An Infrastructure for Searching and Harvesting Complex Image Objects

Nadia P. Kozie\v{v}itch\(^1\), Ricardo da S. Torres\(^1\), André Santanchè\(^1\)
Daniel C. G. Pedronette\(^1\), Rodrigo T. Calumby\(^1\),
Edward A. Fox\(^2\)

\(^1\) Institute of Computing, University of Campinas
Av. Albert Einstein, 1251
Campinas, SP, Brazil
{nadiapk,rtorres,santanche,dcarlos,tripodi}@ic.unicamp.br

\(^2\) Department of Computer Science, M/C 0106, McBryde 114,
Virginia Tech, Blacksburg, VA 24061, USA
fox@vt.edu

Abstract
In order to reuse, integrate, and unify different resources from a common perspective, complex objects (COs) have emerged to facilitate aggregation abstraction and to help developers to manage heterogeneous resources, and their components. In particular, complex image objects (ICO) play a key role in different domains, due to their large availability and integration with datasets, metadata, and image manipulation software. Applications which manage ICOs still lack support by mechanisms for processing and managing data, creating references, making annotations, searching by content, harvesting, and organizing their components. Examples of common services that need support in these applications include (i) Content-Based Image Retrieval (CBIR); (ii) the combination of visual with textual retrieval; and (iii) the harvesting of COs. This paper presents the design and implementation of a CO-based infrastructure comprising these three services, focusing on the developer view of service integration as components. Our infrastructure relies on a specific component technology - Digital Content Component (DCC) - to wrap the complex image object and to encapsulate CBIR related tasks. Other contributions rely on the use of rerank and rank-aggregation approaches for combining visual and text retrieval, and the integration of our DCC-based framework with Open Archives Initiative mechanisms to support metadata harvesting and the “discovery” of the digital objects by other applications.

Key-words: Complex Object, Content-Based Image Retrieval, Digital Content Component, Open Archives Initiative Protocol for Metadata Harvesting