VideoAnywhere: A System for Searching and Managing Distributed Heterogeneous Video Assets[1]

Amit Sheth, Clemens Bertram and Kshitij Shah

Large Scale Distributed Information Systems Lab
Department of Computer Science, University of Georgia
415 Graduate Studies Research Center
Athens, Georgia 30602-7404 USA
amit@cs.uga.edu, http://lsdis.cs.uga.edu

Abstract

Visual information, especially videos, plays an increasing role in our society for both work and entertainment as more sources become available to the user. Set-top boxes are poised to give home users access to videos that come not only from TV channels and personal recordings, but also from the Internet in the form of downloaded and streaming videos of various types. Current approaches such as Electronic Program Guides and video search engines search for video assets of one type or from one source. The capability to conveniently search through many types of video assets from a large number of video sources with easy-to-use user profiles cannot be found anywhere yet. *VideoAnywhere* has developed such a capability in the form of an extensible architecture as well as a specific implementation using the latest in Internet programming (Java, agents, XML, etc.) and applicable standards. It automatically extracts and manages an extensible set of metadata of major types of videos that can be queried using either attribute-based or keyword-based search. It also provides user profiling that can be combined with the query processing for filtering. A user-friendly interface provides management of all system functions and capabilities. *VideoAnywhere* can also be used as a video search engine for the Web, and a servlet-based version has also been implemented.

1 Introduction

Current technology allows the integration of various home entertainment devices such as the TV, VCR and computers into an integrated system. TV tuner cards for the PC enable the user to tune into his or her favorite TV programs on the computer, and many Web sites offer downloadable or streaming videos that can be played back with a minimum of additional software requirements.

On the other hand, there are many people who are unfamiliar with a PC environment or are more comfortable with the TV when it comes to watching the latest news or episodes of "South Park", using the remote control as their only input device to switch between channels and home videos. Also, the notion of a "movie night" is still tightly coupled with the TV screen and the VCR; not many people would consider spending an evening with the family in front of a computer, and that not only because of the difference in the screen size.

The advent of set-top boxes has provided the home user with easy access to the vast resources of the Internet. It is now possible to combine all three sources of videos – TV, home recordings, and the World Wide Web (WWW) – in one system that home users of almost every age are familiar and comfortable with.

Dealing with such a huge number of movies, news clips, shows and other types of videos is a non-trivial task that requires an efficient media-asset management. *VideoAnywhere* is such a manage-

[1] This research work was partially funded by LG Electronics Research Center of America, Princeton Junction, New Jersey, U.S.A. (http://www.lgerca.com) under the UGA Proposal RR81764F.

ment system (Figure 1). It provides the needed system, syntactic and structural interoperability between the systems that support information (specifically video asset) providers and information consumers. A sufficient amount of content-independent, content-based and content-descriptive metadata of all videos that are known to the system are stored in a metabase so that the desired videos can efficiently be searched and found within little time.

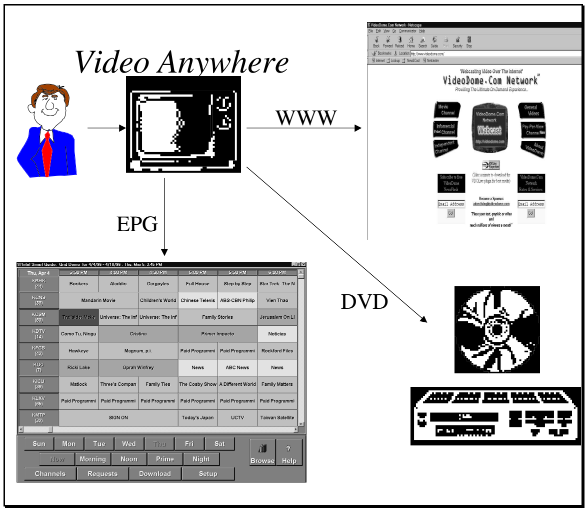

Figure 1: VideoAnywhere: Connected to all major asset sources

The current video search engines are more or less restricted to one source – that is TV, home video, or the WWW – or one type of video. That has in many cases the advantage of being able to provide more advanced search and retrieval options to the user who profits from highly specialized and user-friendly interfaces when searching for interesting assets.

In contrast, the more general search engines and libraries that aim at indexing assets from many different sources have to deal with the problem of multiple domains. Christel points out that "there may not be enough domain knowledge to apply to the user's query and to the library index in order to return only a very small subset of the library to the user matching just the given query." [4]

For instance, scour.net allows keyword-search for many different media types, but does not support attribute-searches that make MCA's interface so useful. A lot of research has been done on content-based video retrieval, similarity queries, and the like. *VideoAnywhere* is aimed at combining many positive aspects of the existing approaches with the usefulness of personalized video content into one user-friendly system. It demonstrates the use of innovative Internet technologies such as XML and agents to build a comprehensive and useful tool for the future home-based video management that

- is designed to access video assets of different types from all possible sources,
- offers a combination of attribute and keyword searches,
- provides a convenient user interface that is oriented towards a typical home user environment, and
- provides a user profile management that helps the user to efficiently exploit the available resources.

In the process, *VideoAnywhere* also uses an extensible architecture that addresses some of the issues related to system, syntactic, structural and semantic interoperability [22] using the latest technology and standards. It combines the capabilities of three relevant areas:

- video search techniques: these include Electronic Program Guides, video data management and search systems from Excalibur and Virage, and digital libraries for video, such as Informedia,
- metadata for video: these includes several research projects (e.g., see [23] for two examples), the "Program and System Information Protocol for Terrestrial Broadcast and Cable" (A/65) for TVs, as well as MPEG-4 and MPEG-7,
- user profiling ([13]): several examples of viewer interest profiles can be found on the Internet ([14]).

As part of the longer-term InfoQuilt project (*http://lsdis.cs.uga.edu/proj/proj.html*) it plays a major role in a system that not only integrates keyword, attribute, and content-based search, but also supports semantic correlation of digital heterogeneous assets.

It is worth noting that Virage's new video search engine [32] can complement *VideoAnywhere:* Virage does not support the efficient attribute search, but is able to segment videos and extract annotations of videos that can be used for fulltext search queries. Segments play an important role not only in montages but also when browsing a video by displaying only the keyframes and annotations of those segments.

2 System Architecture

VideoAnywhere is completely written in Java and extensively uses XML. The system supports users that are connected to all three major video sources, although only first or first two are relevant in its use as a search engine of video assets on the Web:

- the World Wide Web – through Internet connection
- TV programs – through the TV and/or downloaded EPGs
- home-made or recorded videos – through a DVD player and/or a VCR

For the purpose of providing metadata of videos or any other kind of media, using XML is one way of ensuring syntactic and structural interoperability. We expect that in the future many content providers will switch to this way of providing information about their assets.

In the *VideoAnywhere* system most of the configuration files are written in XML. These files are parsed using IBM's "XML for Java" API [35] which implements all important aspects of XML that must be dealt with according to the W3C organization.

Syntactic interoperability does not necessarily guarantee *semantic* interoperability, however. This problem can be addressed by providing ontological mappings, which – in many cases – will be supplied by the users themselves. XML namespaces facilitate automating the creation of such mappings. For the lightweight, persistent metabase we are currently using the object-oriented database POET 5.1 [16].

2.1 The Video Content Agent

The Video Content Agent is the heart of *VideoAnywhere*. Figure 2 visualizes its role in the whole architecture. It maintains the interaction between the metabase and the various video sources and keeps the metabase up-to-date. In order to do so, it needs to perform three basic tasks at the same time:

1. New assets have to be added to it on a constant basis.
2. Existing links to the various Web based assets have to be validated.
3. Expired assets and their occurrences in the montages have to be removed from the metabase.

In addition, it runs a daily query against the metabase to get the ten most interesting assets from the metabase according to the user profile.

To find and add new assets, the Video Content Agent has to go to the known EPGs and Web sites to collect the metadata that describe the available assets. However, a number of issues and challenges arise in the gathering of metadata for the assets:

- Each video provider, EPG's as well as Web sites, structures the metadata in a different way.
- The amount of provided metadata varies widely.
- Different sites have different naming conventions.
- The structure of a particular Web site may change over time.
- New providers come on the market with interesting video offers.

VideoAnywhere handles these issues in the following ways.

For each provider of metadata (EPG's, Web sites, home videos) exists a specialized; for instance, Web based extractors examine the Web pages that belong to a particular Web site. We can speed up the effort of developing or modifying an extractor using an Extractor Development toolkit. More importantly in near future, use of XML can obviate the need for extractors for all syntactic and structural issues. A conversion process, the "Extractor Ontology Mapper", maps the XML tags used at the asset source to the tags that are used in our system.

The metadata that the extractors have found is returned in an XML stream to the Video Content Agent, which creates a new asset and fills in the attribute values according to the XML elements read from the stream. In our current prototype, it is the Video Content Agent who asks the extractors whether new metadata is available. Instead of this "Pull-technology" it is also possible to deploy a "Push-technology", where extractor agents would notify the Video Content Agent every time new information is there. That would require the extractor agents to be autonomous and check the Web site on their own for new metadata. The Video Content Agent would in this scenario simply wait for messages to come in. We are focusing on the former solution in this paper. However, both alternatives have up- and downsides depending on whether there are many home-based metabases and Video Content Agents or only a few metabases supporting search engines on Web portals such as Yahoo, Netcenter, or Excite. We are considering a multi-agent system deploying a rich agent communication language for the next version of *VideoAnywhere* to investigate these issues further. In this paper, we do not describe the other system components for brevity.

Although we currently support a comprehensive query capability involving keyword and attribute-based components, it can be extended to support content-based search by adopting our Black Box Approach using a third party visual information retrieval engine [21]. Additionally, the system also supports asset storage management (which can be used to cache the results of a query or profile based filtering) and montages.

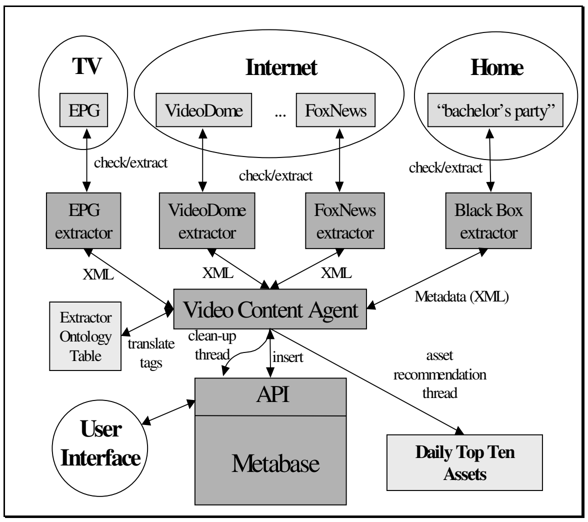

Figure 2: Architecture

Often the user will want to watch parts of different movies that share a common thread, for instance, all touchdown scenes of the Georgia Bulldogs of the last season, or clips of birthdays of one particular person through the years. Rather than merging whole assets into one (very) long asset, it makes sense to compile a so-called montage out of segments of various assets. Let us assume that a collection of Bulldog games has been segmented and annotated with the transcript of the reporters' conversations. All it takes to create an impressive "touchdown show" is to search the annotations for "touchdown", browse through the query results and add all relevant segments to a new montage. In the MontageManager, the segments can be re-arranged if desired, touchdown scenes of the opponents can be found and removed, and the Bulldog fan has an asset he will be proud to present to anyone or watch himself.

3 Future Developments

Two keys extensions of the *VideoAnywhere* system that we expect to complete in the near future are to support the extended information on TV (the tags V, S, L, D, FV which give more detail about the program) and automated semantic tag mapping in the Video Content Agent.

Another extension of the system we consider for the short-term future is a content-based video browsing and searching feature that allows the user to intelligently skim through a (possibly long) video to determine its relevance to the user. To accomplish this we will need software (such as Virage) that segments and indexes the video, and adapt our black-box approach to use third-party indexing and content processing technologies as described in [21]. Video segmentation supports browsing and building montages. Indexing a video allows for content-based queries in addition to already existing keyword and attribute queries. In content-based queries users search for specific patterns, structures, textures, or colors in an image or video. Once MPEG-7 becomes widely accepted it will be possible to look for objects in a video in a much more advanced way.

Finally, in the related project of InfoQuilt, we are also working on a system that deals with a broad variety of digital media and information content (text, traditional databases, images, video, etc.). It utilizes a multi-agent system for intelligent retrieval and management of distributed heterogeneous information. While *VideoAnywhere* manages "only" video data, InfoQuilt abstracts from the level of media types and takes searching one step further, enabling the Internet user to look for interesting data *at a semantic level*. For that purpose the concept of MREF (Metadata REFerence link) has been introduced to support media-independent, sematic-based information correlation [20]. The basic idea of an MREF is to specify a query in a database-independent way. The attribute-search uses a number of ontologies that describe and correlate the entities of interest. Descriptions of those entities constitute the terms of the attribute query (see example in the Appendix). The complete MREF consisting of a combination of keyword, attribute, and content-based search is sent by a local User Agent to a broker, which forwards it to known Resource Agents. These agents are wrappers around a metabase like the one used by *VideoAnywhere*, but holding references to possibly multiple types of digital media. They convert the MREF into a local database query, run the query, and send the results back to the broker and the User Agent, which then displays them in a suitable way.

The work described in this paper, however, constitutes the initial video related capabilities in the longer-term InfoQuilt project (for more information, visit *http://lsdis.cs.uga.edu/proj/proj.html*).

4 Conclusion

It is conceivable to use just the Internet part of *VideoAnywhere* and deploy it as a Video Search Engine for the World Wide Web (for example, see the use of the Junglee engine in the shopping option at Yahoo!). In order for such an engine to be successful, as many content providers as possible have to provide the necessary extractors and register with the engine, or we have to develop an extractor toolkit to rapidly develop extractors for new sources. Our experience shows that the process of developing extractors can be streamlined (further facilitated by an extractor development toolkit) and new extractors can be developed in a matter of hours. Such a model has worked adequately in the past, such as for Junglee, especially when dealing with a few tens of major or important content providers. Alternatively, considering that the content providers know their own metadata the best, they would not have to invest a significant amount of resources into such an extractor, and as a result the Internet world would have a rather powerful video search tool that can be used from many Web portals.

The research that has been done for this paper and the above potential uses reveal that it is highly desirable for the Web video world to come up with a standard that provides both a minimal set of video attributes (comparable to the Dublin Core Metadata set [5]) and a way of structuring these data. The expected impact of XML on the Web world suggests the use of XML tags for this purpose. For the selection of relevant attributes, one has to find a set that is on the one hand not too specific (e.g., not everybody is interested in camera angles) but that on the other hand covers all information that is important to the general audience. We have investigated this issue and plan to discuss it in future. For the most important cases, this can entirely do away with the need to manually write extractors.

One of the implementations supports, among other things, searching the Web for various video types (e.g., movies from videodome.com, hollywood.com, etc.; news clips from foxnews.com, abcnews.com, etc.). Another version of the system called *MusicAnywhere* supports search and access of music on the Web.

Acknowledgement

Vineet Mahajan and Subhrajit Ray have made significant contributions as members of the project team. Interactions with LGERCA's Satish Thatte and Gomer Thomas have also influenced this work.

Bibliography

[1] Advanced Television Systems Committee (ATSC), http://www.atsc.org

[2] ATSC paper A/65, available at ftp://ftp.atsc.org/pub/Standards/A65

[3] Bresnan Communications, http://www.bresnan.com

[4] M. G. Christel, "Addressing the Contents of Video in a Digital Library", in: "Electronic Proceedings of the ACM", San Francisco, 1995

[5] Dublin Core Metadata Element Set: Reference Description. Office of Research and Special Projects, Online Computer Library Center, Inc 1997, Dublin, Ohio. http://www.oclc.org:5046/research/dublin_core

[6] Electronic TV Host, http://www.tvhost.com/~tvhost/index.html

[7] Excalibur RetrievalWare, Excalibur Technologies, http://www.excalib.com

[8] Gist EPG, http://www.tv1.com/tv

[9] A. Hampapur and R. Jain, "Video Data Management Systems: Metadata and Architecture", in "Multimedia Data Management, Using Metadata to Integrate and Apply Digital Media", pp. 245-286, A.. Sheth and W. Klas, Eds., McGraw-Hill, 1998

[10] Informedia Digital Video Project, http://www.informedia.cs.cmu.edu

[11] Intel Smart TV, http://www.hig.com

[12] V. Kashyap, K. Shah, and A. Sheth, "Metadata for building the MultiMedia Patch Quilt" in "Multimedia Database Systems: Issues and Research Directions", S. Jajodia and V.S.Subrahmaniun, Eds., Springer-Verlag, 1995

[13] W. Klippgen, T.D.C. Little, G. Ahanger, and D. Venkatesh, "The use of metadata for the rendering of personalized video delivery", in "Multimedia Data Management, Using Metadata to Integrate and Apply Digital Media", pp. 287-318, A.. Sheth and W. Klas, Eds., McGraw-Hill, 1998

[14] MCA Home Entertainment, http://www.mca.com/home/VL/index.html

[15] Motion Picture Expert Group (MPEG), http://www.mpeg.org

[16] POET, http://www.poet.com

[17] Prevue Online, http://www.prevue.com

[18] G. Salton, "Automatic Text Processing – The Transformation, Analysis and Retrieval of Information by Computer", Addison-Wesley, 1989

[19] Scour.net, http://www.scour.net

[20] K. Shah and A. Sheth, "Logical Information Modeling of Web-accessible Heterogeneous Digital Assets", Proc. of the Forum on Research and Technology Advances in Digital Libraries," (ADL'98), invited paper, Santa Barbara, CA, April 21-24, 1998. pp. 266-275

[21] K. Shah, A. Sheth, and S. Mudumbai, "Black Box Approach to Image Feature Manipulation used by Visual Information Retrieval Engines", The Second IEEE Metadata Conference September, 1997

[22] A. Sheth, "Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics", in Interoperating Geographic Information Systems, M. Goodchild, M. Egenhofer, R. Fegeas and C Kottman (Eds.), Kluwer, 1998.

[23] A. Sheth and W. Klas, Eds., Multimedia Data Managing, Using Metadata to Integrate and Apply Digital Media, McGraw-Hill Publishers, 1998.

[24] A. Sheth, K. Shah, K. Parasuraman, S. Mudumbai, "Searching Distributed and Heterogeneous Digital Media: The VisualHarness Approach," Proc. of the IFIP 2.6 Working Conference on Database Semantics (DS-8):
'Semantic Issues in Multimedia Systems,' Rotorua, New Zealand, January 1999.

[25] Timecast, http://www.timecast.com

[26] Real Networks, http://www.real.com

[27] Resource Description Framework (RDF), http://www.w3c.org/RDF

[28] Verity, http://www.verity.com

[29] VideoBrowser, WorldWide Broadcasting Network, http://www.wbnet.com

[30] VideoDome.com Network, http://www.videodome.com

[31] VideoSeeker, http://www.videoseeker.com

[32] Virage, http://www.virage.com

[33] WebSEEk, University of Columbia, http://www.ctr.columbia.edu/webseek

[34] XML, http://www.w3c.org/XML

[35] XML for Java, http://www.alphaworks.ibm.com/formula/xml

[36] Yahoo!, http://www.yahoo.com

[37] H.J. Zhang, C.Y. Low, S.W. Smoliar and J.H. Wu. "Video Parsing, Retrieval and Browsing: An Integrated and Content-Based Solution", in: Proc. of ACM Multimedia '95, San Francisco, 1995.