Incorporating Information Extraction in Database Queries

Yoav Nahshon, Liat Peterfreund, Stijn Vansummeren, WebDB16 (Best Paper)
Advisor: Prof. Benny Kimelfeld

Information extraction (IE) is the identification, and consequent or concurrent classification and structuring into semantic classes, of specific information found in unstructured data sources, such as natural language text, making the information more suitable for information processing tasks.

Spannerlog is a declarative language for information extraction that extends Datalog. It defines relations both over strings and spans and therefore enables to reason about the structure of a string as well as its content. The language incorporates different kinds of IE functions such as natural language processing functions and spanners.

IE functions can operate on substrings.