

Title: Distributional–relational database

URL: https://en.wikipedia.org/wiki/Distributional%E2%80%93relational_database

PageID: 53644736

Categories: Category:Computational linguistics, Category:Computer data, Category:Database management systems, Category:Heuristics, Category:Language modeling, Category:Natural language processing software

Source: Wikipedia (CC BY-SA 4.0).

A distributional–relational database , or word-vector database , is a database management system (DBMS) that uses distributional word-vector representations to enrich the semantics of structured data .

As distributional word-vectors can be built automatically from large-scale corpora , [1] this enrichment supports the construction of databases which can embed large-scale commonsense background knowledge into their operations. Distributional-Relational models can be applied to the construction of schema-agnostic databases (databases in which users can query the data without being aware of its schema), semantic search , schema-integration and inductive and abductive reasoning as well as different applications in which a semantically flexible knowledge representation model is needed. The main advantage of distributional–relational models over purely logical or semantic web models is the fact that the core semantic associations can be automatically captured from corpora, in contrast to the definition of manually curated ontologies and rule knowledge bases. [2]

Distributional–relational models

Distributional–relational models were first formalized [3] [4] as a mechanism to cope with the vocabulary/semantic gap between users and the schema behind the data. In this scenario, distributional semantic relatedness measures, combined with semantic pivoting heuristics can support the approximation between user queries (expressed in their own vocabulary), and data (expressed in the vocabulary of the designer).

In this model, the database symbols (entities and relations) are embedded into a distributional semantic space and have a geometric interpretation under a latent or explicit semantic space. The geometric aspect supports the semantic approximation between entities from different databases, or between a query term and a database entity. The distributional relational model then becomes a double layered model where the semantics of the structured data provides the fine-grained semantics intended by the database designer , which is extended by the distributional semantic model which contains the semantic associations expressed at a broader use.

These models support the generalization from a closed communication scenario (in which database designers and users live in the same context, e.g. the same organization) to an open communication scenario (e.g. different organizations, the Web), creating an abstraction layer between users and the specific representation of the conceptual model.

References