20130810 Meeting with Dickson Lukose of MIMOS

Attendees:

Dickson Lukose MIMOS, Renan Souza UFRJ/SLAC, Bebo White and Les Cottrell SLAC.

MIMOS & Semantic Technology

Dickson gave a brief introduction to MIMOS and their interest in Semantic Web Technology (ST).

MIMOS is a company set up by the Malaysian government to do research in Information Technology (IT) and move the technology to benefit the country. It is funded by the Ministry of Science. There are about 700 people at MIMOS. There are 3 divisions: software for ICT (including knowledge interaction, video analytics, wireless and security), hardware (includes micro energy technology), manufacturing.

The ST interest of Dickson's team began in 2007. It was very new then and there was a lot of learning and many experts were invited to MIMOS to assist in the learning process and to advise. They identified a niche area to focus on as being the area of analysis. The team has 3 major areas: Reasoning, Analysis and knowledge interaction. They deal with unstructured data including multi-media. Their serious work began in 2009 and they now have ~ 20 papers and now are working on moving some technology developed to local industry and government.

One of Dickson's current interests is how to locate, index and search the relevant repositories form the millions of repositories that will exist. In his view the Semantic web is in its early stages like the web was in the early 1990s.

PingER

Les Cottrell introduced PingER.

PingER is a Internet performance monitoring project going back to 1995. It use the simple ubiquitous Internet ping tool to make active end-to-end measurements every 30 minutes. It was originally set up to provide monitoring for High Energy Physics worldwide collaborations. Starting this century it was extended to quantify the Digital Divide in Internet performance for regions and countries of the world.

There are now ~ 90 monitoring sites in 21 countries making measurements to over 800 remote hosts in over 160 countries of the world. In all there are over 8000 monitor-remote host measurement pairs with the measurements being made every 30 minutes. The measurement data from the monitors is gathered nightly by two archive/analysis/presentation sites at SLAC and at NUST in Islamabad, Pakistan. There are direct metrics including: round trip time (min/avg/max), jitter, loss, unreachability, and out of order packets. In addition there are derived metrics including throughput, Mean Opinion Score, and directivity (identifies the directness of the connection between 2 nodes at known locations. Directness values close to one mean the path between the hosts follows a roughly great circle route. Values much smaller than 1 mean the path is very indirect.)

The results are available publicly via the web. It has been used for: identifying problems (in particular last mile problems); identifying sites that need upgrading; setting expectations for a collaboration such as High Energy Physics or the Comprehensive Nuclear-Test-Ban Treaty Organization; setting up and verifying Service Level Agreements; deciding whether VoIP is going to work well enough for a phone conference; deciding where to base a software effort; evaluating the effects of major events such as the Japanese earthquake/tsunami, the cable cuts in the Mediterranean, uprisings such as the "Arab Spring"; quantifying the "Digital Divide", deciding which replicated server (such as Hotmail) to use to respond to a user; providing analysis and guidance to decision makers, funding agencies and politicians on he network performance in their area; etc.

Casting PingER data into RDF format

Renan is from UFRJ in Rio Da Janeiro, Brazil. He has been an intern at SLAC for 3 months. He leaves SLAC on the 18th August.

PingER has a huge amount of data and, until the conclusion of this project, the easiest way to retrieve the data is through Pingtable. Pingtable provides a friendly web interface to retrieve PingER raw data (millions of text files) and load it into a human readable HTML page. However, this is not a web standard and crossing PingER data to generate very specific information may not be possible or extremely difficult using the existing way to retrieve PingER data. Renan has therefor put together a project to cast the analyzed PingER data into RDF hence providing Linked Open Data Access. This enables accessing and analyzing the data using the Sesame Open Source framework.

The RDF data base is populated nightly. It contains 17 million triples. It has records of all the monitor-remote pairs for over 10 metrics aggregated by day for the last 365 days, as well as aggregated monthly and yearly going back to 1998. There are web forms enabling the user to select the source and destination, the metric(s), the time window etc. Then by means of SPARQL queries the data is shown as time series.

To demonstrate how the data can be used for Mashups, Renan has developed a Google map based application that presents the PingER performance seen from SLAC to world universities, together with the size of the university (endowment, students etc.) and university type (private, public, ...). The University data is taken from DBpedia and the location data from Geonames.

Going Forward

Dickson is happy to be involved with the PingER project. As part of this, it would be good to install the PingER monitoring software in a host at MIMOS to instantiate a monitoring host there. This would add measurements from MIMOs to the world, with an emphasis on sites in Malaysia and S. E. Asia. This data could be mined to understand and make case studies on the Internet's performance between MIMOS and the rest of the world, and also the performance within Malaysia. The PingER performance data can also be mined to provide mashups with other country data such as Gross National Product per person, computers per person, cell phone license per person, wired phones per person, fraction of populace getting tertiary education etc.

To assist in installing the PingER monitoring application at MIMOS, Les will send an invitation to Dickson with all the details. Dickson will take this to the MIMOS IT Engineering group to recommend they install the application.

Also Renan will provide Dickson with documentation on the PingEr Linked Open Data project and RDF access to PingER data.