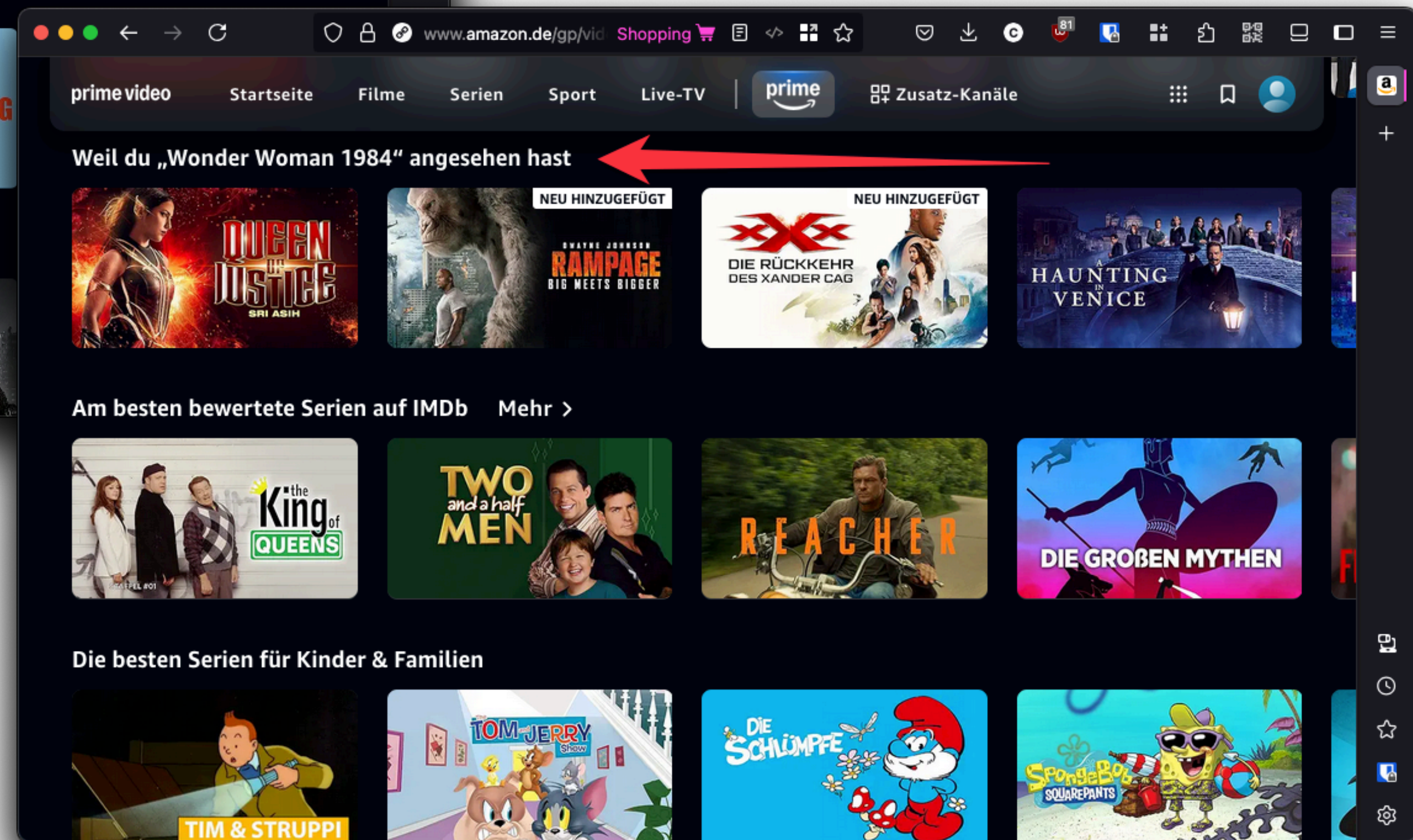
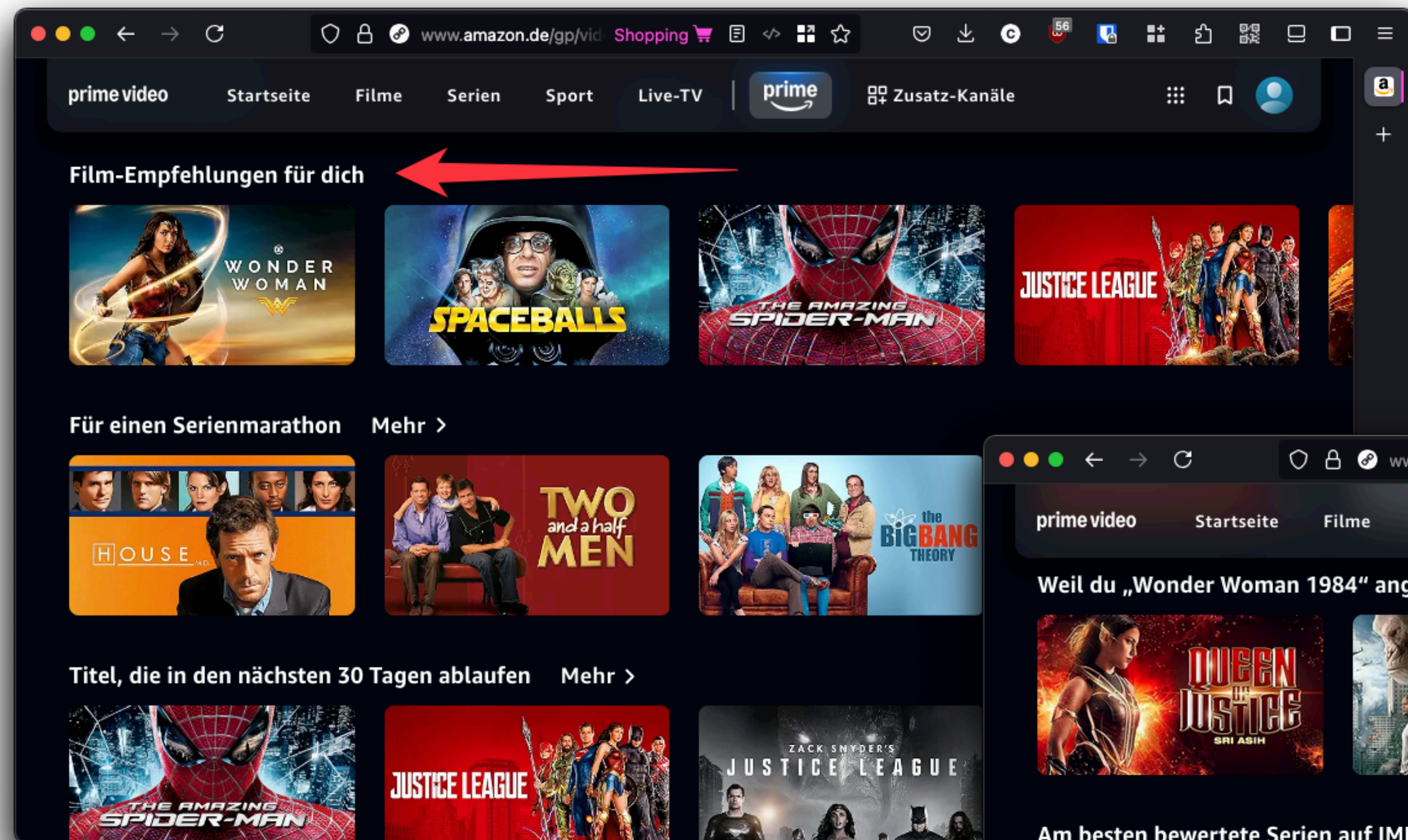


Vektordatenbanken: Grundlagen



Motivation & Lernziel

Was klassische SQL-Datenbanken (noch) nicht können

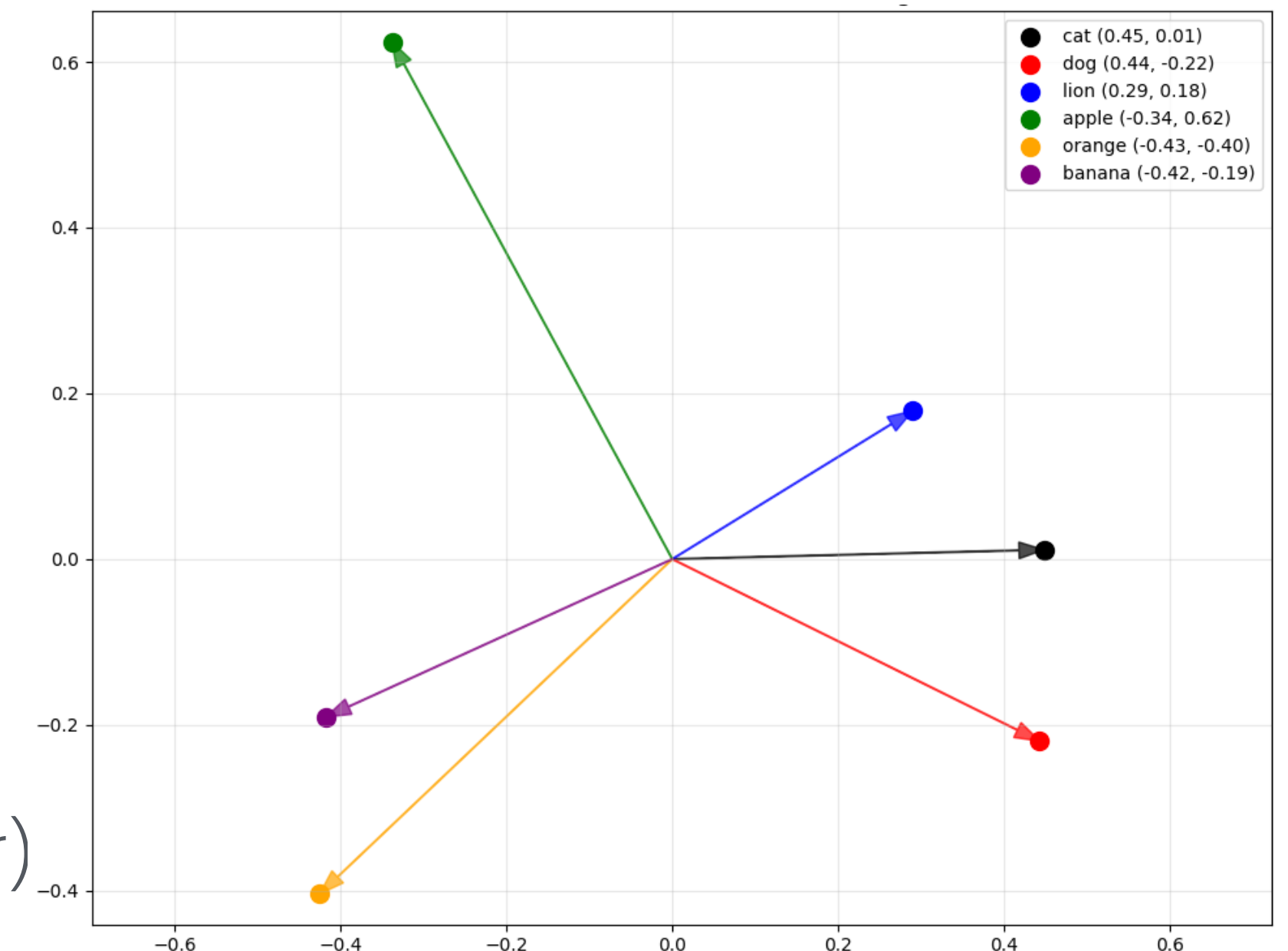
Movie-ID	Title	Overview
...
199	Star Trek: First Contact	The Borg, a relentless race of cyborgs, are on a direct course for Earth. Violating orders to stay away from the battle, Captain Picard and the crew of the newly-commissioned USS Enterprise E ...
343	Harold and Maude	The young Harold lives in his own world of suicide-attempts and funeral visits to avoid the misery of his current family and home environment. Harold meets an 80-year-old woman named Maude ...
680	Pulp Fiction	A burger-loving hit man, his philosophical partner, a drug-addled gangster's moll and a washed-up boxer converge in this sprawling, comedic crime caper. ...
773	Little Miss Sunshine	A family loaded with quirky, colorful characters piles into an old van and road trips to California for little Olive to compete in a beauty pageant.
1368	First Blood	When former Green Beret John Rambo is harassed by local law enforcement and arrested for vagrancy, the Vietnam vet snaps, runs for the hills and rat-a-tat-tats his way into the ...
7326	Juno	A young girl named Juno gets herself pregnant and tries to stand on her own, but soon learns a few lessons about being grown up.
...

Ausschnitt aus einer Film-Datenbank basierend auf [1]

Semantische Suche

Bedeutung als Vektoren: Die Idee hinter Word2Vec

- > 2013 von einem Google-Team vorgestellt [2]
- > Idee
 - > Jedes Wort wird als (hochdimensionaler) **Vektor** repräsentiert
 - > Wörter, die in ähnlichen Kontexten verwendet werden, werden durch "ähnliche" Vektoren repräsentiert
 - > Berechnung der Vektorenrepräsentation erfolgt durch neuronale Netze, die auf Google News Korpus trainiert wurden (eingeschränkt auf die 1 Million häufigsten Wörter)
- > Dieses Vorgehen wird als **Embedding** bezeichnet



2D-Visualisierung von Wort-Vektoren

Semantische Suche

Wie misst man Ähnlichkeit?

- > Wie kann die Ähnlichkeit von Wörtern quantifiziert werden?
 - > Naheliegende Idee: Je näher zwei Vektoren beieinander liegen, desto ähnlicher sind die Inhalte
 - > Mögliche Metriken

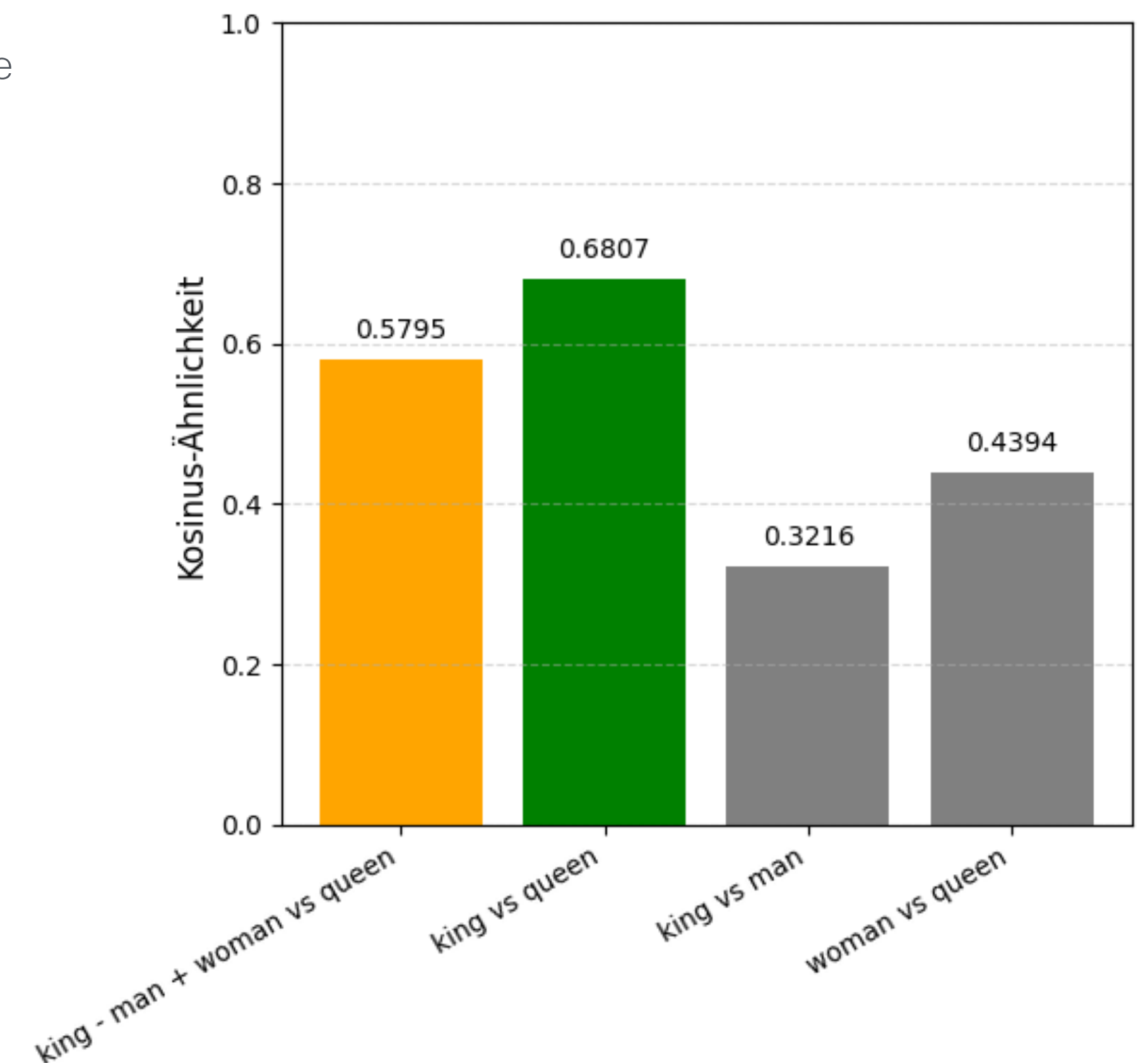
> Skalarprodukt: $\vec{a} \cdot \vec{b} = \sum_{i=1}^n a_i b_i$

> Kosinus-Ähnlichkeit: $\cos(\theta) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$

> Interessante Erkenntnis: 👑 - 👤 + 👩 ≈ 👑

> Verwendung für Suche in Datenbanken (Pseudo-Code):

```
SELECT movie, cosine_similarity(embedding, query) AS similarity
FROM movies
ORDER BY similarity DESC
LIMIT 10
```



Kosinus-Ähnlichkeit von Wort-Vektoren

Anhang & Referenzen

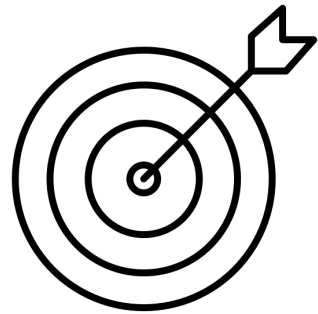
Referenzen

- [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.

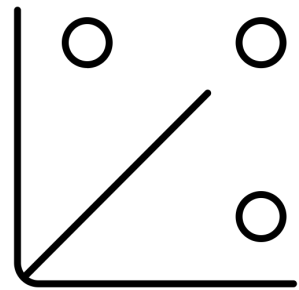
Referenzen

- [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.

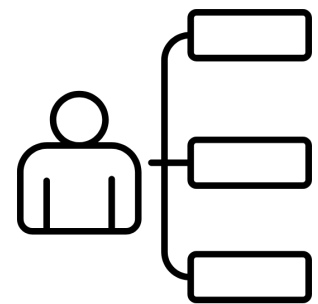
Quellenangaben für Icons



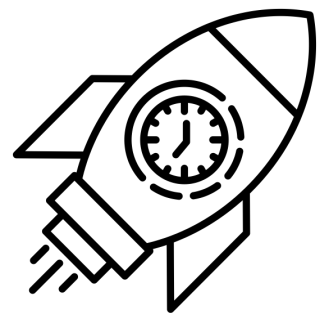
Created by Barracuda from the Noun Project, CC BY-3.0 Attribution License,
<https://thenounproject.com/icon/target-873334/>



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/>