Chapter 41
Describing and Selecting Collections of Georeferenced Media Items in Peer–to–Peer Information Retrieval Systems

Daniel Blank
University of Bamberg, Germany

Andreas Henrich
University of Bamberg, Germany

ABSTRACT

The ever-increasing amount of media items on the World Wide Web and on private devices leads to a strong need for adequate indexing and search techniques. Trends such as personal media archives, social networks, mobile devices with huge storage space, and networks with high bandwidth capacities make distributed solutions and in particular Peer-to-Peer (P2P) Information Retrieval (IR) systems attractive. On the other hand, when designing effective media search applications, various search criteria have to be addressed. Hereby, geospatial information is frequently used as well as other criteria, such as text, audio or visual media content, and date and time information.

In this chapter, the authors outline how collections of georeferenced media items can be indexed and searched in P2P IR systems. They discuss different types of P2P IR systems and focus in detail on an approach based on collection description and selection techniques. This approach tries to adequately describe and select collections of georeferenced media items. Finally, the authors discuss its broad applicability in various application fields.

INTRODUCTION

In recent years, the availability—and with it the usefulness—of geospatial metadata has increased dramatically. Digital cameras and mobile phones are nowadays often equipped with GPS sensors at affordable cost. Hence, such devices are able to capture georeferenced information in the personal lives of millions of people from all over the world. In addition, geo-tagging tools with rich user interfaces have emerged in different domains and large geo-tagging initiatives try to georeference textual resources such as in case of Wikipedia. As a consequence, an increased importance of geospatial information in the context of search can be recognized.

Obviously, *geospatial information* is not the only search criterion. When searching for media items other criteria such as *textual content*, *timestamps*, and *(low-level) audio or visual content information* can be used as well—often in an integrative way. A combination of these criteria can allow for the effective retrieval of text, image, audio, and video documents.

As we all know, the amount of media items on the World Wide Web and on private devices steadily increases. Service providers such as Flickr, YouTube, or Facebook (Beaver, et al., 2010) have to maintain huge hardware infrastructures in order to keep up with the tremendous increase in data volumes. So far, it is unclear if existing server-centered solutions will also suit our needs in the future. Hence, a need for alternative indexing and search techniques might arise.

Peer-to-Peer (P2P) Information Retrieval (IR) systems consist of computers from all over the world. These computers can act as both clients and servers. By applying a scalable P2P IR protocol, a "service of equals" for the administration of media items can be established in contrast to existing client/server-based solutions. No expensive infrastructure has to be maintained and idle computing power in times of inactivity can be used to maintain, analyze, and enrich media items.

P2P IR systems offer the benefit that media items can remain on individual devices since there is no need for storing them on remote servers hosted by third party service providers. Crawling which consumes large amounts of web traffic (Bockting & Hiemstra, 2009) can thus be avoided. In addition, dependency from service providers acting as informational gatekeepers can be reduced, because they are no longer able to decide which information can be retrieved or accessed and which cannot. In times of a strong market concentration in internet search and social network applications as well as public debates addressing the privacy of data, P2P IR could offer some benefits.

As our primary use case, georeferenced images are administered in a P2P IR system. The images of a certain user are stored locally on the user's personal device(s) and a scalable P2P IR protocol is applied in order to facilitate retrieval. An image can hereby be described by various criteria: textual metadata, (low-level) visual content features, a timestamp, and a geographic coordinate. Personal media collections containing multiple images can thus be represented by corresponding collection descriptions allowing for efficient and effective collection selection when processing a given query. We assume in the following that at least some of the images of a peer are geo-tagged. A resource description capturing the geographic footprint of an image collection can thus be generated from the set of georeferenced images. In this chapter, we focus on geospatial query processing. In particular, we address geospatial k nearest neighbor (k-NN) queries—finding the k closest media items according to a given query location.

In literature, many approaches for P2P IR can be found. The following section entitled *Peer-to-Peer Information Retrieval Systems for Geospatial Search* will give an overview on different types of P2P IR systems and outline how georeferenced media items can be indexed in a P2P setting in general. We additionally describe how a comprehensive indexing of the abovementioned search criteria (geospatial, textual, date and time, and audio or

Related Content

Detection of Urban Expansion by using DMSP-OLS, Landsat Data and Linear Spectral Unmixing Method
Cihan Uysaland Derya Maktav (2015). *International Journal of 3-D Information Modeling (pp. 58-67).*
www.irma-international.org/article/detection-of-urban-expansion-by-using-dmsp-ols-landsat-data-and-linear-spectral-unmixing-method/138263

A Component-Based 3D Geographic Simulation Framework and its Integration with a Legacy GIS
Zhen-Sheng Guoand Yuzuru Tanaka (2013). *Geographic Information Systems: Concepts, Methodologies, Tools, and Applications (pp. 1216-1230).*
www.irma-international.org/chapter/component-based-geographic-simulation-framework/70501

Sharing Environmental Data through GEOSS
Gregory Giuliani, Nicolas Ray, Stefan Schwarzer, Andrea De Bono, Pascal Peduzzi, Hy Dao, Jaap Van Woerden, Ron Witt, Martin Benistonand Anthony Lehmann (2013). *Geographic Information Systems: Concepts, Methodologies, Tools, and Applications (pp. 1260-1275).*
www.irma-international.org/chapter/sharing-environmental-data-through-geoss/70504

Drivers and Barriers to the Use of Building Information Modelling in India
Anil Sawhneyand Prateek Singhal (2013). *International Journal of 3-D Information Modeling (pp. 46-63).*
www.irma-international.org/article/drivers-and-barriers-to-the-use-of-building-information-modelling-in-india/99617

Modelling the Spatial Distribution of the Anopheles Mosquito for Malaria Risk Zoning Using Remote Sensing and GIS: A Case Study in the Zambezi Basin, Zimbabwe
Francis Danquah Ohemengand Falguni Mukherjee (2015). *International Journal of Applied Geospatial Research (pp. 7-20).*
www.irma-international.org/article/modelling-the-spatial-distribution-of-the-anopheles-mosquito-for-malaria-risk-zoning-using-remote-sensing-and-gis/122801