

Vektordatenbanken: Details

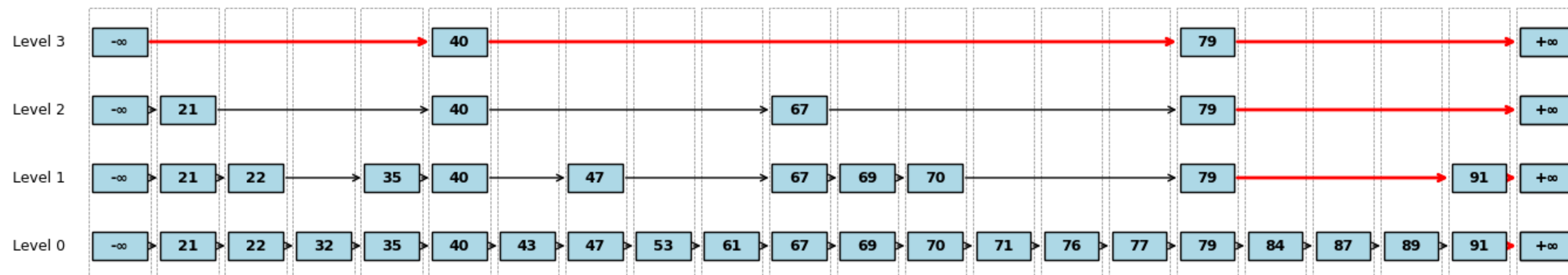


Ich will die 10 ähnlichsten Filme zu
"Pulp Fiction"
in den ca. 23,3 Mio. Titeln der IMDB finden.

Semantische Suche

Wie findet man schnell den nächsten Nachbarn?

- > Die k ähnlichsten Elemente zu finden wird als **k-Nearest-Neighbor-Search** bezeichnet
- > Problem
 - > Klassische Indexstrukturen wie B-Trees benötigen eine Ordnung der Elemente (z.B. Zahlen, Zeichenketten)
 - > Naives Vorgehen für Vektoren ist nicht praktikabel
 - > Vergleiche jeden Vektor mit allen anderen
 - > Komplexität: $O(n)$
- > Ziel
 - > Ähnlichsten Elemente ohne alle Vektoren zu vergleichen
 - > Gewisse Ungenauigkeit zulassen (**Approximate Nearest Neighbor (ANN)**)
- > Lösung-Idee: **Skip-Listen** [3]
 - > mehrstufige, zufällig erzeugte Listen mit Sprungverbindungen
 - > schnelleres Suchen (durchschnittlich $O(\log n)$) und schnelleres Einfügen (durchschnittlich $O(\log n)$) als bei einfachen Listen



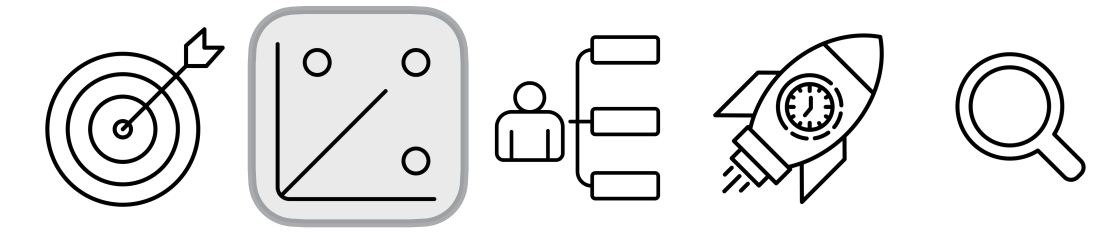
Suche in eine Skip-Liste

Demo

Skip-Listen

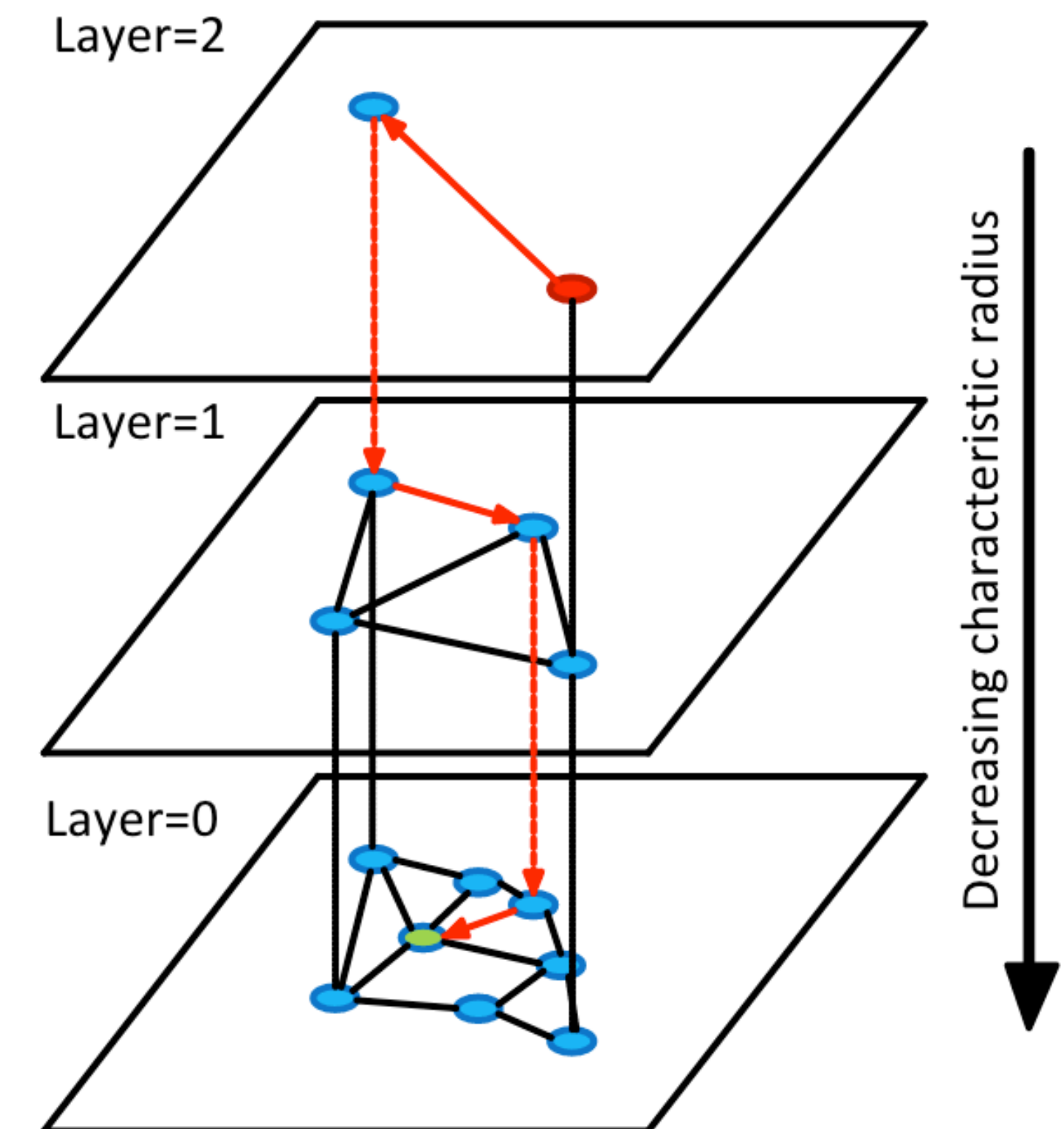
Code und Notebooks verfügbar unter: <http://drumm.sh/dhbw/notebooks>

Semantische Suche



Wie findet man schnell den nächsten Nachbarn... in Datenbanken?

- > Lösung
 - > **Hierarchical Navigable Small World (HNSW) Graphs** [4]
- > Idee
 - > Graph-basierte Indexstruktur für schnelle Approximate Nearest Neighbor (ANN)-Suche
 - > ähnlicher Grundgedanke wie bei Skiplisten
 - grobe Orientierung oben, feine Navigation unten
- > In vielen Vektordatenbanken verfügbar
- > Alternative Ansätze sind z.B. Locality-Sensitive Hashing [9]



Visualisierung der HNSW-Idee aus [4]

Anhang & Referenzen

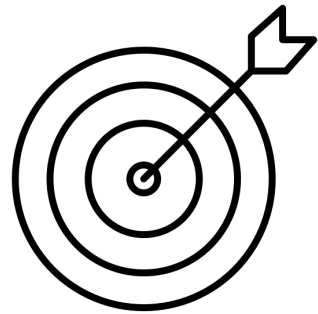
Referenzen

- [1] R. Banik, 'The Movies Dataset'. Accessed: Jun. 17, 2025. [Online]. Available: <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>
- [2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, 'Efficient Estimation of Word Representations in Vector Space', presented at the International Conference on Learning Representations, arXiv, 2013. doi: 10.48550/ARXIV.1301.3781.
- [3] W. Pugh, 'Skip lists: a probabilistic alternative to balanced trees', Commun. ACM, vol. 33, no. 6, pp. 668–676, Jun. 1990, doi: 10.1145/78973.78977.
- [4] Y. A. Malkov and D. A. Yashunin, 'Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs', IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 4, pp. 824–836, Apr. 2020, doi: 10.1109/TPAMI.2018.2889473.
- [5] R. Guo et al., 'Manu: a cloud native vector database management system', Proc. VLDB Endow., vol. 15, no. 12, pp. 3548–3561, Aug. 2022, doi: 10.14778/3554821.3554843.

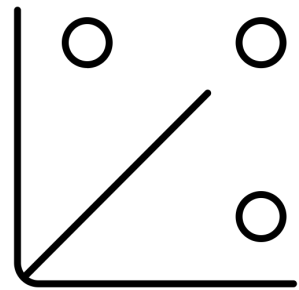
Referenzen

- [6] P. Lewis et al., 'Retrieval-augmented generation for knowledge-intensive NLP tasks', in Proceedings of the 34th International Conference on Neural Information Processing Systems, in NIPS '20. Red Hook, NY, USA: Curran Associates Inc., 2020.
- [7] M. Stonebraker and A. Pavlo, 'What Goes Around Comes Around... And Around...', SIGMOD Record, vol. 53, no. 2, pp. 21–37, Jun. 2024.
- [8] 'SAP HANA Cloud, SAP HANA Database Vector Engine Guide'. Accessed: Jun. 24, 2025. [Online]. Available: <https://help.sap.com/docs/hana-cloud-database/sap-hana-cloud-sap-hana-database-vector-engine-guide/vector-function-reference>
- [9] J. J. Pan, J. Wang, and G. Li, 'Vector Database Management Techniques and Systems', in Companion of the 2024 International Conference on Management of Data, Santiago AA Chile: ACM, Jun. 2024, pp. 597–604. doi: 10.1145/3626246.3654691.

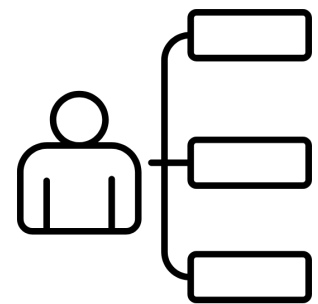
Quellenangaben für Icons



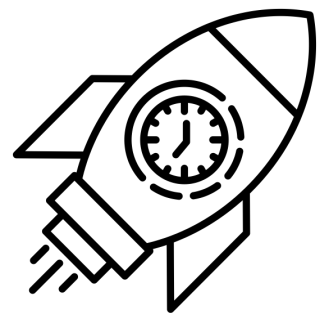
Created by Barracuda from the Noun Project, CC BY-3.0 Attribution License,
<https://thenounproject.com/icon/target-873334/>



Created by Lakshisha from the Noun Project, CC BY-3.0 Attribution License,
<https://thenounproject.com/icon/coordinate-system-1231245/>



Created by Olifernes Tejeros from the Noun Project, CC BY-3.0 Attribution License,
<https://thenounproject.com/icon/use-case-7206637/>



Created by Sumarni from the Noun Project, CC BY-3.0 Attribution License,
<https://thenounproject.com/icon/future-6295390/>



Created by Bestdesignmarket from the Noun Project, CC BY-3.0 Attribution License,
<https://thenounproject.com/icon/magnifier-2039253/>