TOWARD INTELLIGENT BUSINESS INTEGRATION: THE SEMANTIC WEB

James J. Lee, Seattle University, leej@seattleu.edu
Surinder Kahai, SUNY at Binghamton, kahai@binghamton.edu
Yong Jin Kim, Sogang University, yongjkim@sogang.ac.kr
Ben B. Kim, Seattle University, bkim@seattleu.edu

ABSTRACT

The semantic web allows a flexible exchange of trustworthy content representations across heterogeneous platforms by explicitly formalizing knowledge rather than embedding it in program codes. This paper formulates the conceptual framework from technology, the recursive self-organizing architecture. It is not merely for the world of the semantic web, but it also envisions the directions that we are heading for communication integration through communication instantiation. New communication technology will provide new business opportunities by integrating knowledge, formalized on the front-end arsenals, and rendering autonomous and intelligent business integration. Thus, the focus of systems designers is shifted from standardizing how to connect pre-designed systems with centralized control to institutionalizing how systems interact in an ad hoc manner without centralized control.

Keywords: Business intelligence, Semantic web, Web 2.0, Process integration, Service integration, Communication integration, Business integration

INTRODUCTION

Business integration is touted to add considerable value to companies in terms of efficiency and effectiveness of processes, applications, and communications. In particular, recent development in e-business allows companies benefit from integration both by extending existing business practices and by adding new perspectives to the roles of business partners. The basic premise of business integration through information technology is that any part of business, including business processes and content, can be digitized and embedded into information systems. To achieve integration, businesses are currently deploying technologies such as enterprise systems, EDI, and web services. Yet, interoperability difficulties among heterogeneous information systems prevent business integration from immediate agenda of companies. There have been efforts to provide information systems interoperability semantically in knowledge level, but we focus the framework of the semantic web on communication level due to the semantic web as a useful alternative to integration technologies with its inexpensive and flexible nature [Kajan and Stoimenov, 2005]. For the semantic web being useful, however, there must be tremendous efforts from all participants to assert semantics over existing web sites, yet there needs a conceptual framework to access standard ontologies. Therefore, in this article, we classify and compare the different methods of achieving business integration to highlight useful aspects of the semantic web for it. Then, we propose a conceptual framework of the semantic web to help firms execute business integration without worrying about the standardization of business data, processes and technologies.

Process Integration

In general, there are two common solutions for business integration. The first is process integration which involves the use of proprietary technology platforms such as Enterprise Resource Planning (ERP) systems and Electronic Data Interchange (EDI) systems. Process integration is important to the automation of both back-end and front-end processes. As a consequence of the popularity of this solution, there is increasing pressure on both vendors and users to standardize process models. Integrating business processes, however, is designed and implemented only through human interventions because the integrated process, built on agreements between the parties involved, is determined at design time and remains static. Process integration, therefore, is inflexible. In addition to this, the human intervention also makes process integration costly and time consuming.

Service Integration

The second solution is the “Web Services” framework that facilitates a dynamic approach to service integration at the application level. Within the service integration solution, back-end business processes are encapsulated within services that business partners can employ to integrate their processes. In other words, the web services framework refers to an autonomous distributed integration that can occur dynamically by combining Internet-accessible services at run-time based on a predefined pattern. Accordingly, though companies integrate their services at run-time, encapsulated services need predefined standards, which can limit flexibility.

Communication Integration
With recent development in information technology, business integration is heading toward an automated communication mode [Berners-Lee, 1998]. Organizational information is abstracted into the systems and communicated with other systems with very little human intervention. This means intelligent search and query capability using inference model instead of reference model currently applied on web systems. Inference model pushes business intelligence through advance information and communication technologies (less human intervention) while reference model relies on current hyper-text link structure (significant human intervention). This new trend brings the third solution, communication integration where machine-to-machine communication is instantiated at run-time without human intervention to give rise to a *communication instantiation*. This intelligent business integration tends to increase the role of information systems while deemphasizing the importance of human interventions in both design-time and run-time of systems integration, as well as human roles in organizational structure. The main problem with communication integration is how to develop the conceptual framework that integrates the ontologies of the semantic web and database of web systems. Because an ontology describes the meaning and relationships of terms of the data definitions from a specific database, there needs a conceptual model to describe ontologies of the semantic web, considering an ontology as a physical model.

Figure 1: Business Integration Topology.

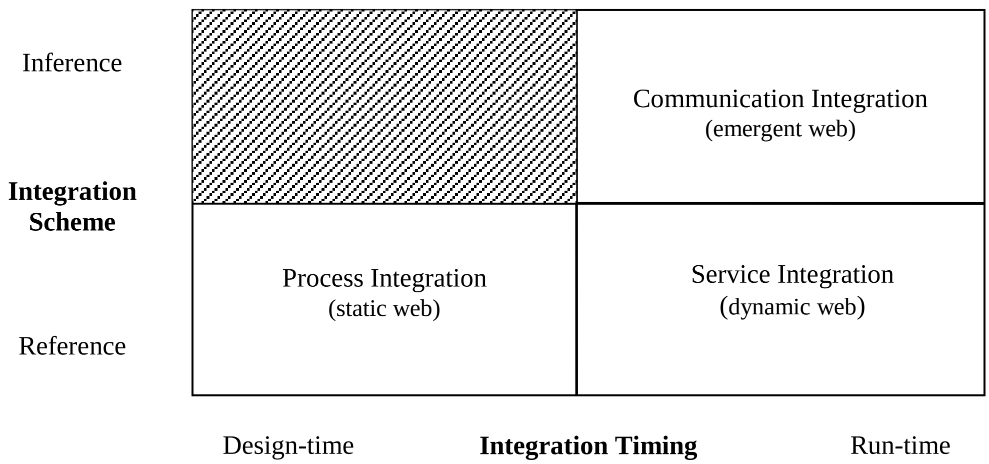


Figure 1 shows how the different types of business integration methods fit into a 2x2 matrix based on the dimensions of integration scheme and integration timing. Integration scheme refers to the ways of integrating different businesses processes while integration timing represents when different systems are being integrated. Reference model requires predefined process design from human intervention while inference model achieves autonomous communication with standardized communication protocols. Therefore, human intervention is no longer needed in the inference model. Traditional approach in systems integration has always occurred in design-time of systems implementation. Using the Web technologies where the direction of WWW is moving to dynamic web environment proposed service-oriented architecture. This movement has been fruited in the trend of Web 2.0 such as web services, Ajax, RSS, etc. At the same time, new efforts are motivated and initiated to evolve the current foundation of WWW. Instead of the use of hyperlink feature to the fullest, the evolutionary structure, such as the Semantic Web, was proposed to make the Web experience more intelligent. Therefore, communication integration can come into existence at run-time without human intervention.

The most recent movement toward communication integration, as mentioned above, is ***the semantic web***. Technically speaking, this shift began in the late 90's when eXtensible Markup Language (XML) gained popularity. As TCP/IP gave the freedom of speech to among hardware and networks, XML provides standard communication protocol among software. This idea is inherited to the semantic web, aiming to build a new intelligent world on the present World Wide Web by bringing new technologies such as Resource Description Framework (RDF) XML that serve to integrate communication at run-time [Berners-Lee, 1998, 2001][Shadbolt, 2006].

Compared to the service integration that requires manual adjustment to coordinate run-time services, communication integration implemented with the semantic web facilitates autonomous and intelligent connections between systems. We discuss this issue in the next section by examining the structure and information systems (IS) aspects of the semantic web and by comparing the semantic web with Web 2.0. This examination will help us develop a deeper understanding of how to prepare for the next wave of web science.

THE SEMANTIC WEB AND ITS STRUCTURE

The semantic web is referred to as "a web of actionable information" where various terms are logically connected to be interoperable between systems [Shadbolt et al., 2006]. It aims to provide services based on machine-understandable web resources so that business integration through machine internetworking and communication is facilitated. In the current state of the semantic web, there are various ways of specifying business rules developed by a variety of communities of interest [Hendler, 2001]. Here, the specification of business rules is defined as an ontology that includes taxonomy and a set of inference rules [Berners-Lee et al., 2001]. The variety of ontologies in the fields of artificial-intelligence and Web science prevents computers from effectively communicating with each other through the semantic web. Thus, tremendous business knowledge accumulated around the web in the form of data warehouses, XML schemas and documents, and other metadata repositories cannot be effectively utilized [Frankel et al., 2004]. To make the semantic web interoperable and render the accumulated knowledge usable, a common conceptualization of the requirements for the semantic web is needed. This conceptualization would serve as a tool for framing the requirements for successful application of the semantic web. In that sense, it would be similar in purpose to the entity-relationship model and the Unified Modeling Language, both of which are common tools in the IS area for representing business requirements. This article provides a theoretical framework based on the theory of deep structure to develop a tool for conceptualizing knowledge requirements for the semantic web based on the logical view of a business. A business in any industry would be able to use our conceptualization to translate their business knowledge into the semantic web, regardless of the physical ontologies that are adopted.

The Theory of Deep Structure and Information Systems Conceptualization

Web 2.0 is ill-suited for automated information processing due to the unavailability of semantics for machines to make inferences. Thus, building semantics on the Web brings the currently dormant web data and its relations to life. Proponents of the semantic web propose "the use of markup language to annotate data with semantic labels so that machines can identify content meaning and use rules for manipulating semantic information appropriately" [Flake et al., 2003]. According to Shadbolt et al. [2001], ontologies in the semantic web have deep and shallow structures. Often found in science and engineering, deep ontologies involve building and developing conceptual specifications to classify complex sets of properties of objects. Shallow ontologies, on the contrary, explain the relations between terms. Shallow ontologies consist of a relatively small number of unchanging terms that help organize a large amount of data. Examples include terms like customer, account number, etc.

However, Shadbolt et al.'s classification of deep and shallow ontologies only reflects physically implemented business objects which differ from one industry to another. Therefore, there is a big gap between business meaning and business objects as manifested in the heterogeneity of ontologies. What we need for business integration is a common conceptual framework that can be shared across business practices. To this end, we introduce the theory of deep structure from the field of linguistics to conceptualize the ontology of ontologies in the semantic web. This theory was pioneered by Noam Chomsky in linguistics, and was introduced in the information systems (IS) field by Wand and Weber [1995]. The theory of deep structure consists of three structures: deep, physical, and surface structures [Wand, 1996]. Table 1 shows the three structures.

Table 1: The Theory of Deep Structure

Structure	IS Aspects	Examples
Deep Structure	Business Objects	Values, Beliefs, Norms
Physical Structure	Implementation Technology	Technological protocols, Practices
Surface Structure	Business Meaning	Thoughts, Notions, Abstract ideas

Deep structure in the information systems context refers to business objects with rules on which information systems can be built and which govern individual behavior and interactions. It implicitly refers to values, beliefs, and norms that are important to an organization. Physical structure refers to the technical means employed for the implementation of business objects. It consists of the technological protocols embedded in information systems. These technological protocols reflect on social interactions, providing work practices to be used in actual work sites. Surface structure refers to business meaning of business objects used in information systems. It represents the thoughts, notions, or abstract ideas generalized and institutionalized via the interface by users from particular instances of interaction. The concept of deep structure can be applied to the semantic web to enable a better understanding and design of the semantic web.

The Semantic Web Structure

The World Wide Web (W3) Consortium defines the semantic web as follows:

> *The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF).*

The above W3 Consortium statement describes the basis of the semantic web as RDF. Semantics are built in XML and Resource Description Framework (RDF) using Uniform Resource Identifiers (URIs). URIs, XML, and RDF can set the semantics on the web and thus formulate ontologies, inferences, and logic of the semantic web, deep structure. There is less ambiguity in conceptualizing *deep structure* because of its clarity of practices in URIs, XML, and RDF.

Ontologies, inferences, and logics determine how to interact between the semantic webs in actual practice. As discussed in the previous section, ontology is the specification of a conceptualization, which defines terms and their relationships in a formal manner. In the case of the semantic web, ontologies are situated in different fields of web businesses so that their use is limited to the interested community. It is not possible to externalize all the knowledge in a specific domain to be shared through the specifications. In the current web, various technologies such as SOAP, RSS, Ajax along with URLs, HTML, and XML are developed and deployed.

In *physical structure*, many models are being proposed such as model-driven semantic web, semantic web services, and semantic blogging. Physical structure so far faces difficulties in the rationalization of actual data-sharing practice due to its practical infancy. The basic approach to physical structure is applying the concept of the semantic web to existing practices. Model driven semantic web, for example, is to enhance the capability of MDA® of the Object Management Group (OMG). Others use model driven business in the semantic web as an application of technologies to help organizations improve and transform their current practices. Adding semantics to the current web structure, people pioneer wider opportunities to enhance the use of current web environment. Semantics render intelligence and efficiency to the Web because the use of semantics leads to less traffic between nodes in the networks to provide meaningful information.

Table 2: The Web Structures

Structure	Examples in Web 2.0	Examples in Semantic Web
Deep Structure	- URLs - SOAP - HTML - RSS - XML - AJAX	- Ontologies - Inferences - Logic
Physical Structure	- Search engines - Blogging - Wikipedia	- Model driven semantic web - Semantic web services - Semantic blogging
Surface Structure	- Self-organiz***ed*** architecture - Collective intelligence - Syndication	- Self-organiz***ing*** architecture - Recursive intelligence - Syndication

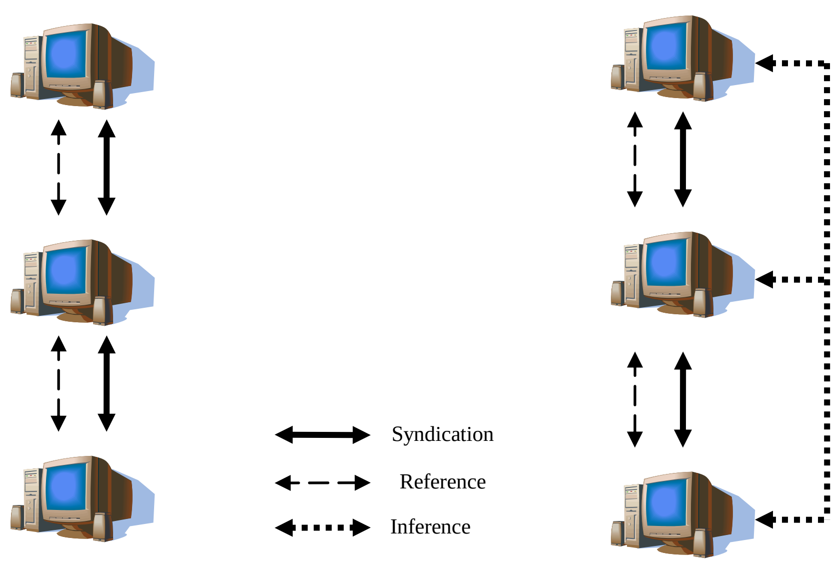
For deep and physical structure, it is relatively easy to picture what the semantic web is and what it does. The most ill-defined structure of the semantic web is *surface structure*. What kind of interface can be drawn from the semantic web as surface structure? In other words, what interfaces can we possibly develop in the human mind? Can different ontologies share the same interfaces? The answer is yes. The surface structure provides a common framework to help users build their own but shared world view of the semantic web. Table 2 summarizes different characteristics of the semantic web and Web 2.0.

With regard to surface structure, the semantic web basically involves the self-*organizing* web while Web 2.0 concerns the self-*organized* artifact [Flake et al, 2003]. Web 2.0 is regarded as a collection of hyperlinks directed to other users. On the contrary, the semantic web has a self-organizing architecture where each member is semantically related to other users. Thus, the semantic web envisions recursive intelligence so that it infers the meanings that other members intend, while Web 2.0 applies collective intelligence, referring to the meanings of others. However, regardless of inference or reference scheme, surface structures of both the conventional and the semantic web share a common feature, syndication [O'Reilly, 2005]. The current web community is geared towards syndicating data outwards, not controlling standardization with the other end of connection.

Web Business Intelligence

One of the anticipated benefits of the semantic web is that direct machine-to-machine communication can replace human end-user interaction of current web applications. While the current web requires users to connect to applications, the web agent architecture for the semantic web enables applications to connect to other applications. The semantic web, therefore, is a key technology in enabling business models to move from manually handled B2B to more intelligent B2B.

Figure 2: The Architecture of the Web



a) Web 2.0
Collective Self-Organized Architecture

b) Semantic Web
Recursive Self-Organizing Architecture

Internet technology has evolved from a primitive information exchange to a complex information communication and even can be extended to knowledge management. Traditional client/server architecture, the backbone of Internet technology, was mainly applied to exchange information through World Wide Web, establishing connections between clients and servers. Internet technology has been limited to documents based on hyper linkable relations, but it evolves into the self-*organized* web as millions of web publishers add, delete, move, and change their pages and links every day. With the introduction of the semantic web, process and service layers of systems can be integrated in the self-organizing structure. The self-*organizing* structure means there is no need for hyperlink distribution property. Rather, it focuses on communication instantiation through the syndicating recursive intelligence. The key to the recursive intelligence in the self-organizing architecture is its inference engine, while the collective intelligence in the self-organized architecture comes from the reference to outside sources.

As a result, the semantic web establishes a logical roadmap between trust-worthy reputations in web sites with semantics instead of direct recommendations from physical connections like hyperlinks in Web 2.0. Thus, the Internet is viewed as meta-networking in the semantic web while it is inter-networking in Web 2.0. As shown in Figure 2, the collective self-organized architecture consists of reference connections between client and server and that is the only way of communication. This causes congestible web traffic, burdening the maintenance of efficient networks. However, recursive self-organizing architecture adds inference connections, releasing the congestion of the Internet traffics as well as transforming the Web into an intelligent medium. Recursive self-organizing architecture is a new design paradigm that excludes human intervention. Instead, a machine can establish intelligent connections to the recursive self-organizing network. Thus, the intelligent connections append the collective inferences of all networks' processes, data, information, and knowledge on the semantic webs through communication instantiation. Communication instantiations gradually establish intelligent and efficient recursive self-organizing networks.

Advancement in Web technology gives the cutting edge to business practices. The Semantic Web architecture promises more intelligent and efficient web structure. In a short period of time, organizations have experienced the scheme of integration, starting from process integration and now to service wave. What the semantic web is proposing is not only the current service movement but also the future wave of communication integration where search and query information on the web is intelligent enough for organizations not worrying about time-consuming predefined business integration. It has the vision, but is also bit too early for business organizations to accept conceptual reflexibility from the current practice, service integration. Table 3 shows the road map between logical views of the Internet architecture, IS aspects of integration type, and practices of business integration.

Table 3. Road Map to the Business Intelligence

Web Architecture	Web Structure	Business Integration
References	Static Web	Process Integration
References Syndication	Dynamic Web	Service Integration
References Syndication Inference	Emergent Web	Communication Integration

TOWARD INTELLIGENT BUSINESS INTEGRATION

As the use of the web evolves, developing web algorithms intelligent enough to infer semantics from Web 2.0, without implementing the semantic web, may increase the complexity of web business, making autonomous and intelligent business integration almost impossible. This is primarily because considerable human interventions are required at design-time and thus it is difficult to fix problems after the web algorithms are implemented. The semantic web allows a flexible exchange of unambiguous content representations across heterogeneous platforms by explicitly formalizing knowledge rather than embedding it in program codes [Eberhart, 2003]. As we conceptualized in this article, the IS view point of the recursive self-organizing architecture is not merely for the world of the semantic web, but it also envisions the directions that we are heading for communication integration through communication instantiation. This new vision will provide new business opportunities by integrating knowledge, formalized on the front-end arsenals, and rendering autonomous and intelligent business integration. Thus, the focus of systems designers is shifted from standardizing how to connect pre-designed systems with centralized control to institutionalizing how systems interact in an ad hoc manner without centralized control.

An intelligent system (AI) is appealing because a lot of business processes can be automated through it. Intelligent data (the semantic web) is even more fascinating because a more complete sense of business integration can be achieved through it. The emergent web promises business integration in the next level without the painful process of standardization. It will allow intelligent communications between web businesses regardless of different systems, legacy applications and various means of business processes [Sayah, 2004].

REFERENCES

1. Berners-Lee, T. (1998) "Semantic web road map", http://www.w3.org/DesignIssues/Semantic.html (current July. 15, 2008).
2. Berners-Lee, T., J. Hendler, and O. Lassila. (2001) "The Semantic Web", Scientific America.com, http://www.ryerson.ca/~dgrimsha/courses/cps720_02/resources/Scientific%20American%20The%20Semantic%20Web.htm (current July. 15, 2008).
3. Eberhart, A. (2003) "Towards semantically enriched business logic", *Electronic Commerce Research and Applications*, (2), pp. 288-301.
4. Kajan, E., and L. Stoimenov. (2005) "Toward an ontology-driven architectural framework for B2B", *Communications of the ACM*, (48)12, pp. 60-66.
5. Flake, G. W., D. M. Pennock, and D. C. Fain. (2003) "The self-organized web: the Yin to the semantic web's Yang", *IEEE Intelligent Systems*, (July/Aug), pp. 72-86.
6. Frankel, D., P. Hayes, E. Kendall, and D. McGuinness. (2004) "The model driven semantic web", 1st International Workshop on the Model-Driven Semantic Web (MDSW 2004), http://www.sandsoft.com/edoc2004/FHKM-MDSWOverview.pdf#search=%22semantic%20web%20paper%22 (current July. 15, 2008).
7. Hendler, J. (2001) "Agents and the semantic web" *IEEE Intelligent Systems Journal*, (March/April), http://ieeexplore.ieee.org/iel5/9670/19905/00920597.pdf?tp=&isnumber=&arnumber=920597 (current July. 15, 2008).
8. O'Reilly, T. (2005) "What is web 2.0: design patterns and business models for the next generation of software", http://www.oreillynet.com/lpt/a/6228 (current July. 15, 2008).
9. .Sayah, J. Y., and L. Zhang. (2004) "On-demand business collaboration enablement with web services", *Decision Support Systems*, (40), pp. 107-127.
10. Shadbolt, N., W. Hall, and T. Berners-Lee. (2006) "The semantic web revisited", *IEEE Intelligent Systems Journal*, (May/June), http://eprints.ecs.soton.ac.uk/12614/1/Semantic_Web_Revisted.pdf (current July. 15, 2008).
11. Wand, Y. (1996) "Ontology as a foundation for meta-modelling and method engineering", *Information and Software Technology*, (38), pp. 281-287.
12. Wand, Y. and R. Weber. (1995) "On the deep structure of information systems", *Information Systems Journal*, (5)3, pp. 203-223.