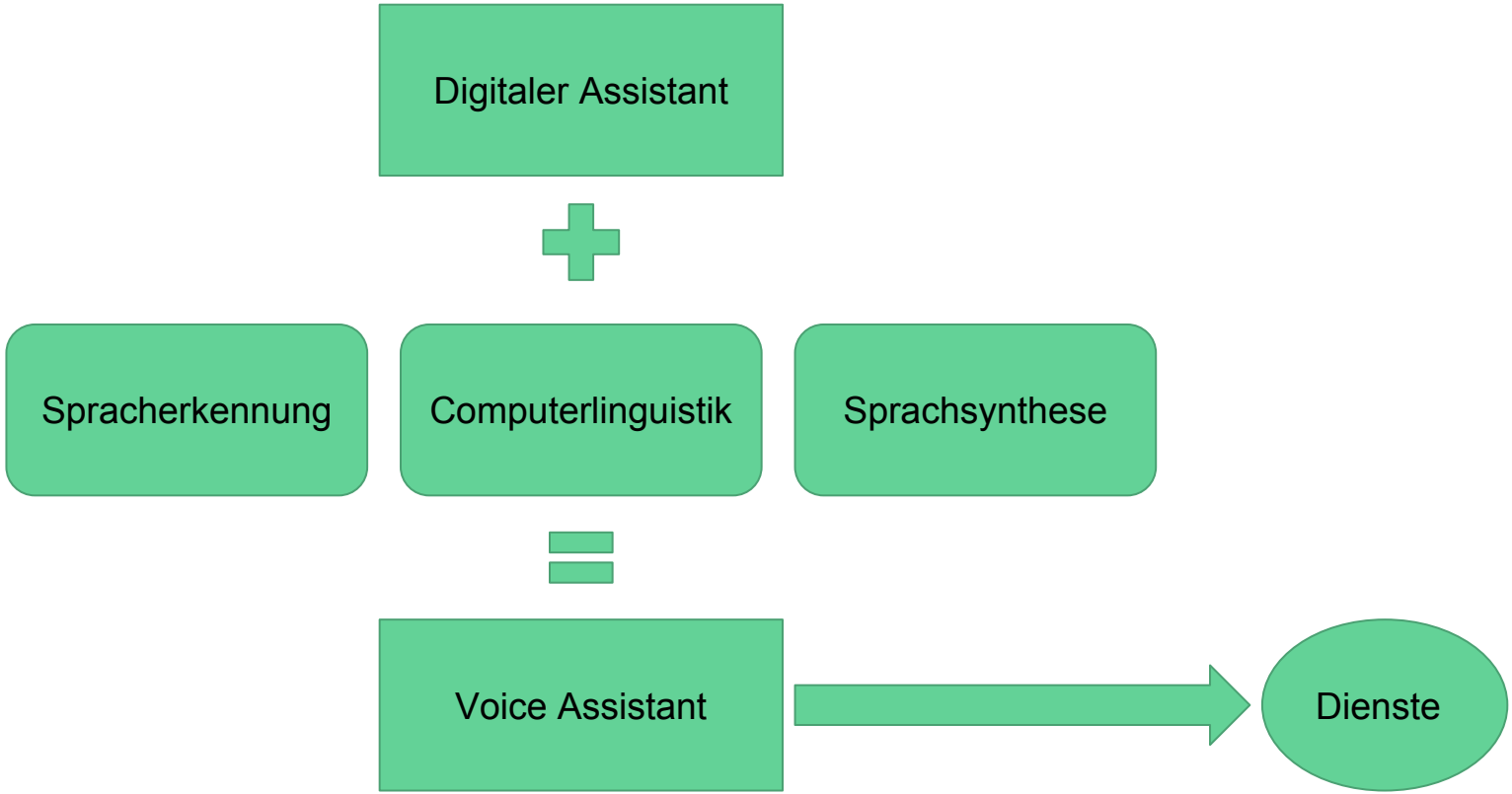
# Voice Assistants

# Inhalt

- Was ist ein Voice Assistant
- Generelle Funktionsweise eines Voice Assistants
- Fallbeispiel Alexa
- Fallbeispiel Google Home
- Gegenüberstellung Alexa und Google Home

# Was ist ein Voice Assistant?

# Was ist ein Voice Assistent



# Digitaler Assistent

- Auch Intelligente Persönliche Assistenten (IPA) genannt
- Softwarelösungen mit Spracherkennung/analyse
- Helfen bei Suche nach Informationen
- Übernehmen einfache Aufgaben
- Führen Dialog mit Anwender
  - Dialog orientiert sich immer mehr an normalem Gesprächsfluss

# Spracherkennung (SE)

- Teilgebiet der Angewandten Informatik und Computerlingustik
- Macht Automaten gesprochene Sprache zugänglich
- Hilft besonders Computern bei automatischer Datenerfassung
- Zwei Arten:
  - Sprecherunabhängig
    - Jeder Nutzer wird direkt erkannt
    - Wortschatz von einigen tausend Wörtern
  - Sprecherabhängig
    - Erfordert kurzes Training des Systems
    - Wortschatz von über 300.000 Wörtern

# Computerlinguistik (CL)

Computerlinguistik erforscht die maschinelle Verarbeitung natürlicher Sprachen. Sie erarbeitet die theoretischen Grundlagen der Darstellung, Erkennung und Erzeugung gesprochener und geschriebener Sprache durch Maschinen.

- Universität München

Erfassung von Sprache ist auf zwei Arten möglich:

- Schallinfomationen (akustisch)
- Buchstabenketten (textuell)

# Computerlinguistik (Saarbrücker Pipelinemodell)

1. Spracherkennung: Umwandlung der Schallinformation in Text, falls nötig.
2. Tokenisierung: Segmentierung der Buchstabenkette in Wörter und Sätze.
3. Morphologische Analyse: Extraktion grammatischer Informationen und Rückführung der Wörter auf Grundform.
4. Syntaktische Analyse: Analyse der einzelnen Wörter eines Satzes auf ihre strukturelle Funktion.
5. Semantische Analyse: Zuordnung der Bedeutung einzelner Sätze. Kann viele Einzelschritte enthalten.
6. Dialog- und Diskursanalyse: In Beziehung setzen aufeinander folgender Sätze.

# Sprachsynthese

- Künstliche Erzeugung der menschlichen Sprechstimme
- Ermöglicht durch Text-to-Speech (TTS) System
- Zwei Methoden:
  - Signalmodellierung
    - Erzeugung durch Zugriff auf Sprachaufnahmen einer Stimme
  - Formatsynthese
    - Vollständig digitales Erzeugen der Stimme
- Größtes Hindernis: Natürliche Sprachmelodie

# Wozu werden Voice Assistants genutzt

- Audiobooks hören
- Informationen anfragen
- Reservierungen durchführen
- Gegenstände auf eine Einkaufsliste hinzufügen
- Mathematische Berechnungen durchführen
- Musik abspielen

Momentan sind die bekanntesten VAs: Siri (Apple), Alexa (Amazon), Google Now, Google Assistant, Cortana (Microsoft)

# Entwicklungsgeschichte Voice Assistants

## AUDREY

- Erstes SE-Gerät
- 1952 entwickelt
- Erkannte einzelne Ziffern
- Lange Abstände zwischen einzelnen Ziffern

# Entwicklungsgeschichte Voice Assistants

## IBM-Shoebox

- Erstes kommerzielles SE-Gerät
- 1961 released
- Erkannte 16 Wörter:
  - die Ziffern 0-9
  - minus, plus, subtotal, total, false und of
- Führte mathematische Operationen durch

# Entwicklungsgeschichte Voice Assistants

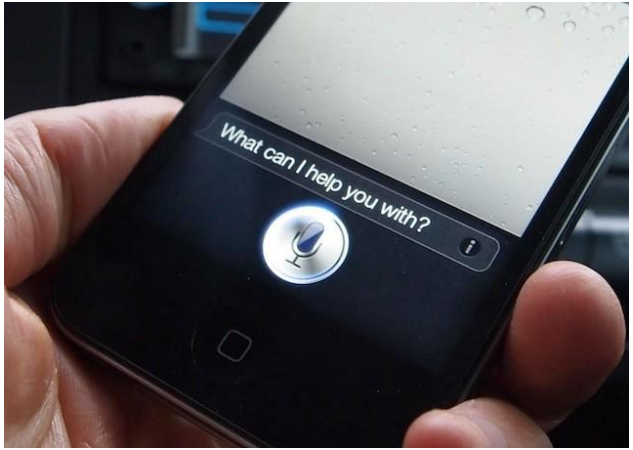
## Harpy

- 1970 released
- Erkannte bis zu 1000 Wörter
- STT Diktiergerät
- 10 Jahre später wurden ganze Sätze erkannt
  - Genutzt wurde das Hidden Markov Model
  - Hidden Markov Model berechnet die Wahrscheinlichkeit dass ein bestimmtes Wort auf ein anderes folgt

# Entwicklungsgeschichte Voice Assistants

Siri

- 4. Oktober 2011 released
- Feature des iPhone 4S
- Erster Voice Assistant
- Nutzte erstmalig Datenübertragung an einen Server um Eingaben zu verarbeiten

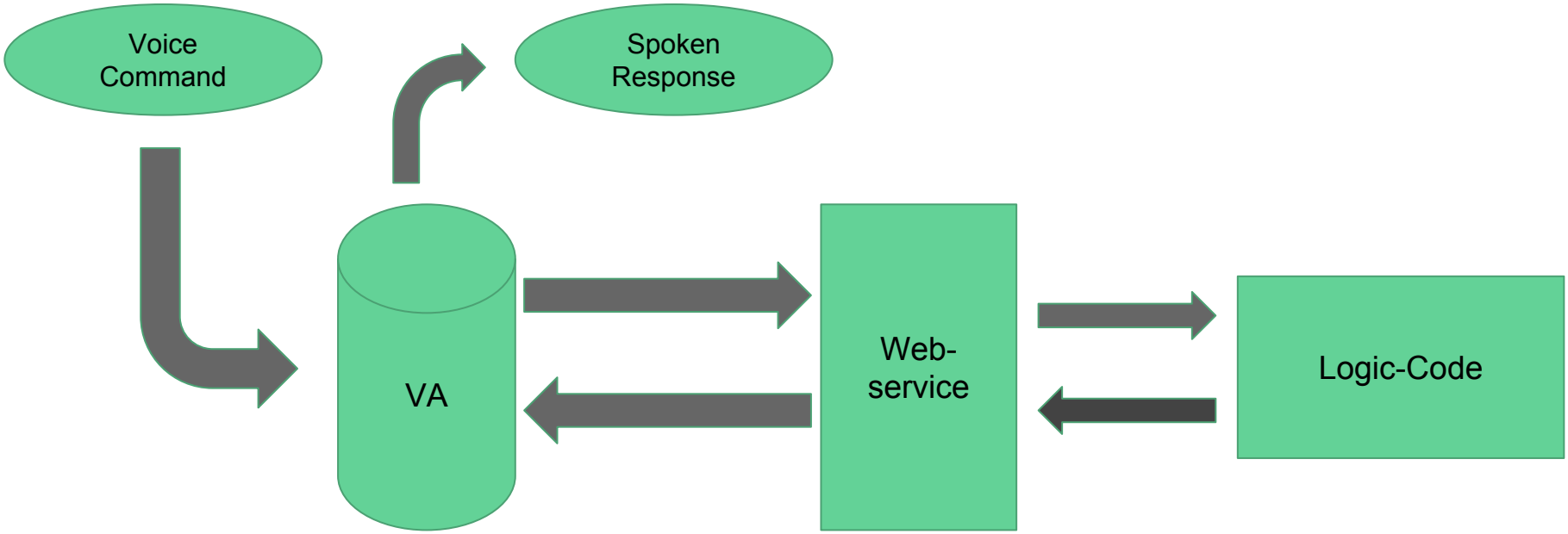


# Entwicklungsgeschichte Voice Assistants

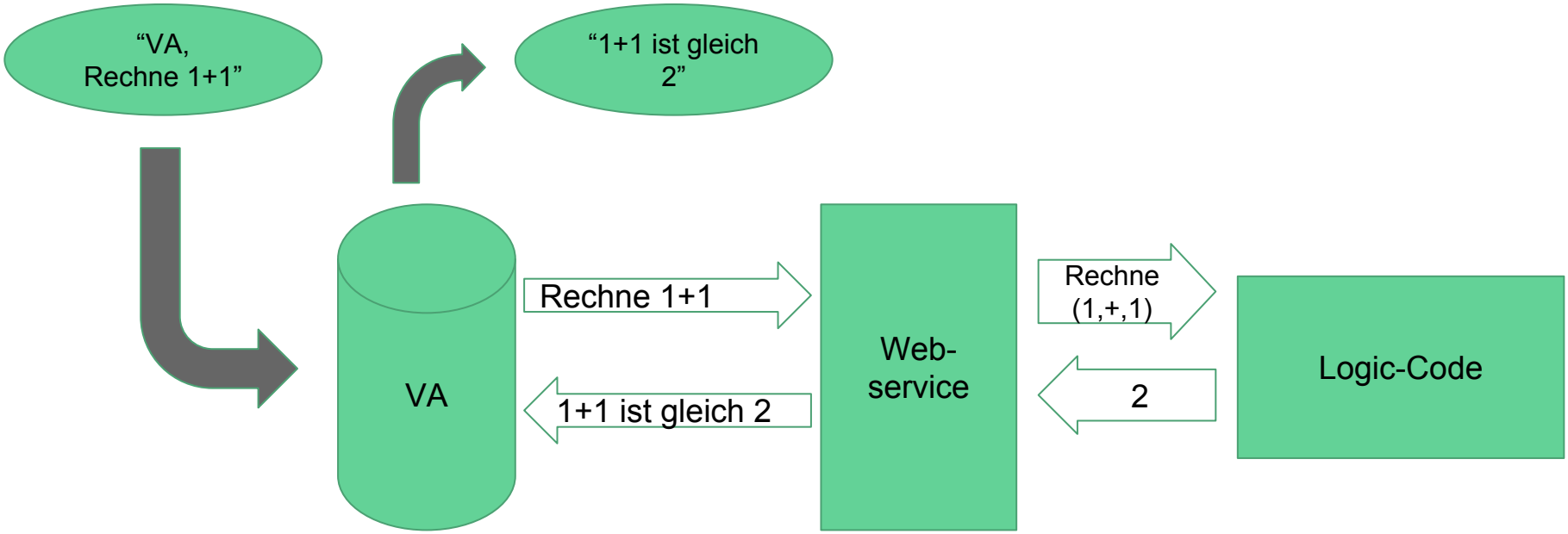
Alexa / Google Home

- 6. November, 2014  / 4. November 2016 released
- Erste eigenständige Voice Assistants
- Erweiterbarer Funktionsumfang


# Wie funktioniert ein Voice Assistant?




4. [illegible] gefunden
ggf. nötige Parameter (CL)



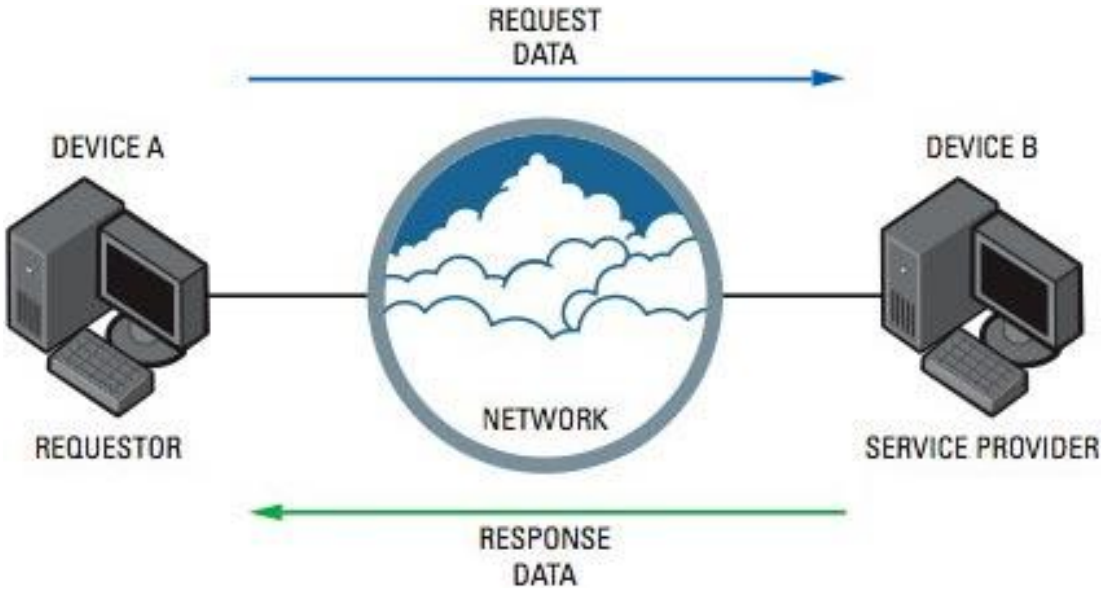

4. [illegible] gefunden
ggf. nötige Parameter (CL)

# Wie funktionieren Voice Assistenten

1. Endgerät wird durch Codewort aktiviert (SE)
2. Sprache wird auf dem Endgerät erkannt (SE)
3. Sprache wird an einen Anwendungsserver gesendet
4. Anwendungsserver erkennt den Anwendungsfall und ggf. nötige Parameter (CL)
5. Anwendungsserver verarbeitet Anfrage
6. Anwendungsserver sendet Antwort an Endgerät
7. Endgerät gibt Antwort wieder (TTS)

# Webservice

## Allgemein



# Webservice

## Amazon Alexa

- Sendet Anfragen an AWS (Amazon Web Service)
- Dieser nutzt Dienste zur Verarbeitung
- AWS Dienst Lambda führt Programmcode aus.
- AVS (Alexa Voice Service) lernt mittels Machine Learning über AWS.

## Google Home

- Webservice kann auf eigenem Server oder z.B. auf einem Server des Google Cloud Projects laufen.
- Node.js Backend führt Dialogflow aus.
- Ein Agent (Anwendung) reagiert auf die Anfrage.

# Logic Code

## Amazon Alexa

- Benutzt Utterances (Aussagen)
- Utterances verbinden die erkannte Spracheingabe mit den Intents
- Intents
  - Definiert mittels JSON
  - Programmiert in versch. Sprachen
- Intents bilden die entsprechenden Funktionen ab (mit/ohne Parameter)

## Google Home

- Actions on Google verwaltet Apps für den Google Assistant
- Durch Utterances werden Intents erkannt (unterstützt durch Machine Learning)
- Aus Utterances können Parameter/Entities herausgelesen werden
- Text Response oder Webhook

# Amazon Alexa

# Amazon Alexa

- Allgemein
- Funktionsweise
- Hardware
- Skills(Erweiterbarkeit)
- Datenschutz/Sicherheit

# Allgemein

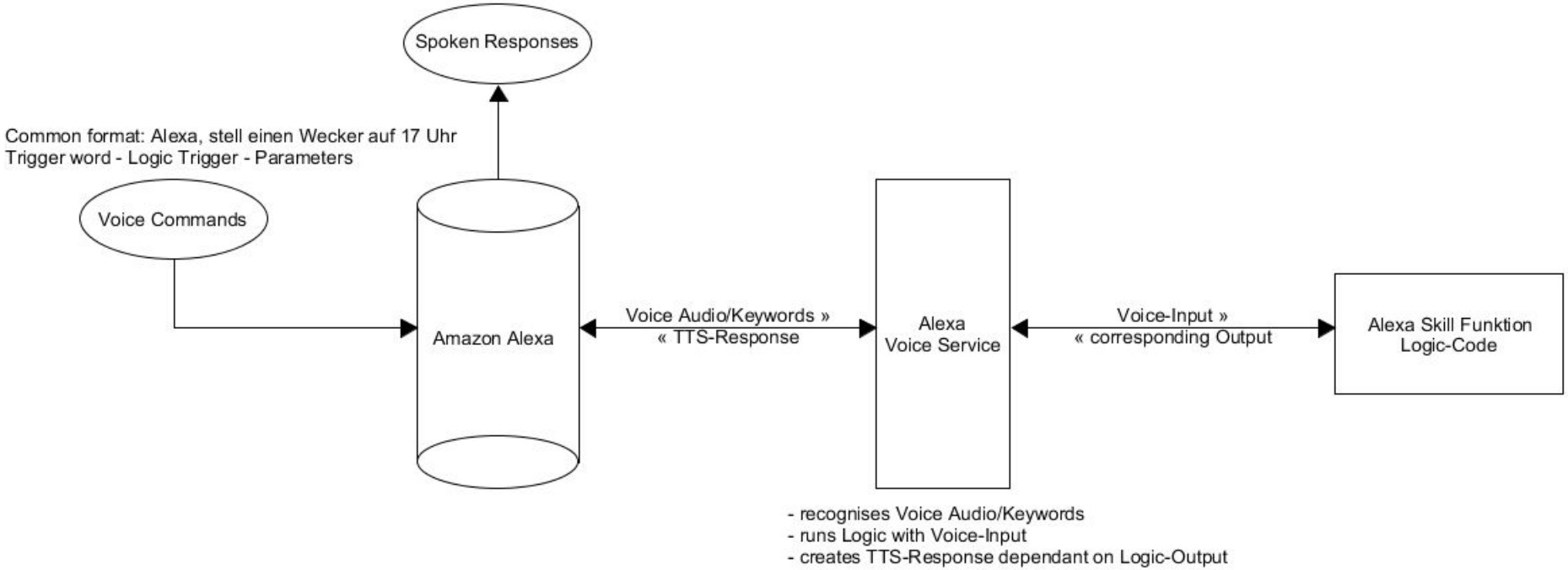
- Voice Assistant von Amazon
- Veröffentlichung 6. November, 2014 (USA)
- Veröffentlichung 26. Oktober, 2016 (Deutschland)
- Momentan veröffentlichte Geräte:
    - Echo Spot (Alexa Lautsprecher mit kleinem Display)
    - Echo Buttons (“Smart Home” Buttons)
    - Echo Connects (Verbindet Alexa mit dem Telefon für Anrufe)
    - Echo Show (7” Display - Anzeige von Videos und anderen Infos)
    - Echo , Echo Dot und Echo Plus
    - Alexa Sprachfernbedienung (FireTV)

# Funktionen/Funktionsweise

- Jede Funktion ist ein “Skill”
- Bietet vorinstallierte Skills und einen Skill Store für Drittanbieter
- Kann Musik abspielen
- Spracherkennung und TTS(Text-to-Speech)
- Reagiert auf Schlagwörter(Alexa,Echo,Amazon oder Computer)
- Stimmen werden *noch* nicht unterschieden
- Von Beginn an kann es: Musik abspielen, Wetter mitteilen, Alarm setzen uvm.

# Funktionen/Funktionsweise

## Amazon Alexa VA Architektur



# Funktionen/Funktionsweise

Typischer Ablauf eines “Gesprächs”:

1. Nutzer sagt einen Satz mit dem Keyword z.B. Alexa
2. “Alexa” wird lokal erkannt, der Rest der Anfrage wird an den Alexa Voice Service übertragen
3. Alexa Voice Service verarbeitet die Anfrage und sendet alles verarbeitet durch die Utterances weiter an die Intents des passenden Skills
4. Skill erstellt eine Antwort als Text, AVS betreibt TTS und teilt dem Nutzer die Antwort mit

# Hardware

Echo Dot

- Volle Alex Voice Service Funktionalität
- 7 Mikrofone/1 Lautsprecher
- Arbeitet als Smart Home Hub
- Hardwaretasten für die Lautstärke
- Und Mikrofon

Preis: 59,99 €

# Hardware

Echo

- Ähnlich wie Echo Dot
- Größere und bessere Lautsprecher
- 63 mm-Woofer
- 16 mm-Hochtonlautsprecher

Preis: 99,99 €

# Hardware

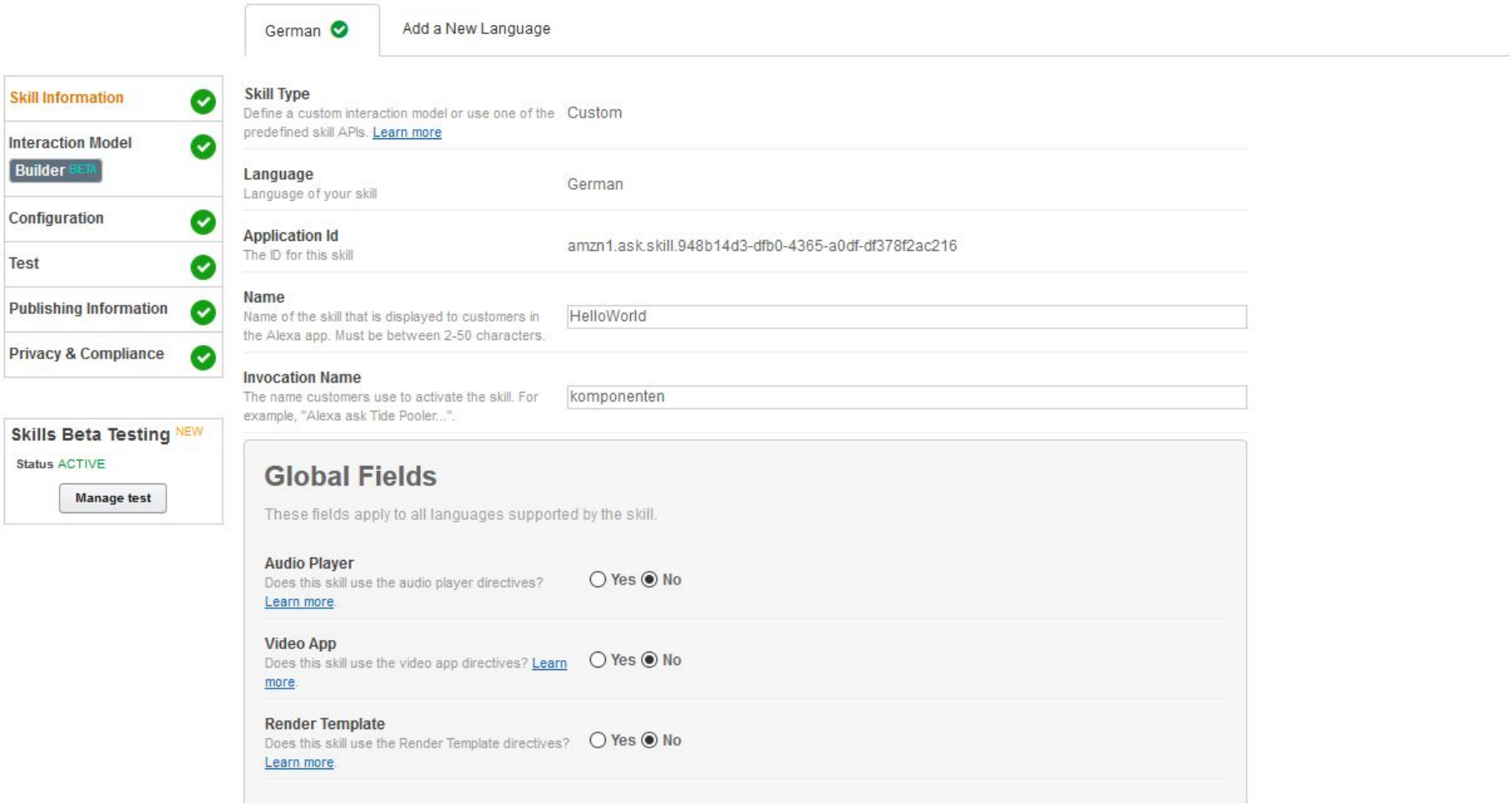
Echo Plus

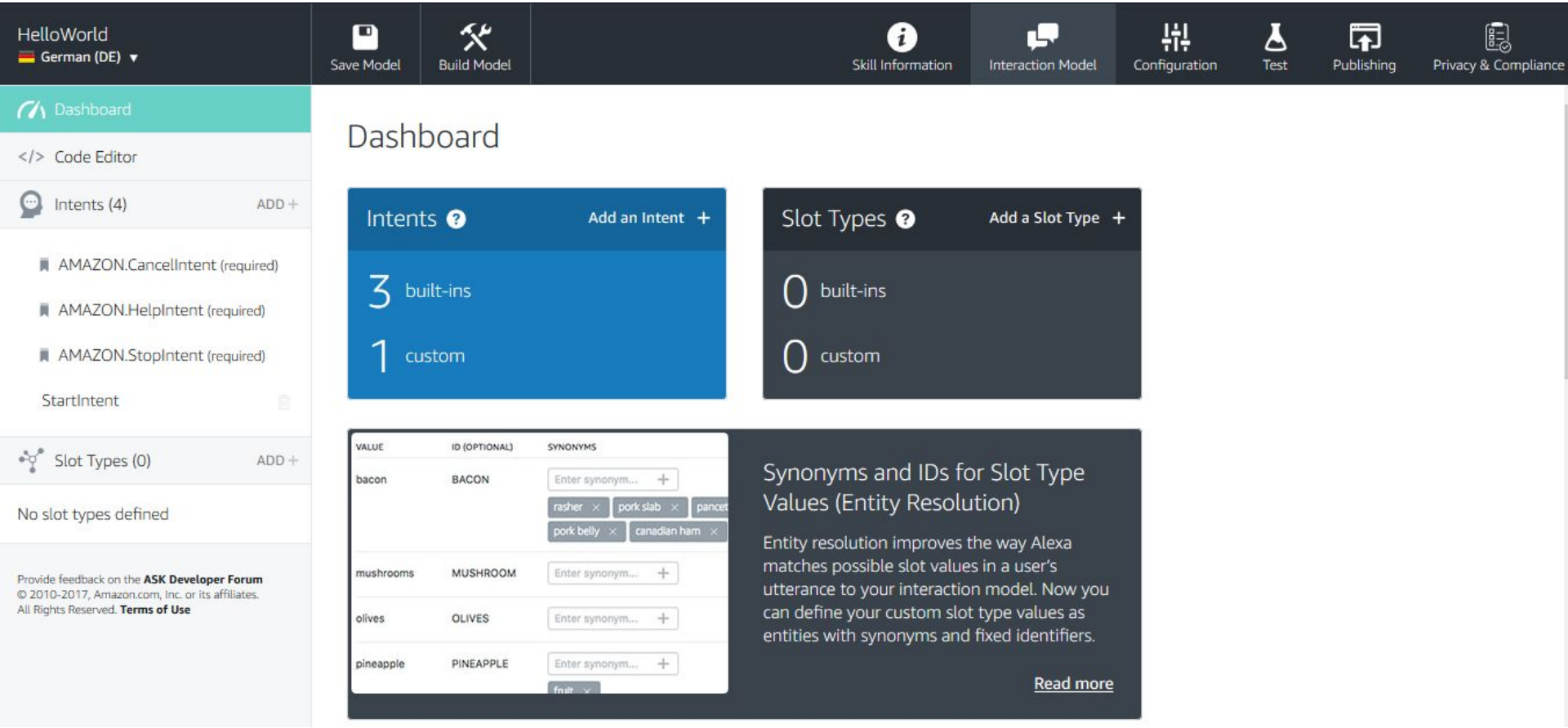
- Funktionalitäten wie Echo Dot und Echo
- integrierten ZigBee-Smart Home-Hub
- 63 mm-Woofer
- 20 mm-Hochtonlautsprecher

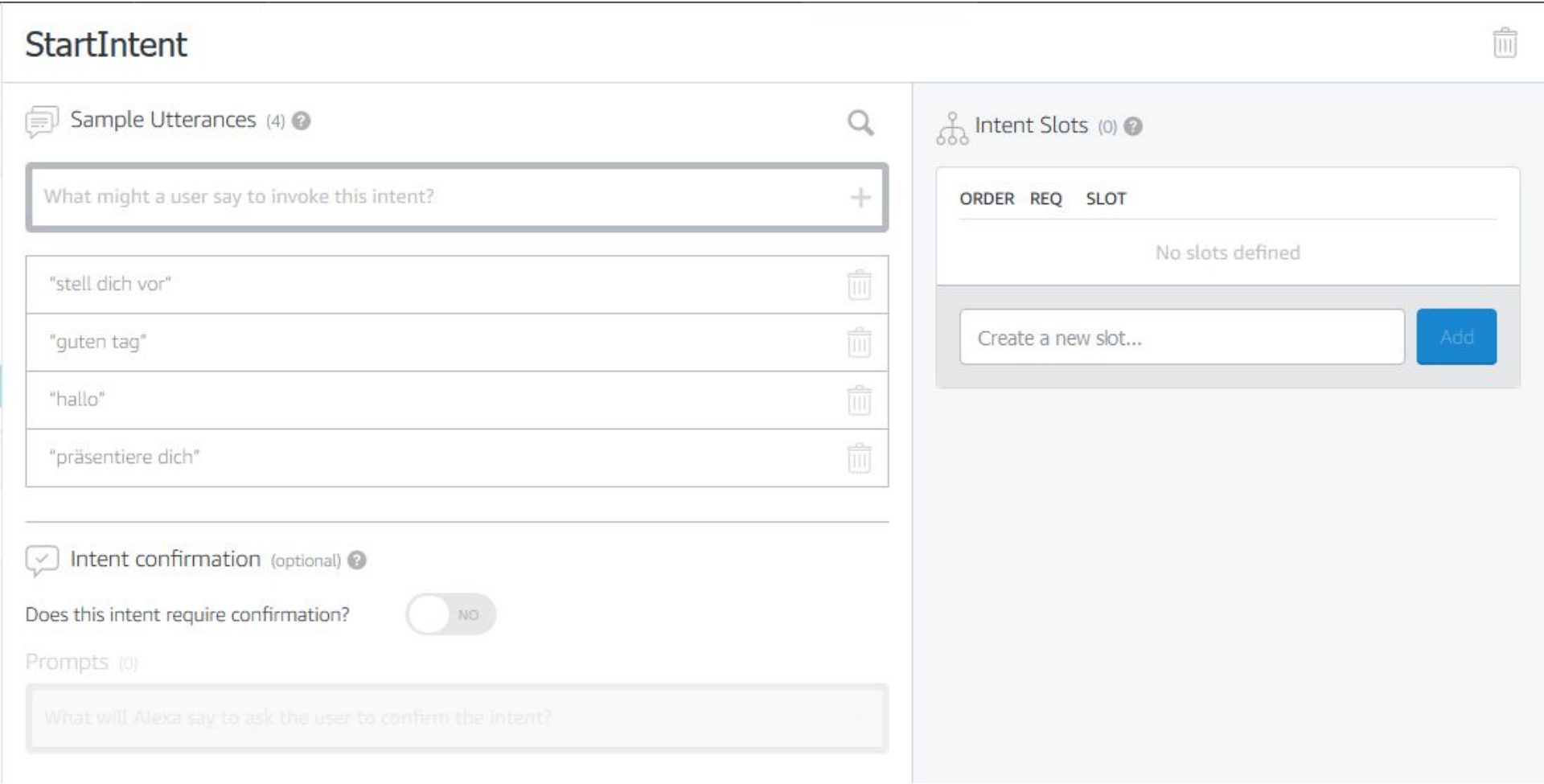
Preis: 149,99 €

# Skills(Erweiterbarkeit)

- Alexa lässt sich durch Skills erweitern
- Es gibt einen Skill-Store mit aktuell über 10 Tausend Skills
- Ein Skill besteht aus:
  - Utterances
  - Intents
  - Funktion(z.B. AWS Lambda)
- Utterances und Intents werden bei Amazon erstellt und auch bereitgestellt
- Funktion kann von einem eigenen Server oder über AWS Lambda bereitgestellt werden
- Server muss öffentlich erreichbar sein


German
Add a New Language
Skill Information
Interaction Model
Builder BETA
Configuration
Test
Publishing Information
Privacy & Compliance
Skills Beta Testing NEW
Status ACTIVE
Manage test
Skill Type
Define a custom interaction model or use one of the predefined skill APIs. Learn more
Custom
Language
Language of your skill
German
Application Id
The ID for this skill
amzn1.ask.skill.948b14d3-dfb0-4365-a0df-df378f2ac216
Name
Name of the skill that is displayed to customers in the Alexa app. Must be between 2-50 characters.
HelloWorld
Invocation Name
The name customers use to activate the skill. For example, "Alexa ask Tide Pooler...".
komponenten
Global Fields
These fields apply to all languages supported by the skill.
Audio Player
Does this skill use the audio player directives? Learn more.
Yes No
Video App
Does this skill use the video app directives? Learn more.
Yes No
Render Template
Does this skill use the Render Template directives? Learn more.
Yes No


HelloWorld
German (DE)
Save Model
Build Model
Skill Information
Interaction Model
Configuration
Test
Publishing
Privacy & Compliance
Dashboard
Code Editor
Intents (4)
ADD
AMAZON.CancelIntent (required)
AMAZON.HelpIntent (required)
AMAZON.StopIntent (required)
StartIntent
Slot Types (0)
ADD
No slot types defined
Provide feedback on the ASK Developer Forum
© 2010-2017, Amazon.com, Inc. or its affiliates.
All Rights Reserved. Terms of Use
Dashboard
Intents
Add an Intent
3 built-ins
1 custom
Slot Types
Add a Slot Type
0 built-ins
0 custom
VALUE
ID (OPTIONAL)
SYNONYMS
bacon
BACON
Enter synonym...
rasher
pork slab
pancet
pork belly
canadian ham
mushrooms
MUSHROOM
olives
OLIVES
pineapple
PINEAPPLE
fruit
Synonyms and IDs for Slot Type Values (Entity Resolution)
Entity resolution improves the way Alexa matches possible slot values in a user's utterance to your interaction model. Now you can define your custom slot type values as entities with synonyms and fixed identifiers.
Read more


StartIntent
Sample Utterances (4)
What might a user say to invoke this intent?
"stell dich vor"
"guten tag"
"hallo"
"präsentiere dich"
Intent confirmation (optional)
Does this intent require confirmation?
NO
Prompts (0)
What will Alexa say to ask the user to confirm the intent?
Intent Slots (0)
ORDER REQ SLOT
No slots defined
Create a new slot...
Add

```json
{
  "languageModel": {
    "intents": [
      {
        "name": "AMAZON.CancelIntent",
        "samples": []
      },
      {
        "name": "AMAZON.HelpIntent",
        "samples": []
      },
      {
        "name": "AMAZON.StopIntent",
        "samples": []
      },
      {
        "name": "StartIntent",
        "samples": [
          "präsentiere dich",
          "hallo",
          "guten tag",
          "stell dich vor"
        ],
        "slots": []
      }
    ],
    "invocationName": "komponenten"
  }
}
```

# HelloWorldAlexa

Qualifiers ▼ Actions ▼ Select a test event.. ▼ Test Save

File Edit Find View Goto Tools Window

Environment

- HelloWorldAlexa
  - node_modules
  - index.js
  - package.json

index.js

```
107
108 const handlers = {
109     'LaunchRequest': function () {
110         this.emit('GetFact');
111     },
112     'StartIntent': function () {
113         this.emit('GetFact');
114     },
115     'GetFact': function () {
116         // Get a random space fact from the space facts list
117         // Use this.t() to get corresponding language data
118         const factArr = this.t('FACTS');
119         const factIndex = Math.floor(Math.random() * factArr.length);
120         const randomFact = factArr[factIndex];
121
122         // Create speech output
123         const speechOutput = this.t('GET_FACT_MESSAGE') + randomFact;
124         this.emit(':tellWithCard', speechOutput, this.t('SKILL_NAME'), randomFact);
125     },
126     'AMAZON.HelpIntent': function () {
127         const speechOutput = this.t('HELP_MESSAGE');
128         const reprompt = this.t('HELP_MESSAGE');
129         this.emit(':ask', speechOutput, reprompt);
130     },
131     'AMAZON.CancelIntent': function () {
132         this.emit(':tell', this.t('STOP_MESSAGE'));
133     },
134     'AMAZON.StopIntent': function () {
135         this.emit(':tell', this.t('STOP_MESSAGE'));
```

# Google Home

# Google Home

- Allgemein
- Funktionen/Funktionsweise
- Hardware
- Erweiterbarkeit
- Actions

# Allgemein

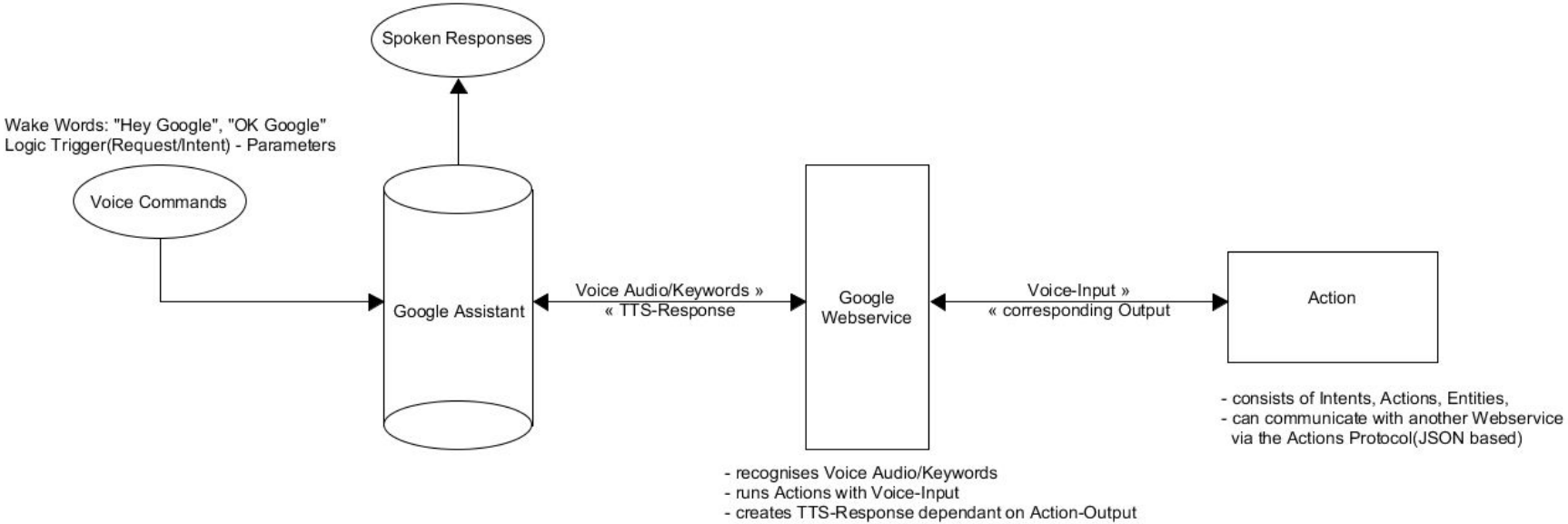
- Von Google entwickelter Smart Home Lautsprecher
- Persönlicher Voice Assistant
- Veröffentlichung USA: 4. November 2016
- Markteinführung Deutschland: 8. August 2017 mit deutscher Stimme
- Momentan 3 verschiedene Google Home Geräte
  (Home, Home Mini, Home Max)

# Funktionen/Funktionsweise

- Beinhaltet eine Reihe von Funktionen von Google und Drittanbietern
- Mehrere Google Home Geräte können zum synchronen Abspielen von Musik verwendet werden
- Bis zu 6 Benutzer können an der Stimme erkannt und unterschieden werden
- Weckworte: “*Hey Google*” oder “*OK Google*”
- Integrierten *Google Assistant* gibt es auch auf anderen Geräten (z.B. Smartphones)
- Allgemeine Funktionen werden erweitert durch *Actions on Google*
- Interaktion mit dem Assistenten über Konversationen (möglichst natürlicher Austausch)

# Funktionen/Funktionsweise

## Google Voice Assistant Architektur



# Funktionen/Funktionsweise

Typischer Ablauf:

1. Nutzer fordert Assistenten auf eine gewisse Aktion auszuführen
2. Assistent sendet Anfrage an Actions on Google
3. Actions on Google gibt passende App zurück
4. Assistent fragt den Nutzer ob er die App ausführen will
5. Nutzer bestätigt mit “Ja”
6. Assistent stellt die App vor
7. Assistent übergibt die Konversation an die App

# Hardware

## Google Home

- 149€
- Farbige Status-LEDs an der Oberfläche
- Kapazitiver Berührungssensor zum Starten und Stoppen von Musik oder anpassen der Lautstärke
- Mute-Button für Mikrofon an der Rückseite

# Hardware

## Google Home Mini

- 59€
- Gleiche Fuktionalität
- Berührungssensor zum Anpassen der Lautstärke
- Weiße Statuslichter scheinen durch den Stoff an der Oberseite
- Mute-Schalter für Mikrofon an der Rückseite

# Hardware

## Google Home Max

- 400$
- Stereo Lautsprecher (einschließlich zweier Hochtöner und Subwoofer)
- Magnetisch befestigbarer Ständer für vertikale Ausrichtung
- Beinhaltet *Smart Sound* (nutzt machine learning um sich an die Umgebung anzupassen)

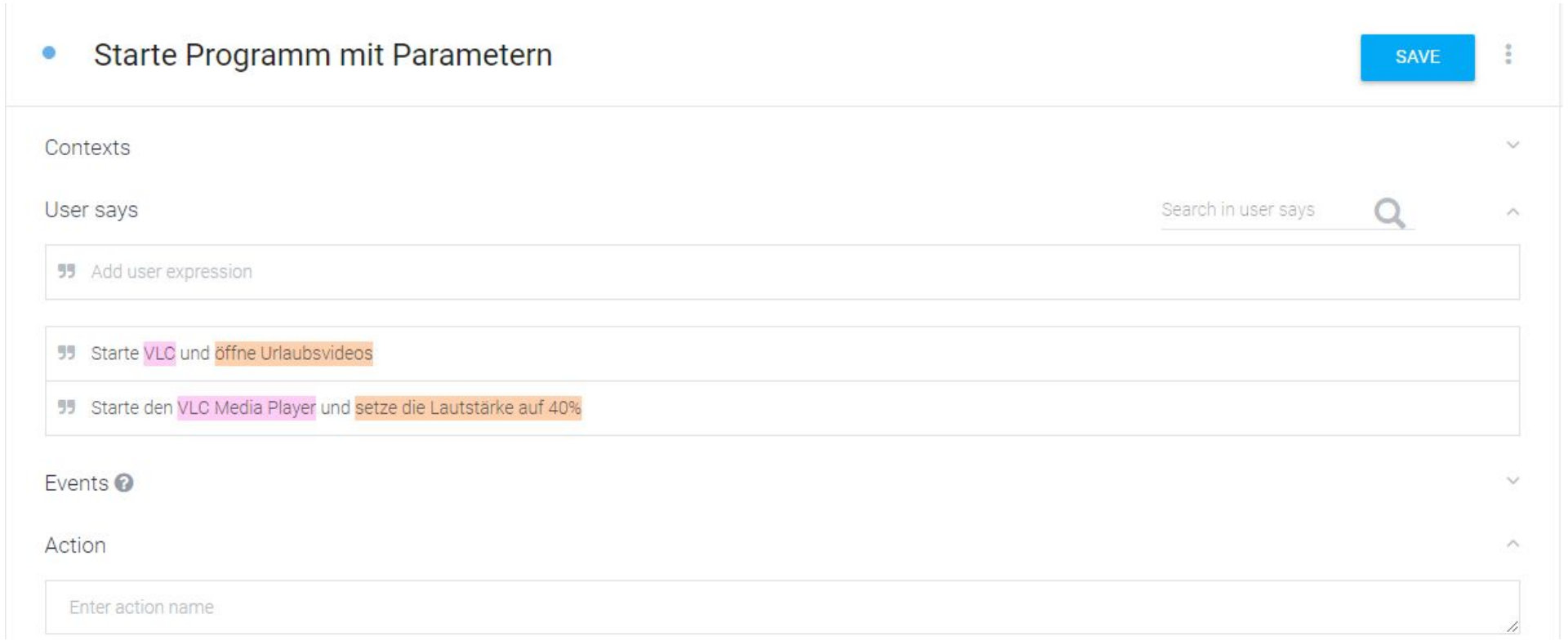
# Erweiterbarkeit

- Google bietet verschiedene Optionen zum Entwickeln von Apps
- Templates
  - Beinhalteten vorgefertigte Konversationen
- Actions SDK
  - Bietet keine NLU (natural language understanding)
  - Für simple Aktionen mit geringer Input Varianz
- Dialogflow
  - Funktionalität der Actions SDK in einfach zu nutzender Web IDE
  - Beinhaltet NLU engine

*Im folgenden wird der Aufbau der Actions in Dialogflow beleuchtet*

# Actions

- Intents
  - Sind Hauptbestandteil der Actions
  - Werden aus Utterances (Äußerungen des Nutzers) bestimmt
  - Können mehrere verschiedene Parameter enthalten, die als Entities von der NLU erkannt werden
- Entities
  - Kategorien in denen Objekte eines Typs vordefiniert werden können
  - Erleichtern die Erkennung von relevanten Daten
  - Einzelne Objekte können mit Synonymen angegeben werden
- Fulfillment
  - Neben einfacher Text Response kann das erkannte Intent an einen Web service gesendet werden um eine Response zu erhalten


Starte Programm mit Parametern
SAVE
Contexts
User says
Search in user says
Add user expression
Starte VLC und öffne Urlaubsvideos
Starte den VLC Media Player und setze die Lautstärke auf 40%
Events
Action
Enter action name

| REQUIRED | PARAMETER NAME | ENTITY | VALUE | IS LIST |
|---|---|---|---|---|
| ☐ | Programm | @Programm | $Programm | ☐ |
| ☐ | zusaetzliche_Parameter | @sys.any | $zusaetzliche_Parameter | ☐ |
| ☐ | Enter name | Enter entity | Enter value | ☐ |

+ New parameter

Response

DEFAULT +

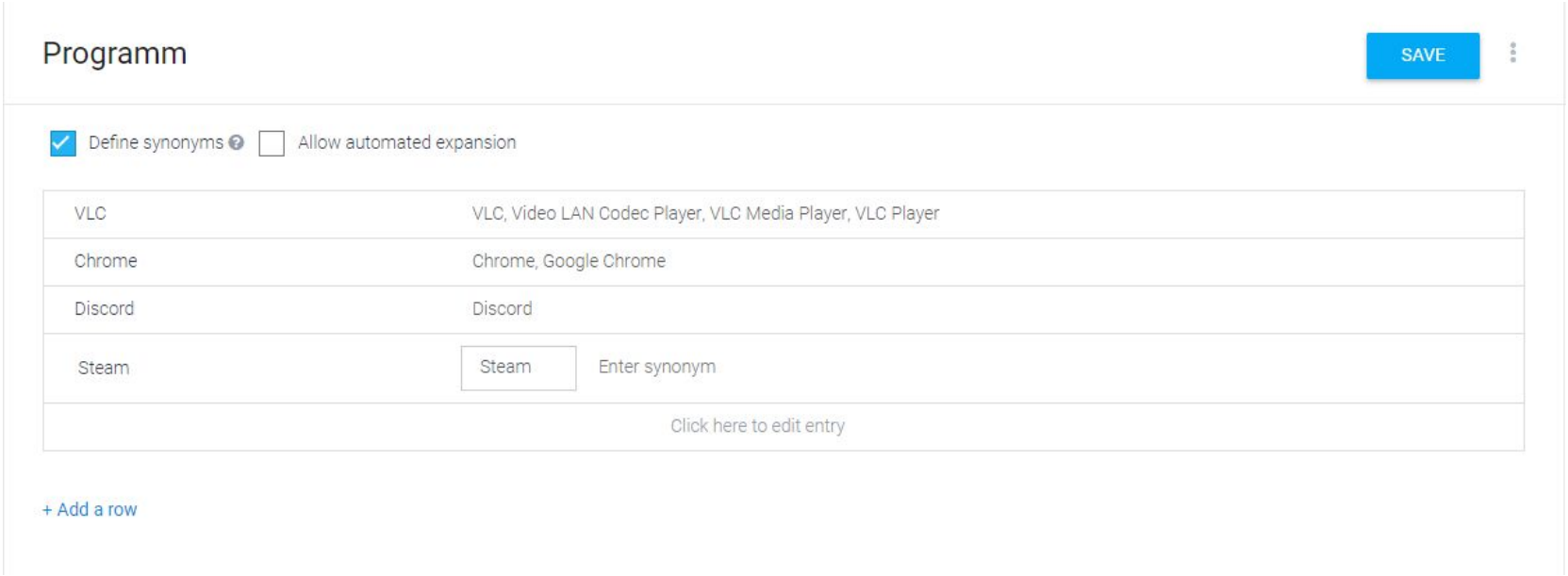
Text response

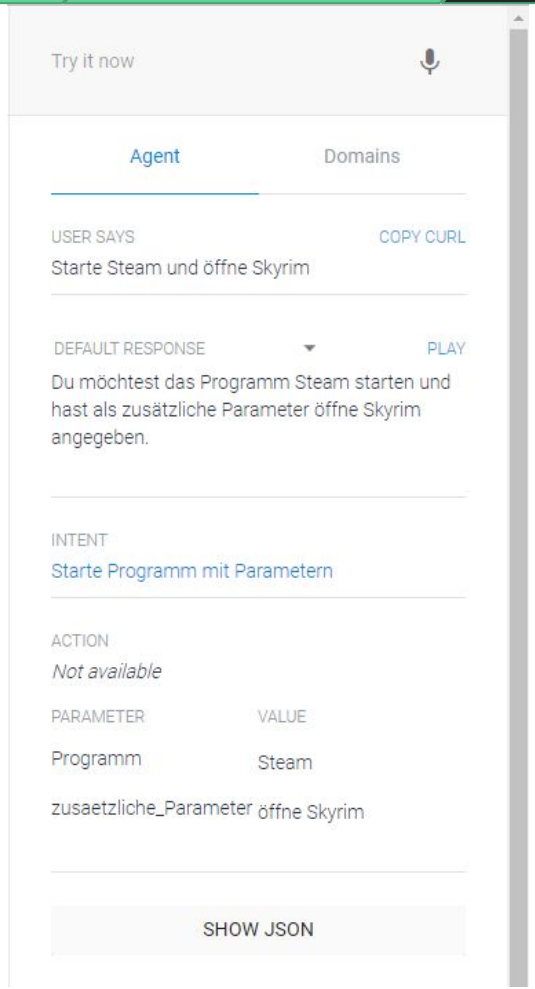
1 Du möchtest das Programm $Programm starten und hast als zusätzliche Parameter $zusaetzliche_Parameter angegeben.

2 Enter a text response variant

ADD MESSAGE CONTENT

Fulfillment


Programm
SAVE
Define synonyms
Allow automated expansion
VLC
VLC, Video LAN Codec Player, VLC Media Player, VLC Player
Chrome
Chrome, Google Chrome
Discord
Discord
Steam
Steam
Enter synonym
Click here to edit entry
+ Add a row


Try it now
Agent
Domains
USER SAYS
COPY CURL
Starte Steam und öffne Skyrim
DEFAULT RESPONSE
PLAY
Du möchtest das Programm Steam starten und hast als zusätzliche Parameter öffne Skyrim angegeben.
INTENT
Starte Programm mit Parametern
ACTION
Not available
PARAMETER
VALUE
Programm
Steam
zusaetzliche_Parameter
öffne Skyrim
SHOW JSON

JSON


```
{
  "id": "3fe48c5d-b29d-4f29-b87d-dedae8fc7532",
  "timestamp": "2017-12-15T21:44:57.781Z",
  "lang": "de",
  "result": {
    "source": "agent",
    "resolvedQuery": "Starte Steam und öffne Skyrim",
    "action": "",
    "actionIncomplete": false,
    "parameters": {
      "Programm": "Steam",
      "zusaetzliche_Parameter": "öffne Skyrim"
    },
    "contexts": [],
    "metadata": {
      "intentId": "b1575a12-89b9-4e5f-a3e0-18f1e5aa6390",
      "webhookUsed": "false",
      "webhookForSlotFillingUsed": "false",
      "intentName": "Starte Programm mit Parametern"
    },
    "fulfillment": {
      "speech": "Du möchtest das Programm Steam starten und hast als zusätzliche Parameter öffne
        Skyrim angegeben.",
      "messages": [
        {
          "type": 0,
          "speech": "Du möchtest das Programm Steam starten und hast als zusätzliche Parameter
            öffne Skyrim angegeben."
        }
      ]
    },
    "score": 1
  },
  "status": {
    "code": 200,
    "errorType": "success",
    "webhookTimedOut": false
  },
  "sessionId": "dfa19fe3-5f7b-4705-8dc6-4c4c2847754e"
}
```


CLOSE COPY

# Gegenüberstellung
# A. Alexa / G. Home

| Alexa | Google Home |
|---|---|
| Wahrscheinlichkeitsbasiertes Matching | KI hilft bei Ergänzung der Utterances |
| “Alexa”,”Echo”,”Computer”,”Amazon” | “Hey Google”, “OK Google” |
| Einkaufen direkt per Sprachbefehl | Google Suche, andere Google Dienste |
| Über 10.000 Skills(Erweiterungen) | Bislang wenige Actions |
| Musik in Planung, Sprache mittels ESP-System (Echo Spatial Perception) | synchronisiertes Abspielen von Musik |

- Google Home und Alexa haben viele gleiche Features
- Kein direkter und aussagekräftiger Vergleich möglich/nötig