



# IBM WATSON + PYTHON

Mineração de dados para inserção de conhecimento  
no IBM Watson em uma abordagem educacional

Luiz Eduardo F. Bentes - [luizfbentes@gmail.com](mailto:luizfbentes@gmail.com)



C . E . S . A . R



**Luiz Bentes**  
**Engenheiro de Testes - CESAR Manaus**

**Estudante de Engenharia de  
Computação na Universidade do  
Estado do Amazonas**

**Membro do Laboratório de  
Sistemas Inteligentes - LSI**

# A expectativa



# A realidade



# O que é um chatbot?

“São softwares que funcionam dentro de aplicações de mensagens.”

“Em um comércio, chatbots são programas utilizados para simular a conversação humana”





“De maneira simples, *bots* são uma nova interface do usuário, ou seja, uma nova maneira de expor serviços de software através de uma interface conversacional.”

# Projeto Pedagógico do Curso

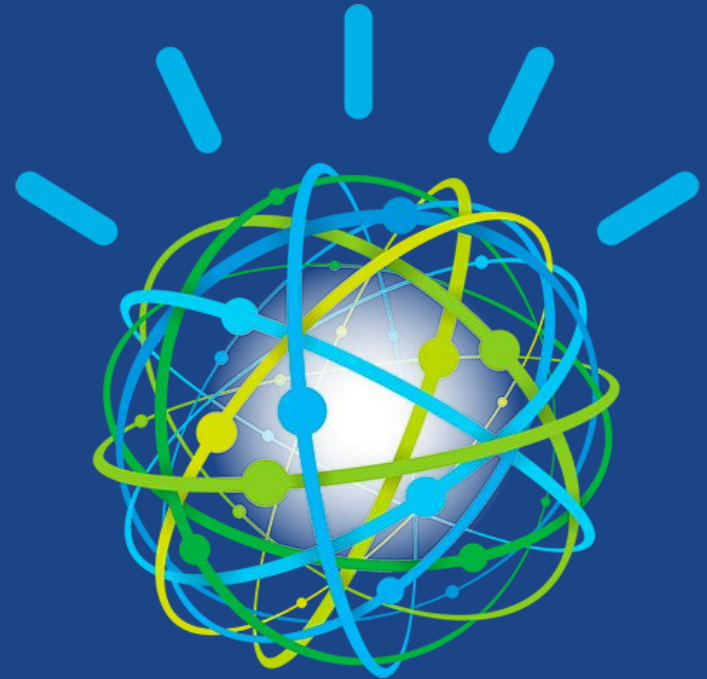
- Documento construído pela gerência do curso de graduação
- Vigente conforme o **Decreto nº 3.860/2001** e a Lei de Diretrizes e Bases da Educação Nacional
- **171 páginas** (PPC de Engenharia de Computação da Universidade do Estado do Amazonas)

## PPC - Projeto Pedagógico do Curso

- Apresentação do documento
- Dados institucionais da UEA
- Sistema de Administração Acadêmica
  - Sistema Curricular
  - Matrícula
  - Aproveitamento de Estudos
  - Regime Letivo
- Contextualização do Curso
  - Dados de identificação do curso
  - Matriz Curricular
  - Turno de Funcionamento
  - Estágio Supervisionado
  - Atividades Complementares
  - Iniciação Científica
  - Monitoria
  - TCC
- Apêndices
  - Ementário
  - Corpo Docentes
  - Regulamentos de Estágio/TCC/Atividades Complementares

# IBM WATSON - Computação Cognitiva

- Combinação das habilidades superiores de um computador com as habilidades inerentes aos seres humanos
- Primeira plataforma referente a Computação Cognitiva
- As principais frentes de estudo dos Sistemas Cognitivos são:
  1. Entendimento de **Linguagem Natural**;
  2. Geração e teste de hipóteses;
  3. Aprendizado dinâmico





**Será que isso poderia ajudar  
mesmo o aluno?**

**Bora fazer um “forms” e ver**

Conhecendo o “Problema”

# Google Forms + Python

- ❑ 21 perguntas
- ❑ Perguntas sobre o PPC (documento)
- ❑ Perguntas sobre o conteúdo do PPC

- ❑ 61 respostas obtidas
- ❑ 30% dos alunos responderam
- ❑ Vamos a análise de dados...

## Engenharia de Computação -- Pesquisa de conhecimento sobre o Plano Pedagógico do Curso

Olá! Esse questionário é parte do estudo observacional para desenvolvimento do Trabalho de Conclusão de Curso do aluno Luiz Eduardo Fernandes Bentes sob a orientação da Profa. Dra. Eloá B. Guedes, ambos da Escola Superior de Tecnologia da Universidade do Estado do Amazonas.

Quem pode participar:

- Alunos de Engenharia de Computação da Universidade do Estado do Amazonas, cursando qualquer período letivo

Ao longo deste questionário, os participantes serão convidados a responderem perguntas sobre o PPC do curso de Engenharia de Computação atualmente vigente na instituição. O sigilo dos nomes será preservado em todas as etapas da divulgação do resultado. Ao clicar SIM na pergunta a seguir, você concorda com os termos de participação nesta pesquisa.

Sua participação é muito importante, pois será útil para colaborar no desenvolvimento de uma solução que pode ajudar você e a outros alunos do curso! Pedese gentilmente que as perguntas sejam respondidas de maneira honesta e transparente, para reforçar os resultados e as análises posteriores.

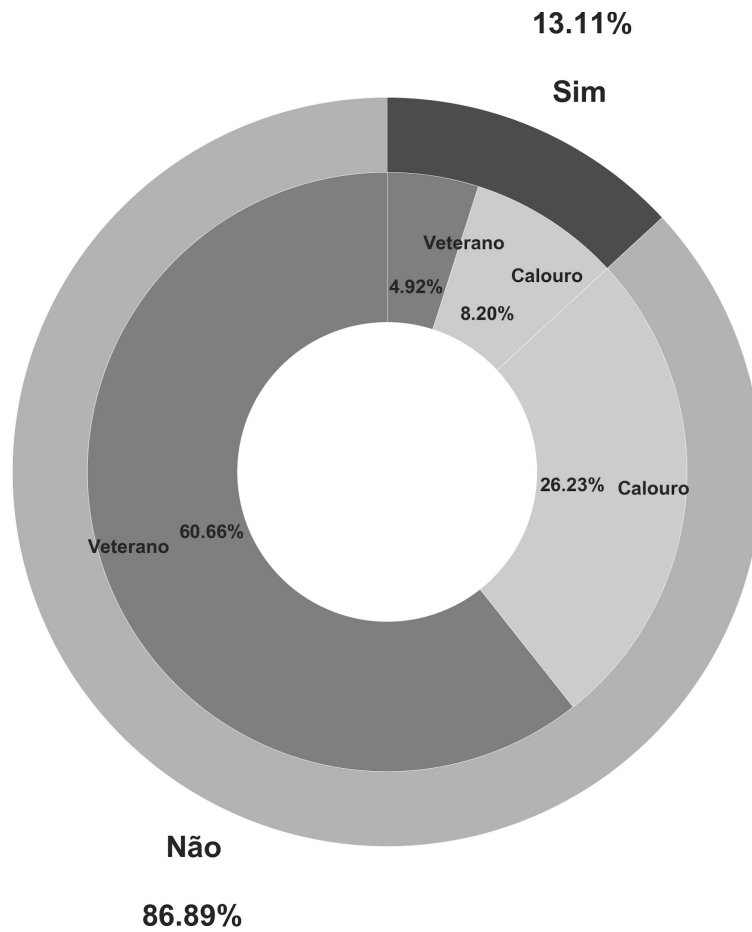
Qualquer dúvida, entre em contato:

Luiz Bentes ([lefb.eng@uea.edu.br](mailto:lefb.eng@uea.edu.br))

Eloá B. Guedes ([ebgcosta@uea.edu.br](mailto:ebgcosta@uea.edu.br))

Laboratório de Sistemas Inteligentes  
Universidade do Estado do Amazonas.

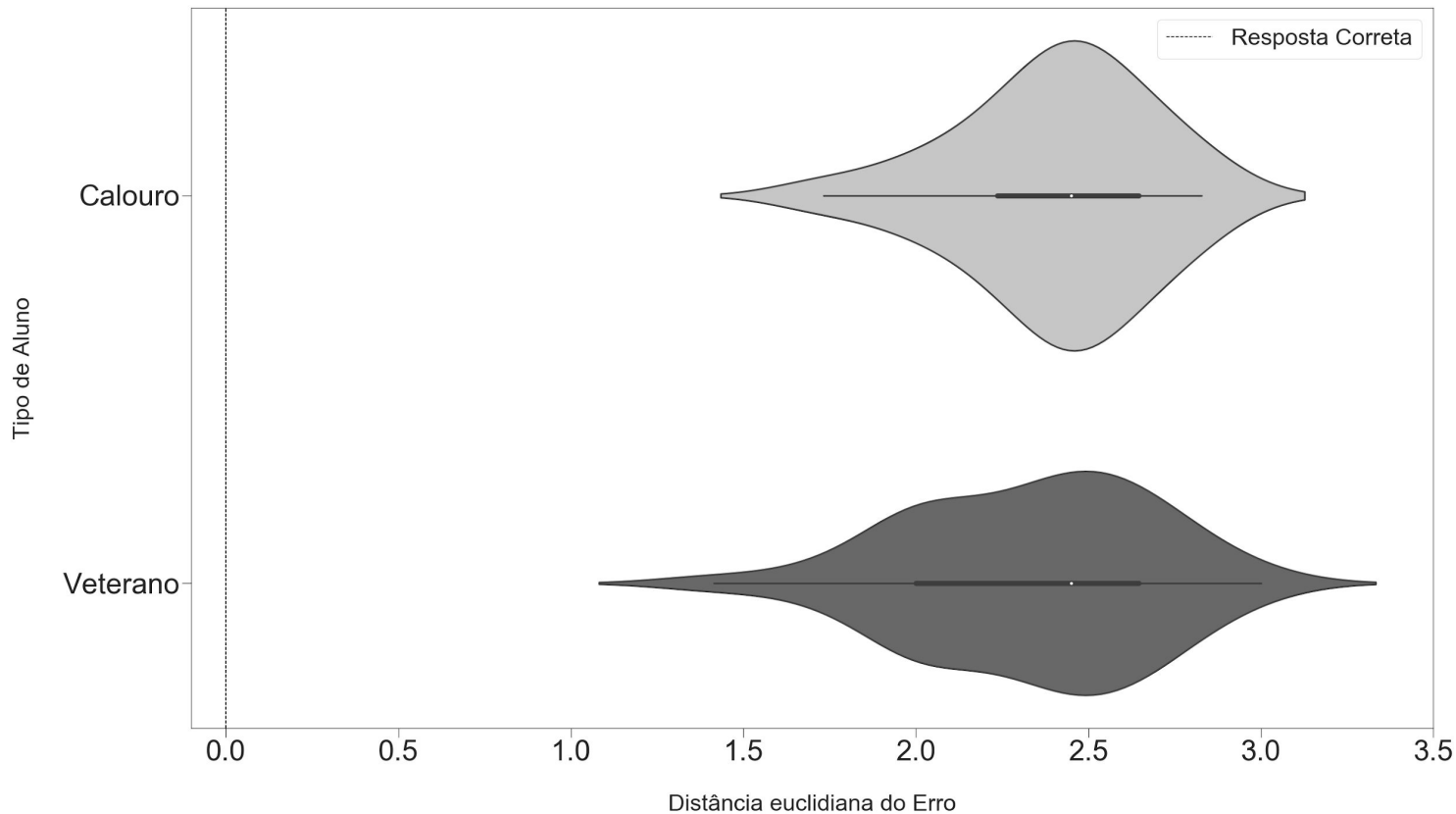
# Resultado sobre a pergunta: “Conheço o PPC do curso de Engenharia de Computação”



Considero que o PPC do curso aborda os seguintes tópicos (marque todas as possíveis respostas):

- Disciplinas da grade de computação
- Regras do estágio obrigatório
- Resultados do ENADE (3 últimos anos)
- Área de atuação dos professores
- Calendário Acadêmico do ano letivo
- Dados históricos do Curso
- Contato dos professores
- Laboratórios disponíveis
- Regras de avaliação para Mestrado/Doutorado
- Tecnologias utilizadas no curso
- História da UEA

# Resultado sobre a pergunta: “Considero que o PPC do curso aborda os seguintes tópicos (marque todas as possíveis respostas)”

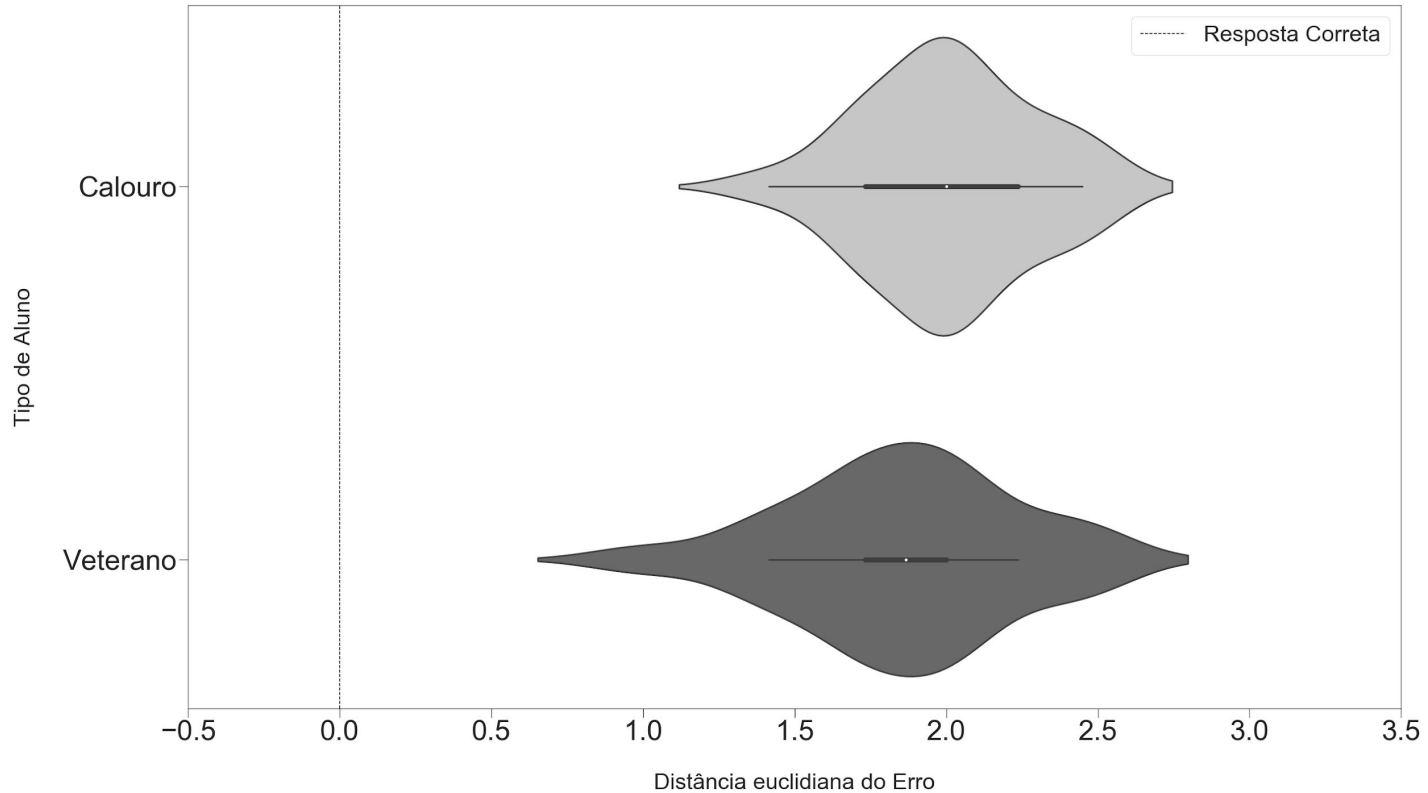


Considero que as 120h de atividades complementares possam ser integralizadas com as seguintes atividades (marque todas as opções que julgar possíveis):

- Iniciação Científica
- Monitoria de Laboratórios
- Doação de Sangue
- Curso de Idiomas
- Disciplinas Optativas
- Estágio Obrigatório
- Serviço Militar
- Visita Técnica
- Desenvolvimento de Protótipos
- Palestras
- Congressos



Resultado da pergunta: “Considero que as 120h de atividades complementares possam ser integralizadas com as seguintes atividades...”



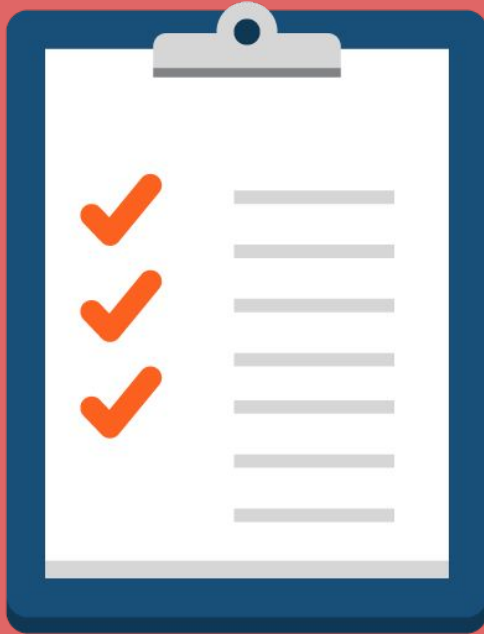
**E tudo feito com Python**

**E tudo feito com Python**



# Modelando o Chatbot

# Checklist do Chatbot



- ❑ Onboarding **simples e explicativo**
- ❑ Conteúdo do PPC de forma coesa e de fácil acesso
- ❑ Estrutura de conversa capaz de simular uma **conversa realística.**
- ❑ Está disponíveis a **eventuais mudanças** conforme o PPC vigente.
- ❑ Fácil acesso ao *chatbot* via aplicativos de mensagens

# Onboarding

Refere-se a primeira interação do chatbot;

Responsável por introduzir as funcionalidades, objetivos e formas de utilização do *chatbot*





Engenharia de Computação  
GRADUAÇÃO - BACHARELADO

Manaus

Apresentação    Currículos    Docentes    **Discentes**    Coordenação    Secretaria    Calendário

Ingresso: INTEGRAL - 2019/1

| Matrícula  | Nome                              | Forma de Ingresso | Situação  |
|------------|-----------------------------------|-------------------|-----------|
| 1915080050 | Adan Thierry Souza Dos Santos     | Vestibular        | Ativo     |
| 1915080001 | Alejandro Da Rocha Loureiro       | Vestibular        | Ativo     |
| 1915080049 | Brayan Santos Da Silva            | Vestibular        | Ativo     |
| 1915080060 | Caio Vinícius Pereira Moura       | SIS               | Ativo     |
| 1915080053 | Davi Rodrigues De França          | Vestibular        | Ativo     |
| 1915080032 | Eduardo De Arruda Lira            | SIS               | Ativo     |
| 1915080063 | Fernando Araújo Alves Filho       | SIS               | Ativo     |
| 1915080064 | Gabriel Silva Pacheco De Carvalho | SIS               | Cancelado |
| 1915080011 | Gabriella Rabelo Monteiro         | Vestibular        | Ativo     |
| 1915080065 | Giovanna Souza Teodoro            | SIS               | Ativo     |

pandas  
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

Python e Selenium para navegação nas abas

Captura das partes importantes

Criação do DataFrame

|           | <b>matricula</b> | <b>nome_completo</b>                | <b>entrada</b> | <b>status</b> |
|-----------|------------------|-------------------------------------|----------------|---------------|
| <b>1</b>  | 1915080050       | Adan Thierry Souza Dos Santos       | Vestibular     | Ativo         |
| <b>2</b>  | 1915080001       | Alejandro Da Rocha Loureiro         | Vestibular     | Ativo         |
| <b>3</b>  | 1915080049       | Brayan Santos Da Silva              | Vestibular     | Ativo         |
| <b>4</b>  | 1915080060       | Caio Vinicius Pereira Moura         | SIS            | Ativo         |
| <b>5</b>  | 1915080053       | Davi Rodrigues De Franca            | Vestibular     | Ativo         |
| <b>6</b>  | 1915080032       | Eduardo De Arruda Lira              | SIS            | Ativo         |
| <b>7</b>  | 1915080063       | Fernando Araujo Alves Filho         | SIS            | Ativo         |
| <b>9</b>  | 1915080011       | Gabriella Rabelo Monteiro           | Vestibular     | Ativo         |
| <b>10</b> | 1915080065       | Giovanna Souza Teodoro              | SIS            | Ativo         |
| <b>11</b> | 1915080057       | Guilherme Melo Dos Santos           | Vestibular     | Ativo         |
| <b>12</b> | 1915080052       | Gustavo Faraco De Andrade Rodrigues | Vestibular     | Ativo         |
| <b>13</b> | 1915080066       | Ian Mateus Ribeiro De Souza         | SIS            | Ativo         |
| <b>14</b> | 1915080034       | Igor Carvalho Costa                 | SIS            | Ativo         |
| <b>15</b> | 1915080014       | Joao Filipe Da Silva Brandao        | Vestibular     | Ativo         |



# Capturar de Informações do PPC

## Disciplinas:

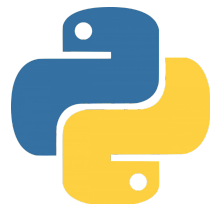
- Código da disciplina
- Ementa
- Créditos
- Pré-requisitos
- Período

The diagram illustrates the process of capturing information from a PDF document. A central icon of a document with a red 'PDF' label is positioned between two boxes. A red arrow points from the PDF icon to the 'Disciplinas' box on the left, and another red arrow points from the PDF icon to the 'Docentes' box on the right. The 'Disciplinas' box contains a list of five items: 'Código da disciplina', 'Ementa', 'Créditos', 'Pré-requisitos', and 'Período'. The 'Docentes' box contains a list of four items: 'Currículo Lattes', 'Formação', 'Contato', and 'Área de atuação'.

PDF

## Docentes:

- Currículo Lattes
- Formação
- Contato
- Área de atuação



Reg[ular]  
\*  
Ex[pression]

| SIGLA  | COMPONENTE CURRICULAR     | CARGA HORÁRIA |     |      |     | Nº DE CRÉDITOS | Pré-Req   |
|--|---------------------------|---------------|-----|------|-----|----------------|-----------|
|  |                           | CHT           | CHP | CHES | THC |                |           |
| ESTCMP031  | Inteligência Artificial I | 60            | 30  | -    | 90  | 5.4.1          | ESTCMP030 |
| EMENTA   |                           |               |     |      |     |                |           |
| Introdução. Resolução de problemas. Métodos de busca. Conhecimento e raciocínios precisos e imprecisos. Aprendizagem. Comunicação. Ação. Aplicações.   |                           |               |     |      |     |                |           |
| OBJETIVO   |                           |               |     |      |     |                |           |
| Conhecer as questões básicas e natureza da pesquisa e desenvolvimento na área. Analisar os aspectos relacionados a aquisição, representação e exploração do conhecimento para a produção de sistemas inteligentes em diferentes domínios de aplicação.   |                           |               |     |      |     |                |           |
| BIBLIOGRAFIA BÁSICA  |                           |               |     |      |     |                |           |
| <ol style="list-style-type: none"><li>LUGER, G. F.; STUBBLEFIELD, W. A.; Artificial Intelligence - Structures and Strategies for Complex Problem Solving. 3a ed. 1998</li><li>RUSSEL, S.; NORVIG, P. (2010) Artificial Intelligence - A Modern Approach. Editora Pearson Education. 3a. ed.</li><li>ARTERO, A. O. Inteligência Artificial - Teoria e Prática. Editora Livraria da Física, 2009</li></ol> |                           |               |     |      |     |                |           |
| BIBLIOGRAFIA COMPLEMENTAR  |                           |               |     |      |     |                |           |
| <ol style="list-style-type: none"><li>BRAGA, Antonio de Padua; LUDERMIR, Andre Ponce de Leon F. De Carvalho; BERNARDA, Teresa. Redes Neurais Artificiais - Teoria e Pratica, 2a edição. Editora LTC, 2007.</li><li>Tim JONES. Artificial Intelligence: A Systems Approach. Jones and Bartlett Publishers, 2008.</li></ol>  |                           |               |     |      |     |                |           |

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



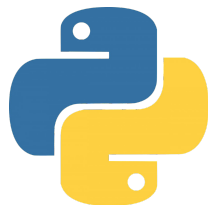
Python e Expressões Regulares para captura de informações



PPC do Curso (Apêndice A)



Criação do DataFrame



Reg[ular]  
\*  
Ex[pression]

| ITEM | NOME DO DOCENTE                    | GRADUAÇÃO                     | TITULAÇÃO | LINK PLATAFORMA LATTES (CNPq)   | REGIME DE TRABALHO |
|------|------------------------------------|-------------------------------|-----------|---|--------------------|
| 1    | Áurea Hiléia da Silva Melo         | Processamento de Dados – UFAM | Mestre    | <a href="http://lattes.cnpq.br/0243106634406881">http://lattes.cnpq.br/0243106634406881</a>   | 40h                |
| 2    | Carlos Maurício Serodio Figueiredo | Engenharia Elétrica – UFAM    | Doutor    | <a href="http://buscatextual.cnpq.br/buscatextual/visualiza_cv.do?id=B508867">http://buscatextual.cnpq.br/buscatextual/visualiza_cv.do?id=B508867</a> | 40h                |
| 3    | Danielle Gordiano Valente          | Processamento de Dados - UFAM | Mestre    | <a href="http://lattes.cnpq.br/8744374911033949">http://lattes.cnpq.br/8744374911033949</a>   | 40h                |
| 4    | Danielle Pompeu Noronha            | Processamento de Dados - UFAM | Mestre    | <a href="http://buscatextual.cnpq.br/buscatextual/visualiza_cv.do?id=C327980">http://buscatextual.cnpq.br/buscatextual/visualiza_cv.do?id=C327980</a> | 40h                |

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

Python e Expressões Regulares para captura de informações

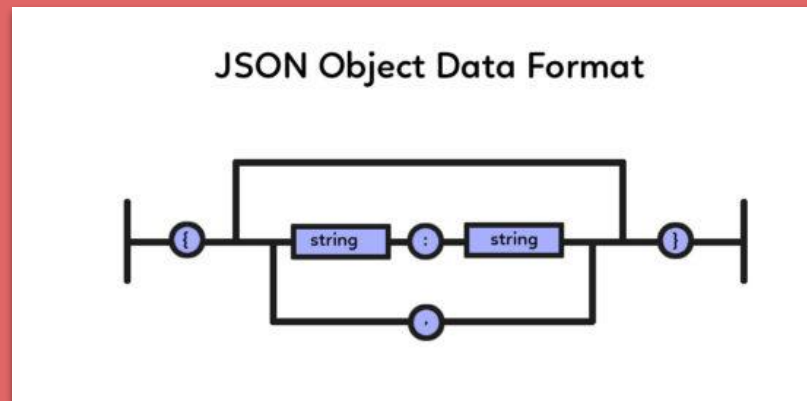
PPC do Curso (Apêndice B)

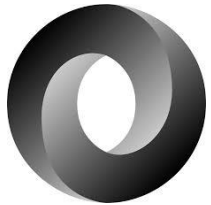
Criação do DataFrame

|   | nome                                     | codigo    | ch_teorica | ch_pratica | ch_curricular | ch_total | credito | pre_requisito | periodo | ementa  |
|---|--|-----------|------------|------------|---------------|----------|---------|---------------|---------|---|
| 0 | Álgebra Linear I                         | ESTBAS001 | 60         | 00         | 00            | 60       | 4.4.0   | -             | 1       | Matrizes.<br>Determinantes.<br>Sistemas lineares. V...  |
| 1 | Cálculo I                                | ESTBAS002 | 90         | 00         | 00            | 90       | 6.6.0   | -             | 1       | Limite e continuidade.<br>A derivada. A regra da<br>... |
| 2 | Introdução à Engenharia                  | ESTBAS003 | 30         | 00         | 00            | 30       | 2.2.0   | -             | 1       | Engenharia como<br>Ciência e Tecnologia;<br>Atuação ... |
| 3 | Introdução às Ciências do Ambiente       | ESTBAS004 | 30         | 00         | 00            | 30       | 2.2.0   | -             | 1       | Ciência Ambiental:<br>Conceito de Ciência<br>Ambient... |
| 4 | Química Geral                            | ESTBAS005 | 60         | 00         | 00            | 60       | 4.4.0   | -             | 1       | Estados da Matéria.<br>Estrutura atômica.<br>Teoria ... |
| 5 | Comunicação e Expressão                  | ESTBAS006 | 60         | 00         | 00            | 60       | 4.4.0   | -             | 1       | A linguagem. A língua<br>e a fala; variedades<br>lin... |
| 6 | Física I                                 | ESTBAS007 | 60         | 00         | 00            | 60       | 4.4.0   | -             | 1       | Sistemas de medida.<br>Mecânica: movimento<br>em uma... |
| 7 | Introdução à Programação de Computadores | ESTCMP009 | 30         | 30         | 00            | 60       | 3.2.1   | -             | 1       | Fundamentos da<br>resolução de<br>problemas (Lógica...  |
| 8 | Álgebra Linear II                        | ESTBAS008 | 60         | 00         | 00            | 60       | 4.4.0   | ESTBAS001     | 2       | Espaço vetorial.<br>Dependência e<br>independência l... |

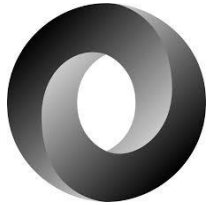
# Conversão de Dataframe para JSON

- ❑ Entidades e Fluxos de Diálogo provindo dos DataFrames
- ❑ Intenções construídas tendo em vista a utilização dos alunos
- ❑ Versionamento



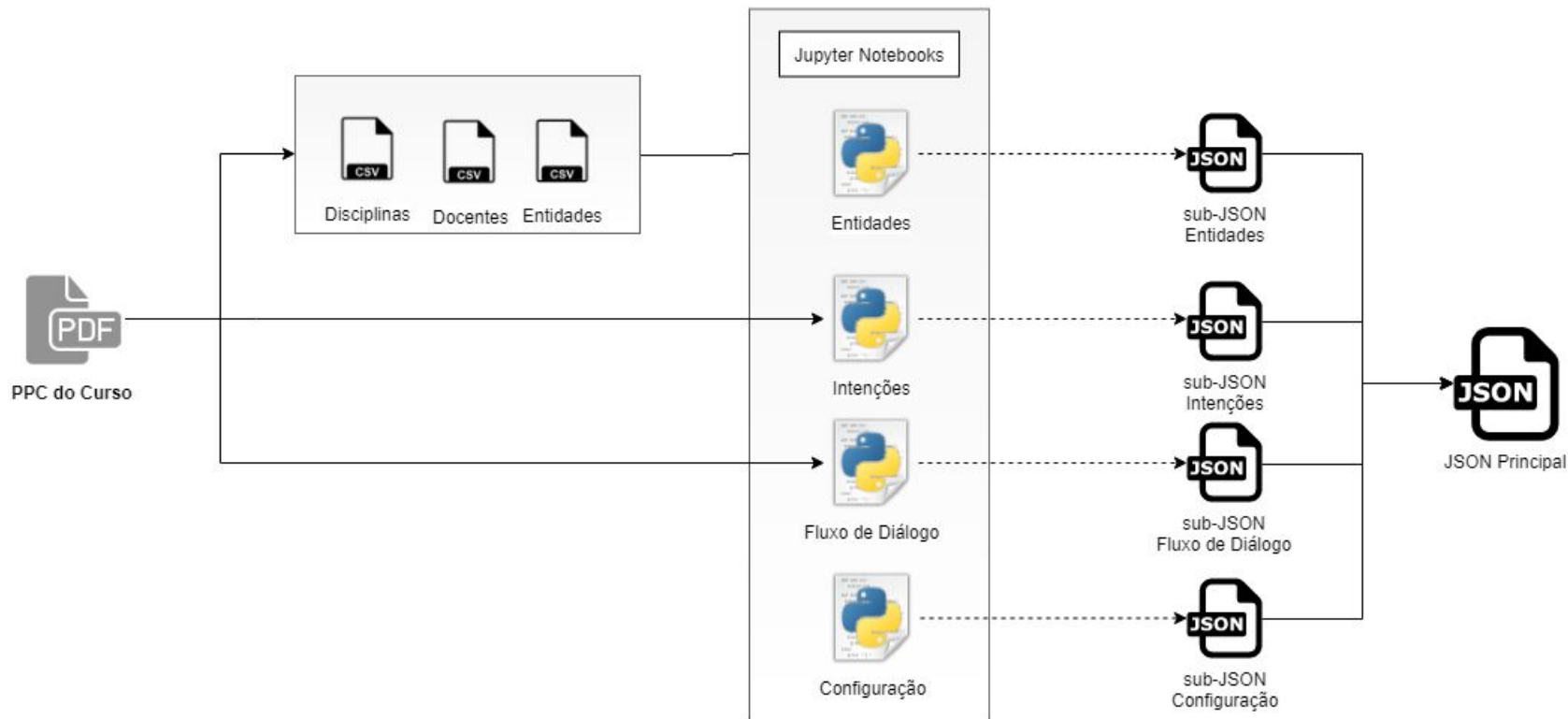


```
1 "entities": [{
2     "entity": "disciplinas",
3     "values": [
4         {
5             "type": "synonyms",
6             "value": "Introducao a Programacao de Computadores",
7             "synonyms": [
8                 "ESTCMP009",
9                 "LP1",
10                "IPC"
11            ]
12        },
13        ...
14    ]
15 }
```



```
1 "intents": [  
2   {  
3     "intent": "saber-carga-horaria-disciplina",  
4     "examples": [  
5       {  
6         "text": "Qual a carga horaria de ESTCMP009?"  
7       },  
8       {  
9         "text": "qual a ch de algebra 1?"  
10      },  
11      {  
12        "text": "Carga Horaria de LP1?"  
13      },  
14      {  
15        "text": "Qual a carga horaria da materia"  
16      }  
17    ],  
18    "description": "Usuario deseja sabe Carga Horaria"  
19  },  
20  ...  
21 ]]
```

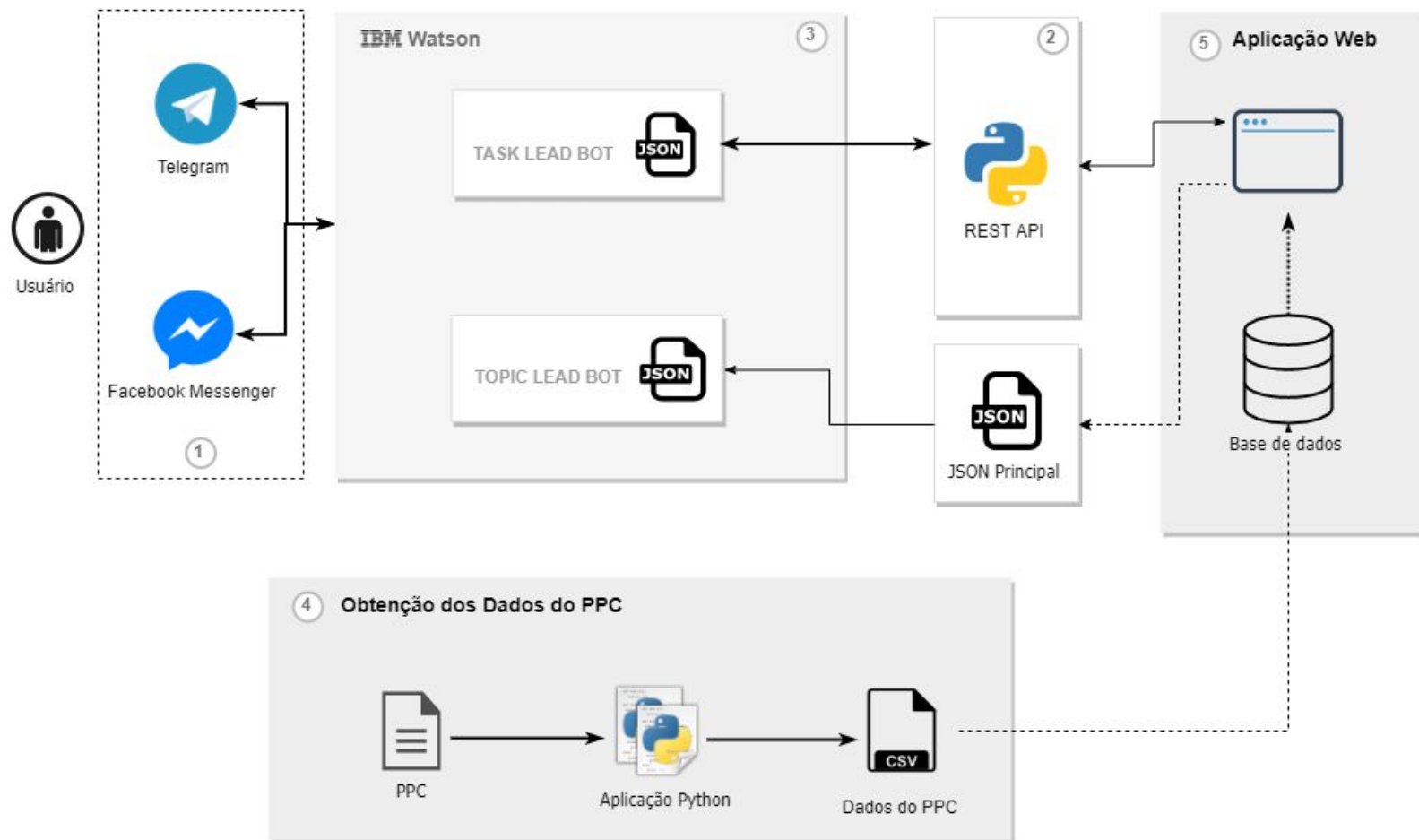
# Parsing para JSON





# Arquitetura do Chatbot

# Arquitetura do Chatbot - Visão Geral



# Muito Obrigado!



/luizbentes



luizbentes@gmail.com



@pydatamaneus

@pyladiesmanaus