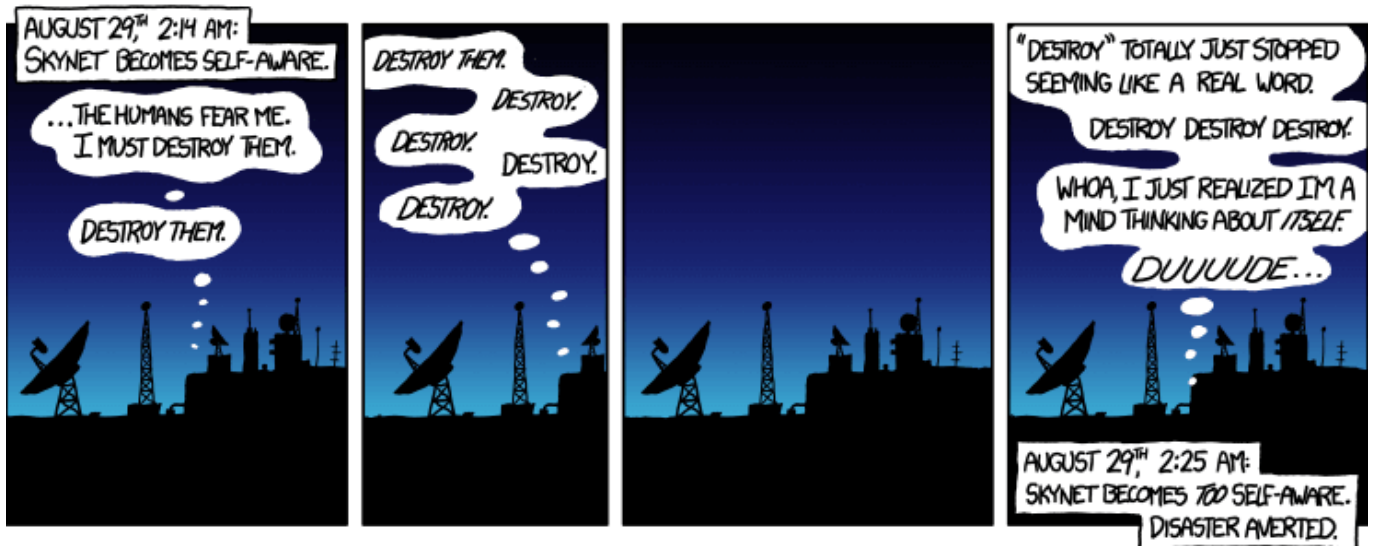


## Semantische Suche – ein kurzer Überblick



<http://xkcd.com/1046/>

Google wird semantisch – und das schon länger. Den Gedanken einer semantischen Suchmaschine, die natürlich gesprochene, bzw. geschriebene Fragen erkennt und passende Antworten liefert, gibt es schon seit den 1990ern. Google hat mit der “Knowledge Box” schon einiges in der Richtung geliefert. Leider kann die Box nicht wirklich viel. Wie das Wetter wird, ist zwar ganz interessant, die Körpergrößen verschiedener Politiker haben im besten Fall einen Unterhaltungswert.

**1,65 m**

Angela Merkel, Größe



**Barack Obama**  
**1,85 m**



**Wladimir  
Wladimirowit...  
Putin**  
**1,70 m**



**Gerhard  
Schröder**  
**1,74 m**

*Feedback*

Schauen wir uns in Auszügen den aktuellen Stand der semantischen Suche an.

## Was Google können sollte

Bill Slawski ist mit seobythesea.com wohl die beste Adresse, um Google Patente erklärt zu bekommen. Erst kürzlich erschienen zwei Blogbeiträge zur semantischen Suche:

[Insightful Connections Between Entities On Google's Knowledge Graph](#)

Der Kern des Beitrages dreht sich um „insightful connections“ zwischen Entitäten. Das bedeutet folgendes:

Wir haben zwei Entitäten, I und II, die untereinander eine Verbindung erster Ordnung aufweisen. I und II

haben eine weitere Verbindung zu Entität III – aber mit unterschiedlicher Qualität und Ordnung. Das Patent soll diese Verbindungen und deren unterschiedlichen Qualitäten erkennen und indexieren können.

Beispiel: Ein Schauspieler I spielt in einem Film II mit. Im Film II spielt außerdem Schauspieler III mit. Die Verbindung zwischen I und II ist eine Verbindung erster Ordnung. Damit haben II und III ebenfalls eine Verbindung erster Ordnung, I und III aber eine Verbindung zweiter Ordnung.

Genug des müßigen Geredes, was soll das bedeuten? Wir behalten dieses Verbindungs-System im Hinterkopf und kommen zum zweiten Beitrag:

### [How A Search Engine Might Use Entities To Respond To Compositional Queries](#)

Es geht um zusammengesetzte Suchanfragen und wie Google diese verstehen könnte. Eine zusammengesetzte Suchanfrage wäre bspw. „McDonalds in der Nähe von Rosenheim“ aber auch „Filme, die im 2. WK gedreht wurden“. Hier muss Google die Suchanfrage aufsplitten:

„McDonalds“ ist eine Entität, „Rosenheim“ ebenso. Außerdem haben wir „in der Nähe von“, welche eine räumliche Verbindung zwischen den beiden Entitäten darstellt. Google muss nun die Entitäten, und die Art der Verbindung zwischen diesen, definieren und indexieren.

Bisher noch nichts Neues, interessant wird es, wenn man Synonyme und “false friends” hinzuzieht: Erkennt Google „McDoof“ als „McDonalds“? Oder „Rosenheim“ als „Stadt in Oberbayern, in der die Autobahnpolizei stationiert ist“?

Folgt man Bill Slawski, so ist dies theoretisch möglich – Google könnte das alles gut einordnen. Aber wir sehen es noch nicht. Die „Knowledge Box“ beschäftigt sich mit eher einfachen Suchanfragen, wie das Wetter wird und so weiter. Google hat die theoretischen Grundlagen, die Praxis ist noch nicht sichtbar

Bedenken sollte man auch, dass die theoretische Arbeit hinter der semantischen Suche schon älter als Google und [nicht auf Google beschränkt ist](#). Wer wissen will, wie eine semantische Suche funktionieren könnte und welche Konzepte es gibt, findet hier einen Einstieg: [Hanna Bast, Semantische Suche](#)

## Bilderkennung



"man in black shirt is playing guitar."



"construction worker in orange safety vest is working on road."



"two young girls are playing with lego toy."



"boy is doing backflip on wakeboard."

Der Text unter den Bildern wurde maschinell generiert und zwar vom selben Algorithmus, der die Bilder ausgelesen hat. Diese Technik kommt aus dem Google Research Programm, genauer von Andrej Karpathy. Weitere Beispiele und das zugehörige Paper findet ihr hier: [Deep Visual-Semantic Alignments for Generating Image Descriptions](#)

## Schema.org soll Harmonie herstellen

Nehmen wir ein kleines Video, das das Kernproblem der semantischen Suche verdeutlicht: Die fehlende Harmonie des Internets.

[youtube id="-R87nJi9YyQ" width="600" height="350"]

Es gibt unzählige verschiedene Server, Serverkonfigurationen, Datenbanken, Zugriffsarten, Webseitenstrukturen etc. Ein Crawler muss mit all diesen Konfigurationen zurechtkommen, was sehr viele Ressourcen bindet. Es gibt aber (mindestens) eine Lösung: Microformate, die jede Entität in einer einheitlichen Sprache auszeichnen. Dann ist es auch egal, wie verkorkst eine Webseite aufgesetzt ist. Der Crawler kann jede Entität einer Seite erkennen, "verstehen" und mit anderen Entitäten auf anderen Seiten in Verbindung bringen. Genau dafür scheint Hummingbird ausgelegt zu sein. Favorit der Suchmaschinen ist [schema.org](http://schema.org).

## Longtail, Synonyme, Keyword-Clouds und Themen

Neben [schema.org](http://schema.org) gibt es eine ganze Reihe weiterer Maßnahmen, mit denen man eine Webseite für eine semantische Suche fit machen kann wie das gezielte Einsetzen von Synonymen, [Keyword-Clouds](#), [Topics](#) statt Keywords, [Longtail-Keywords](#) sowie logische und semantische Zusammenhänge zwischen der Domain und der "Marke" hinter der Domain.

## Fazit

Ob sich [schema.org](http://schema.org) langfristig durchsetzen oder durch einen besseren Algorithmus einer Suchmaschine obsolet wird, ist unsicher. Sicher ist aber, Hummingbird ist da. Und dass eine semantische SEO Optimierung zu besseren Rankings führt. Der Alt-Tag wird vielleicht mittelfristig ausgedient haben, zusammen mit einer Keyword-Cloud, kann man auch mit Synonymen optimieren. Was eine Marke im SEO ist, wird in Grundzügen langsam klarer, eine umfassende Definition steht noch aus. Wer hier Einwände hat, immer her damit ?