**Table 1:**

Comparison of methods for extracting data from web tables

|  |  |  |  |
| --- | --- | --- | --- |
| **Limitation** | **Way to overcome the limitation, represented in [27]** | **Proposed way to overcome the limitation (Qurma + TableProcessor)** | **Example** |
| The lack of contextual information | The Semantic table interpretation (STI) engine calculates a measure of similarity for all embedded concepts and contexts. | The Qurma tool [29] uses data from Wikipedia, which looks at which topic the term is most likely to refer to. | Qurma enables automatic Wikipedia searches to uncover the shared concept: 'Rivers flowing into Balkhash. |
| Dependency on table metadata | STI uses a similarity metric, such as cosine similarity or dot product. | The Qurma tool uses 10 heuristics to avoid relying on metadata | An example of one of the heuristics is the rule for automatic recognition of coordinates. |
| Dependence on specific knowledge graphs | TableMiner+ utilizes OpenIE for domain-independent structured information retrieval. | The Qurma tool extracts tables across domains using the CleanArchitecture method. | The tool searches the entire Wikipedia and can retrieve data from any subject area. |
| Identification of a limited set of annotations | STILTool assesses annotations from various approaches on T2Dv2 and Limaye200 Gold Standard tables. | The TableProcessor [25-30] tool, having received a new concept, adds it to the system, and models the semantic interdependence. | New names, previously unknown to the system, are placed in a special buffer zone and offered to the user for manual confirmation or rejection. |
| Lack of implementation | MantisTable implements two modern approaches (T2Dv2 and Limaye200) and compares with a baseline method that uses different annotation steps for concepts, data types, and predicates. | The QURMA system's Fluent Design interface allows users to specify a document's URL for table retrieval.  TableProcessor has a ready-made script on GitHub where any user can process their table. | 1. Qurma: https://github.com/Kyrmasch/Sorge 2. Table Processor: https://github.com/Igriva/TableProcessor |