JLIS.it
Italian Journal of Library and Information Science
Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione

Introducing Multimedia Information Retrieval to libraries

Roberto Raieli

MIR cultural references

Documents and knowledge for today's welfare

The press, and printed texts, have broadly monopolized culture, knowledge circulation and memory preservation for centuries, but the communication entrusted to images and sounds was never really given the second place, thanks to the role that images and sounds have always maintained, throughout history, in oral and visual culture. During the 20th century, however, a true revolution took place and spread throughout society, characterized by technology innovations regarding media of all kinds, as well as their development to digital form, dissemination via the Web and interaction with the user. The *information object* resulting from this process is quite complex, but well established and easily definable in everyday life: a digital, multimedia, interlinked and interactive *resource*, almost always on the Web. This object, although it escapes overall definitions – such as the typical *multimedia* or the broader *hypermedia* – has definite placements in today's knowledge society.

JLIS.it Vol. 7, n. 3 (September 2016)
ISSN (online) 2038-1026 | ISSN (print) 2038-5366
DOI: 10.4403/jlis.it-11530

The true innovation is provided by the digital form, which enhances the potential of the multimedia object, making it a powerful communication resource tailored to the needs of the times, really usable in a non-linear way, allowing its fast spreading throughout society for the common progress of knowledge. If technologies can not be the only ones decisive of freedom from property and economic barriers, they can and must be the simplifiers of access, and this will help disseminating digital information anyway. Documentation and Library and Information Science (LIS) have a primary focus in this, as specific task of their theoretical predictions and technical planning. In the area of information and knowledge management, LIS may very well foresee and plan the development and use of information technologies for universal *welfare*.[1]

So, in the contemporary panorama of *knowledge society* – much more advanced than the *information society* – Documentation and LIS have to ask many pressing questions on the potentialities and effectiveness of technologies and services for the knowledge organization and management, as well as questioning the adequacy of management systems for multimedia databases, digital libraries and archives, considering their large Web application.

In libraries, archives and museums, new tools for the organization and mediation of their increasing amount of multimedia digital resources are crucial. However, multimedia systems and services conception and architecture still reveal a contradiction in the organizational logic, despite the radical changes that have transformed documents in full multimedia

[1] In this direction is *The Lyon Declaration* (August 2014): http://www.lyondeclaration.org.

resources. If searching and retrieving a *written* document by means of visual or sonorous language is not possible, likewise retrieving documents consisting in *sounds* or *figures* through descriptive texts cannot be considered an effective method. On the contrary, it should appear a waste of time to look for the photo of a coloured landscape through a complicated word description of the desired tonalities, rather than submitting a sample of the colours to a special search system.

By the standpoint that Documentation and LIS shall have in considering the new society and new technologies, the limits of operating according to the logic and terms of a traditional Information Retrieval (IR) perspective should appear evident. In IR traditional practice every kind of document search is carried out under the conditions of a query in *textual* language, but by now it is necessary to define broader criteria for the Multimedia Information Retrieval (MIR). So, every kind of digital resource can be processed through the elements of language, or *meta-language*, appropriate to its own nature.

Within the general and *organic* methodology of the MIR can be distinguished: a system of Text Retrieval (TR), based on textual information for the processing and search of textual documents; a method of Visual Retrieval (VR), designed on visual data for the search of visual documents; a method of Video Retrieval (VDR), founded on audiovisual data for the processing of videos; and a criterion of Audio Retrieval (AR), based on sonorous data for the processing and the retrieval of audio documents.

This vision is actually suitable to the handling of multimedia documents for the improvement of the services to users. Thus, in databases where the content of the documents is substantially a text, using access keys that are terms and strings extracted – *from the inside* – from that same content is obvious and appropriate.

Instead, in databases of images or sounds, attributing – *from the outside* – a textual description to different contents appears simplifying and inaccurate. And moreover, though the method of analyzing concepts and attributing them a terminological descriptor is often suitable for texts, the same method for images or audiovisuals is not equally effective – since the subjective limits in gathering their intimate concepts are greater, and these are rather indescribable by terms.

The MIR system – as an holistic whole of the TR, VR, VDR and AR systems – is structured on the fundamental principles of a methodology of analysis and search based on the *content* of the documents, defined as Content Based Information Retrieval (CBIR).[2] Within the CBIR logic, analysis and search methods are defined as *content-based*. These are founded on the use of storage and retrieval keys of the same nature as the *concrete* content of the resources they are applied to. These keys are based on a language appropriate to every resource typology, able to point consistently to the concrete content, as well as to the meaning aspects of a certain document.

MIR theory and the Library and Information Science

Experimentation and use of MIR technologies are already well developed within computer engineering, artificial intelligence, computer vision, or audio processing fields, while the interest in the methodological and operational revolution of MIR, and the reflection on its conceptual development, still have to be introduced among librarians, archivists, documentalists and information managers. The LIS context still has the opportunity to welcome the discussion, addressing the general development

[2] In several interpretations CBIR is "Content Based Image Retrieval".

of MIR systems according to the LIS needs, at a time when MIR databases and interfaces are in the testing phase. This is a must for Documentation and LIS: developing this cultural and technological revolution by meeting the information and knowledge needs of the society, by interpreting problems in describing, classifying, indexing and retrieving documents and information in new systems.

In this *advocacy* of MIR reasons, some remarks by Sara Pérez Álvarez (2006) about the interest of the Documentation in CBIR are still very useful. The scholar writes that the goal of CBIR systems is the automation of all processes of analysis and search. CBIR aims at implementing an analysis and retrieval method considering simultaneously all facets of multimedia documents: those related to the *meaning* and those related to the *content*. From the LIS standpoint, therefore, the "joined" approach techniques are to be deepened, as they represent an ideal way for documentation processes: in fact they consider both the semantic and the formal nature of images, such as videos and sounds. How to adapt these document processing methods is a problem to be solved by qualified documentation professionals, as it regards the characteristics of the documents to be represented, the timing and quality of the response to the query, the users' information needs and their expectations.

Engineering research within the CBIR held more technical issues and algorithmic computation related to the content, while semantic issues, users behavior, dialogue interfaces, remain a LIS prerogative. Therefore – according to Pérez Álvarez – the Documentation science must lead the "human dimension" into CBIR studies, by focusing on users, their mental categories, their search strategies, and their overall needs when interacting with the systems. The whole body of knowledge and practices

belonging to Documentation, developed over time, plays a specific role in the multidisciplinary set that is the basis for the research on MIR. Only from this perspective, and disseminating the MIR vision, we can push forward the studies on MIR itself. From the early stages of CBIR in the Documentation field the need for a genuine alliance involving documentalists and engineers, and other experts, was felt, according to the principle of the convergence of skills (Cawkell 1993; Enser 1995, 2000).

Relying on these alliances, even today the most pressing issue is an ambitious, courageous and *utopic* experimentation – even risking to fail – to be performed in libraries, archives and museums. This must be very contagious, reaching documentation centers of radios, televisions, laboratories, industries and other really well equipped bodies, where there may be a great interest for applications and results of experimentation.

Analysis and indexing of digital multimedia documents

The ground of MIR and the content-based indexes

The Information Retrieval system, compared to the new conceptions of CBIR, is defined as the *term-based* system for indexing and searching. In the classical setting, a number of attempts have been made to evolve IR systems to the new needs of users and the requirements of multimedia documentation. These attempts often have resulted in highly complex and difficult solutions, that hardly succeed in managing today's information-searching panorama, also revealing an internal *crisis* in the existing system. The weakness these experimentations have in common are the difficulties to renew the textual retrieval

principles (Williamson and Beghtol 2003; Kovács and Takács 2014).

Only a content-based perspective will coherently approach the formal, dynamic, figurative, sonorous contents etcetera – without failing to consider the textual contents with the same coherence. The main criterion for the *contentual*[3] analysis of the documents is to directly found the means for handling and searching on the basis of the true content of each of them, be it text, figure, sound or a whole richly and variously combined.

If a *conceptual* IR system, relying on the development of a terminological culture, can be effective in processing a mainly textual set of documents, a *formal* search and retrieval system is rather determinant in the application to multimedia documents, founding upon the concrete perceptive abilities of every user.

A lot of query strings for multimedia databases, digital libraries, archives, museums, or also the Web, attempting to fully express users' information needs, aim to a search definition that goes over the details definable with precise terms constructions or with few elaborated sentences, pointing to qualities *proper* of the content. If the simplest queries, not specified about spatial compositions, actions, or expressive forms, can be satisfied in the area of term-based systems, more complex query strategies require a completion with further operations that, with the traditional methods and tools, not always bring about the results the user expects. A system of MIR is more helpful, since the query formulation does not have to be forced within the limits of the textual language, but it can be inputted as it is naturally

[3] In the Oxford English Dictionary "contentual" is: "belonging to, or dealing with, content".

produced, directly in visual, sonorous, audiovisual, and textual means.

This will be possible only by analyzing and indexing documents not according exclusively to the *terminologically* reportable or translatable data – semantically – but also by structuring a sort of index directly constituted by the *concrete* and *formal* data – contentually – of the documented objects. However, the concept of *indexing* in this context must be understood in its wider sense. It has to be referred to a methodology of creating the database index – and the documents' metadata in general – through extraction from non-textual documents of elements that are not terms and are not translatable into terms. A content-based index will be *made* of the data with which the machine operates for reproducing images, sounds, or words contained in the documents.

The sense of the problem can be schematized with a simple – very known and used – example of Visual Retrieval[4] (Enser 1995, 2008). A search system that forces to set terminological strings is not useful to someone who desires to retrieve images having a certain combination of forms and colours, remembered through *sensibility* without memory of the image typology, of the author or the title. Any combination of phrases will fail the retrieval goal, as it will go in circles around the presumed meaning of the desired image, and only the name of the author or the title of the work could help, as terms included in the indexing set. Indexing or classification data refer to another system setting, of an *intellectual* and specialist kind, and they seem to be abstract data relating to

[4] The entire no. 1, vol. 5 (2016), of the *International Journal of Multimedia Information Retrieval* is dedicated, as a special issue, to "visual information retrieval".

the image, useful only when they are known before the search (e.g. Figure 1).[5]

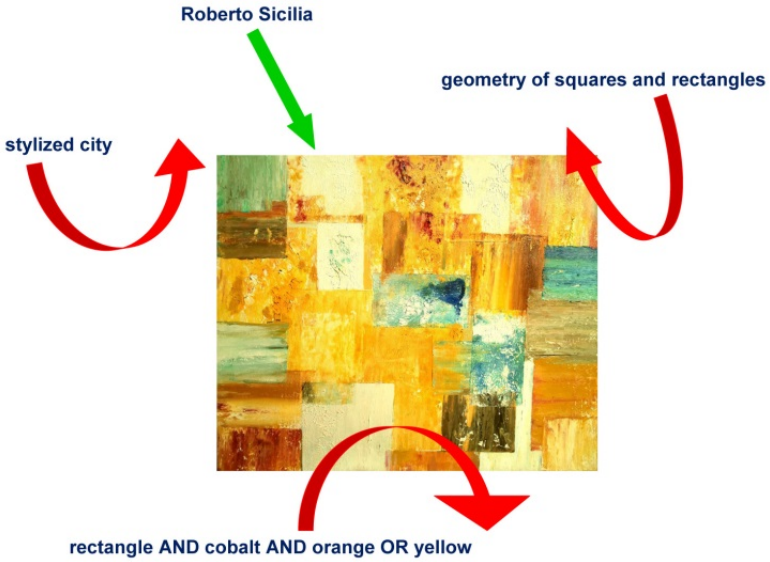

Figure 1: An example of textual-visual search

Otherwise, if the system can be searched by proposing the combination of textures or shapes and colours that the user imagines, or vaguely remembers, it is possible to go directly to the contentual *core* of the document concretized by the image (e.g. Figure 2).

[5] The painting in the examples is by Roberto Sicilia (*Città*, 1988, oil on canvas, 50x60 cm).

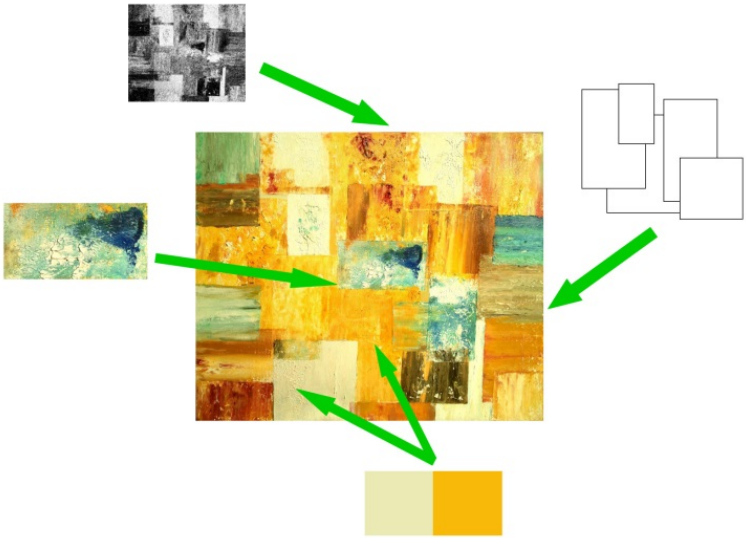

Figure 2: An example of visual-visual search

The precursors of MIR

Several authors have laid for many years the groundwork of the content-based handling problem for multimedia resources. Michael Buckland is among the researchers who first gave clear indications to revolutionize traditional formulations (Buckland 1991). In a fascinating recognition on the origins of the Documentation theory, the scholar shows as the attention to non-textual materials is not at all a recent issue, but it began with the first steps of the discipline. In the first half of the 20th century, Paul Otlet (1934) pointed out the need to define what the "document" is in technical meaning, as object of the "documentation", establishing that it – other than a simple book

or article of periodical – can also have the form of a "three-dimensional object", being the documental representation of any "expression of the human thought". In this respect, also Suzanne Briet (1951) is recalled. Briet explains, in a famous example, that if an "antelope" that runs in the forest cannot be a document, it soon becomes one as it is captured and exposed in a zoo, and a paper describing its characteristics can be considered only a derived and secondary document. These examples show well how to interpret as a document only the textual, or mainly textual, objects is restrictive.

Elaine Svenonius also is among the first researchers to understand the new problem of the indexing languages (Svenonius 1994). The scholar develops her theory starting from the *limits* of the conceptual indexing of "nonbook materials", and concentrates on the sense of creating a "subject indexing model". Svenonius wonders: "what then is a subject?", questioning on what really is the subject of a work or of a document, what is its meaning, and how is it possible to index it. The subject in IR methodologies has been often identified with what a work treats, with its "theme" that may be a concept, a situation or a thing. Indeed, a large part of the problems related to the indexing of multimedia materials rise from such immediate identification. For example, even if the subject of a picture is definable as the theme shown by the figures, this is really far from defining it with completeness. Visual or sonorous languages use expressive ways in which the subject is directly *implied* with the *materiality* of the forms, and in which it cannot be identified if abstracted from the contentual context – as it can happen with oral or textual language because of their mostly *intellectual* character.

Furthermore, the expressiveness of the work is concentrated essentially in the "thingness" of the whole object in which the

work consists, related to its concrete content, all in one with the *physicality* of the expressive medium: pictorial, sculptural, photographic, musical, or other kind. Therefore, if subject indexing is possible and profitable when a language – visual, audio, audiovisual, or also textual – is used for merely *documentary* purposes, to suppose that having verified the usefulness of such indexing in certain contexts this can be an effective method in all information contexts would be a mistake.

Principles and meaning of the MIR

Story and development of the art

In defining a methodology accorded to the *classical* parameters of information searching, Information Retrieval has always adopted a fundamentally *user-centred* perspective, focusing on the conceptual, interpretative and terminological ways with which users describe and handle every kind of information. So, Documentation has gone far from the physical, objective and formal principles of the automatic modalities of data organization, storage and retrieval. However, in the last twenty years, the growing importance of multimedia documents and the new tools offered by digital technologies have determined the creation of multimedia databases of higher complexity in comparison to traditional systems. For this reason, researches on the possibility to start a formal multimedia indexing, and especially on the deep and true nature of multimedia queries, have been developed to establish the best search techniques for the new multimedia digital libraries, archives and the Web.

Otherwise, the increasing use of IR both in commercial and scientific circles has restimulated the interests in the area of the Information Science that, unlike Library Science and

Documentation, has faced the various problems from a *computer-centred* perspective, defining processing and evaluation techniques for the raw constitutive data of documents' contents. In LIS it has been possible to combine *computer-centred* perspective without opposing the *user-centred* one, but considering all the interests of the *user*.

The debut of the CBIR, in the 1990s, was founded on image processing and on *computer vision* studies (Kato 1992; Del Bimbo 1999). Highly relevant in view of the beginnings is a Peter Enser's (1995) comprehensive essay. Enser analyzes theoretical and practical issues associated with the "pictorial information retrieval". He underlines that the majority of image databases, according to IR, are structured by "translating" into terms visual contents and their access keys. The scholar stigmatizes as a "sacrifice of the message in favour of the medium" this exclusively terminological processing of documents, which gives rise to a series of problems in representating and indexing the figurative content. The query, which must be expressed terminologically, can aim only at matching the textual "surrogates" of visual documents – subjects, keywords, index terms, titles or captions. Even when such search yields results, indexing all the terms required for describing an image will never be exhaustive, and often the *qualities* of a visual object do not fall into any linguistic category. So, a valid image retrieval system must be based on the CBIR logic, directly handling the visual content, surpassing conventional term-based treatment founded on descriptors.

John Eakins (1996) proposes one of the first frameworks for image retrieval, classifying visual queries into a series of levels of complexity. Then he discusses how new analysis and search systems can address users' needs at each level. Automatic CBIR

techniques can already meet many of these needs at the level of the "primitive features" search – color, shape, texture – and will soon be able to act at the level of the "logical attributes" – kind, typology, appearance. The scholar, anyhow, remains skeptical that CBIR systems can achieve a good role at the level of "abstract characteristics" – class, meaning.

William Grosky (1997) draws some general conclusions on this development, setting a synthetic theoretical definition. The researcher points out the principles of such a handling of multimedia data: a process allowing the transition from the "real objects", belonging to the world of the daily experience, to the "data models" of these objects. The content-based data model represents the properties of the things, their relationships and the operations defined over them, and such "abstract concepts", nevertheless, inside it are *translated* in digital data, physically situated in the database system. This way, through the data model mediation, queries and other operations referrable to the true objects and their context can be turned into operations on the abstract representations of such objects, and then these operations are turned into operations on the digital data translating the abstract representations in the *language* of the electronic system.

In the late 1990s, attention to video documents started an important progress in the handling of visual documents involving also movements, speaking and sounds, pushing research towards a more complex kind of multimedia documentation. A book by Frederick Lancaster (2003) treats the theme in a comprehensive way, recapitulating the whole IR possible development inside the term-based structure, until it reaches the content-based perspective. According to many authors dealing with CBIR theories – as Edie Rasmussen, Howard Besser or Sarah Shatford

Layne – Lancaster confirms, close to the importance of "word-based descriptions" for representing document characteristics of conceptual and semantic "high-level", the possibility to store and retrieve visual objects through "intrinsic features as colour, shape, and texture", characteristic elements of representational "low level". So, every search system makes available in "hybrid" way all the means that users require for planning a query, for still images and dynamic video documents, also without knowing a query vocabulary, also interrogating the fluctuating Web (Lancaster 2003, 215-233).

Exposing "sound databases" and "music retrieval" system issues, the scholar makes similar reasonings and reviews of studies and researchers – as Lie Lu, Stephen Downie and Donald Byrd. The objective of modern formulation of the music retrieval is: "answer music queries framed musically", that is to use the content-based method for searching sonorous pieces by sonorous elements (Lancaster 2003, 237-244).

At the beginning of the 21st century, investigating specific MIR matters has been possible, such as the improvement of processing algorithms able to calculate a huge number of variables. The way forward now is: constructing new specific and effective indexes of multimedia data; developing high-level analysis and query systems for large amounts of data; setting robust results evaluation and ranking systems also interacting with user specifications; and, finally, development of analysis and search paradigms able to relate the *automatic* objective representations of the machine with the *intellectual* sophisticated analysis by the human (Deb 2004; Gast *et al.* 2013).

The *evaluation* of such technology is an ultimate matter. To establish an utility-centred research focus is critical, bridging the so called "utility gap", or the distance between users' expectations

and real systems usefulness (Hanjalic 2012). Specific methods and protocols of evaluation set for MIR systems are necessary, allowing to appraise the advantages and the ineffectiveness of methods and systems, the user satisfaction related to procedures and results, and all the possibilities of development and improvement.[6]

Beyond this, since the information process effectiveness is largely influenced by the *interaction* of the operator with the system, a lot has to change also relating to the user, in sight of a greater friendliness, and of a smarter and faster satisfaction of information demands (Linckels and Meinel 2011). The whole system for approaching multimedia databases must be reset, on the basis of the demands to define the query also through visual and sound data, by operations developing in continuous interaction between human and computer. A branch of the researches on multimedia systems has to study the user behaviour, concrete needs and real search demands. Among studies about MIR effectiveness for users, a successful branch was the English one, in which the work of Peter Enser was predominant (Enser and Sandom 2003; Enser *et al.* 2005). Many researchers are occupied with analysis and diffusion of tests and surveys submitted in documentation centres, libraries or archives (Venters *et al.* 2004, 338-342), focused on verifying the usefulness of MIR interactive methods, and the *active learning* of the system arising from user's relevance feedback (Thomee and Lew 2012; Nikzad and Abrishami 2014).

Even these studies have brought CBIR researchers to stigmatize as "semantic gap" the discovered *semantic* ineffectiveness of

[6] See the web site of TREC Video Retrieval Evaluation: http://www-nlpir.nist.gov/projects/trecvid.

search systems, on the contrary, based only on the automatic content processing, which tend not to consider the level of the meaning. So, the semantic approach cannot be neglected by a content-based system, and a complete system of MIR must allow to develop every search with all the means that the user wants. A MIR system must *understand* the user's requests through both contentual and conceptual specifications, ability resumed in "bridging the semantic gap" (Enser 2008).[7]

Very relevant for the stabilizing and the growing significance of MIR studies is the foundation in 2012 of the *International Journal of Multimedia Information Retrieval*, aiming to present achievements both in semantic and in contentual processing of multimedia (IJMIR 2012). Anyway, one of the great challenges for the future is the need to move from the *academic* and experimental state of MIR systems to a practical and commercial phase, favouring cooperation between research and industry.[8]

Finally, the *commercial* successes in image and sound retrieval are to consider. Google Goggles is a smartphone app developed by Google labs, from around a decade, allowing someone to photograph or film objects and places to send a content-based query, getting a Google page with a list of related results (Google Goggles 2016). Google images, then, is the application of content-based tecnology to the common Google interface that reproposes, improved, a system already tried around 2000: the true novelty, compared to existing image search pages, is that

[7] See also the web site of the Semantic Media Network: http://semanticmedia.org.uk.

[8] Some laboratories and research groups are at least to be mentioned, and among them especially: MediaMill, http://www.science.uva.nl/research/mediamill; Viper, http://viper.unige.ch/doku.php/home.

someone can upload personal patterns and figures, and start searches using materials that are not already in an index.[9] The true commercial successes of content-based applications, however, are SoundHound (2016) and Shazam (2016), Audio Retrieval systems for years perfecting their application to mobile phones and smartphones, exploiting the widespread interest in the world of music to which they apply music recognition techniques.

Scopes, goals and effectiveness of MIR

Information Retrieval is a system for analysing and searching, through *terms*, mainly textual documents, which can be applied to visual, audio and video documents. Multimedia Information Retrieval is proposed as a general system for processing and retrieving, through *texts*, *images* and *sounds*, documents of every kind or full multimedia. Nevertheless, such a clarity is for a large part still to be reached.

In short, the MIR revolution is founded on the definition and application of a storage and retrieval technology that directly handles the *concrete content* of every document typology: using the same expression language of a given document, and employing processing and search modalities every time appropriate to its specific textual, visual, audio or audiovisual content, beyond any abstract mediation of a linguistic and intellectual kind.

Considering the significance of the *organic complex* of the four MIR specific methodologies – TR, VR, VDR and AR – to reach a good level of precision in multimedia documents retrieval all

[9] See the image search page at: https://www.google.it/imghp?hl=it&tab=wi&ei=MM3-VKDlCYn4Uqmug7AP&ved=0CBYQqi4oAg

modalities need to interact, inside a single system, according to a univocal principle. A single search interface is required, allowing a query formula which, combining images and texts, sounds and terms, is able to search very complex resources, whose contents extend beyond all the levels of *sense* and *meaning*, where semantic definitions do not have less importance than contentual characteristics (Menard and Smithglass 2014).

Since we have illustrated a simple example of VR (see section *The ground of MIR*), we have now to underline briefly the VDR and AR specificity. Video Retrieval resources processing has something in common with VR, but handling audiovisuals requires taking into consideration elements such as time, movements, transformations, editing, camera movement and, often, sound and text data. VDR processing runs by the extraction of *video-abstracts* characterized by spatio-temporal factors, supplemented by information on textual data relating to the written and the spoken in the video (Jiang *et al.* 2013).[10]

Audio Retrieval methods differ because an audio data stream is mainly connoted by *tempo-related* properties, and properties relating to frequency and sound characteristics such as tone, pitch, timbre, melody and harmony. In the audio resources processing, AR techniques have something in common with the whole MIR, but specialising under specific sonorous aspects. This even means working directly with contentual elements and

[10] The entire no. 1, vol. 4 (2015), of the *International Journal of Multimedia Information Retrieval* is dedicated, as a special issue, to "video retrieval".

concrete objects, as *ineffable* as sounds may seem, without excessive mediation of terms (Casey *et al.* 2008).[11]

Critical numerical matters

MIR systems show a series of open problems, with several consequences related to information searching and management (Lew *et al.* 2006; Mittal 2006). The main problem is always to develop the content-based method for the handling of any multimedia resource. The advantages of a more suitable system of document management have to be so evident that MIR will naturally replace the traditional IR architecture.

A major critical question, anyway, remains: related to the practical and individual *human* goals of information searching, what effectiveness can the icy *numerical* procedures of content-based systems have? The whole research for computational algorithms and data processing which can be not only mathematically *efficient* but also pragmatically *effective* actually tends to the overcoming of the distance between human and computer, taking into account information qualities expected by the human operator (Yoshitaka and Ichikawa 1999; Maybury 2012).

If the mechanical and absolute efficiency of the numerical processes can be certain, not the same can be said about their usefulness in answering the needs of every end user. The mathematical and direct operations automated by the computer are without the errors produced by human evaluation and mediation of documents and contents, but they are also deprived of the peculiar flexibility and intelligence of the human in

[11] The entire no. 1, vol. 2 (2013), of the *International Journal of Multimedia Information Retrieval* is dedicated, as a special issue, to "hybrid music information retrieval".

interpreting aspects not objectively evident. However, content-based and excessively numerical methods are not always really appropriate to satisfy the increased demands of scholars and experts, such as common users. If MIR systems show a certain validity in the case of a direct and *contentual-objective* approach to the document, they present a certain narrowness in the case of a theoretical and *intellectual-interpretative* approach.

The *sense* of an object represented in a document, indeed, has to be gathered in its true *totality*: in the simultaneous consideration of its several sensible and intellectual qualities. The interpretation of a multimedia object has a considerable value in the search process when information demands go beyond the *perceptive* characteristics of the object – automatically calculable by the computer – and reach the level of the *semantic* realization – definable only by the human. The content-based query needs to be knowledge-assisted: which means that the user has to query the system also with a subjective description of the information demand. Consequently, the use of semantic terms created by the human operator can be very useful to show both to the user and to the system what the mathematical analyses of an example model cannot directly gather.

Reconciliation between semantic and contentual principles

The main critical issues raised by the MIR possible innovation may meet in a conclusive matter. Establishing that there will never be an ultimate solution for the *contradictions* and the *gaps* of the relationship between the cognitive and cultural demands of the human and the numerical and automatic responses of the system, it is possible at least to define a perspective of *collaboration* between the information seeker and the tools to analyze and search for the information itself. The solution for the conflict

between *conceptual* and *concrete* means of accessing to information – or between term-based and content-based systems of processing – can be only a solution of *organic integration* among the principles and the methodologies of analysis, search and retrieval that constitute the only apparently incompatible semantic and contentual areas.

A large part of the international literature indicates as *semantic gap* the distance between the *high-level* conceptual-semantic representation of an object – proper of human knowledge – and the *low-level* formal-contentual denotation – belonging to the machine automatic processes. The semantic gap is defined as the *not coincidence* between the information that can directly be drawn out from a document and the different interpretation that the same data can receive by every user in every specific situation. This is a very critical matter for MIR development: since the meaning of a multimedia resource is rarely explicit, the system purpose is to help overcome the *void* between the simplicity of the document processing offered by the computer and the rich semantic expectations of the user.

The representative levels of a document vary from the lower level, composed by the simple extraction of its *raw data* immediately taken by the computer, up to the higher level, constituted by the *semantics* that it carries as they are realized by users. Users come to the higher level formulating requests of documents with an intellectually refined value, endowed with attributes of meaning assigned thanks to a cultural context of reference, impossible to identify without the semantic-terminological support (Hare *et al.* 2006). The traditional IR systems actually deal with this kind of searches, with all the limits of the conceptual abstraction, but this informative level is the

most difficult to reach for content-based systems, founded on the semiotic consideration, more than semantic, of the document.

A widely proposed solution for *bridging* the gap is considering the use of the guides for navigating in the Semantic Web: the *ontologies*. However improved, an ample set of annotations and data related to a resource is far from representing it in its semantic richness, which seems, instead, to be representable positioning the resource within an ontology (Hare *et al.* 2006). The appeal to ontologies in MIR systems, therefore, makes it possible to explicitly state part of the meaning of a document, and this enables to formulate the query also through concepts, continually integrating the content-based search tools that revolve around the objects immediately *seen*. This way, the multimedia query can be semantically completed, since ontology tools are able to represent both the meanings of the objects with their relationships in a document, and the meaning of the whole document in a context (Mallik and Chaudhury 2012).

Integrating ontologies in MIR systems, nevertheless, a certain *rigor* seems to be residual in these conceptual tools, and it can propose again the problem of the rigidity and the abstractness of the typical IR schemes. To avoid such a risk, ontologies can be combined with the *folksonomies*, and tags directly assigned by end-users. Folksonomies represent an important element of comparison, since they are often valid cues for metadata definition or for information-search strategies. In this direction goes a discussion started by the same founders of the Web 2.0, the Semantic Web, and related organizing structures (Shadbolt *et al.* 2006; Guy and Tonkin 2006; Yang 2012).

Following MIR principles, every user has the possibility to search freely, allowing the system to *learn* on the spot new information about the searched resources, integrating and widening its

interpretative abilities. The integration between the semantic tools of ontologies and folksonomies, contemporarily integrated into the content-based tools of CBIR, can bring to the conciliation of many oppositions between the principles of the semantic-interpretative and the contentual-objective information handling, in the general organicity of the MIR.

Introducing MIR theory and methodology

Synthesis of the foundations

Debated the reduced effectiveness of an exclusively terminological search metodology applied to the new and advanced multimedia databases a difference of principles can be highlighted between the IR and the MIR. It is clear, now, in what sense MIR methodologies, coherent with the concrete *content* of the handled documents, are defined as content-based, as opposed to the traditional systems founded upon terminological *descriptors* of such content, named term-based.

This does not imply, however, the rejection of the *conceptual* interpretation and representation of the documented content and of the document. Considering the semantic limits of the content-based system, an appropriate intellectual intervention in the organization and search for documents is necessary, to define the *meanings* beyond the *feelings*, to specify the query strategies and to increase the retrieval possibilities.

It is necessary to define an *organic* approach integrating *contentual* and *semantic* ways to documents: this approach will always be valid for all kinds of multimedia resouces, take into account univocally their concrete and conceptual representability and accessibility, and consider contentual-objective and intellectual-interpretative information needs. Documents, of whatever true nature they be,

can be always inserted in logical interrelated spaces, to be searched without influences inside such *semantic* positions with the *semiotic* methods appropriate for each one. [12]

The more advanced MIR systems can be very useful in supporting both theoretical research and creative practice, as a tool for professionals or a guide for general users (Beaudoin 2016). Users can always fully resort to their own *intelligence* and *sensibility*, to their own *creative* abilities and *imagination*, interacting with a system inclined to welcome unpredictable variations of the search way and to *understand* the human strategy, learning from the seeker's behaviour.

Concerning the organic complex of MIR methodologies, in order to reach a good level of reliability, the coexistence of all retrieval modalities is essential. The different procedures operate better in continuous and organic interaction, in a single and *holistic* query interface (Ah-Pine *et al.* 2015). The new systems need to be prepared to accommodate together all traditional and innovative solicitations, of IR, MIR and Semantic Web: from the descriptive and conceptual, to the contentual and semantic ones – to the *comprehensive* ones of linked data. Allowing several search strategies – combining terms, concepts, words, figures, movements, sounds, classes and codes – is critical for searching very complex resources, whose *knowledge content* extends throughout all levels of sense and meaning.[13]

---

[12] An introduction to the paradigms experimented for planning and applying MIR systems, and more technical specifing, are in an author's book (Raieli 2013, 134-171).

[13] The example figure is composed with images of the *Madonna Sistina* by Raffaello Sanzio (1513-14, oil on canvas, 265x196 cm), a photo of the actor

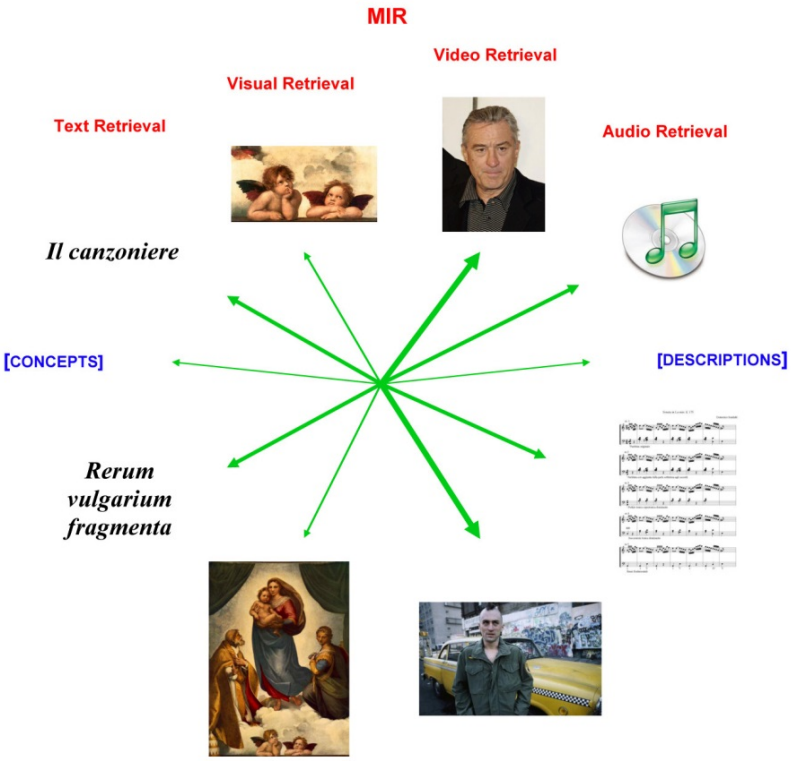

Figure 3: The MIR organic system

Conclusions and relaunch

If, in the current perspective, the MIR ultimate aim is the *automatic* contentual and semantic retrieval of multimedia information and resources, bridging the semantic gap is still the

---

Robert De Niro, and a shot with Robert De Niro from the film *Taxi driver* (1976) by Martin Scorsese.

main challenge (Jiang *et al.* 2016; Tan and Ngo 2016). To build this bridge, it will be necessary to develop machines able to reach high-level meanings starting from low-level features, and to set algorithmic processes also capable of simulating the connections of the human brain (Xu and Wang 2015).[14] Hypothesizing that automatic systems can reach the refined semantic-interpretative level proper to human beings, however, is quite difficult. This cognitive level is largely *logical*, but it is enhanced by ineffable, or *tacit*, knowledge elements, by inexplicable intuitions, by perceptive emotions. The gap between human and machine, in essence, remains, and it can be addressed only through the organic *collaboration* of the two ultimately different views.

Referring to the general possibilities of the Semantic Web, one thing is to construct logical connections among text strings of *linguistic* meanings, another thing is to interpretate, not only formally and logically, but also *emotionally*, multimedia resources. Thus, even the use of ontologies and linked data, only partially bridges the chasm between appearance and essence of a multimedia object, placing the object in a useful class, but never completely discovering the enigma of its *true* interpretation.

All this must not remain in the dimension of *utopia*, it is achievable through a real interest in applying and disseminating content-based technologies. So, even the complex MIR systems – as the systems for navigating in the Semantic Web – can be transformed from *élite* instruments into common technological tools used by the masses. It is just the user interface that has the task of transforming computer language into a language

[14] About this issue, more than ten years old, see the Vol. 4 (2015), no. 2, of *the International Journal of Multimedia Information Retrieval*, dedicated, as a special issue, to "concept detection with big data".

understandable by common people, without any loss of effectiveness in information handling (Castellucci 2004).

Developments in society towards knowledge intended as a *commons* have necessitated the commonality of information systems and resources, have made imperative technological *democratization* of access. For this reason we have already to think beyond the Semantic Web, where also the spirit of the *semiotic* access, immediately intuitive, sensitive, has a definite place, to favor the approach to knowledge of increasingly wide circles of citizens, even if they have little possibilities for studying or developing intellectual attitudes.

References[15]

Ah-Pine, Julien, *et al.* 2015. "Unsupervised visual and textual information fusion in CBMIR using graph-based methods." *ACM Transactions on Information Systems* 33 (3): 1–31.

Beaudoin, Joan E. 2016. "Content-based image retrieval methods and professional image users." *Journal of the Association for Information Science & Technology* 67 (2): 350–365.

Briet, Susan. 1951. *Qu'est-ce que la documentation?* Paris: EDIT.

Buckland, Michael K. 1991. "Information Retrieval of more than text." *Journal of the American Society for Information Science* 42 (8): 586–588.

Castellucci, Paola. 2004. "George Boole: il pensiero dietro la maschera." In *L'organizzazione del sapere: studi in onore di Alfredo Serrai*, ed. by M. T. Biagetti, 55–69. Milano: Bonnard.

Casey, Michael A., *et al.* 2008. "Content-based Music Information Retrieval: current directions and future challenges." *Proceedings of the IEEE* 96 (4): 668–696.

Cawkell, Antony E. 1993. *Indexing collections of electronic images: a review*. London: British Library.

Deb, Sagarmay. (ed.). 2004. *Multimedia systems and Content-Based Image Retrieval.* Hershey: Idea Group.

Del Bimbo, Alberto. 1999. *Visual Information Retrieval.* San Francisco: Kaufmann.

Eakins, John P. 1996. "Automatic image content retrieval: are we getting anywhere?" In *Proceedings of third International Conference on Electronic Library and Visual Information Research*, 123–135. London: Aslib.

[15] Websites last accessed 30 april 2016.

Enser, Peter G. B. 1995. "Pictorial information retrieval: progress in documentation." *Journal of Documentation* 51 (2): 126–170.

---- 2000. "Visual image retrieval: seeking the alliance of concept-based and content-based paradigms." *Journal of Information Science* 26 (4): 199–210.

----. 2008. "Visual image retrieval." *Annual review of information science and technology* 42 (1): 1–42.

Enser, Peter G. B., and Christine J. Sandom. 2003. "Towards a comprehensive survey of the semantic gap in visual image retrieval." In *Image and Video Retrieval: CIVR 2003 Proceedings*, 291–299. Berlin: Springer.

Enser, Peter G. B., *et al.* 2005. "Surveying the reality of semantic image retrieval." In *8th International Conference on Visual Information Systems Proceedings*, 177–188. Berlin: Springer.

Gast, Erik, *et al.* 2013. "Very large scale nearest neighbor search: ideas, strategies and challenges." *International Journal of Multimedia Information Retrieval* 2 (4): 229–241.

Google Goggles. 2016. https://support.google.com/websearch/answer/166331?hl=it&ref_topic=25275.

Grosky, William I. 1997. "Managing multimedia information in database systems." *Communications of the ACM* 40 (12): 73–80.

Guy, Marieke, and Emma Tonkin. 2006. "Folksonomies: tidying up tags?" *D-Lib Magazine* 12 (1). http://www.dlib.org/dlib/january06/guy/01guy.html.

Hanjalic, Alan. 2012. "New grand challenge for Multimedia Information Retrieval: bridging the utility gap." *International Journal of Multimedia Information Retrieval* 1 (3): 139–152.

Hare, Jonathon S., *et al.* 2006. "Mind the gap: another look at the problem of the semantic gap in image retrieval." In: *Proceedings*

of Multimedia Content Analysis, Management and Retrieval 2006, 75–86. San Jose: SPIE.

International Journal of Multimedia Information Retrieval: IJMIR (2012–). London: Springer. http://link.springer.com/journal/13735.

Jiang, Lu, *et al.* 2016. "Text-to-video: a semantic search engine for internet videos." *International Journal of Multimedia Information Retrieval* 5 (1): 3–18.

Jiang, Yu-Gang, *et al.* 2013. "High-level event recognition in unconstrained videos." *International Journal of Multimedia Information Retrieval* 2 (2): 73–101.

Kato, Toshikazu. 1992. "Database architecture for content-based image retrieval." In: *Image Storage and Retrieval Systems: SPIE Proceedings vol. 1662*, 112–123. San Jose: SPIE.

Kovács, Béla Lóránt, and Margit Takács. 2014. "New search method in digital library image collections: a theoretical inquiry." *Journal of Librarianship and Information Science* 46 (3): 217–225.

Lancaster, Frederick W. 2003. *Indexing and abstracting in theory and practice*. Urbana Champaign: University of Illinois.

Lew, Michael S. 2006. "Content-based Multimedia Information Retrieval: state of the art and challenges." *ACM Transactions on Multimedia Computing, Communications and Applications* 2 (1): 1–19.

Linckels, Serge, and Christoph Meinel. 2011. *E-librarian service: user-friendly semantic search in digital libraries*. Berlin: Springer.

Mallik, Anupama, and Santanu Chaudhury. 2012. "Acquisition of multimedia ontology: an application in preservation of cultural heritage." *International Journal of Multimedia Information Retrieval* 1 (4): 249–262.

Maybury Mark T. 2012. *Multimedia information extraction*. New York: Wiley-IEEE.

Menard, Elaine, and Margaret Smithglass. 2014. "Digital image access: an exploration of the best practices of online resources." *Library Hi Tech* 32 (1): 98–119.

Mittal, Ankush. 2006. "An overview of multimedia content-based retrieval strategies." *Informatica* 30 (3): 347–356.

Nikzad, Mohammad, and Hamid Abrishami Moghaddam. 2014. "An incremental evolutionary method for optimizing dynamic image retrieval systems." *International Journal of Multimedia Information Retrieval* 3 (1): 41–52.

Otlet, Paul. 1934. *Traité de documentation*. Bruxelles: Mundaneum.

Pérez Álvarez, Sara. 2006. "Aproximación al estudio de los sistemas de recuperación de imágenes 'CBIR' desde el ámbito de la Documentación." *Documentación de las ciencias de la información* 29: 301–315.

Raieli, Roberto. 2013. *Multimedia Information Retrieval: theory and techniques*. Oxford: Chandos.

Shadbolt, Nigel, *et al.* 2006. "The Semantic Web revised." *IEEE Intelligent Sistems* 21 (3): 96–101.

Shazam. 2016. http://www.shazam.com.

SoundHound. 2016. http://www.soundhound.com.

Svenonius, Elaine. 1994. "Access to nonbook materials: the limits of subject indexing for visual and aural languages." *Journal of the American Society for Information Science* 45 (8): 600–606.

Tan, Chun-Chet, and Chong-Wah Ngo. 2016. "On the use of commonsense ontology for multimedia event recounting." *International Journal of Multimedia Information Retrieval* 5 (2): 73–88.

Thomee, Bart, and Michael S. Lew. 2012. "Interactive search in image retrieval: a survey." *International Journal of Multimedia Information Retrieval* 1 (2): 71–86.

Venters, Colin C., *et al.* 2004. "Mind the gap: Content-Based Image Retrieval and the user interface." In *Multimedia systems and Content-Based Image Retrieval*, ed. by S. Deb, 322–355. Hershey: Idea Group.

Williamson, Nancy J., and Clare Beghtol (eds.). 2003. *Knowledge organization and classification in international Information Retrieval.* New York: Haworth.

Xu, Lei, and Xiaoguang Wang. 2015. "Semantic description of cultural digital images: using a hierarchical model and controlled vocabulary." *D-Lib Magazine* 21 (5/6). http://www.dlib.org/dlib/may15/xu/05xu.html.

Yang, Sharon Q. 2012. "Tagging for subject access." *Computers in libraries* 32 (9): 19–23.

Yoshitaka, Atsuo, and Tadao Ichikawa. 1999. "A survey on content-based retrieval for multimedia databases." *IEEE Transactions* 11 (1): 81–93.

RAIELI ROBERTO, Università Roma Tre. roberto.raieli@uniroma3.it.

Raieli, R. "Introducing Multimedia Information Retrieval to libraries". JLIS.it. Vol. 7, n. 3 (September 2016): Art: #11530. DOI: 10.4403/jlis.it-11530.

ACKNOWLEDGMENT: The author acknowledges with thanks Matilde Fontanin, University of Trieste, for her help in reviewing the English version of the paper.

ABSTRACT: The paper aims to introduce libraries to the view that operating within the terms of traditional Information Retrieval (IR), only through textual language, is limitative, and that considering broader criteria, as those of Multimedia Information Retrieval (MIR), is necessary. The paper stresses the story of MIR fundamental principles, from early years of questioning on documentation to today's theories on semantic means. New issues for a LIS methodology of processing and searching multimedia documents are theoretically argued, introducing MIR as a holistic whole composed by content-based and semantic information retrieval methodologies. MIR offers a better information searching way: every kind of digital document can be analyzed and retrieved through the elements of language appropriate to its own nature. MIR approach directly handles the concrete content of documents, also considering semantic aspects. Paper conclusions remark the organic integration of the revolutionary contentual conception of information processing with an improved semantics conception, gathering and composing advantages of both systems for accessing to information.

KEYWORDS: Multimedia Information Retrieval; Content-Based Image Retrieval; Content description; Multimedia documents; Semantic gap.

Date submitted: 2015-09-28

Date accepted: 2016-03-06

Date published: 2016-09-15