Abstract: The vast amounts of data available on the Web present unique opportunities, but are often extremely hard to work with due to their scale, noisiness, and heterogeneity. In this talk, I discuss novel algorithms that address the challenge of making sense of both structured data and unstructured text on the Web. One major focus is reliably matching equivalent items across different Web sources, including Wikipedia and domain-specific databases, which we solved using scalable graph-based algorithms and linear optimization techniques. Building on this, I discuss methods to harvest taxonomic and semantic information about entities and concepts in over 100 languages, which led to UWN/MENTA, the largest database of its kind. Finally, I present Web-scale text analytics methods that allow us to collect additional common-sense knowledge that is useful in natural language understanding tasks. Before concluding, I outline several applications of this work, including query interfaces and reasoning engines.