TRANSCODING INTERNET CONTENT FOR HETEROGENEOUS CLIENT DEVICES

John R. Smith, Rakesh Mohan and Chung-Sheng Li

IBM T.J. Watson Research Center,
30 Saw Mill River Road, Hawthorne, NY 10532
{jrsmith,rakesh,csli}@watson.ibm.com

ABSTRACT

There is a growing diversity of client devices that have access to the Internet. However, much of the content on the Internet cannot be handled by the devices that have limited communication, processing, storage and display capabilities. In order to improve the utility of a wide range of client devices, we propose a network-based solution for transcoding Internet content. The system uses an InfoPyramid for representing and transcoding video, images, audio and text. The InfoPyramid manipulates the content along the dimensions of fidelity and modality, and aggregates the methods for content analysis, translation, filtering and selection. The InfoPyramid utilizes a policy engine, which incorporates user and publisher preferences, various transcoding policies, device descriptions, and real-time network constraints in order to adapt the Internet content to the client devices.

1. INTRODUCTION

Many new devices, such as personal digital assistants (PDAs), hand-held computers, Internet-ready phones, and WebTVs, are gaining access to the Internet. The capabilities of these devices to receive, process, store and display Internet content varies widely. Given the variety of client devices, it is difficult for Internet content publishers to tailor content to the individual devices.

Internet content publishers do not have many options for customizing the content. In some cases, the publishers manually generate secondary, text-only versions of Web pages that the users can select instead of the originals. Other mechanisms within the HTTP protocol allow the client to specify some client attributes, such as the preferred language of the user, or the image, video, and audio formats supported by the client device. Using this information, the content server can automatically select and deliver content that is compatible with the client device and the user.

In the case of client devices with minimal capabilities, such as pagers and cell phones, special markup languages are being developed, such as HDML ([1]). However, mechanisms still need to be developed to automatically convert Internet content into these formats. The emerging XML markup language may improve the capabilities for adapting content since separate style sheets can be developed for client devices that determine how the content is displayed [2].

1.1. Related work

Recently, several systems have been developed for adapting Internet content to client devices. Fox, et al., developed a system for distilling, or compressing images that pass through an Internet proxy [3]. Other commercial systems such as Intel's Quick Web [4] and Spyglass' Prism [5] similarly compress the images that pass through the Internet service provider to speed-up download time.

We have developed a content-based system for transcoding images [6]. This system retrieves and analyzes images in the Internet and classifies them into image type and purpose classes in order to select appropriate methods for transcoding them. Since the system is limited to transcoding images, we propose a more powerful solution for Internet content that includes video, images, text and audio. The system transcodes the content along the dimensions of fidelity and modality in order to better adapt it to a large variety of client devices.

1.2. Outline

In this paper, we propose a system for transcoding Internet content. In Section 2, we present an overview of the Internet content transcoding system. In Section 3, we present the InfoPyramid framework for representing and manipulating video, images, audio and text. In Section 4, we describe processes for selecting the versions of the content to adapt the content to the clients according to capabilities and preferences. Finally, in Section 5, we describe the deployment of the transcoding system in the Internet as transcoding proxies.

2. INTERNET CONTENT TRANSCODER

Figure 1 illustrates the proposed Internet content transcoding system. The system retrieves and analyzes the Internet content and ingests it into the InfoPyramid format. A policy engine gathers the capabilities of the client, the network conditions and the transcoding preferences of the user and publisher. This information is used to define the transcoding options for the client. The system then selects the output versions of the content and uses a library of content analysis, filtering, translation and manipulation routines to generate the content to be delivered to the client device.

The Internet content transcoding system may be deployed at the server, proxy or client. Deployed at a proxy, the system is able to retrieve the Internet content, analyze

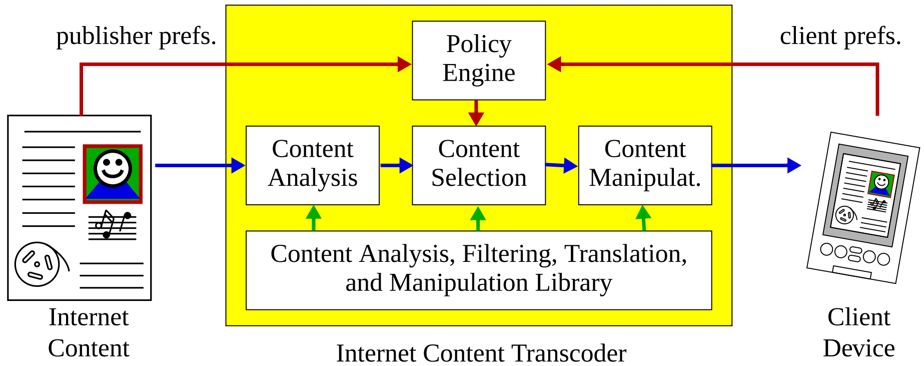

Figure 1: The Internet content transcoding system adapts the content to the capabilities of the client devices.

and transcode it, and deliver the results to the client, on-the-fly. Deployed at the server, the system can be used in the content publication process. The system can pre-materialize the alternate versions of the Internet content and store them at the server. In this case, the system merely selects the versions of the content to deliver to the client. In some cases, the transcoding system can be deployed at the client to customize the content display, such as according the user preferences, as long as the client has sufficient capabilities.

3. INFOPYRAMID FRAMEWORK

The InfoPyramid provides a general framework for handling the Internet content. It allows specialized methods to be plugged-in for analyzing, filtering, translating and manipulating the Internet content. As depicted in Figure 2, the InfoPyramid develops a conceptually redundant representation of the Internet content that aggregates multiple versions of the content along the dimensions of modality (video, image, text, and audio) and fidelity (which includes summarized and compressed versions) [7]. The translation and summarization methods generate the alternate versions of the content as needed.

3.1. Translation and summarization

The translation methods convert the content between modalities, such as, text to audio, or video to images. On the other hand, the summarization methods generate versions within the same modality, but with different fidelity. For example, the summarization methods compress the images, summarize text, and extract and re-animate the key-frames from video. The translation and summarization methods can be cascaded to change both the modality and fidelity of the content. In this case, a video can be converted to a small, reduced-color image.

For each of the modalities, we describe some of the summarization and translation methods that could be used to change the fidelity and modality of the content, respectively.

1. Images:
   - Fidelity – Spatial size reduction, color depth reduction, lossy compression [6].
   - Modality – Images to text: related text, embedded text, semantic labels [8].
2. Video:
   - Fidelity – Spatial size, temporal size, playback rate, bit-rate.
   - Modality – Video to images: key-frames; Video to audio: sound track; Video to text: closed captions, embedded text.
3. Text:
   - Fidelity – Key-term extraction, text summarization, document headings extraction.
   - Modality – Text to audio: speech synthesis, Text to text: language translation (i.e., English to French).
4. Audio:
   - Fidelity – Bit-rate reduction, stereo to mono.
   - Modality – Audio to text: speech recognition.

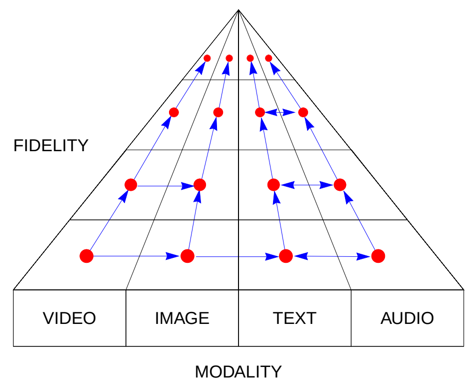

Figure 2: The InfoPyramid aggregates multiple representations of the content along the dimensions of fidelity and modality and unifies the methods for manipulating the content.

In order to evaluate content alternatives, the system could assign content value scores to each of the content alternatives. Using the content value scores, the system is able to optimize the selection of the content according to the device capabilities, preferences and network conditions.

4. TRANSCODING SYSTEM

We propose that the transcoding system utilizes a policy engine to evaluate the alternatives for adapting the content to the client devices.

4.1. Policy Engine

The policy engine would gather the capabilities of the client and the transcoding preferences of the user and publisher, and sense the network conditions to define the transcoding options for the client. In order adapt the Internet content to these devices, the transcoding proxy generates and selects versions of the content according to the policies, network and device constraints, and preferences.

4.1.1. Content value scores

The InfoPyramid system provides the mechanism for assigning content value scores to the alternate versions of the content. In some cases, the content value scores are derived automatically by measuring the loss in fidelity that results from translating or summarizing the content. For example, the content value scores can be linked to the distortion introduced from compressing the images or audio. Otherwise, the content value scores can be tied directly to the methods that manipulate the content.

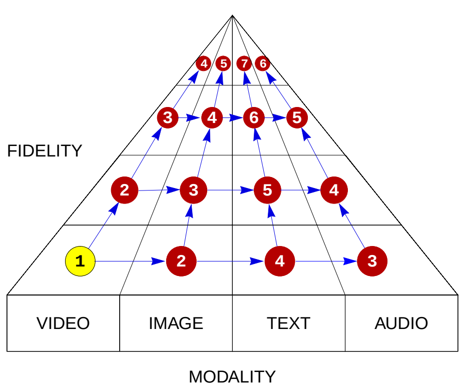

Figure 3: Example Internet content value scores assigned for video.

Figure 3 illustrates examples of content value scores that can be assigned for transcoding video. In this example, the original video has the highest content value. Each manipulation of the video along the dimension of fidelity or modality alters the content value. For example, converting the video to a sequence of images results in a small reduction in content value. Converting the video to a highly compressed audio track produces a higher reduction in the content value.

4.1.2. User and publisher preferences

The content value scores comprise only part of the information that can be used in the content selection process. Both the publisher and user may have preferences for how the content is transcoded. Figure 4 illustrates some example preferences for transcoding images. In this example, the preferred versions for the image content are the low-fidelity versions of the image, and the translations to text.

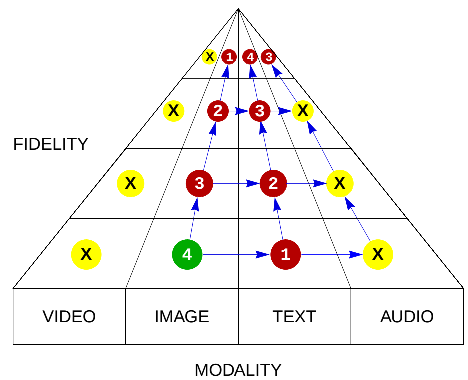

Figure 4: Internet content format preferences may be established by the publisher or user. This example illustrates the preferred content alternatives for images. Device constraints may further score or eliminate some content alternatives.

4.1.3. Device constraints

Various constraints of the devices affect the selection of the content. For example, many devices cannot handle video. In this case, the corresponding content alternatives can be eliminated (see Figure 4). Overall, the display, storage and processing capabilities of the client devices eliminate the selection of individual versions of the content. These also place constraints on the set of selections for a Web document. For example, if the device has a local storage of only 16 KB, then this places a hard limit on the total size of the versions of the content selected.

4.1.4. Network constraints

Similarly, the network would place constraints on the content selection process. In general, the network constraints introduce a trade-off between content data size and load time. For example, if the user specifies a maximum load time for a page, then to accommodate this load time, the transcoder system must sense the end-to-end bandwidth to derive the target data size for the content. Then, the system can select the content of maximum content value that is within the target data size.

4.2. Content Selection

We investigate more closely how the system could optimize the selection of the content alternatives. Consider the Web document with two objects, A and B, depicted in Figure 5. Let $A_{ij} = T_{ij}(A)$ and $B_{kl} = T_{kl}(B)$ transcode A and B,

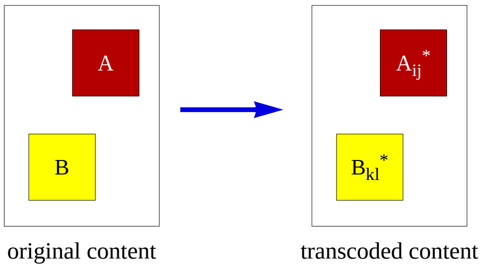

Figure 5: Example of content selection for a Web document consisting of two objects, A and B.

respectively, to target modalities i and k, and fidelities j and l. Let $V(A_{ij})$ and $V(B_{kl})$ be the content value scores for A_{ij} and B_{kl}, and let $S(A_{ij})$ and $S(B_{kl})$ be the data sizes of A_{ij} and B_{kl}. Finally, let S_T be the maximum data size for the content, which may have been derived from the user's specified maximum load-time and the network conditions.

4.2.1. Maximum content value

The content selection process selects A_{ij}^* and B_{kl}^* with maximum content value for a target data size as follows:

$$\begin{aligned} V(A_{ij}^*) + V(B_{kl}^*) &= \max_{i,j,k,l}(V(A_{ij}) + V(B_{kl})), \text{ and} \\ S(A_{ij}^*) + S(B_{kl}^*) &\leq S_T. \end{aligned} \quad (1)$$

4.2.2. Minimum load time

Given a minimum acceptable content value V_T, the content alternatives A_{ij}^* and B_{kl}^* of minimum data size are given by:

$$\begin{aligned} S(A_{ij}^*) + S(B_{kl}^*) &= \min_{i,j,k,l}(S(A_{ij}) + S(B_{kl})), \text{ and} \\ V(A_{ij}^*) + V(B_{kl}^*) &\geq V_T. \end{aligned} \quad (2)$$

4.2.3. Device constraints and preferences

By extending this optimization process, the content selection system could incorporate the user ($U(A_{ij}^k)$) and publisher ($P(A_{ij}^k)$) preferences, and client device constraints ($D(A_{ij}^k)$), to best adapt the content. In this case, the total score of each item k in the Web document is given by

$$Z(A_{ij}^k) = \omega_v V_{ij}^k + \omega_u U_{ij}^k + \omega_p P_{ij}^k + \omega_d D_{ij}^k,$$

where ω_v, ω_u, ω_p, and ω_d are weighting factors assigned by the user or system.

5. REAL-TIME, NETWORK-BASED TRANSCODING

The transcoding system can be implemented in the network in the form of a transcoding proxy (TP) to perform transcoding on-the-fly, see Figure 6. The transcoding proxy system is designed to have a high bandwidth connection to the content provider. In most cases, the proxy has a low bandwidth connection to the client. As a result, reducing the amount data through compression and transcoding at the proxy results in faster delivery, even when accounting for the added time for content analysis, selection and transcoding [6].

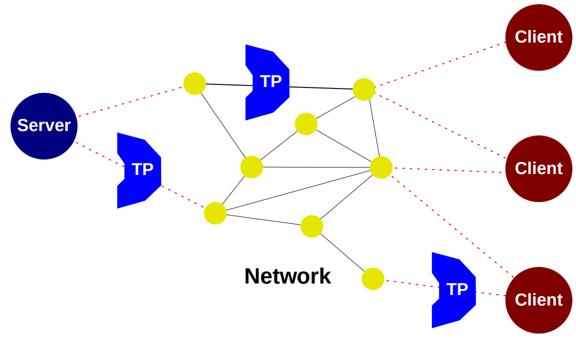

Figure 6: Network-based transcoding of Internet content using transcoding proxies (TP).

6. SUMMARY

We proposed a system for the network-based transcoding of Internet content in order to improve the accessibility of a wide range of client devices to the content on the Internet. The transcoding system retrieves, analyzes, and ingests the Internet content into an InfoPyramid representation. The InfoPyramid provides a conceptual framework for manipulating the content along the dimensions of modality and fidelity. The transcoding system selects the content from the InfoPyramid by assessing the various content alternatives to adapt the Internet content to the client devices. In this way, a wide range of client devices gain access to the large amounts of content on the Internet that is otherwise beyond their capabilities.

7. REFERENCES

[1] Unwired Planet. Handheld device markup language specification. Technical Report Version 2.0, Unwired Planet, Inc., April 1997.

[2] C. M. Sperberg-McQueen T. Bray, J. Paoli. Extensible markup language (XML), W3C recommendation. Technical Report Version 1.0, W3C, February 1998.

[3] A. Fox, S. D. Gribble, E. A. Brewer, and E. Amir. Adapting to network and client variability via on-demand dynamic distillation. In ASPLOS-VII, Cambridge, MA, October 1996.

[4] Intel Quick Web. http://www.intel.com/quickweb.

[5] Spyglass-Prism. http://www.spyglass/products/prism.

[6] J. R. Smith, R. Mohan, and C.-S. Li. Content-based transcoding of images in the Internet. In IEEE Proc. Int. Conf. Image Processing (ICIP), Chicago, Il, October 1998. Submitted.

[7] C.-S. Li, R. Mohan, and J. R. Smith. Multimedia content description in the InfoPyramid. In IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing (ICASSP), Seattle, WA, May 1998. Special session on Signal Processing in Modern Multimedia Standards.

[8] J. R. Smith and S.-F. Chang. Visually searching the Web for content. IEEE Multimedia Mag., 4(3):12 – 20, July–September 1997.