**Explore FREDDY: Fast Word Embeddings in Database Systems**

**Word Embeddings in Database Systems**

Language learning methods (word2vec, fastText) extract semantic word relations → **Word Embeddings**

Importing word embeddings in a relational database system → Enables inductive reasoning on text values.

**Word Embedding Operations**

- **SELECT** keyword
- **FROM** keywords
- **ORDER BY** cosine_similarity('comedy', keyword)
- Comedy, sitcom, dramedy, comic, satire, ...

Cosine_similarity(varchar t1, varchar t2): Calculating the cosine similarity of two tokens.

**System Architecture**

- Extended Postgres Database System: FREDDY
- Extension with novel Word-Embedding Operations (UDFs)
- Index structures of word embeddings as database relations
- Different search methods for different operations (non-exhaustive, exhaustive and exact search) based on product quantization

**Search Methods**

- Post verification and batch-wise execution according to query demands

**Product Quantization for Fast Similarity Search**

- Idea: Reduce the computation time of the Euclidean square distance through an approximation by a sum of precomputed distances

Quantizer functions $q$ assign subvectors $u_i(y)$ to centroid $[c_1, ..., c_k]$.

$$y = (y_1, ..., y_d, y_{d+1}, ..., y_{2d}, ..., y_{d+k}, ..., y_{3d})$$

$$u_i(y) = (u_{i1}(y), ..., u_{id}(y), ..., u_{id+k}(y), ..., u_{i3d}(y))$$

**Distance Calculation**: 
Calculation of approximate distances by sums of precomputed squared distances

$$d\left(u_i(x), q_j(u_j(y))\right)^2$$

**IVFADC** (Inverted File System with Asymmetric Distance Computation)

- Non-exhaustive search reduces the amount of distance computations
- But: Not applicable for all operations

**Challenges**

- Integrate operations in SQL
- Sufficient performance to execute multiple operations for one query during runtime → approximated nearest neighbor search
- Accomplish different demands on precision and execution time

Contact: michael.guenther@tu-dresden.de  
https://wwwdb.inf.tu-dresden.de/research-projects/freddy/