Table understanding is typically easy for humans, yet difficult to automate. It involves more than just identifying header and data cells and associating category labels with each data cell. Table understanding must also align both table category labels and table data values with domain knowledge. It is furthermore essential, although little studied so far, to fully exploit auxiliary information like table titles, footnotes, and parenthetical unit designators. Finally, the output of a table analysis system must be suitable for organizing and querying the collected knowledge. Formalizing and unifying these processes will lead to deeper insights about table understanding and facilitate partially or wholly automating it.

To this end, we propose the development of an ontology of tables—what they are and how table understanding can be automated: Based on this table ontology, we approach automating table understanding as follows: Via grid tilings and layout geometry, we discover each table's topology—its properties that are independent of layout and remain invariant under transformation into an abstract conceptualization. After executing the appropriate transformation, we semantically enrich the conceptualized table by aligning it with domain knowledge. Specifically, we linguistically ground its concepts and the relationships among these concepts, and explicate all implied constraints over these concepts and relationships. We then extend the analysis of single tables to collections of related tables enabling users to harvest and organize information in a domain of interest covered by the table collection. As a practical demonstration of the power of deep table understanding, we intend to process 2000 random samples from a large class of tables we call "grid tables" found on institutional geopolitical websites.

For practical significance of the proposed research, we note that decision makers of all kinds (scientists, engineers, business executives, military leaders, and public officials) have a constant need to gather and organize information scattered among hundreds of web tables. We pursue the objective of automated alignment of the information contained in heterogeneous yet domain-specific table collections in an effort to aid decision makers reach the best possible formative and summative decisions.

Intellectual Merit. We intend to establish complementary ontologies *for* defining tables and *for* processing tables, and to use them to grow populated, domain-specific ontologies of information extracted *from* tables. For an important class of information-dissemination systems (i.e., grid tables), the application of a table ontology to develop an ontology of the messages carried by that system (i.e., the content of grid tables) is a new paradigm in ontology engineering. Simultaneous conceptualizations of table structure/processing and of table content should synergistically lead to a greater understanding of ontologies themselves and of automated ontology learning. Furthermore, reverse-engineering semi-structured data into organized, queriable knowledge structures expands the frontiers of document engineering while bypassing some of the problems of natural-language understanding.

Broader Impact. A practical way to harvest and organize information from tables would benefit every community that uses tables to disseminate information. Beyond its direct application to aid decision makers, this research also promotes education and cross-fertilization between universities. Our research teams at BYU and RPI have always included undergraduate and graduate students from diverse cultural and geographic backgrounds, experience, gender, and ethnicity. As we guide their professional development, we will continue to emphasize non-technical aspects such as the ethos of experimental work, statistical integrity, collaborative versus competitive research, respect for the intellectual property of others, and publication etiquette.

Keywords: automated table processing, table tessellation, information harvesting from web tables, semantic enrichment, table structure ontology, table task ontology.