Web Science: Conceptualization of the Semantic Web

Jinyoul Lee, State University of New York at Binghamton, Binghamton, NY 13902, USA; E-mail: jylee@binghamton.edu
Surinder Kahai, State University of New York at Binghamton, Binghamton, NY 13902, USA; E-mail: kahai@binghamton.edu
Yong-Jin Kim, State University of New York at Binghamton, Binghamton, NY 13902, USA; E-mail: ykim@binghamton.edu
Bonn-oh Kim, Seattle University, USA; E-mail: bkim@seattleu.edu
Paitin Katerattanakul, Western Michigan University, Kalamazoo, MI 49008, USA; E-mail: pairin@wmich.edu

INTRODUCTION

As the use of Web technologies has increased tremendously, the web business is getting attentions from business organizations. Originated from artificial intelligence (AI) tradition, the syllogism based Semantic Web has a great deal of hype. To introduce the semantic web, this paper is the initial effort to conceptualize the promises of the semantic web in the context of information and communication technologies. Especially, this paper emphasizes the science aspect of the semantic web, introducing theoretical approach on its meaning.

A promising and ambitious statement can be found from the inventor of Web technology and the director of W3 Consortium, Tim Berners-Lee (1998), that "One of the major obstacles to this has been the fact that most information on the Web is designed for human consumption…the Semantic Web approach instead develops languages for expressing information in a machine processable form." As we all can perceive, data on the Web is not in a processable form for any application. Thus, what the semantic web seeks is data in a reprocessable form from the the web. XML and web service have been in the same track of this idea. Even combined with XML and web services, the semantic web intends to serve more intelligently who use the web technologies.

Through the theoretical conceptualization in this paper, the merits of the semantic web are revisited in terms of web science.

THE SEMANTIC WEB

The semantic web is "a web of actionable information – information derived from data through a semantic theory for interpreting the symbols" (Shadbolt et al., 2006). The semantic theory explains "meaning" to which the logical connection of terms establishes interoperability between systems. The aim of the semantic web is to provide services based on the machine-understandable web resources so that the business integration through machine internetworking and communication is facilitated. To effectively communicate through the semantic web, there needs to be a common conceptualization, called *ontologies*. Ontology is defined as the science or study of being. In the artificial-intelligence and Web science, however, it refers to the specification of a conceptualization, which defines terms and their relationships in a formal manner (Hendler, 2001). That is, ontology in the web science is a document or file that formally defines the relations among terms including taxonomy and a set of inference rules (Berners-Lee et al., 2001).

Ontologies, the basis for the semantic web, consist of various forms of knowledge such as entity-relationship (ER) models, unifying modeling languages (UML), data warehouses, XML schemas and documents, and other metadata repositories (Frankel et al., 2004). ER diagram and UML, based on the theory of sets and relations, advocate standardization on conceptualization. On the contrary, there is no single ontology for the semantic web. The semantic web rather diversifies ontologies because of its flexibility in semantic expression.

IS VIEW OF THE SEMANTIC WEB

According to the proponents of the semantic web, the conventional web is ill suited for automated information processing due to the unavailabilty of semantics for machines to infer. Building semantics on the Web brings the meaning to the web data and its relations. Proponents of semantic web, therefore, 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). To do this, the semantic web uses ontologies. Though there is no standard ontology elected, ontologies in the semantic web have deep and shallow structure (Shadbolt et al., 2001). Often discovered in science and engineering, deep ontologies involve building and developing conceptual specifications toclassify complex sets of properties of objects. Shallow ontologies, on the contrary, explain the basic relations in terms of geospatial information. Shallow ontologies consist of a relatively small number of unchanging terms that help organize a large amount of data. The examples include the terms like customer, account number, etc.

However, these two views of ontologies are not enough to understand all forms web business. This classification of the semantic web does not provide conceptual framework which can help to build the semantic web. Therefore, we introduce the theory of deep structure from the field of linguistics because it can append the meaning of the semantic web in information and communication aspects. This theory was pioneered by Noam Chomsky in linguistics, and was introduced in IS by Wand and Weber (1995). The theory of deep structure consists of three structures: deep, surface, and physical structures (Truex and Baskerville, 1998). Table 1 summaries the relationships between the terms of structure, definition, and meanings. In addition to this theory, we interpret three structures into core, concept, and framework (Table 1).

A core is a central and often foundational part usually distinct from the enveloping part by a difference in nature. Deep structure defines the intention of information systems to be modeled by real world systems. It implicitly means value, beliefs, and norms that are important to organization and IS. Thus, deep structure has rules that govern individual behavior and interactions. We will use the term 'core' in

Table 1. Conceptualizations of IS: Core, concept, and framework

Conceptualization	Structure	Definition	Meaning
Core	Deep Structure	Intention	Rules, Values, Beliefs, Norms
Concept	Surface Structure	Institution	Mind, Thought, Notion, Abstract
Framework	Physical Structure	Interaction	Categories, Terms, Practices

this study with the same meaning of deep structure to view IS. Second, a concept is something conceived in the mind, thought, notion, or an abstract or generic idea generalized from particular instances. Instantiation occurs from interactions between actors. In the context of IS, actors interact with IS. Thus, institutions built by actors through interactions can be formalized to concepts. The second structure, surface structure, explains this nature of concept. Surface structure can be explained as an interface between the IS and its users' organizational environments. Third, a framework is a basic conceptual structure (as of ideas) and a skeletal, openwork, or structural frame. The third structure, physical structure, belongs here. It refers to the technologies used to implement information systems. Thus, it determines the protocols that apply to the perceptions of social interaction as collectives by observers. Information system provides such categories and terms to be used in actual work places.

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 core of the semantic web is well established as appeared in the W3 Consortium statement. This is a definition of semantic web in which semantics is represented in XML and Resource Description Framework (RDF) using URIs. Thus, URIs, XML, and RDF represent the deep structure of semantic web. There is less ambiguity in this structure because it is clearly defined in general terms. Framework, however, raises difficulties in a rationalization of actual data-sharing practice. It determines the physical structure of semantic web, interaction between the semantic webs in actual practice. In other words, ontologies, inferences, and logics of semantic web explain what physical structure of semantic web is about. Until now, the use of ontolgies is limited to the interested community meaning that there will be multiple ontologies of different fields of web businesses. The hardest conceptualization of semantic web is surface structure. What concepts can be drawn from semantic web? For core and framework in Table 2, it is relatively easy to picture what semantic web is and what it does. However, surface structure directly points institutions of semantic web. What institutions can we possible expect? Can different ontologies share same institutions? As we listed in Table 2, partial understanding explains the evolution of semantic web from large scale to medium scale systems while transformability applies to the situation from small scale to medium scale. Intertwingling basically pinpoints the concept of "self-organized web" (Flake et al, 2003). However, to develop the governing concept of the semantic web that describes different ontologeis in a big picture, we will envision the surface structure of semantic web more rigorously in the following section.

RECURSIVE SELF-ORGANIZING ARCHITECTURE

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

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 the Internet technology, was mainly applied to exchange information through World Wide Web, connecting clients and servers. The Internet technology, however, has been limited to documents to documents based on hyper linkable relations. With the introduction of the semantic web, the processes of applications are combined in the self-organizing nature. The conventional web is based on the Internet as 'inter-networking'. The semantic web, however, view the Internet as 'meta-networking', where it is named as the recursive self-organizing architecture while the conventional web uses client/server architecture. The conventional client/server architecture consists of direct connections between client and server. However, the semantic web is emphasizing the logical structure using semantics, **recursive self-organizing architecture**. Recursive self-organizing architecture is a new design paradigm that explains there is no need for a direct relation between machines. Instead, a machine may establish a connection to the recursive self-organizing network. Thus, the recursive self-organizing architecture appends the collective inferences of all networks' processes, data, information, and knowledge on the semantic webs. In other words, institutions are structured in semantic web and can be hosted, as established on recursive self-organizing network. It is also recursive because of its inference engine.

As discussed earlier, the relation structure has a different meaning in the semantic web compared to the conventional web. In the conventional web, hyper links implicitly infer TCP/IP protocol that is shown to the users as logical addresses for the connection between client and server. The relation between the links is static as the link between web pages embedded in html files. In the semantic web, the relation between web pages is not explained by hyper links; rather it is abstracted as "instantiation" between machines. Not like hyper links, instantiation infer freedom of connection to any other ontologies.

Table 3 summarizes the design paradigms of the conventional web and the semantic web.

Table 2. Conceptualizations of the Semantic Web

Conceptualization	Structure	Definition	Examples
Core	Deep Structure	Intention	• URIs • XML • RDF
Concept	Surface Structure	Institution	• Partial understanding • Transformability • Intertwingling
Framework	Physical Structure	Interaction	• Ontologies • Inferences • Logic

Table 3. Design paradigms of the Web

	The Conventional Web	The Semantic Web
Logics	Inter-Networking	Meta-Networking
Relation Building Structure	Client/Server Architecture	Recursive Self-Organizing Archtecture
Information Orientation	Server-Centirc	Server-Decentric
Communication Mode	Static	Instant

Figure 1. Dynamics in enterprise integration

Structure View	Process Integration	Service Integration	Communication Integration
Relation View	Static	Dynamic	Instant

IMPLICATIONS

Back in the 1990's companies like SAP AG, Oracle, Baan, PeopleSoft and J. D. Edwards created a multi-billion dollar business with ERP technology that automated and connected what had once been disparate parts of corporations – human resources, manufacturing processes, inventory supply and financial planning. These companies rode the wave of the corporate BPR (business process re-engineering) trend that gained steam in the middle of the decade. Along the way, the ERP industry began to get saturated and view its growth struggle due to its focus on internalization. The main reasons for this were the Internet revolution and the surprising speed with which e-business began to change the way business was done, i.e., externalization of enterprise. Almost immediately, businesses have started to become Web centric.

In technical tradition, three solutions are accepted in current externalization efforts. First, process integration solutions have become commonplace in today's market, but the current vendor emphasis is on proprietary offerings, not standards. Process integration is critically important to the automation of both internal and enterprise systems, so both vendors and users will come under increasing pressure to standardize process models and protocols as relational database model became a de facto. The second important integration model is the "Web Services" framework model, promising a dynamic approach to application integration. The idea behind the autonomous distributed integration pattern is that integration can occur dynamically by combining Internet-accessible services at run-time according to a predefined pattern. Finally, an internet-based, hosted integration service is emerging as a cost-effective way for mid-market companies to participate in integration. These services join trading partners by employing integration hubs with transformation, routing and message management services. Lease/rental and per-transaction payment models provide an alternative to software purchase and enable smaller players to participate in value chains with much larger partners.

Although the integration market has been characterized by many small innovative start-ups, the essential e-business integration problem is an enterprise-level problem. Companies that started out adopting EAI technology to connect ERP systems with a few other systems, or synchronizing databases or a data warehouse, are quickly moving toward e-business solutions. E-Business infrastructures are complex, big-ticket items. Large organizations are more comfortable partnering with large vendors that can provide full solutions. IBM is considered a leader in this space, even though the components of its solution are not yet fully integrated. Other vendors making end-to-end integration include BEA, NEON (New Era of Networks), STC, TIBCO and Mercator.

To implement e-business technically means to determine how to implement in the three areas as discussed previously, integrations of processes, applications, and communications. To move this topic to the organization level, components of organizations must by re-visited and re-structured in the lights of modern IT-enabled organizations. Organizations have been collection of people with same goals and their relationships. With enterprise externalization of modern IT-enabled organizations equipped with advanced information and communications technologies, it adds one more values on integrations in processes, applications, and communications. E-business, in many cases, not just extends existing business practices but also include new perspective of how to integrate different levels of processes, services, and communications. Thus, the use of e-business imposes different roles in relations.

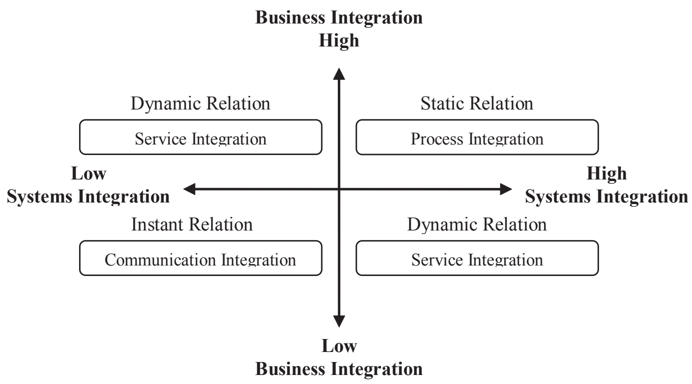
Now, it is possible that all organization contents can be stored, manipulated, and exchanged in any form of information systems. Examples are any relational database systems, ERP systems, and more recently web stores. Any part of organizational contents including business processes can be separated and digitized into information systems. Even, these systems can be related with other systems. These relationships, however, had been established only among humans. Thus, integrated processes are still static. Due to advance networking and the Internet, businesses information systems can be connected through communications networks. But, wait. What happened to the human members of organizations? In traditional settings, organizations used to be all about members and their relations. This at present is very tough to define due to the rapid changing environment but it alternately puts into two directions human roles in IT organizations. The first direction is rather negative effects on the importance of human roles. The traditional roles of humans are digitized into the IT systems. Thus, valuable assets, relations, which have an ontological meaning of members in their organizations are drastically diminished. Dramatic effects were many incidents of lay-offs in the late 1990s and early 2000s. This changes can reflected in dynamic relation in service integration structure. The second direction is now being gradually discovered in the field, establishing the new types of relations to systems instead of the ones using humans. Traditional strong cohesive bondage with other members is now shifting to information systems because organizational contents are abstracted into the systems and those can be communicated with others. This new trend builds new type of communication, communication instantiation (machine-to-machine instantiations).

Business integration in terms of processes, applications, and communications is touted to add considerable value to companies (Figure 1). In particular, recent development in e-business allows companies to enjoy the benefits of integration both by extending existing business practices and by adopting new perspective to the roles of business partners. The basic premise of the business integration through information technology is that any part of business including business processes and content can be digitized and embedded into information systems.

Figure 2 shows a graphical illustration of two different dimensions; business integration and systems integration. Business integration occurs when two parties coordinate their businesses while systems integration means the level of configuration to adjust systems. It also shows the nature of relation between systems; static, dynamic, and instant. Technically speaking, this shift has already begun from late 90's when eXtensible Markup Language (XML) gained popularity.

There are organizational impacts from business integration. Because the traditional human roles are digitized into the information systems, process integration, such

Figure 2. Business integration topology



as ERP or EAI, has been reflected in many incidents of lay-offs in the late 1990s and early 2000s. As a result, new relations built in process integration replace the traditional strong cohesive bondage between humans with the role-based modular integration. Toward service and communication integration, companies have hard time to find business implications till now. Service integration is promising but it has a drawback that it needs manual adjustment to coordinate run-time services. The proponents of the semantic web claim that communication integration can solve the problem of service integration.

CONCLUSION

The use of the Web is evolving, even it is self-evolved. Without semantic web, it is also possible to write a scenario that envisions Web algorithms which are intelligent enough to infer semantics from the current Web structure. However, it would also highly increase the complexity of Web business, making business integration heterogeneous. As shown in the enterprise externalization, communication instantiation movement shares the same idea, machine-to-machine instantiation, by the recursive self-organizing architecture. In other words, semantic web has strong business implications. For future study, the semantic web conceptual model we developed here can be mapped to the physical model, discussing recursive self-organizing architecture.

Web science is described as "a science that seeks to develop, deploy, and understand distributed information systems, systems of humans and machines, operating on a global scale" (Shadbolt et al., 2006). This paper discusses the conceptualization, dynamics, and topology of the semantic web in web business. This is an initial step to analyze the essence of practical technologies using the lens of science in business context.

REFERENCES

Berners-Lee, T. (1998). Semantic web road map. http://www.w3.org/DesignIssues/Semantic.html.

Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web. Scientific America.com, May 17.

Flake, G. W., Pennock, D. M., and Fain, D. C. (2003). The self-organized web: the Yin to the semantic web's Yang. IEEE Intelligent Systems, July/Aug, pp. 72-86.

Frankel, D., Hayes, P., Kendall, E., and McGuinness, D. (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

Glaser, B. G. and Strauss, A. L. The discovery of grounded theory: strategies for qualitative research, New York: Aldine Publishing Company, 1967.

Hendler, J. (2001). Agents and the semantic web. IEEE Intelligent Systems Journal, March/April.

Shadbolt, N., Hall, W., and Berners-Lee, T. (2006). The semantic web revisited. IEEE Intelligent Systems Journal, May/June.

Truex, D. P., and Baskerville, R. (1998). "Deep structure or emergence theory: contrasting theoretical foundations for information systems development", Information Systems Journal, 8, pp. 99-118.

Wand, Y. and Weber, R. (1995). "On the deep structure of information systems", Information Systems Journal, 5, 3, pp. 203-223.

Related Content