Chapter 7

Continuous Media Web:
Hyperlinking, Search and Retrieval of Time-Continuous Data on the Web

Silvia Pfeiffer, CSIRO ICT Centre, Australia

Conrad Parker, CSIRO ICT Centre, Australia

André Pang, CSIRO ICT Centre, Australia

ABSTRACT

The Continuous Media Web project has developed a technology to extend the Web to time-continuously sampled data enabling seamless searching and surfing with existing Web tools. This chapter discusses requirements for such an extension of the Web, contrasts existing technologies and presents the Annodex technology, which enables the creation of Webs of audio and video documents. To encourage uptake, the specifications of the Annodex technology have been submitted to the IETF for standardisation and open source software is made available freely. The Annodex technology permits an integrated means of searching, surfing, and managing a World Wide Web of textual and media resources.

INTRODUCTION

Nowadays, the main source of information is the World Wide Web. Its HTTP (Fielding et al., 1999), HTML (World Wide Web Consortium, 1999B), and URI (Berners-Lee et al., 1998) standards have enabled a scalable, networked repository of any sort of

information that people care to publish in textual form. Web search engines have enabled humanity to search for any information on any public Web server around the world. URI hyperlinks in HTML documents have enabled surfing to related information, giving the Web its full power. Repositories of information within organisations are also building on these standards for much of their internal and external information dissemination.

While Web searching and surfing has become a natural way of interacting with textual information to access their semantic content, no such thing is possible with media. Media on the Web is cumbersome to use: it is handled as dark matter that cannot be searched through Web search engines, and once a media document is accessed, only linear viewing is possible — no browsing or surfing to other semantically related documents.

Multimedia research of the recent years has realised this issue. One means to enable search on media documents is to automate the extraction of content, store the content as index information, and provide search facilities through that index information. This has led to extensive research on the automated extraction of metadata from binary media data, aiming at bridging the semantic gap between automatically extracted low level image, video, and audio features, and the high level of semantics that humans perceive when viewing such material (see, e.g., Dimitrova et al., 2002).

It is now possible to create and store a large amount of metadata and semantic content from media documents – be that automatically or manually. But how do we exploit such a massive amount of information in a standard way? What framework can we build to satisfy the human need to search for content in media, to quickly find and access it for reviewing, and to manage and reuse it in an efficient way?

As the Web is the most commonly used means for information access, we decided to develop a technology for time-continuous documents that enables their seamless integration into the Web's searching and surfing. Our research is thus extending the World Wide Web with its familiar information access infrastructure to time-continuous media such as audio and video, creating a "Continuous Media Web".

Particular aims of our research are:

- to enable the retrieval of relevant clips of time-continuous documents through familiar textual queries in Web search engines,
- to enable the direct addressing of relevant clips of time-continuous documents through familiar URI hyperlinks,
- to enable hyperlinking to other relevant and related Web resources while reviewing a time-continuous document, and
- to enable automated reuse of clips of time-continuous documents.

This chapter presents our developed **Annodex** (annotation and indexing) technology, the specifications of which have been published at the IETF (Internet Engineering Task Force) as Internet-Drafts for the purposes of international standardisation. Implementations of the technology are available at http://www.annodex.net/. In the next section we present related works and their shortcomings with respect to our aims. We then explain the main principles that our research and development work adheres. The subsequent section provides a technical description of the Continuous Media Web (CMWeb) project and thus forms the heart of this book chapter. We round it off with a

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/continuous-media-web/25972

Related Content

Test Zone Search Optimization Using Cuckoo Search Algorithm for VVC
Suvojit Acharjeeand Sheli Sinha Chaudhuri (2022). *International Journal of Multimedia Data Engineering and Management (pp. 1-16).*
www.irma-international.org/article/test-zone-search-optimization-using-cuckoo-search-algorithm-for-vvc/314574

Creating Global Classrooms Using Universal Design for Learning
Victoria Brown (2018). *Digital Multimedia: Concepts, Methodologies, Tools, and Applications (pp. 145-168).*
www.irma-international.org/chapter/creating-global-classrooms-using-universal-design-for-learning/189472

Learning by Doing via Game Making
Jae Yeob Jungand Hyung Sung Park (2011). *Gaming and Simulations: Concepts, Methodologies, Tools and Applications (pp. 1354-1366).*
www.irma-international.org/chapter/learning-doing-via-game-making/49453

Credibility Analysis for Online Product Reviews
Min Chenand Anusha Prabakaran (2018). *International Journal of Multimedia Data Engineering and Management (pp. 37-54).*
www.irma-international.org/article/credibility-analysis-for-online-product-reviews/220431

Network-Based Information System Model for Research
Jo-Mae B. Maris (2005). *Encyclopedia of Multimedia Technology and Networking (pp. 756-761).*
www.irma-international.org/chapter/network-based-information-system-model/17325