

SEO-TIPPS AUS DER PRAXIS

SICHTBARKEITSBOOST: REDIRECT-MAPPING AM BEISPIEL VON THALIA UND WELTBILD



Redirect-Mapping ist eine essenzielle SEO-Aufgabe, sei es bei der Umstellung der eigenen Domain auf ein neues System oder bei der Übernahme bestehender Websites. Es bezeichnet den Prozess, bei dem alte URLs einer Website auf neue URLs weitergeleitet werden. Das Ziel ist, Besucher und Suchmaschinen auf die richtige Zielseite zu leiten, ohne dass wertvolle Rankings oder Nutzer verloren gehen. Bei fremden Domains, die einer anderen Struktur folgen, kann dies herausfordernd sein. Noch dazu, wenn es um große Mengen an Seiten geht, die sich redaktionell kaum überblicken lassen. Bei der Weiterleitung von Weltbild auf Thalia wurden mehrere Millionen URLs betrachtet und die wichtigsten Seiten zielgerichtet weitergeleitet.

Die relevanten Seiten richtig priorisieren

Für Deutschland bildeten die 73 vorhandenen Sitemaps von weltbild.de die Ausgangsbasis. 70 davon entstammten dem Produktbereich und konnten wiederum jeweils bis zu 100.000 URLs enthalten. Aufgrund der großen zu erwartenden Datenmenge wurde zur Analyse überwiegend die Python IDE Jupyter Labs genutzt, da Arbeitsblätter in Excel und CSV-Dateien in Numbers auf etwa eine Million Zeilen beschränkt sind. Damit ließen sich die Produkt-Sitemaps von weltbild.de herunterladen, entpacken und automatisiert aneinanderfügen. In Summe ergab das ca. 1,5 Millionen Produktseiten sowie mehrere Tausend Kategorie- und Ratgeberseiten. Man sollte bei der Auflistung der URLs auch daran denken, dass es bei der Ursprungsdomain noch

„Broken Backlinks“ geben kann. Das sind 404-Seiten mit starken Backlinks, die es wert sind, weitergeleitet zu werden. Ein Offpage-Check mit Ahrefs oder den Link Research Tools (LRT) ist hier sinnvoll.

Im nächsten Schritt wurden Duplikate entfernt, stichprobenhaft Statuscodes gecrawlt und Daten aus den gängigen SEO-Tools zu Sichtbarkeit und Backlinks ergänzt. Wenn der Zugriff auf die Google Search Console vorhanden ist, können auch große Mengen Klicks und Impressionen über die zugehörige API abgefragt werden (zum Beispiel mit Google Apps Script).

Bei großen Online-Shops gibt es häufig duplizierte Seiten und nicht alle Inhalte müssen weitergeleitet werden. Es ist daher sinnvoll, die URLs in Gruppen einzuteilen, abhängig davon, wie viel Prozent des gesamten Traffics bzw. der

Sichtbarkeit sie erzielen. So kann man abhängig von der Zeit und den eigenen Ressourcen bei dem Redirect-Mapping schrittweise vorgehen und richtig priorisieren.

Produktseiten mit hoher Genauigkeit abgleichen

Produktseiten lassen sich in den meisten Fällen über den Abgleich der ISBNs (Bücher) oder EANs (oftmals Medienprodukte wie Filme, Musik oder Videospiele) zuordnen. Dazu wurden die ausgewählten Produktseiten von weltbild.de gecrawlt und deren Produktnummern extrahiert. Es wäre zu aufwendig und unnötig gewesen, noch zusätzlich große Mengen an Produktseiten von thalia.de zu crawlen. Die Seiten lassen sich automatisch anhand der Produktnummer über die interne Suche abfragen.

Man verwendet den URL-Parameter aus der internen Suche seiner Zieldomain, fügt die Produktnummer ein und generiert eine Liste von URLs (zum Beispiel domain.com/suche?sq=Produktnummer). Diese Liste kann man mit einem Webcrawler durchlaufen (zum Beispiel Screaming Frog). Dadurch wird der Link zu der Trefferseite aus der internen Suche extrahiert. Zum Abgleich kann man auch noch die passende Meldung für „keine Ergebnisse“ scrapen, damit man bei Produkten ohne Treffer sicher ist, dass der Crawl funktioniert hat. So konnten wir bereits einen Großteil der Produkte mit einer Genauigkeit von annähernd 100 % matchen.

Die Domain weltbild.at wurde abgeschaltet

Etwas herausfordernder war, dass die Domain weltbild.at zu diesem Zeitpunkt bereits abgeschaltet war. Die URLs, unter denen die österreichischen Sitemaps aufrufbar waren, schienen in etwa denen des deutschen Pendants zu entsprechen. Daher wurde die Wayback Machine (web.archive.org) verwendet und nach kurzer Zeit konnten alle Sitemaps identifiziert werden. Danach wurden auch diese Sitemaps mit Python automatisiert heruntergeladen und ausgewertet. Unter der Annahme, dass viele Seiten zwischen Deutschland und Österreich gespiegelt wurden, erfolgte der Abgleich anhand der URLs ohne den länderspezifischen Domainnamen.

Kategorien und Ratgeberseiten matchen mit KI

Es kamen auch KI-gestützte Methoden für das Redirect-Mapping zum Einsatz. Dafür wurden



Abb. 1: Erste Auswertung der URL-Anzahl für die Sitemaps von Weltbild DE & AT

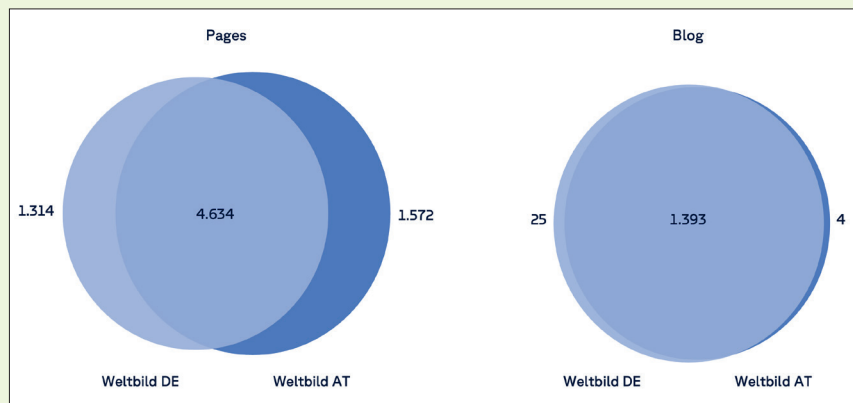


Abb. 2: Die Venn-Diagramme zeigen Beispiele für die Schnittmenge der relativen URL-Pfade zwischen Weltbild DE & AT.

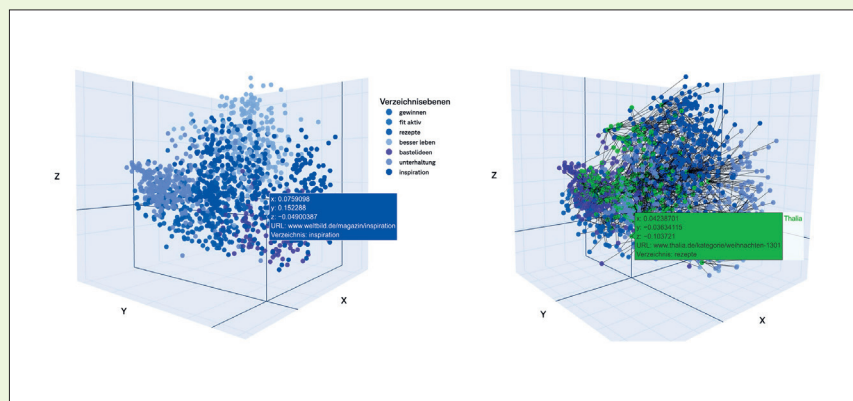


Abb. 3: Links: vereinfachte 3-D-Darstellung der OpenAI-Embeddings generiert aus den Weltbild-Ratgeberseiten; rechts: passende Thalia-Zielseite in Grün mit Verbindungslinie zu Weltbild

zunächst alle Kategorie- und Ratgeberseiten der Website gecrawlt, um ihre Inhalte wie Texte, Überschriften und andere Informationen zu sammeln.

Vektor Embeddings spielen hier eine zentrale Rolle: Moderne

Machine-Learning-Modelle, die an riesigen Datensätzen trainiert wurden, erkennen semantische Zusammenhänge zwischen Wörtern und Texten. Embeddings bestehen aus hochdimensionalen Vektoren, wobei ähnliche Inhalte durch ähnliche

SEO-TIPPS AUS DER PRAXIS

Index	Weltbild_URL	Thalia_URL
0	www.weltbild.de/magazin/bastelideen	www.thalia.de/kategorie/kreativ-basteln-4780
1	www.weltbild.de/magazin/bastelideen/anleitungen/adventskalender-mit-salzteig-anhaengern/B1xMFMISv	www.thalia.de/kategorie/fuer-kinder-31405
2	www.weltbild.de/magazin/bastelideen/anleitungen/adventskalender-ideen-fuer-kinder/Byv7CG5Pr	www.thalia.de/kategorie/fuer-kinder-31405
3	www.weltbild.de/magazin/bastelideen/anleitungen/advent-advent-gesteck-selber-basteln/SyfG2YmJj	www.thalia.de/kategorie/adventskraenze-27244
4	www.weltbild.de/magazin/bastelideen/anleitungen	www.thalia.de/kategorie/kreativ-basteln-4780
5	www.weltbild.de/magazin/bastelideen/anleitungen/aquarell-malen-ideen-fuer-anfaenger/rkcggTHla	www.thalia.de/kategorie/aquarellmalerei-32409
6	www.weltbild.de/magazin/bastelideen/anleitungen/adventskalender-selber-basteln/BygnbYIQ	www.thalia.de/kategorie/fuer-kinder-31405
7	www.weltbild.de/magazin/bastelideen/anleitungen/basteln-mit-kindern/BkEpyewyd	www.thalia.de/kategorie/kinderzimmer-29385
8	www.weltbild.de/magazin/bastelideen/anleitungen/brandmalerei-heisser-diy-bloggertipp-von-dekotopia/rkDZheTJ4	www.thalia.de/kategorie/deko-17926
9	www.weltbild.de/magazin/bastelideen/anleitungen/diy-adventskalender-fuer-kinder/rJGmn4UK	www.thalia.de/kategorie/fuer-kinder-31405

Abb. 4: Die ersten zehn Seiten aus dem Abgleich der Ratgeberseiten

Index	Weltbild_URL	Thalia_URL
390	www.weltbild.de/magazin/gewinnen/der-heimweg-von-sebastian-fitzek-gewinnen/r1vauC0uv	www.thalia.de/kategorie/fitzek-sebastian-16326
407	www.weltbild.de/magazin/gewinnen/gewinnen-sie-den-neuen-fitzek/Bkca3eNwd	www.thalia.de/kategorie/fitzek-sebastian-16326
1249	www.weltbild.de/magazin/unterhaltung/autoren/weltbildliest-stargast-sebastian-fitzek/HkKrekYU	www.thalia.de/kategorie/fitzek-sebastian-16326
1262	www.weltbild.de/magazin/unterhaltung/autoren/versetzt-uns-ger-n-in-angst-und-schrecken-sebastian-fitzek/ryINFJAH	www.thalia.de/kategorie/fitzek-sebastian-16326
1282	www.weltbild.de/magazin/unterhaltung/autoren/sebastian-fitzek-im-interview/ryLKx/WGX	www.thalia.de/kategorie/fitzek-sebastian-16326

Abb. 5: Das Modell ordnet die Inhalte zum Krimiautor Sebastian Fitzek korrekt zu.

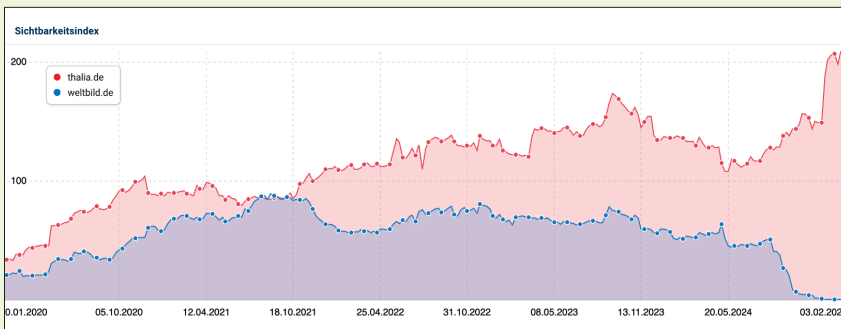


Abb. 5: Das Modell ordnet die Inhalte zum Krimiautor Sebastian Fitzek korrekt zu.

wachstum von Thalia zur Weihnachtszeit im Dezember 2024 deutschlandweit auf Platz 8. Auch die Domain thalia.at lag in diesem Zeitraum auf Platz 7 in Österreich (Quelle: Sistrix, Movers & Shakers DE & AT, Dez. 2024).

Fazit

Um große Mengen an Seiten zu untersuchen, sollte man sich mit einem geeigneten Tool vertraut machen. Es muss eine Vielzahl von Daten innerhalb kurzer Zeit bearbeiten können. Eine vollständige Datengrundlage ist ein wichtiger Faktor für dieses Vorgehen. Auch die Art und Weise, wie die Seiten abgeglichen werden, sollte im Vorhinein feststehen. Bei Produktseiten bietet sich der Abgleich über die interne Suche anhand der Produktnummer an. Insbesondere bei Ratgeberseiten mit hoher Informationsdichte haben wir mit KI-gestützten Tools gute Erfahrungen gemacht.

Zahlenkombinationen dargestellt werden. Diese wurden anhand der API von OpenAI erstellt.

Um die inhaltliche Nähe zwischen den Seiten zu messen, wurde das Google-Modell ScaNN (Scalable Nearest Neighbors) verwendet. Dieses Modell vergleicht die Embeddings und berechnet die Distanz zwischen ihnen. Eine geringe Distanz zeigt an, dass die Seiten inhaltlich nah beieinander liegen.

Die Genauigkeit hing vor allem von der Informationsdichte auf den jeweiligen Seiten ab. Bei den Ratgeberseiten konnte eine Genauigkeit von über 90 % in den Stichproben erzielt werden, während die Quote bei Kategorienseiten geringer war (ca. 80 %).

Ergebnisse bei der Sichtbarkeit

Nach der Übernahme von weltbild.de sowie der Einpflege des Redirect-Mappings verzeichnete die Domain thalia.de einen konstanten Anstieg bei der Sichtbarkeit in Sistrix. Etwa einen Monat nach der Umsetzung der Redirect-Maßnahmen stieg die Sichtbarkeit noch einmal stark an (+56.96 Punkte/+38%). Der Anstieg im Dezember entspricht in etwa dem Verlust von weltbild.de (50.994 Punkte) seit der Abschaltung im August.

Auch das Core-Update und die gesteigerte Crawlfrequenz des Google-Bots begünstigten dies möglicherweise zusätzlich. Damit lag das absolute Sichtbarkeits-