The CRB-SIM model has provided a paradigm for future e-marketplace design, especially for those focusing on semantic integration. Future works would be proposed towards the specific designs of industry-wide e-marketplaces based on CRB-SIM model.

References

- Anderson, M. (1993) 'Collaborative integration in the Canadian pharmaceutical industry', *Environment and Planning A*, Vol. 25, No. 12, pp.1815–1838.
- Anderson, M. (1995) 'The role of collaborative integration in industrial organization: observations from Canadian aerospace industry', *Economic Geography*, Vol. 71, No. 1, pp.55–78.
- Antoniou, G., Skylogiannis, T., Bikakis, A., Doerr, M. and Bassiliades, N. (2007) 'DR-BROKERING: a semantic brokering system', *Knowledge-Based Systems*, Vol. 20, No. 1, pp.61–72.
- Bakos, Y. (1998) 'The emerging role of electronic marketplaces on the internet', *Communications of the ACM*, Vol. 41, No. 8, pp.35–42.
- Bergamaschi, S., Guerra, F. and Vincini, M. (2002) 'A data integration framework for e-commerce product classification', *Proc. of ISWC 2002*, LNCS 2342, Spring, pp.379–393.
- Berry, L., Bartram, L. and Booth, K. (2005) 'Role-based control of shared application views', *Proc. of ACM UIST'05*, USA, pp.23–32.
- Chandler, D. (2009) *Semiotics for Beginners*, 2 November, <http://www.aber.ac.uk/media/documents/S4B/semiotic.html>
- Dani, A.R., Pujari, A. and Gulati, V.P. (2007) 'Strategy proof electronic markets', *ACM Proc. of ICEC'07*, USA, pp.45–54.
- Du, T., Li, E. and Wei, E. (2004) 'Mobile agents for a brokering service in the electronic marketplace', *Decision Support Systems*, Vol. 39, No. 3, pp.371–383.
- Evans, P. and Wurster, T. (1999) 'Getting real about virtual commerce', *Harvard Business Review*, Vols. 11–12, pp.85–94.
- Gabrilovich, E. and Markovitch, S. (2007) 'Computing semantic relatedness using Wikipedia-based explicit semantic analysis', *Proc. of IJCAI-07*, 6–12 January, Hyderabad, India, pp.1606–1611.
- Gary, B. (1989) *Collaborating: Finding Common Ground for Multiparty Problems*, Jossey-Bass, San Francisco, USA.
- Gates, W. and Nissen, M. (2001) 'Designing agent-based electronic employment markets', *Electronic Commerce Research*, Vol. 1, No. 3, pp.239–263.
- Gerdt, P. and Sajaniemi, J. (2006) 'A web-based service for the automatic detection of roles of variables', *ACM Proc. of ITiCSE'06*, Italy, pp.178–182.
- Gruber, T. (1993) 'A translation approach to portable ontologies', *Knowledge Acquisition*, Vol. 5, No. 2, pp.199–220.
- Guan, X. (2008) *Context-based Translation of Constant Concept Values in E-Business*, Master Thesis, University of Macau, Macau.
- Guo, J. (2006a) 'Achieving transparent integration of information, documents and processes', *Proc. of IEEE ICEBE'06*, pp.559–562.
- Guo, J. (2006b) 'Inter-enterprise business document exchange', *ACM Proc. of ICEC'06*, Canada, pp.427–437.
- Guo, J. (2007) 'A term in search of the infrastructure of electronic markets', *Research and Practical Issues of Enterprise Information Systems II Volume 2*, IFIP, Vol. 255, Springer, pp.831–840.
- Guo, J. (2008) *Collaborative Concept Exchange*, VDM Publishing, Germany.
- Guo, J. (2009) 'Collaborative conceptualization: towards a conceptual foundation of interoperable electronic product catalogue system design', *Enterprise Information Systems*, Vol. 3, No. 1, pp.59–94.
- Guo, J. and Sun, C. (2003a) 'Collaborative product representation for emergent electronic marketplace', *Proc. of Bled eCommerce Conf.*, 9–11 June, Bled, Slovenia, pp.847–859.

- Guo, J. and Sun, C. (2003b) 'Context representation, transformation and comparison for ad hoc product data exchange', *Proc. of ACM DocEng'03*, France, pp.121–130.
- Guo, J., Hu, Z., Chan, C-K., Luo, Y. and Chan, C. (2008) 'Document-oriented heterogeneous business process integration through collaborative e-marketplace', *ACM Proc. of ICEC'08*, Article No. 39, Austria.
- Hämäläinen, M., Whinston, A. and Vishik, S. (1996) 'Electronic markets for learning: education brokerages on the Internet', *Communications of the ACM*, Vol. 39, No. 6, pp.51–58.
- Jin, J. and Ahn, G. (2006) 'Role-based access management for ad-hoc collaborative sharing', *Proc. of ACM SACMAT'06*, USA, pp.200–209.
- Keller, A. and Genesereth, M. (1996) 'Multivendor catalogs: smart catalogs and virtual catalogs', *EDI Forum: Journal of Electronic Commerce*, Vol. 9, No. 3, pp.87–93.
- Linche, D. and Schmid, B. (1998) 'Mediating electronic product catalogs', *Communication of the ACM*, Vol. 41, No. 7, pp.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', *Proc. of IEEE HICSS'98*, USA, p.131.
- Luo, Y. (2008) *Semantic Integration of Heterogeneous Online Hotel Information Systems*, Master Thesis, University of Macau, Macau.
- Noia, T., Sciascio, E., Donini, F. and Mongiello, M. (2003) 'A system for principled matchmaking in an electronic marketplace', *Proc. of WWW 2003*, Hungary, pp.321–330.
- Nynachama, M. and Osborn, S. (1999) 'The role graph model and conflict of interest', *ACM Transactions on Information and System Security*, Vol. 2, No. 1, pp.3–33.
- Omelayenko, B. and Fensel, D. (2001) 'A two-layered integration approach for product information in B2B e-commerce', *Proc. of ECWEB-2001*, LNCS 2115, Springer, Germany, pp.226–239.
- Park, J. and Hwang, J. (2003) 'Role-based access control for collaborative enterprise in peer-to-peer computing environments', *Proc. of ACM SACMAT'03*, Italy, pp.93–99.
- Park, J., Sandhu, R. and Ahn, G. (2001) 'Role-based access control on the web', *ACM Transactions on Information and System Security*, Vol. 4, No. 1, pp.37–71.
- Pernici, B. (1990) 'Objects with roles', *Proc. of ACM SIGOIS and IEEE CS TC-OA Conf. on Office Information Systems*, USA, pp.205–215.
- Pitts, V. and Edelson, D. (2004) 'Role, goal, and activity: a framework for characterizing participation and engagement in project-based learning environments', *Proc. of ICLS'04*, ACM Press, USA, pp.420–426.
- Robinson, M. and Bannon, L. (1991) 'Questioning representations', *Proc. of ECSCW'91*, Amsterdam, pp.219–233.
- Saussure, F. (1966) *Course in General Linguistics*, McGraw-Hill Book Company, New York.
- Schmidt, K. (1994) 'The organization of cooperative work: beyond the 'Leviathan' conception of the organization of cooperative work', *Proc. of ACM CSCW'94*, USA, pp.101–112.
- Segev, A. and Beam, C. (1999) 'Brokering strategies in electronic commerce markets', *Proc. of ACM EC'99*, USA, pp.167–176.
- Seland, G. (2006) 'System designer assessments of role play as a design method: a qualitative study', *Proc. of ACM NordiCHI 2006*, Norway, pp.222–231.
- Strijbos, J., De Laat, M., Martens, R. and Jochems, W. (2005) 'Functional versus spontaneous roles during CSCL', *Proc. of CSCL'05*, ACM Press, Taiwan, pp.647–656.
- United Nations (1987) *EDI for Administration, Commerce, and Transport (EDIFACT)*, United Nations.
- Veit, D. (2004) *Matchmaking in Electronic Markets*, LNCS 2882, Springer.

- Wang, S., Zheng, S., Xu, L., Li, D. and Meng, H. (2008) 'A literature review of electronic marketplace research: themes, theories and an integrative framework', *Information Systems Frontiers*, Vol. 10, No. 5, pp.555–571.
- Yu, J., Reddy, Y., Liu, S., Yao, Z., Bharadwaj, V., Reddy, S. and Kankanahalli, S. (2008) 'Adaptive collaboration in an ever-changing environment', *Multiagent and Grid Systems*, Vol. 4, No. 2, pp.125–140.

Notes

¹<http://www.sftw.umac.mo/~jzguo/pages/demo.html>

²<http://www.ibm.com/developerworks/library/specification/ws-bpel/>