Exploiting Users' Map Annotations

John Krumm – Microsoft Research, Redmond, WA USA
(Work with Lakshmi Mummidi of Microsoft in Hyderabad India.)

Position Paper for Workshop on Volunteered Geographic Information
December 13-14, 2007

As a company with a consumer level, "Live Search Maps" Web site (http://maps.live.com/), Microsoft is in a position to gather geographic information from a large number of users. One feature that enables this ability is Live Search Maps' "Collections" feature. Collections allow users to create sets of geographically anchored pushpins, each annotated with text, URLs, or photos. Some of the best collections are highlighted at http://www.passthepoi.com/, which includes "100 Points of Interest in Central Park", China's "Five Great
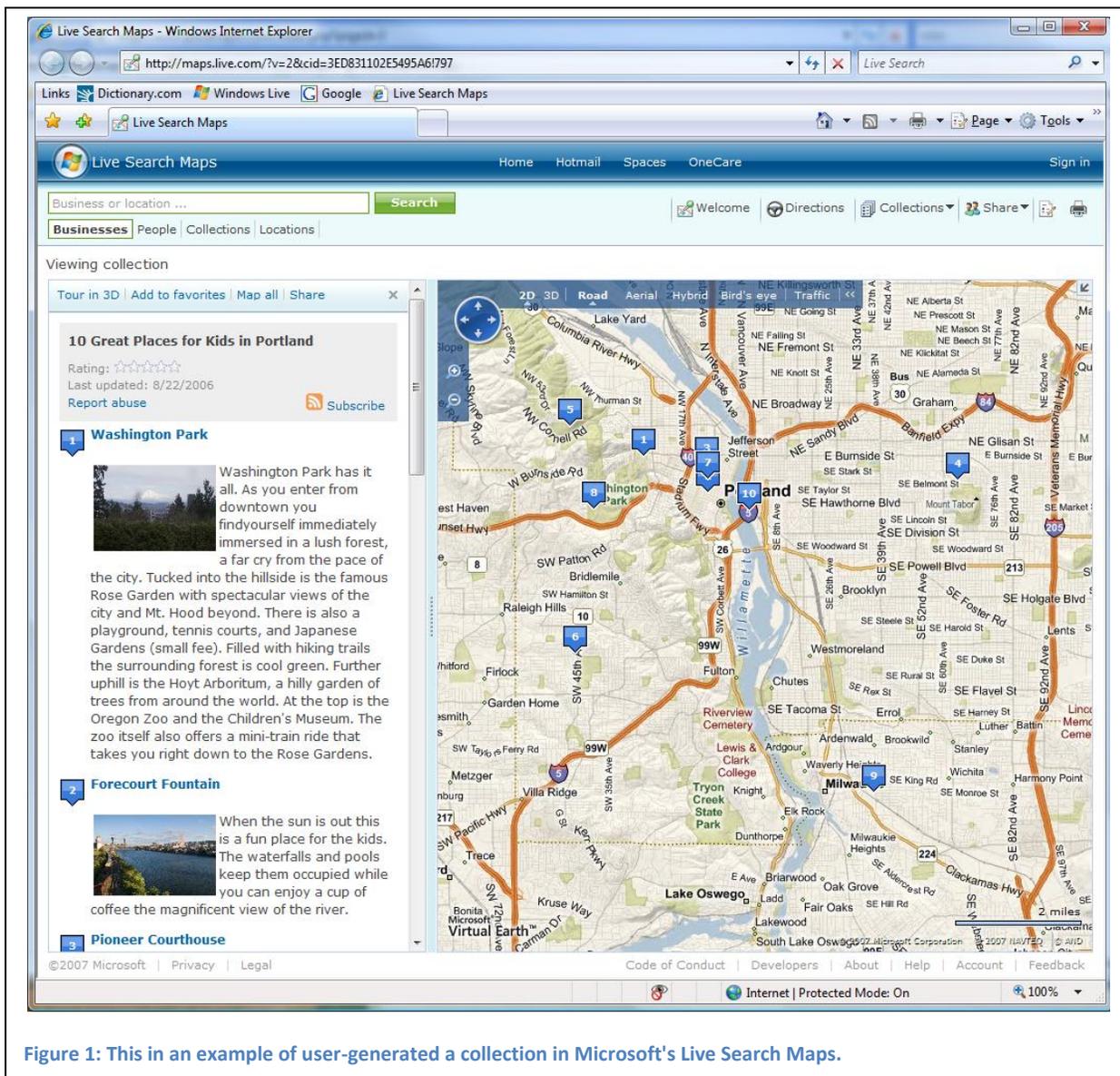

Figure 1: This in an example of user-generated a collection in Microsoft's Live Search Maps.

Mountains", and "10 Great Places for Kids in Portland", shown in Figure 1.

We currently have a few hundred thousand public collections, which altogether give us over one million annotated pushpins with latitude/longitude. The number of public collections is growing by many thousand per month.

User-annotated points like this can be exploited for a variety of uses. Two that we have attempted are to assess the prominence of existing landmarks and to find new landmarks that should be added to the map. Assessing the prominence of existing landmarks is relatively easy. We have a database of landmarks, and we check to see how often the landmark name is mentioned in the title or description of a collection pushpin. The more often the landmark is mentioned, the more prominent it is. As an example, we looked at a subset of pushpins in the Seattle, WA USA area and found the list in Figure 2 as the ones most often mentioned. Information like this can be used to determine which landmarks to show at different zoom levels, as a way to describe more obscure locations in terms of prominent ones (e.g. 0.5 kilometers east of Pike Place), and as a way to pick landmarks to give driving directions.

We have also developed an algorithm for finding new landmarks from collections. In brief, this proceeds by first making geographic clusters of pushpins, extracting all possible one-, two-, and three-word phrases from the associated text, and processing these phrases to find which ones are mentioned frequently in the cluster (e.g. "Space Needle"), but not very often outside the cluster. Comparing phrases outside the clusters helps eliminate a huge number of common, un-landmark-like phrases (*e.g.* "over there", "here we see", *etc.*) Using some simple machine learning techniques, we have found many sensible new landmarks that do not appear in our regular database of landmarks. We verified the relevance of the new landmarks with a small user study.

Figure 2: This shows the prominence of Seattle landmarks as determined by the number of times they were mentioned in a subset of pushpins from the area.