

# Data Engineering Issues in E-Commerce and Services: DEECS 2007 Workshop Summary

Martin Hepp University of Innsbruck Innsbruck, Austria mhepp@computer.org	Mehmet Sayal Hewlett-Packard Labs Palo Alto, CA, USA mehmet.sayal@hp.com	Sang-goo Lee Seoul National University Seoul, Korea sglee@snu.ac.kr	Juhnyoung Lee IBM T. J. Watson Research Center Hawthorne, NY, USA jyl@us.ibm.com	Junho Shim Sookmyung Women's University Seoul, Korea jshim@sookmyung.ac.kr
--	---	---	--	--

## ABSTRACT

In this paper, we summarize the presentations and discussions of the 3rd International Workshop on Data Engineering Issues in E-Commerce and Services (DEECS 2007), which was held in conjunction with the 8th ACM Conference on Electronic Commerce (EC'07) on June 12, 2007, in San Diego, California.

## 1. Data Engineering on a WWW Scale

As part of their e-business activities, enterprises create, share, recombine, and process an ever increasing amount of data and documents that range from simple transactional data to complex business process definitions. In many respective processes, multiple spheres overlap, e.g. (1) inter- with intra-organizational data sources in collaborative processes, (2) master data with transactional data, (3) regular data with behavioral aspects of processes, or (4) facts with normative assertions. Since business entities are members of multiple value chains, same as data sources are often used in multiple contexts and processes, there are often conflicting requirements on the representation of such data.

All this in combination makes e-business data engineering a very challenging task, already at the level of a single company but even more in value chains. Also, respective modeling choices may have long-lasting and far reaching impact, and affect (1) operational efficiency, (2) business process agility, (3) the range of analytical tasks for decision support, and (4) strategic options. Eventually, those choices determine the current and future degree of automation in content and process integration.

While e-business standards in general help, standardization alone does not solve the problem. This is because producing a consensual representation takes time, consumes resources, and constrains an entity's ability to capture individual details. Moreover, the cost of implementing and enforcing the standards is sometimes prohibitive for small businesses. Proprietary representations, on the other hand, hamper interoperability, and complicate B2B integration. In service-oriented architectures (SOAs), this complexity is further increased by the behavioral dimension of e-business interactions, e.g. services choreography and orchestration.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DEECS 2007 Workshop, June 13, 2007, San Diego, California, USA.  
Copyright 2007 ACM 1-58113-000-0/00/0004...\$5.00.

The DEECS workshop series aims at providing a venue for discussion and the exchange of ideas on *data and knowledge engineering* issues in the dynamic environment of e-business, enterprise computing, and business services and transformation.

## 2. Current Relevance

Within the past five years, the ontology research community in computer science has brought to maturity a comprehensive set of foundational technology components, and this both at the conceptual level and in the form of prototypes and software. This includes, among other assets, ontology engineering methodologies, standardized ontology languages, ontology engineering tools, and other infrastructure like APIs, repositories, and scalable reasoners, plus a plethora of work for making the Deep Web and computational functionality in the form of Web Services accessible at a semantic level. However, the amount of visible applications and convincing showcases is still limited, in particular in e-business domain.

In particular, there exist many pressing challenges in managing e-business data that would obviously benefit from the higher level of abstraction and the increase in automation that semantic technologies can offer. The recent call for "Services Science" as a much needed new interdisciplinary field of research [1] is one of the most outstanding examples.

Unfortunately, the dissemination of the state of the art in SW technology into the respective communities is still to be improved, and the complexity of real-world data engineering issues is often not considered in the Semantic Web community.

In parallel to that development, data and knowledge engineering in the business world has become more pressing than ever, because

1. most business processes are now extremely collaborative and distributed,
2. the division of labor and thus the specificity of parts and materials has increased,
3. the geographical distribution of partners has grown, and
4. there is very little room for ambiguities and deviation, e.g. with regard to product characteristics or logistical aspects.

But data engineering for e-business is more complex than just solving the representation aspect, as it combines challenges from multiples spheres: First, the economic dimension, in particular incentives for individual actors to contribute and to comply, and the plethora of problems in price-finding between economic actors. Second, managing risk and uncertainty, i.e., how to establish e-business procedures that execute automatically despite incomplete information. Third, theoretical and practical challenges of acquiring necessary information, e.g. by human

experts, by dedicated agents that harvest Web data (crawlers), or by machine-learning approaches over existing data. And finally, we face the enormous challenge of combining results into practically useful applications.

### 3. Submissions and Review Process

Following a Call for Papers disseminated widely over popular mailing lists and other channels, we received a broad spectrum of papers for review. In total, we received ten submissions, of which five full papers and one position paper were eventually accepted for presentation and publication in the post-proceedings. All papers were reviewed by at least two members of the Program Committee, most even by three members. In the end, we were able to compose a workshop program that addresses data engineering challenges from the various branches of the problem domain.

### 4. Summary of the Workshop

The workshop was organized as a full-day event in conjunction with the ACM Conference on Electronic Commerce (EC'07). Besides the presentations and discussions of the accepted papers, there was a keynote speech and a concluding panel discussion on the state of the art in data engineering for e-business. In the following, we briefly summarize the presentations.

#### 4.1 Keynote Speech

Dr. Juhnyoung Lee (IBM Research) gave an interesting keynote speech about the Service Science, Management and Engineering (SSME) initiative that is currently led by IBM. SSME is a new multi-disciplinary effort that aims at integrating several established research areas such as computer science, operations research, engineering, management sciences, business strategy, social and cognitive sciences, and legal sciences. Dr. Lee pointed out that global markets are shifting from agriculture and manufacturing to service-based economies. He also pointed out that the U.S. Bureau of Labor Statistics employment projections forecast that employment growth will continue to be concentrated in the service-providing sector of the economy. With major industrialized nations having now more than 75% of their GDPs related to the services sector, and developing nations closely behind, governments and industry need talented, skilled staff in the services field. Dr. Lee indicated that SSME has gained the attention of several universities around the world, and has been offered by several institutions as certificate programs already. The goal of the SSME discipline is to make productivity, quality, sustainability, learning rates and innovation rates more predictable across the service sector.

This has been a highly interactive speech. One major concern in the audience was the reason why SSME has to emerge as a new discipline, rather than being supported by individuals from different disciplines. One may reply that computer science has emerged from applied mathematics in a similar way. As the computational advantages of computers increased over time, computer science started to be recognized as a new discipline, rather than being a sub-topic in the mathematics department. Dr. Lee also mentioned that there exists an important gap between the business and IT fields of research. Important decisions are made at the business level but it is difficult for engineers to provide useful technologies at the higher business levels unless they can support business data analysis using software tools that are

familiar for business managers. This is just one example where professionals who understand multiple disciplines can contribute significantly. Businesses that provide services to their customers can benefit from individuals who have been educated under the SSME discipline.

Dr. Lee mentioned that the industry seems to be ahead of academia and government organizations as service providers. Therefore, an initiative like SSME needs strong support from industry, especially large corporations, since those are the ones who most benefit from the service industry. Also, it was pointed out that partnerships among industry, academia, and government will be critical to fuel the progress of the service-based economy of the future.

#### 4.2 *BestChoice* SRM: A Simple and Practical Supplier Relationship Management System for e-Procurement

Dongjoo Lee presented a practical SRM system called *BestChoice SRM*. It enhances a conventional SRM with the simple adaptability of *BestChoice* which is a decision support system for evaluating suppliers with the aid of visually-supported analysis. This will allow seamless integration with existing systems and permits flexible strategy changes as market conditions shift. The key to this flexibility is template-based segmentation. It was mentioned that popular software tools such as SAS, SAP, and i2 suffer from an unsatisfying integration with existing processes and systems. The presented approach measures the strategic importance of sourcing groups, calculates the utility of suppliers, and provides a mechanism to produce numeric values for decision making by using models and methods such as MAUT (Multi-Attribute Utility Theory) and AHP (Analytical Hierarchy Process).

#### 4.3 Diversification of Risk based on Divided Tasks in Large-scale Software System Manufacture

Dr. Tokuro Matsuo presented a method that provides resource allocation by auctioning available resources. The first part of the presentation explained that certain items are dividable into smaller pieces, and some other items are indivisible. Another important notion is the reversibility of the division, in other words, the capability to re-compose divided pieces. Dr. Matsuo indicated that software can be divided into smaller pieces which can be integrated to form a complete software system. Namely, software is divisible and this operation is also reversible. Dr. Matsuo presented a method for auctioning modules of a software product by sellers (or developers). He argued that the overall risk can be reduced by dividing a software development effort across multiple developers using the auction system. In case a developer goes bankrupt, the majority of the software can then still be built by the remaining developers. Therefore, the time it takes to re-develop the portion of the bankrupted developer may be much smaller than an alternative case in which the whole software is developed by a single developer. However, of course the additional overhead of resources for dividing and recomposing a software project must be considered.

#### **4.4 DeepBot: A Focused Crawler for Accessing Hidden Web Content**

Manuel Álvarez presented *DeepBot*, a focused crawling engine that can extract data from the “hidden Web”. The hidden Web is the part of the Web content that contains valuable information which is “hidden” behind the query forms of online databases, and/or is dynamically generated by technologies such as Javascript. The proposed tool can model and parse Web forms by measuring the visual distances and angles between text components. This is an interesting and practical approach especially because it does not assume any particular structure or format on the Web forms. It simply learns the labels and their corresponding form components through an automated analysis. The crawled documents are assigned to domains using confidence and specificity values. The specificity value is calculated manually.

#### **4.5 Short-Time Approximation on Combinatorial Auctions – A Comparison on Approximated Winner Determination Algorithms**

Dr. Naoki Fukuta presented research work that compares four approximation algorithms for combinatorial auctions, and investigates the parameter settings under which each of these algorithms performs better. A combinatorial auction is an auction in which bidders can place bids on combinations of items, or “packages”, rather than just individual items. Combinatorial auctions are a popular market mechanism that has an important

effect on electronic markets and policy making. Due to market limitations, it is usually not possible to calculate the optimal winners of a combinatorial auction. Therefore, several approximation strategies have been proposed previously. The comparisons presented by Dr. Fukuta are important for understanding the circumstances under which each approach performs better. The comparison especially focused on cases in which there are many bids, but a limited amount of time available to calculate the approximation of the optimal bid.

#### **5. Conclusion and Outlook**

In the course of the workshop, we have seen an interesting set of research results addressing a variety of problems of data engineering in Web-centric e-business. The contributions range from theoretical work to prototype implementations and even software close to commercial usage. All this work can be seen as building blocks for the future infrastructure for Web-based, services-centric commerce.

The organizers would like to thank all members of the program committee for the valuable reviews, the ACM EC’07 organizers for their great support, and all contributors for their papers and for the lively discussion.

We are looking forward to the next DEECS workshop.

#### **6. References**

- [1] H. Chesbrough and J. Spohrer, "A research manifesto for services science," *Communications of the ACM*, vol. 49, pp. 35-40, 2006.