



De missie van Semanta is het verkleinen van de digitale afstand tussen de bedenkende mens en de zoekende mens.

De enorme hoeveelheid data die Openai, Google en Bing en anderen hebben verzameld heeft een boost teweeg gebracht die door sommigen een revolutie genoemd wordt. Het jargon dat daarvoor gebruikt wordt is een GPT op zichzelf. Ik zal alle termen voorlopig even achterwege laten en mij beperken tot 4 hoofdthema's.

De mens staat daarin centraal en is het begin- en eindpunt van de (digitale) kennisoverdracht, die door een (virtuele)gesprekspartner ontsloten kan worden door toegang tot de kennis die door de bedenkende aan het internet toevertrouwd is.



Bedenkende mens

Verwijst naar het vermogen van mensen om na te denken over zichzelf, hun omgeving en de wereld om hen heen.



Kennis

Internet als een uitgebreide bron voor het verwerven van kennis



Gesprekspartner

Komt voor in de contexten van conversatie, taalbots, kennis, en beroepen binnen de documenten



Zoekende mens

Het fundamentele menselijke streven naar het verwerven van kennis en inzicht.

Bedenkende mens

De bedenkende mens heeft in de afgelopen 30 jaren bijna 2 miljard websites gebouwd die ieder op zich bruikbare informatie kunnen bevatten voor het maken van Tiny Language Models.. Daarvan heeft Semanta er meer dan 300 verwerkt met een totale grootte van 2GB.

Van iedere [website bij Semanta](#) is de website omgezet naar een verzameling tekstbestanden, die de basis vormen voor digitaal gereedschap. Gereedschap waarmee iedere website voorzien kan worden van een op [chatbase.co](#) gebaseerde digitale assistent zoals [Edrobot](#) en door het [Semanta college](#) gebouwde gereedschappen.

Google

hoeveel websites heeft google verwerkt

Het is onmogelijk om het exacte aantal websites te bepalen dat Google heeft verwerkt. De zoekmachine crawlt voortdurend het internet en indexeert nieuwe websites en pagina's.

We kunnen echter wel enkele schattingen maken:

1. Aantal websites op het internet:

- Op 31 december 2022 waren er meer dan **1,97 miljard websites**. (<https://www.internetlivestats.com/>)
- Google heeft **niet alle** websites op het internet geïndexeerd. Sommige websites zijn nieuw, offline of geblokkeerd voor Google.

2. Aantal geïndexeerde websites:

- In 2022 had Google **meer dan 60 biljoen pagina's** in zijn index. (<https://blog.google/>)
- Het exacte aantal websites is onbekend, maar het is waarschijnlijk **een fractie van het aantal pagina's**.

3. Crawlsnelheid van Google:

- Google crawlt voortdurend het internet, met een snelheid van **miljarden pagina's per dag**. (<https://blog.google/>)

Conclusie:

- Google heeft **een enorm aantal websites** verwerkt, waarschijnlijk **meer dan 100 miljard**.
- Het exacte aantal is onbekend en verandert voortdurend.

Extra informatie:

- U kunt het aantal geïndexeerde pagina's van uw website controleren met **Google Search Console**. (<https://search.google.com/search-console/about>)
- U kunt meer te weten komen over hoe Google websites crawlt en indexeert in de **Google Search Console Help Center**. (<https://support.google.com/webmasters/?hl=en>)