# WEBCAP: A Capacity Planning Tool for Web Resource Management

Sami Habib and Maytham Safar
Kuwait University
Computer Engineering Department
P. O. Box 5969 Safat 13060 Kuwait
{shabib, maytham}@eng.kuniv.edu.kw

## ABSTRACT

A staggering number of multimedia applications are being introduced every day. Yet, the inordinate delays encountered in retrieving multimedia documents make it difficult to use the Web for real-time applications such as educational broadcasting, video conferencing, and multimedia streaming. The problem of delivering multimedia documents *in time* while placing the least demands on the client, network and server resources is a challenging optimization problem. The WEBCAP is ongoing project that explores applying capacity planning techniques to manage or tune the Web resources (client, network, server) for optimal or near optimal performance, subject to minimizing the retrieval cost while satisfying the real-time constraints and available resources. The WEBCAP project consists of four software modules: object extractor, object representer, object scheduler, and system tuner. The four modules are connected serially with 3 feedback-loops. In this paper, we focus on how to extract objects from multimedia document and how to represent them as object and operation flow graphs while maintaining precedence relations among the objects.

### Categories and Subject Descriptors

H.3.4 [Systems and Software]: Performance Evaluation.

**General Terms:** Algorithms, Design, Performance.

**Keywords**: Multimedia, Scheduling, Capacity-Planning.

## 1. INTRODUCTION

The Internet (World Wide Web) has been experiencing a phenomenal growth in terms of increasing numbers of servers, and increasing numbers of users that want to access the huge amount of information distributed at vast geographical sites. The main characteristics of the WWW are: distributed information system, heterogeneity, frequent changes, large in size, and non-uniformity of information access.

The software tools that are automatically managing and tuning (design or redesign) Web resources are lacking behind due to the complexity and diversity of the problem. However, WEBCAP uses well-known techniques, which are based on the approaches of a capacity planning tool. Capacity planning has been established for managing and tuning resources in operation research for decades; moreover, it views as a heuristic process aimed at satisfying predicted client needs through the use of predicted technological changes in the most cost-effective and timely manner [4].


*WWW 2005*, May 10-14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

WEBCAP project involves in multidisciplinary subjects such as object representation, object retrieval/caching, operation scheduling, and web resource tuning. Both authors have intensive background in these areas, especially object caching [8][9], and object scheduling and system tuning [5][6]. To our knowledge, most of the published related work is limited to object scheduling without considering the operation scheduling within Web system [3][10]. In this paper, we introduce WEBCAP as a capacity planning tool that can ***guide*** the distributed systems designer/manager to schedule multimedia objects over distributed resources. Also, we focus on a multimedia object extraction and representation.

## 2. AN OVERVIEW OF WEBCAP PROJECT

The WEBCAP consists of four modules as shown in Figure 1. The object extractor (OE) is an automating process of identifying the precedence relations among the objects inside Web document of different contents. For example, with static web page content, some use of *regular expressions* (e.g., XPath expressions), special *mark up tags*, or *HTTP headers* could be required to provide minimal hints that helps to decide the precedence relations between the objects. While with dynamically generated Web page content, RSS libraries can help in deciding the precedence relations between the objects in the Web page. Finally, with RSS or XML feed content, the process of identifying the precedence relations among the objects in a feed can be automated by adding new RSS Modules [7], where namespaces are used to describe a space for our own extensions.

The object representer (OR) is a 2-level process, where in first level a multimedia document is modeled as an *object flow graph* (OFG). The node in OFG represents a multimedia object such as a text, still image, audio, or video. The edge in OFG represents the precedence relations among objects based on Allen's temporal properties [1]. In the second level, the object flow graph is mapped into an operating flow graph (OPFG) including all operations (fetch, transmit, decode, and render) that are needed to fetch an object from the server, transmit it through the network, decode and render it at the client side.

The object scheduler (OS) takes an OPFG as an input and determines the optimal starting and finishing times for all operations, subject to the available Web resources while maintaining all precedence relations among the objects.

If the object scheduler fails to satisfy the precedence relations, then the system tuner (ST) is invoked to either re-manage system resources (Fig. 1, loop 1), reconfigure document representation (loop 2) or relax precedence relations (loop 3).

## 3. ASIAN TSUNAMI DISASTER EXAMPLE

To illustrate the object extraction and representation, we use a news clip that shows the impact of the earthquake on Malaysia,

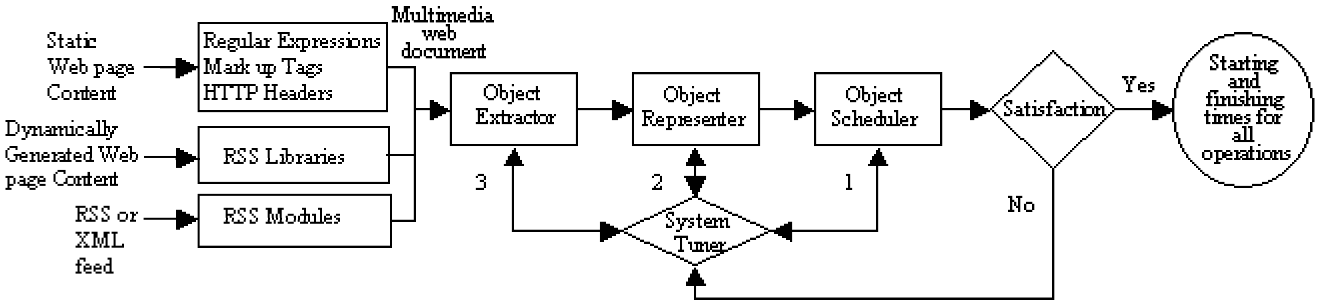


Figure 1. An overview of the WEBCAP project.

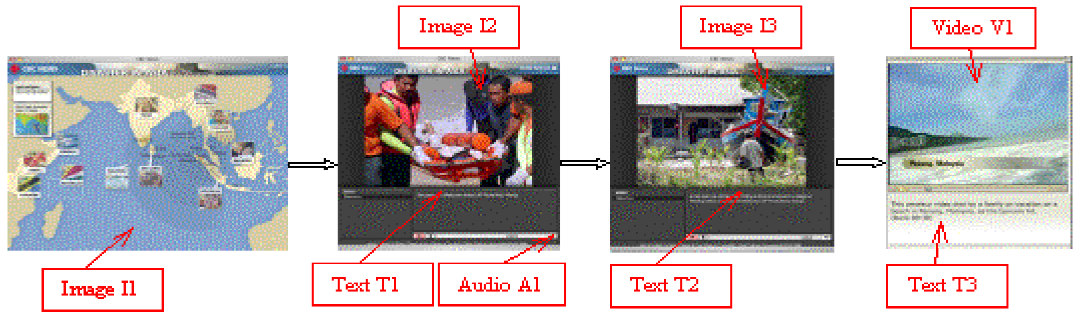


Figure 2. The Asian Tsunami disaster example.

where dozens of people and properties were swept away from beaches on the island of Penang [2]. Figure 2 shows a web-page that consists of 8 media objects (such as still image, text, audio, and video) with precedence relations among the objects. Image 1 displays the location of the disaster area in Malaysia. Then, Image 1 is replaced by Image 2 showing rescuers on a Malaysian beach and a textual description of it (Text 1). At the same time an audio is played (Audio 1) briefing news about the disaster. Image 3 displays a man looking at a fishing boat sitting in front of a house in a village on Penang Island. Note that the text description (Text 2) is not displayed until Image 3 is completely displayed. This is to make sure that Text 2 is not displayed with the wrong image (Image 2). Once Text 2 is displayed and Audio 1 is played completely, a new text (Text 3) is displayed and a video is played that shows an amateur video shot by a family on vacation on a beach in Penang as the tsunami hit. Figure 3 illustrates the object flow graph of the Asian Tsunami disaster and the operation flow graph for one object (T1). The executed operations on each object can differ in their numbers and execution times depending on the object type and the available resources.

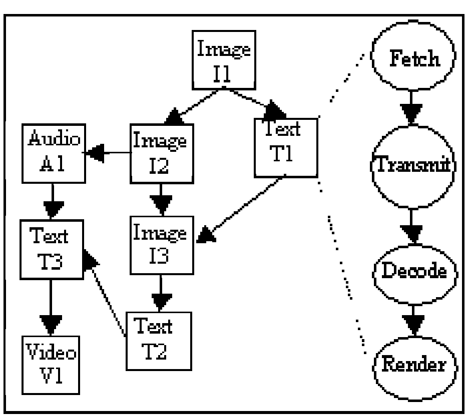


Figure 3. An object flow graph with a sample of operations.

## 4. CONCLUSION AND FUTURE WORK

We have introduced an ongoing project called WEBCAP that ultimately manages (reconfigure or redesign) Web resources for real time multimedia document retrieval while satisfying all users and systems constraints. We are working on unifying and automating the extraction and representation of precedence relations among objects contained in distributed documents with different formats. This work is supported by Kuwait University, Research Grant No. EO 05/04.

## REFERENCES


[1] J. Allen, "Maintaining Knowledge about Temporal Intervals," Communications of the ACM, vol. 26, no. 11, 1983.

[2] CBC News, http://www.cbc.ca

[3] J. Courtiat, C. Santos, C. Lohr, and B. Outtaj, "Experience with RT-LOTOS, a Temporal Extension of the LOTOS Formal Description Technique," Computer Communications, Vol. 23, No. 12, p. 1104-1123, 2000.

[4] V. Gupta, "What is Network Planning?" IEEE Communications Magazine, vol. 23, no. 10, October 1985, pp. 10-16.

[5] S. Habib, C. P. Ravikumar and A. C. Parker, "Storage Allocation and Scheduling Problems in Web Caching Applications," NLANR Web Cache Workshop, Boulder, Co.,USA, 1997.

[6] S. Habib, "Synthesis and Optimization of Application-Specific Intranets," Ph.D. Dissertation, University of Southern California, LA, CA, USA, 2001.

[7] IBM Developer Works. http://www-106.ibm.com/developerworks/

[8] M. Safar, "Efficient MBC-based shape retrieval and spatial querying," Ph.D. Dissertation, University of Southern California, LA, CA, USA, 2000.

[9] M. Safar, "Classification of Web Caching Systems," In the Proceedings of International Conference on WWW/Internet. Lisbon, Portugal, 2002.

[10] J. Xu, W. Lee, and J. Liu, "Scheduling Web Requests in Broadcast Environments," The Thirteenth International World Wide Web Conference, 2004.