Lev Manovich

From Reading to Pattern Recognition

The emergence of social media creates a radically new opportunity to study cultural processes and dynamics. For the first time, we can follow imagination of hundreds of millions of people –th images and videos they create and comment on, the conversations they are engaged in, the opinions, ideas, and feelings they express.

Until now, the study of the social and the cultural (individual beings, individual artifacts, and larg groups of people/artifacts) relied on two types of data: "shallow data" about many (statistics, sociology) and "deep data" about a few (psychology, psychoanalysis, anthropology, art history; methods such as "thick description" and "close reading"). However, the rise of social media alor with the computational tools that can process massive amounts of data makes possible for a fundamentally new approach for the study of human beings and society. *We no longer have to choose between data size and data depth*. Rather than having to generalize from small samples or rely on our intuition, we can study exact cultural patterns formed by millions of cultural texts. I other words, the detailed knowledge and insights, which before can only be reached about a fev texts, can now be obtained about massive collections of these texts.

In 2007, Bruno Latour summarized these developments as follows: "The precise forces that mo our subjectivities and the precise characters that furnish our imaginations are all open to inquirie by the social sciences. It is as if the inner workings of private worlds have been pried open because their inputs and outputs have become thoroughly traceable." (Bruno Latour, "Beware, your imagination leaves digital traces", Times Higher Education Literary Supplement, April 6, 2007.)

But how do you "read" through billions of Twitter posts, blogs, Flickr photos, or YouTube videos practice? That is, how do you *read for patterns*?

Today people use a variety of software tools to select the content of interest to them from this massive and constantly expanding universe of cultural texts and conversations. These tools include search engines, RSS feeds, and recommendation systems. But while these tools can he you to find what to read, they do not show the larger patterns across this universe.

Computer scientists and media companies use a different set of tools and techniques that allow for the detailed study of such patterns. They employ statistical data analysis, data mining, information visualization, and visual analytics. They also have access to substantial computatio resources needed to analyze massive data sets. For example, many companies use "sentiment analysis" to study the feelings which people express about their products in blog posts. Recent publications in computer science investigated how information spreads on Twitter (data: 100 million tweets), what qualities are shared by most favored photos on Flickr (data: 2.2 million photos), and what geotagged Flickr photos tell us about people's attention (data: 35 million photos).

What if everybody had access to such techniques? At present, this requires knowledge of advanced topics in computer science and statistics. However, with the right tools, anybody shou be able to at least explore large image collections and notice interesting patterns. At Software Studies Initiative (softwarestudies.com), we have been developing such software tools, and testing them on sets of different types of cultural images ranging from all 4535 covers of *Time* magazine (1923-2009) to one million manga pages. Currently we are using these tools to study video remixes on YouTube, images from deviantart.com, and spatial patterns in Second Life. W plan to release all tools as open source shortly.

[May 2011]