In preparing this special issue on structure discovery, I was motivated by the observation that *unstructured data* is often not unstructured at all. It may possess a rich (though perhaps irregular) structure. Often, it is the absence of **known** structure rather than the absence of structure that leads to a data-set being characterized as "unstructured" or "semi-structured". Web data and data represented in XML (a semi-structured data model) is often (though certainly not always) highly structured. Similarly, I have found in my own work on integrating and transforming structured data sources, that effective integration exploits not only the given ("imposed") structure of a schema, but also implicit structure within the data. These observations were reinforced by an entertaining keynote by Joe Hellerstein at the recent 2003 SIGMOD WebDB Workshop that elaborated on techniques that exploit "engineered" structure in data (as done in traditional database management where a data design or schema is imposed and enforced) and those that exploit found structure (as done in information retrieval where unstructured documents are managed without a constricting schema). However, data rarely resides neatly in one of these two camps. Unstructured data may possess structure that could be exploited in its management. Similarly, the imposed schema of a structured data-set may not always do it justice as it may be incomplete or even inaccurate.

The articles in this special issue explore the question of how to find new and unknown structure in data. Notice that I have purposely not defined precisely what I mean by structure. Each article takes a different view of this, focusing on structure that can be exploited in different classes of data management applications. The articles also present techniques drawn from a variety of disciplines. It is my hope that collecting these short position statements together in one volume will illustrate the problems, permit comparisons, highlight synergies, and inspire more work in what I believe to be a very promising research direction.