SeerSuite
By Brian Lieberman and Robert Wagner
CS 4624, Virginia Tech
Blacksburg, VA
05/06/13
Tarek Kanan
Project Aim

There are many projects involving digital libraries and research, but there has not been much work with attempting to implement simultaneous multilingual support.

The goal of this project is to be able to crawl documents in a multilingual fashion using SeerSuite and to get systems here, as well as in Qatar, working together.

We are using CentOS machines here, as well as in Qatar, and we are going to port the configuration we establish here to their machines.
SeerSuite Structure

Digital Libraries and Multilingual Data

created by

Web Crawlers

such as

SeerSuite

stores

runs on

Apache Tomcat

viewed by

CiteSeerx Application

contains

Axis 2

imported by

CiteSeerx Application

runs

DOI Server

Data

indexed and searched by

Solr

Languages

such as

Arabic

English

MySQL

Accesses
Progress

- Set up UNIX OS and local dependencies such as Java, VPN, and remote access. The dependencies are required programs to make the base OS work correctly.

- Installed dependencies for SeerSuite installation which included: Tomcat, Solr, Axis2, Ant, MySQL. These dependencies are required for running the web app and for storing data, etc.

- Configured the installed dependencies and copied in resources necessary for SeerSuite to run

- Set up a local repository folder to hold crawled documents

- Configured the SeerSuite files to point to the correct paths that we set up during installation

- Built the SeerSuite application and deployed it
Challenges and Solutions

- Acquiring Red Hat Linux licences forced switch to CentOS
- Installing JRE for VPN, resolved with switch to CentOS
- Getting Solr to run under Apache Tomcat, worked with student from Penn State and in depth research to resolve.
- Physical access to lab, was solved with installation of TeamViewer for remote access.
- Configuring SeerSuite settings to access databases and installed server applications, corrected login info fixed issue.
- Communicating with SeerSuite experts, we got new experts and hopefully will have more success with them
Conclusion

- Successfully built and deployed web application
- Currently working on importing sample data and crawling documents and talking with people from Penn State to help.
- Need to get in touch with more people from Penn State that can be helpful
- Once data is imported, application will be fully functional
- Documentation includes instructions for reproducing installation in Qatar
Questions?