International Journal of Geographical Information Science
Vol. 22, No. 3, March 2008, 315–336

Research Article

Location approximation for local search services using natural language hints

S. SCHOCKAERT*, M. DE COCK and E. E. KERRE
Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281, S9 9000, Gent, Belgium

(*In final form 13 August 2007*)

Local search services allow a user to search for businesses that satisfy a given geographical constraint. In contrast to traditional web search engines, current local search services rely heavily on static, structured data. Although this yields very accurate systems, it also implies a limited coverage, and limited support for using landmarks and neighborhood names in queries. To overcome these limitations, we propose to augment the structured information available to a local search service, based on the vast amount of unstructured and semi-structured data available on the web. This requires a computational framework to represent vague natural language information about the nearness of places, as well as the spatial extent of vague neighborhoods. In this paper, we propose such a framework based on fuzzy set theory, and show how natural language information can be translated into this framework. We provide experimental results that show the effectiveness of the proposed techniques, and demonstrate that local search based on natural language hints about the location of places with an unknown address, is feasible.

Keywords: Geographical information retrieval; Web intelligence; Fuzzy set theory; Local search

1. Introduction

Local search services, such as Google Maps[1], Yahoo! local[2] and MSN search local[3], allow users to search for a particular business within a certain geographic context. A user may, for example, be interested in restaurants, hotels, grocery stores, dentists, etc., which are located close to some user-specified address. Currently, such queries are evaluated against a fixed list of businesses. This allows local search services to display a high degree of accuracy, and to interact with users through a very convenient interface: query results are presented in an intuitive way, including maps that show the locations of the retrieved businesses, driving directions, user reviews, etc. Hence, it should come as no surprise that local search services have become increasingly popular.

However, local search services work in a way that is very different from traditional search engines, which use crawlers that continuously search the web for new information. The sole use of a static, structured knowledge base gives rise to a

*Corresponding author. Email: Steven.Schockaert@UGent.be
[1] http://local.google.com/
[2] http://local.yahoo.com
[3] http://search.msn.com/local/

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2008 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810701626277

number of important limitations. First of all, it restricts the coverage of the system, as many businesses will not be contained in the knowledge base, even if there are websites that contain useful information about their location. Moreover, it limits the information covered to the kind traditionally contained in the well-known yellow pages, while it is exactly the inclusion of more dynamic, ephemeral information that would bring local search services to their full potential (Himmelstein 2005). Consider, for example, a system that would also be able to deal with queries such as *give me information about concerts in Seattle during the next few weeks*, which also involves a temporal constraint.

A second limitation is related to the use of landmarks in queries. Rather than asking for *restaurants near 219 4th Ave N, Seattle*, a user might want to know about *restaurants near Space Needle*. Although the locations of important landmarks are generally available in gazetteers, many place names are not supported in current systems, e.g. because the gazetteers used do not contain any information about the place, or because the place is known by different names, not all of which are contained in the gazetteers. A related problem is the support for neighborhood names. As the boundaries of neighborhoods are usually vague, gazetteers tend to contain either no information at all about neighborhoods, or only a centroid, i.e. a location considered to be the center of the neighborhood. Consequently, local search services provide no, or very limited, support for queries such as *restaurants in Seattle's Belltown neighborhood*.

A promising solution to the aforementioned problems is to augment the available structured information with information extracted from the web. On one hand, this could be information extracted from semi-structured data. Many lists of hotels, restaurants, attractions, etc. are available on the web. Usually, it is quite easy to extract from such lists the relevant names and corresponding addresses, either by writing a wrapper manually, or by using automated wrapper induction techniques (Eikvil 1999, Kushmerick 2000). However, most information on the web is still in unstructured form, i.e. free natural language text. Because it may be very hard to find the address of a particular business or landmark in (unstructured) web pages, we can sometimes only rely on hints in natural language sentences about their location. We may know, for example, that some hotel is located in Belltown, within walking distance from Pike Place Market, and a few blocks away from Space Needle. While we cannot derive the exact location of the hotel from this, we may be able to approximate its location accurately enough to estimate the relevance w.r.t. a given query. To obtain such an approximation, we need a computational framework in which the spatial extent of a neighborhood like Belltown, and nearness information such as *within walking distance from x*, and *a few blocks away from y*, can be represented.

The aim of this paper is to show that local search using natural language hints about the location of places, is feasible. In particular, we show how nearness information in natural language, and information about the surrounding neighborhood of a place, can be translated into fuzzy restrictions, and how such fuzzy restrictions can be used to estimate the location of a place with an unknown address. In the next section, we present an overview of existing work related to the interpretation of nearness and the automatic construction of footprints. Section 3 deals with the construction of a knowledge base containing relevant geographical information, which will be used for experiments throughout the paper. Next, in Section 4, we show how nearness information in natural language can be translated