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, started in January 2007. 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) complex search, retrieval and ranking in distributed environments, such as P2P networks and (2) media analysis and enrichment for search. 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 [3]. 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 model — an important example of a non-standard medium is music, which plays a relevant role within SAPIR and is addressed in Section 4 where a module of the prototype developed within the project is illustrated.

[1] https://sysrun.haifa.il.ibm.com/sapir/index.html

Uncertainty and media 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 [2]. 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.

3 Architectural Approach to Distributed Retrieval

The approach to modeling multimedia distributed retrieval introduced in the previous section has been the starting point for designing a software architecture, called SPINA (Superimposed Peer Infrastructure for iNformation Access). SPINA aims at being independent of both the underlying network infrastructure and the media of the documents stored in the network; it is focused on exchanging statistics about the features extracted from the indexed documents and aggregating the resources according to the level hierarchy; it selects peers and routes query by the degree of belief that a peer or a document store relevant information. Finally, it aims at integrating the software mod-

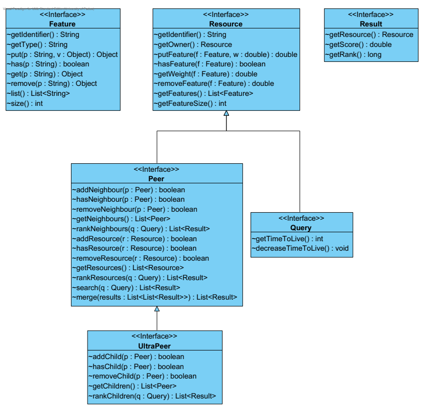

Fig. 1. SPINA conceptual architecture.

ules developed by us in SAPIR. The current status of the design of SPINA is depicted in Figure 1. In particular:

- `Resource`: represents the basic unit carrying information, i.e. generic multimedia content, and is a container for `Feature`s together with their weights;
- `Peer`: is a container of `Resources`. A `Peer`: is capable of ranking its own `Resource`s with respect to a given `Query` and providing information about their `Feature`s; may have neighbour `Peer`s; is capable of ranking its own neighbours with respect to a given `Query`; may be owned by another `Peer` which actually is an `UltraPeer`;
- `UltraPeer`: manages and organizes a group of `Peer`s and is capable of rank them with respect to a given `Query`.

4 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. A common approach is to compute an *audio fingerprint* directly from a recording in digital format [1], with the goal of providing the user with metadata about the recording – title, composer, year, genre, and so on. The prototype presented in this demo extends the concept of audio fingerprint, because the system is able to generalize the information of an audio recording and to recognize alternative versions of the same recording.

The approach is based on previous work on music identification on audio to score matching [5] and audio to audio matching [4]. As for fingerprinting, the methodology is based on a collection of labeled works, stored in a database, and tagged with relevant metadata. For each work in the database, an Hidden Markov Model (HMM) is automatically built, where each state in the HMM is labeled by a musical event of the musical work (notes, chords, rests). The transition probabilities model the fact that musical events are ordered in time, which correspond to a topology which is called *left to right* because only self-transitions – modeling the duration of an event – and transitions towards the next event are allowed. This is an important characteristics, which from quadratic becomes linear in the number of states.

Labeling states with musical events allows us to define their observation probabilities according to the characteristics of these events. In particular, the spectrum of the audio signal to be recognized has been statistically modeled considering that each musical event is related to the presence of peaks at particular frequency bands, which correspond to the first harmonics of the musical events. The approach can be exemplified considering that, for each event, a bank of bandpass filters is computed, each filter centered on the expected harmonics of the events. The observations are then computed measuring the amount of energy output by the filterbank divided by the overall energy.

Identification is carried out using standard techniques, based on the use of forward probabilities. Alternative approaches have been tested, as described in [5], taking into account the particular paths across the HMM states, yet the use of forward probabilities is the best compromise between computational cost and identification results. Experimental results with a collection of 50 unknown recordings and a database of 200 recordings showed that 90% of the analyzed recordings were ranked among top 3 positions, and the 66% of them were correctly identified. Moreover, only 4% returned the correct match after the first 10 positions. The mean average precision for all the 50 recordings was 78.9%.

A prototype system has been designed and developed to test the architecture. A user-friendly graphical interface have been developed on the top of the identification routines. The current demonstrator, which will be installed on a laptop computer, allows us to carry out an identification task in about 0.3 seconds. After having uploaded the system with an unknown music file, the user is presented with a list of candidate recordings which are sorted in order of relevance and may be played for verifying if they have been correctly identified. The main interface is reported in Figure 2. Current research is oriented towards the integration of the standalone prototype with the distributed P2P architecture, by spreading the database among the peers, which compute a partial match with the unknown recording. Such a solution will give high scalability and parallelism to the model thus increasing considerably the performance in term of speed and robustness.

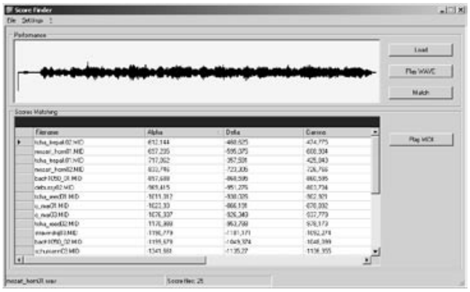

Fig. 2. The prototype system.

Acknowledgments

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

1. P. Cano, E. Batlle, T. Kalker, and J. Haitsma. A review of audio fingerprinting. *Journal of VLSI Signal Processing*, 41:271–284, 2005.
2. N. Fuhr. Probabilistic models in information retrieval. *The Computer Journal*, 35(3):243–255, 1992.
3. M. Melucci and A. Poggiani. A study of a weighting scheme for information retrieval in hierarchical peer-to-peer networks. In *Proc. of the European Conf. on IR Research (ECIR)*, LNCS 4425, Springer, pages 136–147, 2007.
4. R. Miotto and N. Orio. Recognition of music performances through audio matching. *Italian Research Conf. on Digital Library Systems*, 2007.
5. N. Orio. Musifind: a System for the Automatic Identification of Music Works. In *Atti del XV Convegno Nazionale su Sistemi Evoluti per Basi di Dati*, 2007.