

ABSTRACT

Dynamic conversion of legacy web-based data
for real time XML-based programming

Christopher Ottley

A significant issue on the Internet today is that of the conversion of legacy web-based data to a form accessible to modern applications by way of a standards based XML API. A cost effective, dynamic means of providing this interface would aid both the development of new applications accessing legacy web systems and the migration of existing web systems to more modern versions. The design and implementation of an application and API that offers such an XML-based solution is presented.

The solution explored, called BunnyX, sits as a proxy between the legacy web-based data and the XML-aware applications. It dynamically converts HTML pages into XML according to structure and functionality attributes. It uses XML and Java technology to provide flexible interaction with the remotely located data and logic. It also uses a custom-developed Reducing Mountain Algorithm (RMA) which is a simple, fast, structural reduction algorithm that provides a simplified but related view of the data present in the HTML page requested.

The behavior of the RMA and proxy were verified by a specifically developed test suite. The produced API's speed, flexibility and quality were also evaluated in comparison to applications developed using legacy web technology. It was found that the XML versions and services provided a very flexible API, but the quality of the API was dependent on each specific HTML page and its speed was less favorable due to processing done and XML packaging overhead.

The purpose of this study is to determine a flexible and cost effective means of dynamically accessing web-application logic and data for data extraction and scripted logic tasks. The objective is to provide the design and implementation of an API that would allow for easy migration of existing web applications and access to both data and remotely accessible logic.

Keywords: XML, Web Service, Dynamic, Proxy, Application, RMA.