ABSTRACT: It is often challenging to specify queries against a relational database since SQL requires its users to know the exact schema of the database, the roles of various entities in a query, and the precise join paths to be followed. On the other hand, keyword search is unable to express many desired query semantics.

In the real world, people ask questions in natural language, such as English. Theoretically, natural language interfaces for databases (NLIDBs) have many advantages over other widely accepted query interfaces. For example, a typical NLIDB would enable naive users to specify complex, ad-hoc query intent without training. Not surprisingly, an NLIDB is regarded by many as the ultimate goal. Despite these advantages, in real world applications, NLIDBs have not been widely adopted.

In this dissertation, we investigate the construction of NLIDBs, specifically from the following three aspects:

1. A natural language query is inherently ambiguous and some ambiguities may be too hard for computers to resolve. Can a system collaborate with users to achieve satisfactory reliability without burdening the user too much?

2. The interpretation process can be considered as a mapping from a natural language query to the correct point in the semantic coverage of the NLIDB. Can the mapping process get easier by carefully defining the semantic coverage?

3. Can an NLIDB work when no training examples are available, collect the user behavior data as the training examples and improve itself from real usage?

In this dissertation, we provide affirmative answers to the above questions in the form of new query mechanism designed, techniques provided and systems constructed.

Chair: H. V. Jagadish