Distributed Information Retrieval and Automatic Identification of Music Works in SAPIR

M. Agosti, E. Di Buccio, G.M. Di Nunzio, N. Ferro, M. Melucci, R. Miotto and N. Orio

Department of Information Engineering, University of Padua, Italy
{agosti, dibuccio, dinunzio, ferro, melo, miottori, orio}@dei.unipd.it

1 The Context

The Search in Audio-visual content using Peer-to-peer Information Retrieval (SAPIR) project\(^1\) is an EU IST FP6 research project that aims at developing theories and technologies for the next-generation search techniques, which will effectively and efficiently deliver relevant information from exponentially growing distributed and very dynamic multimedia databases.

The SAPIR consortium includes experts from industry and academia, and it will provide major innovations for powerful peer-to-peer (P2P) search on audio-visual content. It will be based on a scalable, completely decentralized, largely self-organizing P2P system where peers act both as client and servers and the users produce audio-visual content using multiple devices – ultra-peers may act as service providers which maintain indexes and provide search capabilities. The Information Management System research group of the Department of Information Engineering of the University of Padua is partner of the SAPIR consortium and it is mainly concerned with (1) media analysis and enrichment for search, and (2) complex search, retrieval and ranking in distributed environments, such as P2P networks. This work reports on the results that have been reached so far and that have been mapped on an initial prototype.

2 Modeling Distributed Retrieval

The representation of the uncertain nature of the retrieval process and of the routing mechanisms which characterize a distributed system is a key issue when modeling a distributed and heterogeneous information retrieval (IR) system, such as the one in the P2P network designed within SAPIR. The uncertainty is basically caused by the very limited knowledge about the documents and collections which store the data relevant to the information needs represented as free multimedia queries by the end users.

A probabilistic model would naturally allow for dealing with uncertainty. In addition, the multimedia character of the data requires a general probabilistic

\(^1\) https://sysrun.haifa.il.ibm.com/sapir/index.html
model. These two requirements, i.e. uncertainty and heterogeneity, can be met if the layered approach proposed in the model reported here can be adopted. When a probabilistic model is adopted for modeling IR, three layers can be distinguished: event space, representation and description [3]. The layered approach allows for separating conceptual modeling (event space) from representation and description. This separation allows for defining different event spaces, different representations for a single event space, and different descriptions for a single representation.

The event space contemplates the set of the original resources of interest – in the context of SAPIR, these resources are documents, peers, ultrapeers and queries. The set of resources are then combined for defining different event spaces accordingly to the architecture of the P2P network – for example, the hybrid network architecture would comprise three levels of resources: media objects, peers and ultrapeers. Two event spaces are currently studied within SAPIR. The first event space is a set of tuples of resources at different levels; this approach allows for modeling routing mechanisms as well as the distribution of information across many peers. The second approach consists of defining independent event spaces for different resource levels; in this way an event can be represented with a higher number of degrees of freedom.

An IR system represents these entities because of its limited computational capabilities for understanding the content. This is the reason for introducing the representation and description layers. The representation layer is implemented as random variables on the event space of the conceptual layer, while a description level is implemented as estimators of the random variables. Representation and description may depend on the medium. This layered approach extends the previous work on P2P-IR reported in [4], and it is crucial for addressing the retrieval of multimedia and distributed information. An important example of a non-standard medium is music, which plays a relevant role within SAPIR; Section 3 illustrates some issues concerning music IR and the prototype being developed within the project.

3 A Prototype System for Automatic Music Identification

The application of automatic identification of music works ranges from digital right management to automatic metadata extraction, and to music access and retrieval. Different techniques of automatic identification could be exploited for music IR; all of them should be robust to distortions, additional noise, A/D and D/A conversions, and compressions.

The first, common approach is audio fingerprint directly from a recording in digital format [2]; a fingerprint is a unique set of features that can be seen as a content-based signature that summarizes an audio recording. Audio fingerprinting systems are normally not able to generalize the features and to identify different performances of the same music work. On the other hand, the identification of a music work may also be carried out without linking the process to a particular performance. The music identification of broadcasted live performances,
for instance, may not benefit from the fingerprints of other performances, since most of the acoustic parameters may be different.

An alternative approach is *audio watermarking* which exploits research on psychoacoustics for embedding an arbitrary message—the watermark—in a digital recording without altering the human perception of the sound [1]. The message can provide metadata about the recording (such as title, author, performers), the copyright owner, and the user that purchases the digital item. On the other hand, watermarking techniques require that the message is embedded in the recording before its distribution, a situation that can be applied only on newly released material.

The collection of labeled works is presented to the system for the off-line extraction of some features. These are stored in a database and can be linked to a tag or other meta-data relevant to each recording. At the identification step, the unlabeled audio is processed in order to extract some features that are compared somehow with the features stored in the database. A list of the most likely matches is then returned at the end of the process. In a distributed P2P architecture, the database could be figured out as spread among the different peers. Each peer should store a part of the whole database and then compute a partial match with the unknown recording. The databases could be distributed in order to collect together similar recordings, following a set of criteria (music genre, author, etc.). The ultra-peer will then compute the final rank, by elaborating the ranks returned by each peer. Such a solution will give high scalability and parallelism to the model thus increasing considerably the performance in term of speed and robustness.

A novel methodology for the identification of music works from the recording of a performance, yet independently from the particular performance has been proposed and described in [5]. The methodology is based on an *audio-to-audio* matching where identification is performed using hidden Markov models. The approach can be considered a generalization of audio fingerprinting, because the relevant features used for identification are not linked to a particular performance of a music work. This methodology extends previous work on music identification based on audio to score matching [6], where performances were modeled starting from the corresponding music scores. The methodology fits with a distributed architecture because of its high interoperability.

A prototype system has been designed and developed in order to test the architecture. All the routines have been written in ANSI C++, with no particular optimization for the particular machine and for a distributed architecture. Tests on the computational time have been carried out on a laptop computer, with an Intel CPU Centrino Dual Core at 1.66 GHz, with 1 GB of RAM, mounting a Microsoft Windows XP operative system. Tests have been done figuring out a situation of a peer storing a part of the database composed of 200 recordings, all of them representing tonal Western music. Final results returned the 90% of the analyzed recordings ranked among top 3 positions, and the 66% of them were correctly identified. With this quite low level equipment, the average time needed to identify an audio excerpt was about 65 seconds, meaning that the
identification task for each recording of the database was about 0.3 seconds. These results may show that the identification task becomes feasible also when larger collections of competing audio recordings are used.

A user-friendly graphical interface have been developed on the top of the identification routines for providing the users with a tool to identify unlabeled recordings. For each unknown audio, the users are presented with a list of candidate recordings which are sorted for relevance and may be played for verifying if they have been correctly identified, an example is reported in Figure 1. The prototype can be useful for organizing the content of personal music collections, which is the reason because the methodology has been tested using a simple hardware/software configuration.

Acknowledgements

The work reported in this work has been partially supported by the SAPIR project, as part of the Information Society Technologies (IST) Program of the European Commission (Contract IST-045128).

References