

# Lecture 7

## Intelligent Agents

COMP 474/6741, Winter 2024

### Introduction

- Intelligent Conversational Agents
- Examples
- Design process

### Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

- Question Classification
- NLP Pipelines
- Designing the Knowledge Base
- Linking Documents and Knowledge Graphs

### Conclusions

- Example: Watson
- Summary

### Notes and Further Reading

René Witte  
Department of Computer Science  
and Software Engineering  
Concordia University

## 1 Introduction

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

## 2 Search-Based Bots

### Search-Based Bots

Pattern-Matching Bots

## 3 Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## 4 Grounding-Based Bots

### Conclusions

Example: Watson  
Summary

## 5 Conclusions

Notes and Further Reading

## 6 Notes and Further Reading

# What Air Canada Lost In ‘Remarkable’ Lying AI Chatbot Case

Marisa Garcia Senior Contributor @

I offer an insider's view of the business of flight.

[Follow](#)

Feb 19, 2024, 06:03am EST



## Introduction

- Intelligent Conversational Agents
- Examples
- Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

- Question Classification
- NLP Pipelines
- Designing the Knowledge Base
- Linking Documents and Knowledge Graphs

## Conclusions

- Example: Watson
- Summary

## Notes and Further Reading

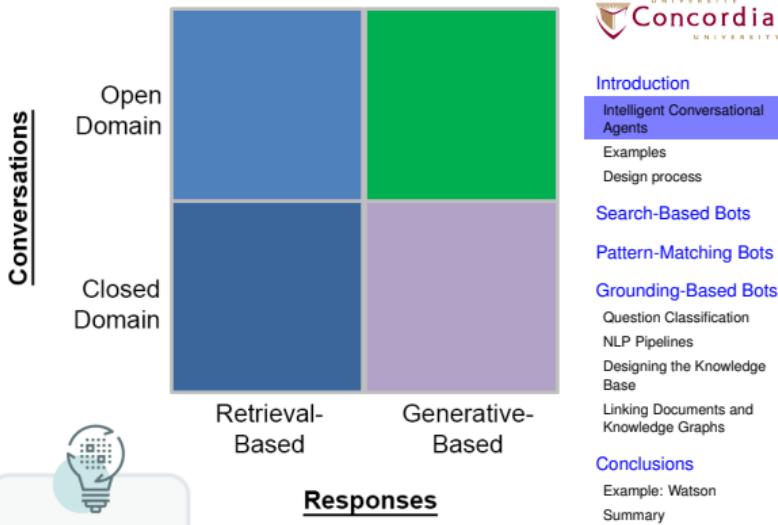
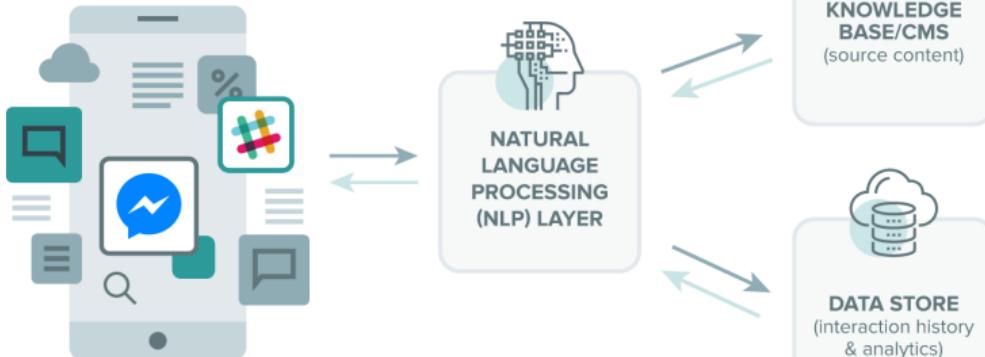
# Intelligent Conversational Agents

a.k.a. Dialog System or Dialog Engine

René Witte



- A **software program** that can interpret and respond to statements made by users in a **natural language**
- Different types of chatbots
  - Generic vs. Goal-oriented
  - Retrieval vs. Generative (Deep Learning)
- Similar architecture, different stacks



Introduction

Intelligent Conversational Agents

Examples

Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification  
NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

Conclusions

Example: Watson Summary

Notes and Further Reading

# Apple's "Knowledge Navigator" Vision (1987)

René Witte



Apple Knowledge Navigator Video (1987)



<https://www.youtube.com/watch?v=umJsITGzXd0>

## Introduction

Intelligent Conversational Agents

Examples

Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson

Summary

## Notes and Further Reading

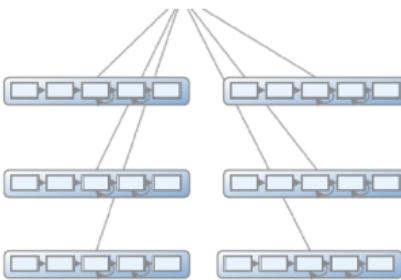
## Web Services Directory

| All                     | Viewing 1 to 1377 of 1377 APIs      |
|-------------------------|-------------------------------------|
| Advertising (15)        |                                     |
| Answers (5)             |                                     |
| Blog Search (7)         |                                     |
| Blogging (21)           |                                     |
| Bookmarks (16)          |                                     |
| Calendar (5)            |                                     |
| Chat (13)               |                                     |
| Database (12)           |                                     |
| Email (30)              |                                     |
| Enterprise (40)         |                                     |
| Events (14)             |                                     |
| Fax (2)                 |                                     |
| Feeds (12)              |                                     |
| File Sharing (6)        |                                     |
| Financial (77)          |                                     |
| Food (3)                |                                     |
| Games (21)              |                                     |
| Government (30)         |                                     |
| Internet (90)           |                                     |
|                         |                                     |
| API                     | Description                         |
| Google Maps             | Mapping services                    |
| Flickr                  | Photo sharing service               |
| YouTube                 | Video sharing and search            |
| Amazon eCommerce        | Online retailer                     |
| Twitter                 | Microblogging service               |
| eBay                    | Online auction marketplace          |
| Microsoft Virtual Earth | Mapping services                    |
| del.icio.us             | Social bookmarking                  |
| Google Search           | Search services                     |
| Yahoo Maps              | Mapping services                    |
| Yelp                    | Local user reviews and city guides  |
| hostip.info             | IP lookup                           |
| Netvibes                | Personalized home page with widgets |
| PayPal                  | Online payments                     |
| Rhapsody                | Online music services               |
| WeatherBug              | Weather forecast services           |



## Web Services and APIs

<https://tomgruber.org/writing/semtech09.htm>



## Domain & Task Models



## Guided Dialog



### Introduction

Intelligent Conversational Agents

Examples  
Design process

### Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

### Notes and Further Reading

# Siri Presentation

René Witte



A screenshot of a Vimeo video player. The video frame shows a man with a beard and short hair, wearing a dark green striped shirt, speaking. He is standing in front of a dark background with a vertical yellow light strip on the right. The video player interface includes a play button, a progress bar showing 43:18, and a control bar with volume and settings icons. Above the video, the Vimeo logo and navigation links (Product, Solutions, Watch, Pricing) are visible. To the right of the video are search, login, and new video buttons. On the far right of the player are five interaction icons: a heart, a clock, a document, and a paper airplane. Below the video, the title "KEYNOTE: The Game Changer: Siri, a Virtual Personal Assistant" is displayed, along with a timestamp of "11 years ago | More". A "Semantic Web" channel badge and a "Follow" button are also present. To the right of the video, there is a "Related Videos" section with a thumbnail for another video titled "KEYNOTE: Th... Semantic Web".

KEYNOTE: The Game Changer: Siri, a Virtual Personal Assistant

11 years ago | More

Semantic Web Follow

Related Videos

Autoplay next video

 KEYNOTE: Th... Semantic Web



## Introduction

Intelligent Conversational Agents

Examples

Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson

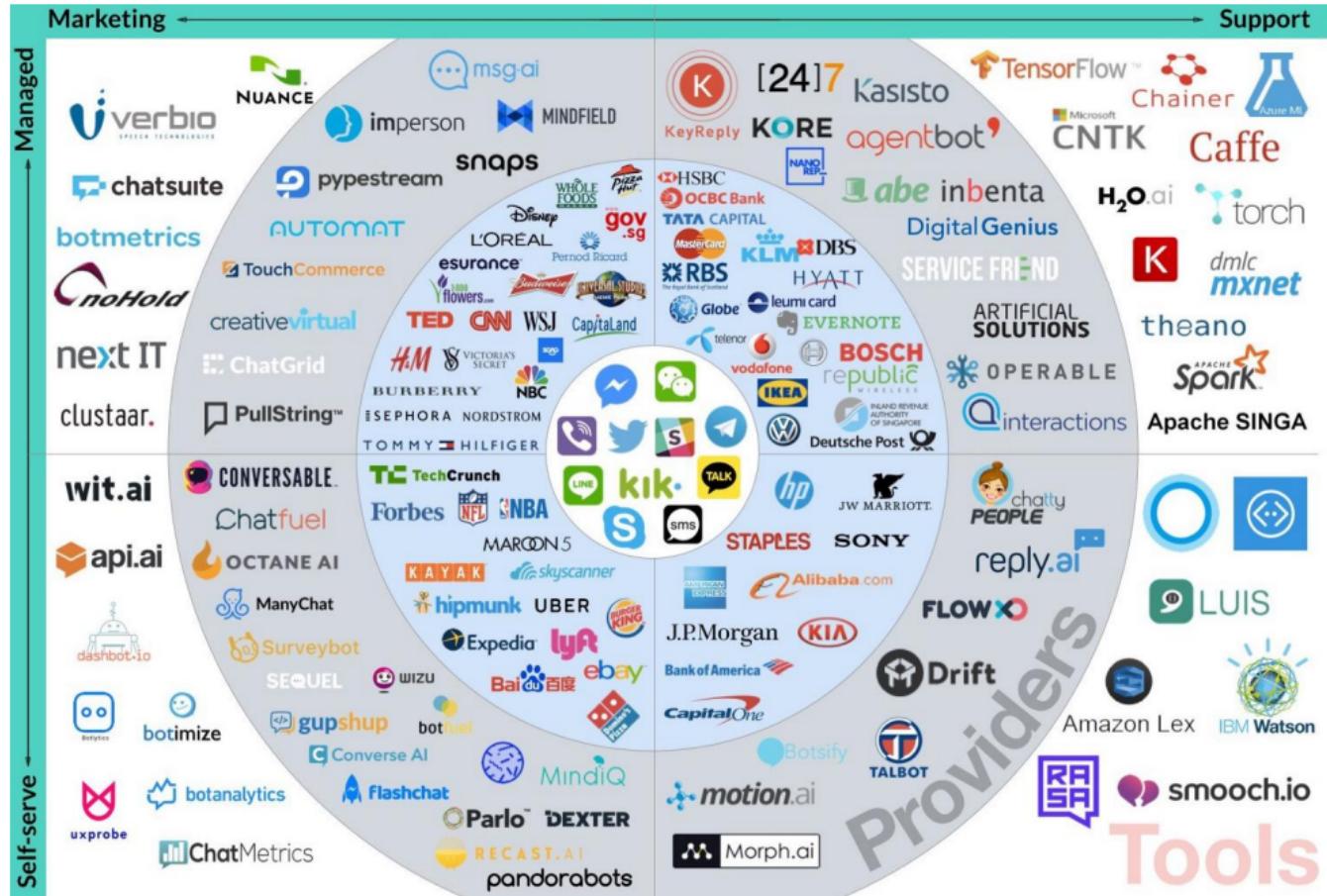
Summary

## Notes and Further Reading

<https://vimeo.com/5424527>

# The Chatbots Landscape (2017)

René Witte



## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

## Modern approaches

Pattern matching: Regex matching and response templates (canned responses)

Grounding: Knowledge graphs and inference on those graphs

Search: Text retrieval

Generative: Statistics and machine learning

## Hybrid approaches

Using multiple (or all four) techniques in one bot

### Introduction

Intelligent Conversational Agents

Examples

Design process

### Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson

Summary

### Notes and Further Reading

## Examples

**Question answering:** Google Search, Alexa, Siri, Watson

**Virtual assistants:** Google Assistant, Alexa, Siri, MS paperclip

**Conversational:** Google Assistant, Google Smart Reply, Mitsuki Bot

**Marketing:** Twitter bots, blogger bots, Facebook bots, Google Search, Google Assistant, Alexa, Allo

**Customer service:** Storefront bots, technical support bots

**Community management:** Bonusly, Slackbot

**Therapy:** Woebot, Wysa, YourDost, Siri, Allo

### Introduction

Intelligent Conversational Agents

Examples

Design process

### Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

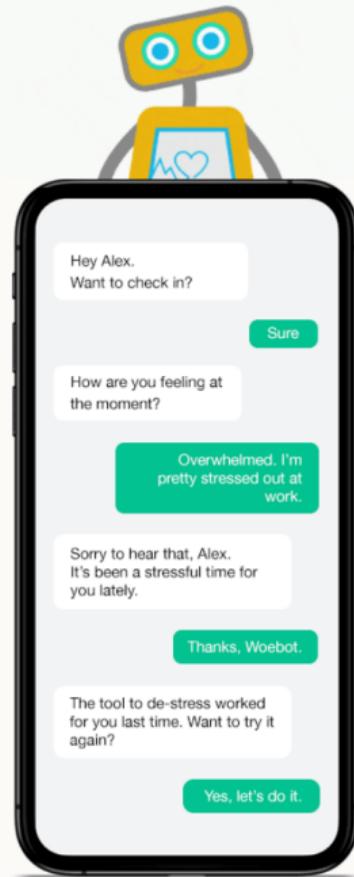
### Conclusions

Example: Watson

Summary

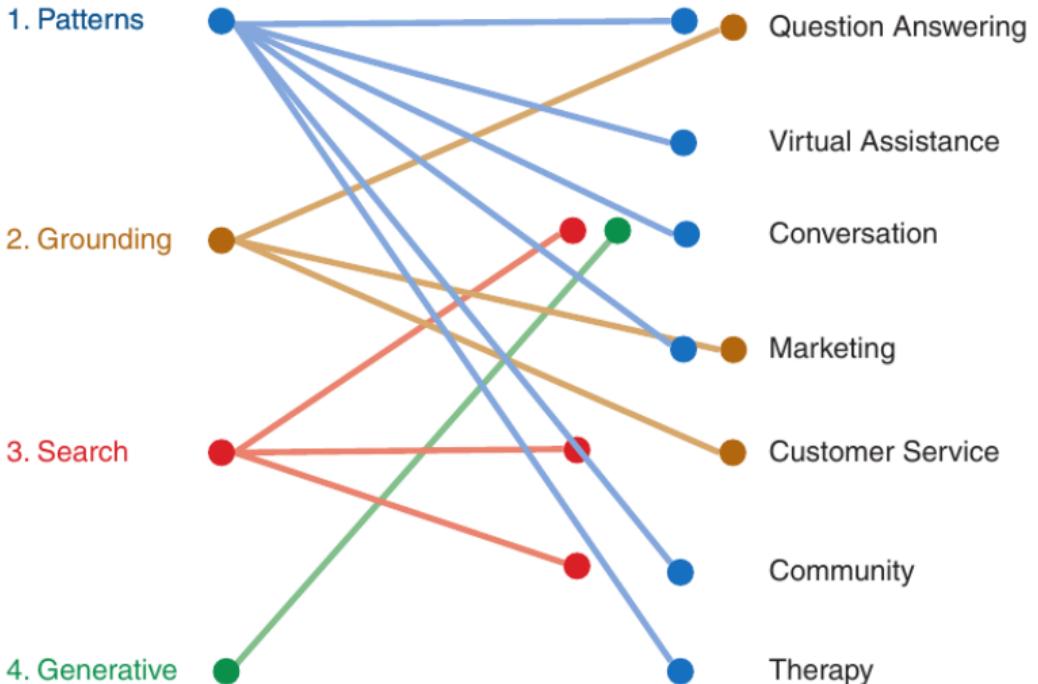
### Notes and Further Reading

## Tiny conversations to feel your best



# Chatbot techniques used for some example applications

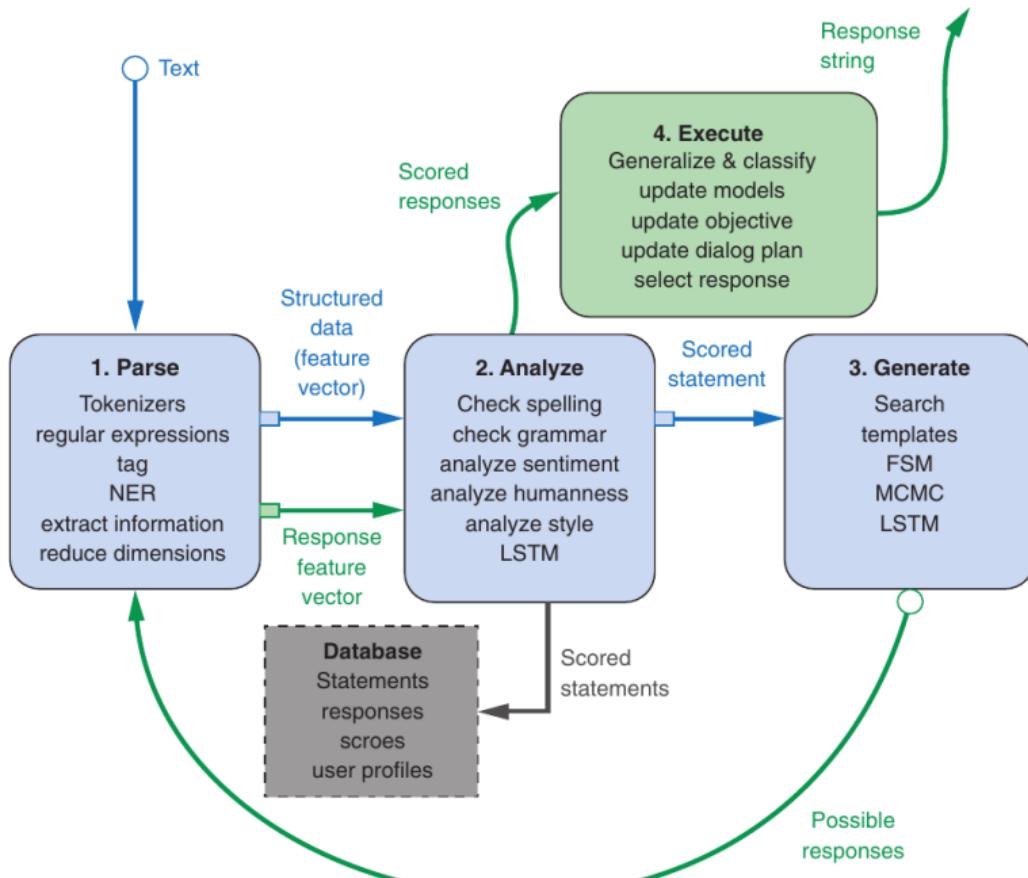
René Witte



- Introduction**
- Intelligent Conversational Agents
- Examples
- Design process
- Search-Based Bots
- Pattern-Matching Bots
- Grounding-Based Bots
- Question Classification
- NLP Pipelines
- Designing the Knowledge Base
- Linking Documents and Knowledge Graphs
- Conclusions
- Example: Watson Summary
- Notes and Further Reading

# Chatbot recirculating (recurrent) pipeline

René Witte



## Introduction

Intelligent Conversational Agents

Examples

Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson

Summary

## Notes and Further Reading

## Bots promoting movies, TV shows, video games, ...

- HBO promoted “Westworld” with “Aeden”
- Sony promoted “Resident Evil” with “Red Queen”
- Disney promoted “Zootopia” with “Officer Judy Hopps”
- Universal promoted “Unfriended” with “Laura Barnes”
- Activision promoted “Call of Duty” with “Lt. Reyes”

Call of Duty: Infinite Warfare's first victory: 6M bot messages on Facebook

MIKE MINOTTI @TOLKOTO MAY 3, 2016 11:40 AM



The leader of the Settlement Defense Front in Call of Duty: Infinite Warfare.

Image Credit: Activision

### GAMESBEAT SUMMIT

The most intimate gaming event of the year

Los Angeles

April 28 - 29

Register Today

<https://venturebeat.com/2016/05/03/call-of-duty-infinite-warfares-first-victory-6m-bot-messages-on-facebook/>

### Introduction

Intelligent Conversational Agents

### Examples

Design process

### Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

Question Classification  
NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

Notes and Further Reading

[Introduction](#)

Intelligent Conversational Agents

[Examples](#)

Design process

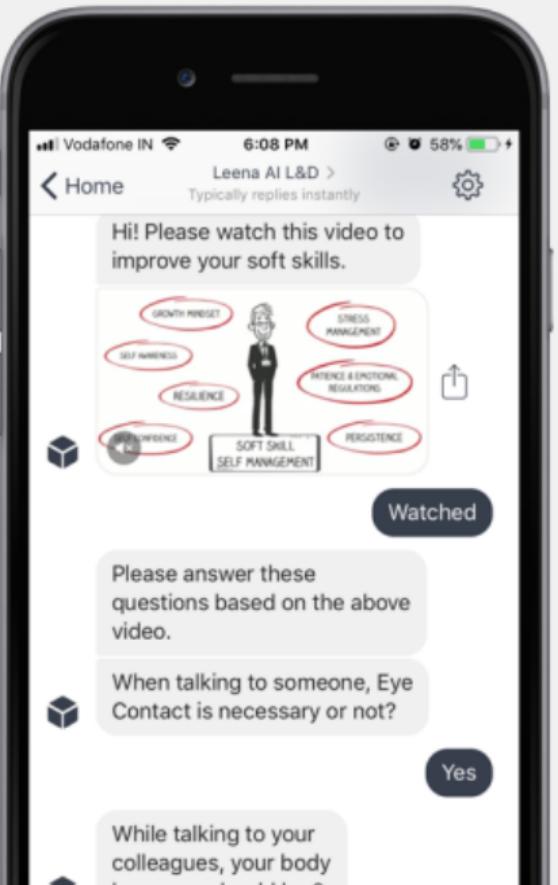
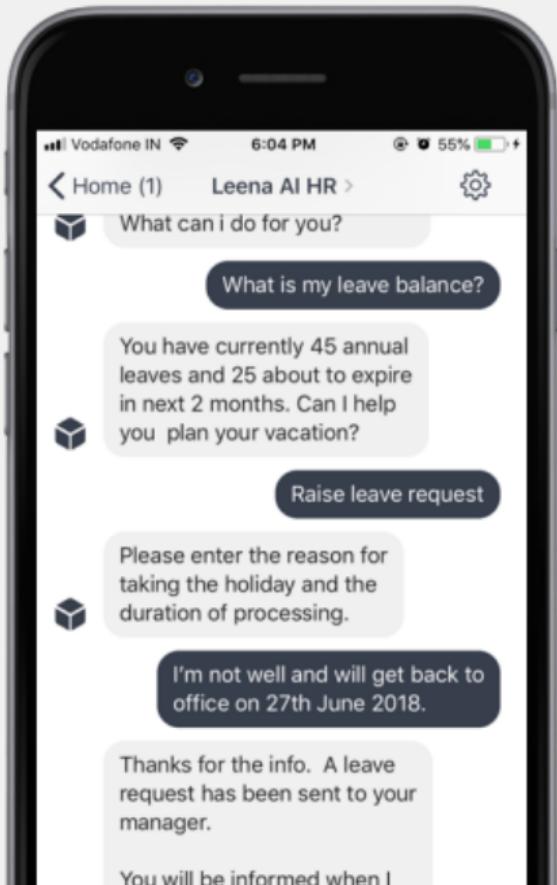
[Search-Based Bots](#)[Pattern-Matching Bots](#)[Grounding-Based Bots](#)

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

[Conclusions](#)Example: Watson  
Summary[Notes and Further Reading](#)

## Introduction

Intelligent Conversational Agents

### Examples

Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson

Summary

## Notes and Further Reading



**CONTACT : +91 99990 54014**  
**[sales@leena.ai](mailto:sales@leena.ai)**

▶ ▶! 🔍 2:11 / 2:13



Leena AI Customer Story: Reliance Capital

<https://www.youtube.com/watch?v=aiuDC2OSIYE>

## NEWS

# Y Combinator Graduate & HRTech Startup Leena AI Raises \$2 Mn Seed Fund



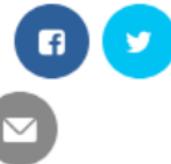
Bhumika Khatri

Inc42 Staff

20 Sep '18 • 4 min read

## SHARE STORY

232 SHARES



- Leena AI has raised funds from the investors from US, Latin America, Indonesia, Middle East and India
- Kunal Bahl and Rohit Bansal (founders of Indian ecommerce platform Snapdeal) also invested in the round
- The company plans to use the funds for product development and international expansion

[Introduction](#)[Intelligent Conversational Agents](#)[Examples](#)[Design process](#)[Search-Based Bots](#)[Pattern-Matching Bots](#)[Grounding-Based Bots](#)[Question Classification](#)[NLP Pipelines](#)[Designing the Knowledge Base](#)[Linking Documents and Knowledge Graphs](#)[Conclusions](#)[Example: Watson Summary](#)[Notes and Further Reading](#)

# Chatbot lawyer overturns 160,000 parking tickets in London and New York

René Witte



**Free service DoNotPay helps appeal over \$4m in parking fines in just 21 months, but is just the tip of the legal AI iceberg for its 19-year-old creator**



▲ DoNotPay has been helping Londoners and New Yorkers fight illegal parking fines, like this one from Lambeth in London. Photograph: Alamy

An artificial-intelligence lawyer chatbot has successfully contested 160,000 parking tickets across London and New York for free, showing that chatbots can actually be useful.

## Introduction

Intelligent Conversational Agents

## Examples

Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson

Summary

## Notes and Further Reading

## Perceived Business Benefits

- “Top 5 Emerging Technologies in 2018” (Gartner)
- Global Market to reach \$1-3B by 2025, CAGR of 25-40%
- Huge benefits across the value chain:
  - Sales & Marketing
  - HR & Operations
  - Service & Payment
  - Retention & Growth

The screenshot shows a web browser window with the URL gartner.com. The page title is "Gartner". There are three blue navigation icons: a menu icon, a search icon, and a user profile icon. The main content is an article titled "Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020".

**Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020**

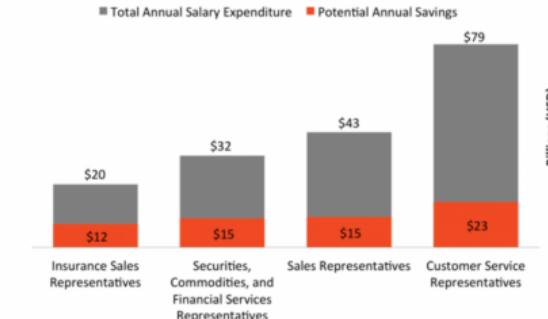
The screenshot shows a web browser window with the URL businessinsider.com. The page title is "BUSINESS INSIDER". The main content is an article titled "80% of businesses want chatbots by 2020".

**80% of businesses want chatbots by 2020**

Business Insider Intelligence Dec. 14, 2016, 10:15 AM



Potential Annual US Salary Savings Created By Chatbots



René Witte



Introduction

Intelligent Conversational Agents

Examples

Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification  
NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading



[Join Extra Crunch](#)

[Login](#)

[Search](#)

[TC Early Stage  
2021](#)

[Startups](#)

[Videos](#)

[Audio](#)

[Newsletters](#)

[Extra Crunch](#)

[The TC List](#)

[Advertise](#)

[Events](#)

—

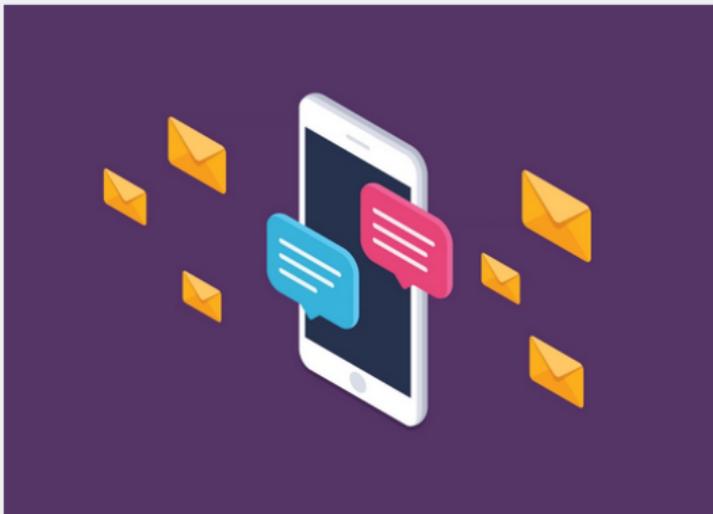
[More](#)

# Chatbot startup Heyday raises \$5.1M



Anthony Ha @anthonyha 6:51 PM GMT-5 • March 10, 2021

Comment



[Image Credits: Getty Images](#)

Montreal-based [Heyday](#) announced today that it has raised \$6.5 million Canadian (\$5.1 million in US dollars) in additional seed funding.

René Witte



[Introduction](#)

Intelligent Conversational Agents

[Examples](#)

Design process

[Search-Based Bots](#)

[Pattern-Matching Bots](#)

[Grounding-Based Bots](#)

Question Classification  
NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

[Conclusions](#)

Example: Watson  
Summary

[Notes and Further Reading](#)

## Realizing Chatbots in Higher Education

Edubots is an ongoing [Erasmus+ Knowledge Alliance](#) project, funded by the European Commission for 2019-2021.

We welcome all educators in Europe to join our learning community to participate in the project activities and make use of the research, learning resources and tools that will be made available.

[Sign up for Pilot](#)[Learn more](#)

<https://www.edubots.eu>



### Introduction

Intelligent Conversational Agents

### Examples

Design process

### Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

Question Classification  
NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

### Notes and Further Reading

## User Stories

- Develop pairs of user questions – expected answers
- Generalize multiple stories with the same theme

## Development

- Identify appropriate technology for specific type of question
- Enhance bot based on approach (e.g., add data for retrieval)

→ Worksheet #6: Task 1

Introduction

Intelligent Conversational Agents

Examples

Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

Conclusions

Example: Watson

Summary

Notes and Further

Reading

# Outline

René Witte



## 1 Introduction

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

## 2 Search-Based Bots

### Search-Based Bots

Pattern-Matching Bots

### Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

### Notes and Further Reading

## 3 Pattern-Matching Bots

## 4 Grounding-Based Bots

## 5 Conclusions

## 6 Notes and Further Reading

▲  
8  
votes  
▼

**Question:** I have a echo & echo dot in my house, can i use this on another wifi system?

**Answer:** Yes you can. As part of the set up process you select the wifi network you want it attached.

By Jon4093 on October 16, 2018

▼ See more answers (5)

## Information Retrieval-based Approach

Given a corpus of previously answered questions

- Create tf-idf vector of the question
- Compute cosine similarity with either:
  - tf-idf vectors of answers (if only answers available)
  - tf-idf vectors of questions (if question/answer pairs available)

## Example Dataset

1.4 million answered questions from Amazon:

<https://jmcauley.ucsd.edu/data/amazon/qa/>

Introduction

Intelligent Conversational Agents  
Examples  
Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading

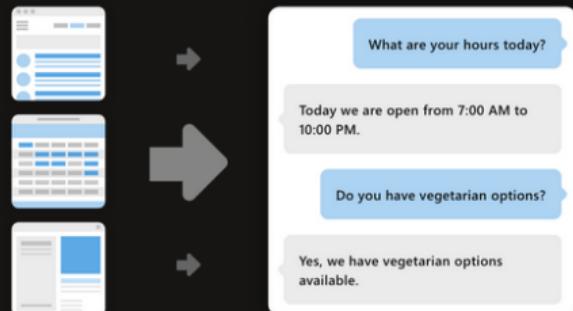
# QnA Maker (Microsoft)

“Low code” approach

Try QnA Maker managed (Preview) - released with deep learning and easy resource management. [Learn more.](#)

# From data to bot in minutes

Build, train and publish a sophisticated bot using FAQ pages, support websites, product manuals, SharePoint documents or editorial content through an easy-to-use UI or via REST APIs.



[Get started >](#)

# Microsoft Azure Cognitive Services for Language



## Get started with Azure Cognitive Services for Language

Use our Natural Language Processing (NLP) features to analyze your textual data using state-of-the-art pre-configured AI models or customize your own models to fit your scenario.



Featured

Extract information

Classify text

Understand questions and conversational language

Summarize text

Translate text

Retrieve the most appropriate answer to questions using question answering (CQA) or classify intents and extract entities for conversational utterances using conversational language understanding (CLU). Use orchestration workflow to create one project that routes queries between multiple CQA and CLU projects. [Learn more about understanding conversational language.](#)



### Answer questions

Use the prebuilt question answering API to get answers to questions over unstructured text.



### Custom question answering

*Next generation of QnAMaker*

Customize the list of questions and answers extracted from your content corpus to provide a conversational experience that suits your needs.



### Conversational language understanding

*Next generation of LUIS*

Classify utterances into intents and extract information with entities to build natural language into apps, bots, and IoT devices.

[Open Conversational language understanding](#)



### Orchestration workflow

Connect and orchestrate CLU, Custom question answering & LUIS projects together in one single project.

[Try it out](#)

[Open Custom question answering](#)

[Open Orchestration workflow](#)

# Example

René Witte

## ↳ Test the knowledge base

QnA Maker GA (stable release)

QnA Maker managed (preview release)

1. In the QnA Maker portal, in the upper right, select **Test** to test that the changes you made took effect.
2. Enter an example user query in the textbox.

I want to know the difference between 32 bit and 64 bit Windows

The screenshot shows the QnA Maker portal interface. At the top, there are two tabs: "QnA Maker GA (stable release)" and "QnA Maker managed (preview release)". Below the tabs, a user query "I want to know the difference between 32 bit and 64 bit Windows" is entered into a text input field. To the left of the input field are two buttons: "Published KB ?" with a checkbox and "Start over". The main area displays the query and the system's response. The response is presented in a blue box with the text: "i want to know the difference between 32 bit and 64 bit Windows". Below this, the system says "Inspect You". A detailed explanation follows: "The terms 32-bit and 64-bit refer to the way a computer's processor (also called a CPU) handles information. The 64-bit version of Windows handles large amounts of random access memory (RAM) more effectively than a 32-bit system. Not all devices can run the 64-bit versions of Windows." At the bottom of the response area, it says "TestSurfaceManual (Test) at 1:30 PM". At the very bottom, there is a text input field with the placeholder "Type your message here ...".

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

### Search-Based Bots

#### Pattern-Matching Bots

#### Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

### Notes and Further Reading

# Outline

René Witte



## 1 Introduction

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

## 2 Search-Based Bots

### Search-Based Bots

## 3 Pattern-Matching Bots

### Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## 4 Grounding-Based Bots

### Conclusions

Example: Watson  
Summary

## 5 Conclusions

Notes and Further Reading

## 6 Notes and Further Reading

# Eliza Example

René Witte



## Eliza + DOCTOR script

If the input sentence is:

*I am very unhappy these days.*

Eliza's response will be:

*How long have you been very unhappy these days?*

## Processing

Keyword:

*I am*

Decomposition pattern:

*I am <whatever>*

Reassembly Pattern:

*How long have you been <whatever>?*

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

### Search-Based Bots

### Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

### Notes and Further Reading

# Pattern-response matching

René Witte



```
"(..)?do you remember(..)": [
    "Did you think I would forget $2?",  

    "Why do you think I should recall $2 now?",  

    "What about $2?",  

    "You mentioned $2",
],  
  
"(..)?my mother(..)": [
    "Who else in your family $2?",  

    "Tell me more about your family",
],  
  
"(..)?why don't you(..)": [
    "Should you $2 yourself?",  

    "Do you believe I don't $2?",  

    "Perhaps I will $2 in good time",
],
```

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## Conclusions

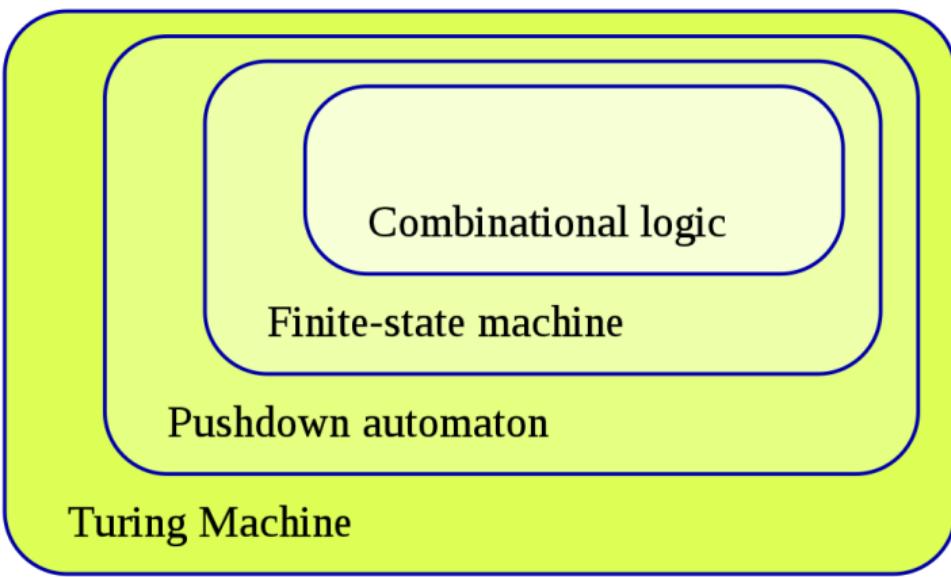
Example: Watson  
Summary

## Notes and Further Reading

## Finite State Machine

A.k.a deterministic finite automaton (DFA).

## Automata theory



Introduction

Intelligent Conversational Agents  
Examples  
Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading

## Basic Concepts ([https://en.wikipedia.org/wiki/Regular\\_expression](https://en.wikipedia.org/wiki/Regular_expression))

**Boolean "or":** A vertical bar separates alternatives. For example, gray|grey can match "gray" or "grey".

**Grouping:** Parentheses are used to define the scope and precedence of the operators (among other uses). For example, gray|grey and gr(a|e)y are equivalent patterns which both describe the set of "gray" or "grey".

**Quantification:** A quantifier after a token (such as a character) or group specifies how often that a preceding element is allowed to occur:

- ? The question mark indicates zero or one occurrences of the preceding element. For example, colou?r matches both "color" and "colour".
- \* The asterisk indicates zero or more occurrences of the preceding element. For example, ab\*c matches "ac", "abc", "abbc", "abbcc", and so on.
- + The plus sign indicates one or more occurrences of the preceding element. For example, ab+c matches "abc", "abbc", "abbcc", and so on, but not "ac".
- {n} The preceding item is matched exactly n times.
- {min,} The preceding item is matched min or more times.
- {min,max} The preceding item is matched at least min times, but not more than max times.

→ Worksheet #6: Task 2

Introduction

Intelligent Conversational Agents  
Examples  
Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading

## Towards Chatbot Frameworks

- Open standard, started in 1995 by Richard Wallace et al.
- Used in the A.L.I.C.E. chatbot (and many others)
- Using XML-based patterns
- AIML kernel loads patterns and responds when match is found

## Example

```
<category>
  <pattern>HELLO *</pattern>
  <template>Hi, human!</template>
</category>
<category>
  <pattern>WHAT IS YOUR NAME</pattern>
  <template>I am ConUBot, your helpful assistant.</template>
</category>
```

Introduction

Intelligent Conversational Agents  
Examples  
Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading

# AIML Examples (v2.0)

René Witte



```
<?xml version="1.0" encoding="UTF-8"?><aiml version="2.0">
<category>
    <pattern>HI</pattern>
    <template>Hi!</template>
</category>

<category>
    <pattern>[HELLO HI YO YOH YO']<br/>
        [ROSA ROSE CHATTY CHATBOT BOT CHATTERBOT]</pattern>
    <template>Hi , How are you?</template>
</category>

<category>
    <pattern>[HELLO HI YO YOH YO' 'SUP SUP OK HEY]<br/>
        [HAL YOU U YALL Y'ALL YOUS YOUSE]</pattern>
    <template>Good one.</template>
</category>
</aiml>
```

(from [LHH19, Chapter 12])

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

## Python Implementations

- E.g., PyAiml, aiml, aiml\_bot
- Support AIML 1.0 only :(

```
<?xml version="1.0" encoding="UTF-8"?>
<aiml version="1.0.1">
<category>
  <pattern>HELLO *</pattern>
  <template>Hi Human!</template>
</category>
<category>
  <pattern>HELLO TROLL</pattern>
  <template>Good one, human.</template>
</category>
</aiml>
```

## Using aiml\_bot

```
import aiml_bot
bot = aiml_bot.Bot(learn="conubot.aiml")
In : bot.respond("Hello_Conubot!!!")
Out: 'Hi_Human!'
```

See <https://pypi.org/project/AIML-Bot/>

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

### Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

### Notes and Further Reading

# Random answers

René Witte



```
<?xml version="1.0" encoding="UTF-8"?>
<aiml version="1.0.1">
  <category>
    <pattern>HELLO *</pattern>
    <template>
      <random>
        <li>Hi Human!</li>
        <li>Hi there!</li>
        <li>Hello!</li>
      </random>
    </template>
  </category>
  <category>
    <pattern>HELLO TROLL</pattern>
    <template>
      <random>
        <li>Good one, human.</li>
        <li>Clever!</li>
      </random>
    </template>
  </category>
</aiml>
```

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

```
<aiml version="1.0.1" encoding="UTF-8">
  <category>
    <pattern>MY DOGS NAME IS *</pattern>
    <template>
      That is interesting that you have a dog
      named <set name="dog"><star/></set>
    </template>
  </category>
  <category>
    <pattern>WHAT IS MY DOGS NAME</pattern>
    <template>
      Your dog's name is <get name="dog"/>.
    </template>
  </category>
</aiml>
```

## Example

My dogs name is Max

That is interesting that you have a dog named Max

...

What is my dogs name?

Your dog's name is Max.

<https://www.devdungeon.com/content/ai-chat-bot-python-aiml#sessions>



### Introduction

Intelligent Conversational Agents  
Examples  
Design process

### Search-Based Bots

### Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

### Notes and Further Reading

```
<card>
  <image>www.png</image>
  <title>Italian Greyhound</title>
  <subtitle>A very good dog</subtitle>
  <button>
    <text>AIML How-To</text>
    <postback>HOW TO</postback>
  </button>
  <button>
    <text>Back To Tour</text>
    <postback>RESUME TOUR</postback>
  </button>
</card>
```



**Italian Greyhound**  
A very good dog

[AIML How-To](#)

[Back to Tour](#)

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

<http://www.aiml.foundation/doc.html>

# Pandorabots (<https://pandorabots.com>)

Cloud-based platform for AIML 2.0 bots

René Witte



pandorabots

MY BOTS bender 3

CLUBHOUSE >

HELP >

Integrations

|                     | Name / AppID | Status | Channel |
|---------------------|--------------|--------|---------|
| Clubhouse           | -            | -      | Add     |
| Web Widget          | -            | -      | 🔒       |
| Amazon Echo         | -            | -      | 🔒       |
| Browser Integration | -            | -      | 🔒       |
| Facebook Messenger  | -            | -      | 🔒       |
| kik- Kik Messenger  | -            | -      | 🔒       |
| Line                | -            | -      | 🔒       |
| Skype               | -            | -      | 🔒       |
| Slack               | -            | -      | 🔒       |
| Telegram            | -            | -      | 🔒       |
| Twilio              | -            | -      | 🔒       |
| Twitter             | -            | -      | 🔒       |

A purple speech bubble icon is located at the bottom right of the table.

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

## Other bot languages

René Witte

### API.ai

Proprietary, intuitive language for dialog specifications

- Dialogue history, location and user preferences
- Developed by startup Speaktoit, released as api.ai in 2014

### Dialogflow

API.ai was bought by Google in 2016 and renamed in 2017 to [Dialogflow](#)

- Powers Google Assistant and other services

The screenshot shows the Dialogflow website. At the top, there's a navigation bar with links for 'Overview' (which is underlined in orange), 'Case studies', 'Docs', 'Blog', 'Pricing', and 'Support'. To the right of the navigation are a search bar, a 'Go to console' button, and a 'Sign in' button. Below the navigation, a large heading reads 'Build natural and rich conversational experiences'. Underneath this heading is a paragraph of text: 'Give users new ways to interact with your product by building engaging voice and text-based conversational interfaces, such as voice apps and chatbots, powered by AI. Connect with users on your website, mobile app, the Google Assistant, Amazon Alexa, Facebook Messenger, and other popular platforms and devices.' At the bottom left is an orange 'Sign up for free' button. On the right side of the page, there's a large promotional image for a video titled 'Intro to Dialogflow'. The image features a woman sitting at a desk with a laptop, and various icons for YouTube, a play button, and social sharing.

Powered by Google's machine learning

Dialogflow incorporates Google's machine learning

Built on Google infrastructure

Dialogflow is a Google service that runs on Google

Optimized for the Google Assistant

Dialogflow is the most widely used tool to build



#### Introduction

Intelligent Conversational Agents  
Examples  
Design process

#### Search-Based Bots

#### Pattern-Matching Bots

#### Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

#### Conclusions

Example: Watson  
Summary

#### Notes and Further Reading

# Amazon Lex

Conversational interfaces for your applications powered by the same deep learning technologies as Alexa

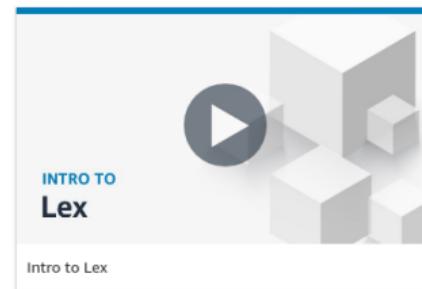
Get started with Amazon Lex

Amazon Lex is a service for building conversational interfaces into any application using voice and text. Amazon Lex provides the advanced deep learning functionalities of automatic speech recognition (ASR) for converting speech to text, and natural language understanding (NLU) to recognize the intent of the text, to enable you to build applications with highly engaging user experiences and lifelike conversational interactions. With Amazon Lex, the same deep learning technologies that power Amazon Alexa are now available to any developer, enabling you to quickly and easily build sophisticated, natural language, conversational bots (*"chatbots"*).

Speech recognition and natural language understanding are some of the most challenging problems to solve in computer science, requiring sophisticated deep learning algorithms to be trained on massive amounts of data and infrastructure. Amazon Lex democratizes these deep learning technologies by putting the power of Amazon Alexa within reach of all developers. Harnessing these technologies, Amazon Lex enables you to define entirely new categories of products made possible through conversational interfaces.

As a fully managed service, Amazon Lex scales automatically, so you don't need to worry about managing infrastructure. With Amazon Lex, you pay only for what you use. There are no upfront commitments or minimum fees.

<https://aws.amazon.com/lex/>



## Introduction

- Intelligent Conversational Agents
- Examples
- Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

- Question Classification
- NLP Pipelines
- Designing the Knowledge Base
- Linking Documents and Knowledge Graphs

## Conclusions

- Example: Watson
- Summary

## Notes and Further Reading

# Outline

René Witte



## 1 Introduction

Introduction

Intelligent Conversational Agents  
Examples  
Design process

## 2 Search-Based Bots

Search-Based Bots

## 3 Pattern-Matching Bots

Pattern-Matching Bots

## 4 Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading

## 5 Conclusions

## 6 Notes and Further Reading

## Adding Knowledge

- Using a **knowledge base** when creating answers
- More scalable than hard-coding every possible answer in patterns

## From the lab exercises...

① *What is <X>?*

E.g., “*What is Concordia University?*”

② ⇒ query DBpedia to retrieve the `rdfs:comment` (in the user's language) of X

*Concordia University (commonly referred to as Concordia) is a public comprehensive university located in Montreal, Quebec, Canada. Founded in 1974 following the merger of Loyola College and Sir George Williams University, Concordia is one of the three universities in Quebec where English is the primary language of instruction.*

[Introduction](#)

[Intelligent Conversational Agents](#)  
[Examples](#)  
[Design process](#)

[Search-Based Bots](#)

[Pattern-Matching Bots](#)

[Grounding-Based Bots](#)

[Question Classification](#)  
[NLP Pipelines](#)  
[Designing the Knowledge Base](#)  
[Linking Documents and Knowledge Graphs](#)

[Conclusions](#)

[Example: Watson](#)  
[Summary](#)

[Notes and Further Reading](#)

# Question-answering Workflow

René Witte



## Introduction

- Intelligent Conversational Agents
- Examples
- Design process

## Search-Based Bots

## Pattern-Matching Bots

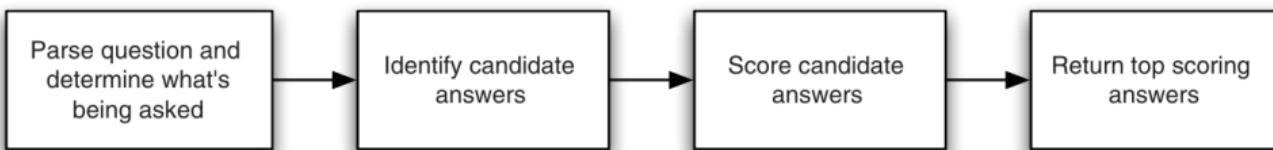
## Grounding-Based Bots

- Question Classification
- NLP Pipelines
- Designing the Knowledge Base
- Linking Documents and Knowledge Graphs

## Conclusions

- Example: Watson
- Summary

## Notes and Further Reading



Copyright 2013 by Manning Publications Co., [IMF13]

## Classifying Questions

Different types of questions require different SPARQL query structures (ASK, SELECT, using COUNT, etc.)

| Answer type (training code) | Example   |
|-----------------------------|---|
| Person (P)                  | Which Ivy League basketball player scored the most points in a single game during the 1990s?            |
| Location (L)                | Which city generates the highest levels of sulphur dioxide in the world?                                |
| Organization (O)            | Which ski resort was named the best in North America by readers of <i>Conde Nast Traveler</i> magazine? |
| Time point (T)              | What year did the Pilgrims have their first Thanksgiving feast?   |
| Duration (R)                | How long did <i>Gunsmoke</i> run on network TV?   |
| Money (M)                   | How much are Salvadoran workers paid for each \$198 Liz Claiborne jacket they sew?                      |

# Question Types (contd.)

René Witte



| Answer type (training code) | Example  |
|-----------------------------|--|
| Percentage (C)              | What percentage of newspapers in the U.S. say they are making a profit from their online site?   |
| Amount (A)                  | What is the lowest temperature ever recorded in November in New Brunswick?   |
| Distance (D)                | What is the approximate maximum distance at which a clap of thunder can be heard?  |
| Description (F)             | What is dry ice?   |
| Title (W)                   | In which fourteenth-century alliterative poem by William Langford do a series of allegorical visions appear to the narrator in his dreams? |
| Definition (B)              | What does the postage stamp cancellation O.H.M.S. mean?  |
| Other (X)                   | How did the banana split originate?  |

Copyright 2013 by Manning Publications Co., [MF13]

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

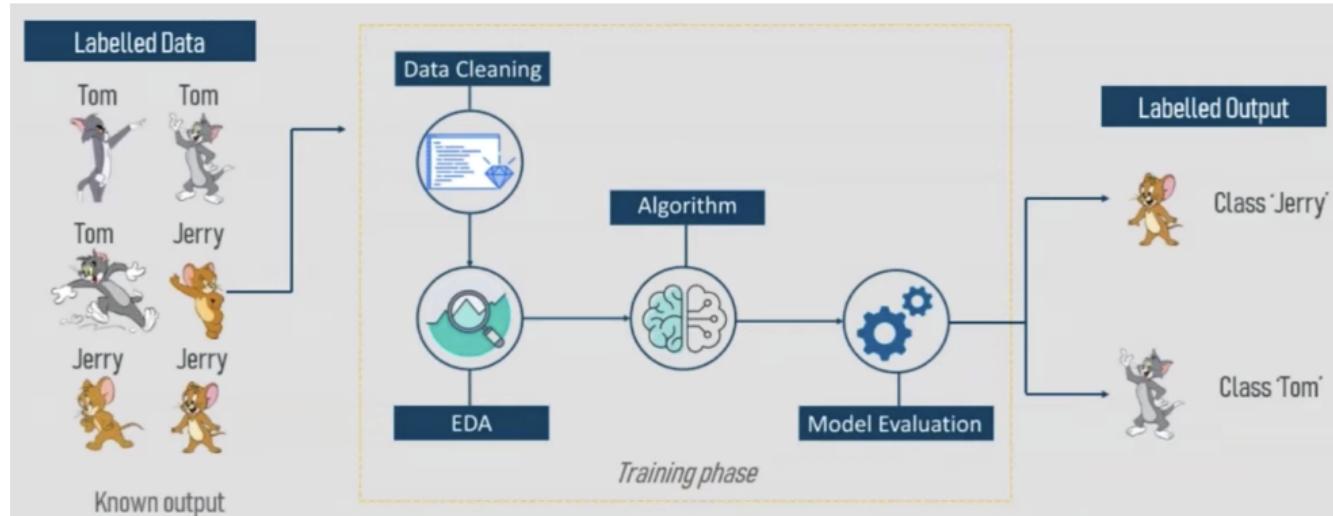
## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

## Applying ML

- What we need is a **classifier** that takes a question as input and returns the type as output
- This is a typical machine learning problem (supervised learning)



### Introduction

Intelligent Conversational Agents  
Examples  
Design process

### Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

### Notes and Further Reading

## Feature Engineering

Need to convert text (here: question) into a **feature vector**

- Could use count or tf-idf vector
- but this results in a high number of dimensions (and possible overfitting)

## Idea: Reduce Dimensions

Can we come up with some other features that can be easily extracted?

- Length (in words/characters)?
- Number of words with capital letters?
- Ends with question mark or not?
- Number of nouns/verbs?

Challenge: find features that facilitate classification.

→ Worksheet #6: Tasks 3 & 4

Introduction

Intelligent Conversational Agents  
Examples  
Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification

NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

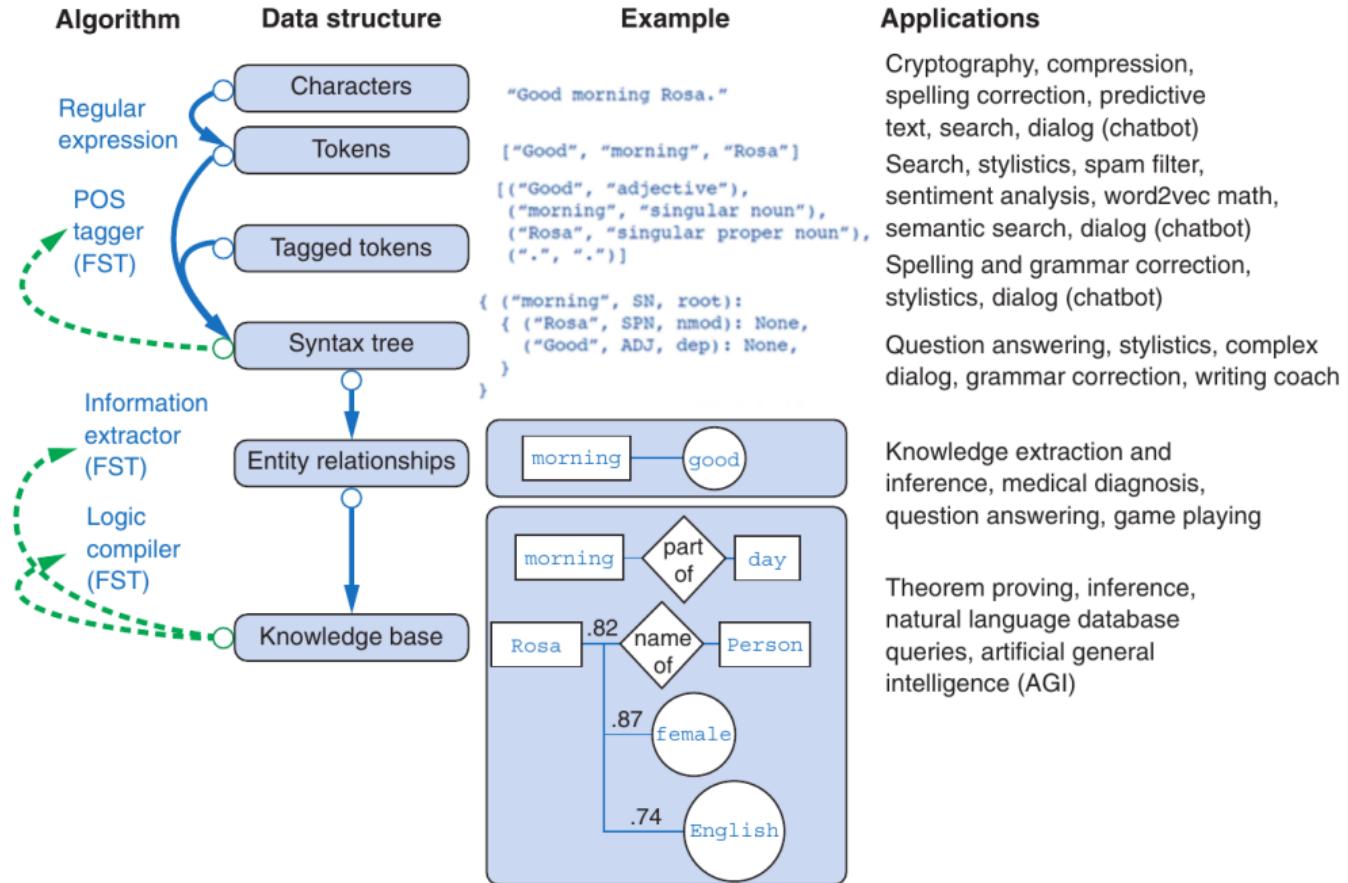
## Some common strategies

**Stemming:** reduce words to their stem (e.g., *students*, *student* ⇒ *student*)

**Stopword removal:** remove stop-words (e.g., *the*, *in*, *an*, *a*, ...)

# Example NLP Pipeline

René Witte



- Introduction
  - Intelligent Conversational Agents
  - Examples
  - Design process
- Search-Based Bots
- Pattern-Matching Bots
- Grounding-Based Bots
  - Question Classification
- NLP Pipelines
  - Designing the Knowledge Base
  - Linking Documents and Knowledge Graphs
- Conclusions
  - Example: Watson Summary
- Notes and Further Reading

## Task

- Split input stream of characters into individual tokens (words, numbers, etc.)
- Done by a [Tokenizer](#) (e.g., the default tokenizer in scikit-learn)

## Tokenization can be difficult...

For example, biomedical documents with complex expressions, chemical formulas, etc.:

- *1,4- $\beta$ -xylanase II from Trichoderma reesei*
- *When N-formyl-L-methionyl-L-leucyl-L-phenylalanine (fMLP) was injected...*
- *Technetium-99m-CDO-MeB [Bis[1,2-cyclohexanedione-dioximato(1-O)-[1,2-cyclohexanedione dioximato(2-O)methyl-borato(2-)N,N',N'',N'',N''',N''''')-chlorotechnetium) belongs to a family of compounds...*

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

### Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

Question Classification  
NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

Notes and Further Reading

## POS Tagging

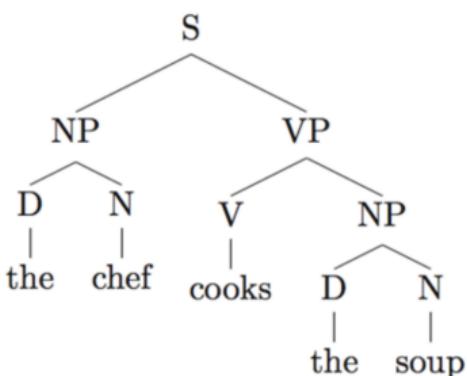
Assign a POS tag (e.g., Noun, Verb, Adjective, Adverb, ...) to each Token:

*The/DT big/ADJ dog/NN*

Can be done reliably, available in NLP libraries (e.g, spaCy or NLTK for Python).

## Parsing

Create a tree representing a sentence's grammatical structure



You can then extract **subject** or **object**, e.g., to use in a SPARQL query.

→ Worksheet #6: Task 5

(ROOT (SBARQ  
    (WHADVP (WRB When))  
    (SQ (VBD did) (NP (NNP princess) (NNP Diana))  
        (VP (VB die))) (. ?)))

Tab. 1: List of identified question types

| Sentence Type | Question Type | Example   |
|---------------|---------------|---|
| SBARQ         | WHADVP        | When was the Battle of Gettysburg?                        |
|               | WHADJP        | How much did Pulp Fiction cost?                           |
|               | WHNP          | Who designed the Brooklyn Bridge?                         |
|               | WHPP          | In which city does the Chile Route 68 end?                |
| S             |               | Show me all books ... . List all basketball players ... . |
| SQ            |               | Is Berlin the German capital?                             |

- [Introduction](#)
  - Intelligent Conversational Agents
  - Examples
  - Design process
- [Search-Based Bots](#)
- [Pattern-Matching Bots](#)
- [Grounding-Based Bots](#)
  - Question Classification
- [NLP Pipelines](#)
  - Designing the Knowledge Base
  - Linking Documents and Knowledge Graphs
- [Conclusions](#)
  - Example: Watson Summary
- [Notes and Further Reading](#)

# Generating SPARQL Queries (II)

René Witte



## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification  
NLP Pipelines

Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

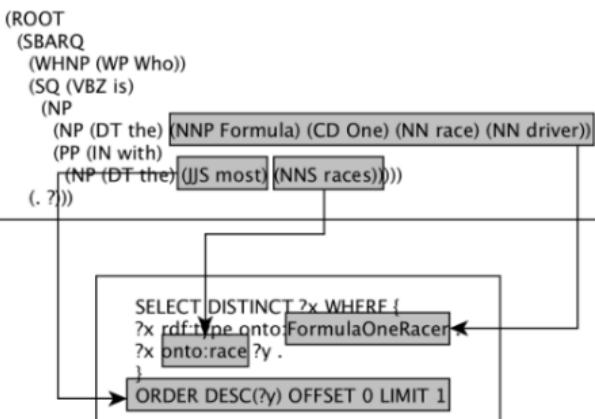
Notes and Further Reading

How many children did Benjamin Franklin have?

```
SELECT ?x WHERE {  
    res:Benjamin_Franklin onto:child ?x .  
}
```

```
SELECT COUNT(DISTINCT ?x) WHERE {  
    res:Benjamin_Franklin onto:child ?x .  
}
```

Who is the Formula One race driver with the most races?



Steinmetz N, Arning AK, Sattler KU. From natural language questions to SPARQL queries: a pattern-based approach. BTW 2019. <https://dl.gi.de/handle/20.500.12116/21702>

→ Worksheet #6: Task 6

[Introduction](#)[Intelligent Conversational Agents](#)[Examples](#)[Design process](#)[Search-Based Bots](#)[Pattern-Matching Bots](#)[Grounding-Based Bots](#)[Question Classification](#)[NLP Pipelines](#)[Designing the Knowledge Base](#)[Linking Documents and Knowledge Graphs](#)[Conclusions](#)[Example: Watson](#)[Summary](#)[Notes and Further](#)[Reading](#)

$$\text{recall}(q) = \frac{\text{number of correct system answers for } q}{\text{number of benchmark answers for } q}$$

$$\text{precision}(q) = \frac{\text{number of correct system answers for } q}{\text{number of system answers for } q}$$

$$F_1\text{-score} = 2 * \frac{\text{recall}(q) * \text{precision}(q)}{\text{recall}(q) + \text{precision}(q)}$$

# The Stanford Question Answering Dataset (SQuAD)

René Witte

SQuAD

Home Explore 2.0 Explore 1.1

Following the Cretaceous–Paleogene extinction event, the extinction of the dinosaurs and the wetter climate may have allowed the tropical rainforest to spread out across the continent. From 66–34 Mya, the rainforest extended as far south as 45°. Climate fluctuations during the last 34 million years have allowed savanna regions to expand into the tropics. During the Oligocene, for example, the rainforest spanned a relatively narrow band. It expanded again during the Middle Miocene, then retracted to a mostly inland formation at the last glacial maximum. However, the rainforest still managed to thrive during these glacial periods, allowing for the survival and evolution of a broad diversity of species.

Which type of climate may have allowed the rainforest to spread across the continent?

*Ground Truth Answers:* the wetter climate may have allowed the tropical rainforest to spread out across the continent. | wetter | wetter

*Prediction:* wetter

What has allowed for the Savanna region to expand into the tropics?

*Ground Truth Answers:* Climate fluctuations during the last 34 million years have allowed savanna regions to expand into the tropics. | Climate fluctuations during the last 34 million years | Climate fluctuations

*Prediction:* Climate fluctuations

During what time did the rainforest spanned a narrow band?

*Ground Truth Answers:* During the Oligocene, for example, the rainforest spanned a relatively narrow band. | Oligocene | Oligocene

*Prediction:* Oligocene

When did it retract to a inland formation?

Introduction

Intelligent Conversational Agents  
Examples  
Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines

Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading

<https://rajpurkar.github.io/SQuAD-explorer/>

→ Worksheet #6: Task 7

## 1 Introduction

Introduction

Intelligent Conversational Agents  
Examples  
Design process

## 2 Search-Based Bots

Search-Based Bots

## 3 Pattern-Matching Bots

Pattern-Matching Bots

## 4 Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

Grounding-Based Bots

Question Classification  
NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading

## 5 Conclusions

## 6 Notes and Further Reading

## Knowledge Base Design

Unlike (relational) databases, knowledge bases do not have a rigid, pre-defined design for a specific application

- Encourage inter-connection of knowledge
- Still helpful to have an idea of possible use cases

## Design Method

- Start with a set of **questions** to guide the design of the vocabulary and data triples
- It must be possible to answer the questions when the knowledge base has been constructed
- These **competency questions** can be seen as a *requirements specification* for a vocabulary or knowledge base

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

### Search-Based Bots

### Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

Notes and Further Reading

## Examples

Some possible competency questions for a university knowledge graph:

- Q1: Which courses are offered at University  $X$ ?
- Q2: Which topics are covered by a course  $C$ ?
- Q3: Which professors teach at university  $X$ ?
- Q4: What are the prerequisites for a given course  $C$ ?
- Q5: Which universities offer courses that cover topic  $T$ ?

## Note

These mostly follow a single pattern

*Which [CLASS1] [PROPERTY] [CLASS2] ?*

Introduction

Intelligent Conversational Agents  
Examples  
Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines

Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading

# Competency Question: Patterns

René Witte



| ID | Pattern                                | Example   | PA | RT    | M     | DE   |
|----|--|---|----|-------|-------|------|
| 1  | Which [CE1] [OPE] [CE2]?               | Which pizzas contain pork?  | 2  | obj.  |       |      |
| 2  | How much does [CE] [DP]?               | How much does Margherita Pizza weigh?                                 | 2  | data. |       |      |
| 3  | What type of [CE] is [I]?              | What type of software (API, Desktop application etc.) is it?          | 1  |       |       |      |
| 4  | Is the [CE1] [CE2]?                    | Is the software open source development?                              | 2  |       |       |      |
| 5  | What [CE] has the [NM] [DP]?           | What pizza has the lowest price?                                      | 2  | data. | num.  |      |
| 6  | What is the [NM] [CE1] to [OPE] [CE2]? | What is the best/fastest/most robust software to read/edit this data? | 3  | both  | num.  |      |
| 7  | Where do I [OPE] [CE]?                 | Where do I get updates?   | 2  | obj.  |       | spa. |
| 8  | Which are [CE]?                        | Which are gluten free bases?  | 1  |       |       |      |
| 9  | When did/was [CE] [PE]?                | When was the 1.0 version released?                                    | 2  | data. |       | tem. |
| 10 | What [CE1] do I need to [OPE] [CE2]?   | What hardware do I need to run this software?                         | 3  | obj.  |       |      |
| 11 | Which [CE1] [OPE] [QM] [CE2]?          | Which pizza has the most toppings?                                    | 2  | obj.  | quan. |      |
| 12 | Do [CE1] have [QM] values of [DP]?     | Do pizzas have different values of size?                              | 2  | data. | quan. |      |

Towards Competency Question-Driven Ontology Authoring, [https://link.springer.com/chapter/10.1007/978-3-319-07443-6\\_6](https://link.springer.com/chapter/10.1007/978-3-319-07443-6_6)

CQ Archetypes (PA = Predicate Arity, RT = Relation Type, M = Modifier, DE = Domain-independent Element; obj. = object property relation, data. = datatype property relation, num. = numeric modifier, quan. = quantitative modifier, tem. = temporal element, spa. = spatial element; CE = class expression, OPE = object property expression, DP = datatype property, I = individual, NM = numeric modifier, PE = property expression, QM = quantity modifier)

Introduction

Intelligent Conversational Agents  
Examples  
Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification  
NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

Conclusions

Example: Watson Summary

Notes and Further Reading

## Example: African Wildlife Ontology (AWO)

Competency question awo\_6 “Which plants eat animals?”

Verify using a SPARQL query:

```
SELECT DISTINCT ?eats
WHERE {
    ?eats rdfs:subClassOf awo:plant, [
        a owl:Restriction ;
        owl:onProperty awo:eats;
        owl:someValuesFrom awo:animal
    ] .
    FILTER(?eats != owl:Nothing)
}
```

See [https://www.researchgate.net/publication/338424817\\_Dataset\\_of\\_ontology\\_competency\\_questions\\_to\\_SPARQL-OWL\\_queries\\_translations](https://www.researchgate.net/publication/338424817_Dataset_of_ontology_competency_questions_to_SPARQL-OWL_queries_translations)

## Another approach: SHACL

There is also a separate standard, the *Shapes Constraint Language (SHACL)* (<https://www.w3.org/TR/shacl/>), see [HHDF20] for an example.

→ Worksheet #6: Task 8

Introduction

Intelligent Conversational Agents  
Examples  
Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines

Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

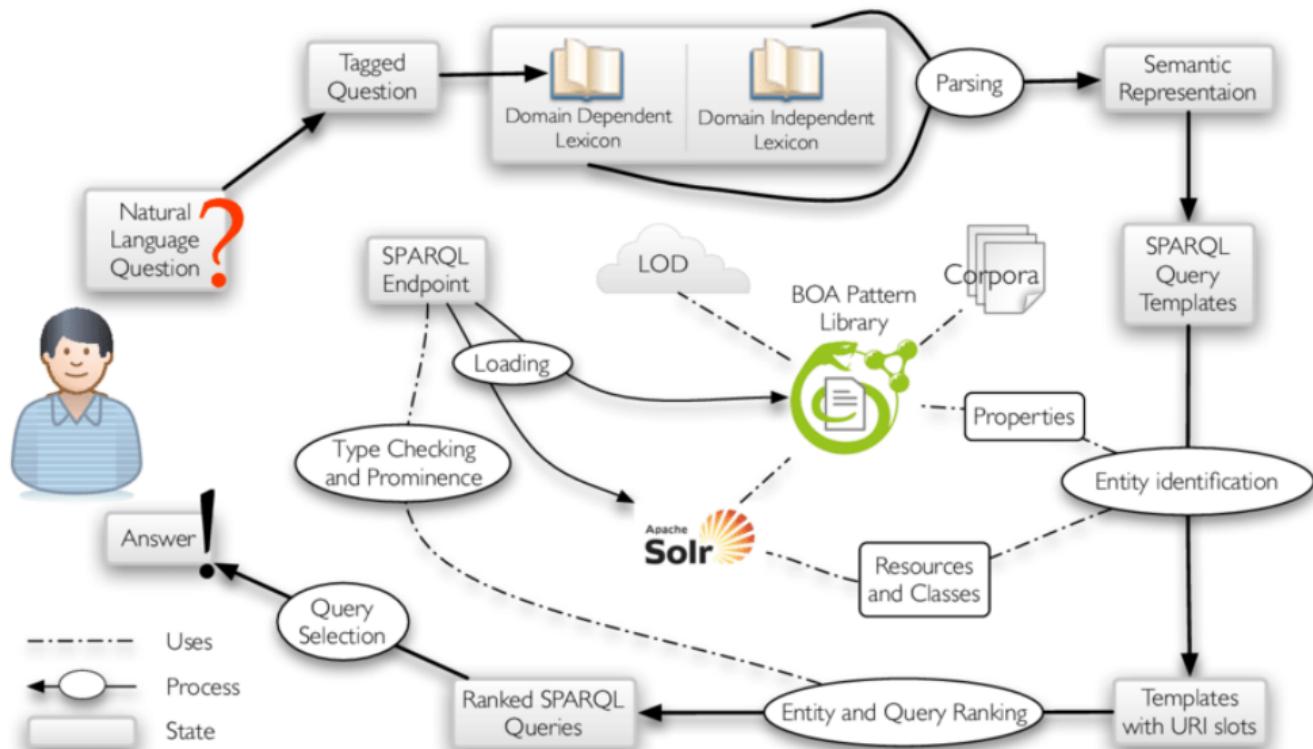
Conclusions

Example: Watson  
Summary

Notes and Further Reading

# Generic SPARQL Query Generator

René Witte



Unger C, Bühlmann L, Lehmann J, Ngonga Ngomo AC, Gerber D, Cimiano P. Template-based question answering over RDF data. In *Proceedings of the 21st international conference on World Wide Web 2012 Apr 16 (pp. 639-648).*

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification  
NLP Pipelines

Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

# Outline

René Witte



## 1 Introduction

Introduction

Intelligent Conversational Agents  
Examples  
Design process

## 2 Search-Based Bots

Search-Based Bots

## 3 Pattern-Matching Bots

Pattern-Matching Bots

## 4 Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base

Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

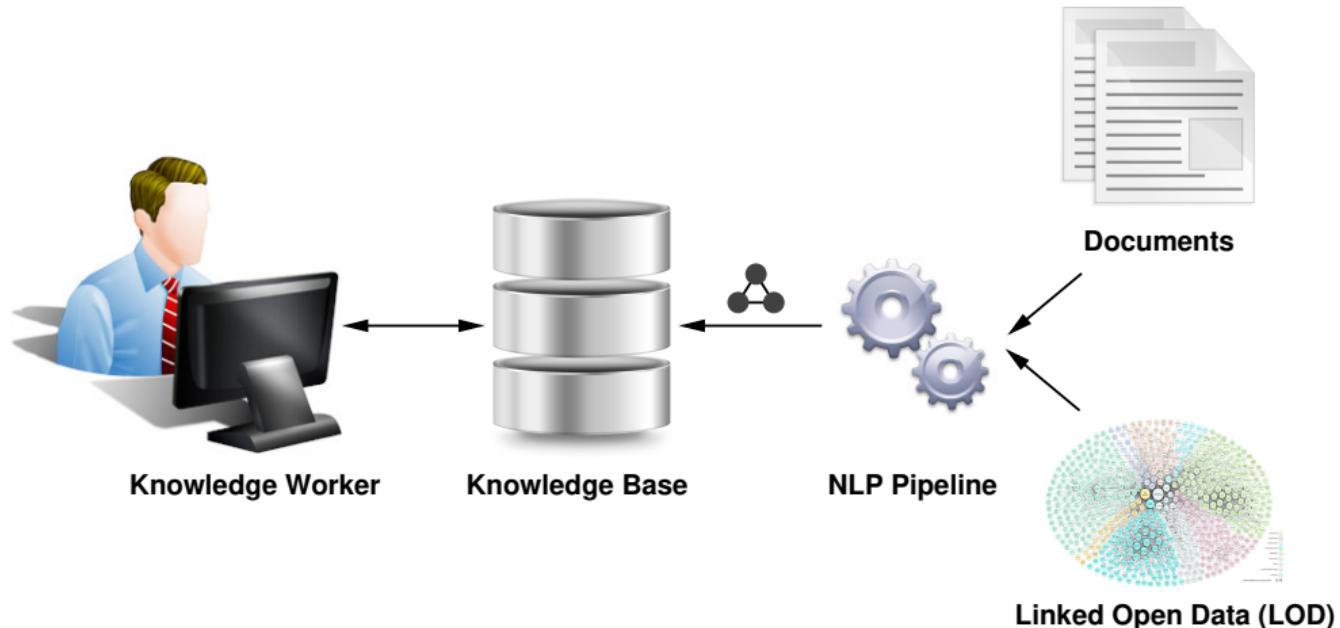
Notes and Further Reading

## 5 Conclusions

## 6 Notes and Further Reading

# Knowledge Management Architecture 3.0

René Witte



## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base

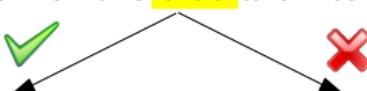
## Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

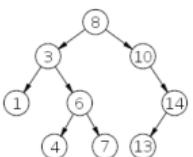
## Notes and Further Reading

"The proposed approach takes advantage of both the efficient computation of the tree architecture ..."



[http://dbpedia.org/resource/Tree\\_\(data structure\)](http://dbpedia.org/resource/Tree_(data%20structure))

<http://dbpedia.org/resource/Tree>



## Introduction

- Intelligent Conversational Agents
- Examples
- Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

- Question Classification
- NLP Pipelines
- Designing the Knowledge Base

## Linking Documents and Knowledge Graphs

## Conclusions

- Example: Watson Summary

## Notes and Further Reading

“What is a group of moose”  
tap to edit

It's Jared Padalecki.

# Jared Padalecki

American actor

Jared Tristan Padalecki (born July 19, 1982) is an American actor. He is best known for his role as Sam Winchester on Supernatural. He grew up in Texas and rose to fame in the early 2000s after appearing on the



[Introduction](#)

[Intelligent Conversational Agents](#)  
[Examples](#)  
[Design process](#)

[Search-Based Bots](#)

[Pattern-Matching Bots](#)

[Grounding-Based Bots](#)

[Question Classification](#)  
[NLP Pipelines](#)  
[Designing the Knowledge Base](#)

[Linking Documents and Knowledge Graphs](#)

[Conclusions](#)

[Example: Watson](#)  
[Summary](#)

[Notes and Further Reading](#)

## PermID: Connecting Data to the World

Search our Linked Data Graph by Name, Ticker or RIC



*"PermID plays a key role in empowering the creation of an efficient, transparent, and orderly marketplace for the crowdfinancing industry."*

LUAN COX, CEO OF CROWDNETIC

Register Now



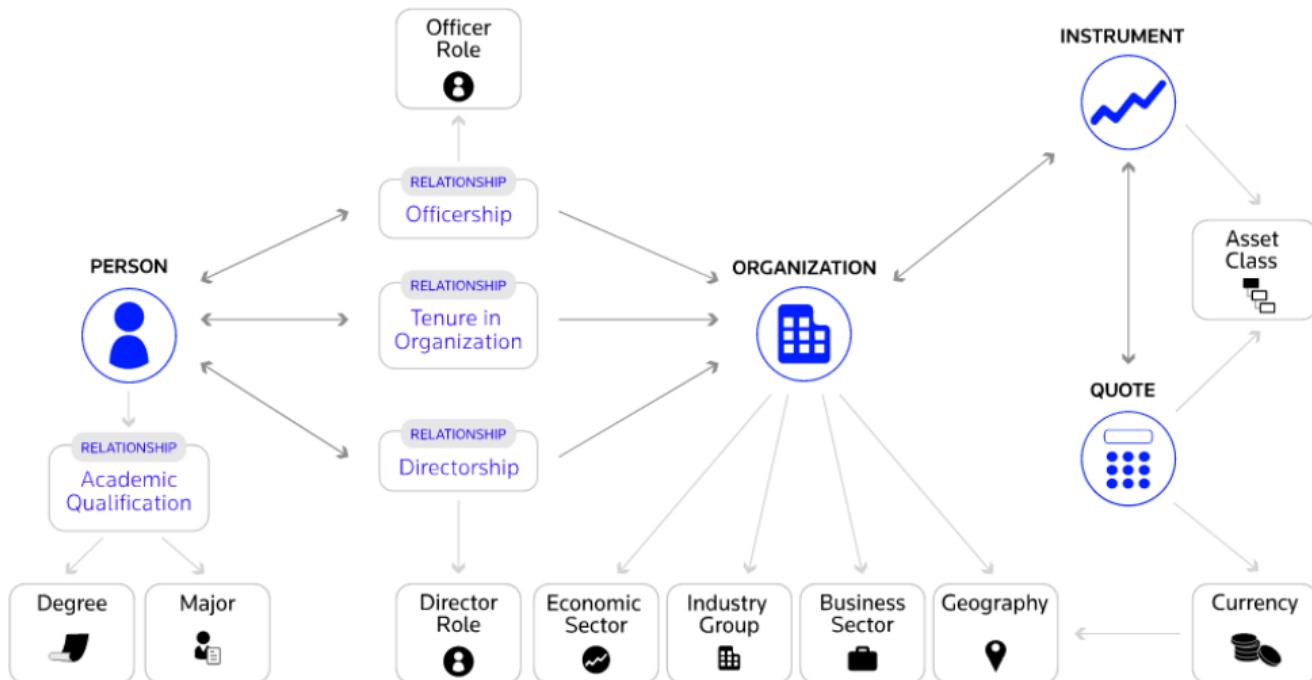
### What is PermID?

Data has more potential value when it can be shared or opened. It can be used by a number of different stakeholders and partners, within or outside an organization, for a variety of applications to gain new analytical insight and to build new products and services. And in order to effectively use data, it's important to understand how it connects to the real world. That's why LSEG is making available its Permanent Identifiers, or PermIDs, and the associated entity masters and metadata to the market. PermIDs are open, permanent and universal identifiers where underlying attributes capture the context of the identity they each represent.

Explore the site to access the PermIDs available through a free and open license. [Click here](#) to register.

# PermID Linked Data Graph

René Witte



→ Worksheet #6: Task 9

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base

## Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading



# DBpedia Spotlight

## Shedding light on the web of documents

It is a tool for automatically annotating mentions of DBpedia resources in text, providing a solution for linking unstructured information sources to the Linked Open Data cloud through DBpedia.



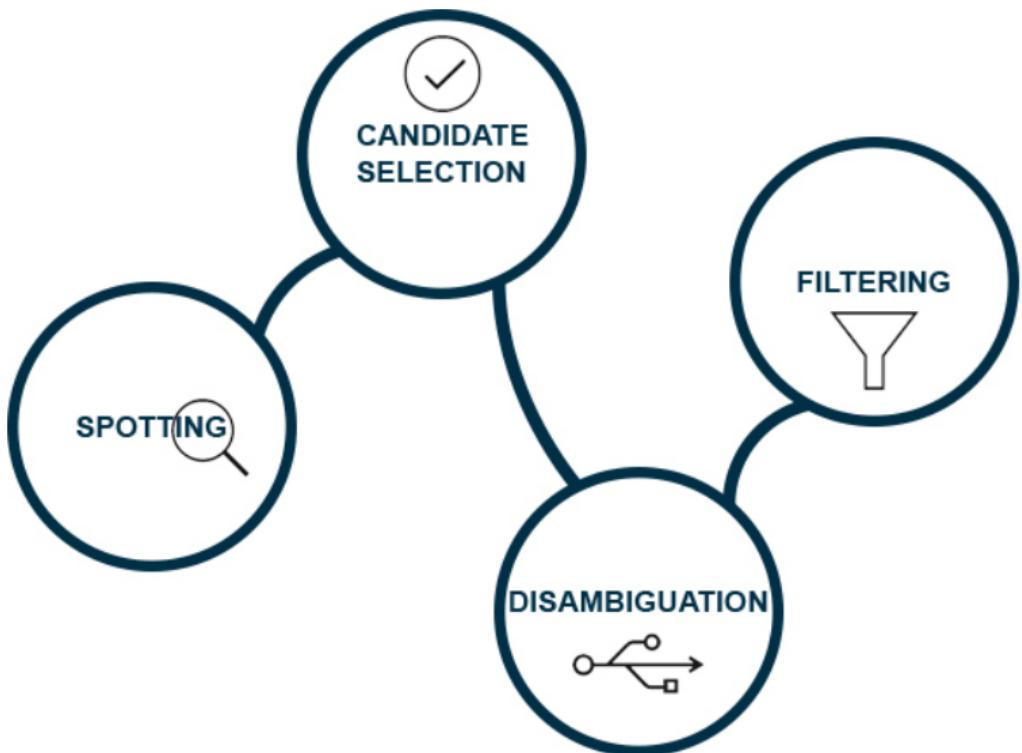
[About](#)   [Documentation](#)   [Acknowledgements](#)   [Publications](#)   [Contact](#)



## How Does the Tool Work?

With a four step approach, DBpedia Spotlight performs named entity extraction, including [entity detection](#) and [name resolution](#). It can also be used for [named entity recognition](#), amongst other [information extraction](#) tasks.

[Learn More](#)



<https://www.dbpedia-spotlight.org/>

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

## Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base

## Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

# DBpedia Spotlight: Disambiguation

René Witte



The screenshot shows the DBpedia Spotlight interface. At the top is the DBpedia Spotlight logo, which includes a stylized orange and yellow brain-like graphic above the text "DBpedia Spotlight". Below the logo are several input fields and controls. On the left, there is a "Confidence:" slider set to 0.15, a checked checkbox for "n-best candidates", and a text input field containing the sentence "Paris Hilton is visiting the Hilton in Paris.". To the right of these are "Language: English" dropdown menus and "SELECT TYPES..." and "ANNOTATE" buttons. At the bottom right of the main area is a "BACK TO TEXT" button. A note at the bottom states: "This demo uses the statistical DBpedia Spotlight web service at <https://api.dbpedia-spotlight.org/en>".

→ Worksheet #6: Task 10

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base

### Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

## Approach

Spotlight uses a **statistical model** for disambiguation

- Using the context of a word/phrase
- Learned from the Wikipedia pages including the links (for multiple languages)

A 2012 paper reports a 85.1% accuracy for English (with 7.4% missing links)

## Discussion

- Modern approaches for disambiguation use **word embeddings** (discussed later in the course), with better performance
- However, *Spotlight* is a stable, easy-to-install tool and useful for first experiments

*Improving Efficiency and Accuracy in Multilingual Entity Extraction,*  
<https://doi.org/10.1145/2506182.2506198>

[Introduction](#)

[Intelligent Conversational Agents](#)  
[Examples](#)  
[Design process](#)

[Search-Based Bots](#)

[Pattern-Matching Bots](#)

[Grounding-Based Bots](#)  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base

[Linking Documents and Knowledge Graphs](#)

[Conclusions](#)

[Example: Watson](#)  
[Summary](#)

[Notes and Further Reading](#)

# Outline

René Witte



## 1 Introduction

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

## 2 Search-Based Bots

### Search-Based Bots

## 3 Pattern-Matching Bots

### Pattern-Matching Bots

## 4 Grounding-Based Bots

### Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## 5 Conclusions

### Conclusions

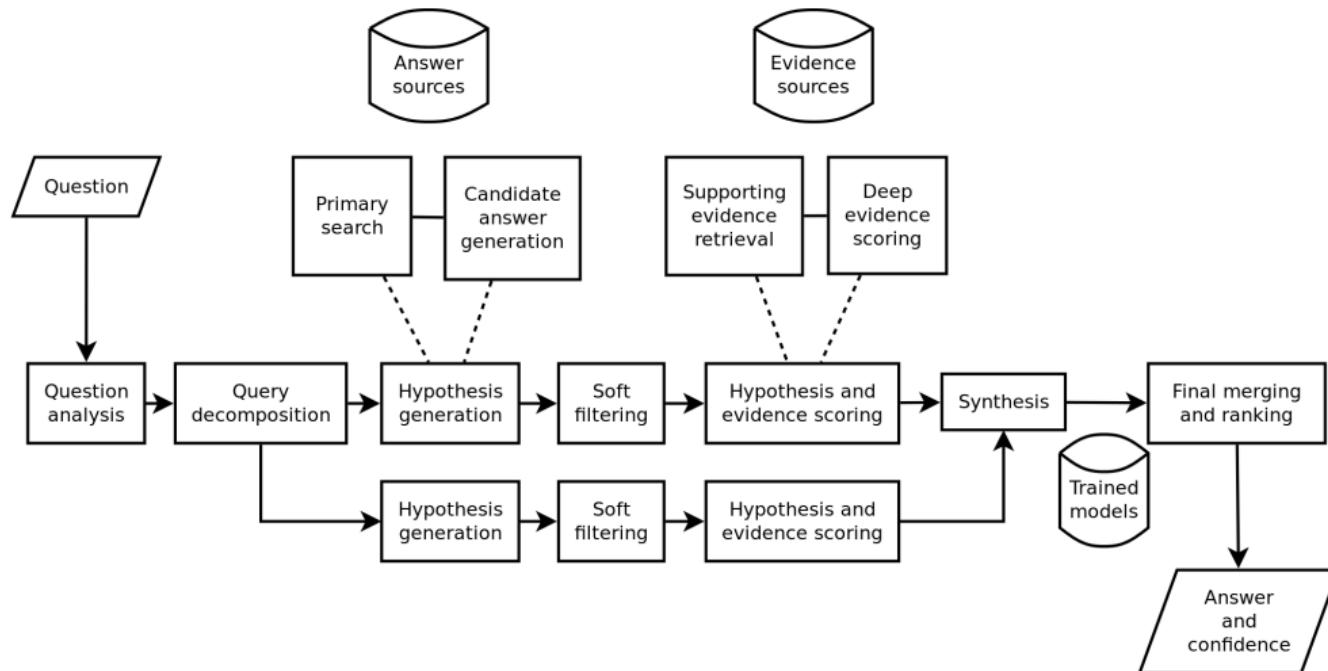
Example: Watson  
Summary

Example: Watson  
Summary

## 6 Notes and Further Reading

### Notes and Further Reading

Watson is a type of question-answering (QA) system, first developed 2006–2011



#### Introduction

- Intelligent Conversational Agents
- Examples
- Design process

#### Search-Based Bots

#### Pattern-Matching Bots

#### Grounding-Based Bots

- Question Classification
- NLP Pipelines
- Designing the Knowledge Base
- Linking Documents and Knowledge Graphs

#### Conclusions

- Example: Watson
- Summary

#### Notes and Further Reading

2011 *Jeopardy!* competition: 2,880 POWER7 threads and 16 terabytes of RAM

Introduction

How Watson Answers a Question in Four Steps

00,000 ms

Actual Processing Time



Introduction

Intelligent Conversational Agents

Examples

Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification

NLP Pipelines

Designing the Knowledge Base

Linking Documents and Knowledge Graphs

Conclusions

Example: Watson

Summary

Notes and Further

Reading

The first person mentioned by name  
in 'The Man in the Iron Mask'  
is this hero of a previous book  
by the same author.



#IBMWatson

IBM Watson: The Science Behind an Answer

[https://www.youtube.com/watch?v=ZbjTtCG3\\_X0](https://www.youtube.com/watch?v=ZbjTtCG3_X0)

# Summary of Chatbot Approaches

René Witte



| Approach   | Advantages   | Disadvantages  |
|------------|--|--|
| Grammar    | Easy to get started<br>Training easy to reuse<br>Modular<br>Easily controlled/restrained           | Limited “domain”<br>Capability limited by human effort<br>Difficult to debug<br>Rigid, brittle rules   |
| Grounding  | Answers logical questions well<br>Easily controlled/restrained                                     | Sounds artificial, mechanical<br>Difficulty with ambiguity<br>Difficulty with common sense<br>Limited by structured data<br>Requires large scale information extraction<br>Requires human curation |
| Retrieval  | Simple<br>Easy to “train”<br>Can mimic human dialog  | Difficult to scale<br>Incoherent personality<br>Ignorant of context<br>Can’t answer factual questions  |
| Generative | New, creative ways of talking<br>Less human effort<br>Domain limited only by data<br>Context aware | Difficult to “steer”<br>Difficult to train<br>Requires more data (dialog)<br>Requires more processing to train   |

Copyright 2019 by Manning Publications Co., [UH#19]

## Introduction

Intelligent Conversational Agents  
Examples  
Design process

## Search-Based Bots

## Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## Conclusions

Example: Watson  
Summary

## Notes and Further Reading

→ Worksheet #6: Task 11

# Outline

René Witte



## 1 Introduction

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

## 2 Search-Based Bots

### Search-Based Bots

### Pattern-Matching Bots

### Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

## 3 Pattern-Matching Bots

### Conclusions

Example: Watson  
Summary

## 4 Grounding-Based Bots

### Notes and Further Reading

## 5 Conclusions

## 6 Notes and Further Reading

## Required

- [LHH19, Chapter 12] (Dialog engines)

## Supplemental

- [IMF13, Chapter 8] (Building a QA System)
- [RN10, Chapter 26] (Philosophical Foundations)

### Introduction

Intelligent Conversational Agents  
Examples  
Design process

### Search-Based Bots

### Pattern-Matching Bots

Grounding-Based Bots  
Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

### Conclusions

Example: Watson  
Summary

### Notes and Further Reading

## References

René Witte



- [HHDF20] Marvin Hofer, Sebastian Hellmann, Milan Dojchinovski, and Johannes Frey.  
The New DBpedia Release Cycle: Increasing Agility and Efficiency  
in Knowledge Extraction Workflows.  
In *Semantic Systems. In the Era of Knowledge Graphs*, pages 1–18,  
Cham, 2020. Springer International Publishing.  
[https://link.springer.com/chapter/10.1007/978-3-030-59833-4\\_1](https://link.springer.com/chapter/10.1007/978-3-030-59833-4_1).
- [IMF13] Grant S. Ingersoll, Thomas S. Morton, and Andrew L. Farris.  
*Taming Text: How to find, organise, and manipulate it.*  
Manning, 2013.  
<https://concordiauniversity.on.worldcat.org/oclc/772977853>.
- [LHH19] Hobson Lane, Cole Howard, and Hannes Max Hapke.  
*Natural Language Processing in Action.*  
Manning Publications Co., 2019.  
<https://concordiauniversity.on.worldcat.org/oclc/1102387045>.
- [RN10] Stuart Russell and Peter Norvig.  
*Artificial Intelligence: A Modern Approach.*  
Prentice Hall, 3rd edition, 2010.  
<https://concordiauniversity.on.worldcat.org/oclc/359890490>.

Introduction

Intelligent Conversational Agents  
Examples  
Design process

Search-Based Bots

Pattern-Matching Bots

Grounding-Based Bots

Question Classification  
NLP Pipelines  
Designing the Knowledge Base  
Linking Documents and Knowledge Graphs

Conclusions

Example: Watson  
Summary

Notes and Further Reading