Martin Flanagan: Accelerating investigative discovery

Abstract

Since the beginning of the computing era, a key problem has been how to let 'ordinary' people ask arbitrarily deep and meaningful questions of large collections of data without being forced to resort to programming experts.

The historical 'solution' was to have teams of programmers writing 'applications' for 'users' - an expensive approach whose deliverable becomes rapidly out of date leading to the well known silo problem.

The Semantic Web is purposed to replace the increasingly cumbersome nature of this model - using Ontologies to describe concepts and their relationships, questions can be now phrased in terms that represent the questioners' domain of expertise in familiar language, with software then translating that into the computer code needed to retrieve the answers.

The Semantic Discovery System (SDS) is InSilico Discovery's software product that implements this 'simple questions/relevant answers' Semantic Web vision, and is especially acclaimed for its unique ability to efficiently retrieve the answers from multiple distributed and disparate sources of the organisation's internal production data - as well of course as external Web sources.

Gartner call this the Corporate Semantic Web - Organisations understandably want to leverage the huge value existing in production systems, but without being forced to do a mass data migration to a new architecture - they want a "Semantic Web Bridge". This bridge is the value SDS supplies - the organisational data remains in situ but it can now fulfil two purposes simultaneously - continuing to serve day to day production needs but also now supporting ad hoc research queries.

SDS achieves this capability by using Semantic Web technologies (OWL, SPARQL, RDF etc) to represent a logical view of the world coupled with a Federated Query system to retrieve the physical data that will always reside in situ inside databases, files, proprietary systems etc.

SDS has been built over a 10 year period (in collaboration with Universities of Pennsylvania, Manchester and GSK) based on referenceable implementations at major Pharmaceutical companies.
This paper discusses capabilities of the SDS product family and plans for the immediate future.

**About the speaker**

Martin has over 20 years experience in delivering software solutions and services. He began his professional career at the age of 14 by writing the first educational software for McGraw-Hill commercially available for the Commodore 64. Since that time, he has been a developer of software solutions and advocate of emerging technology adoption. Martin brings extensive experience in the delivery of semantic data integration to the management team and has assisted such organisations as the Australian Bureau of Statistics, Paramount Pictures and the Queensland University of Technology. Martin has held various engineering, consulting and management positions with software companies such as Compuware Corporation, Open Environment Corporation, Jarrah Technologies and geneticXchange Inc.