The Current Issue

We database folk love our structured data. The field started as an effort to solve the business data processing problem. It is hard to imagine something more (rigidly) structured than an 80 column card having fixed length fields. Precursor database systems, called formatted file systems, were little more than cards on tape. Early database systems were liberated from the 80 column constraint by the increased flexibility of tape and disk, where records could actually be arbitrary sizes. Further liberation occurred when the database systems started to deal with variable length fields.

Despite the preceding advances, and the subsequent introduction of data independence, where one did not have to know the physical representation of data in order to query and update it, we continued to rely on complete knowledge of how data was laid out, what the attribute names and types were, indeed complete syntactic knowledge of the data. But times change. Our prior world of structured data depended upon our controlling the definition of the data and hence its structure. Today's web-ish, XML-ish world is clearly not under our control. So if the database field is to contribute to this new world, we need to meet it at least half way. One way to do this, so as to leverage our strengths in structured data, is to discover or impose structure on this new world's data.

The current issue presents a number of ways of attacking the problem of structure discovery. Given the great diversity of data, there is probably no one golden road to success in this endeavor. The current issue explores a number of interesting and widely varying approaches. Renee Miller, who herself structures her share of data, has brought together an interesting cross-section of methods for finding the structure that is in data. I want to thank Renee for the fine job she has done in producing an issue on structure discovery that should stimulate lively discussions in many parts of our field.