



# Designing the Content Analyzer of a Travel Recommender System



Carla Binucci, Felice De Luca, Emilio Di Giacomo\*, Giuseppe Liotta, Fabrizio Montecchiani

Dipartimento di Ingegneria, Università degli Studi di Perugia, Italy

## ARTICLE INFO

### Article history:

Received 27 March 2017

Revised 16 June 2017

Accepted 17 June 2017

Available online 19 June 2017

### Keywords:

Content Analysis

Travel Recommender Systems

Content-based Approach

## ABSTRACT

Content-based travel recommender systems suggest touristic attractions based on a best match between users' preferences and a given set of points of interests, called POIs for short. When designing such systems, a critical aspect is to equip them with a rich enough knowledge base that, for each POI, indicates how much the POI is relevant for a set of possible topics of interests, also called TOIs for short. This paper focuses on the problem of designing the Content Analyzer of a content-based travel recommender system. The Content Analyzer is a module that receives as input a set of POIs and a set of TOIs and it computes the relevance of each POI with respect to each TOI. The proposed approach is unsupervised, fully automatic, and it relies on publicly available sources of information. We describe an implementation of the technique in a system called Cicero and present an experimental evaluation of its effectiveness against a ground truth generated by experts.

© 2017 Published by Elsevier Ltd.

## 1. Introduction

Planning a trip by taking into account personal preferences often becomes a time-consuming and difficult task, given the overwhelming amount of information available on a large variety of digital sources (such as institutional web sites, travel blogs, travel guides, etc.). A *travel recommender system* guides a tourist through this large space of possible options by matching touristic and leisure attractions (technically called *Points of Interest*, or *POIs* for short) with traveler's interests. As pointed out in an early work of [Staab et al. \(2002\)](#) and in a more recent survey of [Borrás, Moreno, and Valls \(2014\)](#), the most common travel recommender systems use a content-based approach, in which the user expresses her needs by associating values to a set of attributes and the system matches these needs with a given set of available POIs. As depicted in [Fig. 1](#), the architecture of a content-based travel recommender system consists of three main modules: a *Content Analyzer*, a *Profiler*, and a *Matching Module*. The Content Analyzer gathers information from various sources and it computes a suitable description of the POIs that is stored in a *Knowledge Base*. The Profiler constructs users' profiles by collecting and analyzing data representative of their interests. The Matching Module outputs a *travel recommendation*, i.e., a ranked list of those POIs that are most suitable for the users' profiles. A lim-

ited list of content-based travel recommender systems includes, for example, ([Brilhante, Macedo, Nardini, Perego, & Renso, 2015](#); [Cenamor, de la Rosa, Núñez, & Borrajo, 2017](#); [Györödi, Györödi, & Dersidan, 2013](#); [Huang & Bian, 2009](#); [Lee, Chang, & Wang, 2009](#); [Lim, Chan, Leckie, & Karunasekera, 2015](#); [Martínez Santiago, Ariza López, Montejo-Ráez, & Ureña López, 2012](#); [Montejo-Ráez, Perea-Ortega, García-Cumbreras, & Martínez-Santiago, 2011](#); [Ruotsalo et al., 2013](#); [Vansteenwegen, Souffriau, Berghe, & Oudheusden, 2011](#)).

We focus on the widely adopted keyword-based vector space model to represent users' profiles and POIs' descriptions ([de Gemmis, Lops, Musto, Narducci, & Semeraro, 2015](#)). According to this model, the POIs' description and the users' profiles are represented as two vectors in an  $n$ -dimensional space, where each dimension corresponds to a *Topic of Interest*, or *TOI* for short. A TOI is a keyword, that describes a subject on which a traveler can find interesting attractions in the region; examples of TOIs could be *History*, *Religion*, *Art*, and so on. The travel recommendation is then computed by means of a suitable matching technique between these two vectors. For example, suppose that the profile of a tourist in Rome defined on the TOIs *History*, *Religion*, and *Art* has higher scores on *History*, then the Matching Module will return a list of POIs where the Colosseum and the Roman Forum are ranked higher than the National Gallery of Modern and Contemporary Art.

A critical aspect in the design of an effective content-based travel recommender system is to create a rich enough Knowledge Base with a suitable description of a large set of POIs. The description of a POI in the Knowledge Base should encode how much this POI is relevant for the set of TOIs presented to the user.

\* Corresponding author.

E-mail addresses: [carla.binucci@unipg.it](mailto:carla.binucci@unipg.it) (C. Binucci), [felice.deluca@studenti.unipg.it](mailto:felice.deluca@studenti.unipg.it) (F. De Luca), [emilio.digiacomio@unipg.it](mailto:emilio.digiacomio@unipg.it) (E. Di Giacomo), [giuseppe.liotta@unipg.it](mailto:giuseppe.liotta@unipg.it) (G. Liotta), [fabrizio.montecchiani@unipg.it](mailto:fabrizio.montecchiani@unipg.it) (F. Montecchiani).

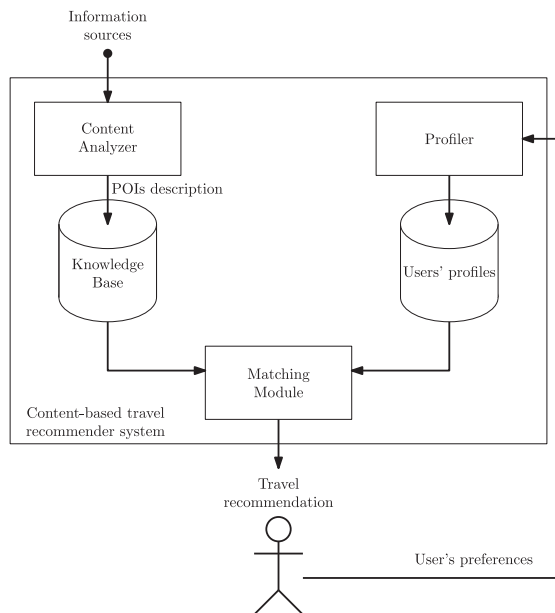


Fig. 1. High level architecture of a content-based travel recommender system.

Obtaining such a description for each POI often relies on a manual classification which is an expensive task (see, e.g., Batet, Moreno, Sánchez, Isern, and Valls (2012); Gavalas et al. (2015); Gavalas and Kenteris (2011); Lim et al. (2015); Lucas et al. (2013); Martínez Santiago et al. (2012); Meehan, Lunney, Curran, and McCaughey (2013); Savir, Brafman, and Shani (2013); Umanets, Ferreira, and Leite (2014); Vansteenwegen et al. (2011)). Also, if the list of TOIs changes, a different description for the POIs may be needed. Thus the maintenance of these systems is an expensive task and it may become particularly critical in those contexts where territorial policy makers want to offer their tourists TOIs that vary based on available seasonal attractions and events. For example, suppose that in a certain period there is a special event in a region, such as a music festival; a regional policy maker can decide to introduce the new TOI *Music* in the system so that tourists who visit the region because of the festival will find POIs that are particularly interesting for them.

This paper focuses on the design of the Content Analyzer for a travel recommender system. Motivated in part by a research grant of the Umbria Region of Italy<sup>1</sup> (see also Binucci, Didimo, Liotta, Montecchiani, & Sartore (2013)), we study the following problem: Given a geographic region  $\mathcal{R}$  and a set of TOIs (defined by a policy maker who wants to promote the touristic attractions of a territory), compute the relevance of the POIs in  $\mathcal{R}$  with respect to the given TOIs. We propose a graph-based algorithmic method specifically tailored for the context of a travel recommender system and we embed such method in a content analyzer called Cicero<sup>2</sup>. The main features of Cicero are as follows.

- Cicero receives as input a region  $\mathcal{R}$  and a set of TOIs  $T = \{t_1, t_2, \dots, t_n\}$ ; it retrieves a set of POIs  $P = \{p_1, p_2, \dots, p_m\}$  in region  $\mathcal{R}$  and it computes as output a *description* for each POI. The description of a POI  $p_i \in P$  is an array whose  $j$ -th element is a numeric value in the range  $[0, 1]$  that represents the relevance of  $p_i$  with respect to TOI  $t_j$ .
- The computation of the descriptions of the POIs is unsupervised and automatic. Since it is unsupervised, it does not

require a training set of reliable data, which can be difficult to obtain. The fact that it is automatic allows us to handle large geographic regions with thousands of POIs, for which a manual assignment would be impractical or too expensive.

- Cicero relies on publicly accessible Web resources and does not use an ontology. Ontologies are often used to represent (and reason about) the tourism domain knowledge (see, e.g., Castillo et al. (2008); Lee et al. (2009); Moreno, Valls, Isern, Marin, and Borrás (2013); Ruotsalo et al. (2013); Wang, Zeng, and Tang (2011)). However, ontologies for travel recommender systems are typically designed ad-hoc and built manually, which can be a time-consuming and labor-intensive task. Instead, the proposed technique computes a description of each POI by executing a shortest path algorithm on a suitable concept network that is extracted by automatically crawling Wikipedia<sup>3</sup> and OpenStreetMap<sup>4</sup>.

The rest of this paper is organized as follows. A critical discussion about the main differences and similarities between the approach of Cicero and existing literature in the areas of travel recommender systems and of semantic technologies can be found in Section 2. In Section 3 we present a reference architecture upon which we designed and implemented our system. In Section 4 we discuss the main principles and methods behind Cicero. We describe how public encyclopedic sources can be used to construct a network of concepts that is then exploited to compute the descriptions of the POIs with respect to the given TOIs. To have an indication about the effectiveness of the proposed techniques, in Section 5 we:

1. Use Cicero to create a concept network for two Italian touristic cities, namely Rome and Perugia; the network is computed with respect to three popular TOIs that are particularly relevant for the two chosen cities, namely *Art*, *History*, and *Religion*.
2. Evaluate some structural properties of the created network that affect the effectiveness and the efficiency of our approach.
3. Use the network to create a Knowledge Base and compare the results against a ground truth defined by professional touristic guides and experts of the two cities. While for their sizes and number of available POIs the two chosen cities have different characteristics, the experimental analysis suggests that the algorithmic technique behind Cicero can be effective in both scenarios.
4. Realized a proof-of-concept implementation of a content-based recommender system that uses the computed Knowledge Base.

Finally, conclusions and future research directions are discussed in Section 6.

## 2. Related Work

The research in this paper naturally relates with both the literature about (travel) recommender systems and the literature about semantic relatedness analysis. In this section we briefly recall some of the most relevant references of these research areas and spend a few more words about those contributions that adopt an approach similar to ours.

Several travel recommender systems have been described in the literature; as already pointed out in Section 1, the majority of them (see, e.g., Brillhante et al. (2015); Cenamor et al. (2017); Györfödi et al. (2013); Huang and Bian (2009); Lee et al. (2009); Lim et al. (2015); Martínez Santiago et al. (2012); Montejo-Ráez et al. (2011); Ruotsalo et al. (2013); Vansteenwegen et al. (2011))

<sup>1</sup> POR-FESR Project "TRART: Telematic Representation-Augmented Reality-Territories".

<sup>2</sup> <http://www.felicedeluca.com/cicero/>.

<sup>3</sup> <https://www.wikipedia.org/>.

<sup>4</sup> <https://www.openstreetmap.org/>.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات