sonera plaza | medialab

Multimedia Search Engines White Paper

Sonera Plaza Ltd
MediaLab

www.medialab.sonera.fi
info@medialab.sonera.fi

January 4, 2002

Copyright © 2002 Sonera Plaza Ltd MediaLab
This document is provided *as is,* without any warranty of any kind.
The opinions expressed herein are subject to change without notice.
All trademarks are recognized as the property of their respective owners.

Multimedia Search Engines

1 INTRODUCTION

Search engines have always been among the Web's most popular sites. Traditionally, search engines have concentrated on text material, that is, Web pages, but there already exist many search engines for media files, such as MP3 audio and image files.

In this white paper, we take a look at multimedia search engines, search techniques, products and service providers. By a multimedia search engine we mean an application that helps the user to find downloadable multimedia files or streams, such as audio, video, animation, and images. We concentrate on multimedia search engines that operate in or are suitable for the Web environment. We start chapter 2 by examining how search engines are built and how they work. We then discuss some techniques that can be used to make multimedia content searchable. Chapter 3 introduces four different types of real world multimedia search engines while chapter 4 presents five providers of multimedia search technology.

2 MULTIMEDIA SEARCH TECHNOLOGY

2.1 How search engines work

In a simplified model, a search engine consists of four core components: a spider, a parser or indexer, a query engine, and a Web interface as presented in Figure 2-1.

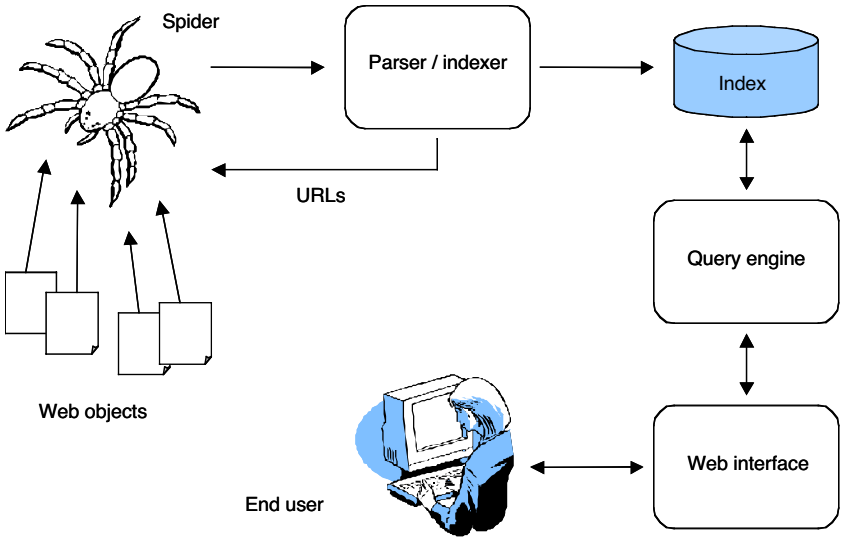

Figure 2-1. Search engine architecture

The spider, also called a robot or a crawler, is the heart of a Web-based search engine. It is an autonomous Web client which automatically makes connections to Web servers and requests objects. The response is checked and if the request is successful, the object is then fetched and indexed. In the indexing phase, words from the textual objects are saved in the index along with other information such as word locations. The search engine index created is similar in concept to the index found in books. While a book's index provides page references for a particular word, the search engine's index contains words along with references to the objects containing those words. As we will see later in this document, also non-textual objects can be indexed. The

 © 2001 Sonera Plaza Ltd MediaLab

index is later used to find out which pages contain the query terms entered by the user. The fetched object is also analyzed for new links and the new URLs found are then fed to the spider.

The query engine consults the search index and by using clever algorithms decides which objects match the user query best. For example, the query engine can boost pages with query terms occurring in the title part of the page, or pages with many referring links.

Usually users don't directly interact with the query engine; they make queries via a Web interface. The Web interface takes the query from the user and sends it to the query engine in a format understood by the engine. The query engine returns the results in a suitable format, for example, XML. The Web interface then parses the results and presents them to the user.

To keep the index up-to-date, the spider must repeatedly check the objects for modifications. Fortunately HTTP – the protocol used between Web clients and servers – has If-Modified-Since style requests to help with this.

2.2 Excluding objects from the index

Spiders can basically collect and index everything publicly available on a Web server. If one wants to disallow the gathering of some – for example copyrighted – material, the robots exclusion protocol [1] can be used. Well-behaved spiders do not fetch objects which are disallowed according the standard. For example, a webmaster can add a "robots.txt" file shown in Table 2-1 in the root directory of a Web server to exclude the file /pics/bike.jpg from Google Image Search. HTML Meta tags can also be used for this purpose.

Table 2-1. An example of a robots.txt file

```
User-agent: Googlebot-Image
Disallow: /pics/bike.jpg
```

2.3 From Text Search to Multimedia Search

Search engines have traditionally concentrated on text documents. Nowadays, information of increasingly audiovisual nature is available on the Web and multimedia search engines are needed to find that material. In the following sections, several different techniques for making multimedia content searchable are presented.

2.3.1 Checking the outgoing links

HotBot[1] was one of the first search engines which presented simple multimedia features. It allows for the restriction of the results for pages containing a given media type or technologies. For example, the user can select for the search to return only pages containing audio, video, JavaScript, ActiveX or a given file suffix. Analyzing a page's outgoing links and the file types they point to during the indexing phase achieves this behavior.

Another example, the MediaSearch[2] prototype developed at Sonera Plaza MediaLab, takes this approach a little bit further. By using MediaSearch it is not only possible to restrict the results to pages containing the appropriate media, but it's also possible to actually see the appropriate media links on a page in real time. For example, when restricting results for pages containing animation, MediaSearch presents a "Show Animation on this page" link after each hit on a result page. If the link is clicked, MediaSearch fetches the page and parses the hyperlinks in the

[1] http://www.hotbot.com

[2] Currently at http://www.medialab.sonera.fi/mediasearch/

 © 2001 Sonera Plaza Ltd MediaLab

correct format in the HTML. MediaSearch supports over 30 file formats including common formats for digital audio, image, video and animation. If the restriction was made for images and the fetched page contains images of type GIF or JPEG, a scaled image is shown in a browser window. Other MediaSearch features include automatic playlist creation; playlists are generated for found MP3, RealMedia and Windows Media files so that it is possible to listen or watch media files in a sequence. All queries are run against a third party index which is created and continuously updated by indexing robots which retrieve HTML pages. Audio, image, video or other media files are not indexed in this database. Only the file types the HTML page refers to are saved in the index.

2.3.2 Analyzing the referring page

Many image searches on the Web nowadays use an approach that is more sophisticated than what was described earlier; they can analyze and index the text on the page adjacent to the image, the image link text, text in the HTML alt tag, file name or file path name. At least Google also uses algorithms to remove duplicate images. Similarly, this approach can also be used with other media files such as audio and video. Even though these search engines do not "look inside" the media files, they can give quite relevant results.

2.3.3 Looking inside the media file

One can go a step further in multimedia indexing by opening the media file itself and trying to mine for textual information in the file. For example, at the end of an MP3 file there is a fixed size 128-byte area called the ID3v1 tag, used for song title, artist, album, year, comment, track number and genre [2]. The newer ID3v2 tagging system [3] has much richer syntax, but it is not as widely used.

Similarly, also other media formats can contain this kind of information. For example, a Portable Network Graphics (PNG) image file can contain textual information such as title, author, description, copyright, creation time, software used, disclaimer, warning, source and comment [4]. Information such as title, author, copyright, description, and keywords can be added during the RealMedia encoding process. Unfortunately, typically only the title, author and copyright fields are completed.

2.3.4 Utilizing metadata

Not all file formats contain meta information, and even if they do, an indexing engine should know how to handle all the different file formats and where to find that information in a file. It would be better, if we had a data model which could be used with different media formats and utilized a rich set of metadata. This is where the metadata model comes in. Unfortunately, there exist more than one metadata format, and someone must naturally provide the metadata before it can be used. Two metadata models are briefly described here: the Dublin Core metadata model has worked well for textual material for several years and MPEG-7 may become the standard for multimedia in the future.

2.3.4.1 The Dublin Core

The Dublin Core (DC) metadata set uses fifteen elements to describe an electronic resource. These elements are title, creator, subject and keywords, description, publisher, contributor, date, resource type, format, resource identifier, source, language, relation, coverage, and rights management. The use of so-called qualifiers can extend these elements. For example, it is possible to have a *Relation.IsPartOf* element to declare that an object is part of a larger object. DC has traditionally focused on text resources, but it can also be used to describe audiovisual resources to some extent. The example taken from [5] and shown below shows a the metadata for a complete video and for one of its scenes by using an extended DC element set.

 © 2001 Sonera Plaza Ltd MediaLab

- *Complete Video Documentary*
  Title = Tu as crie LET ME GO
  Creator = Director: Anne Claire Poirier; Producers: Joanne Carriere, Paul Lapointe
  Subject = Heroin habit; Drug abuse - Social aspects
  Description = "A film both about the heroin problem and also about attachment, loss and the resolution of grief, made by an honored Canadian filmmaker, Anne Claire Poirier, in the wake of the death of her 26 year old daughter, a heroin addict, murdered whilst working as a prostitute."
  Publisher = Canada: National Film Board of Canada, 1998
  Date = 1998
  Type = Image.Moving.Film.Documentary
  Format = 1 videocassette (97 min.) : sd., b+w ; 1/2 in
  Identifier = 362.293
  Source = QVC 362.293 tua vhs
  Relation.HasPart = scene1, scene2, scene3, scene4, scene5,...
  Copyright = National Film Board of Canada

- *Scene1*
  Description.transcript = "The end of the world; a Sunday morning; a young woman dies, strangled in her ramshackle apartment. She was a heroin addict; she was of service to men; she was beautiful; she was my daughter"
  Description.keyframe = "http://www.slq.qld.gov.au/av/scene1.jpg"
  Description.clip = "http://www.slq.qld.gov.au/av/scene1.rm"
  Type = "Image.Moving.Film.Documentary.Scene"
  Format.length = 2min 20 secs
  Coverage.t.min scheme = SMPTE content = 00:00:00;1
  Coverage.t.max scheme = SMPTE content = 00:02:20;25
  Relation.IsPartOf = video_doc

Depending on the indexing application, the metadata presented above can be saved as HTML incorporating the metadata in the Meta tags, or it can be expressed in and transported to other formats such as XML or RDF.

2.3.4.2 MPEG-7

Unlike the Dublin Core, the MPEG-7 data model was designed for multimedia from the very beginning. It is much more complex than the DC model. While former MPEG standards 1, 2, and 4 make content available, MPEG-7 allows for the finding of the content. In other words, MPEG-7 specifies a standard set of descriptors that can be used to describe various types of multimedia information. The description is associated with the multimedia content itself to enable efficient and relevant searching. It is said that MPEG-7 will make the Web as searchable for multimedia content as it is for text today [6 and 7]. MPEG-7 metadata can be expressed in XML and sent separately or with the audiovisual stream. MPEG-7 should make it possible to generate queries like [8]:

- *Play few notes on a keyboard and get in return a list of musical pieces containing (or close to) the requested tune or images somehow matching the notes, e.g. in terms of emotions.*
- *Use an excerpt of Pavarotti's voice and get a list of records or video clips where Pavarotti is singing or video clips where Pavarotti is present.*
- *Draw a few lines on a screen and get in return a set of images containing similar graphics.*

MPEG-7 is an unfinished standard and it's not yet widely used, but at least Singingfish Multimedia Search (see chapter 4) utilizes MPEG-7 description schemes to model the metadata of Internet streaming media [9].

 © 2001 Sonera Plaza Ltd MediaLab

2.3.5 Other analyzing methods

If a media file has metadata associated with it, the search engine's life will be much easier. Other methods for making audiovisual media searchable also exist. Broadcast television can include a digital transcript which can be used in searches. Text recognition technology can be used to extract the title and captions from a video. Speech recognition technology can be used to digitize spoken words on audio tracks. Some search engines can utilize such image and video content as colors and patterns.

3 EXAMPLES OF MULTIMEDIA SEARCH ENGINES

In this chapter, four different types of search engines – image, MP3, MIDI, and video search – are briefly described.

3.1 Google Image Search

http://images.google.com

Google Image Search contains information of more than 250 million Web images. Google analyzes the text on the page adjacent to the image, the image caption and other factors to determine the image content. All Google text search query operators can be used with the image search. In addition, the advanced image search allows restricting the results according to image size (any size, icon sized, small, medium, large, very large, wallpaper sized), file type (any type, JPEG, GIF), coloration (any color, black and white, grayscale, full color) or domains. A filter for mature content is also available, but it only works with pages in English. The result page shows thumbnail images with image name, resolution, size and location.

3.2 Lycos MP3 Search

http://music.lycos.com

Lycos uses FAST's technology for its MP3 search. Many MP3 search engines have been criticized for returning dead links. Lycos makes no exception in this regard, although it attempts to avoid dead links in two ways. First, Lycos does not return files whose connection reliability is known to be poor. This is established by comparing the Lycos results with the technology provider's MP3 results at AllTheWeb.com. Second, Lycos tries to deal with dead links by using a wrapper which checks whether the MP3 file is available at the server. If not, the search presents "Currently unavailable" popup. If it is, the search presents a popup with links to the original MP3 file and to a dynamically generated M3U playlist file.

3.3 MIDI Explorer

http://www.musicrobot.com

MIDI Explorer is a search engine for locating MIDI files and MIDI sites on the Web. It knows of over a million MIDI files. MIDI file names and file lengths together with a link text on HTML pages are retrieved and indexed. File lengths are used to find different MIDI files with the same name. The search has four query types: in *Group by name* the results are organized into groups that have the same name, *Group by length* will sort results by file length, *Show all* presents all results, and *Length* can be used to find files that have a specified file length.

3.4 VastVideo

http://www.vastvideo.com

VastVideo allows the searching of 18 000 video clips in Windows Media format. The videos are divided into 13 categories, so it is also possible to find a clip by browsing the categories. It is also possible to buy a tape of the found video. VastVideo's search language contains some features not found in other search engines: @ matches exactly one alphabetic character, #

 © 2001 Sonera Plaza Ltd MediaLab

matches exactly one numeric character, \ forces the next character to be a literal character, - matches zero or one character, ? matches exactly one character, ~ matches words spelled similarly, and [expr] matches one character from expr.

4 MULTIMEDIA SEARCH TECHNOLOGY PROVIDERS

In this chapter, we introduce six multimedia search technology providers and their services or products. AltaVista, FAST, Inktomi and Singingfish sell their search as a service; they provide a multimedia index and a method for accessing it. Table 4-2 summarizes the features of these services. Convera and Virage sell products for video and audio indexing. It should be noted that this chapter is not a complete list of multimedia search technology providers.

4.1 Service Providers

4.1.1 AltaVista

http://services.altavista.com

AltaVista provides Multimedia Search as an optional package for its Internet Search service. In this model, licensed customers can have access to the index used by the main AltaVista site and containing more than 45 million media files. The query is made by using the HTTP GET method. Query results are available in XML format or in AltaVista's proprietary format. Table 4-1 shows an example output for a video search 'car'. AltaVista's query language contains a wide variety of operators including +, -, "", AND, OR, AND NOT, and NEAR. Multimedia search is divided into Image, Audio and Video searches which are further divided into basic and advanced search. Image search supports GIF, JPEG and PNG files. Search can be restricted to photos, graphics or buttons/banners. In addition, it is possible to return color or black and white images or both. Audio search allows searching for AIFF, AU, MIDI, MP3, WAV, Windows Media or Real formats. Video search allows searching for AVI, MPEG, QuickTime, Windows Media or Real formats. Audio and video results can be restricted by clip duration. Multimedia Search contains family filter which tries to exclude offensive material when turned on.

4.1.2 Fast Search & Transfer (FAST)

http://www.fast.no

FAST Multimedia Search is based on the same core technology as the AllTheWeb.com search engine and it uses the same index for queries. Multimedia Search is also capable of indexing and hosting licensed content provided by customers. Multimedia Search is accessed via FAST's query API and results are returned in XML, so customers can create their own look and feel for the search. The search allows for the simultaneous searching of image, video and audio files or one media type at a time. Supported media formats include JPEG, GIF, BMP for images; AIFF, MP3, MIDI, RealAudio, WAV for audio; and AVI, AVI-DivX, QuickTime, MPEG for video. With images, it is also possible to restrict the search for color, grayscale or line art. With audio files, it is possible to restrict the search for stereo or mono. With audio and video search, there is an option to search for streams or downloadable files, or both. The search includes an offensive content reduction filter which reduces the amount of offensive material, but only works with pages in English. According to FAST Multimedia Search index contains information of approximately 70 million media files.

 © 2001 Sonera Plaza Ltd MediaLab

Table 4-1. An example of XML results from AltaVista video search taken from [10]

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE Val SYSTEM http://partners.altavista.com/static/iss/iss.dtd>
<Val>
 <MmSearch>
    <MmHeader>
       <Match>874</Match>
       <Available>875</Available>
       <InputParam>car</InputParam>
    </MmHeader>

    <MmEntry>
      <Title>Dude, Where's My Car</Title>
      <Abstract></Abstract>
      <Size>0</Size>
      <URL>http://movies.vidnet.com/perl/altavista/mpge_nf.pl…</URL>
      <FrameCount>0</FrameCount>
      <Similar>n</Similar>
      <Mmpage>4449224523525</MmPage>
      <ClipTitle></ClipTitle>
      <ClipAuthor></ClipAuthor>
      <ClipCopyright></ClipCopyright>
      <ClipAbstract></ClipAbstract>
      <ClipFormat>asf</ClipFormat>
      <ClipDuration>264</ClipDuration>
      <ClipChannels>1</ClipChannels>
      <ClipSampleBits>16</ClipSampleBits>
      <ClipSampleRate>8000</ClipSampleRate>
      <MmType>8</Mmtype>
      <MmClass>2051</MmClass>
      <MmSize>4366></MmSize>
      <MmHeight>120</MmHeight>
      <MmWidth>160</MmWidth>
      <MmFileName>dude56.asx</MmFileName>
      <MmID>132863815</MmID>
      <VideoFrameRate>10</VideoFrameRate>
      <VideoDepth>32</VideoDepth>
   </MmEntry>
  …
 </MmSearch>
</Val>
```

4.1.3 Inktomi

http://www.inktomi.com

Inktomi provides Media Search as an additional service for its normal text search [11]. Currently Inktomi uses Singingfish's (see below) content for its Media Search, so it is possible to search the same media formats as with Singingfish Multimedia Search: MP3, QuickTime, RealMedia and Windows Media. It is not possible to search for images. Queries are made and results are returned by using the Inktomi Data Protocol (IDP). It is possible to search for MP3

 © 2001 Sonera Plaza Ltd MediaLab

files with artist, song title, genre, year, or album name. If these metawords are used in queries, the results are automatically restricted to files that support the ID3 tag. The search engine stores the results of five attempts to crawl a media file and the value stored reflects the success level of those attempts. Queries can be restricted to files whose connection reliability value is on a required level.

4.1.4 Singingfish

http://www.singingfish.com

Singingfish, a subsidiary of Thomson multimedia, develops a multimedia search and sells it as a service. Singingfish has retrieved and indexed over 18 million streams and media files. The database contains links to Microsoft Windows Media formats, Apple QuickTime, MP3, and the following RealNetworks formats: RealVideo, RealAudio and RealFlash. The files are divided into eight categories by an automated classification system. The search can be restricted by category, stream length and bit rate. The demonstration search at singingfish.com shows category name, stream name, length, bit rate and copyright information with the results. It also contains a family filter option. Singingfish's customers include metasearch services such as Dogpile and MetaCrawler. Also, RealNetworks' RealOne Player uses Singingfish's search technology to allow end users to find digital audio and video files on the Internet.

Table 4-2. Features of multimedia search services

	AltaVista	FAST	Inktomi	Singingfish
Number of media files	45 million	70 million	18 million	18 million
Query language	Excellent	Good	Good	Not known
Simple / advanced search	Yes / yes	Yes / yes	Customizable	Yes / yes
Audio formats	AIFF, AU, MIDI, MP3, RealAudio, WAV, Windows Media	AIFF, MIDI, MP3, RealAudio, WAV	MP3, RealAudio, Windows Media	MP3, RealAudio, Windows Media
Image formats	GIF, JPEG, PNG	BMP, GIF, JPEG	None	None
Video formats	AVI, MPEG, QuickTime, Real, Windows Media	AVI, AVI-DivX, MPEG, QuickTime	QuickTime, Real, Windows Media	QuickTime, Real, Windows Media
Content filter switch	Yes	Yes	No	Yes
Content categorization	No	No	No	Yes
Clip duration restriction	Yes	Yes	No	Yes
Bit rate restriction	No	No	No	Yes
Query API	XML + own	XML	IDP	Not known

4.2 Product Suppliers

4.2.1 Convera

http://www.convera.com

Convera's Screening Room can be used to capture, search for, index, and publish video content. With it users can capture video; browse visual summaries; catalog content using metadata, annotations, closed caption text, voice sound tracks; and search for video clips using text and image clues. Convera also develops the RetrievalWare product for indexing and searching text. Its add-on, Internet Spider, can be used to retrieve, index and search multimedia files. The query language supports all standard Boolean operators. APIs are available for C, C++ and Java.

 © 2001 Sonera Plaza Ltd MediaLab

4.2.2 Virage

http://www.virage.com

Virage develops the VideoLogger software which simultaneously encodes and indexes video. VideoLogger can investigate an analog or digital video signal, segment the video and generate a storyboard of keyframe images according to changes in the visual content. It can extract closed caption text or teletext from the signal. VideoLogger can also identify faces, on-screen text and numbers in the video by using Virage plug-ins. Face identification is done by mapping facial features in a video to user-defined face libraries. Results are added to the VideoLogger index. A text recognition plug-in can detect text such as sport scores, names, and pricing information, in a video frame. Using the Virage Audio Recognition plug-in, VideoLogger also listens to the audio signal and converts spoken words, speaker names and audio types into searchable text. This information index is used to search for and retrieve the video clip. The speaker identification component recognizes 300 speakers regardless of the words or language spoken and it is also possible to add new speakers to the system. An audio classification component identifies sounds in the video signal and registers where they occur. It is possible to recognize audio types such as applause or laughter. VideoLogger supports RealVideo, Windows Media, QuickTime and MPEG. It runs on Sun Solaris and Windows NT/2000.

5 SUMMARY

Search engines have concentrated on indexing textual material in the past. As more and more audiovisual content is freely available on the Web and more users have broadband access to the Internet, multimedia search engines have become attractive.

Basic multimedia search engines analyze the outgoing link types of a page. Today, many public spider-based multimedia search engines do not "look inside" the media files, but they analyze the referring page and index the text on it.

Other methods for making audiovisual media searchable also exist. Broadcast television can contain a transcript which can be used in video searches. Text recognition and speech recognition technologies can be used to extract titles, captions and spoken words.

Metadata helps in constructing multimedia indices and generating more relevant search results. The Dublin Core metadata model has worked well for textual material and it can be used for multimedia content to some extent. Time will tell whether MPEG-7 will be the definitive metadata standard for multimedia content or if it fails due to its own complexity.

Multimedia search engine technology providers typically sell their tools in two ways; as a product or as a service. In the former model the customer installs, configures and runs the software by himself. In the latter model, the customer gets access and protocols to the provider's media index.

Convera and Virage have quite similar products. Service provider FAST has the biggest media index, but AltaVista has the most extensive query language. Singingfish Multimedia Search and Inktomi Media Search do not contain images, thus their indices are smaller. The service with the biggest index is not automatically the best service. Instead, one should concentrate on the number of unbroken links in the index and relevant results returned.

 © 2001 Sonera Plaza Ltd MediaLab

REFERENCES

[1] Robots Exclusion. Available from http://www.robotstxt.org/wc/exclusion.html [Accessed November 2, 2001]

[2] ID3 made easy. Available from http://www.id3.org/id3v1.html [Accessed October 29, 2001]

[3] ID3v2 made easy. Available from http://www.id3.org/easy.html [Accessed October 29, 2001]

[4] PNG (Portable Network Graphics) Specification, Version 1.2. Available from http://www.libpng.org/pub/png/spec/ [Accessed October 30, 2001]

[5] Hunter Jane, Mather Tim. Veggie – An Indexing, Browsing, Search and Retrieval System for Audiovisual Libraries. Available from http://archive.dstc.edu.au/RDU/staff/jane-hunter/INFOG99/paper.html [Accessed November 7, 2001]

[6] Martinez José M. (ed). Overview of the MPEG-7 Standard (version 5.0). Singapore, March 2001. Available from http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm [Accessed November 5, 2001]

[7] Day Neil. Martinez José M. Introduction to MPEG-7 (v3.0). Sydney, July 2001. Available from http://mpeg.telecomitalialab.com/public/mpeg-7_introduction.zip [Accessed November 5, 2001]

[8] MPEG-7 Frequently Asked Questions. March 2000. Available from http://mpeg.telecomitalialab.com/faq/mp7.htm [Accessed November 5.2001]

[9] Rehm Eric. Representing Internet Streaming Media Metadata using MPEG-7 Multimedia Content Description Schemes. Singingfish.com. Available from http://www.acm.org/sigs/sigmm/MM2000/ep/rehm/ [Accessed November 6, 2001]

[10] Internet Search Services. Version 3.02. Search Engine Interface Description. AltaVista Company. August 20, 2001. Available from http://www.altavista.com/static/pdf/ISS302.pdf [Accessed November 9, 2001]

[11] Inktomi Web Search product information. Available from http://www.inktomi.com/products/search/web.html [Accessed November 6, 2001]

 © 2001 Sonera Plaza Ltd MediaLab