2013 IEEE 14th International Conference on Mobile Data Management

Querying the Web with Local Intent

Christian S. Jensen
Department of Computer Science
Aarhus University
Denmark
csj@cs.au.dk

Motivation

In step with the rapid proliferation of mobile devices with Internet access, the web is increasingly being access by mobile-device users on the move. Further, it is increasingly possible to accurately geo-position mobile devices, and increasing volumes of geo-positioned content, e.g., web pages, business directory entries, and microblog posts, are becoming available on the web. In short, an increasingly mobile and spatial web is fast emerging. This development enables web queries with local intent, i.e., keyword-based queries issued by users who are looking for web content near them. In addition, it implies an increasing demand for query functionality that supports local intent.

Talk Content

A spatial web query takes user-location data and user-supplied keywords as arguments and returns web content that is spatially and textually relevant to these arguments [1]. Many different kinds of relevant spatial web queries may be envisioned. Based on recent and ongoing work by the speaker and his colleagues, the talk covered here offers an account of efforts that aim to invent and enable new spatial web querying functionality that is meaningful, easy to use, meets perceived user needs, and can be supported efficiently. The talk covers the ideas and concepts underlying different kinds of query functionality; it outlines techniques capable of supporting the different kinds of functionality; and it presents new challenges yet to be addressed.

Query Functionality

The essence of the problem setting is simple to formalize. We are interested in querying a set of objects $\mathcal{D}$ that are available on the web. A *spatial web object* $p = \langle \lambda, \psi \rangle \in \mathcal{D}$ has two attributes: a geo-location (λ) and a text value (ψ).

Many different kinds of queries on such object can be considered. In what follows, focus is on queries that involve ranking. The prototypical query $q = \langle \lambda, \psi, k \rangle$ takes a location λ, keywords ψ, and a number k as arguments. It retrieves k objects in $\mathcal{D}$ that are close to λ and relevant to ψ. The following function is used to rank the objects.

$$rank_q(p) = \alpha \frac{\|q.\lambda \; p.\lambda\|}{maxD} + (1 - \alpha) \frac{tr_{q.\psi}(p.\psi)}{maxP}$$

The terms capture the Euclidean distance betwen query and object and the relevance of the object text to the query keywords. Both terms are normalized to the unit interval, and parameter α controls the relative importance of the two. As an alternative, we may define $rank_q(p) = \|q.\lambda \; p.\lambda\| / tr_{q.\psi}(p.\psi)$, which avoids the α parameter.

We may consider continuous queries issued by moving users. Then the location argument λ changes continuously, meaning that the ranking of the objects in $\mathcal{D}$ also changes and must be kept up to date as long as the query is active. The λ-free ranking function has been used in this setting.

Another direction is to take co-location into account when ranking objects. An object that is relevant to a query should gain in relevance if it is located close to other objects that are also relevant to the query. This makes it possible to favor clusters of relevant objects.

Next, it is possible to retrieve relevant groups of objects rather than relevant individual objects, as above. This general direction offers many possibilities for interesting query functionality. Collective queries return sets of objects that together satisfy (contain) the query keywords and that, e.g., are close to each other and close to the query or are simply close to the query location. Top-k group queries return groups G of objects according to a ranking function with the following format.

$$rank_q(G) = \alpha \frac{\beta ||q.\lambda, G|| + (1 - \beta) diam(G)}{maxD} + (1 - \alpha) TR^G_{q.\psi}(G)$$

Here, parameter β controls the balance between the distance to a group and the compactness of the group, and function $TR^G_{q.\psi}(\cdot)$ gives preference to large groups over small groups.

Acknowledgments

The author's research as covered here was supported in part by the Geocrowd Initial Training Network funded by the European Commission as an FP7 - People's Marie Curie Action. The author would like to thank all colleagues with whom he collaborated on the topics covered here.

References

[1] X. Cao, L. Chen, G. Cong, C. S. Jensen, Q. Qu, A. Skovsgaard, D. Wu, and M. L. Yiu, *Spatial Keyword Querying*, ER 2012, pp. 16–29.

978-0-7695-4973-6/13 $26.00 © 2013 IEEE
DOI 10.1109/MDM.2013.101