fast SBP™

Demystifying multimedia search

Online and on-demand rich media is a burgeoning industry with new strategies for market share growth. Effective multimedia search combines powerful technologies to create a foundation for these strategies.

A business white paper
by FAST Search Best Practices

What is multimedia search?

Rich media documents, such as songs and videos, are difficult to automatically understand. The search technology used to tackle this challenge, leveraged by content owners and viewers embracing new monetisation models, opens up a whole new world of innovative applications.

Look around you. Rich media – audio, video, images, animations – is everywhere: on terrestrial, cable and satellite television; on FM, satellite and digital radio; in landline, mobile and voice over IP telephone calls; in online videos and radio stations; in streaming videos, podcasts, DVDs, home videos, voicemails… The explosion in digital rich media content is poised to dwarf traditional digital formats, which means that information retrieval technologies, including search engines and personal search tools, need to adapt to cater for the growing professional and consumer demand in multimedia.

The recent proliferation of broadband and high-speed mobile access is making rich media widely accessible. Therefore, in contrast to traditional linear programmed broadcast, consumers are now favouring "always-on, anywhere, anytime, any device" access to video and audio. Somewhat cautiously, multimedia content owners are beginning to embrace these monetisation models. A new business landscape is forming. For example, Triple Play strategies from the telcos, and pressure from independents such as Apple, are justifying business models around the distribution of copyrighted videos and music.

One of the keys to these distribution models is search as the primary means to link users to the relevant content. At its most general, a multimedia search application enables users to retrieve any digital rich media content, be it audio files with music or spoken word, images, animations or videos. For example, such an application provides users with the ability to ask the following questions, or be alerted when new content matching these criteria is available:

- Give me a list of all videos filmed by our field crews within a ten mile radius of the Parliament between the 10th and 15th of January.
- Let's watch the episode of "Friends" where they sit in the coffee shop and talk about Rachel's baby.
- What is the most relevant information we have on the intranet about how to use the new manufacturing machinery (PDF manuals, training videos, FAQ database, web pages etc…)?
- Take me to yesterday's news where the anchorman mentions the World Cup in the same segment where the Prime Minister discusses hooligans.
- How many of the phone calls received in our call centre last week where from unhappy customers? Which operators are best at retaining unhappy customers?
- Where can I download a "Crazy Frog" ringtone compatible with my mobile phone?

The technologies and methods used to answer these questions have different degrees of complexity, and a subsequent variation in the quality of results.

In the text world documents are simply defined by the words used, and they can then be searched by those words. Multimedia includes images, sounds, music and speech. Therefore a blend of sophisticated techniques is necessary to correctly understand the meaning of them, without which search would not be possible. This paper will describe some of the methods used in the industry to solve this complex puzzle, and explore how these new methods enable different applications to link multimedia content with the people who want it.

Example Applications

"Multimedia search" is not just one application, but a whole family of them that all have in common the ability to return rich media content in their results.

To begin with, good, basic search must be format-agnostic, returning the most appropriate documents regardless of the source. Therefore high-end search platforms must be multimedia enabled. For example, corporate intranets commonly contain conference call recordings and webcasts along with presentations, memos and other text-based materials. Enterprise search solutions should index these and make them available. In the same way, when a desktop search tool finds information about "Britney Spears" on a personal computer, the results could include songs and music videos from a personal digital media collection, or voicemails regarding a concert ticket application.

In addition to this, search-centric applications specifically designed for rich media distribution and consumption are becoming more prevalent. These include video web search and rich media search (and management) for content owners and syndicators. There are also broadcast monitoring and call centres applications, and applications for the various devices that can be used to view content: computers, televisions, iPods, Sony PSPs and in-car satellite navigation touch-screens. For example, many cell phones are now geographically aware and have built-in cameras. This enables product search from a mobile by scanning the bar code of a product, or using object or character recognition: a true multimedia-to-multimedia search.

Video web search

As monetization models for rich media evolve around search-driven advertising, subscriptions and content sales, there is more and more of an emphasis on video search, aggregation and distribution from traditional web players. Whereas there may be a small number of dominant web search engines, the video web market remains fragmented. Many owners are attempting to keep their content within walled-gardens and free video is generally of uneven quality and reliability, swamped by clips of dubious value. Spidering the internet for video can be difficult as links are often dynamic, intentionally obfuscated, or only accessible via peer-to-peer networks.

In the face of this complexity, crawling technologies are quickly adapting to the paradigm of video. Once the technicalities of quality assessment, treating embedded media links, streams, short-lived videos and so on are overcome, there is still the key issue of understanding the content. Transcripts will often be ineffectual in videos where the dialogue is minimal and non-descriptive. For example, in a movie's action-packed chase scene, the participants are unlikely to say that they are racing around in cars. Nor will they describe the type, colour or location of their vehicles, which is probably what viewers want to search on. Image analysis, optical character recognition, link text and contextual information are required. Combining these methods properly can be done manually on small sets of hand-picked and controlled sites. On the other hand this becomes a complex and ill-defined problem when set against the whole of the worldwide web.

The next challenge, after different information related to the subject of a file is obtained and stored, is how to correctly blend them to judge relevancy. How do you compare one video where the transcript, judged to be 80% accurate, contains the phrase "Britney Spears" versus another where she is mentioned on the page where the video was embedded? All the available layers of information must be merged into one unified representation. In conjunction with this novel ways of expressing search queries will aid both usability and search precision. For example taking a photo of a restaurant with your mobile phone to find reviews before going in to eat; or whistling a tune to find the ringtone.

Furthermore a key part of search applications is the generation of teasers describing why each result is relevant. Web users are unlikely to download large video files unless they believe them to be useful. Therefore video thumbnails, dynamic storyboards, metadata, extracted entities and dynamic contextual information must be appropriately generated and presented for a successful search experience.

There are also many challenges around the architecture and scaling of an index of billions of multimedia objects, covered below in more detail.

So far no video search engine has been highly acclaimed. Services that rely on primitive metadata or only on transcription and phonetic information are unable to meet users' needs. The jury is still out on who will dominate the video search space, but success will no doubt depend on the right combination of technology and relevancy, partnerships with content owners, marketing budgets and the whims of web surfers.

Content distribution

A second area of multimedia search is the applications provided by the owners and syndicators of rich media content. This is a key use of the technology, as there is a proven consumer demand for pay-for-content, as seen by the success of mobile ringtone sites and legal music downloads. In particular video-on-demand services are taking on increased importance as the audience for traditional media products such as programmed TV and DVDs flattens or declines. State of the art search capabilities are key to the strategies behind the growth of these services.

There are fewer crawling challenges here: the focus is on providing very high relevancy on targeted content. Use cases include standard online search and distribution (either free or paying), such as movie trailer or news channel websites. This functionality is also central to video-on-demand services and the software inside set-top-boxes, PVRs (Personal Video Recorders, i.e. Sky+ or TiVO) and personal media management software such as iTunes. As consumers are presented with a wider array of possible programs to watch and songs to listen to, the ability to find what they're looking for, or to be suggested new content that they will enjoy, is essential.

The type of information generally searchable in these scenarios will include speech-to-text transcription, in particular for news, educational and factual video and radio where the spoken word is directly related to the subject matter. For entertainment related content, such as movies, sitcom episodes, and music, it is valuable to have good descriptions, reviews, and metadata (genre, actor, director, singer, songwriter). Also collaborative filtering (i.e. people who listen to Britney Spears also like Madonna) is pertinent. Finally, image analysis technology can also provide valuable descriptive information.

Above and beyond monetisation, there is a demand for multimedia search tools for video archive management and browsing. Production companies and television channels own large amounts of video. The ability to quickly find a useful piece of footage for inclusion in a news bulletin or documentary is critical to them, both in terms of reducing production times and costs, and leveraging archives to avoid recreating or purchasing content.

In addition video archives can then be easily used to create targeted programs for delivery on emerging broadband and mobile platforms. In these scenarios transcripts, human annotations and metadata from legacy systems are used. Metadata generated at capture time (location of the camera crew, date, technical filming information) is also valuable for more structured searching.

Broadcast and Call Monitoring

High-end search platforms are increasingly used to power discovery and data-mining applications. Driven by the ability to review large amounts of information efficiently, they are at the heart of tools for needle-in-the-haystack fact-finding, analysis of news flow and ‘on-the-fly’ analytics.

Multimedia content is a manually intensive medium to review. Monitoring five television channels and three radio stations 24/7 would require eight people in each shift listening and watching continuously, a tough yet menial task for a method yielding poor scalability and consistency. Automating this is achieved by processing broadcasts or phone-calls through the content refinement and enrichment sub-system of a search platform. Since they are designed to deal with large volumes of unstructured information, the parallel monitoring of multiple channels is possible in a scalable way.

Bird’s eye view analysis of trends and behaviour can then be powered by novel search derivative applications based on flexible search platforms. These combine entity extraction, navigation, sentiment analysis and reviewer personalisation. In relation to broadcast monitoring, concrete applications include threat assessment and the tracking of political and economic situations. In call-centres they include customer satisfaction and agent performance supervision. Naturally the multimedia data is often best combined with other data sources (customer feedback, blogs, websites and news feeds) for optimal results.

On top of this, real-time alerting is much in demand. From managers needing to know as soon as one of their call centre agents slips up and gives the wrong information, to politicians being aware that the opposition party is issuing an aggressive statement, this is a pivotal application of multimedia search platforms. Knowing the facts as soon as something critical airs enables people to make optimal decisions, be it running an election campaign, or choosing what to watch on the television one evening.

How multimedia search works

Standard internet and enterprise search is built on four pillars: treating the corpus of information as discrete documents; using a distributed architecture for optimal performance; taking the free text and metadata to derive each document’s “meaning”; and using a relevancy model to best match each query to the information in the index.

Multimedia search faces basic challenges to this paradigm: “documents” are not clearly defined; particularly large data volumes are involved; what each object talks about is not simple to extract; and relevancy models need to take more variables into account.

Documents

What is a rich media document? A news program could be one document; but from a semantic point of view it may be more relevant to break it down further into individual news stories. Sensibly splitting up large videos and streams is therefore critical for accurate relevancy. At the backend of a search implementation, a web crawler, database or other connector will pull down multimedia documents. These should then be broken up contextually, for example by identifying scene changes and speaker changes. Streaming videos and continuous broadcasts clearly need to have many boundaries defined, such as advertising breaks as well as scenes and speakers. With this in mind a search can return not just relevant broadcasts, but the actual matching scenes which searchers can directly navigate to.

Additionally, parts of a program, such as the originally-broadcast sponsorship and advertising, may need to be removed or replaced before redistribution of the content. All of these requirements necessitate providing the content creator or aggregator with a set of tools to manipulate or segment the data prior its submission into a search index.

Above and beyond this simple unit model, new technologies, such as Contextual Insight, allow multi-layered models of data to be stored. Rather than artificially breaking streams up into simple “documents”, they can be marked up in many scopes within the same index, allowing users to dynamically choose which granularity (program, scene, speaker etc) to search against.

Infrastructure

From an architectural point of view, multimedia search presents many challenges for service distribution and scaling. Monitoring continuous channels or streams, and attempting to index the billions of multimedia objects on the web, will generate large volumes of information. These transcripts, thumbnails and metadata will amount to petabytes of data which need to be indexed, stored and managed effectively. Therefore both the data processing sub-system and core search engine architecture must be built on highly scalable and robust technology. In particular the index must fully support a multi-server installation with node synchronisation and a unified relevancy model.

It should be noted that the parsing of multimedia content, described below, can be computationally expensive. Processing and indexing must be done in parallel and asynchronously, so that new content can become searchable as soon as possible, but valuable information from time-intensive modules can be added later.

In addition security and access control mechanisms are of concern. So if a video web search engine wants to make money from premium content it needs to provide full support for Digital Rights Management.

Meaning

The third challenge for search technologies is defining the meaning of each multimedia element, without which it is impossible to match queries to results.

One of the elements often used is existing embedded metadata: filename, closed captioning, format, duration, size, creation date, resolution, sampling rate, etc. The audio track of multimedia content also contains valuable information that can be extracted. This is done using music versus speech categorisation, wide vocabulary speech-to-text transcription, speaker identification, or acoustic and phonetic analysis. All this is time-aligned to the original footage.

From the video part of multimedia content, image processing can be used to perform further classification and annotation, for example in character recognition of text displayed within videos.

One concern is that these methods often produce noisy results. Semantic and factual extraction can then be used to increase the quality of data augmentation stages. Some providers will also apply manual tagging to files. This is costly and sometimes of inconsistent quality. Nonetheless the approach is favoured by content creators that have a tight control of the production of their multimedia assets, and can be an important part of certain relevancy models.

Contextual information when available is also essential to refine the description of content. For example, in a web setting, a mixture of transcripts, link-text, the text of the container web-page, filename and video quality along with popularity of the source channel may all be used to “understand” and rank videos and images. In a video distribution application, externally contained metadata such as title, description or user reviews which are present in the records management system will be taken into account.

Relevancy

All the above is used to describe the meaning of multimedia data. This information is then fed into the search platform's scoring and ranking system.

Powerful multidimensional ranking models are the optimal technology choice here. They use all the different metadata elements, taking into account the accuracy and expressiveness of each, along with other global attributes such as the quality and authority of the source, and the freshness of the data.

In the entertainment industry user interaction models are also key: people are likely to be looking for more popular videos or the most downloaded ringtones; and user feedback can be used for collaborative filtering. For example, an application that manages playlists can use implicit usage information anonymously to drive recommendations. If one person is subscribed to episodes of C.S.I. and Lost, and another to 24 and C.S.I. they may also enjoy 24 and Lost respectively. This provides an immediate up-sell opportunity to the content provider.

In a TV news search application, relevancy will primarily use transcripts, freshness, the source's authority and an OCR of text displayed onscreen. It may also be useful to tag scenes based on a bank of images of high profile people, such as politicians and celebrities. On the other hand, for a ringtone search, link text, file attributes, popularity, profit margins, audio quality and handset compatibility will be combined to find the song that the person is most likely to want.

The future of multimedia search

Online and on-demand rich media is a burgeoning industry, where the economic models are still being debated. The majority of technical challenges are understood and the necessary ingredients are available. No single combination has yet become the de facto method for the distribution of these complex and diverse data types. On the other hand with the correct mixture of technologies the exercise of building an effective multimedia search application becomes an attainable goal rather than an alchemist's dream.

The components needed include advanced mining of audio and video, such as speech-to-text and meta-data extraction, combined with hierarchical segmentation of streams to provide contextually aware navigation. In addition the information from the surrounding context (webpages, metadata stores) is extremely valuable in defining the meaning of an element.

For each vertical application of multimedia search – the web space, enterprise content management or the entertainment industry – optimal relevancy models must be derived. This will leverage all the above primitives, along with next generation discovery frameworks such as Contextual Insight and self-learning from usage feedback. Also central are all the tools provided by best in class search platforms: linguistic and statistical analysis, semantic and lexical data annotation and so on. The correct blend of these will highlight the resonance of content around certain themes and concepts, identifying it as relevant to particular searchers.

The final keystone to a highly successful rich media application is the business model. The power play between content owners and would-be distributors is still unresolved; piracy concerns are not yet fully alleviated with current digital rights management offerings; and the role of traditional broadcast channels equipped with advanced set-top boxes, versus that of telecommunications organisations providing on-demand services to computers and mobile phones, is yet to be defined.

Recent mergers in the industry and bold press releases have revealed many ambitions from companies hoping to dominate the market. Time will tell who will succeed. Nonetheless whatever conclusion the multimedia wars reach, there is no doubt that media users will in the end benefit from a wealth of choice and hugely flexible access to rich media. Search is then set to play a core role in helping them find what it is they actually want to watch from the vast catalogues of content available.

About FAST SBP™ (Search Best Practices)

SBP consulting is a highly focused transfer of search knowledge and experience from FAST to its prospects and customers. SBP workshops aim to help enterprises realize the full potential of search, by creating optimal strategic, functional and technical roadmaps, delivered in the form of business model, solution and architecture designs.

For any feedback or questions related to this paper, please contact us at sbp@fastsearch.com.

Fast Search & Transfer
www.fastsearch.com
info@fastsearch.com

Regional Headquarters

The Americas
+1 781 304 2400

Europe, Middle East & Africa (EMEA)
+47 23 01 12 00

Japan
+81 3 5511 4343

Asia Pacific
+612 9929 7725