A Wiki for Knowledge in Open Vocabulary, Executable English Over Distributed Data

Position Paper for the Workshop on eGovernment and the Web
18 - 19 June 2007
United States National Academy of Sciences
Washington DC, USA

Adrian Walker
Reengineering
Bristol, CT 06011-1412, USA
internet.business.logic@gmail.com

Abstract

This workshop seeks directions for improving the delivery of eGovernment services through the use of Web technologies, via usability, and via interoperation of information on the web.

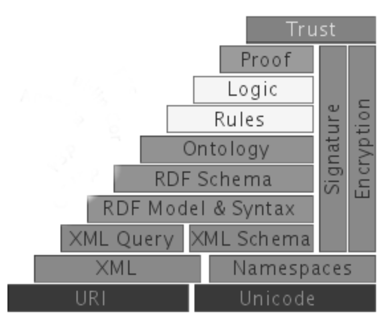
I describe some emerging technology that works as a kind of Wiki for gathering collective intelligence written in open vocabulary, largely open syntax, *executable* English. The knowledge so gathered contains a high level of semantics that can be used directly for interoperation of networked databases. The technology is in a system called Internet Business Logic, or IBL for short. The system is live on the Web (Reeng 2007), and as befits a Wiki, shared use is free.

I describe an online example about mining data to find possible meetings of arms dealers, and another example about supply chains in the oil industry.

Since the collective knowledge is in English, it is indexed by Google and other search engines. This means that writing new executable English knowledge can start with a search for knowledge that can be re-used. In this sense, the IBL is positioned to support new patterns of gathering and using collective executable English knowledge on the Web.

1. A Wide Technical View of Semantics

The well known "layer cake" diagram (Berners-Lee 2004) outlines a high level agenda for work on the Semantic Web:



There is much current work in progress under the heading of "Semantics", as in the interleaving of metadata with data in RDF or OWL, stemming from (Berners-Lee et al, 2001). Such work fits into the layers from XML to Ontology in the above diagram. It may be useful to think of this as "data semantics", or *Semantics1*.

In the diagram, there are boxes labeled Rules, Logic, Proof and Trust. The IBL can be viewed as one way of meeting some of the requirements indicated by those boxes, with an online system that combines *Semantics1* with two further kinds of meaning, as follows.

Semantics2 specifies what conclusions a reasoning engine should be able to infer from any set of rules and facts (Walker 1993), using a logical model theory (Apt et al 1988) that takes into account the semantics under which databases are used in practice. *Semantics3* concerns the meaning of English concepts at the author- and user-interface. The design of the IBL rests on making *Semantics1*, *2*, and *3* work together.

Adding and integrating the three semantic dimensions has the potential not only to support aspects of the Semantic Web, but also to ease some significant problems in database interoperation in government and commercial IT. According to (Forrester 2005): *Aligning IT strategy with business strategy has been one of the top three issues confronting IT and business executives for more than 20 years. Polling of CIOs and business executives conducted in 2004 revealed that aligning IT and business goals remains their No. 1 or 2 priority*.

2. The Author and User Point of View

From the author and user point view, the IBL is simply a collection of browser pages. You can specify an application in the form of business rules in open vocabulary, largely open syntax, executable English. As you do this, the system will check what you write for gaps in meaning, and will suggest how to proceed. You can then use the browser to run the application, and to get English explanations of the results at the government or business level.

Unlike “controlled English” approaches, such as in (Kuhn 2007), the IBL needs no external dictionary or grammar maintenance, yet the English semantics are strict. As you write business rules, you can type in new terms, such as government acronyms and phrases, and use them immediately.

From the information in the business rules, the IBL can automatically generate and run complex SQL queries and transactions to interoperate networked databases. It can then explain the results, in English, at the business level.

Since the author- and user-interface is a browser, teams can share the work of writing and testing their collective knowledge, and the teams can be distributed geographically and in time. Teams can also form ad-hoc around particular subjects.

Also, since the knowledge is in English, it is indexed and searched by Google and other engines, and is thus easier to find for re-use than application knowledge that is locked up in, say, Java.

3. Examples

Here are two examples that are on the web, where you can view, run and change them. Further examples are online at (Reeng 2007). You are cordially invited to write and run your own examples.

3.1 An Arms Dealer Meeting Example

Suppose that we have data about the departure and arrival times of airline flights, and also the passenger lists for the flights. We may be interested in finding out whether certain passengers could have held a meeting in or near an airport, based on data in a table like this:

a person traveling as this-name	left this-airport1	this-daygmt1	on this-flight	and arrived this-airport2	at this-daygmt2
Mohammed Al-Bisri	AMM	200610031300	BA125	LHR	200610040830
Arif Durrani	MNL	200610051800	SQ364	LAX	200610060900
Samir Hakim	LAX	200610062300	SQ364	LHR	200610071030
…..					

To find out about possible meetings, we can write several rules into a browser, including:

```
a person traveling as some-name was likely in or near some-airport from some-daygmt1 to some-daygmt2
a person traveling as some-other-name was likely in or near that-airport from some-daygmt3 to some-daygmt4
that-name comes before that-other-name alphabetically
there is an overlap between that-daygmt1 to that-daygmt2 and that-daygmt3 to that-daygmt4
for that-name at that-airport the latest meeting start is some-startgmt
for that-name at that-airport the earliest meeting end is some-endgmt
---------------------------------------------------------------------------------------------------
we found that-name could have met with the following person(s) near that-airport from that-startgmt to that-endgmt
          that-other-name
```

Running the rules gets results including:

```
we found Arif Durrani could have met with the following person(s) near LAX from 200610060900 to 200610061945
          Mohammed Al-Bisri
          Samir Hakim
```

To view and run the example, you can point a browser to (Reeng 2007) and select ArmsDealerMeeting1.

3.2 An Oil Industry Supply Chain Example

When a geographic region has a demand for a quantity of an oil product, it is in general possible to meet the demand using a number of equivalent products. Many factors influence the proportions of component products that are combined to make an optimal supply chain decision. The factors include the season of the year, the locations of available equivalent products, and the availability of suitable and timely transportation.

For our example (Kowalski and Walker 2005), we project that the target region NJ will need 1000 gallons of product 'y' in October. We then ask what alternative routes and modes-of-transportation (truck, train, boat, pipe) do we have to get that product to the region. Next we ask whether there's a refinery nearby that can produce the base product for finished product 'y'. With all of that, we finally say that we need a delivery plan that is optimized to deliver on time, make a profit, and beat the competition. However, if there is not enough of product 'y', then, depending on the region and the customers, product 'x' or 'z' will do as well; they're just variations of 'y' using different additives. But they'll only do just as well in region NJ for the season including October. This makes sales projections and marketing more complicated, but also gives us more competitive flexibility.

One of the rules for the supply chain task is shown below.

```
estimated demand some-id in some-region is for some-quantity gallons of some-finished-product
    in some-month of some-year
for demand that-id for that-finished-product refinery some-refinery can supply some-amount gallons of some-product
for demand that-id the refineries have altogether some-total gallons of acceptable base products
that-amount / that-total = some-long-fraction
that-long-fraction rounded to 2 places after the decimal point is some-fraction
-------------------------------------------------------------------------------------------------------------
for estimated demand that-id that-fraction of the order will be that-product from some-refinery
```

As of the writing of this paper, typing “demand estimated some-id” (without quotes) into Google finds the rules on the world wide web.

Here is a fragment of the SQL that is automatically generated from the rules:

```
select distinct x6,T2.PRODUCT,T1.NAME,T2.AMOUNT,x5 from
T6 tt1,T6 tt2,T5,T4,T3,T2,T1,T6,
(select x3 x6,T6.FINISHED_PRODUCT x7,T6.ID x8,tt1.ID x9,tt2.ID x10,sum(x4) x5 fromT6,T6 tt1,T6 tt2,
((select T6.ID x3,T3.PRODUCT1,T1.NAME,T2.AMOUNT x4,T2.PRODUCT fromT1,T2,T3,T4,T5,T6,T6 tt1,T6 tt2 where
T1.NAME=T2.NAME and T1.REGION=T6.REGION and T2.MONTH1=T4.MONTH1 andT2.MONTH1=T6.MONTH1 and
T2.PRODUCT=T3.PRODUCT2 and T4.MONTH1=T6.MONTH1 and T3.PRODUCT1=T6.FINISHED_PRODUCT and
T3.SEASON=T4.SEASON and T3.SEASON=T5.SEASON and T4.SEASON=T5.SEASON and T6.ID=tt1.ID and T6.ID=tt2.ID
and tt1.ID=tt2.ID)  union ....
```

It would be difficult to write such SQL reliably by hand, or to manually validate a result that the system has found. As a way of establishing trust, the system can explain each result in step-by-step, hypertexted English, at the business or scientific level. To view and run the example, you can point a browser to (Reeng 2007) and select Oil-IndustrySupplyChain1MySql1.

4. Summary

This position paper argues that we can make progress in some of the workshop directions for usability, and for interoperation of information on the web, by taking a wide technical view of “semantics”. The view is that data semantics (*Semantics1*) is necessary, but not sufficient, and should be integrated with application semantics consisting of (a) inference semantics (*Semantics2*) and (b) the meaning of English concepts at the author- and user-interface (*Semantics3*).

I described some emerging technology for this purpose. It works as a kind of Wiki for gathering collective intelligence written in open vocabulary, largely open syntax, *executable* English. The technology is different from the "controlled English" approach. The knowledge that is gathered contains application semantics that can be directly used for interoperation of networked databases. The technology is in a system called Internet Business Logic that is live on the Web. Shared use of the system is free.

Since the collective knowledge is in English, it is indexed by Google and other search engines. This means that writing some new executable knowledge can start with a search for knowledge can be re-used. In this sense, the IBL is positioned to support new patterns of gathering and using collective executable English knowledge on the Web.

The system can also automatically generate and run SQL to interoperate networked databases. The generated SQL can be too complex to write or check reliably by hand, but its results can be explained in English**,** at the business level.

5. References

Apt, Krzysztof R.; Howard A. Blair, and Adrian Walker. Towards a theory of declarative knowledge. In Jack Minker, editor, Foundations of Deductive Databases and Logic Programming, pages 89--148. Morgan Kaufmann, Los Altos, CA, 1988.

Berners-Lee, T. 2004. www.w3.org/2004/Talks/0412-RDF-functions/slide4-0.html .

Berners-Lee, T; J. Hendler, O. Lassila. 2001. The Semantic Web, Scientific American, May 2001, pp. 28-37.

Forrester 2005. "IT and Business Alignment: Are we there Yet?", Forrester Research Report, www.forrester.com .

Kowalski, Ted and Adrian Walker. 2005. Oil Industry Supply Chain Management Using English Business Rules Over SQL. www.reengineeringllc.com/Oil_Industry_Supply_Chain_by_Kowalski_and_Walker.pdf .

Kuhn 2007. AceRules: Executing Rules in Controlled Natural Language http://attempto.ifi.unizh.ch/site/pubs/papers/kuhn07acerules.pdf .

Reeng 2007. Internet Business Logic system, online at http://www.reengineeringllc.com .

Walker, A. 1993. Backchain Iteration: Towards a Practical Inference Method that is Simple Enough to be Proved Terminating, Sound and Complete. Journal of Automated Reasoning, 11:1-2 .