Editor: Charles Petrie • petrie@stanford.edu

Industrial Semantics and Magic

Charles Petrie • *Stanford University*
Christoph Bussler • *Cisco Systems*

It's time for academic research in semantics to be industrially relevant. This isn't to say that all of the problems of data integration, much less distributed ontologies, are solved — lots of room remains for theses. It is, however, time to put up or shut up (apologies to Bill O'Reilly and Fox News).

We academics claim to know enough to be useful in the young world of the commercial Web and Web services. We see great visions of a world in which Internet computing can do a much better job mediating connections among people. What's stopping us?

Yes, we can say that industry adapts slowly to new ideas, with large companies following start-ups' new technologies. Industry can easily dismiss the idea of really using semantics as blue-sky thinking, but can it really be that industry folk just don't understand the advanced technology and so think it seems like magic? No, it's we academics who are confused about the problem we're solving, as well as the problems that industry would like solved. Some of this confusion concerns the difference between static semantics and process descriptions.

Two Academic Communities Driving Each Other Nuts

Let's start with the confusion among different academic communities working on Web service composition using semantics. At least two, and possibly several, such technical communities exist, each working on different problems and definitions.

AI Service Planning

One distinct community comprises artificial-intelligence types (like Petrie), who work with deductive synthesis technology — logic-based planning, to you.

Arguably, it was Sheila McIlraith's AI planning paper that launched the term "Web service composition,"[1] although she was referring to composing the elements of a *plan*, the most important of which were Web services. In the AI context, a plan is some partial order of actions that changes the world from some initial state to some final goal state. If the plan is successfully executed, the goal should be achieved. (For simplicity and space, we omit discussion of "should.")

A Web service, in this same context, isn't just a Remote Procedure Call, but one with a relatively static description in a standard language. This description should be reachable via a standard Internet protocol and provide at least enough information for using the specified service.[2]

Using this rough definition, we can view Web services as plan operations: they can change the state of the world, and a sufficient description says what must be true prior to using the service as well as what effect the service might have on the world.

This view of Web services makes sense, whereas considering them as just another standard for exchanging messages doesn't warrant a new technical name. Those who say a Web service is a remote procedure that's invokable by a SOAP message are mostly describing SOAP clients. This view — shared by the academic "process" research community, which is closer to industry — drives us AI folks nuts. What's going on?

The Process Problem

The process community typically faces the common business-integration problem: start with multiple processes and integrate them. Processes are defined by message protocols and, following this view, Web services just happen to produce the messages and offer some standard method of describing them, thus facilitating analysis. The process-integration problem is to find a way to integrate these protocols into a super-process (sometimes called a *composite Web service*) that is "correct," meaning the process is deadlock-free and has no unreachable paths. Such process synthesis is very difficult, and most technologies in this context focus on tools and software engineering technologies to semiautomate it.

This also drives AI folks nuts. AI planning can automatically generate processes from collections of Web services, and the resulting processes will be correct — the deductive synthesis itself proves that because it means that the goal state is logically entailed by the process synthesized. Just list the initial state, desired goal state, and the collection of Web services in the various processes. Then say, "abracadabra," and the AI planning algorithm of choice can produce a correct process if doing so is possible. (We omit discussion of what determines whether it's possible.)

Yet, such talk drives the process

continued on p. 94

 Published by the IEEE Computer Society 1089-7801/06/$20.00 © 2006 IEEE

continued from p. 96

community nuts. First, they have no goal state, per se, but rather seek to integrate processes that are ill-defined in terms of final effects. Moreover, the problems are often defined — sometimes contractually and usually because of real business implementations — in terms of integrating the protocols in question. Finally, these protocols sometimes require the exchange of messages that aren't easily expressed as Web services or world states.

One such message exchange occurs in the simplest problem in our Semantic Web Services Challenge Workshop (www.sws-challenge.org). In the purchase-order mediation scenario, the RosettaNet protocol (www.rosettanet.org/Rosettanet/Public/) requires that the last Web service in a process receive a receipt and then finish by responding with acknowledgment. This sort of process is difficult to create with deductive synthesis. (For more on the scenario, see www.sws-challenge.org/wiki/index.php/Scenario:_Purchase_Order_Mediation.)

We're combining the process-integration problem as well as the Web service discovery and, eventually, Web service composition problems — using industrial standards.

The Truth and Reconciliation Workshop

Were they to understand each other, the AI and process communities could learn much from one another. Specifying what a process should accomplish is a good idea, and AI planning needs to incorporate some protocol requirements.

AI planning can also produce processes that solve specific goals, potentially producing use cases that facilitate process integration. Working with the Business Process Execution Language (BPEL; http://www-128.ibm.com/developerworks/library/specification/ws-bpel/), for example, most programmers construct processes or workflows that implement use cases, and then elaborate by hand. AI planning can produce some of these use cases, perhaps incorporating semi-automated process-based semantic matching methods.

As stated before in this column, we really understand each other best when we build something together. That's one of the SWS Challenge's objectives. Another is to take something approximating industrial problems and, hopefully, show that the various academic technologies aren't magic but rather useful techniques that can facilitate industrial problem solving (or at least programming).

Such complexity presents a barrier to participation. We're combining the process-integration problem as well as the Web service discovery and, eventually, Web service composition problems — using industrial standards. The academics must do a lot of work without necessarily making significant scientific advances, and industry folk have too much "real" work to divert their resources to the challenge problems.

Yet, our hope is that participating in this workshop brings real value: the certification of Semantic Web services technologies in solving different levels of difficult problems. We aspire to be the "CE" (Conformité Européenne) of such technologies, for both industry and academia, but certifying functionality rather than safety. Why should anyone read a paper that claims an advance if it hasn't been demonstrated on known problems? On the other hand, if the developer has demonstrated the advance to be real, why shouldn't industry use it?

If this workshop succeeds, with your help, we should see a real reconciliation among the two academic communities because they'll be working on the same problems and talking to each other about them and their technologies.

The Third Person

Advances among those two communities still might not impress the third community very much because real process-integration problems in industry are even more complex. This community is generally skeptical that the technologies described in academic papers can provide much programming relief; in fact, it's often downright scornful that focusing on technology will solve supply-chain (value-network) management problems.

The first problem is with data semantics: different organizations, even inside individual companies, often have different semantics for price, date, tax, product types, and so on. Several academic solutions to this issue actually exist, and some have even been fielded in industry, which suggests that something good will happen here, probably due to links in vertical market standards.

A second, more difficult, dimension of data and process semantics is ownership. Who gets to describe objects' various properties or even assign their unique URIs? This turns out to be a much more difficult problem than you might imagine, especially if we consider methods of distributing ownership rights — a crucial aspect of the problem that none of the academic semantic approaches currently considers.

Third, industry currently executes discovery on a business level (partners, acquisitions, setting up first meetings, and so on). This interaction is embedded deep in current processes, and if a

company has *value* networks (http://en.wikipedia.org/wiki/Value_network), it must build trust on a personal as well as legal level. Because pure mechanistic discovery provides no help, many in industry view academic research's current focus as irrelevant.

Finally, industry faces process semantic issues such as Bussler's power cord example, described in last issue's Peer to Peer column.[3] In this example, a process assumes early on that a line item that costs US$0 in a purchase order has no impact on the process. In reality, however, it must be consolidated at customer accounting systems, which don't recognize this assumption. Defining such process semantics goes beyond simply defining those of Web services — and business problems will simply resist automation in any case.

The set of legal and business problems constituting the context of business interactions is very large. Focusing on narrow technical research issues won't create the brave new vision known as the Northern California ideology, in which large companies no longer exist and most people work as private contractors for ephemeral virtual enterprises.[4,5] At the same time, electronic business standards bodies such as the Organization for the Advancement of Structured Information Standards (www.oasis-open.org) and RosettaNet are trying to address such issues by developing complex specifications for interactions. But not even these will allow complete automation of business-process synthesis.

Real people in companies are the ones with the deep understanding of their processes and partners. Academics and industry researchers would do well to provide technologies that facilitate such knowledge workers in developing interconnected systems. Purely automated approaches won't be feasible for a long time, if ever.

Much current semantic research is irrelevant in this business context. That said, some companies have realized the potential and are quietly developing their own semantic service approaches. To do so, they've formed their own research groups and are starting collaborations with major universities. A major focus is semi-automation, which offers less magic and more practical results.[6,7]

Investigating real industry problems offers academics a vast and challenging array of research problems. The incentive problem, however, is that academics get rewarded for interesting theoretical, rather than practical, advances — unless they start companies. So, we need new mechanisms for bringing the industrial and academic communities together and new incentives for both. The SWS Challenge is only a start. We welcome your suggestions for this initiative, as well as new ways to make progress.

The incentive problem is that academics get rewarded for interesting theoretical, rather than practical, advances — unless they start companies.

References

1. S. McIlraith and T. Son, "Adapting Golog for Composition of Semantic Web Services," *Proc. 8th Int'l Conf. Knowledge Representation and Reasoning* (KR 02), 2002, pp. 482–493; www.ksl.stanford.edu/people/sam/mci-son-kr02.ps.
2. H. Ludwig and C. Petrie, "Cross-Cutting Issues," *Proc. Dagstuhl Seminar on Service-Oriented Computing*, F. Curbera, B.J. Kramer, and M.P. Papazoglou, eds., Internationales Begegnungs und Forschungszentrum (IBFI), 2005; http://drops.dagstuhl.de/opus/volltexte/2006/526/pdf/05462.LudwigHeiko.Paper.526.pdf.
3. C. Petrie, "It's the Programming, Stupid," *IEEE Internet Computing*, vol. 10, no. 3, 2006, pp. 96–95.
4. S. Vaidhyanathan, *The Anarchist in the Library*, Basic Books, 2004.
5. C. Petrie, "Emergent Collectives for Work and Play," *Société de Stratégé AGIR Revue Generale de Stratégé*, Jan. 2005, pp. 146–152; http://www-cdr.stanford.edu/~petrie/revue/.
6. M. Sabbouh et al., *Using Semantic Web Technologies to Enable Interoperability of Disparate Information Systems*, tech. report, Mitre, 2005; www.mitre.org/work/tech_papers/tech_papers_05/05_1025/.
7. J. Raoa et al., "A Mixed Initiative Approach to Semantic Web Service Discovery and Composition — SAP's Guided Procedures Framework," *Proc. IEEE Int'l Conf. Web Services* (ICWS 06), IEEE CS Press, to be published in 2006.

Charles Petrie is a senior research scientist and consulting associate professor at Stanford University. His research interests include concurrent engineering, virtual enterprise management, and collective work. Petrie has a PhD in computer science from the University of Texas at Austin. He is EIC emeritus and a member of *IC*'s editorial board. Contact him at petrie@stanford.edu.

Christoph Bussler is an architect at Cisco Systems, where he is responsible for establishing the service-oriented architecture for one of Cisco's core units. Previously, he was a professor and executive director of the Digital Enterprise Research Institute (DERI) in Ireland. His research interests include Semantic Web services, B2B integration, workflow management, and business-rule systems. Bussler has a PhD in computer science from the University of Erlangen, Germany. Contact him at chbussler@aol.com.