

Coolhunting on the Web and in the Blogosphere

Peter Gloor, Renaud Richardet
MIT Center for Collective Intelligence
pgloor@mit.edu

This paper introduces a new Web and Blog mining approach, which we call “Web Coolhunting”. We make use of the fact that the Web has become a mirror of the real world, breaking latest news through active participation of millions of volunteers on Web sites such as Wikipedia, and political blogs such as dailykos and instapundit. TeCFlow, a dynamic social networking tool, measures popularity and influence of brands and stars over time by looking at their “social embedding” on the Web. Our approach builds a network map with the linking structure of a list of Web sites returned by a Google query. Combining multiple datasets, each containing the linking structure of the Web sites collected through querying Google for the name of a search term (the “star”) permits to find the most central star in a group of stars by comparing the betweenness scores of the different stars. By combining the link maps returned by different searches, we can compare different stars, identifying the ones with the highest betweenness values. These are the most linked, or “talked about” search terms in a given Web or Blog context.

The same combined query and subsequent evaluation by betweenness also permits to find the most relevant Web sites discussing related topics. These Web sites also double up as “kingmakers.” For example, pages on the New York Times, and Wikipedia come up as most relevant without necessarily containing the search terms. In addition, our betweenness-based algorithm is much harder to spam than e.g. Google’s page rank, which is essentially degree-based. We have applied our system to the tracking over time of topics ranging from political candidates and political issues to music bands and Oscar nominations.