

***Better Life for SEOs –
Wie Vektor-Analysen dir
schneller bessere Ergebnisse
liefern***



Die Geschichte von der unperfekt perfekten Domain- Migration

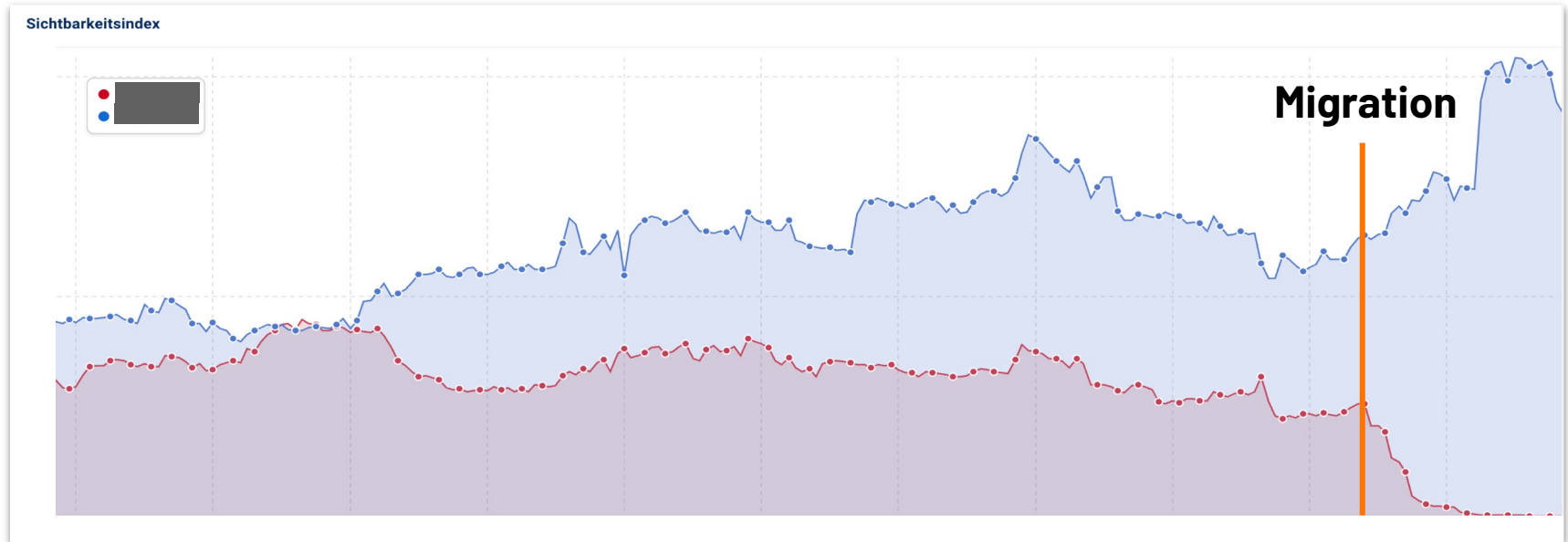




**Ich,
34 Jahre,
SEO Consultant
& nicht mehr
ganz so voller
Vorfreude**



Ende gut, alles gut.



Organische Sichtbarkeit konnte 1:1 übertragen werden

Welche Themen wir uns heute anschauen

- ① Redirect Mapping
- ② Interne Verlinkung
- ③ Query-URL-Matching
- ④ Offtopic-URL-Detection

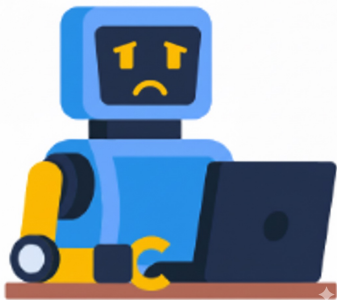


Embeddings

SEO-Analysen

**First things first:
Was sind Vektor-
Embeddings
überhaupt?**

Maschinen “verstehen” Sprache nicht



Der Hund lag zusammengerollt vor der Haustür und wartete geduldig. Bei jedem Geräusch hob er aufmerksam den Kopf, die Ohren gespitzt, die Nase leicht zitternd. Als sich endlich der Schlüssel im Schloss drehte, sprang er auf, wedelte so heftig mit dem Schwanz, dass sein ganzer Körper mitwackelte, und begrüßte sein Herrchen mit freudigem Bellen

Der Hund lag zusammengerollt vor der Haustür und wartete geduldig. Bei jedem Geräusch hob er aufmerksam den Kopf, die Ohren gespitzt, die Nase leicht zitternd. Als sich endlich der Schlüssel im Schloss drehte, sprang er auf, wedelte so heftig mit dem Schwanz, dass sein ganzer Körper mitwackelte, und begrüßte sein Herrchen mit freudigem Bellen



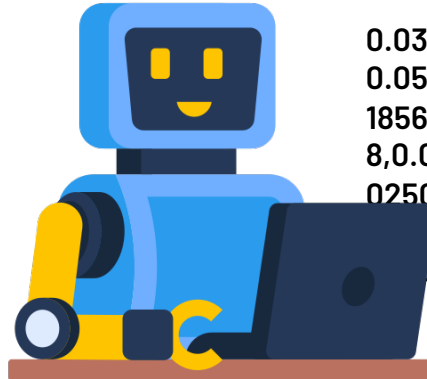
Embedding-Modell

(z. B. SentenceTransformer, OpenAI Embeddings, etc.)



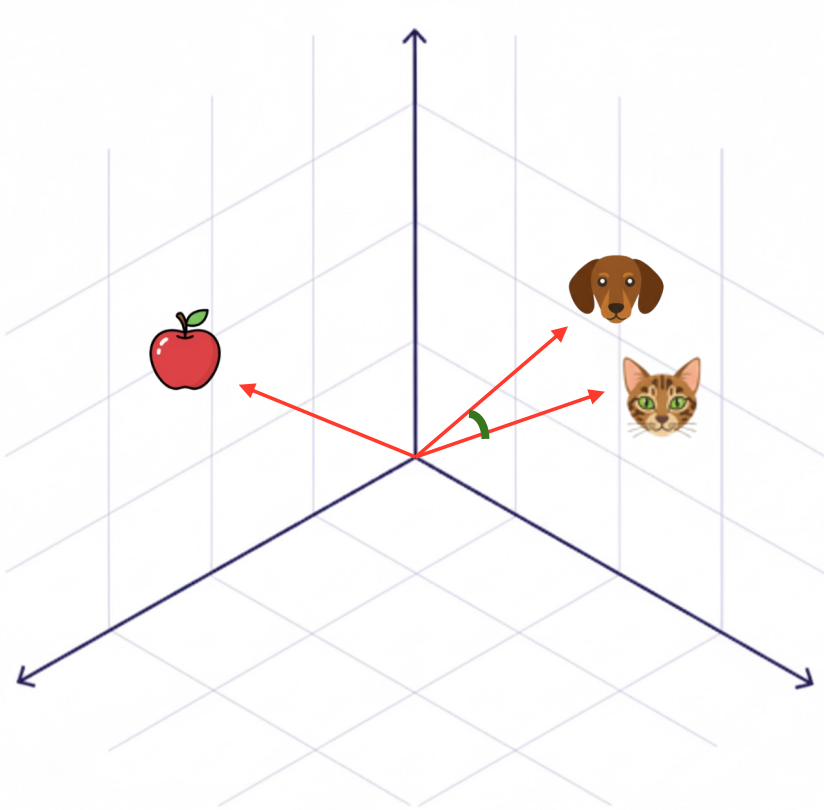
Vektorrepräsentation
des Textes

0.0321007,0.053922765,-0.021664707,-
0.051340055,0.032293856,0.033371683,0.08
185686,0.015691796,0.036515385,0.01811601
8,0.025730614,0.038659245,0.009943789,0.
025016638,0.019693917,0.04457201,0.09978
0.025281407,0.10213552,0.07661153,0.0
8638,0.03336795 (...)



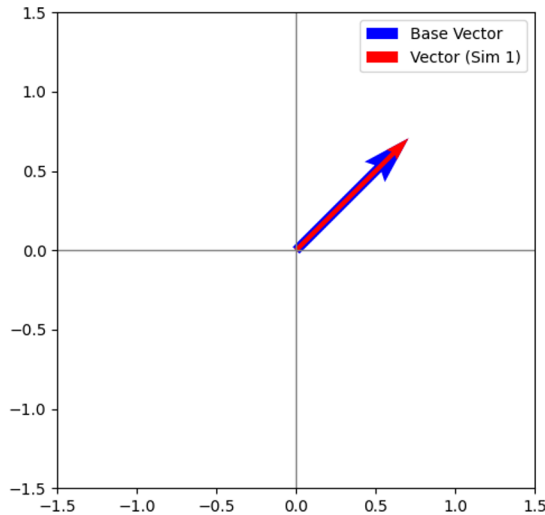
Cosinus Ähnlichkeit

Es gilt: **Je kleiner der Winkel zwischen zwei Vektoren, umso ähnlicher sind sich diese.**



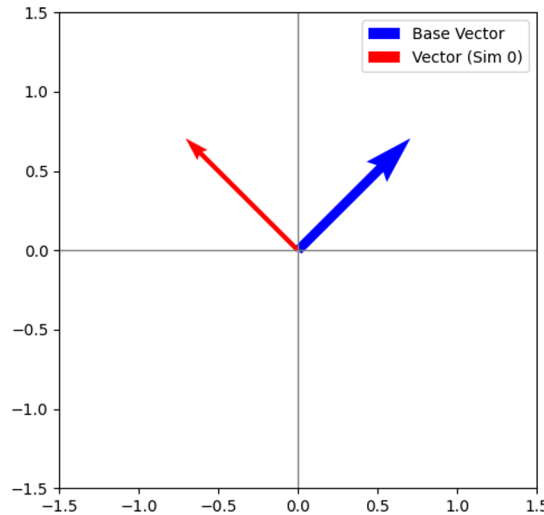
Cosine Similarity: Je kleiner der Winkel, umso ähnlicher

Cosinus Ähnlichkeit = 1



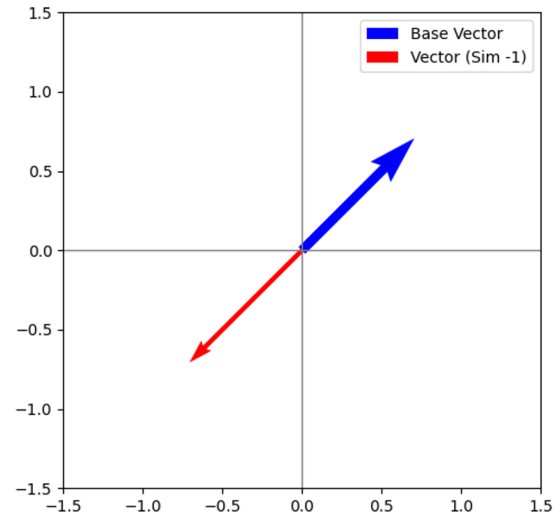
identisch

Cosinus Ähnlichkeit = 0



keine semantische Beziehung

Cosinus Ähnlichkeit = -1

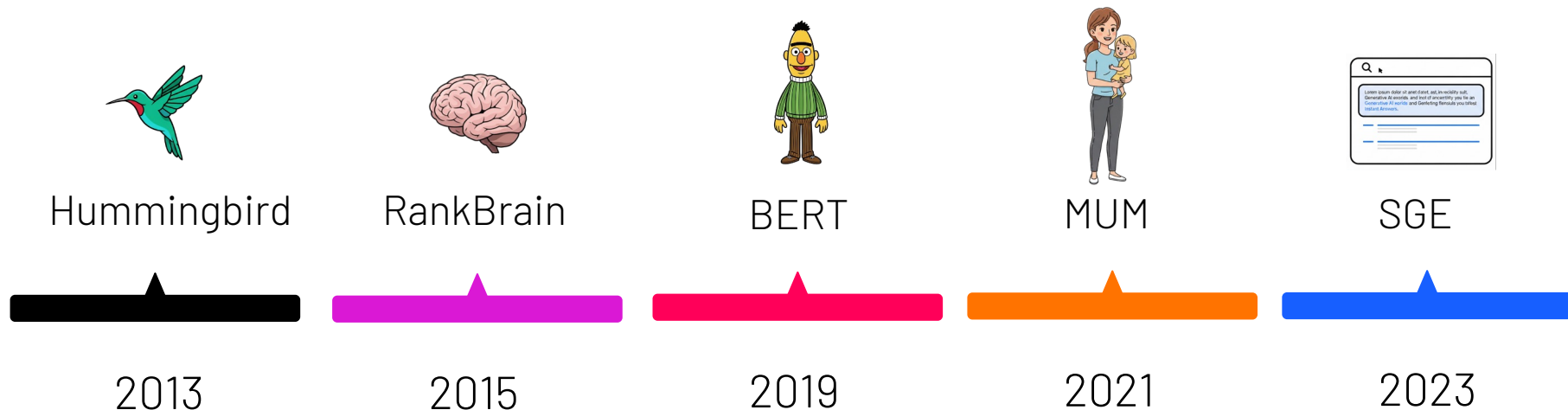


gegensätzlich

**Warum sind SEO-
Analysen auf
Vektor- Embedding
Basis sinnvoll?**

Vektor-Embeddings liefern
einen Einblick, wie **moderne**
Suchmaschinen und auch
LLMs Inhalte verstehen.

"Things not strings" – Googles Weg zur semantischen Suchmaschine



Google hat den Wandel zur semantischen Suche über Jahre hinweg vollzogen

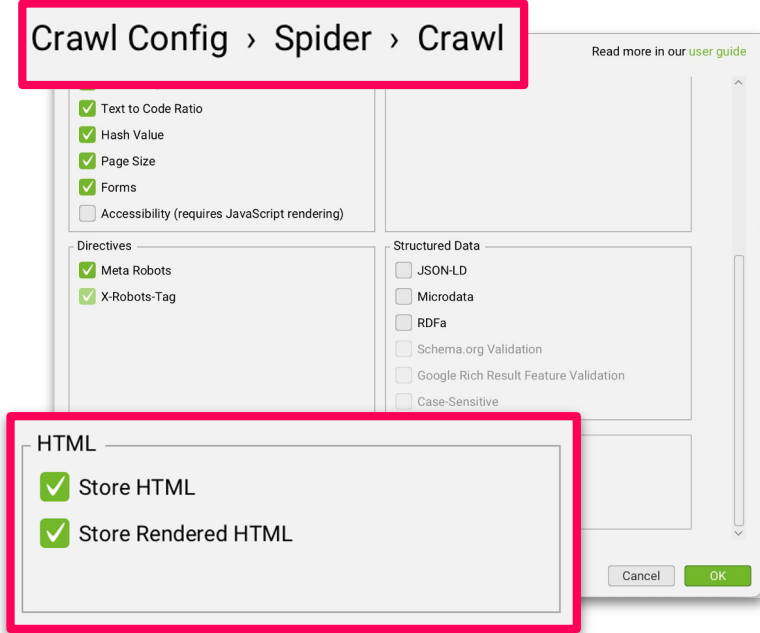
**Wie könnt ihr
Vektor-Embeddings
erzeugen?**

Zum Beispiel im Screaming Frog



Embeddings werden generiert während der Crawl läuft.

HTML speichern

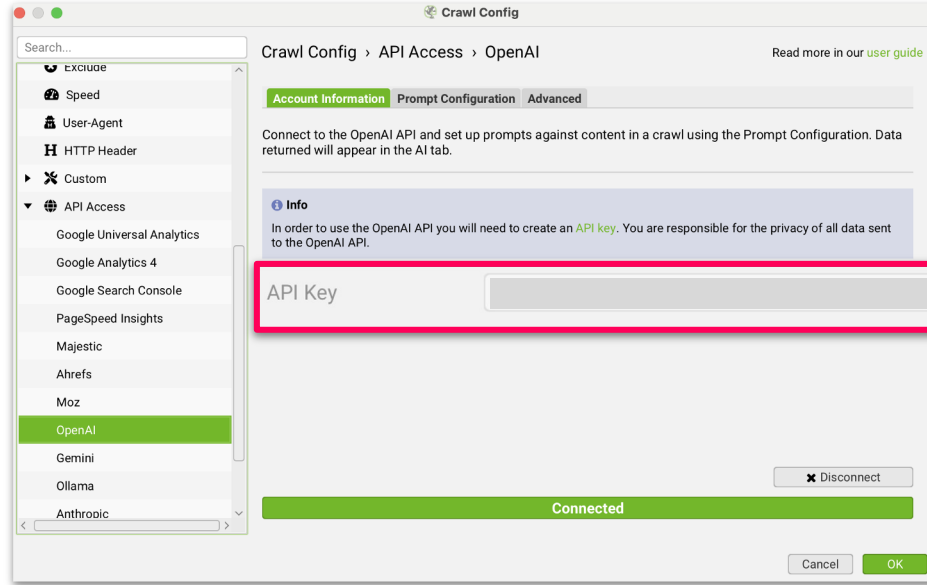


Configuration > Crawl Config

„Store HTML“ anhaken

Wenn ihr mit
JavaScript crawlt,
„Store Rendered“
HTML auswählen

API-Key besorgen und hinterlegen



Configuration > API Access

API-Key eintragen

Falls ihr noch keinen API-Key habt:

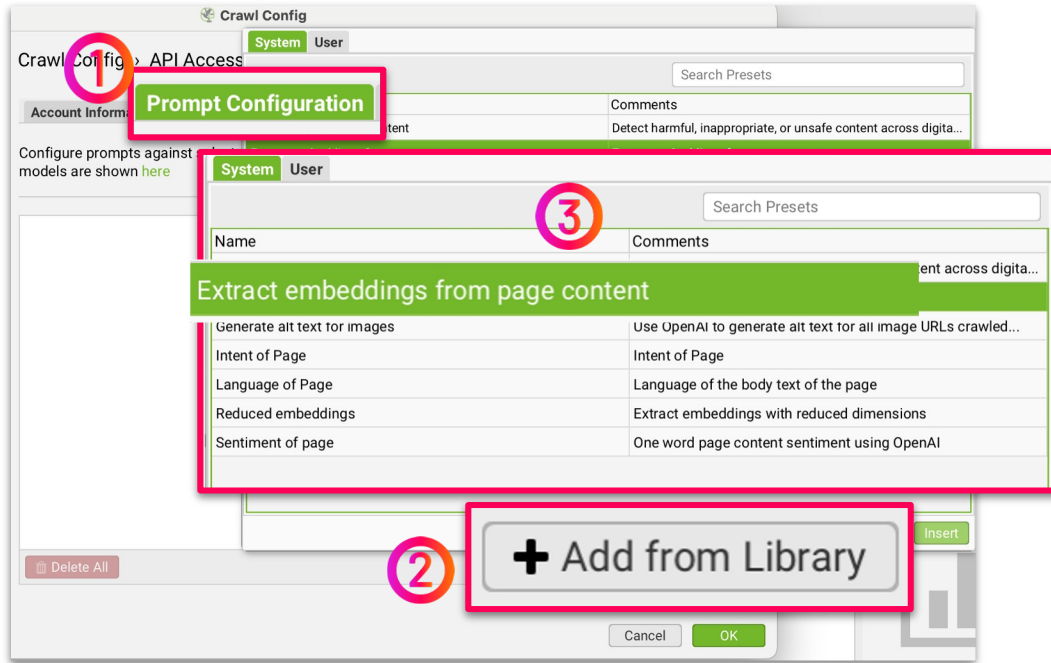


OpenAI



Gemini

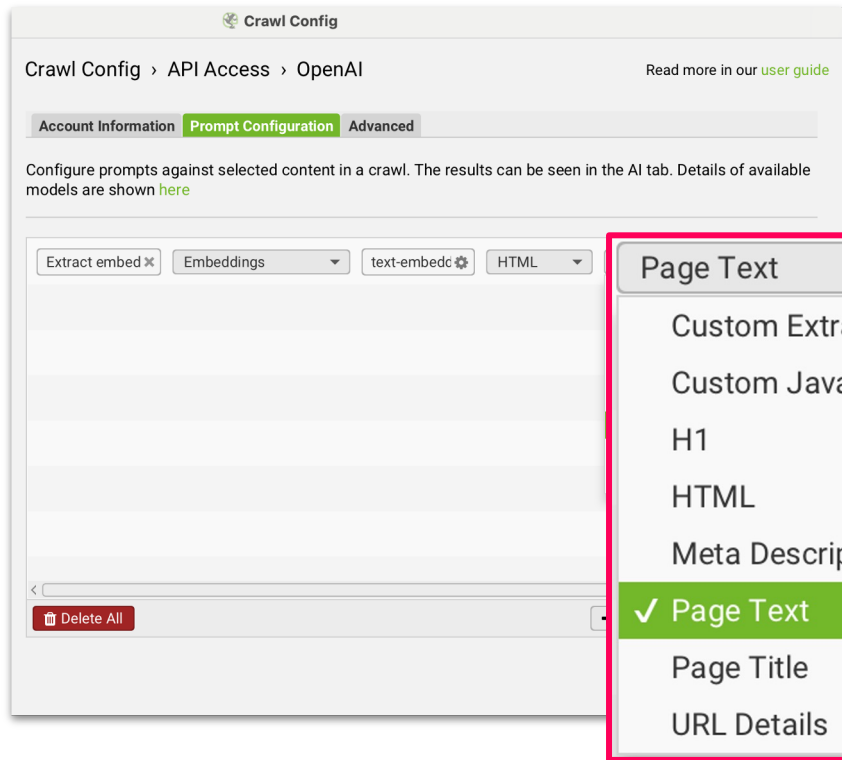
Prompt auswählen



Configuration > API Access

- 1 Tab **Prompt Configuration** auswählen
- 2 + **Add from Library**
- 3 **Extract embeddings from page content** auswählen & Insert

“Embedding-Content” auswählen



Configuration > API Access

Auf **Basis welcher URL-Inhalte** sollen die **Embeddings erstellt** werden?

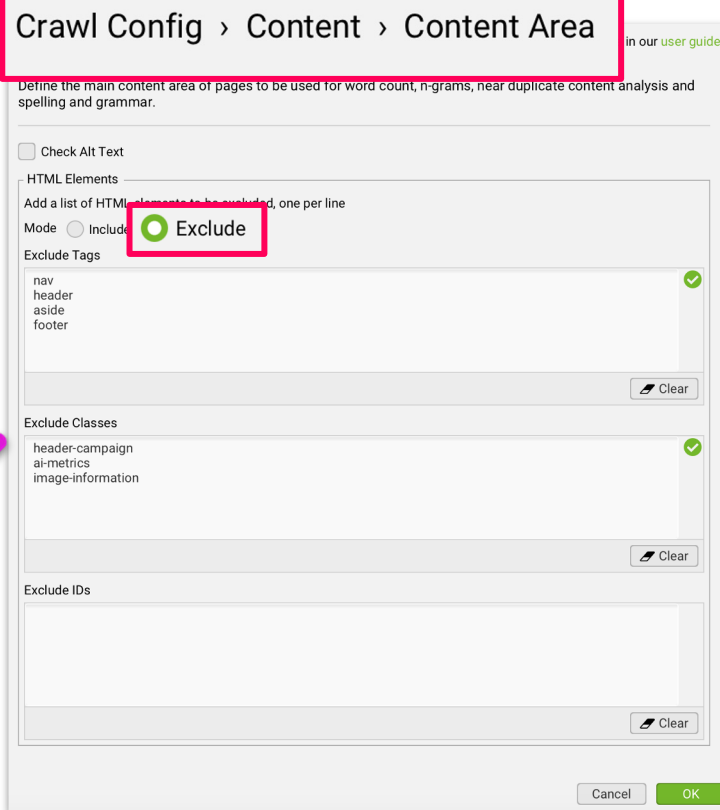
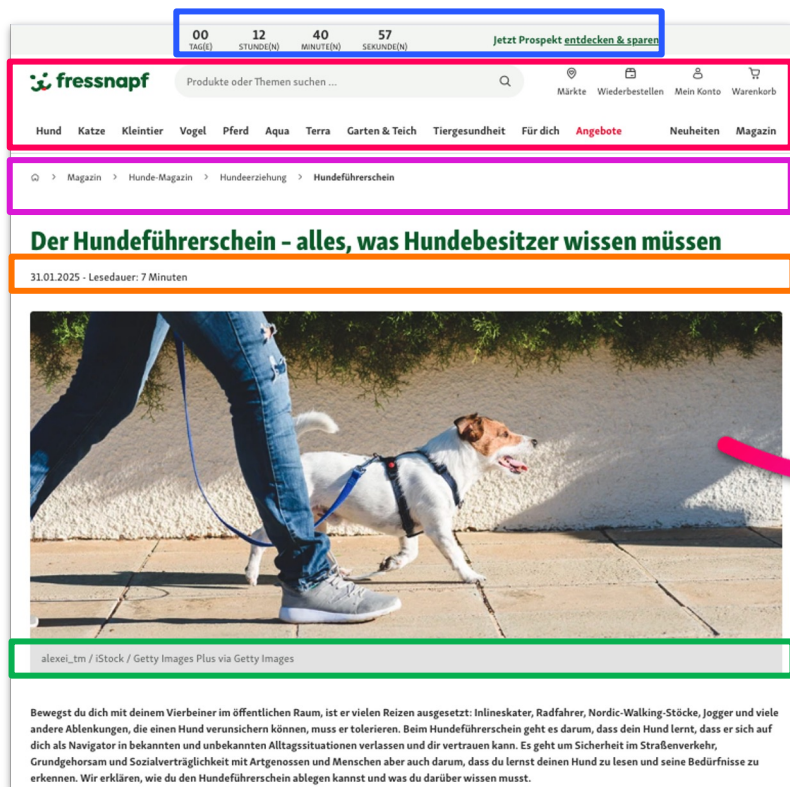
**Wählt euren
Embedding-Content
weise**

Semantische Analysen sind nur so gut wie die Embeddings, auf denen sie basieren



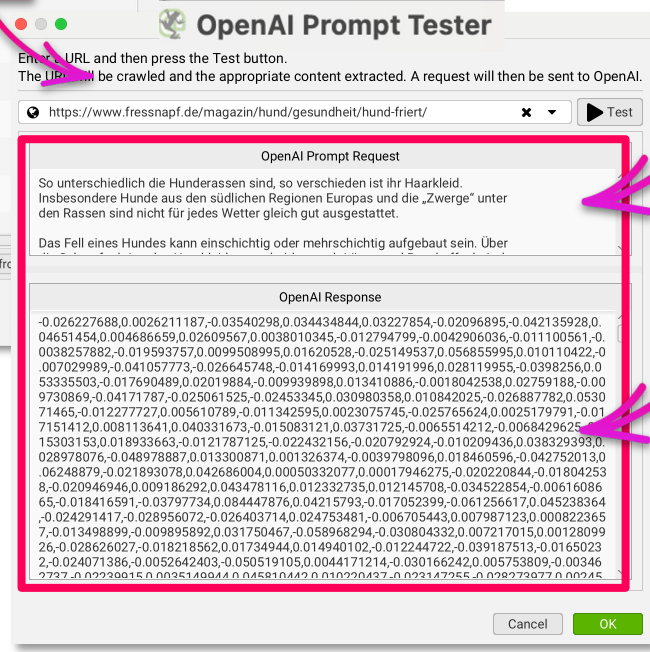
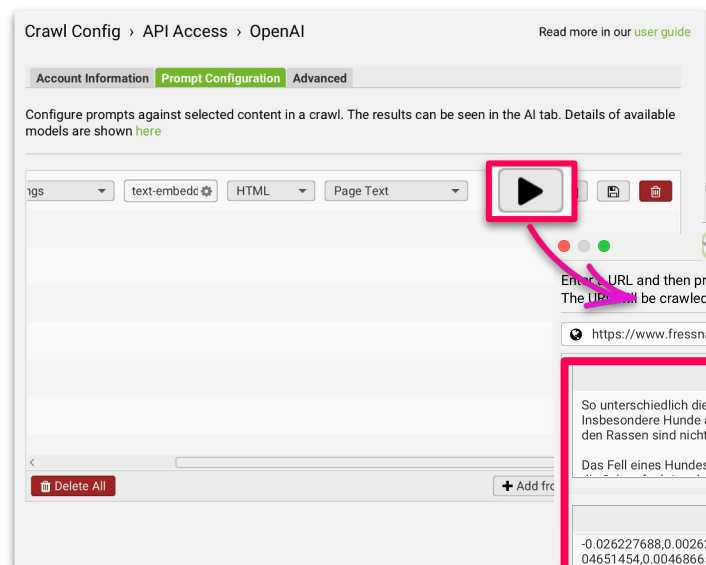
Der Screaming Frog bietet die Möglichkeit, bestimmte **Content Bereiche** – etwa Boilerplate-Content – gezielt vom Crawl **auszuschließen**.

Wählt euren “Embedding-Content” weise



**Überprüft welcher
Content für die
Embedding-Erzeugung
verwendet wird**

(1) Überprüft, welche Inhalte den Embeddings zugrunde liegen



Ans Modell übermittelter
Content

LLM-Response (Embedding)

(2) Überprüft, welche Inhalte den Embeddings zugrunde liegen

The screenshot shows the Screaming Frog SEO tool interface. At the top, the URL bar displays <https://www.fressnapf.de/magazin/>. Below the navigation tabs, a table lists various URLs. A red circle with the number '1' highlights the URL <https://www.fressnapf.de/magazin/katzen/erziehung/>, with the text 'URL auswählen' (Select URL) next to it. Another red circle with the number '3' highlights the 'Visible Content' tab, with the text 'Visible Content' next to it. Below the table, the 'Visible Content' and 'Rendered' sections are shown side-by-side. A red circle with the number '2' highlights the text 'Text für Embeddings' (Text for Embeddings) in the 'Rendered' section. The text in the 'Rendered' section is highlighted in green, and the text 'View Source' is written below it. The bottom of the interface shows various tabs like 'URL Details', 'Inlinks', 'Outlinks', 'Image Details', 'Resources', 'SERP Snippet', 'Rendered Page', 'Chrome Console Log', 'View Source', 'HTTP Headers', 'Cookies', 'Duplicate Details', 'Structured Data Details', and 'Lighthouse Details'.

Row	Address	Content Type	Status Code	Status	Indexability	Indexability Status	Extract
10	https://www.fressnapf.de/magazin/katzen/erziehung/	text/html;charset=utf-8	200		Indexable		0.00345
11	https://www.fressnapf.de/magazin/pferd/	text/html;charset=utf-8	200		Indexable		0.00475
12	https://www.fressnapf.de/magazin/katzen/news/				Indexable		0.00608
13	https://www.fressnapf.de/magazin/kleintiere/beschaeftigung/				Indexable		0.01944
14	https://www.fressnapf.de/magazin/katzen/pflege-hygiene/				Indexable		0.01625
15	https://www.fressnapf.de/magazin/katzen/kitten-entwicklung/	text/html;charset=utf-8	200		Indexable		0.01900
16	https://www.fressnapf.de/magazin/terra/	text/html;charset=utf-8	200		Indexable		0.04197
17	https://www.fressnapf.de/magazin/katzen/kitten-entwicklung/	text/html;charset=utf-8	200		Indexable		0.02385
18	https://www.fressnapf.de/magazin/katzen/kitten-entwicklung/	text/html;charset=utf-8	200		Indexable		0.02672

Selected Cells: 1 Filter Total: 139

Visible Content

Rendered

Die Entwicklung von Katzenbabys

Sie werden so schnell erwachsen – das gilt für Katzenbabys noch mehr als für kleine Menschen. Das Kätzchenalter dauert nur wenige Monate an, während derer die kleinen Katzen rasante Entwicklungsschritte absolvieren. Aber in wie vielen Phasen wird aus dem zahlosen, blinden Fellwürmchen ein selbstbewusster Stubentiger? Lies hier, in welchen Schritten die Entwicklung des Katzenbabys in Wochen abläuft.

Inhalt

- Vor der Geburt: Wie entwickelt sich das Katzenbaby im Bauch der Mutter?
- Endlich auf der Welt: Welche Entwicklungsstufen durchlaufen die Sinne des Kitten?
- Das Kitten und der Rest der Welt: In welchen Stadien verändert sich das Verhalten?
- Vom Katzenbaby zur Jungkatze: Wann ist das Kitten erwachsen?

Vor der Geburt: Wie entwickelt sich das Katzenbaby im Bauch der Mutter?

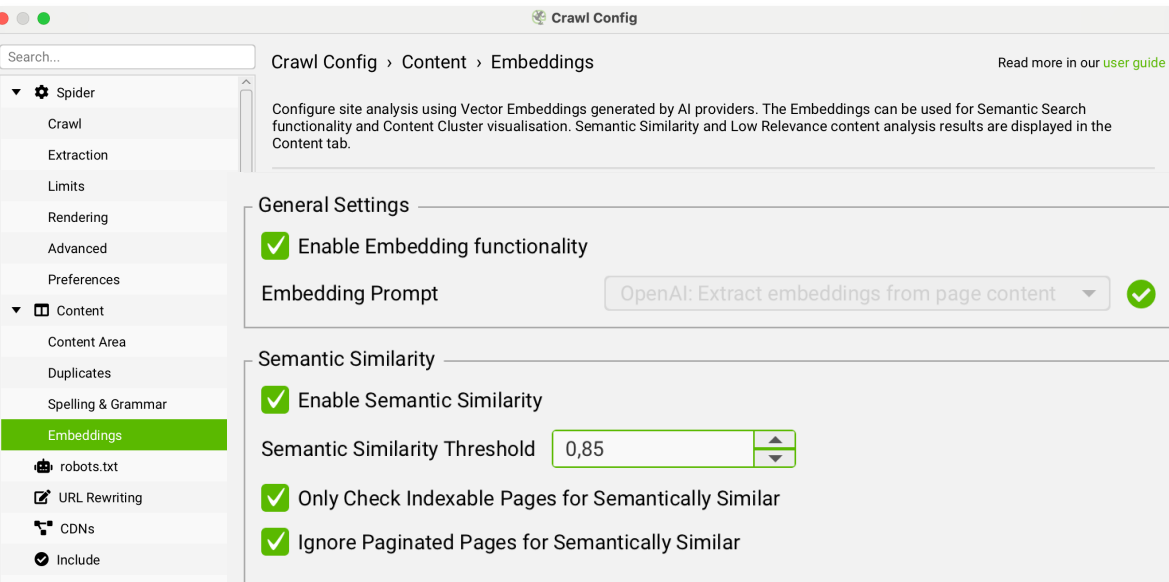
Das Katzendasein beginnt mit dem Stelldichein zwischen Mutterkatze und Kater. Im Moment des Paarungsaktes werden die Eizellen „ausgelöst“, die Kater sind 24 Stunden lang lebensfähig. Eine gegläute Befruchtung ist damit Startschuss für das Katzenleben. Etwa 46 Stunden nach der Befruchtung beginnt

Text für Embeddings

View Source

Bonus-Tipp

Semantisch ähnliche URLs direkt im Screaming Frog berechnen lassen



Configuration > Content > Embeddings

Nachdem ihr den Embedding-Prompt eingefügt habt (vgl. Slide 18), könnt ihr euch theoretisch auch direkt vom Screaming Frog semantisch ähnliche URL-Paare ausgeben lassen

Jetzt aber zu den Use Cases

(1) Redirect Mapping



**Damit ihr euch
beim nächsten
Relaunch nicht
unter der Decke
verstecken müsst**



Es muss **NICHT**
immer
Embedding-
Magic sein.



Welche Informationen stehen euch zur Verfügung?

1 Eindeutige Identifier sind vorhanden und bleiben vor und nach dem Relaunch bzw. der Domain-Migration unverändert, z. B.

- Gleichbleibende URL-Struktur / URL-Slug
- Identischer Title Tag oder H1 etc.
- sku
- ISBN-Nummer
- ...

2 Keine eindeutigen Identifier vorhanden

- wie z. B. bei einer Domain-Migration von zwei unabhängigen Marken

Eindeutige Identifier sind zu priorisieren



Identische
H1 als
Identifier
für
Redirect-
Mapping



**Wenn keine eindeutigen
Identifizier vorhanden sind,
setzen wir auf semantisches
Mapping mit Vektor-
Embeddings.**

Machen wir es konkret



Fiktiver Case:

Fressnapf übernimmt
Zooroyal.

Es sind keine eindeutigen
Identifizier auszumachen.

Werfen wir einen Blick auf die **Magazine**.

Ergebnisse

Redirect-Mapping

Beispiele semantisches Matching

Zu migrierende URL	Best-Match URL	Sim
https://www.zooroyal.de/magazin/hunde/hunde-im-einsatz/rettungshunde/	https://www.fressnapf.de/magazin/hund/sport-spiel/rettungshund/	0,8961
https://www.zooroyal.de/magazin/katzen/ernaehrung-katzen/	https://www.fressnapf.de/magazin/katze/ernaehrung/	0,9183
https://www.zooroyal.de/magazin/aquaristik/aquarienbewohner/	https://www.fressnapf.de/magazin/aquaristik/aquarienbewohner/	0,8738
https://www.zooroyal.de/magazin/hunde/ernaehrung-hunde/duerfen-hunde-kirschen-essen/	https://www.fressnapf.de/magazin/hund/ernaehrung/duerfen-hunde-kirschen-essen/	0,9205
...		

Beispiele semantisches Matching

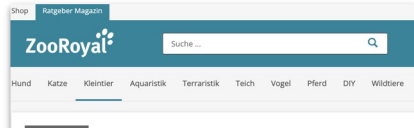
Zu migrierende URL	Best-Match URL	Sim
https://www.zooroyal.de/magazin/hunde/hunde-im-einsatz/rettungshunde/	https://www.fressnapf.de/magazin/hund/sport-spiel/rettungshund/	0,8961
https://www.zooroyal.de/magazin/katzen/ernaehrung-katzen/	https://www.fressnapf.de/magazin/katze/ernaehrung/	0,9183
https://www.zooroyal.de/magazin/aquaristik/aquarienbewohner/	https://www.fressnapf.de/magazin/aquaristik/aquarienbewohner/	0,8738
https://www.zooroyal.de/magazin/hunde/ernaehrung-hunde/duerfen-hunde-kirschen-essen/	https://www.fressnapf.de/magazin/hund/ernaehrung/duerfen-hunde-kirschen-essen/	0,9205
...		

Werte nahe 1 bei der Cosine Similarity stehen für eine **hohe inhaltliche Übereinstimmung**

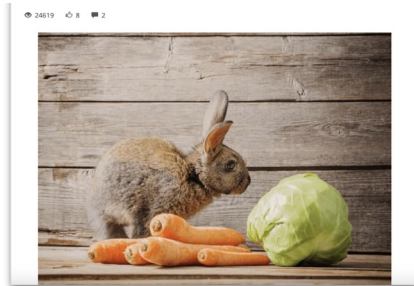
Okay, das hätten wir auch ohne Embeddings hingekriegt – Thank you, **Captain Obvious!**

Zu migrierende URL	Best-Match URL	Sim
https://www.zooroyal.de/magazin/hunde/hunde-im-einsatz/rettungshunde/	https://www.fressnapf.de/magazin/hund/sport-spiel/rettungshund/	0,8961
https://www.zooroyal.de/magazin/katzen/ernaehrung-katzen/	https://www.fressnapf.de/magazin/katze/ernaehrung/	0,8937
https://www.zooroyal.de/magazin/aquaristik/aquarienbewohner/	https://www.fressnapf.de/magazin/aquaristik/aquarienbewohner/	0,8937
https://www.zooroyal.de/magazin/hunde/ernaehrung-hunde/duerfen-hunde-kirschen-essen/	https://www.fressnapf.de/magazin/hund/ernaehrung/duerfen-hunde-kirschen-essen/	0,8935
...		

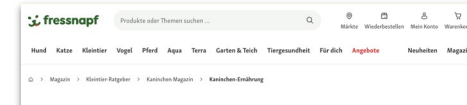




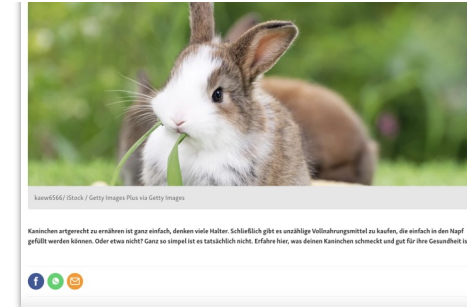
Delikatessen für Kleintiere



Best-Match URL



Kaninchen-Ernährung



	URL-Slug	Title Tag	H1
Zooroyal	/delikatessen-aus-dem-garten/	Delikatessen für Kleintiere aus dem Garten ZooRoyal Magazin	Delikatessen für Kleintiere aus dem Garten
Fressnapf	/kaninchen/ernaehrung/	Kaninchen-Ernährung: Infos & Tipps FRESSNAPF	Kaninchen-Ernährung: von Gräsern und Möhrchen

Semantisches Matching entfaltet hier sein volles Potenzial!

**Noch ein Tipp für
die Arbeit mit
Embeddings**

Viel hilft viel – zumindest bei Embeddings

Zooroyal-URL	Matched Fressnapf-URL basierend auf Title Tag
/magazin/katzen/ernaehrung-katzen/ katze-frisst-nicht /	/katze/ausstattung/ kosten-katze /

Title: **Katze** frisst nicht - Ursachen, **Tipps & Infos** für Katzenhalter



Title: **Wie viel kostet eine Katze? Infos & Tipps** | FRESSNAPF

Viel hilft viel – zumindest bei Embeddings

Die Erfahrung zeigt: Je mehr relevante Inhalte in die Generierung der Embeddings einfließen, desto präziser und belastbarer werden sie. Beispiel:

Zooroyal-URL	Matched Fressnapf-URL bei Embedding basierend auf Main Content	Matched Fressnapf-URL bei Embedding basierend auf Title Tag	Matched Fressnapf-URL bei Embedding basierend auf H1	Matched Fressnapf-URL bei Embedding basierend auf Title, MD und H1
/magazin/katzen/ernaehrung-katzen/ katze-frisst-nicht /	/katze/gesundheit/ katze-frisst-nicht-sensibilitaet /	/katze/ausstattung/ kosten-katze /	/katze/ gesundheit /	/katze/gesundheit/ katze-frisst-nicht-sensibilitaet /



Beim nächsten Relaunch kann euch dieses Tool helfen



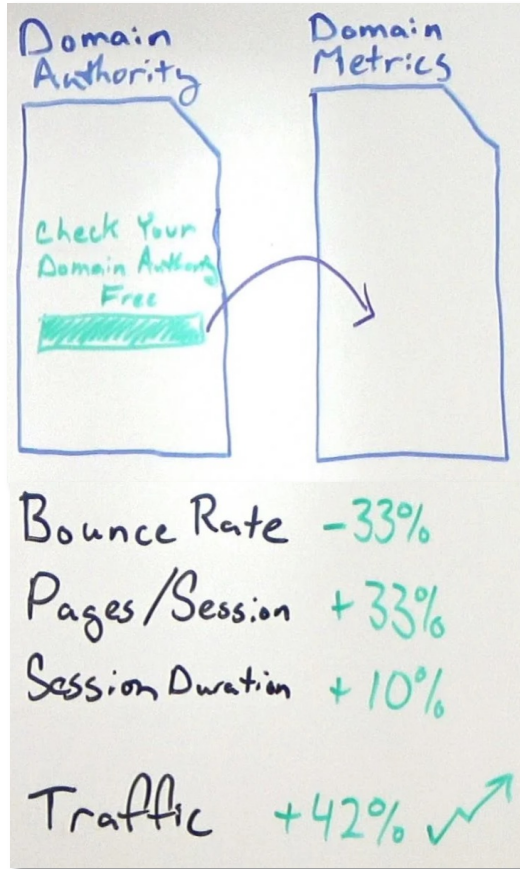
The screenshot shows the ONE Redirector tool interface. At the top is the logo 'one beyond search' in pink and orange. Below it is the title 'ONE Redirector – finde die passenden Redirect-Ziele' with a blue icon of two arrows forming a square. A link '> Was macht das Tool? (Erklärung & Tipps)' is visible. The main section is titled '1. Dateien hochladen'. It contains two identical blocks: 'Datei mit den URLs, die weitergeleitet werden sollen (CSV oder Excel)' and 'Datei mit den Ziel-URLs (CSV oder Excel)'. Each block has a cloud icon and the text 'Drag and drop file here' and 'Limit 200MB per file • CSV, XLSX'.

**In max. 5 Minuten zum
Redirect-Mapping**



[Zum Tool](#)

(2) Interne Verlinkung



Thematisch relevante
interne Links helfen der
verlinkenden **UND** der
verlinkten Seite

Google Patents

Ranking documents based on user behavior and/or feature data

Abstract

A system generates a model based on feature data relating to different features of a link from a linking document to a linked document and user behavior data relating to navigational actions associated with the link. The system also assigns a rank to a document based on the model.

Images (8)

US7716225B1
United States

Download PDF Find Prior Art Similar

Inventor: Jeffrey A. Dean, Corin Anderson, Alexis Battle
Current Assignee: Google LLC

Worldwide applications

← Back to results (User-sensitive PageRank);

Class:

User-sensitive PageRank

Abstract

Techniques are described for generating an authority value of a first one of a plurality of documents. A first component of the authority value is generated with reference to outbound links associated with the first document. The outbound links enable access to a first subset of the plurality of documents. A second component of the authority value is generated with reference to a second subset of the plurality of documents. Each of the second subset of documents represents a potential starting point for a user session. A third component of the authority value is generated representing a likelihood that a user session initiated by any of a population of users will end with the first document. The first, second, and third components of the authority value are combined to generate the authority value. At least one of the first, second, and third components of the authority value is computed with reference to user data relating to at least some of the outbound links and the second subset of documents.

Images (3)

Classifications

US9495452B2
United States

Download PDF Find Prior Art Similar

Inventor: Pavel Berkhin, Usama M. Fayyad, Prabhakar Raghavan, Andrew Tomkins
Current Assignee: Starboard Value Intermediate Fund Lp As Collateral Agent, Altaia Inc

Worldwide applications

2006 · US 2009 · US

Application US12/576,679 events

2009-10-09 · Priority to US12/576,679
2009-10-09 · Application filed by Yahoo Inc until 2017
2010-01-28 · Publication of US20100023513A1
2016-11-15 · Publication of US9495452B2
2016-11-15 · Application granted

Google Patente wie User-Sensitive PageRank oder das "Topical PageRank Patent" unterstreichen dies:

Ein Link vererbt **mehr Power je größer die thematische Nähe zwischen verlinkender und verlinkter Seite ist.**

**Im Kern geht es für uns SEOs
meistens um diese Frage**

**Welche
thematisch
ähnlichen Seiten
verlinken
einander noch
nicht?**



“Früher” sind wir so vorgegangen, um thematisch ähnliche Seiten zu finden

site:<https://www.fressnapf.de/magazin/>
hund barfen -inurl:[hund/ernaehrung/barfen/](https://www.fressnapf.de/magazin/hund/ernaehrung/barfen/)



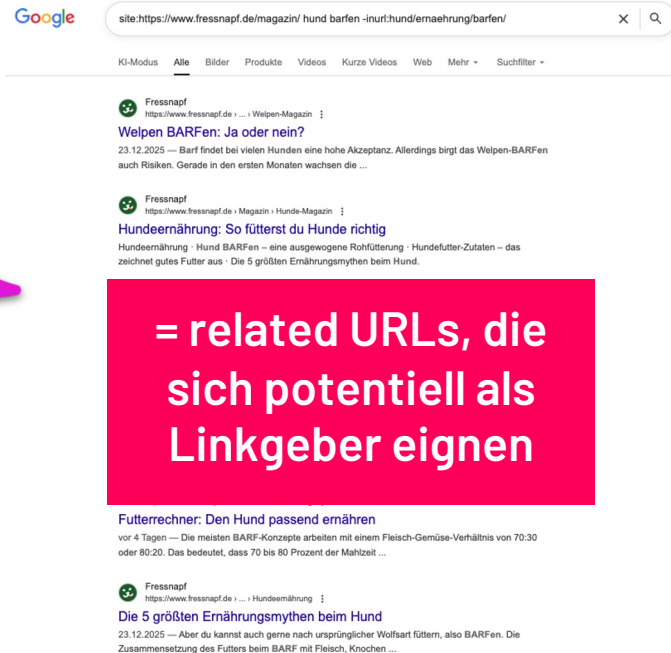
<https://www.fressnapf.de/magazin/> :
Verzeichnis aus dem die related URLs aka
Linkvorschläge kommen sollen



hund barfen: Thema meiner Target URL
(= linkempfangende Seite)



[hund/ernaehrung/barfen/](https://www.fressnapf.de/magazin/hund/ernaehrung/barfen/): Target URL (diese soll natürlich
nicht bei den Linkvorschlägen auftauchen)



**Diesen Umweg / Aufwand
können wir uns jetzt sparen.**

**Wir nutzen Vektor-Embeddings
(und ein Tool).**



Daten sammeln



Vektor-Embeddings

All Inlinks

Linkmetriken je URL

(Unique Inlinks & Outlinks, Link Score)

Inhaltliche Seitenelemente

(Title, MD, H1, ggf. Main Content)



Anzahl Backlinks pro URL

Anzahl Referring Domains
pro URL

2

Dateien im Linktool hochladen (A1)

Einstellungen

Matching-Backend (Auto-Switch bei Bedarf) ⓘ

☒ Exakt (NumPy) ☐ Schnell (FAISS)

Gewichtung (Linkpotenzial)

Das Linkpotenzial gibt Aufschluss über die Luktativität einer URL als Linkgeber.

Interner Link Score 0.30 ⓘ

PageRank-Horder-Score 0.35 ⓘ

Referring Domains 0.20 ⓘ

Backlinks 0.15 ⓘ

Schwellen & Limits (Related URLs)

Ähnlichkeitsschwelle 0.80 ⓘ

Anzahl Related URLs 10 ⓘ

Scope der URL-Auswahl

Welche URLs sollen in die Analysen einbezogen werden? ⓘ

☒ Alle URLs (Standard) ☐ Regex-basiert (nur passende URLs)

Ankertext-Vorschläge (optional, A1)

☐ KI-Ankertext-Vorschläge für Ziel-URLs generieren ⓘ



ONE Link Intelligence

Entwickelt von [Daniel Kremer](#) von [ONE Beyond Search](#) | Folge mir auf [LinkedIn](#) für mehr SEO-Insights und Tool-Updates

> Was du mit dem Tool machen kannst und wie du es nutzt

Welche Analysen möchtest du durchführen?

Mehrfachauswahl möglich

Interne Verlinku... x

FREE

Analyse 1 interne Verlinkungsmöglichkeiten finden – erforderliche Dateien

URLs + Embeddings (CSV/Excel) ⓘ



Drag and drop file here

Limit 200MB per file • CSV, XLSX, XLSM, XLS

Browse files

Bitte lade die Datei für URLs + Embeddings (CSV/Excel).

All Inlinks (CSV/Excel) ⓘ



Drag and drop file here

Limit 200MB per file • CSV, XLSX, XLSM, XLS

Browse files

Bitte lade die Datei für All Inlinks (CSV/Excel).

Linkmetriken (CSV/Excel) ⓘ



Drag and drop file here

Limit 200MB per file • CSV, XLSX, XLSM, XLS

Browse files

Bitte lade die Datei für Linkmetriken (CSV/Excel).

Backlinks (CSV/Excel) ⓘ



Drag and drop file here

Limit 200MB per file • CSV, XLSX, XLSM, XLS

Browse files

Bitte lade die Datei für Backlinks (CSV/Excel).

Ergebnisse

Ergebnisse

Ziel-URL	Related URL	Cosinus Ähnlichkeit	Link von Related URL auf Ziel-URL vorhanden?	Content-Link von Related URL auf Ziel-URL vorhanden?	Link-potenzial
https://www.fressnapf.de/magazin/hund/ernaehrung/ /uebergewicht-hund/	https://www.fressnapf.de/magazin/hund/ernaehrung/ /hundenahrung-rassen/	0.858	ja	ja	0.2518
	https://www.fressnapf.de/magazin/hund/ernaehrung/ /hund-futterumstellung/	0.855	ja	ja	0.1807
	https://www.fressnapf.de/magazin/hund/ernaehrung/ /futterrechner/	0.849	nein	nein	0.0829
	https://www.fressnapf.de/magazin/hund/ernaehrung/ /hundefutter-selber-kochen/	0.849	ja	ja	0.4121
	https://www.fressnapf.de/magazin/hund/ernaehrung/ /ernaehrungsberatung/	0.844	nein	nein	0.5162

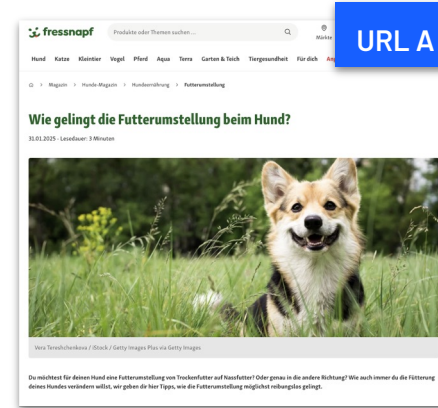
Übergewicht Hund



Cosinus Sim 0,855

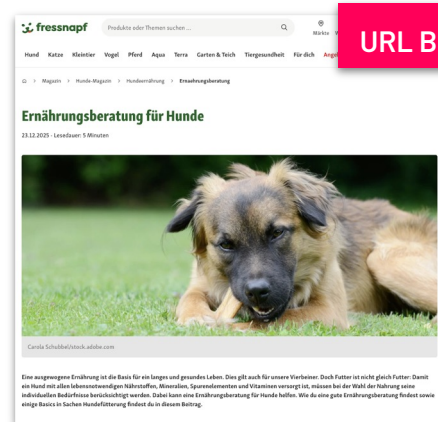
Cosinus Sim
Schwellenwert
0,8

Cosinus Sim 0,844



URL A

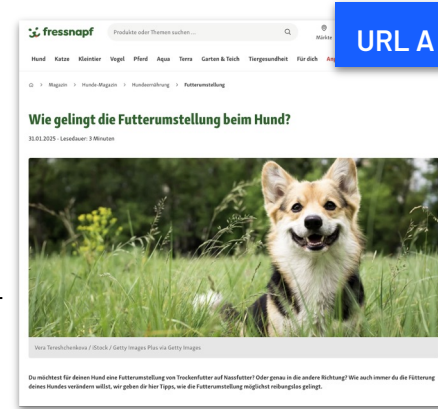
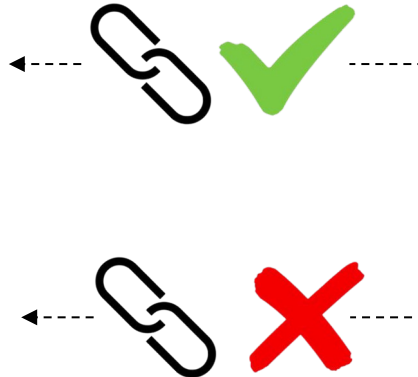
Futterumstellung
Hund



URL B

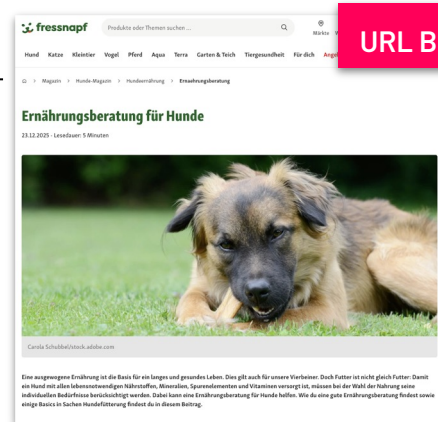
Ernährungsberatung
Hund

Übergewicht Hund



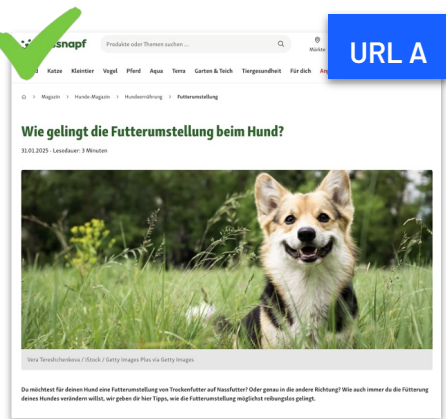
URL A

Futterumstellung
Hund



URL B

Ernährungsberatung
Hund



URL A



URL B

URL	Anzahl Unique Inlinks	Link Score	Anzahl Backlinks	Anzahl Referring Domains	Link-potenzial
Futterumstellung Hund	40	3	1	1	
Ernährungsberatung Hund	144	8	4	3	

Fehlende Potenziallinks auf einen Blick

Ziel-URL	Related URL	Cosinus Ähnlichkeit	Link von Related URL auf Ziel-URL vorhanden?	Content-Link von Related URL auf Ziel-URL vorhanden?	Link-potenzial
https://www.fressnapf.de/magazin/hund/ernaehrung/uebergewicht-hund/ 	https://www.fressnapf.de/magazin/hund/ernaehrung/hundenahrung-rassen/	0.858	ja	ja	0.2518
	https://www.fressnapf.de/magazin/hund/ernaehrung/hund-futterumstellung/	0.855	ja	ja	0.1807
	https://www.fressnapf.de/magazin/hund/ernaehrung/futterrechner/	0.849	nein	nein	0.0829
	https://www.fressnapf.de/magazin/hund/ernaehrung/hundefutter-selber-kochen/	0.849	ja	ja	0.4121
	https://www.fressnapf.de/magazin/hund/ernaehrung/ernaehrungsberatung/	0.844	nein	nein	0.5162

Identifiziert eure Linkpotenziale toolbasiert



ONE Link Intelligence

Entwickelt von [Daniel Kremer](#) von [ONE Beyond Search](#) | Folge mir auf [LinkedIn](#) für mehr SEO-Insights und Tool-Updates

>  Was du mit dem Tool machen kannst und wie du es nutzt

Welche Analysen möchtest du durchführen?

Mehrfachauswahl möglich

Interne Verlinku... x Unpassende inte... x



[Zum Tool](#)

(3) Query-URL- Matching



**Query Fan-out-Analyse /
Keyword- oder
Themenrecherche durchgeführt.**

**Ein Haufen neuer Themen
identifiziert.**

Und nun?

**Einfach für jede Query eine
eigene URL anlegen?**

Bitte nicht.



Google liefert die Antwort

CONTAINS HIGHLY CONFIDENTIAL INFORMATION – SUBJECT TO PROTECTIVE ORDER

Google Search Litigation

January 31, 2025 Call with Google Engineer Pandu Nayak

Participants

- Google: Pandu Nayak
- Expert Prof. James Allan
- Williams and Connolly: John McGowan and Colette Connor
- Cornerstone: Vivek Mani, Chanel O'Neill, and Blake Boswell

I. Overview/Intro

Preview of terminology and Google search structure referenced during the call

- **Document:** Google's name for a webpage. "Document" can refer to the webpage itself or the constructed webpage version Google keeps in its database.
- **Signal:** Google uses many signals to come up with the SERP. Top-level signals are aggregates of other signals, called raw signals. Google uses over 100 raw signals. Some signals are developed using machine learning models vs. others that are called, or thought of as, traditional signals. Examples of signals discussed on the call include:
 - o **Q* (pronounced "Q star"), Google's measure of quality of a document;**
 - o Navboost, a measure of how frequently users (subset by location and device type) click on a particular document for a particular query is a traditional signal. Uses most recent 13 months of data;
 - o RankEmbed, one of Google's primary LLM-trained signals;
 - o Twiddlers – re-rank a set of already selected results
 - o PageRank, one of Google's original signals, continues to be one of the factors that goes into the quality signals for pages.
- [REDACTED] combines signals into a single score, which then determines document rank in the list of blue links on the SERP. [REDACTED] was an early Google use of a machine learning model in its search algorithm. [REDACTED] was taught using a process of [REDACTED]

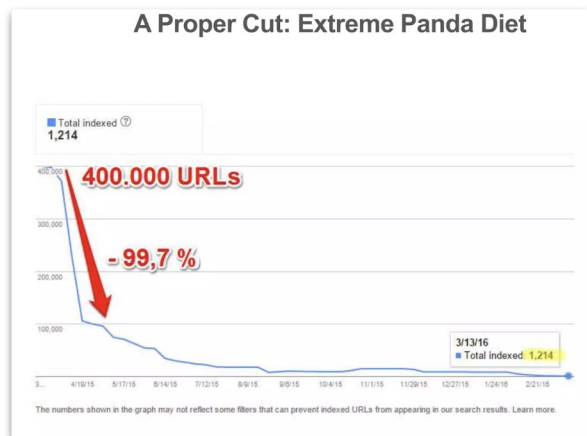
Q* (Q Star)

Viele schwache Seiten (mit niedrigem Q*-Wert) ziehen die starken URLs der Domain mit runter

Rückblick ins Jahr 2019



“ 80 % (actually 95 %) of your website usually is **dead weight**. And it's **pulling down your best pages**.”



**Wie finde ich jetzt aber
heraus, ob es für ein Thema
schon eine thematisch
(sehr) ähnliche URL gibt?**

Toolbasiert relevante URLs für Queries basierend auf semantischer Ähnlichkeit identifizieren

Input-Dateien

Tool

Alle **Queries**



Alle **URLs** inkl. **Main Content**



Option 1

**Die “durchschnittlich”
semantisch ähnlichste URL
finden**

Beispiel



Main Content der URL

Zu Hause „tanken“ leicht gemacht: Mit der Installation einer Wallbox bekommen Sie Ihre eigene E-Tankstelle. Erfahren Sie, wie Sie die Wallbox anmelden können, wie die Installation abläuft und was beim Anschluss zu Hause zu beachten ist. Kann ich nicht auf die Installation einer Wallbox verzichten und mein Fahrzeug an der Haushaltssteckdose laden?

Theoretisch können Sie Ihr E-Auto an einer Haushaltssteckdose (SCHUKO) in der Garage, im Garten oder mit einer Verlängerungsschnur im Haus laden. Nur dauert das nicht nur sehr lange, es besteht auch die große Gefahr von Kurzschlüssen oder Kabelbränden.

...

**“1 Chunk” wird an
Embedding-
Modell
übergeben**

Welcher Main Content passt am besten zur jeweiligen Query?

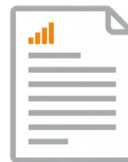
Query

warum schlafen katzen
auf menschen

URL A



URL B



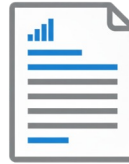
- ✓ Die **Query** und die **Main Contents** der URLs werden **embedded**
- ✓ Matching entspricht: **Welcher ("Gesamt")Main Content passt am besten zur Query?**

Welcher Main Content passt am besten zur jeweiligen Query?

Query

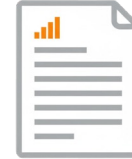
warum schlafen katzen
auf menschen

URL A



SIM 0,921

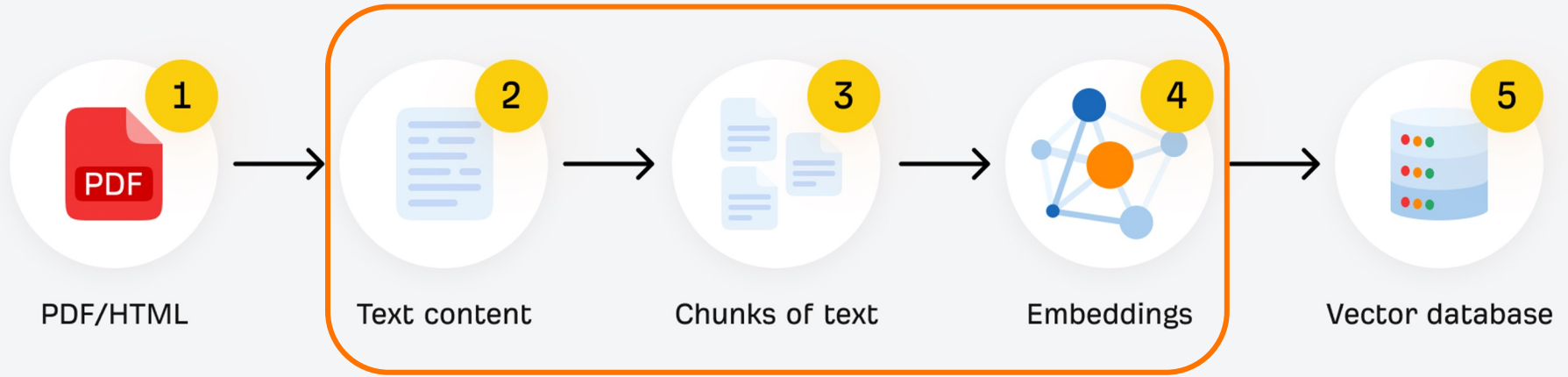
URL B



SIM 0,792

**In der KI-Ära konkurrieren
Absatz (Chunk) gegen Absatz
(Chunk), nicht mehr unbedingt
URL gegen URL**

Der Chunking-Prozess vereinfacht dargestellt



LLMs erstellen eine synthetisierte Antwort



ChatGPT

(...)



6

Abruf und Zusammenführung
relevanter Chunks zur
Beantwortung der User Query



Antwort auf Prompt

Option 2

Einzelne Chunks an Embedding-Modell übergeben

Content nach der **Fixed-size Methode** chunken



Chunk-Größe (Wörter) z. B. 40 **Chunk-Overlap** (Wörter) z. B. 15

Zu Hause „tanken“ leicht gemacht: Mit der Installation einer Wallbox bekommen Sie Ihre eigene E-Tankstelle. Erfahren Sie, wie Sie die Wallbox anmelden können, wie die Installation abläuft und was beim Anschluss zu Hause zu beachten ist. Kann ich nicht

Chunk 1

auf die Installation einer Wallbox verzichten und mein Fahrzeug an der Haushaltssteckdose laden?

Theoretisch können Sie Ihr E-Auto an einer Haushaltssteckdose (SCHUKO) in der Garage, im Garten oder mit einer Verlängerungsschnur im Haus laden. Nur dauert das nicht nur sehr lange, es besteht auch die große Gefahr von Kurzschlüssen oder Kabelbränden.

...

Content nach der **Fixed-size Methode** chunken



Chunk-Größe (Wörter) z. B. 40 **Chunk-Overlap** (Wörter) z. B. 15

Zu Hause „tanken“ leicht gemacht: Mit der Installation einer Wallbox bekommen Sie Ihre eigene E-Tankstelle. Erfahren Sie, wie Sie die Wallbox anmelden können, wie die

Installation abläuft und was beim Anschluss zu Hause zu beachten ist. Kann ich nicht

auf die Installation einer Wallbox verzichten und mein Fahrzeug an der Haushaltssteckdose laden?

Theoretisch können Sie Ihr E-Auto an einer Haushaltssteckdose (SCHUKO) in der Garage,

im Garten oder mit einer Verlängerungsschnur im Haus laden. Nur dauert das nicht nur

sehr lange, es besteht auch die große Gefahr von Kurzschlüssen oder Kabelbränden.

...

Chunk 1

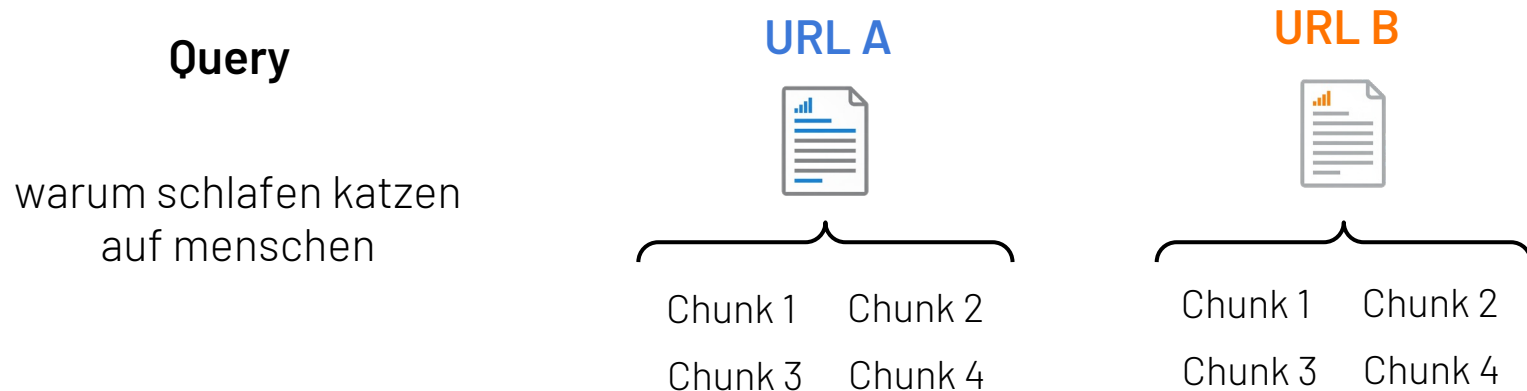
Overlap

Chunk 2

Overlap

Chunk 3

Welcher Absatz (Chunk) passt am besten zur jeweiligen Query – und zu welcher URL gehört dieser?



- ✓ Die **Query** und die **einzelnen Chunks** werden **embedded**
- ✓ Matching folgt einem **“Chunk-first-Ansatz”**

Welcher Absatz (Chunk) passt am besten zur jeweiligen Query – und zu welcher URL gehört dieser?

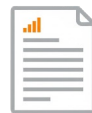
Query

warum schlafen katzen
auf menschen

URL A



URL B



Chunk 1 SIM 0,645

Chunk 1 SIM 0,821

Chunk 2 SIM 0,855

Chunk 2 SIM 0,632

Chunk 3 SIM 0,462

Chunk 3 SIM 0,785

Chunk 4 SIM 0,712

Chunk 4 SIM 0,576

**“wie hoch können
zwergkaninchen springen”**



Beispiel Output "Chunk-Option"

Query	URL	Relevant Chunk	Cosinus-Score Relevant Chunk	Rank URL Chunk Cosinus-Score
wie hoch können zwergkaninchen springen	https://www.fressnapf.de/magazin/kleintiere/kaninchen/zwergkaninchen	damit sich ein Zwergkaninchen auf die Hinterläufe stellen kann, ohne mit den Ohren oben anzustoßen – also mindestens 70 Zentimeter hoch. Im besten Fall erhalten deine Zwergkaninchen regelmäßig Zugang zu einem mobilen Auslauf im Freien oder zu einem dauerhaft befestigten Außengehege. Achte darauf, dass sich deine Zwergge nicht unter dem Zaun durchgraben oder nach oben herausspringen können. (...)	0,686	1

Das Tool ermittelt den **semantisch am besten zur Query passenden Content-Chunk** und die **URL auf der sich dieser befindet**

Die richtige Zwergkaninchen-Haltung: Käfig oder Außengehege?

Grundsätzlich kannst du Zwergkaninchen im Haus oder draußen halten. Eine Kombination der beiden Haltungsverfahren ist ebenfalls möglich. Unverzichtbar ist in jedem Fall eine sichere Unterkunft mit ausreichend Platz. Die Empfehlungen zur Zwergkaninchen-Haltung betragen zurzeit eine Mindestfläche von zwei bis drei Quadratmetern pro Kaninchen. Diese Fläche muss dauerhaft zur Verfügung stehen und hoch genug sein, damit sich ein Zwergkaninchen auf die Hinterläufe stellen kann, ohne mit den Ohren oben anzustoßen – also mindestens 70 Zentimeter hoch.

Im besten Fall erhalten deine Zwergkaninchen regelmäßig Zugang zu einem mobilen Auslauf im Freien oder zu einem dauerhaft befestigten Außengehege. Achte darauf, dass sich deine Zwerge nicht unter dem Zaun durchgraben oder nach oben herausspringen können.



Vom Tool ermittelter relevanter Fressnapf- Chunk

AI Overview

wie hoch können zwergkaninchen springen

KI-Modus Alle Bilder Videos Kurze Videos News Web Mehr ▾ Suchfilter ▾

Übersicht mit KI

Zwergkaninchen können aus dem Stand mühelos **bis zu 70 cm** hoch springen. Aufgrund ihres hohen Sprungvermögens sollten Gehegezäune für die Außenhaltung eine Höhe von mindestens 80 cm bis über 1 Meter aufweisen, um ein Ausbrechen zu verhindern, da sie 70-cm-Zäune häufig überwinden. ⓘ

- **Sprunghöhe:** Etwa 70 cm sind für Zwergkaninchen normal, wobei manche Tiere sogar noch höher springen können.
- **Zaunempfehlung:** Um Zwergkaninchen sicher im Gehege zu halten, sind Zäune mit einer Höhe von mindestens 80 cm (besser mehr) notwendig, besonders um Marder und Füchse fernzuhalten.
- **Wichtige Hinweise:** Da Kaninchen gerne ausbrechen und ihre Sprungkraft nutzen, ist eine Absicherung nach oben (Volierennetz) bei Außengehegen oft sinnvoll, berichten [\(Link: Deutscher Tierschutzbund\)](#) und Kaninchenwiese. ⓘ

Die ganzjährige Außenhaltung von Kaninchen

Achtung: Ein Zwergkaninchen kann mühelos aus dem Stand 70 cm hoch springen.

Deutscher Tierschutzbund ⓘ

Die Haltung von Zwergkaninchen - Deutscher Tierschutz und

Besonders wichtig: Ein Balkon muss ausbruchsicher sein – Zwergkaninchen können...

Deutscher Tierschutzbund ⓘ

AUSSENHALTUNG AUF DEM BALKON

Gerade in Großstädten stehen oft keine Gärten für eine Außenhaltung zur Verfügung. Stattdessen möchten Halter ihre Tiere dann zumindest auf dem Balkon unterbringen. Das ist möglich, es müssen dabei aber alle oben genannten Punkte beachtet werden (**Schutzhütte, Strukturierung, Platzangebot**). Ein im Handel erhältlicher Kaninchen-Gehege reicht keinesfalls für eine tiergerechte Außenhaltung auf dem Balkon aus! Zudem muss der Balkon ausreichend gesichert werden. Die Tiere dürfen nicht nach unten fallen können und für andere Tiere darf es keine Möglichkeit geben, auf den Balkon zu gelangen. **Achtung: Ein Zwergkaninchen kann mühelos aus dem Stand 70 cm hoch springen.**

Die richtige Zwergkaninchen-Haltung: Käfig oder Außengehege?

Grundsätzlich kannst du Zwergkaninchen im Haus oder draußen halten. Eine Kombination der beiden Haltungsvarianten ist ebenfalls möglich. Unverzichtbar ist in jedem Fall eine sichere Unterkunft mit ausreichend Platz. Die Empfehlungen zur Zwergkaninchen-Haltung betragen zurzeit eine Mindestfläche von zwei bis drei Quadratmetern pro Kaninchen. Diese Fläche muss dauerhaft zur Verfügung stehen und hoch genug sein, damit sich ein Zwergkaninchen auf die Hinterläufe stellen kann, ohne mit den Ohren oben anzustoßen – also mindestens 70 Zentimeter hoch.

Im besten Fall erhalten deine Zwergkaninchen regelmäßig Zugang zu einem mobilen Auslauf im Freien oder zu einem dauerhaft befestigten Außengehege. Achte darauf, dass sich deine Zwerge nicht unter dem Zaun durchgraben oder nach oben herausspringen können.

Vom Tool ermittelter
relevanter Fressnapf-Chunk

Thematische Kongruenz

Chunk AIO Top Quelle

AUSSENHALTUNG AUF DEM BALKON

Gerade in Großstädten stehen oft keine Gärten für eine Außenhaltung zur Verfügung. Stattdessen möchten Halter ihre Tiere dann zumindest auf dem Balkon unterbringen. Das ist möglich, es müssen dabei aber alle oben genannten Punkte beachtet werden (**Schutzhütte, Strukturierung, Platzangebot**). Ein im Handel erhältlicher Kaninchenkäfig reicht keinesfalls für eine tiergerechte Außenhaltung auf dem Balkon aus! Zudem muss der Balkon ausreichend gesichert werden. Die Tiere dürfen nicht nach unten fallen können und für andere Tiere darf es keine Möglichkeit geben, auf den Balkon zu gelangen. **Achtung: Ein Zwergkaninchen kann mühelos aus dem Stand 70 cm hoch springen.**

Auch dieses Tool gibt es kostenlos für euch



[Zum Tool](#)

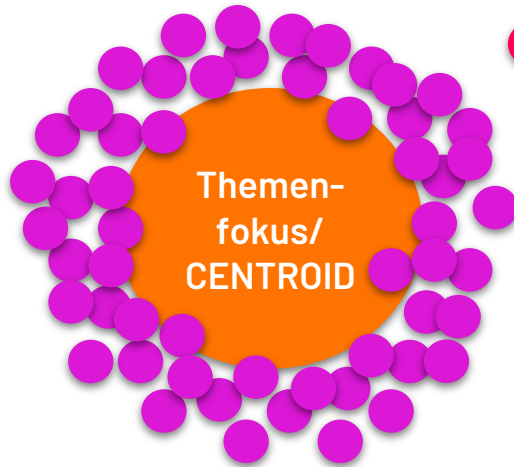
(4) Offtopic URL Detection

Warum ist das wichtig?

In the algo leak, we see that **siteFocusScore**, **siteRadius**, **siteEmbeddings** and **pageEmbeddings** are used for ranking.

What are they?

- **siteFocusScore** denotes how much a site is focused on a specific topic.
- **siteRadius** measures how far page embeddings deviate from the site embedding. In plain speech, Google creates a topical identity for your website, and every page is measured against that identity.
- **siteEmbeddings** are compressed site/page embeddings.



Inhaltliche / Thematische Identität der Domain = Embedding-Mittelwert aller gecrawlten Seiten

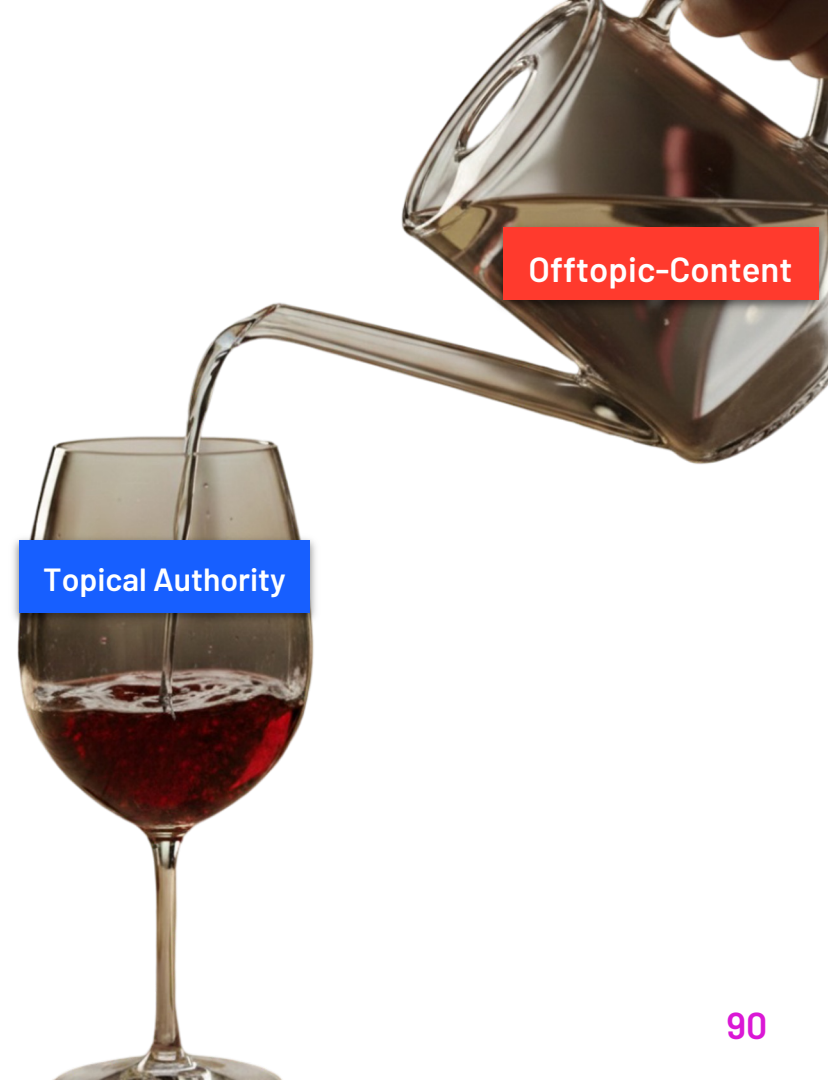
Core Content
Seiten mit geringer semantischer Distanz zum Centroid

Low Relevance / Offtopic Content
Seiten mit hoher semantischer Distanz zum Centroid

Low Relevance / Offtopic Content ist das Wasser in eurem Wein

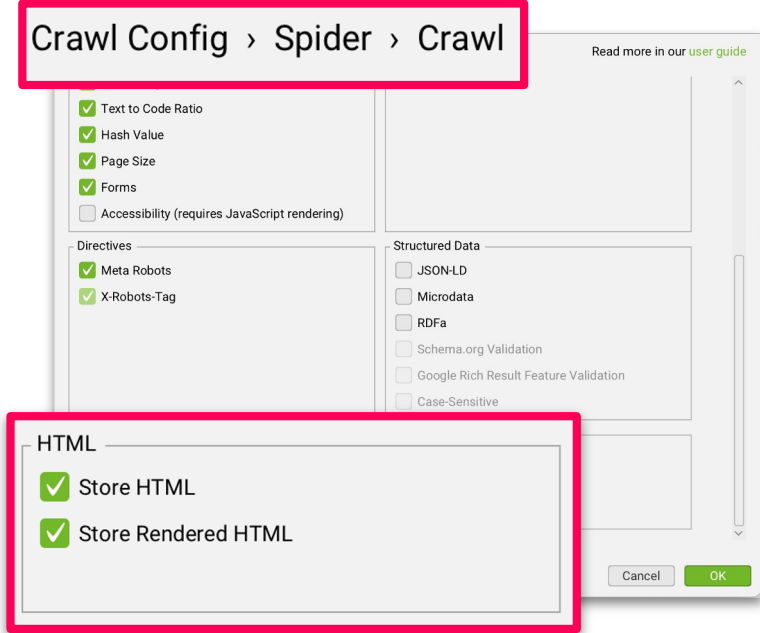
**er verwässert die thematische
Identität und Autorität**

und "verdünnt" oftmals die Wirkung eurer
starken Seiten



Der Workflow um Offtopic-URLs zu finden

HTML speichern

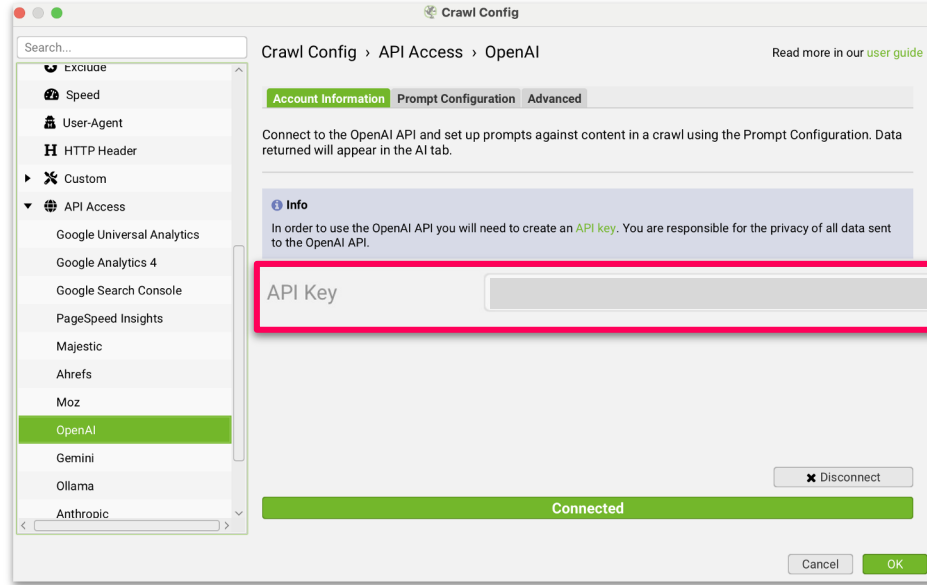


Configuration > Crawl Config

„Store HTML“ anhaken

Wenn ihr mit
JavaScript crawlt,
„Store Rendered“
HTML auswählen

API-Key besorgen und hinterlegen



Configuration > API Access

API-Key eintragen

Falls ihr noch keinen API-Key habt:

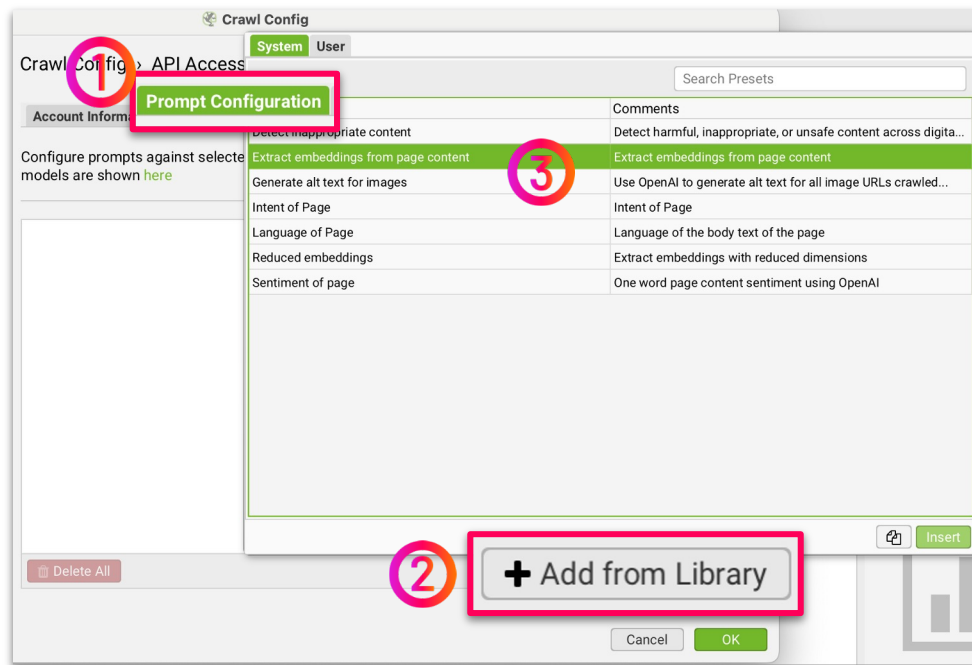


OpenAI



Gemini

Vektor-Embeddings mit dem Screaming Frog erstellen



Configuration > API Access

- 1 Tab **Prompt Configuration** auswählen
- 2 + **Add from Library**
- 3 **Extract embeddings from page content** auswählen & Insert

Der Workflow um Offtopic-URLs zu finden

Crawl Config > Content > Embeddings [Read more in our user guide](#)

Configure site analysis using Vector Embeddings generated by AI providers. The Embeddings can be used for Semantic Search functionality and Content Cluster visualisation. Semantic Similarity and Low Relevance content analysis results are displayed in the Content tab.

General Settings

- ☒ Enable Embedding functionality
- Embedding Prompt

- ☒ Only Check Indexable Pages for Semantically Similar
- ☒ Ignore Paginated Pages for Semantically Similar

Embedding Filter Rules

No Embedding Filter Rules configured

[Delete All](#) [Move Up](#) [Move Down](#) [Add](#)

Low Relevance

- ☒ Enable Low Relevance
- Low Relevance Threshold
- ☐ Only Check Indexable Pages for Low Relevance
- ☐ Ignore Paginated Pages for Low Relevance

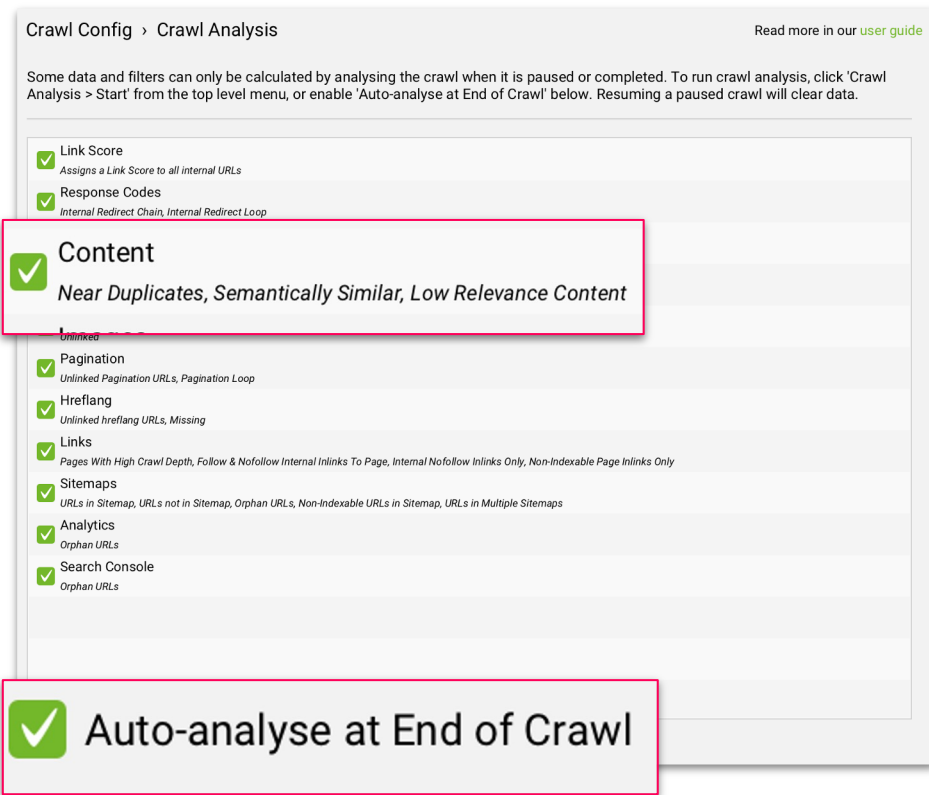
[Cancel](#) [OK](#)

Crawl Config > Content > Embeddings

Enable Embedding functionality
anhaken

Enable Low Relevance anhaken und
Low Relevance Threshold vergeben.
0,6 bedeutet hier z. B. jede URL mit
einer Cosinus Ähnlichkeit von $< 0,6$
zum Centroid gilt als Offtopic.

Der Workflow um Offtopic-URLs zu finden



Crawl Config > Crawl Analysis

Auto-analyse at End of Crawl
anhaken

Alternativ kann die Analyse nach Beendigung des Crawls manuell über Screaming Frog Menüleiste "**Crawl Analysis > Start**" ausgelöst werden

Crawl starten

The screenshot shows the Screaming Frog SEO Spider 23.0 interface. The top bar indicates the current URL is <https://www.fressnapf.de/>. The main table lists crawled URLs, their content types, status codes, and indexability. A summary table on the right shows the distribution of content types, with HTML being the most prevalent at 97.71%. A donut chart at the bottom right visualizes the internal link distribution, showing that HTML links account for the vast majority of the total links.

Row	Address	Content Type	Status Code	Status	Indexability
1	https://www.fressnapf.de/	text/html; charset=utf-8	200		Indexable
2	https://www.fressnapf.de/c/marken/premiere/	text/html; charset=utf-8	200		Indexable
3	https://www.fressnapf.de/c/marken/more/	text/html; charset=utf-8	200		Indexable
4	https://www.fressnapf.de/c/marken/multifit/	text/html; charset=utf-8	200		Indexable
5	https://www.fressnapf.de/aktionen-angebote/neuheiten/	text/html; charset=utf-8	200		Indexable
6	https://www.fressnapf.de/c/katze/katzenschlafplaetze/	text/html; charset=utf-8	200		Indexable
7	https://www.fressnapf.de/ueber-uns/anfahrt/	text/html; charset=utf-8	200		Indexable
8	https://www.fressnapf.de/c/marken/multifit/?q=%3A%3Aclsf-en...	text/html; charset=utf-8	200		Non-Indexable
9	https://www.fressnapf.de/c/terra/	text/html; charset=utf-8	200		Indexable
10	https://www.fressnapf.de/c/asia/fischfutter/?n=%3A%3Ahand	text/html; charset=utf-8	200		Indexable

Content Type	Count	Percentage
HTML	30,769	97,71%
JavaScript	91	0,29%
CSS	1	0%
Images	0	0%
Media	0	0%
Fonts	12	0,04%
XML	1	0%
PDF	0	0%
Plugins	0	0%
Other	553	1,76%
Unknown	62	0,2%

Internal Link Distribution:

- HTML: 97.71%
- JavaScript: 0.29%
- CSS: 0%
- Images: 0%
- Media: 0%
- Fonts: 0.04%
- XML: 0%
- PDF: 0%
- Plugins: 0%
- Other: 1.76%
- Unknown: 0.2%

Spider Mode: Active | Average: 0,15 URL/s. Current: 0,70 URL/s. | Completed 31.489 of 42.422 (74,22%) 10.933 Remaining

Ergebnisse

Ergebnisse

ernal External Security Response Codes URL Page Titles H1 Canonicals Links Custom Search Custom JavaScript Meta Description H2 Content			Overview Issues Site Structure Segments Response Times API Spelling & Grammar		
Low Relevance Content			URLs % of Total		
Row	Address	Semantic Relevance Score ▲			
1	https://www.edeka.de/ernaehrung/tests-rechner/index.jsp	0,557187	Over 70 Characters ⓘ		
2	https://www.edeka.de/ernaehrung/beauty/windelfrei.jsp	0,576813	Multiple ⓘ		
3	https://www.edeka.de/ernaehrung/beauty/babyakne.jsp	0,578335	Non-Sequential ⓘ		
4	https://www.edeka.de/ernaehrung/familienzeit.jsp	0,587263	Content		
5	https://www.edeka.de/ernaehrung/beauty/rizinusoel.jsp	0,588564	All		
6	https://www.edeka.de/ernaehrung/genussthem/baguette-rezepte.jsp	0,592430	Exact Duplicates ⓘ		
7	https://www.edeka.de/ernaehrung/bewusste-ernaehrung/vanadium.jsp	0,593005	Near Duplicates ⓘ		
			Low Relevance Content		
			7		
			1,23%		

Es liegen insgesamt **7 potentielle Offtopic-URLs** im Ernährungsverzeichnis vor

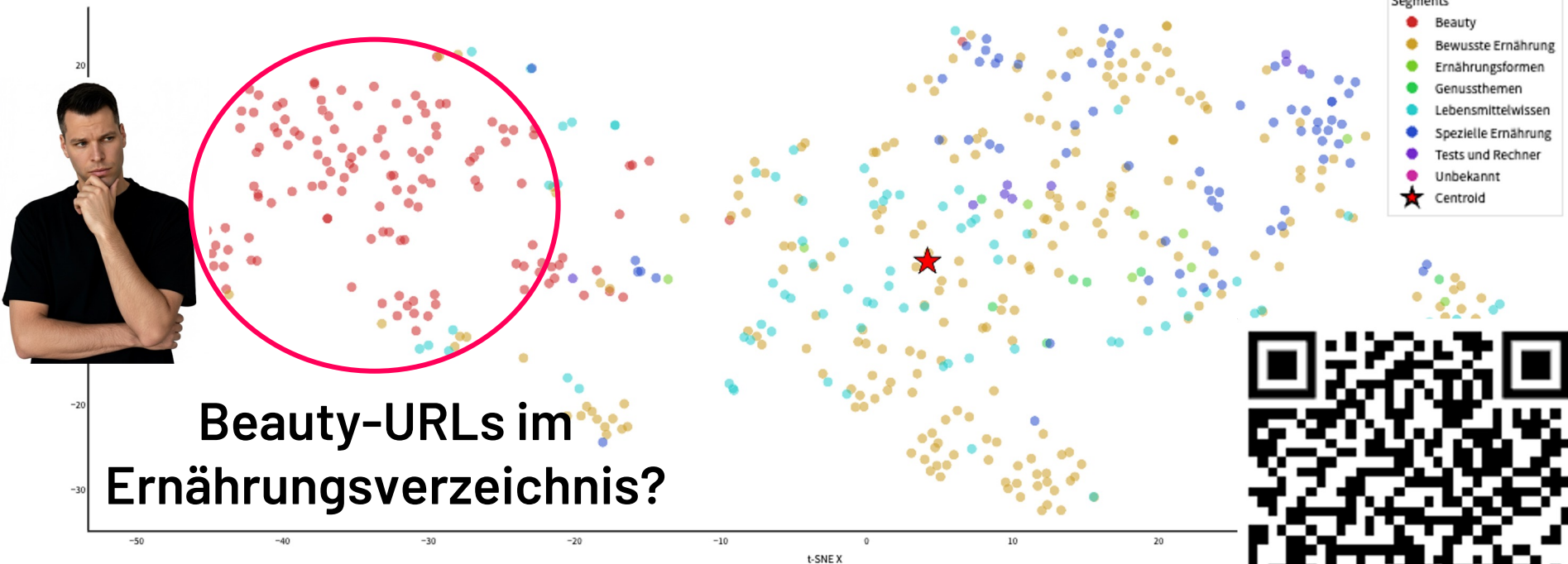


Endlich windelfrei: Tipps rund um das Töpfchentraining

Bye-bye, Wickeltisch! Viele Eltern wünschen sich die letzte Windel sehnsüchtig herbei. Doch wann wird es Zeit, mit dem Töpfchentraining zu beginnen, und wie hilfst du deinem Liebling beim Trockenwerden? Wir klären auf!

Semantische Content-Struktur visualisieren

t-SNE der Seiten-Embeddings



[Zum Tool](#)

**Was tun mit
Offtopic-Content?**



Wasser in Wein verwandeln

z. B. Offtopic Content überarbeiten



2

Wasser entfernen

z. B. Offtopic URLs löschen /
weiterleiten / de-indexieren



3

Wein-Wasser- Mischung tolerieren

z. B. Impressum, Über Uns Seite,
Hubseiten



Empfehlung: Schuster bleib bei deinen Leisten*

*so gut es geht



Fazit



**Wer 2026 ohne Vektor-Analysen arbeitet,
optimiert an der (SEO-)Realität vorbei.**

**Vektor-Embeddings bringen uns näher
heran, wie Suchmaschinen und LLMs
Inhalte wirklich verstehen.**

Give it a try.

***Vielen Dank fürs
Zuhören!***



Daniel Kremer

onebeyondsearch.com

