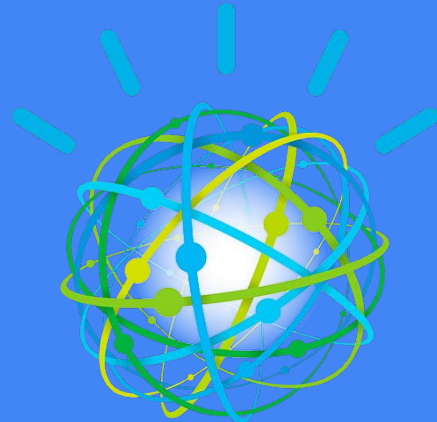
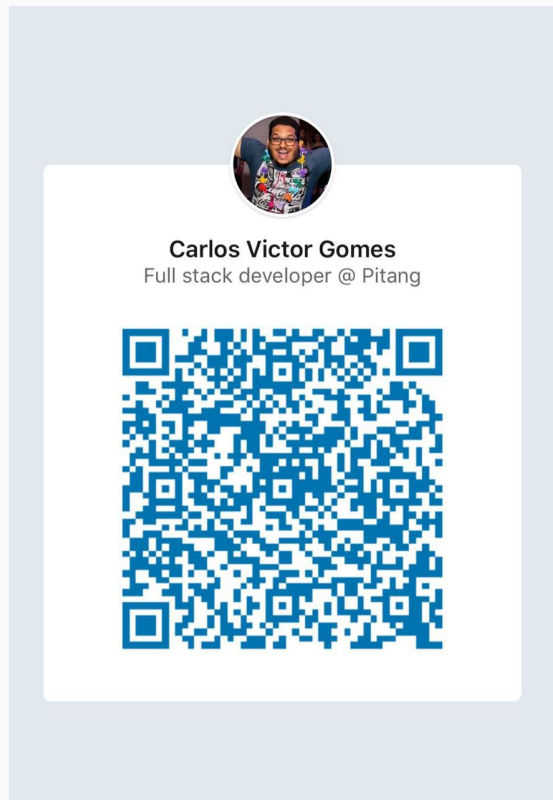


CURSO IBM WATSON

Carlos Gomes

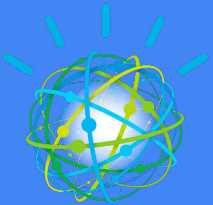


about.me :)



- 9 anos de experiência como Desenvolvedor Fullstack;
- No momento estou trabalhando na **Pitang** no projeto da **SKY** com foco no desenvolvimento de **Chatbots**;
- Também realizo consultoria e desenvolvimento de sistemas pela minha própria empresa **CactoWeb**;

VISÃO GERAL DO CURSO



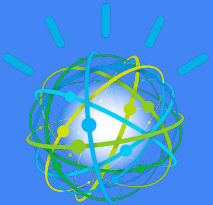
O que nós vamos aprender?

- Seção 1:
 - Apresentação da plataforma IBM Watson;
- Seção 2:
 - Introdução ao IBM Watson;
- Seção 3:
 - Criar sistemas de conversação usando o serviço de conversação do watson;
- Seção 4:
 - Treinar modelos de reconhecimento de imagem utilizando o Watson Visual Recognition;

Pré-requisitos

- Computador: Windows, Mac ou Linux;
- Conhecimento básico de APIs RESTful, JSON, Node.js;
- Conta no IBM Cloud;
- Não é necessário background de *Machine Learning* ou *NLP*;
- Paciência 😊
- Vontade de aprender;

APRESENTAÇÃO DA PLATAFORMA IBM WATSON



Seção 1

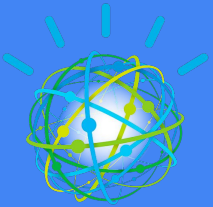
- Fundamentos;
- Apresentando o IBM Watson;
- A plataforma IBM Watson;
- Adaptando o Watson;
- Exemplos;

Alguns alertas

A plataforma **Bluemix** agora se chama **IBM Cloud**;

O **Watson Conversation** é agora o **Watson Assistant**;

FUNDAMENTOS

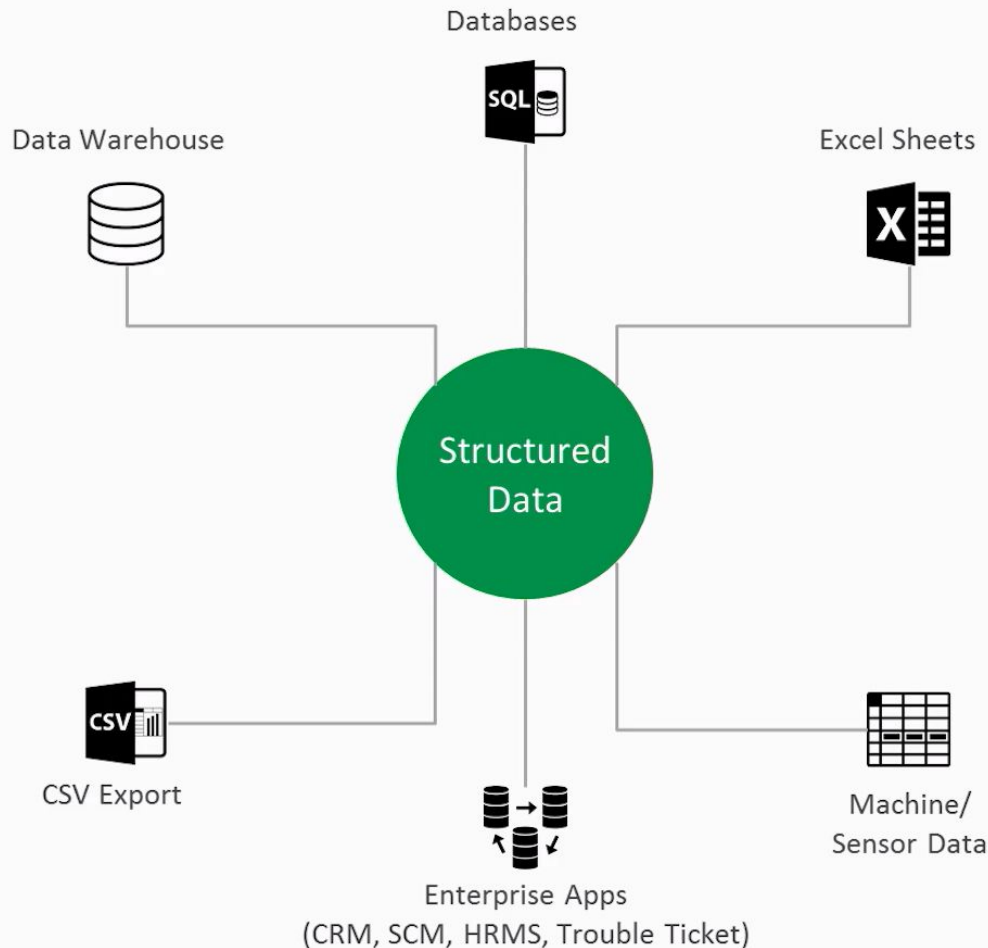


Fundamentos

- Dados estruturados;
- Dados não estruturados;
- *Machine Learning*;
- Recursos e Engenharia de Recursos (*Feature & Feature Engineering*);

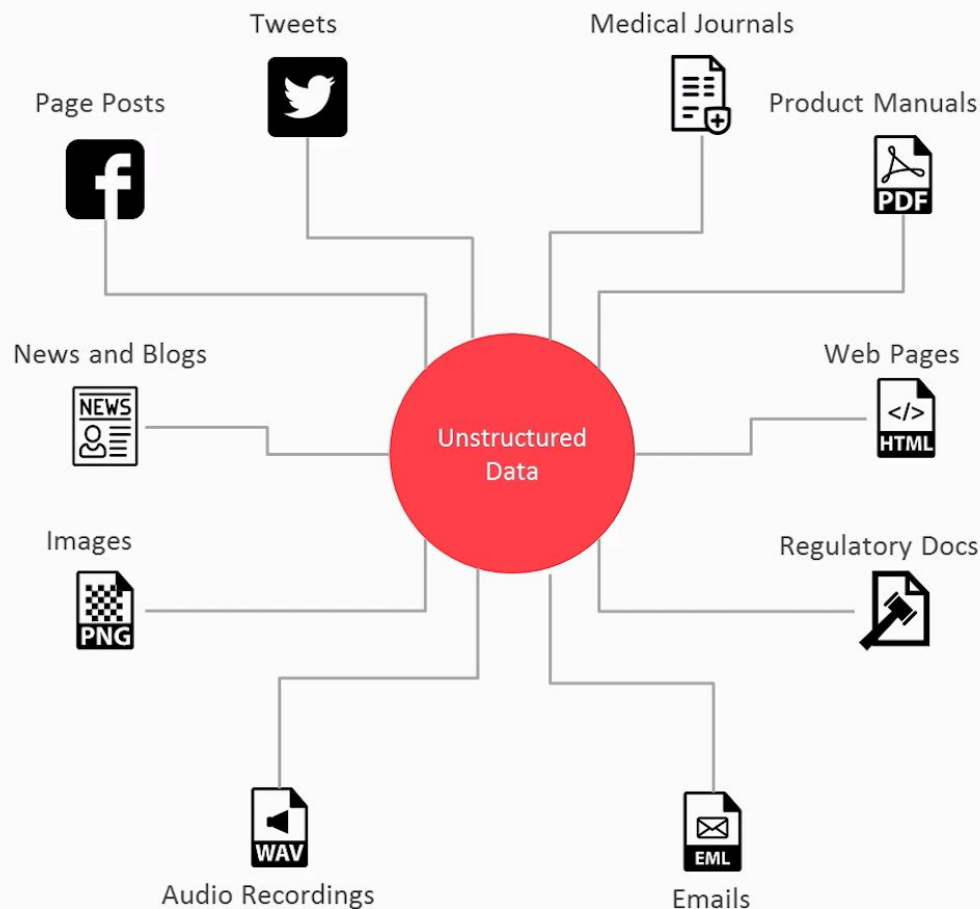
Dados estruturados

- Dados estruturados são destinados ao consumo de máquinas/computadores;
- Eles têm um alto grau de organização e um modelo de dados rigoroso;
- Técnicas conhecidas para armazenar, processar e analisar



Dados não estruturados

- Dados não estruturados são destinados ao consumo humanos;
- Eles não têm necessariamente um modelo de dados formal predefinido;
- Difícil processamento por máquinas;



Comparativo - Dados estruturados x Não estruturados

Coluna	Valor
Fabricante	Ford
Modelo	Mustang
Ano	2008
Data do acidente	11/03/2019
Idade do motorista	35
Gênero do motorista	M

A investigação no local identificou um **Ford** **Mustang** **2008** envolvido em um acidente. O acidente ocorreu às 11:53 de **11 de março de 2019**. O carro estava sendo conduzido por um **homem** de **35 anos**.

90%

Dos dados digitais do mundo são **Não estruturados**

Fonte: 2014 Digital Universe report by IDC/EMC



4,169,819,603

Internet Users in the world



1,671,864,751

Total number of Websites



5,957,980,640

Emails sent **today**



150,643,977

Google searches **today**



142,892

Blog posts written **today**



17,335,769

Tweets sent **today**



160,514,896

Videos viewed **today**
on YouTube



1,859,513

Photos uploaded **today**
on Instagram



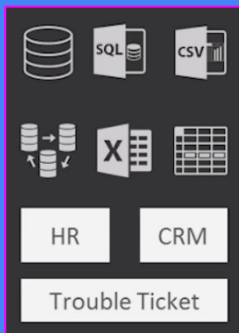
3,094,400

Tumblr posts **today**

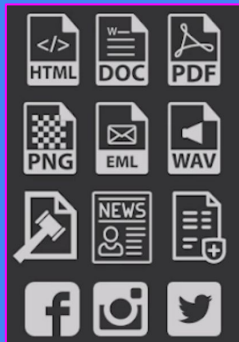
Fonte: screenshot de <http://www.internetlivestats.com/> tirado no dia 10 de Março de 2019

Combine e transforme volumes crescentes de dados estruturados e não estruturados em novos insights facilmente acessíveis

Dados estruturados



Dados não estruturados



Combine, consuma,
limpe e
compreenda todos
os tipos de dados



Treine modelos
com o seu
conhecimento de
domínio



Insights + Resultados

Obtenha insights profundos
e produza resultados
poderosos.

Treine sua IA sobre o que
é importante para sua
indústria tomar as
decisões mais bem
informadas.

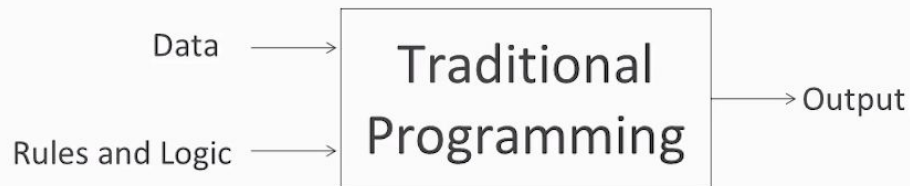
Machine Learning é o campo de estudo que dá aos computadores a capacidade de aprender sem serem programados explicitamente.

- Arthur Samuel

Machine Learning é
essencialmente a ciência de
aprender representações a partir
de dados para fazer previsões e
inferências

Machine Learning

- Na programação tradicional, regras e lógica são explicitamente especificadas;
- É extremamente complexo e difícil de programar à medida que a complexidade dos dados aumenta;
- Um sistema de aprendizagem de máquina é "treinado" em vez de programado explicitamente;
- São apresentados muitos exemplos relevantes para que encaixe em um modelo;



Features (Recursos)

- Recursos representam a característica ou um atributo de um exemplo de treinamento usado para um modelo de aprendizado de máquina;
- Quando um algoritmo de aprendizado de máquina ajusta um modelo aos dados, ele está essencialmente encontra os relacionamentos estatísticos entre os recursos e as variáveis;

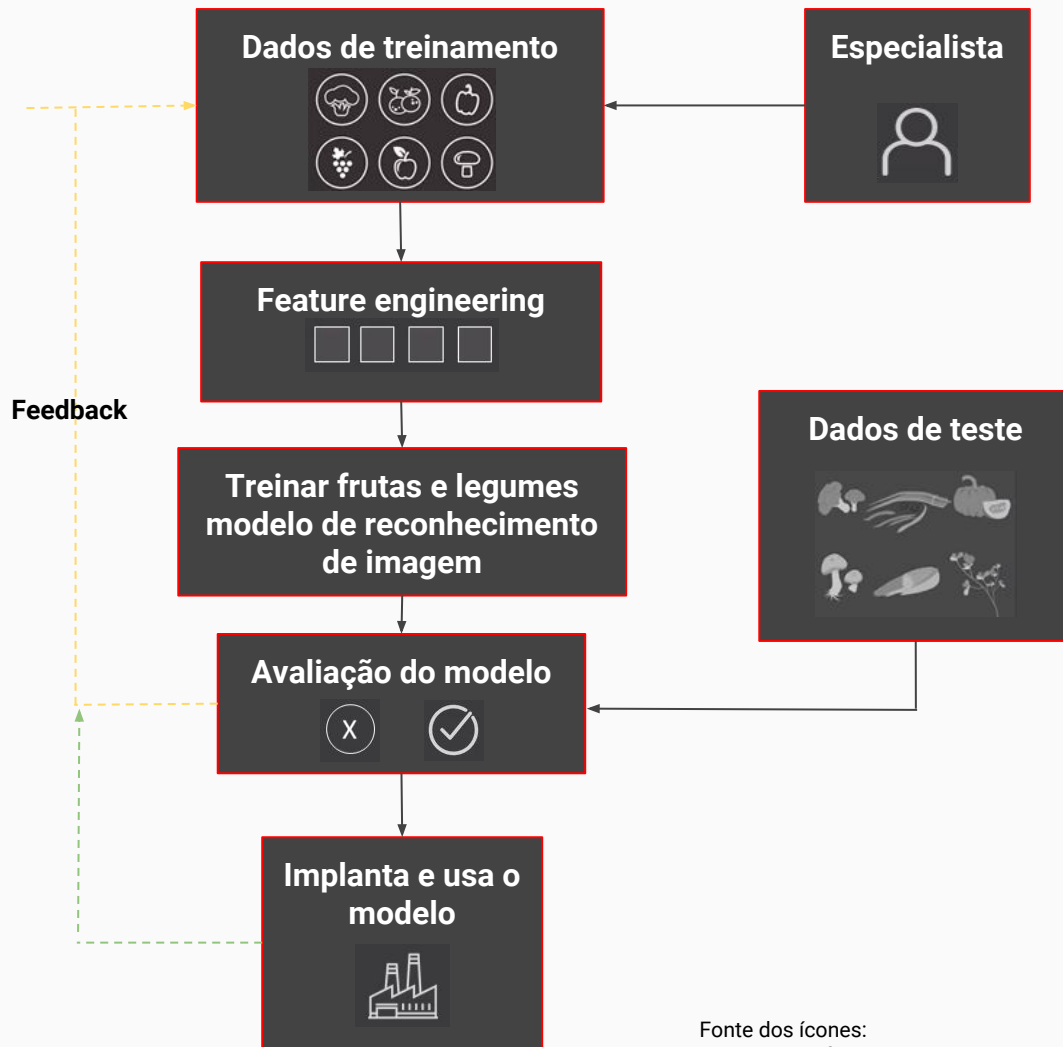
Por exemplo, uma aplicação de ML tentando determinar a probabilidade de doença cardíaca em pacientes.

Quais são alguns dos possíveis recursos?

Gênero	Idade	Altura	Peso	Pressão Sanguínea	Frequência cardíaca em repouso	Histórico médico
--------	-------	--------	------	-------------------	--------------------------------	------------------

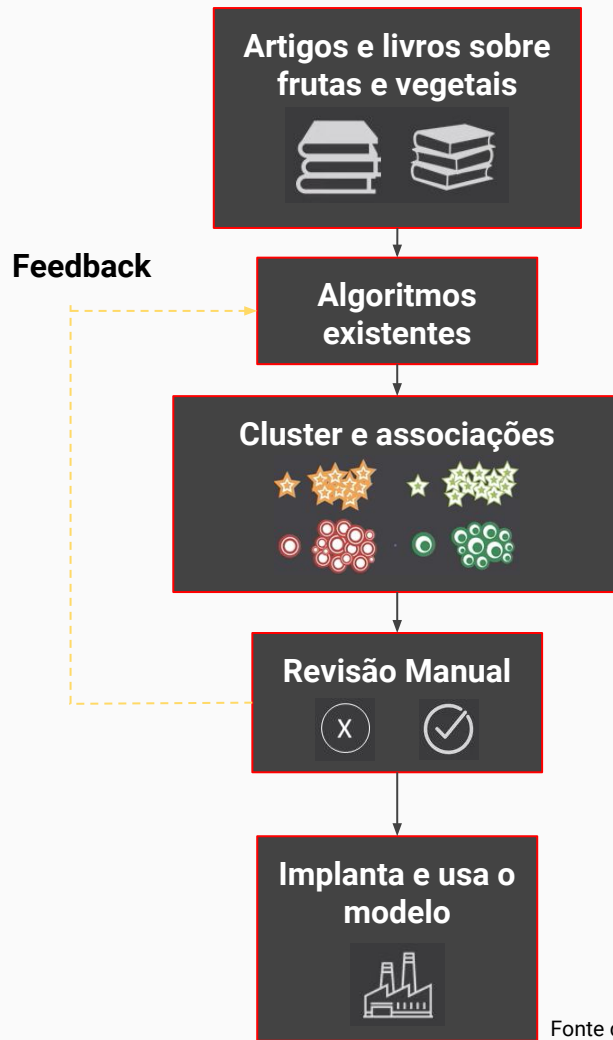
Aprendizagem supervisionada

- Com o aprendizado supervisionado, você fornece dados de treinamento rotulados - entradas e saídas desejadas;
- Um "professor" ou um "especialista no assunto" fornece os dados de treinamento rotulados - uma verdade básica (positivos e negativos);
- Exemplos:
 - Classificando e-mails de clientes
 - Identificando partes de máquinas quebradas a partir de imagens;
 - Compreendendo intenções de linguagem natural;
 - Extração de entidades e relacionamentos de texto
 - Identificando transações de cartões de crédito fraudulentas;



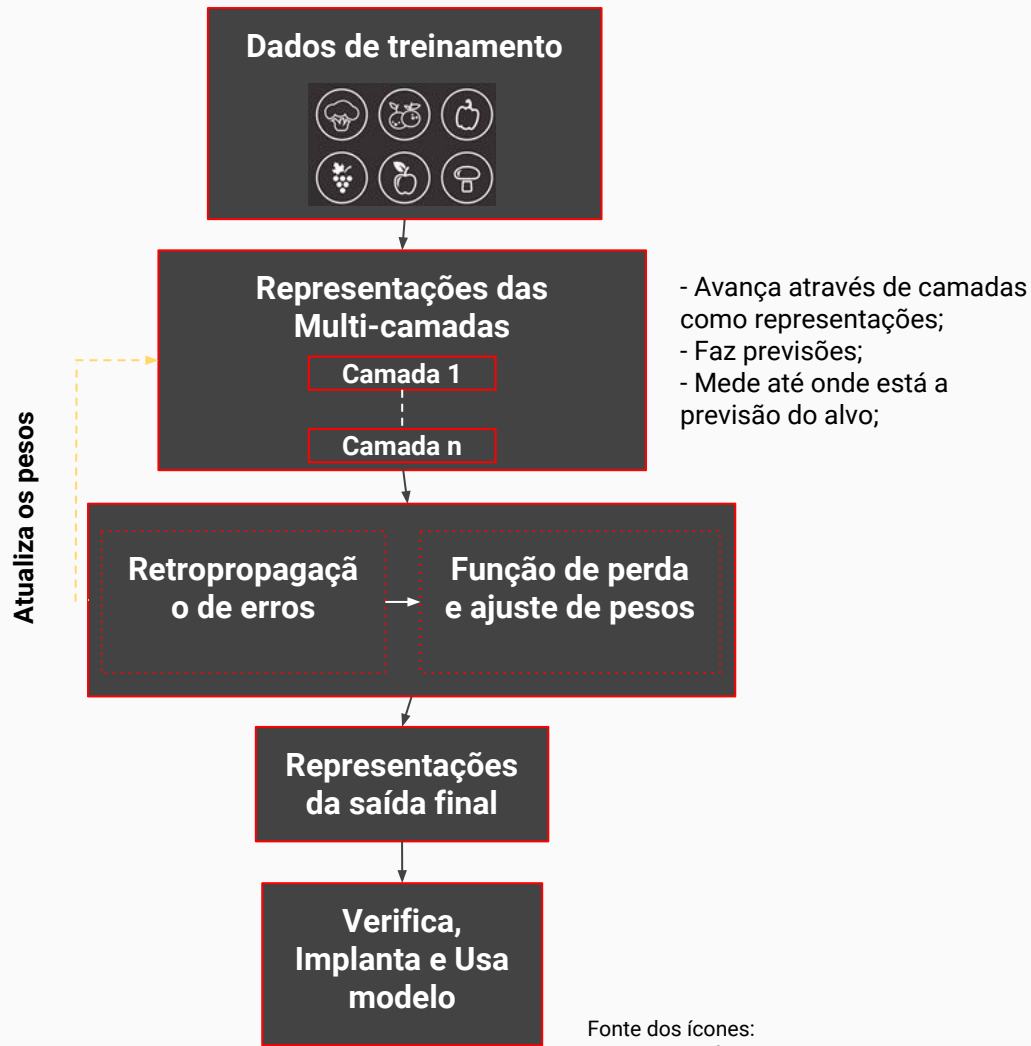
Aprendizagem não supervisionada

- Com aprendizado não supervisionado, o objetivo é aprender padrões no conjunto de dados de entrada;
- Nenhum dado rotulado ou de saída é dado ao algoritmo;
- O objetivo é entender, identificar e inferir estrutura, novos padrões e relacionamentos nos dados;
- Nenhum recurso é mais importante que outros;
- Exemplos:
 - Análise de personalidade psicográfica;
 - Tendências e desvios nas análises de produtos;
 - Recomendações de produtos;
 - Tópicos-chave de um grande número de e-mails;



Deep learning

- *Deep learning* é um subcampo específico de *Machine learning*;
- - O "deep" do *deep learning* refere-se à idéia de camadas sucessivas para representações cada vez mais significativas;
- - *Deep learning* ainda requer exemplos rotulados, mas automatiza completamente a etapa de *feature engineering* (engenharia de recursos);



A API do IBM Watson usa um ou vários algoritmos de aprendizado de máquina e aprendizagem profunda com base na tarefa em questão

Ao treinar a API do IBM Watson para seu próprio domínio ou necessidades específicas, ela cuida da seleção das features necessárias para treinar o modelo.

Em outras palavras, você não estará fazendo feature engineering ao treinar.

Apresentando o IBM Watson

- O que é computação cognitiva?
- Principais características de um sistema cognitivo;
- Tarefas comuns;
- Apresentando o IBM Watson;

Os sistemas cognitivos (também conhecidos como sistemas de inteligência artificial) utilizam uma ou mais tecnologias, como Processamento de Linguagem Natural (PNL), processamento de voz, visão computacional e aprendizado de máquina para permitir que pessoas e máquinas interajam naturalmente. Eles amplificam e aumentam a perícia e a cognição humana.

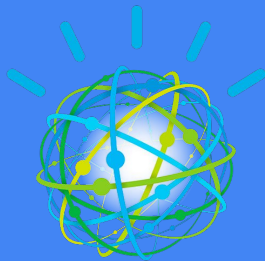
Principais características de sistemas cognitivos/IA



Algumas tarefas populares realizadas por sistemas cognitivos / IA

- Extração de entidade;
- Classificação de intenção;
- Análise de sentimento;
- Expansão de conceito;
- Extração de relacionamento;
- Análise de emoção;
- Detecção de linguagem;
- Resumo de texto;
- Classificação de imagem;
- Busca de imagem por similaridade;
- Transcrição de diálogo;
- Tradução de línguas;
- Ranking de relevância;
- Representação de conhecimento;
- Raciocínio de conhecimento;

O que é o IBM Watson?



O IBM Watson é uma plataforma de inteligência artificial para empresas que é composta por APIs pré-treinadas e treináveis



See



Hear



Speak



Converse



Feel

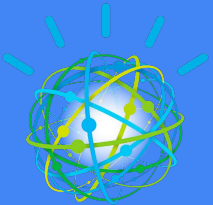


Translate



Discover

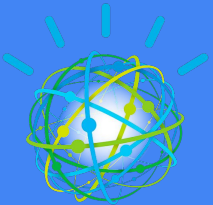
APRESENTANDO O IBM WATSON



Apresentando o IBM WATSON

- Como a API do IBM Watson está disponível;
- APIs do IBM Watson organizadas em 7 categorias;
- Como o Watson aprende;

PLATAFORMA IBM WATSON



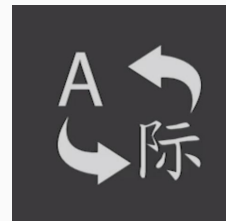
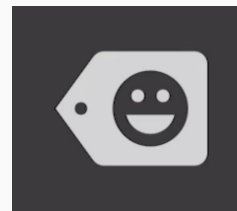
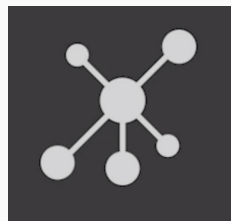
O **Watson** está disponível como um conjunto de **APIs** abertas no **IBM Cloud**.

O **Watson** no **IBM Cloud** permite que você:

- **integre** a poderosa **IA** em seus aplicativos;
- **armazene** os dados;
- **treine** as entradas;
- **gerencie** seus dados com **segurança**

Com Watson você pode:

- Criar chatbots de conversação e agentes virtuais utilizando diferentes idiomas;
- Criar rapidamente um mecanismo de análise cognitiva e de pesquisa cognitiva;
- Treinar classificadores de imagens customizados;
- Adicionar transcrição de fala às suas aplicações;
- Analisar texto para extrair conceitos, entidades, relacionamentos e muito mais;
- Compreender personalidade, emoções e tendências sociais a partir de textos escritos e chats;
- Treinar e adaptar o Watson para suas necessidades exclusivas de negócios e indústria;



Recursos de AI do IBM Watson - Organizados em 7 categorias e disponíveis como APIs no IBM Cloud

AI Assistant



Integrate diverse conversation technology into your application.

[Watson Assistant](#)

Build an AI assistant for a variety of channels, including mobile devices, messaging platforms, and even robots.

Vision



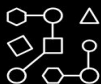
Identify and tag content then analyze and extract detailed information found in an image.

[Visual Recognition](#)

Tag and classify visual content using machine learning.

Recursos de AI do IBM Watson - Organizados em 7 categorias e disponíveis como APIs no IBM Cloud

Data



Collect, organize and analyze your data, then achieve trust, scale and automation across your full AI lifecycle.

[Watson Studio](#)

Build and train AI models, and prepare and analyze data, all in one integrated environment.

[Watson Machine Learning](#)

Use your data to create, train, and deploy self-learning models. Leverage an automated, collaborative workflow to build intelligent applications.

[Watson Knowledge Catalog](#)

Catalog, understand, analyze and share data to drive digital transformation and compliance at scale.

[Watson OpenScale](#)

Operate and automate AI at scale by infusing it with trust and transparency, explaining its outcomes, and eliminating harmful bias.

Recursos de AI do IBM Watson - Organizados em 7 categorias e disponíveis como APIs no IBM Cloud

Knowledge



Get insights through accelerated data optimization capabilities.

[Discovery](#)

Unlock hidden value in data to find answers, monitor trends and surface patterns.

[Discovery News](#)

Access pre-enriched news content in real-time.

[Natural Language Understanding](#)

Natural language processing for advanced text analysis.

[Knowledge Studio](#)

Teach Watson to discover meaningful insights in unstructured text.

Recursos de AI do IBM Watson - Organizados em 7 categorias e disponíveis como APIs no IBM Cloud

Speech



Convert text and speech with the ability to customize models.

[Speech to Text](#)

Easily convert audio and voice into written text.

[Text to Speech](#)

Convert written text into natural-sounding audio.

Language



Analyze text and extract meta-data from unstructured content.

[Language Translator](#)

Translate text from one language to another.

[Natural Language Classifier](#)

Interpret and classify natural language with confidence.

Recursos de AI do IBM Watson - Organizados em 7 categorias e disponíveis como APIs no IBM Cloud

Empathy



Understand tone, personality, and emotional state.

[Personality Insights](#)

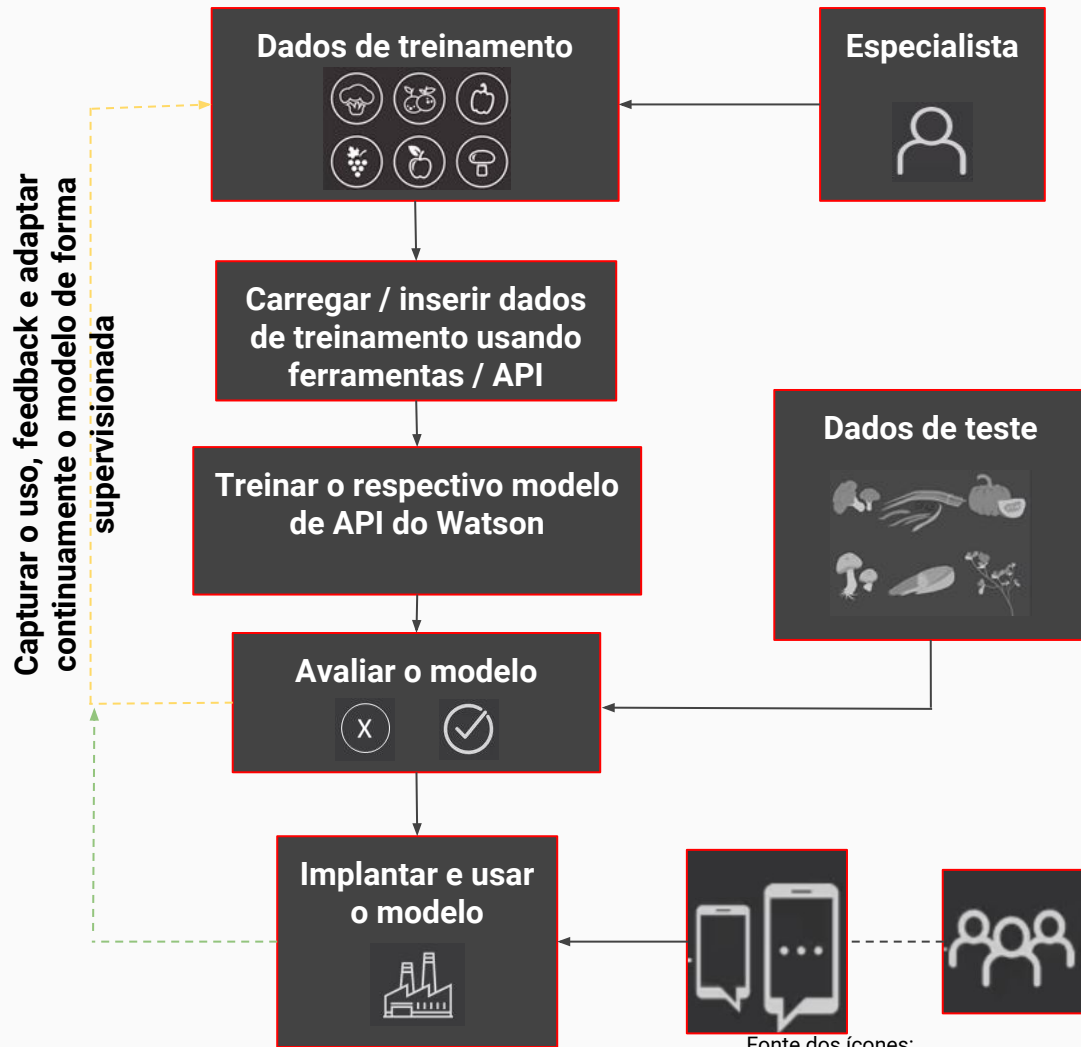
Predict personality characteristics through text.

[Tone Analyzer](#)

Understand emotions and communication style in text.

Como o Watson aprende?

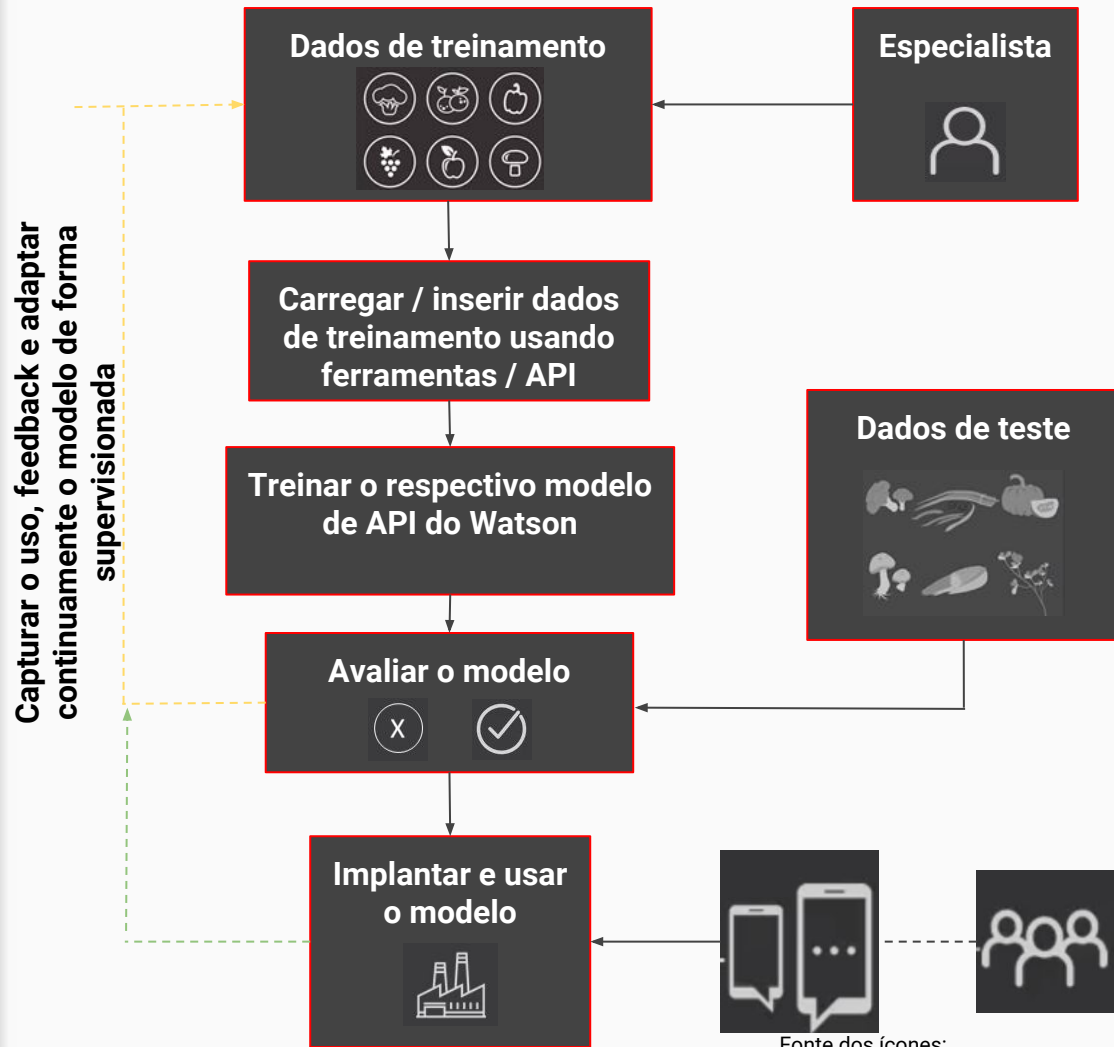
- A maioria das APIs do Watson é pré-treinada.
- Nas APIs que podem ser treinadas, precisamos da verdade fundamental (rótulos) para treinar as respectivas capacidades de maneira supervisionada;
- A verdade dos dados é construída e usada para treinar + testar o modelo;
- Os tipos de dados dependem da função ou tarefa do AI:
 - A classificação de texto ou intenção usa enunciados em linguagem natural como dados de treinamento;
 - O Reconhecimento Visual do Watson usa imagens marcadas positivas e negativas relevantes para treinar um classificador personalizado;
 - O Watson Discovery precisa de conteúdo que seja ingerido no que é conhecido como corpus;
- Toda API do Watson usa um ou vários algoritmos ML ou DL diferentes;



Fonte dos ícones:
<https://www.flaticon.com/>

Como o Watson aprende?

- Tudo começa com desenvolvedores e especialistas de domínio usando as respectivas ferramentas de serviço do Watson, ingerindo o conteúdo;
- Nos bastidores, vários recursos são extraídos e usados para derivar características dos dados de treinamento para gerar o modelo;
- O modelo é avaliado usando um conjunto de testes e, eventualmente, implementado;
- Um aplicativo aproveita o modelo implantado;
- O feedback do modelo implantado é coletado para ajustar continuamente a verdade básica e, portanto, reciclar o modelo;
- A IBM não compartilha esse modelo treinado nem os insights obtidos com ninguém;



Fonte dos ícones:
<https://www.flaticon.com/>

Como o Watson aprende?

O Watson ingere grandes quantidades de dados estruturados e não estruturados

Treinar dados para conceitos, significados semânticos e sutilezas na linguagem

O Watson aprende com os insumos para entender a linguagem natural e extrapolar significados

O cachorro correu pela rua. Ele recebeu um biscoito do seu dono.

O cachorro correu pela rua. Ele recebeu um biscoito do seu dono.

Animal

Ação

Localização

Comida

Pessoa

O hamster correu em sua roda de exercícios. Comeu uma cenoura.

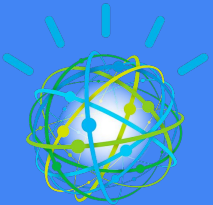
Animal

Ação

Localização

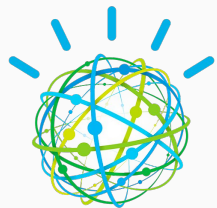
Comida

ADAPTANDO O WATSON



Adaptando o Watson

- Adaptação de domínio do Watson;
- Como os sistemas cognitivos são treinados;



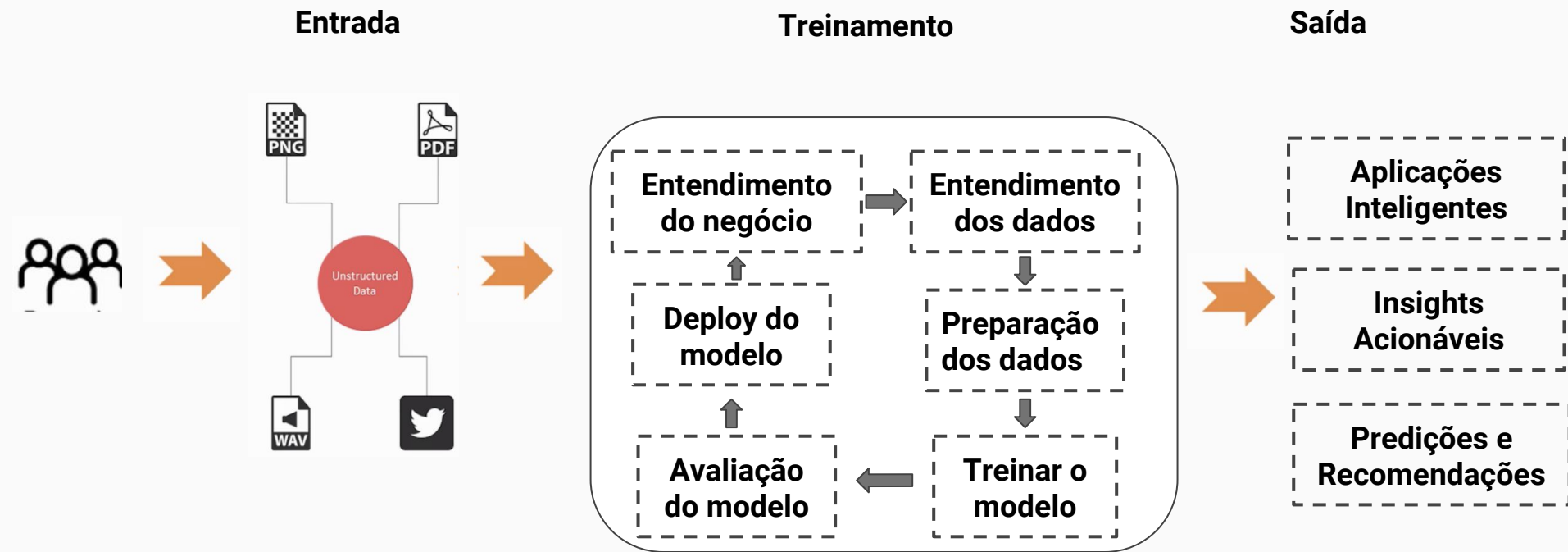
Adaptação de domínio do Watson

- As APIs do Watson vêm pré-treinadas com conhecimento de domínio aberto;
- O Watson pode ser treinado em conjuntos de dados específicos da indústria e do setor;
- Especialistas em domínio treinam as respectivas APIs do Watson para extrair e descobrir insights significativos em textos não estruturados;
- Especialistas em domínio transmitem conhecimento especializado ao Watson e, portanto, o aprendizado acontece não apenas de uma maneira geral;

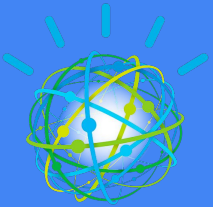
Adaptação de domínio do Watson

- O treinamento começa com uma verdade básica de boa qualidade e normalmente envolve sucessivas rodadas de insumos, treinamento, avaliação e feedback dos especialistas, bem como dos usuários representativos;
- Este exemplo treinado do Watson se torna próprio e não é compartilhado com outros;

Como sistemas cognitivos são treinados ?



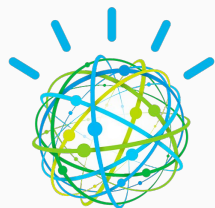
EXEMPLOS



Exemplos de aplicações utilizando Watson

- STAPLES:

- Usando o Assistente do Watson e outros serviços do Watson, a Staples transformou o "Easy Button" em um ecossistema de pedido cognitivo. O Staples Easy System agora remove as dificuldades do processo de pedido, permitindo que os gerentes de escritório façam pedidos onde, quando e como quiserem.
- Fonte: <https://www.ibm.com/watson/infographic/staples/>

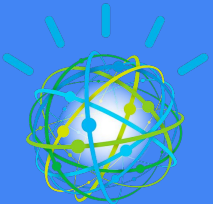


Exemplos de aplicações utilizando Watson

- WOODSIDE:
 - Antes do Watson, os engenheiros da Woodside gastavam até 80% do seu tempo tentando descobrir possíveis soluções ou riscos - e apenas 20% de seu tempo no trabalho real de engenharia. Com o Watson, o tempo gasto na pesquisa foi reduzido em 75%, porque o Watson tem fácil acesso a decisões de conhecimento prévio e aprendizados desenvolvidos pelos próprios funcionários da Woodside.
 - Fonte: <https://www.ibm.com/watson/stories/woodside/>



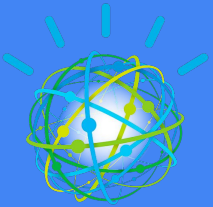
INTRODUÇÃO AO IBM WATSON



Seção 2

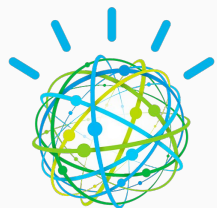
- APIs Watson;
- IBM Cloud;
- Hello Watson Assistant;
- Ambiente de desenvolvimento;

APIs WATSON

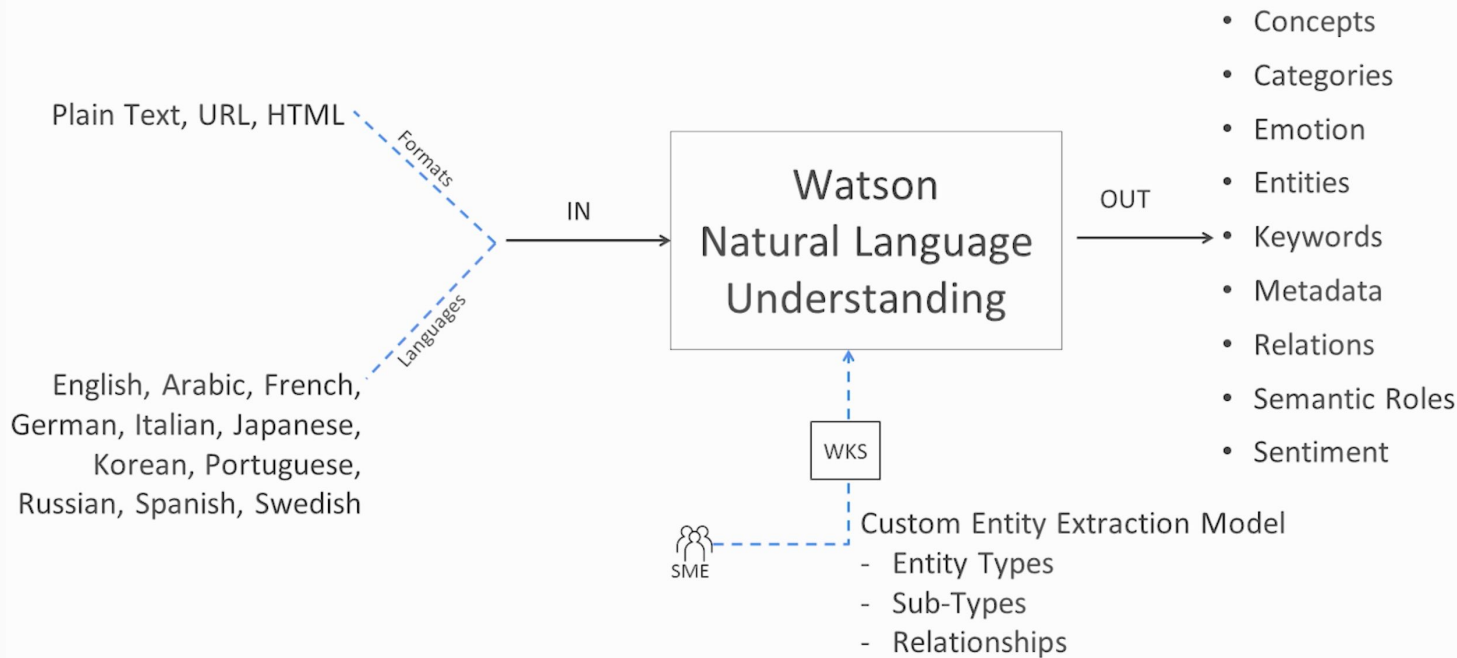


Test Drive

- Watson Natural Language Understanding:
 - <https://www.ibm.com/watson/services/natural-language-understanding>



Natural Language Understanding



IBM Cloud

- Crie sua conta ou faça login;
- No menu navegue até a aba Watson
- Crie seu assistente;
- Vamos fazer um hello world;

Links úteis:

- <http://console.bluemix.net>
- <https://console.bluemix.net/developer/watson/dashboard>
- <https://console.bluemix.net/catalog/services/conversation>

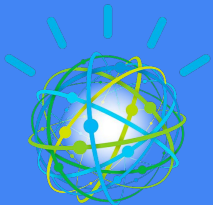
Ambiente de desenvolvimento

- Instalação do NodeJS;
- Utilização do Watson SDK com NodeJS;
- Teste da API do NLU do Watson com NodeJS;
- Lembrando que os SDKs do Watson estão disponíveis para várias linguagens (Python, NodeJS, Java, .NET, Swift, entre outras);

Links úteis:

- <https://github.com/creationix/nvm>
- <https://github.com/watson-developer-cloud>
- <https://gist.github.com/carlosvictor/258bb409756aec70dd1feab6c7878d39>

CRIANDO SISTEMAS DE CONVERSAÇÃO UTILIZANDO WATSON ASSISTANT



Seção 3

- Aprofundar no Watson Assistant;
- Definir intenções e entidades;
- Construir diálogos;
- Avaliar e implantar o modelo;
- Construir um exemplo;
- Melhorar modelos continuamente;

Aprofundando no Watson Assistant

- Sistemas de conversação;
- Noções básicas de Chatbots;
- O que a API faz?
- Capacidades e idiomas suportados;

Sistemas de conversação

- Os sistemas de conversação permitem que os usuários finais se envolvam usando suas formas naturais de expressão:
 - escrita
 - falada
 - visual
 - gestos
- E cumpram suas intenções;

Sistemas de conversação (Ex: Chatbot)

Hi, I am Watson! I can answer questions about common topics for the Westhaven Community Residents. I'm a chronic learner and please help me get better. Ask me something like *"What is the HoA phone number?"*

When are the trash pick up days?

- TRASH and BULK collection days are every Monday and Thursday.
- Trash should be placed in trash bags or trash cans with lids and placed at the end of your driveway prior to 7 am.
- Residential customers are allowed a maximum 10 bags of solid waste per service day. Each bag should not weigh more than 40-pounds.
- Recycling collection day is every Wednesday.

Can you recommend a good plumber?

Recommended plumber is L&S Plumber +1 (877) 577-██████

Os usuários finais interagem e fazem sua pergunta em linguagem natural

- Apresentação do assistente;
- Usuário se expressando;
- Resposta baseada na identificação da intenção do usuário

Sistemas de conversação (1/3)

- Para compreender o que os usuários finais estão dizendo e responder adequadamente ao contexto os **sistemas de conversação** utilizam algumas técnicas, são elas:
 - Processamento de Linguagem Natural;
 - *Machine Learning*;
 - *Deep learning*;
 - Reconhecimento de voz;
- Lembrando que essas técnicas não são as únicas utilizadas, além disso elas também podem ser combinadas;

Sistemas de conversação (2/3)

- As respostas fornecidas por esses sistemas podem variar:
 - Respostas curtas de texto;
 - Passagens relevantes de um grande corpo de texto;
 - Imagens;
 - Vídeos;
 - Uma resposta da chamada de uma API;
 - Execução de ações em nome do usuário;
 - Chamada de algum tipo de processo;

Sistemas de conversação (3/3)

- Devemos lembrar que os sistemas de conversação são previamente treinados e eles também devem aproveitar o contexto do usuário final para fornecer respostas ainda mais significativas e relevantes.

Localização

Idade

Profissão

Gênero

Principais conceitos relacionados a Sistemas de conversação

- Aqui temos alguns termos que sempre estão presentes quando começamos a falar de sistemas de conversação:

Expressões

Bate-papo

Precisão de
classificação

Agentes de
resposta

Análise de
Interação

Fora do assunto

Entidades

Contexto

Diálogo

Intenções

Desambiguação

Principais características de um bom sistema de conversação (1/2)

- Algumas características são importantíssimas na construção de um bom **sistema de conversação**, são elas:
 - Aproveitamento do contexto do usuário, lembra-o da conversa e minimiza a repetição;
 - Utilização de vários agentes de atendimento;
 - Suporte a múltiplas formas de entrada (texto, fala, etc...);
 - Consegue lidar com coloquialismo, erros de ortografia e frases mal formadas;
 - Consegue executar ações em nome dos usuários;
 - Permite interações através de vários canais (Web, Slack, Facebook, Skype, etc...);
 - Consegue identificar a precisão e dizer "não sei" ao invés de tentar fornecer uma resposta errada;
 - Consegue lidar com a conversa de acordo com tom de engajamento do usuário;

Principais características de um **bom** sistema de conversação (2/2)

- Continuando as características...
 - Consegue aproveitar UX específicas dos canais para tornar a experiência mais fácil e agradável para o usuário;
 - Consegue fornecer respostas precisas, curtas e significativas;
 - Consegue suportar desvios do usuário e restabelece a voltar ao tópico principal;
 - Consegue integrar-se com apps de backend, APIs, processos de negócios e etc;
 - Consegue realizar desambiguação das consultas do usuário utilizando o contexto;
 - Possui fluxos de escape embutidas e desvia para especialistas humanos quando necessário;
 - Consegue lidar com **bate-papo** e perguntas **fora do assunto** bem;
 - Consegue aprender (de forma supervisionada) a partir de interações e feedback;

O que o Watson Assistant permite que você faça?

- Crie assistentes;
- Sistemas de conversação baseados em AI
- Conexão para uma variedade de canais;

- Treinar intenções;
- Extrair entidades;
- Fluxos de Diálogo;

- Lidar com a desambiguação;
- Manter e gerenciar o contexto;
- Slots para coletar informações;

- Faça chamadas de API a partir do diálogo;
- Bate-papo e textos longos;
- Implante em vários canais;

- Catálogo de conteúdo pré-treinado;
- Análise de interação;
- Seguro, criptografado;

Recursos do Watson Assistant, aproveitam técnicas de PNL e de *machine/deep learning*



Mais uma vez...

Definição do **Watson Assistant** (1/4)

- Uma das APIs do Watson para treinar e construir sistemas ou aplicativos de conversação baseados em AI;
- Combina *machine learning*, *natural language understanding* e ferramentas integradas para treinar modelos para entender usuários, extrair entidades e criar fluxos de diálogo de conversação;
- Os modelos de compreensão / classificação de intenção podem ser treinados com um pequeno número de exemplos representativos;
- Identificador/extrator entidades personalizadas usando uma combinação de abordagens baseadas em regras e baseadas em padrões;

Mais uma vez...

Definição do **Watson Assistant** (2/4)

- Também fornece entidades de sinônimos pré-construídas para entidades comuns, como data, hora, pessoa, local e assim por diante;
- Fornecer respostas úteis a partir de classificações de intenção;
- Capacidade de armazenar e usar o contexto personalizado que pode ser necessário durante a sessão de conversação, bem como para fornecer respostas personalizadas;

Mais uma vez...

Definição do **Watson Assistant** (3/4)

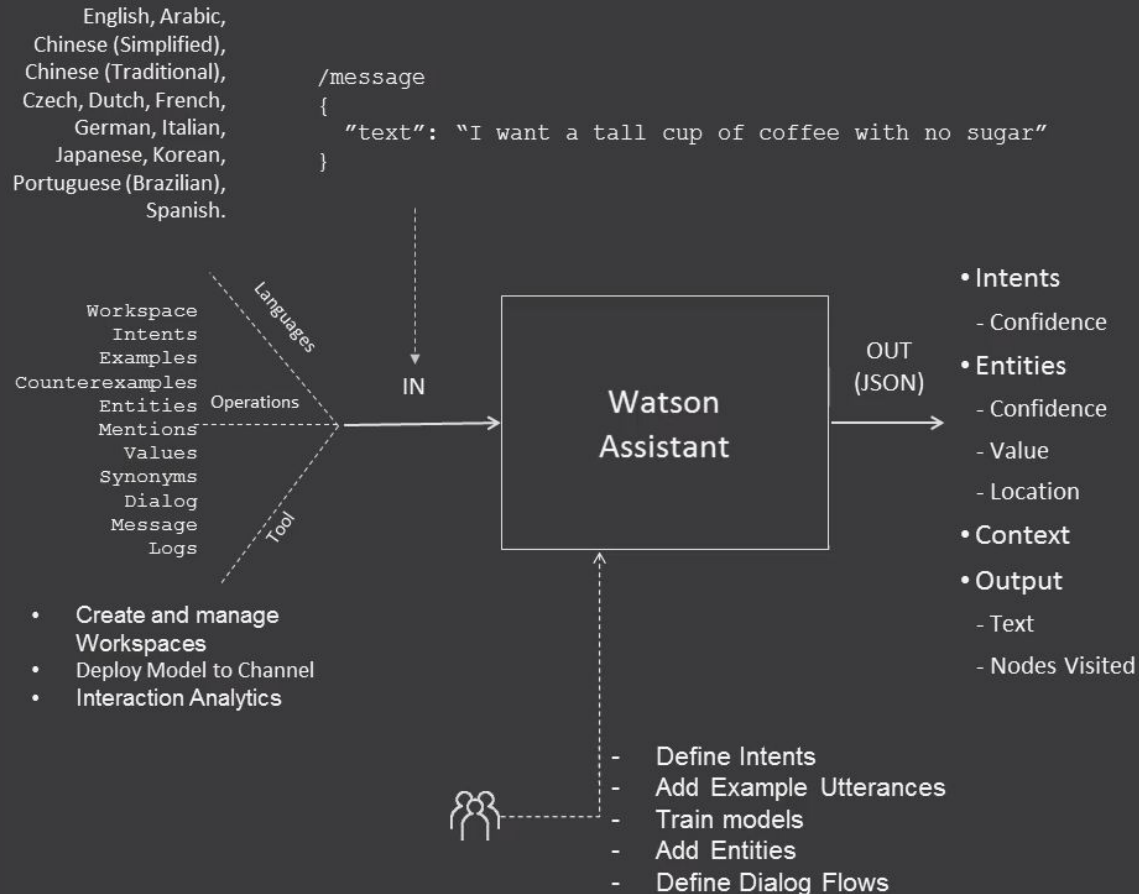
- Via slots, é possível reunir várias informações de um usuário em um único nó. Mesmo que o usuário informe os dados parcialmente, os slots salvam as informações coletadas do usuário e / ou contexto. Ele solicitará ao usuário que forneça apenas as informações ausentes.
- Capacidade de lidar com consultas bate-papo e consultas fora do assunto;
- Começar mais rapidamente a partir do catálogo de conteúdo de pacotes de conteúdo do setor e do atendimento ao cliente já configurado para economizar tempo;

Mais uma vez...

Definição do **Watson Assistant** (4/4)

- O modelo / *workspace* treinado pode ser chamado diretamente como uma API de aplicativos e canais como por exemplo o *Facebook Messenger*;
- Operações de API para gerenciar intenções, entidades e diálogo. Também consulta logs e gerência *workspaces*;
- O Watson Assistant que é executado no IBM Cloud possui controles e diretrizes para o tratamento de Informações Pessoais Identificáveis;

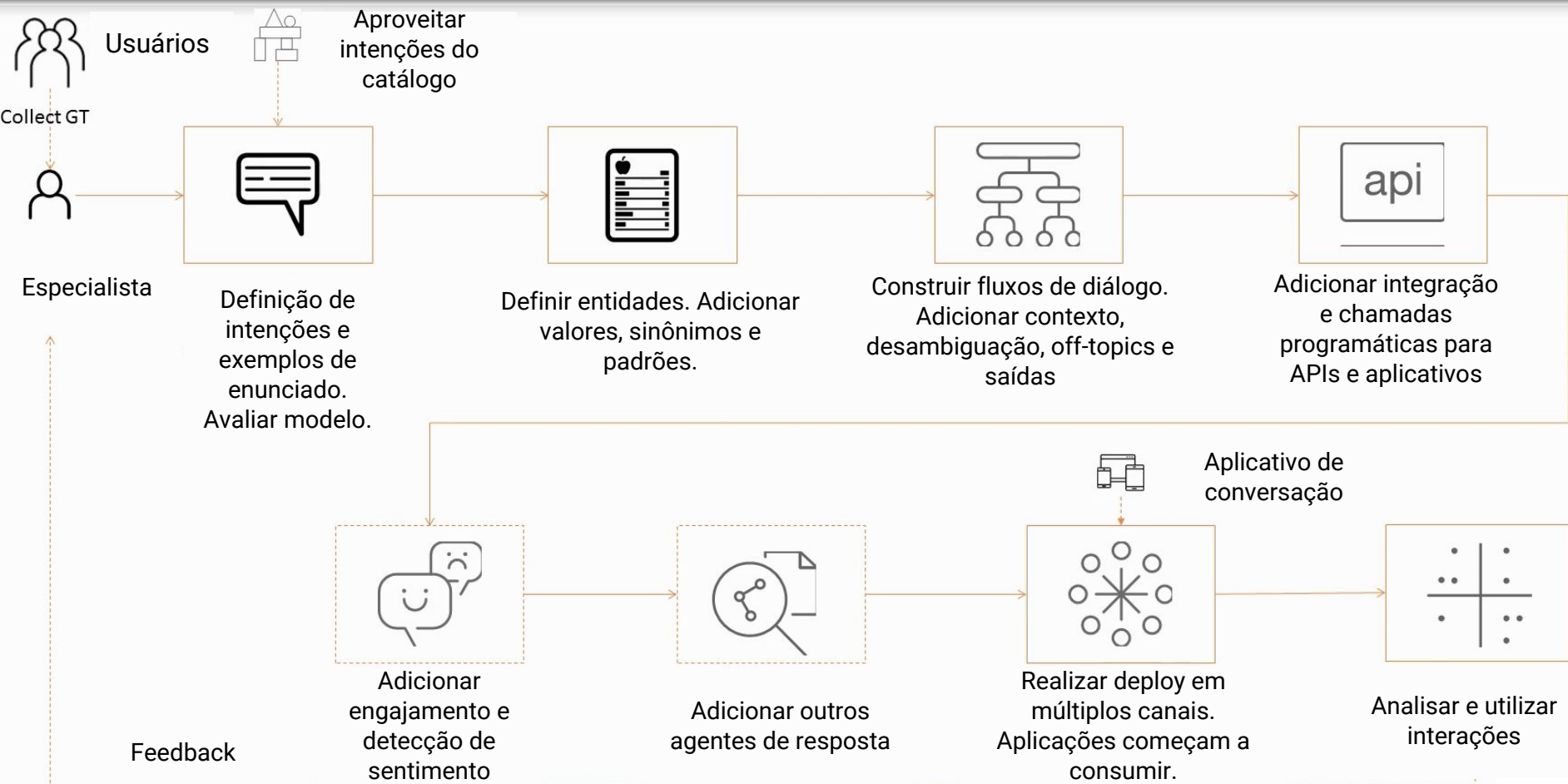
Fluxo Watson Assistant

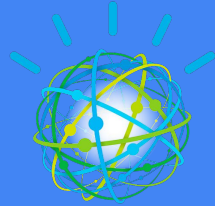


Quais línguas são suportadas?

- O Watson Assistant suporta vários idiomas;
- OBS: a própria interface de ferramentas (descrições, etiquetas e assim por diante) é em inglês;
- Os idiomas de suporte incluem inglês, árabe, chinês, tcheco, holandês, francês, alemão, italiano, japonês, coreano, português (brasileiro), espanhol;
- Recursos individuais do serviço são suportados em maior ou menor grau para cada idioma;

Construindo um sistema conversacional com Watson Assistant





O Watson Assistant fornece uma ferramenta "intuitiva" e fácil de usar para criar e gerenciar *workspaces*, gerenciar entidades intencionais e criar fluxos de diálogo.

Usando a ferramenta, você também pode modelar visualmente e criar um fluxo de diálogo que incorpore as intenções e entidades.

O que a API faz? - Operações da API do Watson Assistant

- Message;
- Workspace;
- Intents;
- Examples;
- Counterexamples;
- Entities;
- Mentions;
- Values;
- Synonyms;
- Dialog nodes;
- Logs

<https://console.bluemix.net/apidocs/assistant>

Alguns casos de uso em que podemos usar o Assistente do Watson

**Autoatendimento
ao cliente**

Assistente de carro

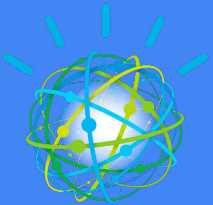
**Assistente de
hospitalidade**

**Automação de
e-mail**

**Automação da
interface de
linguagem natural**

**Automação de
tarefas orientada
por fala**

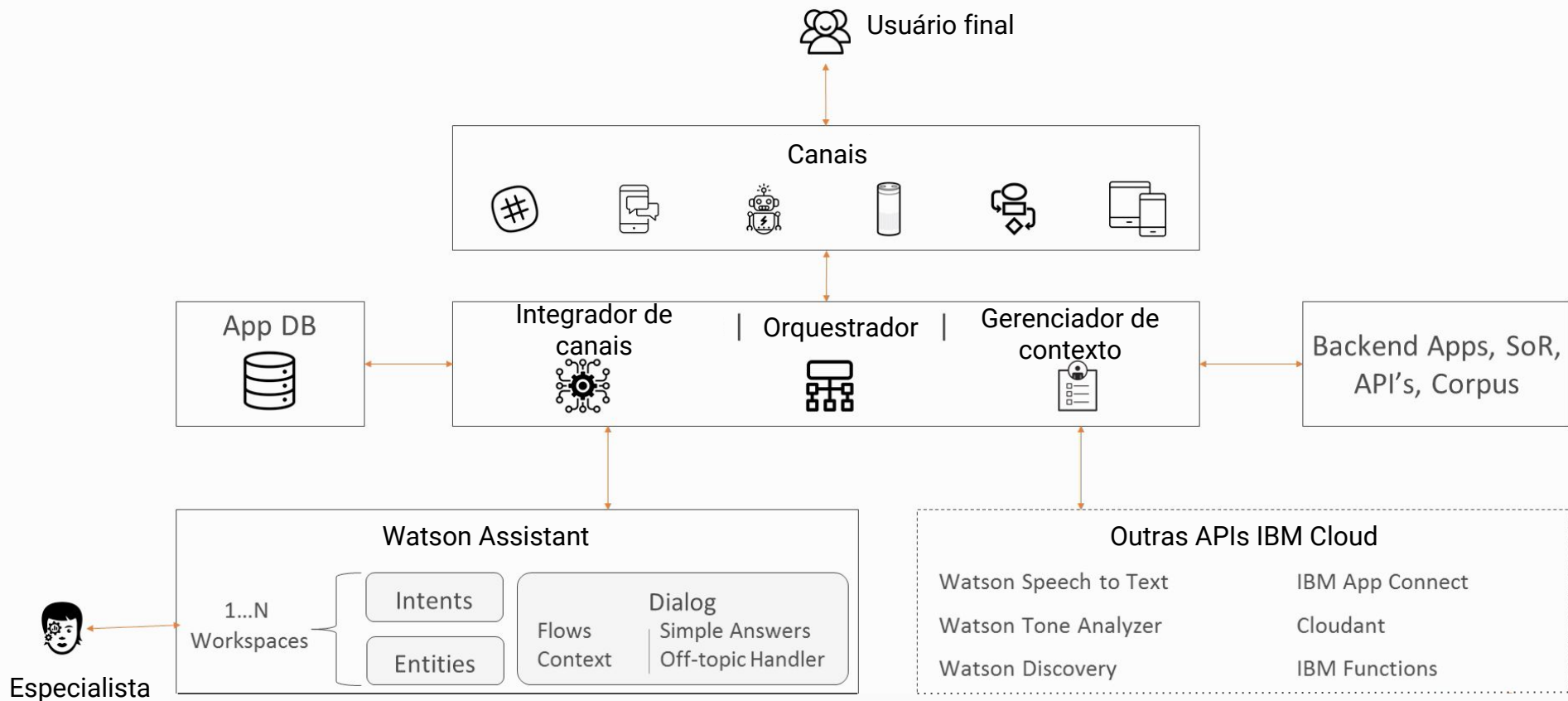
ENTIDADES, INTENÇÕES E *WORKSPACES(SKILLS)*



Entidades, intenções e *workspaces*

- Arquitetura de alto nível,
- Conceitos-chave
- *Workspaces*;

Típica arquitetura de alto nível de um sistema de conversação



Conceitos-chave de sistemas de conversação

- **Intenção:**

- Intenção representa a motivação ou a meta que um usuário final pode ter em mente;
- Uma intenção é expressa como um enunciado de texto;
- A partir do enunciado do texto em linguagem natural, tentamos classificar a intenção e, portanto, o que o usuário final está pedindo;
- Por exemplo, "Pedir bebida" é uma intenção;
- Um determinado sistema de conversação pode ter várias intenções para as quais é treinado;

Conceitos-chave de sistemas de conversação

- **Enunciado (expressão):**

- O usuário expressa sua intenção através de **enunciados** em linguagem natural;
- Por exemplo, o usuário pode perguntar "Eu quero um copo pequeno de café. Sem açúcar e use leite desnatado, por favor." é um exemplo de **enunciado** para a intenção "Pedir bebida";
- Para treinar um modelo de classificação / detecção de intenção, você fornece alguns exemplos de **enunciados** que seu usuário final geralmente usa para indicar seu objetivo;

Conceitos-chave de sistemas de conversação

- **Entidades:**

- Entidades representam os termos e objetos contidos no enunciado;
- Enquanto as intenções permitem que você entenda o objetivo ou a motivação do enunciado em linguagem natural, as entidades permitem que você extraia informações mais específicas do enunciado e forneça o contexto;
- Por exemplo, {Tipo de bebida = café}, {Tamanho da bebida = pequena}, {Ingredientes = [leite desnatado, sem açúcar]} são entidades de "Eu quero um copo pequeno de café. Sem açúcar e use leite desnatado, por favor.";

Workspaces

- **Workspaces** são contêineres para organizar artefatos de desenvolvimento;
- Cada **workspace** contém artefatos, incluindo - **intenções, entidades e diálogo**;
- Cada **workspace** possui uma linguagem declarada / identificada no **workspace**;
- Uma instância do Watson Assistant pode ter vários **workspaces**;
- Existem limites no entanto:
 - Free/Lite: 5 *workspaces*;
 - Standard/Premium: 20 *workspaces*;

Workspaces

- Todo *workspace* vem pré-carregado com um catálogo de "conteúdo pré-construído" que contém artefatos pré-construídos (intenções e entidades) para propósitos de introdução rápida;
- As alterações feitas em um *workspace* (entidades, intenções ou diálogo) acionam automaticamente o processo de treinamento do modelo;
- Vários "org spaces" podem ser criados combinando workspaces dependendo das necessidades;

Workspaces

- Os *workspaces* podem ser organizados dependendo das diretrizes, políticas e procedimentos de TI;
- Pense em cada *workspace* representando uma categoria ou grupo de funções ou habilidades chave;
- Um *workspace* treinado é chamado de um aplicativo externo por meio da chamada da API de mensagem;
- Cada *workspace* possui um identificador exclusivo que precisa ser passado na chamada da API de mensagem;
- Os *workspaces* podem ser exportados / importados como um todo em um arquivo JSON;
- Os *workspaces* podem ser clonados / copiados para criar um novo *workspace*;

Watson Assistant - Instância de serviço 01

Workspace 01

(Autoatendimento **DEV**)

Workspace 02

(Autoatendimento **HML**)

Workspace 03

(FAQ **DEV**)

Workspace 04

(FAQ **HML**)

Watson Assistant - Instância de serviço 02

Workspace 01

(Autoatendimento **PRD**)

Workspace 02

(FAQ **PRD**)

Watson Assistant - Instância de serviço 03

Workspace 01

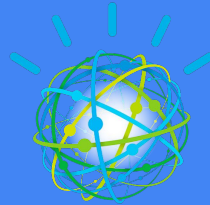
(Autoatendimento **PRD**)
failover

Workspace 03

(FAQ **PRD**)
failover

Embora vários desenvolvedores **possam** trabalhar em um único *workspace*, a qualquer momento, apenas um desenvolvedor pode(deveria) editar itens ou artefatos em um *workspace*:

- Intenções
- Entidades
- Nós de diálogo



Se vários desenvolvedores trabalharem simultaneamente no mesmo artefato, somente as alterações feitas e salvas por último serão aplicadas.



A classificação de texto é um padrão muito comum em IA.

Envolve tarefas de identificar a qual classe um novo enunciado deve se encaixar baseado nos dados de um treinamento prévio.

Intenção (1/3)

- Representa a motivação ou a meta que um usuário final pode ter em mente;
- Os usuários expressam suas intenções por meio de declarações;
- A **classificação** intencional é utilizada para identificar a expressão de linguagem natural e determinar onde ela se encaixa;
- Um sistema de conversação normalmente tem suas várias intenções treinadas com base nos requisitos de negócios;
- Identificar e treinar o conjunto certo de intenções é muito crítico para a adoção bem-sucedida do sistema de conversação;



Intenção = Pedir bebida

"Eu quero um copo pequeno de café.
Sem açúcar e use leite desnatado, por
favor."



Pensem em intenções como verbos

Intenção (2/3)

- Para cada intenção definida, são necessário exemplos representativos de como um usuário final se expressaria;
- O Watson Assistant usa muitos recursos dos exemplos de intenção para obter características do texto. Esses recursos incluem **bigramas**, **vetores de palavras**, **conceitos**, **pontuações** e etc...;
- O Watson Assistant por baixo dos panos usa várias técnicas, incluindo **NLP**, **ML** e **DL**, para extrair os recursos e treinar um modelo de classificação de intenção;



Intenção = Verifique o saldo da conta

Você pode me dizer o saldo da minha conta corrente?"



Intenção = Receber status da reclamação

Como obtenho o meu número de reclamação e verifico o status da minha reclamação?

Intenção (3/3)

- Quando uma declaração do usuário chega, várias estratégias de raciocínio que coletam a linguagem e o contexto do enunciado são aplicadas para classificá-lo para uma intenção;
- Quando uma intenção é reconhecida, o Watson Assistant executará o fluxo de diálogo mais apropriado para responder ao usuário;
- Intenções definidas no conjunto de ferramentas do Assistente do Watson têm um prefixo "#" inserido automaticamente;
- Com base no plano de serviço, há limites para o número de intenções por workspace:
 - Lite / Free: 100 intents / workspace - 25k exemplos no total;
 - Standard: 2000 intents / workspace - 25k exemplos no total;



Intenção = Redefinir senha

Não estou conseguindo logar no meu e-mail. Gostaria de redefinir minha senha.

Nomeando intenções

- Os nomes de **intenções** podem conter
 - Letras;
 - Números;
 - Hífens;
 - *Underlines*;
 - Traços e / ou pontos;
- Não pode conter espaços ou ter mais de 128 caracteres;
- O ideal é você seguir uma convenção de nomenclatura:
 - Exemplo: **VERBO_ENTIDADE_ACAO**;
- Alguns exemplos de intenções:
 - redefinir_senha;
 - verificar_saldo_conta;
 - encontrar_localizacao_filial;

Importando intenções em lote

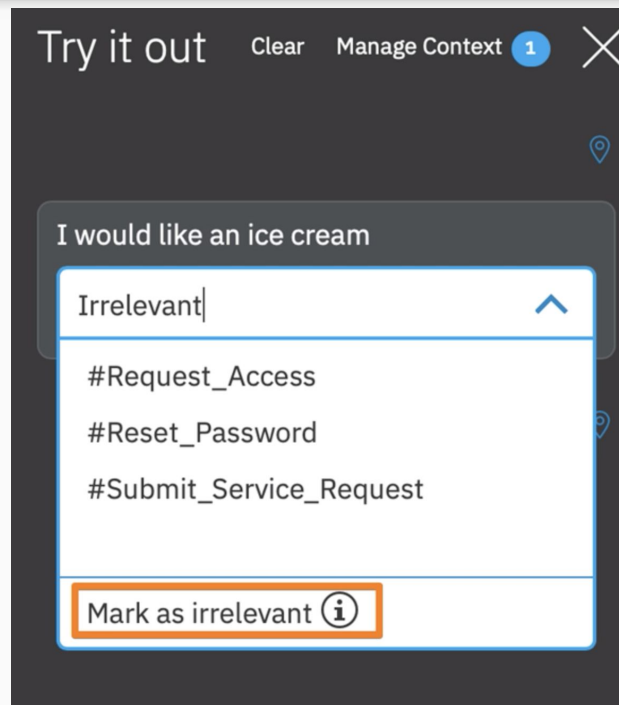
- Se você tiver um grande número de intenções e exemplos, poderá importá-los de um arquivo CSV no Watson Assistant;
- O formato do arquivo CSV deve ser {<EXEMPLO>, <NOME INTENÇÃO>};
- O arquivo CSV deve ser salvo com codificação UTF-8;

Hands on

- Crie um arquivo CSV contendo exemplos e intenções;
- Realize a importação do arquivo na plataforma do Watson Assistant;
- OBS: Tente ser criativo e atinja os seguintes objetivos:
 - Adicione vários exemplos para uma única intenção;
 - Adicione mais de uma intenção;
 - Adicione valores repetidos para algum dos conjuntos (exemplo, intenção);

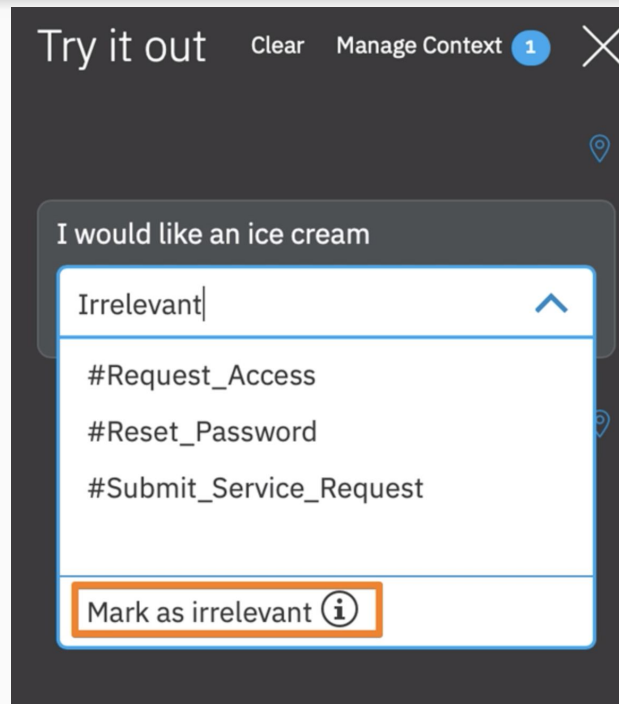
Irrelevante ou off-topic (1/2)

- Ao testar o modelo das intenções dentro da ferramenta Watson Assistant, você pode marcar e indicar uma determinada expressão como "irrelevante";
- Aqueles que são marcados como irrelevantes são salvos como "contra exemplos" no JSON do workspace;
- Estes estão incluídos/vão influenciar no treinamento do modelo;



Irrelevante ou off-topic (2/2)

- Uma vez marcado como irrelevante, você não pode mudar usando o ferramental;
- Tente novamente em "Try it out" e mude a atribuição de intenção;
- Você também pode ter uma intenção "off-topic" explicitamente com exemplos também;
- Mas não misture e combine;



Hands on

- Marque algum enunciado enviado para a plataforma como irrelevante;
- Observe o comportamento após a definição do contra exemplo;

Recomendações para intenções (1/2)

- Comece pequeno com poucas intenções e expanda organicamente + interativamente;
- Organize as intenções em Espaços de Trabalho com base nas funções de negócios que elas executam. Forneça uma descrição detalhada para cada intenção;
- Evite a sobreposição de intenções, bem como exemplos de expressão;
- Os enunciados usados para cada intenção devem ser representados de como os usuários finais perguntariam ou os colocariam. Não inventado;
- Não confie numa população específica de usuários ou PME. Mais variada, melhor.

Recomendações para intenções (2/2)

- Tenha pelo menos 10 exemplos de expressão para treinar cada intenção;
- Melhores respostas são produzidas com mais variedade;
- Limite o número de intenções por workspace;
- Mantenha as intenções simples e não muito refinadas. Exemplo: "Reset_Password" x "Reset_Email_Password" e "Reset_Acc_Password";
- É ideal adicionar off-topic ou intenção não relevante;
- É interessante dividir os exemplos das intenções em **treinamento** e **teste**.
Treinamento para treinar o modelo e teste para avaliar o desempenho do modelo.

Entidade

- Entidades representam os termos e objetos contidos no enunciado;
- As entidades permitem extrair informações mais específicas do enunciado e fornecem o contexto;
- A combinação de intenção e entidades ajuda você a entender melhor o que o usuário está perguntando no contexto;
- Você pode fornecer respostas personalizadas / específicas com base em entidades extraídas da declaração;



Entidades

Ingrediente **Tipo de bebida** **Tamanho de bebida**

"Eu quero um copo pequeno de café.
Sem açúcar e use leite desnatado, por favor."



Pensem em entidades como substantivos

Entidade

- As entidades definidas no Watson Assistant têm um prefixo "@" incluído automaticamente;
- Existem **duas** maneiras pelas quais as entidades são reconhecidas no Watson Assistant:
 - **Baseado em dicionário:** O Watson Assistant procura termos nas declarações do usuário que correspondem aos valores incluindo sinônimos exatamente ou através de um padrão;
 - **Baseado em contexto:** O Watson Assistant não procura por padrão de texto exato ou correspondências. Um modelo é treinado tanto no termo anotado quanto no contexto em que o termo é usado na frase que você anota;



Entidades

Nome de aplicação Tipo de problema

"Não consigo fazer login no Outlook.
Resetar minha senha."

Embora a definição de entidades não seja obrigatória no Assistente do Watson, elas permitem melhorar significativamente o comportamento do sistema final

Entidades baseadas em dicionário

- As entidades baseadas em dicionário no Watson Assistant possuem a seguinte estrutura - {nome da entidade, valor, sinônimo / padrão};
- Existem três tipos de entidades de dicionário;

Dictionary		Annotation ^{BETA}
<input type="checkbox"/>	Entity values (2) ▼	Type
<input type="checkbox"/>	billing	Synonyms
<input type="checkbox"/>	mailing	Synonyms

@contact

Last modified 3 months ago

Try it

Entity name

@contact

Fuzzy Matching Off

Value name

Enter value

Add value

Synonyms

▼

Synonyms

Add synonym...

Show recommendations

Dictionary

Annotation

BETA

Entity values (2) ▼

Type

email

Patterns

^([<>@\\|\\.,:;\\s@*]+(\\.[<>@\\|\\.,:;\\s@*]+)?)\$

phone

Patterns

^[\\+]?([0-9]{3}[0-9]{3})?[-\\s\\.]?[0-9]{3}[-\\s\\.]\$

Entidades baseadas em dicionário

- **Entidade com sinônimos:**

- Uma categoria como "Tipo de bebida", cada um com um ou mais valores (Exemplo: café, refrigerante) e cada valor com um ou mais sinônimos (cafezinho, cappuccino, latte).
- Em tempo de execução, o serviço reconhece os termos no enunciado que correspondem exatamente aos valores ou sinônimos;

- **Entidade com padrões:**

- Uma categoria de termos como uma entidade como "Forma de contato", cada um com um ou mais valores (Exemplo: Email, Telefone) e cada valor é uma expressão regular que define o padrão textual das menções desse tipo de valor.
- Em tempo de execução, o serviço procura padrões correspondentes à expressão regular no enunciado e identifica as correspondências como menções dessa entidade.

Entidades baseadas em dicionário

- **Entidades do sistema:**
 - São entidades de sinônimos pré-construídas pela IBM. Elas cobrem categorias comumente usadas, como **pessoa**, **local**, **números**, **datas** e **horários**.
 - Você simplesmente habilita uma entidade do sistema para começar a usá-la.

Hands on

- Defina pelo menos duas entidades a partir de padrões (regex);
 - Sugestão: e-mail e CPF;
- Habilite entidades de sistema e realize alguns testes;

Nomeando entidades (1/2)

- Os nomes de entidades podem conter letras (em Unicode), números, sublinhados e hífen. Prefixado com @;
- Como mencionado anteriormente, pense em entidades como substantivos.
- Siga uma convenção de nomenclatura, como o próprio objeto (em minúsculas);

Nomeando entidades (2/2)

- Por exemplo: aplicativo, contato, tipo_problema;
- Não inclua espaços no nome.
- O nome não pode ter mais de 64 caracteres;
- Não inicie o nome com a string sys (reservada para entidades do sistema);
- Exemplo de entidades:
 - IBM Notes (aplicativo);
 - Telefone residencial (contato);

Fuzzy matching (1/2)

- Tem a função de reconhecer o termo em um enunciado de entrada **similar** aos valores e sinônimos da entidade;
- Desligado por padrão e, se necessário, deve ser ativado em cada nível de entidade;

Fuzzy Matching ⓘ ☒ On

Fuzzy matching (2/2)

- *Fuzzy matching* funciona, aproveitando técnicas, incluindo:
 - Stemming (derivando);
 - Erros ortográficos para lidar com erros ortográficos do usuário na entrada de enunciados;
 - Correspondência parcial com base na substring;
- OBS: *Fuzzy Matching* pode interferir com palavras definidas como valores em outras entidades.

Fuzzy Matching ⓘ ☒ On

Importar entidades em lote

- As entidades baseadas em dicionário no Watson Assistant possuem a seguinte estrutura - {entity name, value, synonym / pattern};
- Se você tiver um grande número de entidades, valores e sinônimos, poderá importá-los de um arquivo CSV no Watson Assistant;
- O formato do arquivo CSV deve ser {entity, value, synonyms};
- O arquivo CSV deve ser salvo com codificação UTF-8;

Importar entidades via API

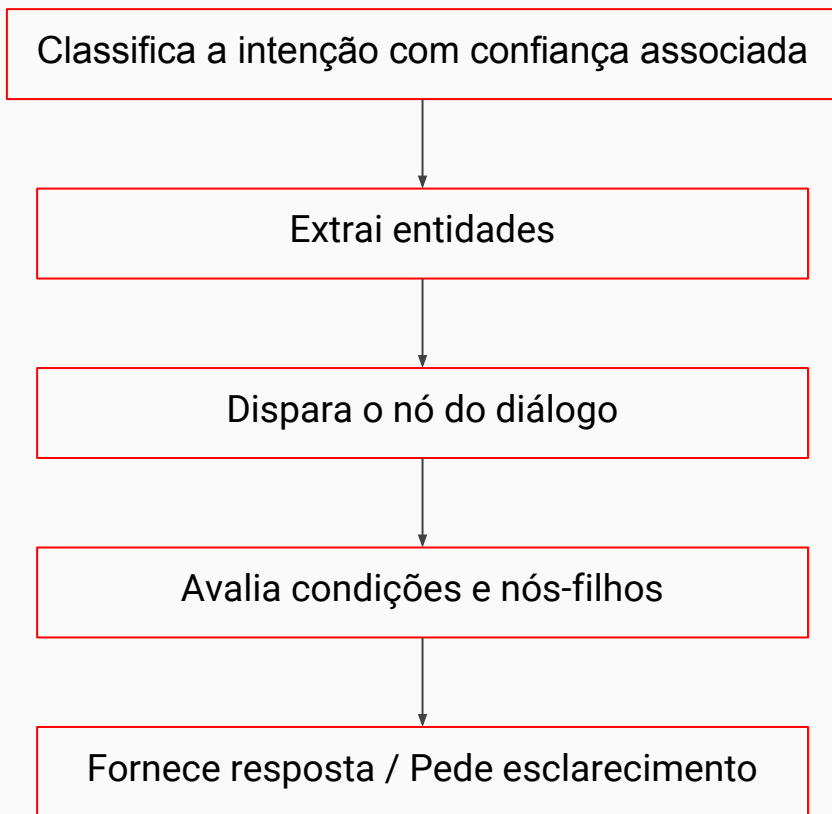
- Você também pode escrever um script para importar via APIs:
 - Faça um POST para `/workspaces/{workspace_id}/entities` para criar uma nova entidade;
 - Seguido por `/workspaces/{workspace_id}/entities/{entity}/values` para criar um novo valor para uma entidade;
 - Seguido por `/workspaces/{workspace_id}/entities/{entity}/values/synonyms` para criar um novo sinônimo para um valor de entidade;

Hands on

- Crie um arquivo CSV contendo entidades e sinônimos;
- Realize a importação do arquivo na plataforma do Watson Assistant;
- OBS: Tente ser criativo e atinja os seguintes objetivos:
 - Adicione vários sinônimos para uma única entidade;
 - Adicione mais de uma entidade;

Processamento de diálogo (1/4)

- Uma caixa de diálogo é representada graficamente / visualmente no Watson Assistant como uma árvore;
- Para cada intenção, **geralmente** há uma ramificação para lidar com a lógica da conversação e o fluxo a ser cumprido;
- Uma ramificação é composta de vários nós;
- Intenções, entidades extraídas e as informações de contexto injetadas são alimentadas na execução do nó;
- Quando uma intenção é reconhecida, o Watson Assistant acionará ou escolherá executar o nó de diálogo apropriado e o fluxo associado para responder ao usuário;



Processamento de diálogo (2/4)

- Cada **nó** de diálogo contém, no mínimo, uma condição e uma resposta;
- A **condição** especifica como o nó deve ser acionado. Pode ser:
 - Intenções;
 - Entidades;
 - Contexto;
 - Outros e em uma combinação;
- As **condições** podem ser baseadas no nome da intenção, entidade, valores de contexto, respostas do usuário e etc;
- **Resposta** é a declaração que o Watson responde ao usuário;
- A resposta pode ser texto, mostrar uma imagem, uma lista de opções ou disparar chamadas de API;

Processamento de diálogo (3/4)

- O fluxo de nós de diálogo é processado pelo serviço do primeiro nó da árvore até o último;
- Percorrendo a árvore, o Watson acionará um nó se a condição for atendida;
- Em seguida, ele se move ao longo do nó acionado para verificar a entrada do usuário em relação a qualquer condição do nó filho;
- Ao verificar os nós filhos, ele se move novamente do primeiro nó filho para o último;
- Ordem dos ramos e nós são importantes. Os nós são avaliados do primeiro ao último;

Processamento de diálogo (4/4)

- Quando o assistente descobre que nenhuma das condições no conjunto atual de nós filho é avaliada como verdadeira, ele volta para a base da árvore / ramificação;
- O Watson continua processando os nós do primeiro ao último. Quando não há condições em nós avaliando como verdadeiro, a resposta do último nó na árvore, que normalmente tem uma condição de **anything_else** que sempre é avaliada como verdadeiro, é retornada;