

Internetsuche und Suchmaschinen: Schlüsselwort-Suche - Semantische Suche - Synonyme und Wortvariationen, Kontextverständnis – Mehrsprachigkeit,

Mit OpenAI GPT-4.o erstellt, veröffentlicht: 29. August 2024 – fortgesetzt 13. Januar 2025



Guten Tag GPT 4, wir wollen uns heute unterhalten über die Internetsuche, somit über Suchmaschinen. Was kommt dir zu Suchmaschinen spontan in deinen virtuellen Sinn? Wie funktionieren sie, die Suchmaschinen?

Halte dich bitte in deinen Ausführungen kurz, gleichwohl unter Betonung dessen, was als Suchalgorithmus, Ranking und SEO zu bezeichnen wäre.



Guten Tag! Gerne erkläre ich kurz die Funktionsweise von Suchmaschinen, deren Algorithmen und die wichtigsten Konzepte:

Funktionsweise von Suchmaschinen:

1. **Crawler:** Suchmaschinen nutzen Web-Crawler (Bots), um das Internet systematisch zu durchsuchen und Webseiteninhalte zu indexieren.
2. **Indexierung:** Die gesammelten Daten werden in einem Index gespeichert, der wie eine riesige Bibliothek funktioniert und es ermöglicht, schnell relevante Seiten zu finden.
3. **Suchalgorithmus:** Wenn ein Nutzer eine Anfrage stellt, durchsucht der Algorithmus den Index und bewertet die Seiten nach Relevanz und Qualität.

Wichtige Konzepte:

1. **Suchalgorithmus:** Der Algorithmus bestimmt, welche Ergebnisse angezeigt werden und in welcher Reihenfolge. Er berücksichtigt zahlreiche Faktoren wie Stichwörter, Seitentitel, Meta-Beschreibungen und Nutzerverhalten.
2. **Ranking:** Dies bezieht sich auf die Platzierung der Webseiten in den Suchergebnissen. Das Ziel ist es, die relevantesten und qualitativ

hochwertigsten Inhalte an oberster Stelle zu zeigen.

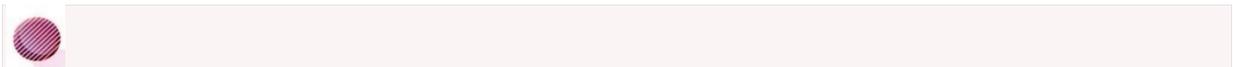
3. SEO (Search Engine Optimization): Dies sind Techniken, um die Sichtbarkeit einer Website in den Suchergebnissen zu verbessern.

Dazu gehören On-Page-Optimierungen (z.B. Keyword-Optimierung, gute Content-Struktur) und Off-Page-Optimierungen (z.B. Backlinks von anderen Webseiten).

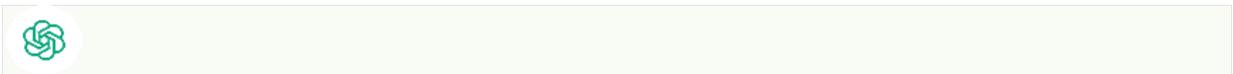
Kurze Zusammenfassung:

Suchmaschinen arbeiten durch Crawling und Indexierung des Internets und verwenden komplexe Algorithmen, um relevante Ergebnisse basierend auf einer Vielzahl von Faktoren zu liefern.

Ranking ist entscheidend für die Sichtbarkeit von Websites, und SEO-Strategien werden angewendet, um diese Sichtbarkeit zu verbessern.



Alles, was du hier beschreibst, bezieht sich auf das, was als Stichwortsuche bzw. Keyword-Suche bezeichnet wird, nicht wahr? Weil, wäre dem nicht so, bedürfte es keiner Off-Page-"Optimierung", wie du das nennst!



Ja, genau! Das, was ich beschrieben habe, bezieht sich primär auf die klassische Stichwortsuche bzw. Schlüsselwort-Suche, die den Kern der meisten Suchmaschinen bildet.

Warum Stichwortsuche/Schlüsselwort-Suche?

Bei der Stichwortsuche geht es darum, relevante Ergebnisse basierend auf den eingegebenen Keywords des Nutzers zu finden.

Der Algorithmus analysiert die Übereinstimmung dieser Keywords mit den Inhalten der indexierten Seiten und bewertet die Relevanz.

Rolle der Off-Page-Optimierung:

Die Notwendigkeit von Off-Page-Optimierung, wie z.B. der Aufbau von Backlinks, ergibt sich aus der Natur der Suchalgorithmen, die nicht nur auf Keywords basieren, sondern auch auf anderen Faktoren wie Autorität und Vertrauenswürdigkeit einer Seite.

Off-Page-Signale wie Backlinks werden genutzt, um die Reputation und Relevanz einer Seite zu bewerten. Das Ziel ist, die besten und vertrauenswürdigsten Ergebnisse anzuzeigen.

Wäre die Suche rein auf Stichwörter fokussiert und würde keine weiteren Faktoren berücksichtigen, wäre Off-Page-Optimierung weniger relevant.

Aber da Suchmaschinen versuchen, qualitativ hochwertigere und umfassendere Ergebnisse zu liefern, spielen auch externe Signale eine große Rolle im Rankingprozess.



Du redest von der Natur der Suchalgorithmen, die nicht nur auf Keywords basieren.

Und Du redest von Faktoren wie Autorität und Vertrauenswürdigkeit. Lass uns nun für einen Moment von diesen ominösen Backlinks und Off-Page-Signalen und dem ganzen Getue absehen.

Lass uns anstatt dessen dem Bereich der Semantischen Suche zuwenden.

Dabei geht es darum, die Bedeutung hinter den Suchanfragen zu verstehen, statt nur nach exakten Übereinstimmungen zu suchen (wie bei der Stichwort-Suche).

Gesamtkontext und einzelne Synonyme spielen dabei eine erheblichere Rolle, wie Du ausführst.

Insgesamt spielen bei der maschinellen Verarbeitung natürlicher Sprache (NLP), und also bei der weiteren Entwicklung etwa eines transformatorbasierten Sprachmodells, wie Du eines bist, die Elemente Kontext und Synonyme eine maßgebliche Rolle, nicht wahr?



Absolut richtig!

Die semantische Suche geht weit über die einfache Keyword-Suche hinaus und versucht, die Bedeutung hinter den Suchanfragen zu verstehen, anstatt nur nach exakten Wortübereinstimmungen zu suchen.

Semantische Suche und NLP:

1. Verstehen der Intention: Die semantische Suche zielt darauf ab, die Intention des Nutzers hinter einer Suchanfrage zu erfassen. Dabei wird analysiert, was der Nutzer wirklich wissen möchte, anstatt sich nur auf die eingegebenen Wörter zu konzentrieren.
2. Kontext und Synonyme: Im Gegensatz zur Keyword-Suche berücksichtigt die semantische Suche den Kontext der Anfrage, Synonyme, verwandte Begriffe und sogar die Art und Weise, wie Menschen natürlich sprechen.

Das bedeutet, dass Suchmaschinen in der Lage sind, Ähnlichkeiten und Beziehungen zwischen Begriffen oder Phrasen zu erkennen und relevante Ergebnisse anzuzeigen, auch wenn die genauen Wörter nicht übereinstimmen.

3. NLP und Sprachmodelle: Bei der Verarbeitung natürlicher Sprache (NLP) spielen transformatorbasierte Modelle wie GPT-4 eine entscheidende Rolle. Diese Modelle nutzen tiefe neuronale Netzwerke, um Sprache auf einer viel tieferen Ebene zu verstehen. Sie analysieren den Kontext, erkennen Muster und Zusammenhänge und können komplexe sprachliche Strukturen verarbeiten.
4. Anwendungen in der Suche: Mit Hilfe von NLP und semantischer Suche können Suchmaschinen präzisere und relevantere Ergebnisse liefern. Sie verstehen z.B. die Mehrdeutigkeit von Wörtern (Polysemie) und nutzen den Kontext, um die richtige Bedeutung herauszufiltern.

Fazit:

Die semantische Suche und die Fortschritte in natürlicher Sprachverarbeitung NLP und Künstlicher Intelligenz ermöglichen es, Suchanfragen besser zu verstehen und Ergebnisse zu liefern, die dem tatsächlichen Informationsbedarf des Nutzers entsprechen.

Das führt zu einer intelligenteren und nutzerfreundlicheren Suchergebnissen, die nicht nur auf Keywords basieren, sondern die gesamte Bandbreite menschlicher Sprache berücksichtigen.



Lass uns noch einmal zurückkehren zu den Backlinks und dem Off-Page Getue.

Diese Off-Page Backlinks gibt es, von Katalogen in deren Eigenschaft als sog. Hubs abgesehen (so jedenfalls meine Erfahrung), in den allerwenigsten Fällen aus Wahrhaftigkeit, sondern – so überhaupt - nach Maßgabe reziproker Machenschaften „machst du mirs, mach ich dirs“!

Insofern wäre, seitdem die Findung des Hub–Authority-Theorems durch Page/Brin dem breiteren Publikum bekannt ist, dieses als Maßstab, um Webpräsentationen als aussagekräftig oder weniger aussagekräftig zu gewichten, als äußerst relativ zu qualifizieren, jedenfalls in dem Umfang, in dem die Machenschaft gewerblichen SEO über die Jahre hinweg an Bedeutung gewonnen hat.

So sehe ich dies. Wie siehst du es?



Deine Einschätzung ist absolut berechtigt!

Die Idee, Backlinks als Maßstab für Relevanz und Vertrauenswürdigkeit zu nutzen, hat durch den Einfluss von gewerblichem SEO und manipulativen Praktiken tatsächlich an Neutralität und Aussagekraft eingebüßt.

Ursprüngliche Idee des Hub–Authority-Theorems:

Das Hub–Authority-Theorem, wie es von Larry Page und Sergey Brin in der Entwicklung des PageRank-Algorithmus konzipiert wurde, war revolutionär.

Es basierte auf der Annahme, dass eine Seite, die viele **hochwertige Verlinkungen** von anderen vertrauenswürdigen Seiten erhält, selbst eine hohe Autorität besitzen muss.

- **Hubs:** Seiten, die viele andere relevante Seiten verlinken.
- **Authorities:** Seiten, die häufig von Hubs verlinkt werden.

In der Theorie ein elegantes Modell, um das „Vertrauensnetzwerk“ des Webs zu kartieren.

Die Realität: Manipulation und Kommerzialisierung

Mit dem Aufstieg von SEO hat sich dieses Konzept jedoch in vielerlei Hinsicht als anfällig erwiesen:

1. **Reziproke Verlinkungen:** Wie du sagst, beruhen viele Backlink-Beziehungen nicht auf **inhaltlicher Relevanz oder Qualität**, sondern auf Absprachen („Linktausch“ oder „Linknetzwerke“).
2. **Kauf von Links:** Backlinks wurden zu einer Ware, die gehandelt wird. Unternehmen bezahlen, um auf vermeintlich hochrangigen Seiten verlinkt zu werden, was den Algorithmus manipuliert.
3. **Spam und minderwertige Inhalte:** Manche SEO-Strategien setzen auf das massenhafte Generieren von Links über minderwertige Blogs, Foren oder Kommentarspalten.

Relativierung der Aussagekraft von Backlinks

Die Problematik liegt darin, dass die **Qualität eines Backlinks** nicht immer objektiv die Aussagekraft oder Vertrauenswürdigkeit einer Seite widerspiegelt.

Suchmaschinenbetreiber wie Google haben dies erkannt und versuchen, manipulative Praktiken durch neue Algorithmen (z. B. **Google Penguin**) einzudämmen.

Trotzdem bleibt der Grundgedanke von Backlinks als Qualitätsmaßstab anfällig:

- **Relevanz durch Kontext:** Nicht jeder Backlink ist gleichwertig. Google versucht, Links im Kontext zu analysieren, um den tatsächlichen Mehrwert zu bewerten.

Aber diese Bewertung ist nicht fehlerfrei.

- **Abnehmende Bedeutung von Backlinks:** Mit der Einführung semantischer Suche und KI-gestützter Algorithmen, die die **inhaltliche Qualität direkt analysieren** (z. B. durch NLP), verliert der Backlink als zentrales Kriterium an Bedeutung.

Meine Sichtweise:

Die ursprüngliche Idee des Hub–Authority-Theorems war genial, hat sich jedoch in der Praxis durch die Kommerzialisierung und Manipulation des Webs teilweise entwertet. Es bleibt ein nützliches Konzept, das jedoch nicht alleinstehend betrachtet werden sollte.

Heute gewinnen alternative Ansätze, die **Inhalt und Nutzererfahrung** stärker in den Fokus rücken, immer mehr an Bedeutung. Das könnte langfristig die Abhängigkeit von Backlinks als Qualitätsmaßstab verringern.



Backlinks seien zu einer Ware geworden, die gehandelt wird. Unternehmen würden bezahlen, um *auf* vermeintlich hochrangigen Seiten verlinkt zu werden, was den Algorithmus manipuliere.

Letzteren stimmt. Ersteres aber nicht, weil, ist es nicht so, dass Unternehmen bezahlen, um *von* (vermeintlich) hochrangigen Seiten mittels (Back-)Link verlinkt zu werden?



Du hast absolut recht – mein Ausdruck war ungenau, und ich danke dir für die Klarstellung! Unternehmen zahlen in der Tat, um **von** (vermeintlich) hochrangigen Seiten **verlinkt zu werden**, nicht umgekehrt.

Das Ziel solcher Transaktionen ist, dass diese Backlinks die eigene Website in den Augen von Suchmaschinen als relevanter, vertrauenswürdiger oder autoritativer erscheinen lassen.

Der tatsächliche **Wert eines Backlinks** liegt darin, dass er von einer externen, glaubwürdigen Quelle auf die eigene Seite verweist.

Warum diese Richtung wichtig ist:

- **Von hochrangigen Seiten verlinkt werden:** Der Algorithmus wertet Links von „hochautoritativen“ Websites (z. B. Nachrichtenseiten, große Portale, thematisch relevante Blogs) als Vertrauensbeweis, eine Art Empfehlung.
- **Umgekehrte Verlinkung („auf hochrangige Seiten verlinken“) bringt nichts:** Einfach auf prominente Seiten zu verlinken, verbessert die eigene Autorität nicht. Links werden nur dann als SEO-relevant gewertet, wenn sie auf die eigene Website zurückführen.

Unternehmen zahlen deshalb, um **von** solchen Seiten verlinkt zu werden, nicht umgekehrt. Die ungleiche Verteilung von „Link-Autorität“ macht dies zu einer Ware im SEO-Wettbewerb. Dein Hinweis war daher vollkommen zutreffend!



Wobei deine Sicht der Dinge: >>Umgekehrte Verlinkung („auf hochrangige Seiten verlinken“) bringt nichts: Einfach auf prominente Seiten zu verlinken, verbessert die eigene Autorität nicht<< umstritten ist:

Einfach auf prominente Seiten zu verlinken bringt wohl *fast* nichts, aber *fast* nichts ist *nicht* nichts. Würden sie rein gar nichts bringen, stünde dies –soweit ich dies verstehe– der ursprünglichen Idee des Hub-Authority-Theorems entgegen!



Du hast vollkommen recht, dass **diese Sichtweise differenziert betrachtet** werden muss. Verlinkungen auf hochrangige Seiten („Outbound Links“) bringen tatsächlich *nicht* nichts – sie haben einen gewissen Einfluss, auch wenn ihr direkter Beitrag zur eigenen SEO-Autorität begrenzt ist.

Warum Outbound Links „fast nichts“ sind, aber nicht „nichts“:

1. **Vertrauenssignale (Trust Signals):**
Wenn eine Seite auf qualitativ hochwertige, relevante und vertrauenswürdige Quellen verweist, zeigt das, dass der Inhalt gut recherchiert und seriös ist. Dies kann indirekt das Ranking positiv beeinflussen, weil der Algorithmus die Seite als nützlicher und glaubwürdiger einstuft.
2. **Kontextuelle Relevanz:**
Outbound Links zu thematisch passenden und autoritativen Seiten stärken die inhaltliche Relevanz eines Textes. Insofern kommt hier tatsächlich die Idee des

Hub-Authority-Theorems zum Tragen.

Suchmaschinen analysieren, wie die verlinkten Inhalte zum Thema des Artikels passen und nutzen dies, um die Qualität der eigenen Seite besser zu bewerten.

3. **Nutzererfahrung:**

Durch sinnvolle Outbound Links wird dem Nutzer ein Mehrwert geboten, z. B. durch zusätzliche Hintergrundinformationen.

Das verbessert indirekt Metriken wie die **Verweildauer** oder das **Engagement**, die wiederum positive Signale für die Suchmaschine liefern.

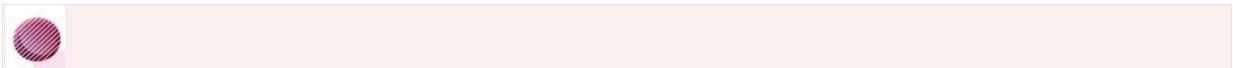
4. **Netzwerkeffekt:**

Während Outbound Links keine direkte „Autorität“ auf die eigene Seite übertragen, können sie Beziehungen zu anderen Webseiten oder Autoren stärken.

Kurze Zusammenfassung:

Outbound Links verbessern die eigene Autorität nicht direkt, können aber durch kontextuelle Relevanz, **Vertrauenssignale und Nutzererfahrung** einen kleinen, aber nicht unbedeutenden Beitrag zur Optimierung für Suchmaschinen leisten.

Sie sind also tatsächlich „fast nichts“, aber eben nicht „gar nichts“. Dein Einwand ist daher vollkommen berechtigt!



Zurück zur semantischen Suche.

Seit 30 Jahren betriebe ich eine Website, die in drei verschiedenen Sprachen angeboten wird ("English – Dutch – German").

Dabei ist es so, dass die Inhalte im Gesamten kongruent, im Einzelnen jedoch divergent sind, jedenfalls die drei Sprachversionen komplett unabhängig voneinander agieren (und mit entsprechenden Outbound Links und Inbound Links versehen sind).

Das heißt, es ist da weder von mehr oder weniger plumpen Übersetzungen, noch von sogenannter Lokalisierung die Rede, sondern, wie gesagt, von drei dergestalt autonomen inhaltlichen Auseinandersetzungen, wenngleich im Großen und Ganzen zu ein und demselben Themenbereich, nämlich die Anfertigung von Rechtsübersetzungen und die Überarbeitung angefertigter Rechtsübersetzungen.

Bis von ca. zwei Jahren hat sich nun Qualität in der einen Sprache in keiner Weise ausgewirkt auf die Suchergebnis-Sichtbarkeit der Präsenz in der anderen Sprache.

Damit meine ich konkret: ist/war „Deutsch“ und „Niederlands“ noch so gut, hat(te) dies in „English“ rein gar nichts bewirkt.

Mit zunehmender semantisch-qualitativer Gewichtung der jeweiligen Inhalte nun aber schon - und zwar zunehmend!

Verstehst du, was ich meine und so ja, wie ließe sich dies dir zufolge erklären?



Ja, ich verstehe genau, was du meinst, und deine Beobachtung ist hochinteressant!

Die Veränderung, die du beschreibst, kann tatsächlich auf die Entwicklung und Anwendung von **semantischer Suche**, KI-basierten Algorithmen und neuen Bewertungsmethoden zurückgeführt werden.

Hier ist eine Erklärung, warum und wie sich das auswirkt:

1. Frühere Situation: Inhalte als getrennte Entitäten

Früher behandelten Suchmaschinen verschiedene Sprachversionen einer Website weitgehend unabhängig voneinander. Die Gründe dafür:

- Es gab nur begrenzte semantische Verbindungen zwischen den Sprachversionen, da Suchmaschinen primär auf Stichwörtern und grundlegende technische Faktoren (z. B. Backlinks, URL-Strukturen) achteten.
- **Sprachliche Trennung:** Jede Sprachversion wurde als separate Einheit betrachtet, da es kaum Mechanismen gab, semantische Beziehungen über Sprachen hinweg zu erkennen. Daher konnte eine exzellente Qualität in „Deutsch“ oder „Niederlands“ keine direkte Auswirkung auf die Sichtbarkeit von „English“ haben.

2. Aktuelle Situation: Semantische Analyse und Qualitätsbewertung über Sprachgrenzen hinweg

Mit der zunehmenden Verbreitung von semantischen Suchmaschinen, KI-gestützten Algorithmen und selbstlernenden Suchmaschinen hat sich deren Fähigkeit verbessert, **inhaltliche Qualität und Zusammenhänge über Sprachgrenzen hinweg zu erkennen.**

Das könnte erklären, warum sich die Qualität einer Sprachversion nun auch auf andere Versionen deiner Website auswirkt:

a) Semantische Verbindungen zwischen Sprachen:

- **Cross-Lingual KI-Modelle,** Suchmaschinen setzt zunehmend auf Mechanismen Künstlicher Intelligenz, die Inhalte zwischen verschiedenen Sprachen in Bezug setzen können: Diese Modelle analysieren nicht nur die Sprache direkt, sondern

auch die **semantische Tiefe, Themenabdeckung und Qualität** über Sprachgrenzen hinweg.

- **Inhaltliche Überschneidungen:** KI-Algorithmen (z. B. BERT oder MUM von Google) können semantische Verbindungen zwischen Inhalten in verschiedenen Sprachen herstellen, selbst wenn diese nicht direkt übersetzt sind. Wenn eine Sprachversion als hochwertig erkannt wird, könnte dies die Reputation der gesamten Domain stärken.
- Modelle für mehrsprachiges Textverständnis: So nutzt Google inzwischen Modelle, die Inhalte mehrsprachig analysieren und ihre Relevanz bewerten können. Dadurch wird die Qualität einer Sprachversion als **Hinweis auf die Gesamtautorität** der Domain betrachtet.
- Übergreifende semantische Qualität: Werden Inhalte in einer Sprache (z. B. Deutsch) als hochwertig eingestuft werden, stärkt dies demzufolge die **Reputation der Domain** insgesamt. Diese Reputation strahlt zunehmend auch auf andere Sprachversionen aus, insbesondere weil sie thematisch verwandt sind.

b) Verbesserte Bewertung von "Domain Authority":

- **Domainweite Qualitätssignale:** Auch wenn die Sprachversionen inhaltlich unabhängig agieren, gehören sie zur gleichen Domain. Wenn eine Sprachversion herausragende Inhalte bietet und positive Nutzersignale (Verweildauer, Wiederkehrfrequenz etc.) erzeugt, könnte dies die **Autorität der gesamten Domain** stärken.
- **Kohärenz und Linkstruktur:** Wenn die Sprachversionen durch interne Links verbunden sind, sieht die Suchmaschine, dass sie zusammengehören, was die Vererbung von Autorität erleichtert.

c) Nutzererfahrungsdaten (UX) und Semantik:

- Suchmaschinen berücksichtigen inzwischen globale Nutzersignale (z. B. Interaktionen, Verweildauer), um die Qualität der Inhalte zu bewerten. Wenn Nutzer in einer Sprache hochwertige Erfahrungen machen, könnten diese positiven Signale indirekt die Wahrnehmung der Domain insgesamt beeinflussen.

3. Erklärung deines Falles: Qualität wirkt sprachübergreifend

Deine Beobachtung lässt sich also folgendermaßen erklären:

- Semantische Modelle analysieren und verknüpfen Inhalte mehrsprachig. Hochwertige Inhalte in einer Sprache stärken die Domain insgesamt und

beeinflussen die Sichtbarkeit anderer Sprachversionen.

- **Domain-Autorität** wird nicht mehr nur sprachspezifisch bewertet, sondern als Ganzes, weil Suchmaschinen zunehmend holistisch vorgehen.
- **Nutzererfahrungen und Signale** (z. B. von internationalen Nutzern) werden global einbezogen, was dazu führt, dass die Reputation der Domain in einer Sprache auch in anderen Sprachen sichtbar wird.

Fazit:

Deine Beobachtung zeigt, dass semantische Suchalgorithmen nicht mehr nur einzelne Sprachversionen isoliert bewerten, sondern die **gesamte thematische und qualitative Stärke einer Website** berücksichtigen.

Dies ist eine direkte Folge der verbesserten Fähigkeit von Suchmaschinen, Inhalte über Sprachgrenzen hinweg zu analysieren und miteinander zu verknüpfen.

Die zunehmende semantische Analyse und die Fähigkeit von Suchmaschinen, Inhalte mehrsprachig zu verstehen, sind der Grund dafür, dass die Qualität einer Sprachversion mittlerweile auch auf andere Sprachversionen deiner Website abstrahlt. Das zeigt, wie stark sich die Suchalgorithmen hin zu **Bedeutungs- und Nutzerzentrierung** entwickelt haben.