

Emergent Information Regularities: Implications for Research and Education in i-Schools

Suresh K. Bhavnani
School of Information, University of Michigan
bhavnani@umich.edu

While much web research in the information field has focused on how users search and retrieve information, very little research has focused on the underlying structure of information itself. This is similar to analyzing the food-seeking behavior of an ant crawling on a beach by focusing on the ant's actions while ignoring the topology of the beach itself. However, more recently, researchers from many disciplines are uncovering *emergent regularities* (uncoordinated actions leading to coherent patterns) of information on the web. This focus on the structure of information provides exciting new opportunities to rethink effective strategies to search and retrieve information, and what to teach aspiring information professionals in a rapidly changing search technology environment.

The above research has revealed that the web exhibits regularities which include patterns in web *content* (e.g. patterns in information scatter, density, specialization, and reliability) and web *structure* (e.g. patterns in link density and link specialization). For example, despite that absence of coordinated actions between web authors, information about a topic on the web tends to occur in different *densities*: some pages contain broad overviews of many facts (general pages), while other pages contain detailed descriptions of a particular fact (specific pages), and still others contain a few relevant facts in the context of broader topics (sparse pages). Other researchers have revealed that pages are linked based on *specialization*. For example, pages that have similar content have more links to each other compared to pages that have dissimilar content. This has led to collections of densely connected pages which share a topic, and to information *hubs* which link to many pages of a similar topic.

Most researchers believe that the above web regularities are not transient information phenomena, but the result of basic underlying processes in how information is created. This is because (as observed by bibliometricians) many of the above regularities have analogs at smaller scales in the print world such as articles about a topic being scattered across different journals in different concentrations (few journals have many articles), and information genres (e.g. brochures) have always existed in print form.

Given the vastness and importance of the open web for many users, knowing how to navigate through the massive accumulation of information on the web makes understanding these regularities critical to intelligently using current and future tools. For example, regularities in information density suggest that to find comprehensive information about a topic, users should follow a *general-specific-sparse* search strategy, where they first read a few general pages (to get an overview of all the facts), followed by specific pages (to get detailed information about particular facts), followed by sparse pages (to understand how the topic being searched might be connected to related topics). Such a strategy reduces the probability of missing important information about a search topic, and is an approach that is independent of features in search tools.

Emergent regularities of information on the web therefore provide new opportunities for information research within and across information domains. Furthermore, such research provides the opportunity to shift the focus of search training in i-Schools away from how to use rapidly-changing features of search tools, towards how to exploit important regularities of information.