**Team Meeting 44**

|  |  |  |  |
| --- | --- | --- | --- |
| **Goals:** | * Educate Team on SORASCS training | | |
| **Required Preparation:** |  | | |
| **When:** | 11/09/2009, 9:00 AM – 12:00 PM | **Where:** | Smith Hall |
| **Attendees:** | Nina, Aparup | | |
| **Scribe:** | Nina, Aparup | **Facilitator:** | Training |

**Minutes**

* Text Data can be in following forms
  + Raw Text data, semi-structure data (email), structured data (.csv file)
* **SORASCS**
  + Open source but not available to public yet.
  + In inception phase.
  + Goal of SORASCS is to help analyst create, save, edit and share workflows. This is currently not possible with existing tools.
  + Incorporate other tools to I/F to SORASCS to create “specialized” workflows (VIBES is an example tool which is used to for generating belief system).
  + SORASCS Server architecture has 3 pieces to it – Registry, Director (workflow engine), Provenence
  + SORASCS system can be broken down into three layers
    - Tools Layer (top level) – this is where our workflow tool sits. Other tools like VIBES sit.
    - Workflow Layer – there is where the workflow execution takes place on SORASCS
    - SORASCS Layer – this is where new CASOS/other tools services could be plugged into SORASCS
  + Roles within SORASCS.
    - Routine Analysts (Field Guy) – should be able to click on a workflow on SORASCS and get the results instantly.
    - Expert Analysts – Responsible for creating and sharing the workflows.
    - Tool Developers– 2 types
      * One who work on creating new tools / providing specialized front end such as VIBES
      * One that use the SORASCS layer to plugin/register new components services into SORASCS.
    - SORASCS is very important for analysts as it will allow them to compose WF using new WF editor that MSE team is building, share WF, execute WF, able to see the analysis steps, save them, replay them in future, when looking at some output from the analysis, analyst will be able to go back and figure out how did he/she get this report. Workflows composition and saving will help them do this.
    - Unusual but extremely important requirement is the workflow tool should be able to handle both Fat and thin clients. Analyst could use some thin client (web services) to start creating workflow and switch to fat client to do some experiments and then come back to the workflow to continue.
    - Other analysis tool besides ORA are UCINET and Pythia (these are non-CASOS tools)
    - Data transformation should be completely transparent to the analyst.
    - Data sharing issue is outside of scope of SORASCS. SORASCS can provide administrative privileges but negotiation between the two agencies on what data to share is done outside of SORASCS boundary.
    - Component organization – based on “types” for example, “text type” for Automap.
    - 90% of the analyst wants to use only default settings of the components. It is not because they don’t care but it is work load issue, training issue and analysts think that default settings are good enough. They want something just good enough cause they are always running on a short time duration to get analysis done
    - Each of the ORA’s reports have the API written so it could be accessed from anywhere.
    - Performance – Execution on some of the ORA components can take from few seconds, to few minutes to 8 hours. The speed is the function of size of the network. 10^6 nodes are some of the longest measure they have seen that take long time to execute.
* **Automap**
  + Broken down into 4 major functionalities.
    - Traditional content analysis (words, text, delete lists)
    - Semantic Network extraction (how are words and concepts linked to gather)
    - Meta Network extraction (organizes the network in who, what, where why relationships nodes).
    - Belief Extraction
  + It is extremely hard to infer the temporal data (time/when)
    - Time the text was generated vs time that the content in the text actually took place. They usually not do this.
* **ORA**
  + Most frequently used reports / network views
    - Cognitive Demand, Eigen Vector, Centrality measure, key entity report, sphere of influence report, group locator report.
    - This is not a prediction tool, only analysis cause data has already happened.
* **Construct**
  + How can I break up the network
  + How can I re-enforce the network
  + Analyst goal is to do above two what-if analysis and they use the ORA reports and network views to help them.