Information Sciences 370–371 (2016) 303–322

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A graph-based approach for visualizing and exploring a multimedia search result space

Umer Rashid[a], Marco Viviani[b], Gabriella Pasi[b,*]

[a] Quaid-i-Azam University Islamabad, 45320, Pakistan
[b] Università degli Studi di Milano-Bicocca / DISCo Viale Sarca, 336 – Edificio U14 – 20126 Milano, Italy

ARTICLE INFO

Article history:
Received 19 January 2016
Revised 5 July 2016
Accepted 28 July 2016
Available online 3 August 2016

Keywords:
Search result visualization
Search result exploration
Multimedia Information Retrieval
Aggregated search
Search result space
Search engines

ABSTRACT

Nowadays, together with the increasing spread of the content available online, users' information needs have become more complex. To fulfill them, users strongly rely on Web search engines, but traditional ways of presenting search results are often unsatisfactory. In fact, Web pages carry information that exists in multiple media formats, such as text, audio, image and video objects. Vertical search engines and medium-specific search services do not provide users with an integrated view of search results. Furthermore, multiple media objects are in most of the cases highly semantically interlinked, but the connections between them are not sufficiently exploited to provide a further exploration of the retrieved objects. To address these issues, in this paper we propose a graph-based approach aimed at providing users with the possibility to dynamically visualize and explore a search result space built over a repository of multimedia documents containing interconnected multiple media objects. To do this, we represent the search result space via a graph-based data model, where both the retrieved multimedia documents and connected relevant media objects are considered. Media objects, among them, are connected via different kinds of similarity relationships, which depend on the low-level features and metadata taken into consideration to access the media objects. The approach and the connected visualization and exploration interface have been implemented and tested on a publicly available dataset, and they have been evaluated by means of a usability test.

1. Introduction

In recent years, a rapid and huge increase in the production of digital data has been observed; for example, Bounie and Gille [4] reported a growth of 233% of the information available on the World Wide Web from 2003 to 2008. This information is in most cases carried by different media types, such as text, audio, image, and video objects [1], which in the literature are referred as *multiple media objects* [25]. *Multimedia documents* can in this context be seen as containers of media objects [21], and a Web page is definitely the most known example of a multimedia document.

Nowadays, Web search engines offer to users the possibility to search both for Web pages (traditional search) and for various media types (vertical search) separately. As a consequence, the results produced by the evaluation of a query are presented to users in separate ranked lists, one for Web pages (multimedia documents) and one for each distinct media type. Recently, *aggregated search* approaches have been proposed to integrate the search results from one or multiple information

* Corresponding author.
E-mail addresses: umerrashid@qau.edu.pk (U. Rashid), marco.viviani@disco.unimib.it (M. Viviani), pasi@disco.unimib.it (G. Pasi).

http://dx.doi.org/10.1016/j.ins.2016.07.072
0020-0255/

sources into a single Web page [5,14]. The rationale behind aggregated search is to present to users integrated *information nuggets* [14], which are either parts/portions of documents or media objects. In this context, the integration of multiple media contents can be provided in two main ways. A first solution consists in the simple aggregation of search results from vertical search engines. A second approach assembles and presents information nuggets based on the possible relationships among them.

Taking inspiration from this second approach, we consider that multiple media objects within multimedia documents could be better integrated in the presentation of search results by taking into account their low-level features and/or the metadata associated with them, and by generating *similarity relationships* among media objects by exploiting this information. In this way, it would be possible to visualize and explore the search result space by means of a *relational* (graph-based) approach.

Following this line of research, in this paper we present a novel way of presenting search results that allows: (*i*) the visualization in an integrated way of both the retrieved multimedia documents and the relevant media objects they contain, and (*ii*) the exploration of the search result space based on the similarity relationships among media objects. The proposed approach relies on a graph-based data model that has been defined to represent a search result space, where nodes are associated with both multimedia documents and media objects, and the considered relationships are: (*i*) the *part-of relationship*, connecting media objects to the multimedia document they belong to, and (*ii*) different *similarity relationships*, connecting multiple media objects among them. The rationale behind the definition of *distinct* similarity relationships is that each media object can be possibly accessed through different *modalities* (i.e., textual, acoustic, and visual) depending on the considered *low-level features* or on the available *metadata* extracted from or associated with it. An image object, for example, could be accessed through a textual modality via the caption associated with it, or a visual modality by considering its color histogram. Similarly, an audio object could be accessed through a textual modality considering some user comments associated with it, or a provided textual transcription of the speech, and through an acoustic modality by extracting its audio histogram. This way, considering the possible access modalities connected to different media objects, we are able to establish (at most) three kinds of similarity relationships among media objects, based on suitable similarity measures that act on the considered low-level features or metadata.

A preliminary introduction of some of the ideas presented in this paper has been described in [22]. With respect to that preliminary work, in this paper:

- we provide a complete formalization of the graph-based data model;
- we exploit multiple similarity relationships over all media types, by focusing on their associated features and metadata, and on the corresponding access modalities (while in the preliminary work only the textual modality was exploited);
- we provide a complete implementation of a Graphical User Interface (GUI), adding the possibility to visualize and explore the search result space on a dynamic graph;
- we test the approach on a publicly available dataset, and we provide user-based evaluations of its usability.

The paper is organized as follows. In Section 2, we discuss the background and the motivations of our research. In Sections 3 and 4, we describe the graph-based data model and its instantiation on a publicly available dataset. In Section 5, we illustrate the GUI and we elaborate a visualization and exploration example. In Section 6, we provide an evaluation of the approach by means of a usability test. Finally, in Section 7, we draw the conclusions and we discuss future research directions.

2. Background and motivations

Most Web search companies provide search engines that allow users to access distinct media repositories, following the so-called *vertical search* paradigm [27]. In this scenario, users select a single media type in which they are interested, they formulate a query, and in most cases the vertical search engine produces a simple ranked list of the retrieved results.

On the Web, there are also a variety of specific media search engines like YouTube,[1] Flicker,[2] and FindSounds[3] that allow, by means of keyword-based queries, to retrieve video, image, and audio objects respectively. Only more recently, general purpose search engines started to focus on the integration of the search results produced by a same query over distinct media types. For example, both Google[4] and Bing[5] provide content-based search for text and image objects; moreover, Google provides a small integrated view to present together highlights of textual, image and other media objects in the results Web page. The research area focusing on the above issue is aggregated search [15]. At the basis of aggregated search there are both the aspects of how to select and represent the objects constituting the search result space, and how to present search results to users in an integrated way. Since the approach proposed in this paper is connected to both aspects, in Section 2.1 we provide a description of the two main classes of approaches to aggregated search, and in Section 2.2 we

[1] https://www.youtube.com/.
[2] https://www.flickr.com/.
[3] http://www.findsounds.com/.
[4] https://www.google.com/.
[5] https://www.bing.com/.