Requirements Definition

CS 4310 Spring 2013

Introduction

Current trends and fluctuations in the Earth's climate have resulted in an increased focus by scientists to study changes in environmental conditions to better understand climate change and associated impacts. Scientists use advanced sensor technology such as meteorological towers, wireless sensor networks, and robotic trams equipped with sensors to collect data at remote research sites. As the amount of data collection instruments introduced in the field increases, so does the volume of environmental sensor data acquired by such instruments. The measurements taken by sensors are discrete samples of physical phenomena and are subject to review of their accuracy dependent on their location.

The Problem

For some scientific projects, instrumentation typically does not include mechanisms to detect *anomalies* as data are collected. In this context, an *anomaly* is a deviation from an expected sensor data value. An anomaly in data does not always represent errors: it may represent environmental variability requiring further analysis. A common practice for environmental scientists is to collect sensor data for extended periods of time, possibly creating numerous individual data files. During data collection, these files are typically not checked to ensure that they adhere to predefined data quality standards. Checking often occurs when the files are transferred to a database. A challenge with this practice is that a large amount of incorrect data can be collected due, for example, to a faulty sensor, and this can be undetected for extended periods of time. Furthermore, if data are identified as incorrect during analysis at a much later time, the data gathering process may have to be repeated.

Repeating data collection is expensive especially when the site is at a remote location. Sensor technology needs to be redeployed and possibly recalibrated; the amount of time required to gather the data can be significant. For time sensitive data that is required for policy decision-making, it might not be even possible to repeat the data gathering process. Because

environmental sensor data can be non-reproducible entities, i.e. the observations at a given time and set of conditions cannot be repeated, the knowledge that could have been obtained from the correct data is not captured.

Requirements Overview. To assist scientists in locating anomalies in scientific sensor data, Cyber-ShARE Center created a tool to specify sensor data properties that can monitor data collected in near real time or data that has been stored in files. The overall purpose of the proposed system is to enhance support for specifying data properties and to apply these properties to monitor sensor data files that may have different formats.

There are two viewpoints in this approach. The notion of using patterns and scope to facilitate specification of commonly used data properties comes from software engineering. The other is the field scientist whose interest is in identifying anomalies in the data set. How can the system facilitate their work in the field?

Your task is to build a system that addresses the following:

- Improve the user interface for the data property specification tool;
- Provide scientists with the ability to create, retrieve, update, or delete sensor data properties using a mobile device;
- Provide the ability to describe the file formats for different types of sensors to allow the system to read a file and check for anomalies using the appropriate properties; and
- Provide the ability to graph the processed data and show where data readings are not satisfying specified data properties.