



White Paper

Zheng Technology

Challenges for Online Research Technologies

The goal of conducting online research is aimed at finding information based on a combination of behaviours and attributes. The challenge is to be able to find anything, and everything, anywhere on the Internet or a large scale enterprise Intranet comprised of heterogeneous sub-networks.

The size of the Internet today makes it practically impossible for any online research system to encompass it in its entirety. Gathered information is already out of date by the time any system has traversed a section of it. Stored information on an enterprise network in a constant state of flux becomes a challenge when scope and content change rapidly.

There is a need for a comprehensive, scalable and customizable online research technology with high performance. The Zheng online research system (Zheng) was been designed to meet that need.

Zheng System Overview

The Zheng¹ technology, invented and developed by ParetoLogic Inc., is an online research system that identifies information within a network, detects and logs attributes of interest on identified information, categorizes and generates reports based on the purpose of a research project.

Major System Components

The Zheng¹ research system consists of four major components:

- Network Navigation Subsystem
- Data Acquisition and Logging Subsystem
- Data Analysis and Adjudication Subsystem
- Data Categorization and Reporting Subsystem

Network Navigation Subsystem

The Network Navigation Process Subsystem is responsible for traversing a network in a manner defined by the nature of the target network and the scope of a research project.

Data Acquisition and Logging Subsystem

Two consecutive steps are performed when Zheng identifies an entity within the scope of a research project. The first step is to acquire data based on collectible attributes and/or behavioural characteristics of the entity. The second step is to log that data so its' original attributes and/or behaviour can be processed further.

Data Analysis and Adjudication Subsystem

The core process of the Zheng research system is the Data Analysis and Adjudication Subsystem. As data is acquired, target entities are analyzed according to the nature and specified goals of the online research project. Zheng incorporates multiple research methods and tools as it records its'

¹ The Zheng technology was named after the famous Chinese mariner Zheng He who sailed from China to many places throughout South Pacific, Indian Ocean, Persian Gulf and distant Africa in seven epic voyages from 1405 and 1433, some 80 years before Columbus' voyages to America. Zheng He's flag ship was 400-foot long compared to Columbus' St. Maria which was 85 feet.

findings. Adjudication on each processed entity draws preliminary conclusions based on the initial analysis findings.

Data Categorization and Reporting Subsystem

Zheng categorizes target entities within the analysis findings by their attributes and/or behaviour and stores the results in a database. Final reports are then generated on the results to fulfill the original purpose for conducting the online research project.

Technologies and Architecture

Zheng has incorporated several technologies to integrate and maximize the efficiency and effectiveness of each subsystem as they collaborate to produce expected results. These technologies consist of Selective, Prioritized and Parallelized Network Traversal; Behavioural-based Data Acquisition and Logging; Hybrid Analysis and Adjudication; Adjustable Combination of Automation and Human Intervention; in conjunction with Componential Design and Scalable Architecture.

Selective, Prioritized, and Parallelized Network Traversal

Zheng meets the demand for an effective and efficient network traversal component by combining three technologies: Selection, Prioritization and Parallelization.

Selective network traversal starts with the manual configuration of the network traversal component. Manual configuration entails setting the probing instructions and establishing the actions it should take when visiting an entity over the network. The selection process can be adjusted to ensure that entities previously visited are not revisited, unless preceded by an existing revisit policy. Due to both the expected courtesy and the bandwidth cost of the network traversal component, the selection technology appropriately confines both the breadth and the depth of the network probing. Selection criteria can be refined further to provide more thorough guidance for network traversal component as the Zheng system begins to produce meaningful and expected results.

Prioritized network traversal dynamically controls and adjusts Zheng's network traversal pattern so probing efforts are concentrated on the most relevant section of the network, and not just a random sample of the network. The prioritization technology also allows manual intervention in supplying the network traversal component with a list of entities of high interest to the research project without restarting the probing process.

Parallelized network traversal allows multiple probing processes to run concurrently in a coordinated manner. Zheng dynamically assigns new entities to different probing sequences. Depending on the intensity and the expected coverage of the network traversal, Zheng can add or remove probing processes without interrupting the rest of the system.

Behavioral-based Data Acquisition and Logging

Fundamental differences exist between attribute-based data acquisition and behavioural-based data acquisition. The former focuses on extracting physical attributes from static objects such as documents and files on a corporate LAN while the latter is used to study the natures of dynamic entities such as web pages with embedded ActiveX controls or Visual Basic scripts. Behavioural-based data acquisition and logging is thus expected to study the activities performed by a target entity, extract key behavioural attributes of these activities and record the attributes in a manner so that the original activities can be resembled, replicated, analyzed and understood. By nature, behavioural-based data acquisition and logging is more sophisticated and challenging.

Zheng is engineered to perform both attribute-based and behavioural-based data acquisition and logging, with an emphasis focused on the latter. The Data Acquisition and Logging Service (Subsystem) observes the interactive relationship between the 'observer', which is Data Acquisition and Logging service, and the 'observable', which is the entity being studied.

Hybrid Analysis and Adjudication Technologies

Designed to be a multipurpose online research system, Zheng incorporates a hybrid Analysis and Adjudication process which utilizes, but is not limited to, the following technologies:

- **Blacklist, Whitelist and Greylist:** the traditional approach of using Blacklist and Whitelist is proven to be effective when it is applied to research subjects with adequate academic and empirical data. In addition, Zheng allows a third list – the Greylist – to be used for further analyzing entities with unknown attributes and/or ambiguous behavior.
- **Lexical Analysis:** Zheng goes beyond just recognizing “key words” or “vocabularies” to identify sequence and states of tokens which contributes to analyzing the behaviors and the activities of entities under research.
- **Digital Signature:** extracting pattern and constructing digital signature empowers Zheng to quickly and accurately identify entities, whose static and/or behavioral attributes contain recurrent patterns.
- **Heuristic Engine:** to further achieve automation and dynamic learning capability, heuristic engine is utilized to store knowledge and procedures of analysis and adjudication from confirmed results.
- **Human Adjudication:** at the discretion of the researchers and enabled by Zheng, human intervention ensures the research effort and process is completely inline with the original goals and purpose of the research project.

By using various combinations of the above technologies, the Analysis and Adjudication process is more thorough, more comprehensive and more scalable.

Componential Design and Scalable Architecture

Componential design makes it easy for Zheng to be constructed, customized and maintained. Each of the four major subsystems can be configured to perform new tasks without impacting existing functionalities or performance of the rest of the system.

An online research project can be configured and initiated within a particular subsystem independent of the entire system. Maintenance work can also be performed separately and on an as needed basis for each subsystem.

Furthermore, Zheng was built on a scalable architecture so that it can accomplish online research projects independent of their scope, intensity or complexity. Scalability and performance of each major subsystem can be individually enlarged and strengthened by dynamically adding more processing capacity.

Adjustable Combination of Automation and Human Intervention

A successful online research system is expected to be highly automated so that it can meet the demand of high-volume processing within any timeframe. Zheng provides full automation by allowing human intervention throughout the operation. Each subsystem supports adjustable combination of automation and human intervention so that researchers can freely choose when they wish to intervene with the online research process, and which parts of the system they choose to interact with.

Putting It All Together

The potentials of Zheng being used as a powerful online research system is better reflected when all four components are constructed and connected together. A fully integrated Zheng system works much like a self-learning system that accumulates knowledge and experience through probing the environment, acquiring information, learning what's more pertinent to the purposes of

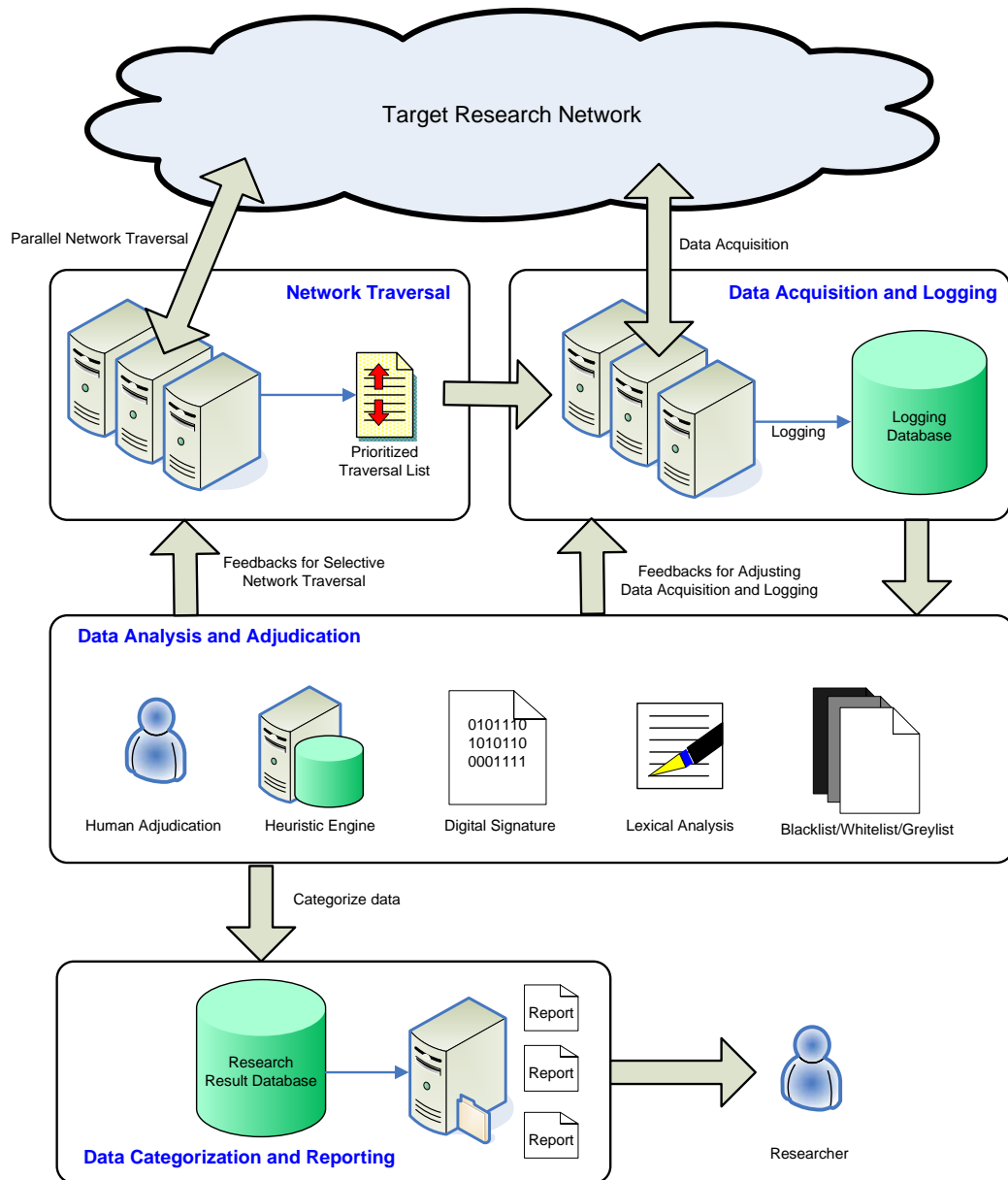


Figure 1: Zheng Online Research Technology

the research and adjust its course to explore segments of the network that potentially contains richer results. Walking through a typical workflow of how Zheng performs a research task better explains this idea.

1. The Network Traversal Subsystem's initial configuration includes a seed list of entities with which to start navigating the network.
2. Working in parallel, multiple traversal processes probes the network concurrently.
3. A prioritized network traversal list is generated according to the preliminary information gathered by the Network Traversal Subsystem. The list is based on the most relevant attributes pertaining to the interest of the research.
4. According to the priority list, the Data Acquisition and Logging service revisits entities on the priority list to collect information by observing and recording how the entities behave.
5. Information collected is then analyzed and adjudicated using a combination of various analysis and adjudication technologies, with the option of involving adjudication performed by real persons.
6. Based on the result of the data analysis and adjudication, feedbacks are given to both the Network Traversal Subsystem and the Data Acquisition and Logging Subsystem. The former can be adjusted to traverse specific areas of the network with a better prioritized entity list while the latter can be tuned to collect data that helps to produce results with higher accuracy.
7. Research results are then stored in database which provides categorization and notations. Various reports can be generated and delivered to the researchers.

Summary

To effectively conduct an online research project over today's massive and fast-changing networks, a solution must address the growing number of difficulties posed by both the challenging nature of the research and the increased expectation of the researchers. These demands call for a solution that directs its research effort to areas of the network with greatest potential, collects appropriate data that can be used to resemble the original behaviors of entities under research, accommodates various combinations of analysis tools and methods and is still of high-performance and easy to maintain. Few commercialized solutions exist to address the gamut of requirements as well as the growing demand.

Zheng adopts a componential design that integrates a few key technologies that are essential to the success of online research projects:

- Selective, prioritized and parallel network traversal
- Behavioral-based data acquisition and logging
- Hybrid data analysis capabilities
- Adjustable automation and human intervention

Further empowered by the scalable architectural design, which leads to high-performance, allows Zheng to meet the demands of the most intensive and comprehensive jobs. Like the nature of many online research projects, where key research steps and processes can be further tuned and adjusted, each subsystem can be customized, modified and strengthened so that Zheng, working as an integrated system, delivers results that fulfill the original demand for the users.

About ParetoLogic Inc.

ParetoLogic is an international software development company headquartered in Victoria, British Columbia, Canada. A member of SIIA, we specialize in providing advanced security applications for enterprise, business and personal computer users. These include custom software solutions for business and government.

Our proprietary Zheng Technology transcends the traditional approach to protection, offering a more robust set of information gathering and dissemination tools enabling preemptive protection previously unavailable.

All information contained in this document is the proprietary information of ParetoLogic, Inc. and is protected by international copyright treaties. This document contains information that is privileged and confidential. Any disclosure, distribution or copying of this document without the prior written consent of ParetoLogic is strictly prohibited under applicable law. ParetoLogic, and Zheng, are registered trademarks or trademarks of ParetoLogic, Inc. in Canada and in other countries. All other trademarks are the property of their respective owners.

Copyright © 2005 ParetoLogic, Inc. All rights reserved.