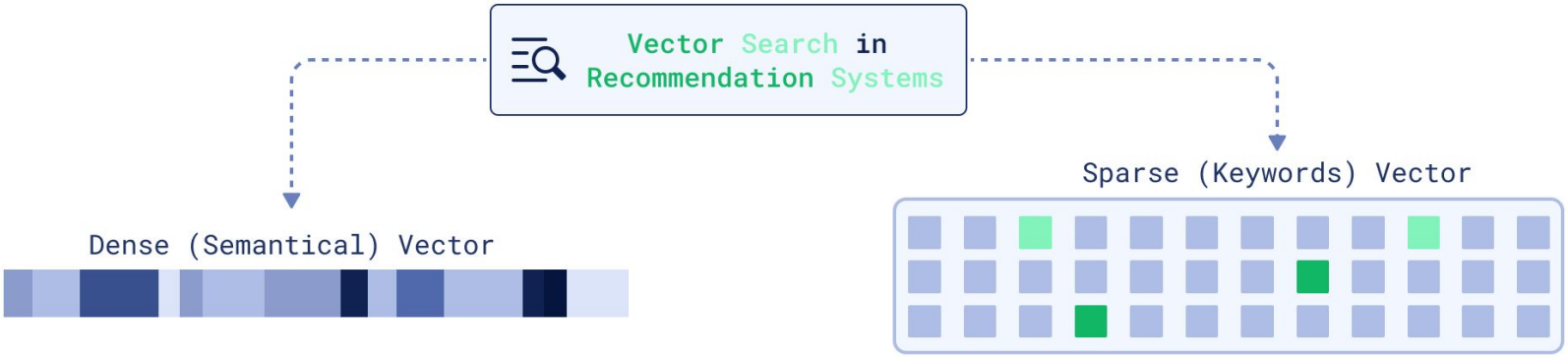
Lightning Talk 1

Enhancing Hybrid Search: *What if BM25 understood meaning?*

Evgeniya Sukhodolskaya,
Developer Advocate,
Qdrant

Hybrid Search

Hybrid Search: Sparse Vectors

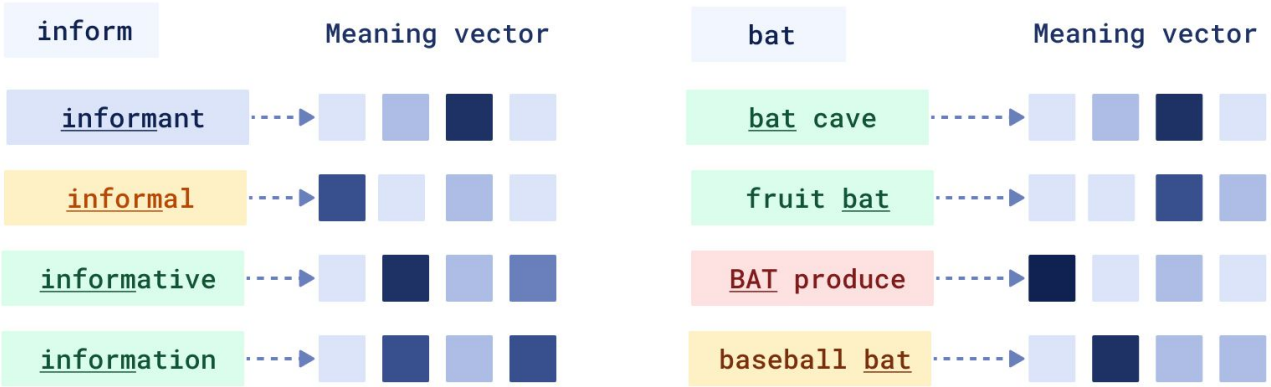


BM25

Sparse Neural Retrieval

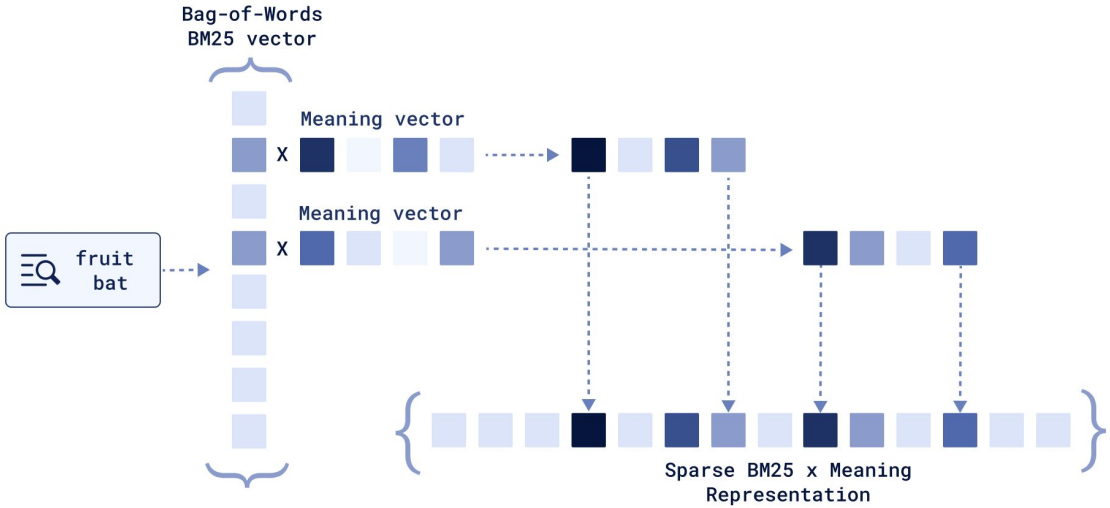
What if BM25 Understood Meaning?

$$\text{score}(D, Q) = \sum_{i=1}^{N} \text{IDF}(q_i) \cdot \text{Importance}_D^{q_i} \cdot \text{Meaning}^{q_i \times d_j}, \text{ where term } d_j \in D \text{ equals } q_i$$

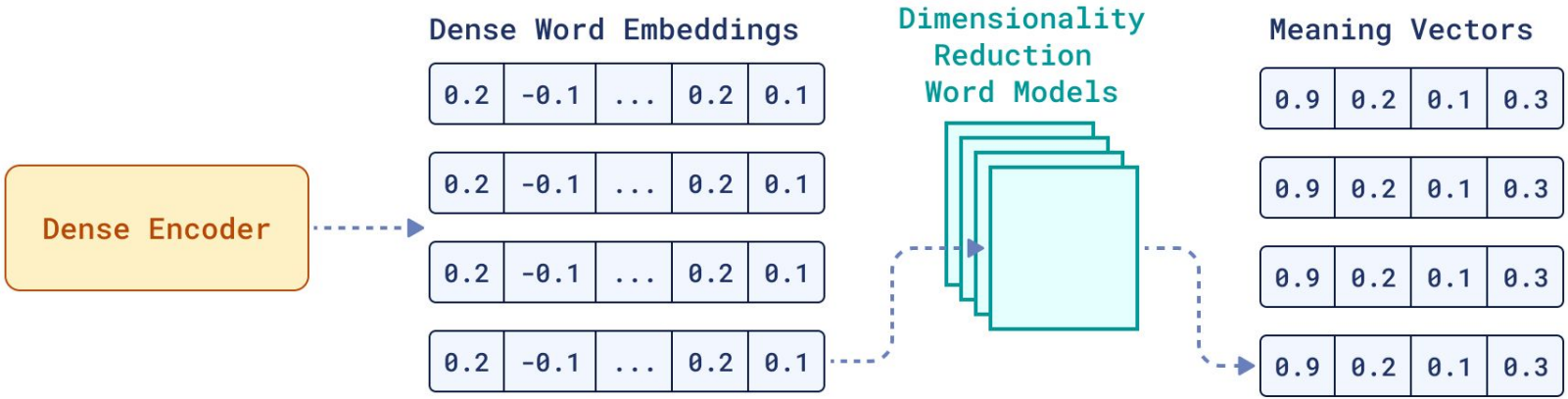


What if BM25 Understood Meaning?

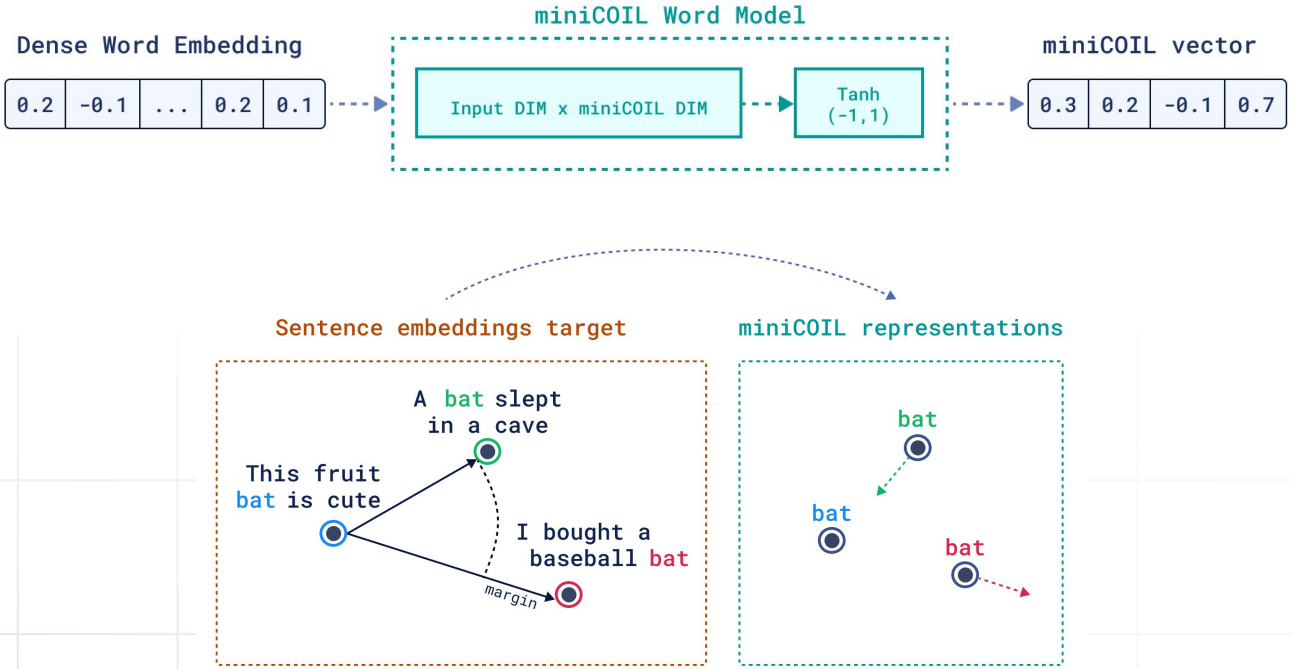
$$\text{score}(D, Q) = \sum_{i=1}^{N} \text{IDF}(q_i) \cdot \text{Importance}_D^{q_i} \cdot \text{Meaning}^{q_i \times d_j}, \text{ where term } d_j \in D \text{ equals } q_i$$



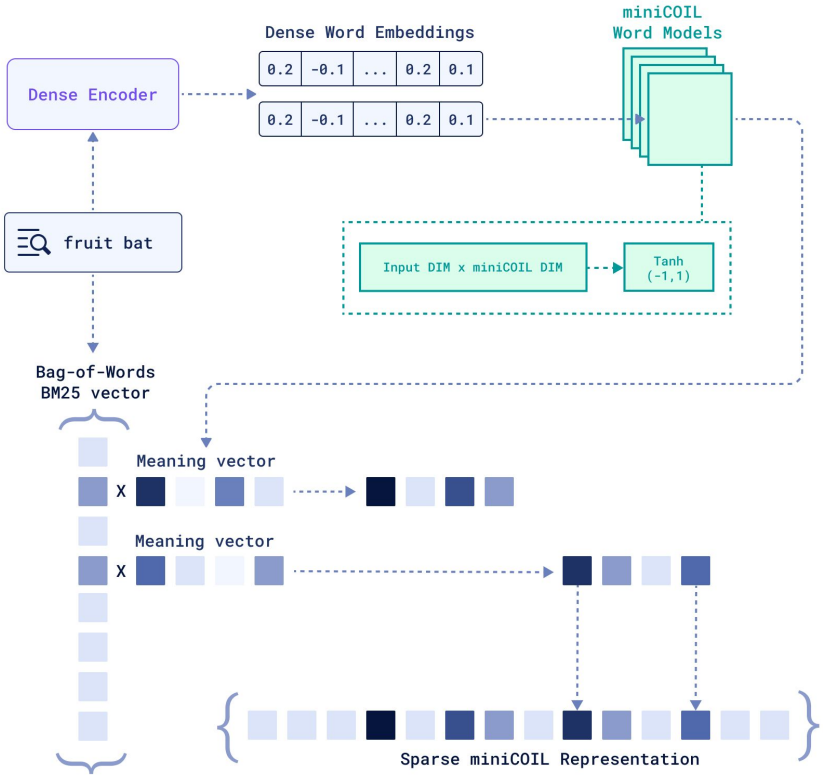
Meaning vectors



Meaning vectors: miniCOIL



Meaning vectors: miniCOIL



Using miniCOIL

miniCOIL works as if BM25 understood the meaning of words & ranked documents based on this semantic knowledge.

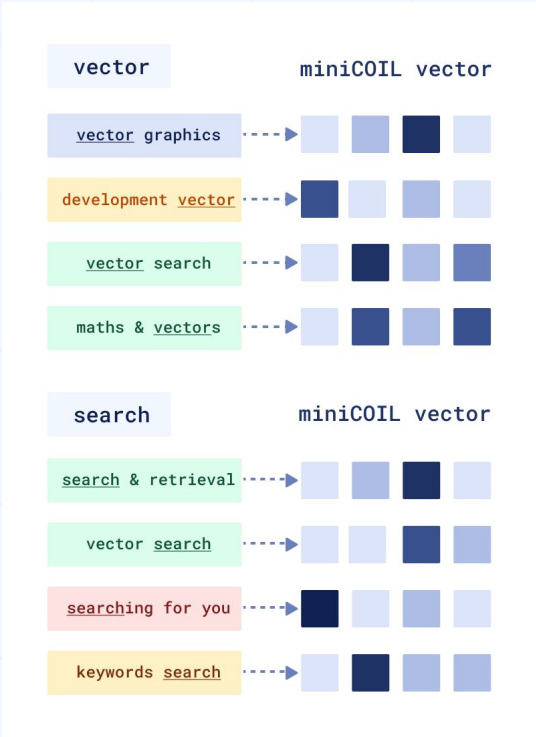
If you also aim for documents semantically similar to the query but expressed in different words, combine **miniCOIL** with **dense encoders**

miniCOIL v1 on HuggingFace

Hybrid Search with miniCOIL

Say, you're searching through book/article/paper titles that contain the words *'vector'* and *'search'*

Code in GitHub Repo



Thank you!