

Concepts for Multimedia Database Exploitation

Richard C. Price

Command and Control Division,
Defence Science and Technology Organisation
PO Box 1500, South Australia, 5108
Email: richard.price@dsto.defence.gov.au

Extended Abstract

The information age is offering new and exciting opportunities for many commercial and Government enterprises. Organisations are now in a position to both receive and collect large volumes of information relevant to their nature of business. New challenges are now emerging, as organisations attempt to find and exploit the important intelligence potentially contained within their large volumes of collected data. A critical factor in an organisations ability to fully exploit this information is their information management policy, practices and procedures.

The first section of this paper proposes an architecture designed to enable an enterprise manage and exploit the increasingly large volumes of information typically held in corporate repositories. Most of the information needing to be processed and analysed is unstructured free text, although increasingly this may include large volumes of audio and video information. Whilst the paper will address techniques for the clustering and visualisation of documents existing in an unstructured form, generally speaking, in order to perform the most valuable forms of analysis, the information will need to be structured prior to the analysis process. In this paper we will discuss the advances being made in the technologies developed to assist the structuring process. In the text domain, the paper will introduce the topic of information extraction and discuss the current research issues currently being addressed both in DSTO and the broader research community. The paper will also discuss DSTOs research activities in the extraction of metadata from audio, concentrating on speaker identification, language identification, gender identification and word spotting.

Once an organisation has established a solid information management foundation, which includes the existence of well maintained structured corporate knowledge bases, it is now well placed to employ the emerging advanced technologies for the visualising and mining of their information.

Finally, the paper will present and discuss a range of data mining and visualisation tools and techniques that DSTO have either developed or acquired to assist organisations uncover interesting patterns of behaviours, trends, associations and links that might exist regarding entities referenced in the raw information.

Keywords: information management, information extraction, named entity extraction, data mining, data visualisation.