

Semanta College

Van chatbots en digitale medewerkers

Nederland Tel: +31-162-313323

Canada Tel: +1-514-331-0572



Van chatbots en digitale medewerkers



- Versiebeheer: 25.08.2021
- Author: Edward A .kool
- Date: 9 maart 2021
- ©Ed Kool 2001-2021



SEMANTA

- Processing pattern
 - Starting point
 - Corpus
 - Internet of Things
 - Talk to a digital Employee



EDDY

- Language teacher
- Internet helper
- How it works



SEMANTUS

Versiebeheer: 25.08.2021

Versiebeheer: 23.01.2021	2
Kennismaking	4
Abstract	4
Inleiding	5
Wat is een chatbot cq gesprekspartner?	7
Chatbot	7
Taalbot.....	7
Digitale gesprekspartner.....	7
Menselijke gesprekspartner.....	7
Xbot.....	8
Scenarioschrijver	8
Kunstmatige intelligentie.....	9
Xbot beleving	9
Conclusie.....	10
Semanta de lerares.....	10
Verwerkingspatroon	10
Uitgangspunt	11
Corpus	11
Internet of Things	11
Spreken met een virtuele gesprekspartner.....	12
Language teacher.....	12
Internet helper	12
Hoe het werkt ?	13

Kennismaking

In 1970 was ik een beginnend programmeur. Cobol, Fortran, Assembler en RPG waren gesneden koek geworden. Later aangevuld met Pascal, PL/1, Basic, C++, Java. Op een zomerse namiddag in 1970, in gedachten verzonken, kreeg ik een visie over de toekomst van programmeertalen en computers in het algemeen. Mijn gedachte stopte bij een computerprogramma dat min of meer menselijk zou kunnen communiceren in het Nederlands.

De gedachte is mij bijgebleven en ja hoor in 1993 maakte ik kennis met het eerste exemplaar van PARS een automatisch vertaalprogramma uit Kharkov in de Oekraïne. In 1994 heb ik dat gedemonstreerd op een bijeenkomst van Surfnet over Kunstmatige Intelligentie.

In 1995 maakte ik verder kennis met Internet en wist ik wat ik wilde: Vertalen met een computer via het Internet. Het duurde tot 2001 voordat er een Nederlandse en Engelse versie waren. Later uitgebreid naar Duits en Pools. Tot 2006 was dit een lucratieve bron van inkomsten.

In 2006 werd ik 58 en dat werd mijn pensioenleeftijd, vanaf die tijd is de combinatie taal en computers mijn dagelijkse activiteit geweest. In eerste instantie nog commercieel met semantisch verrijkte lexicons voor 40 talen. Vanaf 2010 is daar de conversationele chatbot bijgekomen. Tot 2015 met AIML en Pandorabots en daarna eigen programmatuur in HTML, PHP, Javascript en MySQL.

In 2020 was de architectuur voor Semanta gereed en ben ik nu zover dat de Semanta chatbots naast automatische en semi-automatische generatie van vragen en antwoorden, in staat moet worden gesteld daar een coherent en begrijpelijk geheel van te maken.

De chatbot is in staat als digitale gesprekspartner te fungeren. De conversatie tussen mens en humanoid of humanoids onderling wordt vanuit meerdere bronnen tot een begrijpelijk verhaal geconstrueerd.

In tegenstelling tot Google, Microsoft en andere techreuzen is de basis voor Semanta gelegd met lexicons, grammatica en semantische eigenschappen van een taal.

Semanta maakt van tekst begrijpelijke kennis

Abstract

Semanta is het systeem waarmee Lingvistica producten en diensten ten behoeve van het leerproces voor chatbots en digitale representanten ontwikkelt.

Dit document is bedoeld voor iedereen die webbezoekers tegemoet wil treden met een virtuele gesprekspartner. De kennis die daarvoor nodig is, ligt besloten in de eigen website en daarbuiten in het "Internet of things". Semanta stelt een webmaster in staat bronnen tekstueel te analyseren en daarmee de gesprekspartner een integraal onderdeel van de website te laten zijn. In dit artikel wordt ingegaan op de vraagstelling "[Semanta of hoe een computerprogramma zou kunnen leren praten](#)" en het antwoord daarop van Ed Kool.

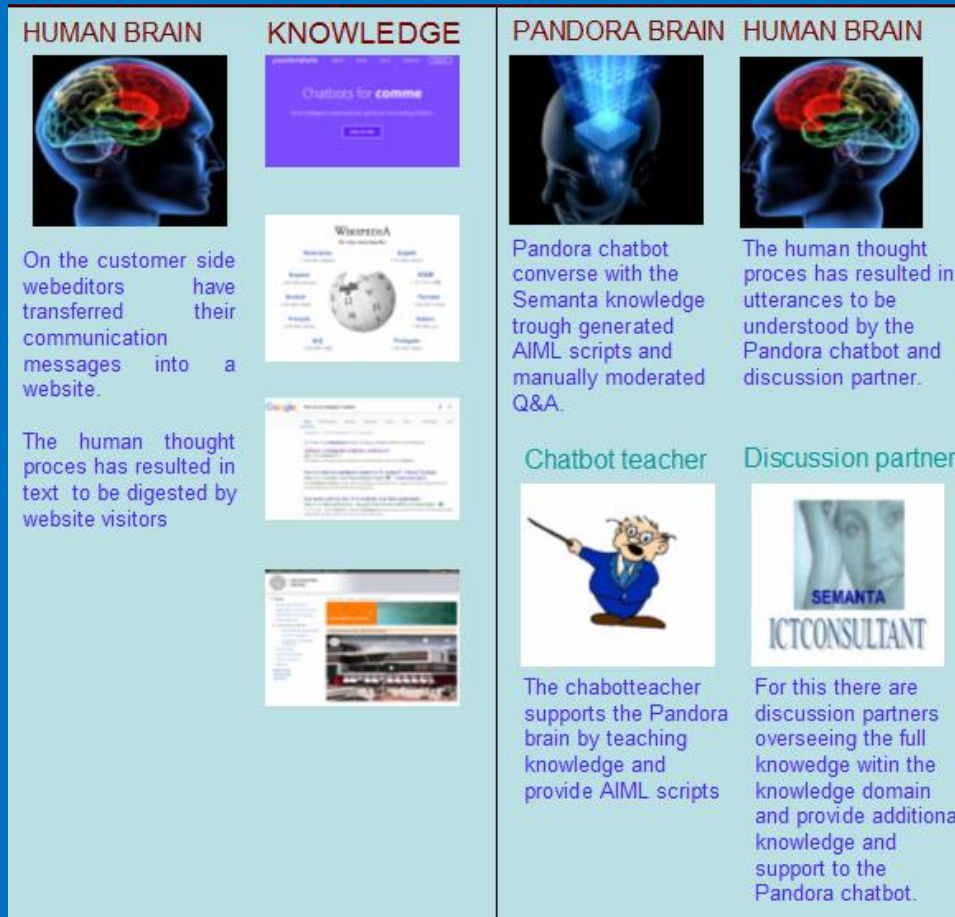
Inleiding

Semanta is het systeem waarmee Lingvistica producten en diensten ten behoeve van het leerproces voor chatbots en digitale representanten ontwikkelt.

Dit document is bedoeld voor iedereen die webbezoekers tegemoet wil treden met een virtuele gesprekspartner. De kennis die daarvoor nodig is, ligt besloten in de eigen website en daarbuiten in het "Internet of things". Semanta stelt een webmaster in staat bronnen tekstueel te analyseren en daarmee de gesprekspartner een integraal onderdeel van de website te laten zijn. In dit artikel wordt ingegaan op de vraagstelling "[Semanta of hoe een computerprogramma zou kunnen leren praten](#)" en het antwoord daarop van Lingvistica.

Alpha versie 15.04.02 was de eerst versie waarmee belangstellenden zijn benaderd. De Alpha versie is voornamelijk gebruikt om interesse te kweken en te toetsen in hoeverre er belangstelling bestaat voor de Semanta Services.

Van chatbots en digitale medewerkers



Figuur 1 Het digitale Semanta proces

Wat is een chatbot cq gesprekspartner?

Chatbot

Lingvistica verstaat onder chatbot een systeem van computerprogramma's, gegevensbanken en procedures waarmee via het internet min of meer menselijk gecommuniceerd kan worden.

Taalbot

Lingvistica verstaat onder taalbot een systeem van computerprogramma's, gegevensbanken en procedures waarmee via het internet gecommuniceerd kan worden. De architectuur van Semanta stelt de webmaster of webeditor in staat de taalbot in te zetten per ondersteunde taal. Op dit moment zijn er taalbots voor het Nederlands, Engels en Russisch.

Digitale gesprekspartner

Lingvistica verstaat onder een digitale gesprekspartner een script dat in staat is om in een of meer kennisdomeinen kennis te vinden en in een bruikbaar antwoord om te zetten. Ieder beroep en dat zijn er meer dan 1250 kan door een gesprekspartner digitaal vertegenwoordigd worden.

Menselijke gesprekspartner

De Menselijke gesprekspartner, **gezichtsherkenning** of de **omgevingstemperatuur**, en kan op basis van externe of interne commando's in **beweging** komen. Dit geldt niet voor gefixeerde processoren als chatbots.

Xbot

De Xbot weet waar hij zich bevindt en kan op commando of op eigen initiatief van de ene ruimte naar een andere gaan en om via de trap naar boven of naar beneden te gaan. De robot kan iemand anders dan Ed Kool verwelkomen weten waar **Ed Kool** is en wat hij op een bepaald moment doet. Door de tijd ingegeven handelingen te verrichten. **Ontbijt** klaarmaken of een **kopje thee** zetten.

Medicijngebruik te initiëren door deze op een vastgelegd tijdstip aan te bieden.

Misschien wel een van de belangrijkste functies is het bewaken van de leefomgeving, wordt Ed op tijd **wakker**, is het gas blijven branden, zijn de lichten uit, is de **verwarming** laag, zijn de gordijnen dicht, is de deur op slot. Als er een calamiteit of afwijking wordt geconstateerd, automatisch bellen. Alle communicatie over het gedrag van Ed Kool tussen de Xbot is tekstueel te duiden.

Scenarioschrijver

Voor Lingvistica is de VGPT scenarioschrijver de eerste gesprekspartner die opgeleid wordt door Semantus om een Webeditor in staat te stellen scenario's voor zijn of haar specifieke Xbot te visualiseren en het scenario en conversatiegedrag op te zetten. Op basis van het Xbotpalet, zie boven kunt u met behulp van Semantus uw eigen scenarioschrijver tot leven brengen.

Kunstmatige intelligentie

De eigenlijke intelligentie zit in de verwerking van taal en met name de semantische aspecten. Semantische aspecten van woordvormen worden bepaald door de context waarin de woordvormen worden gebruikt. Eigenlijk is een robot een mechanisch voorwerp met bionische eigenschappen aangevuld met een processor waarin alle aspecten van een robot samenkomen. Op dit moment zijn produktierobots niet of nauwelijks uitgerust met intelligente spraaktechnologie. Slechts een beperkt aantal robots is in staat het totaal aan gebeurtenissen rond een robot vertaald naar teksten. Dit geldt voor alle signalen van een robot van sensoren, spraakherkenning en gezichtsherkenning, omgevingstemperatuur, vochtigheid, luchtkwaliteit, afmetingen van de ruimte etc. Signalen die net als bij de mens invloed hebben op het **"denkproces"** van de robot, en kunnen in de vorm van tekst aangeboden worden aan het **taalverwerkingsproces** van de robot. .

Xbot beleving

De robot is in staat de handdruk van een mens te meten en vast stellen of de handdruk ferm, normaal of licht was. Door dit in te integreren in de actuele conversatie wordt de Xbot vermenselijkt.

Hetzelfde geldt voor omgevingsgeluiden waarvan de robot boven een aantal decibel in de conversatie het verzoek kan uiten om het geluidsniveau te verlagen.

Bij robots in de zorg doet zich de vraag voor hoe doelmatig dergelijke zorgrobots met de gebruikers kunnen communiceren

Het ontwikkelen van wearable robotics waarvan exoskeletten een extreem voorbeeld zijn kan bijdragen aan het verlichten van fysiek zware beroepen.

Conclusie

Samenvattend is de conclusie dat alle electromechanische signalen van de robot, mits vertaald naar tekst integreerbaar zijn in het taalverwerkingsproces.

Semanta de lerares

De digitale leraar of lerares is een speciale virtuele gesprekspartner. Naast het voeren van een gesprek dient de leraar of lerares in staat te zijn aangeboden informatie in kennis om te zetten. Deze kennisvergroting is een continue proces dat deels automatisch en deels door menselijk handelen uitgevoerd wordt. Lingvistica kan uw medewerkers opleiden om uw digitale docenten te trainen. Semanta is een digitale lerares die andere digitale leraren op kan leiden. De **Pandora-versie** wordt door Semanta gezien als de gesprekspartner waarmee de bezoeker in gesprek gaat en opgeleid is via de

De analyse van de uiting en conversatie-elementen verloopt volgens een vast patroon.

Verwerkingspatroon

- **Tekstselectie.** Losse uiting, tekstbestand, website webpagina's.
- **Vorbewerking.** Splitsing uitingen, woordvormen en segmenten
- **Grammaticale analyse.** Karakter, gevoel en bruikbaarheid
- **Semantische analyse.** Relatie in de context
- **Verrijking.** Aanvullingen op karakter, gevoel en taalkundige kenmerken
- **Vastlegging en conversie.** Afhankelijk van de ontvangende digitale partner is dit proces automatisch of handmatig

Program O

Versie die ter beschikking staat aan de webmaster. De rol van Lingvistica bestaat er uit dat de webmaster wegwijs wordt gemaakt in het toepassen van de semantafunctionaliteit voor zijn of haar website. Omdat virtuele gesprekspartners voorgesteld kunnen worden als lerende virtuele robots, bestaat er ook behoefte aan leraren voor deze robots. Lingvistica speelt daarop in met producten en diensten die het bouwen, onderhouden en exploiteren van chatbots vereenvoudigt en digitale leraren opleidt, die voor het onderwijs van chatbots ingezet kunnen worden.

Uitgangspunt

Om te kunnen "praten" heeft een computerprogramma kennis nodig. Voor Lingvistica is dit kennis in de vorm van teksten. Teksten die in allerlei vormen aangeboden kunnen worden. Iedere uiting via het internet bevat zinvolle informatie.

Voor een website geldt dat Semanta het internet gebruikt als bron om er verwerkbare tekst van te maken. Aan iedere tekstvorm, corpus, tekstbestand, platte tekst of URL, wordt een taal, gespreksonderwerp en kennisdomein toegevoegd, waarmee een gesprekspartner aan de slag kan. De taal waarin de gesprekspartner "praat" wordt bepaald door de internetlocatie van de bezoeker en of de inhoud van de aangeboden tekst.

Is de tekst meer dan 30KB groot, zal het tekstbestand als corpus gekenmerkt worden. Voordat een tekst kan worden verwerkt, dient het bestand opgeladen te worden naar de Semanta server. Dit geldt voor een corpus, tekstbestand en platte tekst. Voor iedere vorm heeft Semanta scripts ontwikkeld. Nadat de tekst is opgeladen zal Semanta de inhoud verdelen in zinnen en zinsdelen, die op hun beurt individueel te bewerken zijn. Van alle zinnen en zinsdelen wordt er grammaticale kennis vastgelegd op basis van de individuele woordvormen.

Corpus

Een gesprekspartner gebruikt het corpus om te converseren met de bezoeker van de website. Het corpus is opgebouwd uit de relaties tussen woordvormen, woordsegmenten, webpagina's. Een corpus bevat "woordelijke" representaties van de uitingvorm over een onderwerp op het internet. Een corpus maakt deel uit van een kennisdomein dat gebruikt kan worden door een of meer gesprekspartners. **Tekstbestand** Bij bestanden tot 30 KB biedt Semanta de mogelijkheid tekstbestanden met de extensies: **TXT**, **DAT**, **AIML** op te laden en voor analyse aan te bieden. De verwerkingstijd is ongeveer 5 secondes per KB tekst.

Platte tekst Tekst tot 1024 tekens wordt door Semanta beschouwd als "platte" tekst. Voor teksten boven dit aantal kunnen in de vorm van een tekstbestand aangeboden worden aan Semanta. Semanta gebruikt "platte tekst" via een dialoog met de bezoeker van een website. In korte vragen en antwoorden wordt de conversatie gevoerd. Bij de verwerking van platte tekst kan ook een URL opgegeven worden waaruit de tekstinformatie wordt onttrokken en wordt aangeboden aan Semanta. De tekst is "raw" en vereist dat de gebruiker een keuze maakt uit de gevonden tekst.

Internet of Things

In het "Internet of things" is ieder object met een IP-adres en voldoende software te representeren als virtuele gesprekspartner. U kunt uw horloge vragen hoe laat het is en de wasmachine laten weten in welk programma de was gedraaid moet worden. In aansluiting daarop kan Semanta met name een bijdrage leveren aan de opzet van tekstuele elementen in het laten genereren van kenniskaarten of Google Knowledge Graphs voor uw website.

Spreken met een virtuele gesprekspartner

Voor de implementatie van spraak in de Semanta services wordt getest met **Nuance** en **ReadSpeaker** als losse services aan te koppelen zijn. Ook Pandora.org voorziet in een sprekende gesprekspartner, waarvoor wij een Semanta versie hebben ontwikkeld. De toenemende vraag om in spreektaal te communiceren is in de architectuur van de Semanta software beantwoord en kan eenvoudig geïmplementeerd worden.

Language teacher

De Semantus is de eerste implementatie van haar rol als LANGUAGE TEACHER waarmee Semanta probeert een antwoord te vinden op de vraag "Kan een computer leren praten?". Lingvistica heeft de basis gelegd voor het positieve antwoord op deze vraag. Op basis van onze Semanta-technologie is het mogelijk om een **niet-Nederlands sprekende bezoeker** vanuit Semanta gereedschappen aan te reiken om vanuit zijn moedertaal Nederlands te leren kennen. Op basis van een enkel woord uit de Nederlandse woordenschat. Wat voor het Nederlands geldt is ook van toepassing op alle andere talen waarvoor wij diensten en producten hebben ontwikkeld.

In overleg met diverse instanties en internetonderzoek heeft Lingvistica in Semanta opgezet om een begin te maken met het voor wat Lingvistica beschouwt als een doorbraak op het gebied van taalbewerking in Nederland.

Internet helper

De virtuele gesprekspartner Scenarioschrijver is de eerste implementatie van haar rol als INTELLIGENT HELPER waarmee Semanta probeert een antwoord te vinden op de vraag "Kan een computer leren praten?". Lingvistica heeft de basis gelegd voor het positieve antwoord op deze vraag.

Op basis van onze Semanta-technologie is het mogelijk om een bezoeker vanuit Semanta gereedschappen aan te reiken om vanuit zijn moedertaal met een gesprekspartner uit een geselecteerd corpus vragen en antwoorden te genereren.

Hoe het werkt ?

Dit gebeurt op twee manieren:

- Semantische kenmerken van individuele woordvormen
- Grammaticale segmenten van 2 tot 5 woordvormen

Afhankelijk van de rolverdeling in het opleidingstraject kan de rol van Semanta, onze eerste **digitale lerares**, ingevuld worden voor een **virtuele gesprekspartner**. Onder leiding van een menselijke **webeditor** kunnen de tools van Semanta ingezet worden om de chatbot, avatar of robot , reproduceerbare kennis bij te brengen. De informatie kan afkomstig zijn uit losse uitingen, teksten, tekstbestanden, websites, wikipedia, bol.com,, google etc. Iedere tekst wordt verondersteld uit een ongestructureerd aantal uitingen te bestaan waarin vragen en antwoorden besloten kunnen liggen. Kennis die uiteindelijk door middel van **intelligente conversaties** tussen de menselijke en en virtuele gesprekspartners ontsloten kan worden.