

# **SAPIR: Towards Large Scale Multimedia Content Search**

Maristella Agosti<sup>1</sup> and Fausto Rabitti<sup>2</sup>

<sup>1</sup> Dipartimento di Ingegneria dell'Informazione,  
University of Padua,  
Padua, Italy

<sup>2</sup> Istituto di Scienza e Tecnologie dell'Informazione (ISTI),  
Consiglio Nazionale delle Ricerche (CNR),  
Pisa, Italy

Existing web search technologies are limited to text-based search, yet still 99% of the information on the web consists of audio-visual content that is searchable only by associated metadata and not by its actual content. This amazing restriction has raised the question of how search technologies can tap into the potential reservoir of information. SAPIR (*Search on Audio-visual content using Peer-to-peer Information Retrieval*) has developed cutting-edge technology to break the barriers and enable search engines to *search large-scale, audio-visual information*, by content. SAPIR's distributed P2P technology has proven to be able to effectively deal with the fundamental scalability issue.

The main aim of SAPIR has been to develop theories and technologies for next-generation search techniques that would effectively and efficiently deliver relevant information in the presence of exponentially growing (i.e., dynamic) volumes of distributed multimedia data. Fundamental to our approach is the development of scalable solutions that address the requirements of future generations of massively distributed data produced in a variety of applications and available on the Web.