

From chatbots and digital employees

Semanta College

From chatbots and digital employees

Netherlands Tel: +31-162-313323

Canada Tel: +1-514-331-0572



From chatbots and digital employees



- Versiebeheer: 23.01.2021
- Auteur: Edward A .kool
- Datum: 9 maart 2021
- ©Lingvistica bv 2001-2021



SEMANTA

- Verwerkingspatroon
 - Uitgangspunt
 - Corpus
 - Internet of Things
 - Spreken met een virtuele gesprekspartner



EDDY

- Language teacher
- Internet helper
- Hoe het werkt



SEMANTUS

Version control: 25. 08.2021

From chatbots and digital employees

Version control: 25. 08.2021	2
Acquaintance	4
Abstract	4
Introduction	5
What is a chatbot or interlocutor?	6
Chatbot	6
Languagebot	6
Digital Employee.....	6
Human Employee	6
Xbot.....	7
Scriptwriter.....	7
Artificial intelligence.....	8
Xbot experience.....	8
Conclusion.....	9
Semanta the teacher.....	9
Processing pattern	9
Starting point	10
Corpus	10
Internet of Things	10
Speaking with a virtual interlocutor	10
Language teacher.....	11
Internet helper	11
How it works ?	12

Acquaintance

In 1970, I was a budding programmer. Cobol, Fortran, Assembler and RPG had become a piece of cake. Later supplemented with Pascal, PL/1, Basic, C++, Java. On a summer afternoon in 1970, lost in thought, I got a vision of the future of programming languages and computers in general. My thought stopped at a computer program that could communicate more or less humanly in Dutch.

The thought has stayed with me and yes in 1993 I was introduced to the first copy of PARS a machine translation prograama from Kharkov in Ukraine. In 1994 I demonstrated this at a surfnet meeting about Artificial Intelligence.

In 1995 I was introduced to the Internet and I knew what I wanted: Translating with a computer via the Internet. It took until 2001 before there were a Dutch and English version. Later expanded to German and Polish. Until 2006, this was a lucrative source of income.

In 2006 I turned 58 and that became my retirement age, from that time on the combination of language and computers has been my daily activity. Initially, commercially enriched lexicons for 40 languages. Since 2010, the conversational chatbot has been added. Until 2015 with AIML and Pandorabots and then own software in HTML, PHP, Javascript and MySql.

In 2020, the architecture for Semanta was ready and I am now ready to enable the Semanta chatbots, in addition to automatic and semi-automatic generation of questions and answers, to make it a coherent and understandable whole.

The chatbot is able to act as a digital interlocutor. The conversation between humans and humanoids or humanoids is reconstructed from multiple sources into an understandable story.

Unlike Google, Microsoft and other tech giants, the foundation for Semanta has been laid with lexicons, grammar and semantic properties of a language.

Semanta turns text into understandable knowledge

Abstract

Semanta is the system with which Lingvistica develops products and services for the learning process for chatbots and digital representatives.

This document is intended for anyone who wants to meet web visitors with a virtual interlocutor. The knowledge required for this is contained in your own website and outside it in the "Internet of things". Semanta enables a webmaster to analyze sources textually and thus

From chatbots and digital employees

let the interlocutor be an integral part of the website. This article discusses the question "Semanta or how a computer program could learn to talk" and the answer to it by Ed Kool.

Introduction

Semanta is the system with which Lingvistica develops products and services for the learning process for chatbots and digital representatives.

This document is intended for anyone who wants to meet web visitors with a virtual interlocutor. The knowledge required for this is contained in your own website and outside it in the "Internet of things". Semanta enables a webmaster to analyze sources textually and thus let the interlocutor be an integral part of the website. This article discusses the question "Semanta or how a computer program could learn to talk" and the answer to it of Lingvistica.

Alpha version 15.04.02 was the first version with which interested parties were approached. The Alpha version has mainly been used to generate interest and to test to what extent there is interest in the Semanta Services.

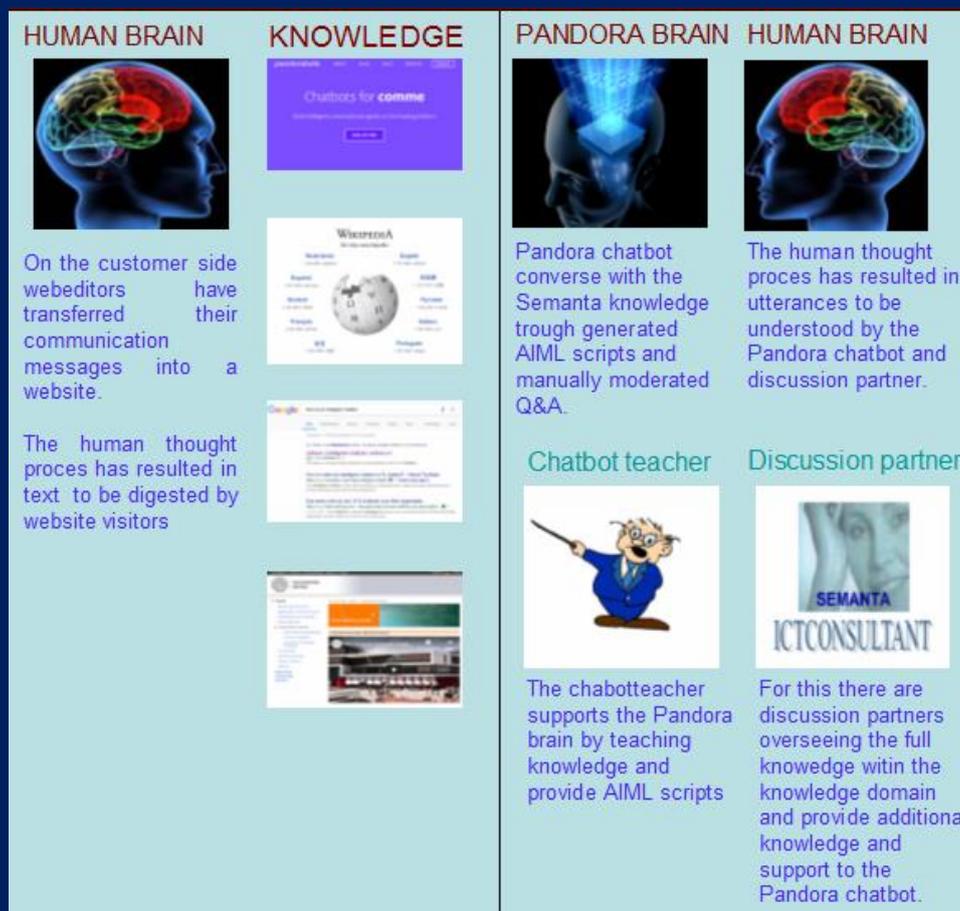


Figure 1 The digital Semanta process

What is a chatbot or interlocutor?

Chatbot

Lingvistica understands chatbot to mean a system of computer programs, databases and procedures with which more or less human communication can be done via the internet.

Languagebot

Lingvistica understands language bot to mean a system of computer programs, databases and procedures with which communication can be done via the Internet. Semanta's architecture allows the webmaster or web editor to use the language bot per supported language. At the moment there are language bots for Dutch, English and Russian.

Digital Employee

Understands a digital interlocutor to mean a script that is able to find knowledge in one or more knowledge domains and to convert it into a usable answer. Every profession and there are more than 1250 can be represented digitally by an interlocutor.

Human Employee

The Human interlocutor, **facial recognition** or the ambient **temperature**, and can move based on external or internal commands. This does not apply to fixed processors such as chatbots.

Xbot

The Xbot knows where it is located and can go from one room to another on command or on its own initiative and to go up or down the stairs. The robot can welcome someone other than Ed Kool knowing where **Ed Kool** is and what he's doing at any given moment. To perform actions dictated by the time. **Prepare** breakfast or make a cup of tea.

Initiate medication use by offering it at a fixed time.

Perhaps one of the most important functions is to monitor the living environment, Ed **wakes upon time**, the gas is kept on, are the lights off, is the **heating** low, are the curtains closed, the door is locked. If an emergency or deviation is detected, call automatically. All communication about the behavior of Ed Kool between the Xbot can be interpreted textually.

Scriptwriter

For Lingvistica, the VGPT screenwriter is the first interlocutor trained by Semantus to enable a Web editor to visualize scenarios for his or her specific Xbot and to set up the scenario and conversational behavior. Based on the Xbot palette, see above you can bring your own scenario writer to life with the help of Semantus.

Artificial intelligence

The actual intelligence lies in the processing of language and especially the semantic aspects. Semantic aspects of word forms are determined by the context in which the word forms are used. In fact, a robot is a mechanical object with bionic properties supplemented by a processor in which all aspects of a robot come together. At the moment, production robots are hardly equipped with intelligent speech technology. Only a limited number of robots are able to translate the total number of events around a robot into texts. This applies to all signals from a robot from sensors, speech recognition and facial recognition, ambient temperature, humidity, air quality, dimensions of the room etc. Signals that, as in humans, influence the **"thought process"** of the robot, and can be offered in the form of text to the language processing **process** of the robot. .

Xbot experience

The robot is able to measure the handshake of a human and determine whether the handshake was firm, normal or light. By integrating this into the current conversation, the Xbot is humanized.

The same applies to ambient sounds of which the robot can express the request above a number of decibels in the conversation to reduce the noise level.

In the case of robots in healthcare, the question arises as to how efficiently such care robots can communicate with users

Developing wearable robotics of which exoskeletons are an extreme example can help to relieve physically demanding occupations.

Conclusion

In summary, the conclusion is that all electromechanical signals of the robot, if translated to text, are integrable in the language processing process.

Semanta the teacher

The digital teacher is a special virtual interlocutor. In addition to conducting a conversation, the teacher or teacher must be able to convert offered information into knowledge. This increase in knowledge is a continuous process that is carried out partly automatically and partly by human action. Lingvistica can train your employees to train your digital teachers. Semanta is a digital teacher who can train other digital teachers. The **Pandora version** is seen by Semanta as the interlocutor with whom the visitor enters into a conversation and is trained via the

The analysis of the expression and conversational elements follows a fixed pattern.

Processing pattern

- **Selection.** Loose utterance, text file, website web pages.
- **Pre-processing.** Split utterances, word forms, and segments
- **Grammatical analysis.** Character, feel and practicality
- **Semantic analysis.** Relationship in context
- **Enrichment.** Additions to character, feeling and linguistic characteristics
- **Capture and conversion.** Depending on the receiving digital partner, this process is automatic or manual

Program O

Version available to the webmaster. The role of Lingvistica is to guide the webmaster in applying the semanta functionality for his or her website. Because virtual interlocutors can be presented as learning virtual robots, there is also a need for teachers for these robots. Lingvistica responds to this with products and services that simplify the construction, maintenance and operation of chatbots and train digital teachers, who can be used for the education of chatbots.

Starting point

In order to be able to "talk", a computer program needs knowledge. For Lingvistica, this is knowledge in the form of texts. Texts that can be offered in all kinds of forms. Every expression via the internet contains meaningful information.

For a website, Semanta uses the internet as a source to turn it into processable text. To each text form, corpus, text file, plain text or URL, a language, topic of conversation and knowledge domain is added, with which a conversation partner can get started. The language in which the interlocutor "talks" is determined by the visitor's internet location and or the content of the text offered.

If the text is more than 30KB in size, the text file will be characterized as a corpus. Before a text can be processed, the file must be uploaded to the Semanta server. This applies to a corpus, text file and plain text. Semanta has developed scripts for each form. After the text has been uploaded, Semanta will divide the content into sentences and phrases, which in turn can be edited individually. Of all sentences and phrases, grammatical knowledge is recorded based on the individual word forms.

Corpus

A conversation partner uses the corpus to converse with the visitor to the website. The corpus is made up of the relationships between word forms, word segments, web pages. A corpus contains "verbatim" representations of the form of expression on a subject on the Internet. A corpus is part of a knowledge domain that can be used by one or more interlocutors. **Text file** For files up to 30 KB, Semanta offers the possibility to upload text files with the extensions: , **TXDAT**, **AIML** and to offer them for analysis. The processing time is about 5 seconds per KB of text.

Plain text Text up to 1024 characters is considered by Semanta as "plain" text. For texts above this number can be offered to Semanta in the form of a text file. Semanta uses "plain text" through a dialogue with the visitor of a website. The conversation is conducted in short questions and answers. When processing plain text, a URL can also be specified from which the text information is extracted and offered to Semanta. The text is "raw" and requires the user to choose from the found text.

Internet of Things

In the "Internet of things", any object with an IP address and sufficient software can be represented as a virtual interlocutor. You can ask your watch what time it is and let the washing machine know in which program the laundry should be turned. In line with this, Semanta can in particular contribute to the design of textual elements in the generation of knowledge maps or Google Knowledge Graphs for your website.

Speaking with a virtual interlocutor

From chatbots and digital employees

For the implementation of speech in the Semanta services is tested with **Nuance** and **ReadSpeaker** if separate services can be linked to. Also Pandora.org provides a speaking interlocutor, for which we have developed a Semanta version. The increasing demand to communicate in colloquial language has been answered in the architecture of the Semanta software and can be easily implemented.

Language teacher

The Semantus is the first implementation of her role as LANGUAGE TEACHER with which Semanta tries to find an answer to the question "Can a computer learn to talk?" Lingvistica has laid the foundation for the positive answer to this question. Based on our Semanta technology, it is possible to provide a **non-Dutch speaking visitor** from Semanta with tools to get to know Dutch from his mother tongue. Based on a single word from the Dutch vocabulary. What applies to Dutch also applies to all other languages for which we have developed services and products.

In consultation with various agencies and internet research, Lingvistica has set up in Semanta to make a start with what Lingvistica considers to be a breakthrough in the field of language editing in the Netherlands.

Internet helper

The virtual interlocutor Screenwriter is the first implementation of her role as INTELLIGENT HELPER with which Semanta tries to find an answer to the question "Can a computer learn to talk?" Lingvistica has laid the foundation for the positive answer to this question.

Based on our Semanta technology, it is possible to provide a visitor with tools from Semanta to generate questions and answers from his mother tongue with a discussion partner from a selected corpus.

How it works ?

This is done in two ways:

- Semantic characteristics of individual word forms
- Grammatical segments from 2 to 5 word forms

Depending on the division of roles in the training programme, the role of Semanta, our first **digital teacher**, can be filled for a virtual **interlocutor**. Under the guidance of a human web **editor**, Semanta's tools can be used to teach the chatbot, avatar or robot, reproducible knowledge. The information can come from individual expressions, texts, text files, websites, wikipedia, bol.com, google etc. Each text is supposed to consist of an unstructured number of expressions in which questions and answers can be contained. Knowledge that can eventually be unlocked through **intelligent conversations** between the human and virtual interlocutors.