BizCQ: Using Continual Queries to Cope with Changes in Business Information Exchange

Wei Tang
NCR Teradata Division
100 N. Sepulveda Blvd.
El Segundo, CA 90245, USA
310-524-6072

wt185001@ncr.com

Kipp Jones
nuBridges, LLC.
Building 400, Suite 632
Atlanta, GA 30328, USA
770-730-3722

kjones@nubridges.com

Ling Liu, Calton Pu
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280, USA
404-385-{1139, 1106}

{lingliu,calton}@cc.gatech.edu

ABSTRACT
In this poster, we propose the framework of BizCQ, a system to apply *Continual Queries* [7][8] on Web-based content to manage information exchanges between two business partners. In this poster, we describe ways to leverage previous research in Web monitoring techniques applied to the everyday problem of managing change within a business environment, and focus on the difficulties of managing changes that are caused by external parties in business-to-business (B2B) information exchanges.

Categories and Subject Descriptors: H.3.1 Content Analysis and Indexing, H.3.3 Information Search and Retrieval (*Information filtering)*, H3.4 Systems and Software H.3.5 Online Information Services

General Terms: Experimentation, Human Factors.

Keywords: Change response, semantic web, information quality, continual query, business-to-business, B2B.

1. INTRODUCTION
Web page monitoring tools [1][7][8][10] have been developed and used to help people keep current as information on the Web has changed. These monitoring systems differ in quality of difference generation and summarization as well as methods of notification delivery. To date, the uses have been focused on keeping people up to date. However, very little work has been done to explore the use for automating and integrating change response to business processes impacted by the changes.

In this poster, we propose a framework to integrate automated change detection and notification systems for response automation. In particular, we apply the WebCQ system [7] to solve real business problems and study the extensions to adapt to changes based on the type, quality, and semantics of the changes detected.

This framework is applied to a problem that has plagued companies that do electronic business-to-business information exchange. This problem centers on the need for systems to remain compliant to specified document formats and protocols while continuously modifying and upgrading the computing systems to support ongoing business needs. In short, it takes energy to continuously update an information exchange system in order to avoid the degradation of that system.

There are a number of hurdles that must be overcome in order to

reach the goal of automating change management. This poster will explore three of these topics: change detection, change response and emerging standards.

2. SYSTEM ARCHITECTURE
The BizCQ framework builds on top of the current WebCQ system in that it incorporates a change response service. In WebCQ [7][8], users establish 'sentinels' as software agents to watch for changes to various Web pages containing information of interest. The difference is that besides the normal email notification method, an external response action can be registered for trusted parties at sentinel installation time.

A response action field (see Figure 1) is used to connect the BizCQ change response subsystem with the client-side response action proxy. In case of detected change happening, the client-side action will be invoked for appropriate responses. For a detailed description of other system components, please see [7][8].
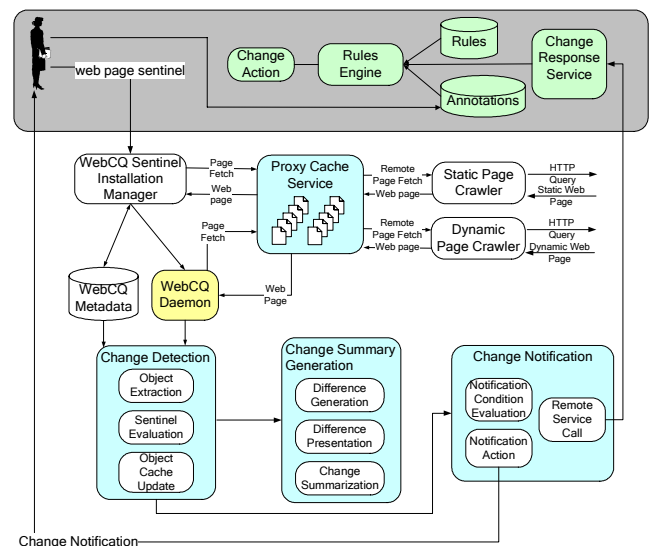

Figure 1. BizCQ Framework Overview

3. CHANGE DETECTION
Change detection is crucial to change monitoring services. The main tasks of change detection include accurately extracting and identifying objects on the target Web page, detecting changes between the page's last visited snapshot and the current copy, and generating visual differences and summarizations. The change detection sub-system is based on work done previously by [7][8].

BizCQ inherits the basic semantics for Web page changes and their associated page sentinel definitions [7][8]. These sentinels are created to track relevant page changes, such as hyperlinks, images, words, phrases, lists, and tables and are the basic building blocks for identifying application domain specific changes.

There are several classes of changes: *Structural changes, Informational changes*, *Presentational changes* and *Semantic changes*. The most important class for this problem domain are semantic changes, which are generally the most difficult to model and capture.

4. CHANGE RESPONSE FOR INFORMATION EXCHANGES

Business information exchanges often involve complex and often manual processes. Because many updates will trigger certain "status" changes, the ability to monitor the changes and respond to them becomes inevitably important. The ability to formulate a response based on the type, semantics, and value of the change is an important step towards automating the monitoring and change response process.

To help the BizCQ system provide relevance, we have proposed an extension to the WebCQ system to provide annotation capability to the definition of sentinels. This annotated information provides hints to both the human respondent as well as the automated response system.

The Web service notification allows the system to dynamically change and grow its capabilities over time. Through this loosely coupled connection, the change monitoring, the change measurement, and the change response can all be modified and expanded without impacting the other subsystems.

5. EMERGING STANDARDS

A robust system must be able to provide a level of flexibility in order to adjust to emerging technology and standards.

We believe that newer mechanisms, which are being used throughout the Internet such as RSS [9], can be leveraged to provide another level of service and scalability to the change-monitoring problem. This XML format provides a mechanism to publish 'channels' of related information and could be used to publish relevant information related to changes on a business partner's Web site.

Scaling now becomes much simpler, and can be modeled with set mathematics along with change characteristics of the site, as:

$$|R| + |\Delta R \cap S| = \# \ of \ HTTP \ Requests$$

Where R is the set of RSS Sentinels, ΔR represent the set of RSS feeds (the changes for that site), and S is the set of the standard Sentinels. This assumes all members of S are members of R.

Determining the type and semantics of a change are another area where standards can play an important role. Each Web site conforms to its own structure, format, and amount of information published and there is very little correlation between these various systems.

One standard that has been proposed but not widely adopted is the gXML (guideline XML) [5] standard. There are several other standards that are being adopted that attempt to help with the problem of inter-business information exchange. These include Electronic Business using eXtensible Markup Language (ebXML) [4] and Universal Description, Discovery and Integration (UDDI) [11]. All of these specifications define common ways to provide a standard method to exchange business messages.

6. RELATED WORK

We are working on expanding the service to cover more types of information sources (semi-structured textual sources including PDF, Doc, XML, etc.) by utilizing more advanced object extraction techniques [6] and conversion tools. Our future work includes extending the framework for Web site and page collection monitoring (similar to [1]). Data mining and machine learning approaches will be incorporated into the framework. We are also exploring and developing tools for better change visualization [2][3]

7. CONCLUSION

The use of a Web-based monitoring system (such as WebCQ) to manage and maintain information exchanges between business partners has considerable value. Further development of the system to ensure accurate results and to assist users in formulating appropriate responses to the changes is needed to continue to increase the value of the monitoring system.

The BizCQ framework proposed in this poster aims at providing an architecture and a set of techniques for information monitoring on the Web or within a corporate space. The change detection system and associated techniques such as difference generation, presentation, and summarization, as well as the automatic response system will give businesses more leverage when coping with constant information changes.

8. REFERENCES

[1] V. Boyapati, K. Chevrier, A. Finkel, N. Glance, T. Pierce, C. Whitmer. ChangeDetector™: A site-Level monitoring tool for the WWW. *In Proceedings of WWW11 Conference*, Honolulu, Hawaii, May 2002.

[2] E. H. Chi, J. Pitkow, J. Mackinlay, P. Pirolli, R. Gossweiler, and S. K. Card. Visualizing the evolution of web ecologies. *In Proceedings of the ACM SIGCHI Conference on human factors in computing systems*, Los Angeles, CA, 1998.

[3] D. B. Crouch, C. J. Crouch, and G. Andreas. The use of cluster hierarchies in hypertext information retrieval. *In Proceedings of Hypertext'89*, Pittsburgh, Pennsylvania, November 1989.

[4] ebXML. http://www.ebxml.org/.

[5] Guideline XML (gXML). http://xml.coverpages.org/gxml.html.

[6] L. Liu, C. Pu, and W. Han. XWRAP: An XML-enabled wrapper construction system for Web information sources, *In Proceedings of the 16th International Conference on Data Engineering* (ICDE), San Diego CA, February 2000.

[7] L. Liu, C. Pu, and W. Tang. WebCQ: detecting and delivering information changes on the web. *In Proceedings of International Conference on Information and Knowledge Management* (CIKM), Washington D.C., November 2000.

[8] L. Liu, W. Tang, D. Buttler, and C. Pu. Information monitoring on the web: a scalable solution. World Wide Web Journal, Kluwer Academic Publishers, 5(4), pp 263-304, 2002.

[9] RSS (RDF Site Summary) 1.0 standard. http://web.resource.org/rss/1.0/.

[10] TrackEngine. http://www.trackengine.com/

[11] Universal Description, Discovery, and Integration (UDDI). http://www.uddi.org