Exploiting Vague Spatial Information in Geographic IR

Steven Schockaert* Martine De Cock
Etienne E. Kerre
Department of Applied Mathematics and Computer Science, Ghent University
{steven.schockaert,martine.decock,etienne.kerre}@UGent.be

Local search services such as Google Maps[1], Yahoo! local[2] and MSN search local[3] have recently become very popular. Unfortunately, these systems still suffer from a very limited coverage; e.g. when asking for pizzerias in Brussels, the closest pizzeria returned is in the Netherlands for Google Maps, in the United States for Yahoo! local, and in Condé sur Escaut, 42.3 miles from Brussels, for MSN search local[4]. A promising way to overcome this limitation of local search services is to augment the available structured geographic data with knowledge deduced from the vast amount of unstructured and semi–structured information available on the web, often expressed in natural language. As outlined below, representing and reasoning with vague spatial information plays a crucial role in this process.

While it is relatively straightforward to automatically find occurrences of restaurants or hotels in unstructured natural language text, e.g. using a named entity tagger, it may be very hard to find the corresponding addresses. Using site–specific wrappers for webpages containing relevant semi–structured information, e.g. large lists of hotels, restaurants, or other businesses, the corresponding addresses can typically be extracted much more easily. However, even if we do find an address for a particular business, the geocoding step, i.e. the conversion from an address to geographical coordinates, might still go wrong (e.g. if the address cannot be recognized). Therefore, the only location information we sometimes have are vague natural language descriptions like "The hotel is located within a 5 minute walk from the Atomium". By representing vague spatial relations such as "within a 5 minute walk from" as fuzzy relations, i.e. mappings of pairs of locations to the unit interval $[0, 1]$ representing the degree to which two locations are within a 5 minute walk from each other, we obtain a computational model in which approximations of the locations of ungrounded places (i.e. places with an unknown location) can be deduced.

Another step which involves vague spatial information is the modelling of the extent of neighborhoods and other vague, non–political regions. Since it is common place to talk about neighborhoods such as Brussels' EU area, local search services should be able to deal with queries involving them (e.g. "restaurants in Brussels' EU area"). On the other hand, because the boundaries of such regions tend to be ill–defined, gazetteers usually lack any relevant location information. Again, the solution lies in the use of vague information extracted from natural language descriptions on web pages. By extracting a list of places that are said to lie in a certain region, and a set of spatial relations that hold with other regions, an approximation of the extent of the region can be obtained. As both the membership of a place to a region, and spatial relations between regions can be vague, a model for spatial reasoning with vague regions is required to ensure the consistency of the knowledge base and to draw useful inferences.

*Research assistant of the Research Foundation – Flanders
[1] http://local.google.com/
[2] http://local.yahoo.com
[3] http://search.msn.com/local/
[4] Search performed on October 9, 2006