

# AI & CHATBOT

Aula 10 – Extração de Conhecimento, NLU e  
Integração com Reddit

**Prof. Daniel Gomes**

Prof. Daniel Petrini

Prof. Guilherme Aldeia

Prof. Henrique Ferreira

Prof. Michel Fornaciali

Prof. Miguel Bozer

FIAP  
GRADUAÇÃO

# Bases de Conhecimento

Visão geral sobre Mineração de Dados

# Introdução a Extração de Conhecimento

- A extração ou recuperação de conhecimento é a tarefa de obter informação em bases de dados estruturadas ou não estruturadas. É uma tarefa ligada a **Mineração de Dados**, uma área da Ciência da Computação que alia técnicas de manipulação de arquivos e dados com Inteligência Artificial e estatística;
- Como parte do conhecimento humano está em linguagem natural (escrito em documentos de texto ou páginas HTML, por exemplo), o **Processamento de Linguagem Natural** é muitas vezes empregado em conjunto, dando origem a área de **Mineração de Texto**;
- Existem várias bibliotecas e técnicas para extrair conhecimento. Ao longo do curso veremos algumas delas;

# Introdução a Extração de Conhecimento



- A informação pode ser armazenada de diversas formas; Em geral temos dois tipos de banco de dados:
  - Relacional (SQL): dados são armazenados em tabelas;
  - Não relacional (NoSQL): dados podem ser armazenados em diferentes formatos;
- Além disso os dados podem estar armazenados de maneira:
  - Estruturada: tabelas
  - Semiestruturada: HTML, XML, JSON
  - Não estruturada: texto, imagens, áudio;

# Introdução a Extração de Conhecimento

- Informação pode estar em diferentes mídias!
- A informação pode estar estruturada, semi-estruturada ou não estruturada!

TABELA

Entrada	$x_1$	$x_2$	...	$x_n$	$y$	$\hat{y}$
1	70.52	30	...	0.584	90	100
2	60.96	27	...	1.254	81	90
...	...	...	...	...	...	...
$k$	97.48	35	...	0.758	122	120

OBJETO

JSON		Raw Data	Headers
Save		Copy	Collapse All Expand All Filter JSON
▼ intents:			
▼ 0:		intent:	"agradecimento"
▼ examples:			
▼ 0:		text:	"Agradeço o serviço"
▼ 1:		text:	"Esse bot tá show, obrigado"
▶ 2:			{...}
▶ 3:			{...}
▶ 4:			{...}
▶ 5:			{...}
▶ 6:			{...}
▶ description:			
▶ 1:			

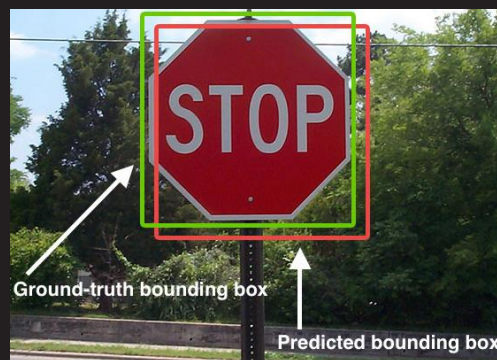
TEXTO

Esse campo de pesquisa ganhou muita notoriedade em 1986, quando David E. Rumelhart e James L. McClelland publicaram um livro que apresentou um modelo matemático computacional capaz de realizar um treinamento supervisionado dos neurônios artificiais. Esse algoritmo é chamado de **Backpropagation** e permite otimizações globais no modelo, sem restrições. Esse algoritmo também foi chamado de regra Delta generalizada, pois foi baseado na regra Delta, algoritmo de aprendizagem das redes Adalines.

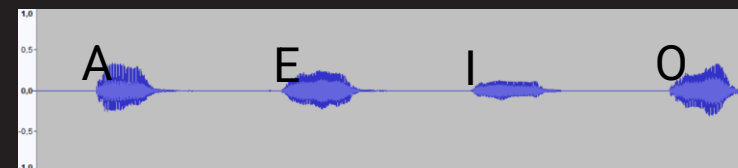
Foi a partir desses trabalhos e da criação de diversos Journals e conferências que muitas instituições fundaram institutos de pesquisas e programas educacionais que estudam redes neurais artificiais e modelos de aprendizagem.

Nos próximos tópicos vamos aprender como a rede neural pode realizar predições através do algoritmo **Feedforward** e o aprendizado ou ajustes dos pesos, com o algoritmo **Backpropagation**. Animado? Vamos começar!

IMAGEM

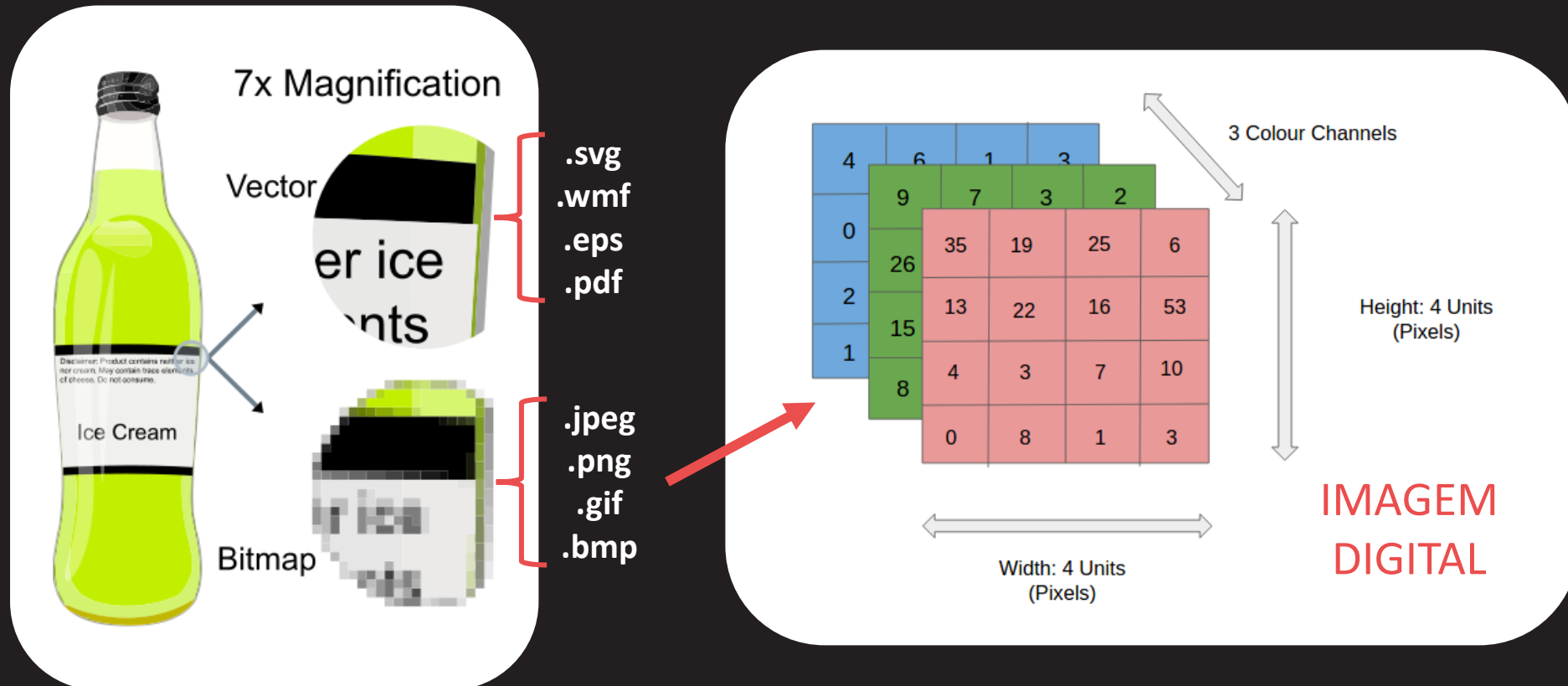


ÁUDIO



# Introdução a Extração de Conhecimento

- Informação pode estar em diferentes mídias!
- A informação pode estar estruturada, semi-estruturada ou não estruturada!



formatos de codificação

# Introdução a Extração de Conhecimento

- Como o texto é representado no computador?
- Strings são objetos em linguagem de programação usados para trabalhar com caracteres;
- Os caracteres (e mais recentemente, emojis) são imagens mapeadas para um código hexadecimal (e binário);
- O mapeamento hexadecimal mais conhecido é o ASCII (American Standard Code II). Para contemplar outras línguas (além do alfabeto latino) e incorporar emojis, temos o Unicode;
- Em memória, o Unicode pode ser UTF-8, UTF-16, UTF-32;

## ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]





# Introdução a Extração de Conhecimento

- Para algoritmos de IA precisamos modificar a forma como texto é representado a partir de análises e transformações;
  - As strings devem ser transformadas em outras representações numéricas;
  - Uma técnica muito usada na área de Processamento de Linguagem Natural é transformar as string em vetores numéricos, uma técnica chamada de **Embedding**;
  - Podemos ter Word Embedding quando representamos palavras por um vetor ou ainda Sentence Embedding quando representamos sentenças por vetores numéricos;
  - Existem várias formas de fazer isso, entre elas:

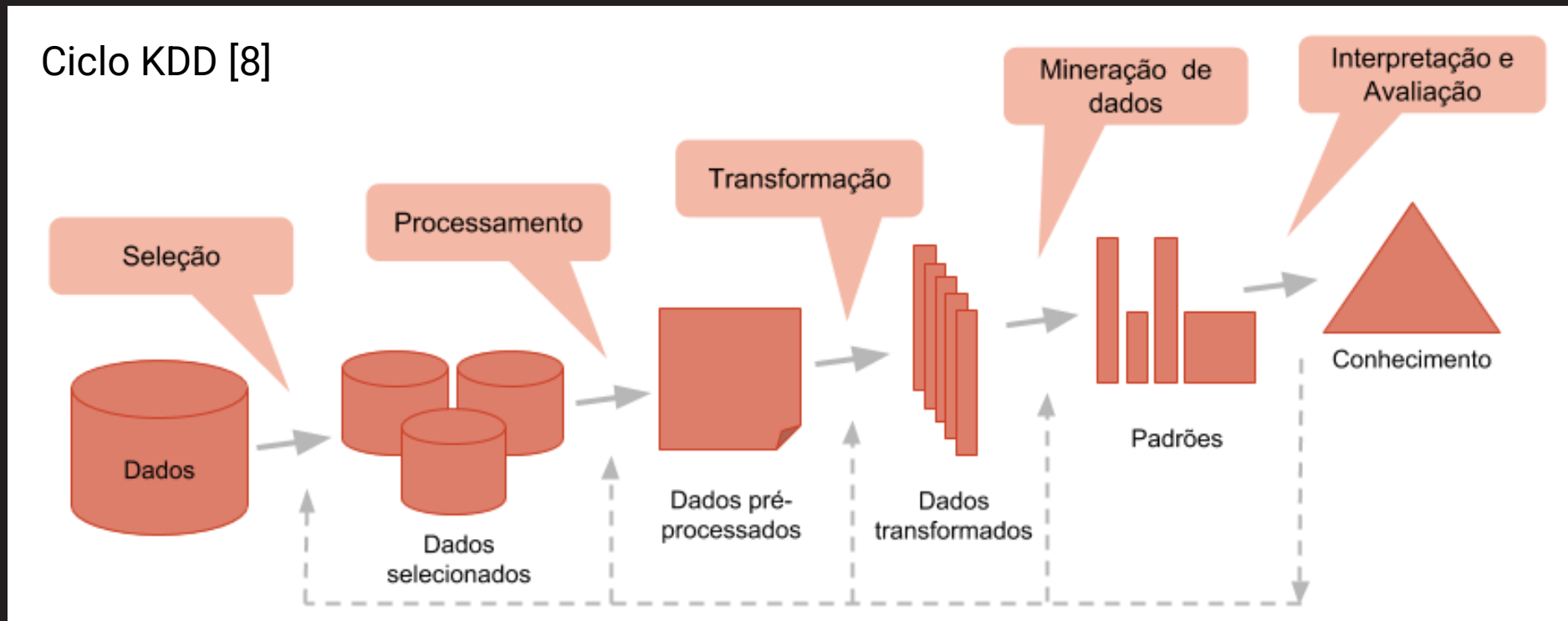
Frases: Bag of Words (BOW) ou o TF-IDF:

texto = "eu vou ao cinema hoje"												
	cada	um	vou	eu	amanhã	cinema	em	hoje	e	a	ao	em
texto_vetor:	0	0	1	1	0	1	0	1	0	0	1	0



# Introdução a Extração de Conhecimento

- Conhecimento é o produto final de uma descoberta em dados;
- Um dos processo de extração de informação de bases de dados é conhecido como **knowledge-discovery in databases – KDD**, e faz parte essencial do trabalho de Cientista de Dados;



# Natural Language Understanding

Uma introdução aos serviços de Mineração de Texto

# Natural Language Understanding

IBM Cloud

Procurar recursos e ofertas...

Q

Catálogo

Documentos

Suporte

Gerenciar

Henrique Ferreir...

Catálogo / Serviços /

Natural Language Understanding

IBM • Data da última atualização: 07/05/2021 • Docs • Docs da API

Criar

Sobre

Selecionar uma localização

Selecionar uma localização

Dallas (us-south)

Selecione um plano de precificação

Os preços exibidos não incluem impostos. Os preços mensais mostrados são para país ou local: [Estados Unidos](#)

Plano	Recursos	Precificação
Lite	<b>30.000 itens NLU por mês</b> 1 Modelo customizado Limite de taxa de API fixa. Consulte o plano Padrão para limite de taxa de API superior NOTA: um item NLU é baseado no número de unidades de dados enriquecidas e no	Grátis

Criar

Incluir na estimativa

Resumo

Natural Language Understanding

Grátis


Localização: Dallas

Plano: Lite


Nome do serviço: Natural Language Understanding-po

Grupo de recursos: Default

# Natural Language Understanding

 IBM Cloud

Procurar recursos e ofertas...




Catálogo


Documentos


Suporte


Gerenciar ▾

Henrique Ferreir...











[Lista de recursos](#) /

Natural Language Understanding-fiap  Ativo [Incluir tags](#) 

[Detalhes](#)

[Ações...](#) ▾

**Gerenciamento**


[Introdução](#)

[Credenciais de serviço](#)

[Plano](#)

[Conexões](#)

Start by viewing the tutorial

[Getting started tutorial](#) 


[API reference](#)


Plan

Lite


Upgrade

Credentials


[Download](#) 

[Show credentials](#) 

API key:

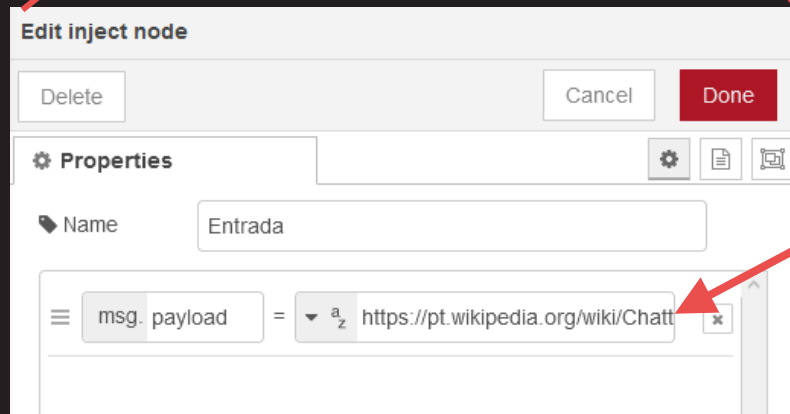
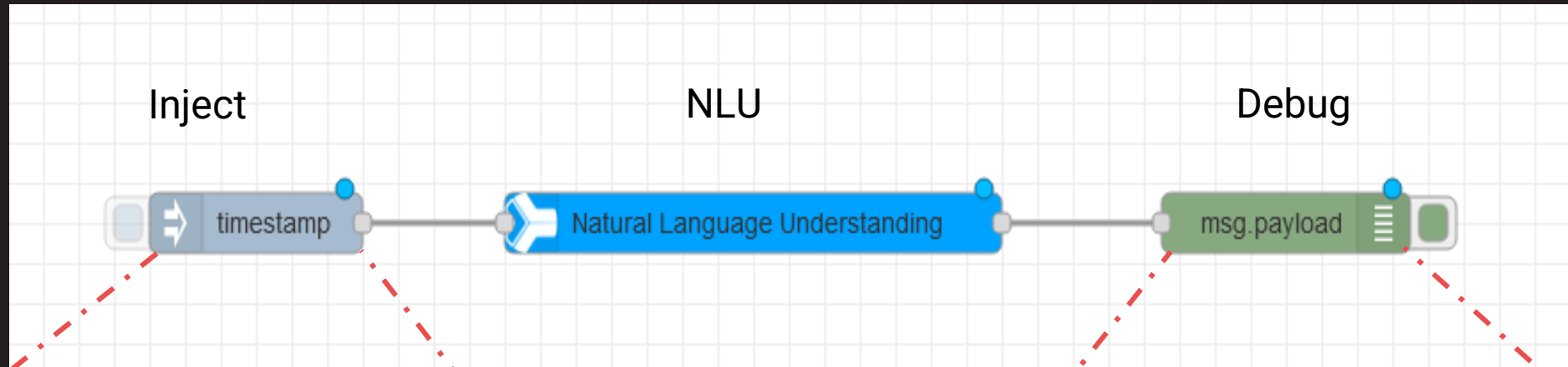
..... 

URL:

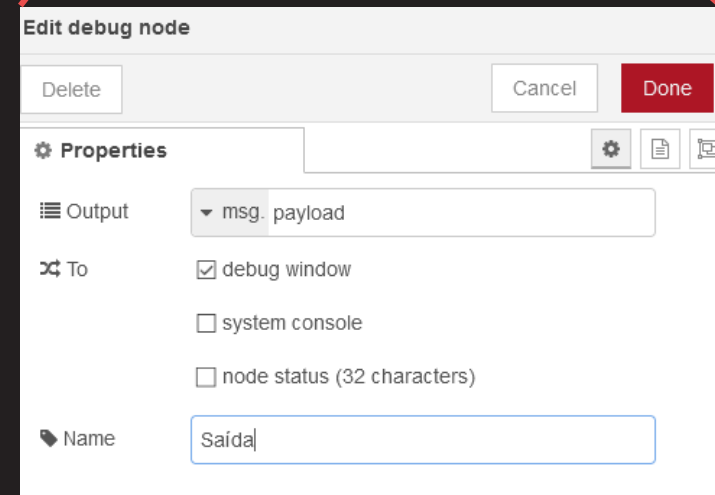
`https://api.us-south.natural-language-understanding.watson.cloud.ibm` 

<  >

# Natural Language Understanding + NodeRED



Vamos extrair dados de um site



# Natural Language Understanding + Node-RED

Edit Natural Language Understanding node

Delete Cancel Done

**Properties**

Name

Username

Password

or

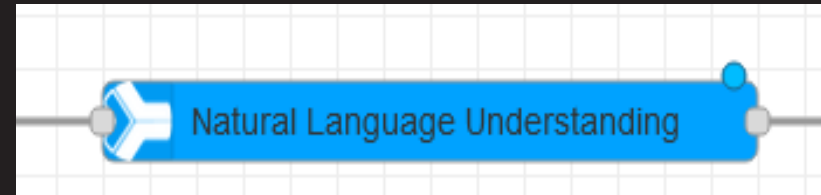
API Key

Service Endpoint

Extract the following features:

<input type="checkbox"/> Categories	<input type="checkbox"/> Keywords
<input type="checkbox"/> Concepts	<input type="checkbox"/> Metadata
<input type="checkbox"/> Document Emotion	<input type="checkbox"/> Relations
<input type="checkbox"/> Document Sentiment	<input type="checkbox"/> Semantic Roles
<input type="checkbox"/> Entities	<input type="checkbox"/> Syntax

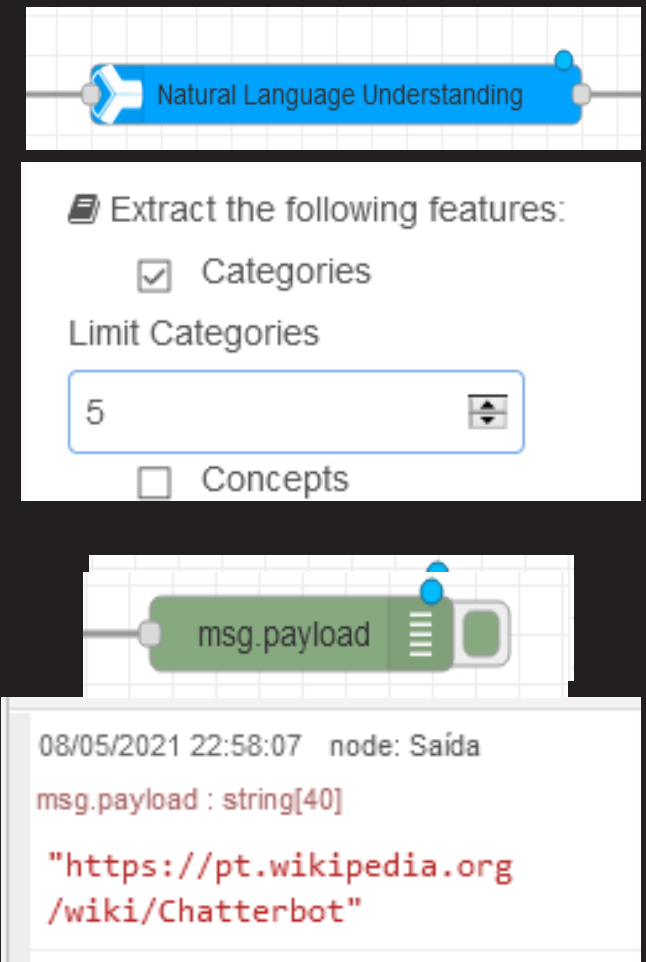
Limit Text Characters



- Credenciais do serviço
- Característica dos dados a serem extraídos. Existem vários tipos de informação que podem ser capturadas, vamos testar algumas;

# Natural Language Understanding + Node-RED

- Vamos começar entendendo o que são **Categorias**. Vamos pedir para o NLU reconhecer 5 categorias no texto do site.
- Todas as Categorias que o NLU pode capturar estão aqui <https://cloud.ibm.com/docs/natural-language-understanding?topic=natural-language-understanding-categories-hierarchy>
- Entretanto, se rodarmos assim, vamos obter no nó de debug como saída exatamente o link de entrada. Como podemos resolver isso?



The image shows a Node-RED workflow. At the top, a blue node labeled "Natural Language Understanding" is connected to a green "msg.payload" node. Below the NLU node, a configuration panel is visible with the following settings:

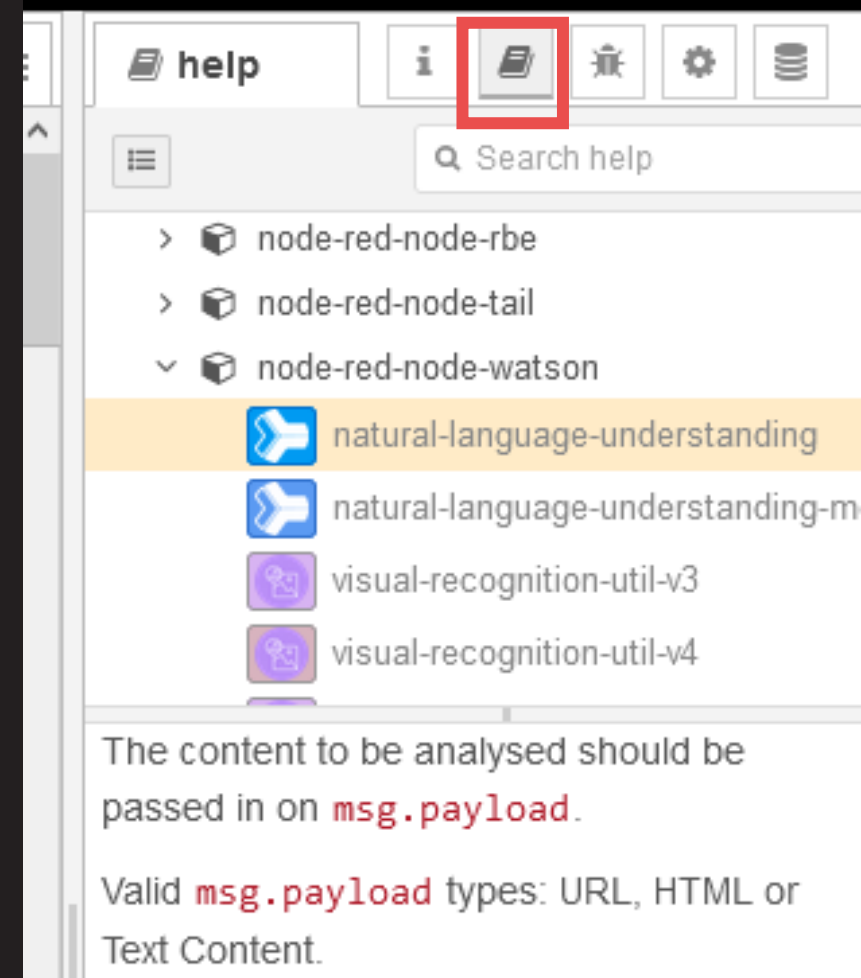
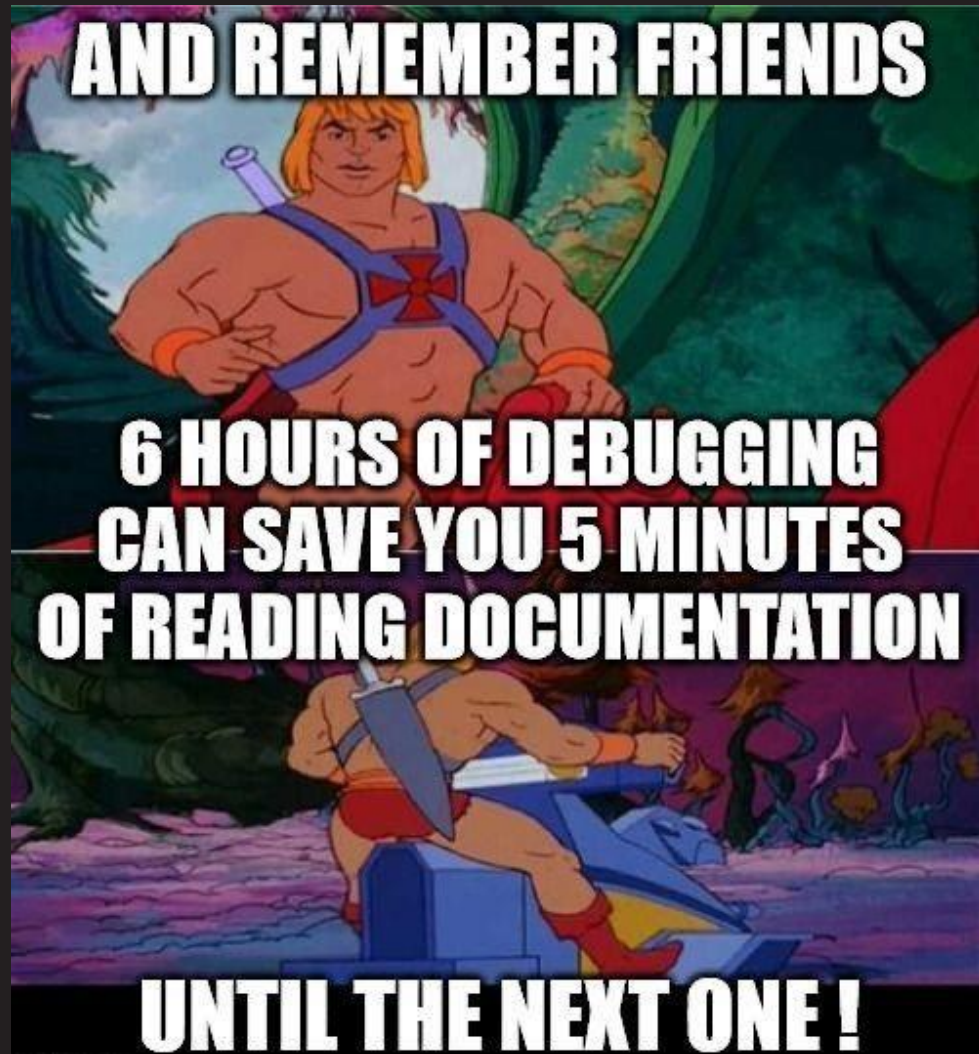
- Extract the following features:
  - ☒ Categories
  - Limit Categories: 5 (in a text input field)
  - ☐ Concepts

Below the configuration panel, a debug console window shows the output of the "msg.payload" node:

```
08/05/2021 22:58:07 node: Saída
msg.payload : string[40]
"https://pt.wikipedia.org/wiki/Chatterbot"
```

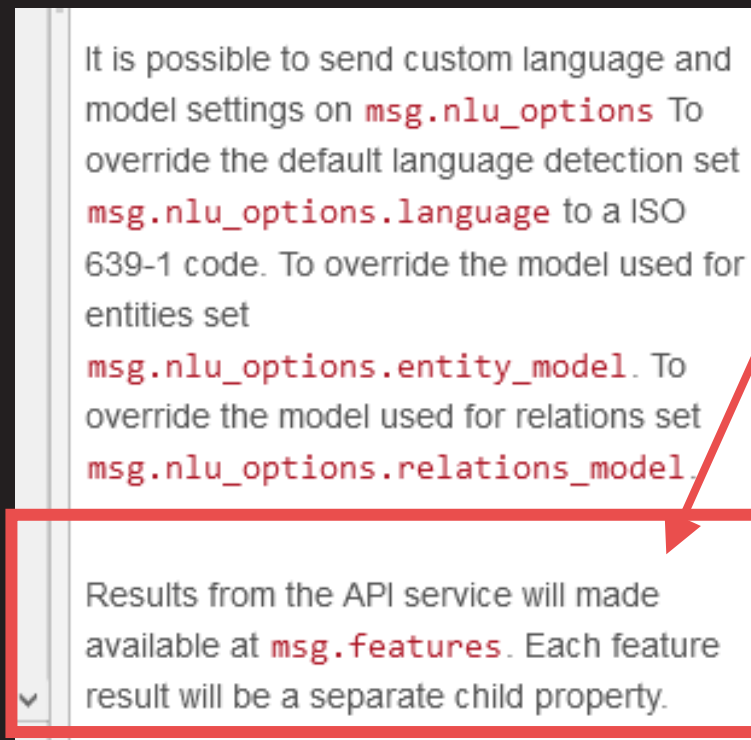
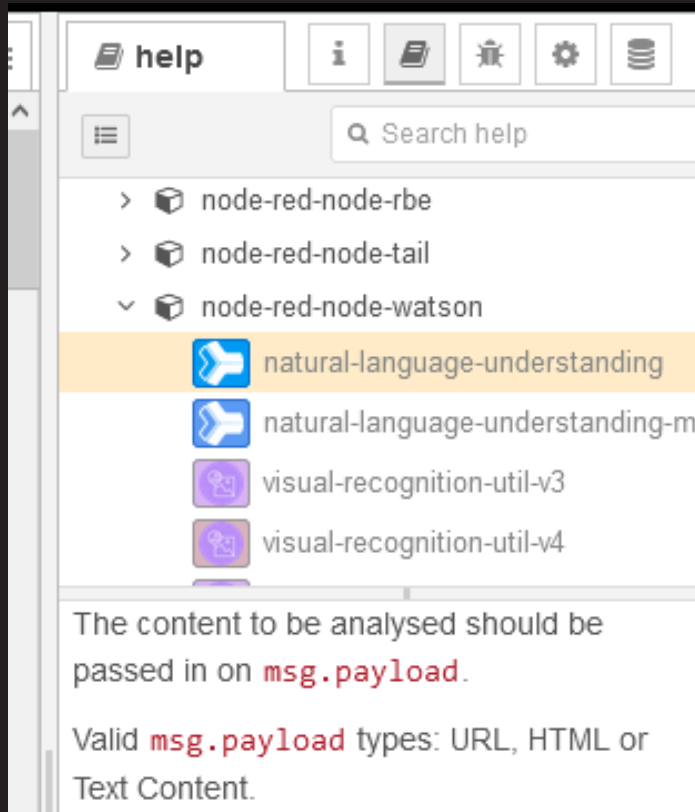


# Natural Language Understanding + Node-RED



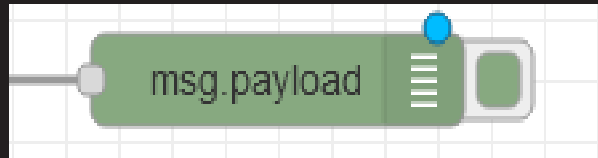
# Natural Language Understanding + Node-RED

- Vamos olhar nas informações sobre o node de NLU. Clique no menu de help na aba direita da interface do Node-RED. Selecione o nó de NLU.



- Observe que a resposta é colocada na variável `features` por padrão;
- Vamos mudar no nó de debug para mostrar essa variável.

# Natural Language Understanding + Node-RED



**Edit debug node**

Delete Cancel Done

**Properties**

Output

To ☒ debug window  
☐ system console  
☐ node status (32 characters)

Name

```
08/05/2021 23:49:08 node: Saída
msg.features : Object
  object
  usage: object
  retrieved_url:
    "https://pt.wikipedia.org
    /wiki/Chatterbot"
  language: "pt"
```

```
  categories: array[5]
    0: object
    1: object
    2: object
    3: object
    4: object
```


# Natural Language Understanding + Node-RED

```
▼ categories: array[5]
  ▼ 0: object
    score: 0.967183
    label: "/technology and computing/internet technology/social network"
  ▼ 1: object
    score: 0.890503
    label: "/technology and computing/software"
  ▼ 2: object
    score: 0.832727
    label: "/business and industrial/business software"
  ▼ 3: object
    score: 0.798418
    label: "/technology and computing/operating systems"
  ▼ 4: object
    score: 0.748668
    label: "/technology and computing/hardware/computer"
```

- Estamos olhando a página de Chatbot da Wikipedia (<https://pt.wikipedia.org/wiki/Chatterbot>);
- Olhe o conteúdo da página e as categorias encontradas pelo NLU. Os resultados fazem sentido?
- Observe o score: ele indica a confiança que o NLU tem de que a página contém essas categorias (1 significa 100% de confiança, de forma que 0.8905 significa 89,05%).

# NLU + Node-RED

- Recuperando os **Conceitos** agora:

 Extract the following features:

<input type="checkbox"/> Categories	<input type="checkbox"/> Keywords
<input checked="" type="checkbox"/> Concepts	<input type="checkbox"/> Metadata
Maximum Concepts	<input type="checkbox"/> Relations
<input type="text" value="8"/>	<input type="checkbox"/> Semantic Role
<input type="checkbox"/> Document Emotion	<input type="checkbox"/> Syntax
<input type="checkbox"/> Document Sentiment	
<input type="checkbox"/> Entities	


Limit Text Characters

```
msg.features : Object
  ▼ object
    ▶ usage: object
      retrieved_url:
        "https://pt.wikipedia.org/wiki/Chatterbot"
      language: "pt"
    ▼ concepts: array[8]
      ▶ 0: object
      ▶ 1: object
      ▶ 2: object
      ▶ 3: object
      ▶ 4: object
      ▶ 5: object
      ▶ 6: object
      ▶ 7: object
```

```
msg.features : Object
  ▼ object
    ▶ usage: object
      retrieved_url:
        "https://pt.wikipedia.org/wiki/Chatterbot"
      language: "pt"
    ▼ concepts: array[8]
      ▼ 0: object
        text: "Robô"
        relevance: 0.991631
        dbpedia_resource:
          "http://pt.dbpedia.org/resource/Robô"
      ▶ 1: object
      ▶ 2: object
      ▶ 3: object
      ▶ 4: object
      ▼ 5: object
        text: "Computador"
        relevance: 0.940756
        dbpedia_resource:
          "http://pt.dbpedia.org/resource/Computador"
      ▶ 6: object
      ▶ 7: object
```

# NLU + Node-RED

- Recuperando **Emoções** presentes no documento:

 Extract the following features:

<input type="checkbox"/> Categories	<input type="checkbox"/> Keywords
<input type="checkbox"/> Concepts	<input type="checkbox"/> Metadata
<input checked="" type="checkbox"/> Document Emotion	<input type="checkbox"/> Relations
Emotion Targets	
<input type="text"/>	
<input type="checkbox"/> Document Sentiment	<input type="checkbox"/> Semantic Roles
<input type="checkbox"/> Entities	<input type="checkbox"/> Syntax

Limit Text Characters

- Infelizmente o NLU não reconhece emoções para textos em português;




```
msg : string[29]
```

```
"unsupported text language: pt"
```

# NLU + Node-RED

- Recuperando **Sentimentos** (ou impressão) presentes no documento:

 Extract the following features:

<input type="checkbox"/> Categories	<input type="checkbox"/> Keywords
<input type="checkbox"/> Concepts	<input type="checkbox"/> Metadata
<input type="checkbox"/> Document Emotion	<input type="checkbox"/> Relations
<input checked="" type="checkbox"/> Document Sentiment	<input type="checkbox"/> Semantic Roles
	<input type="checkbox"/> Syntax

Sentiment Targets

☐ Entities

Limit Text Characters

- Reconhece se o texto é “positivo”, “negativo” ou “neutro”:

```
msg.features : Object
  ▾ object
    ▶ usage: object
    ▾ sentiment: object
      ▾ document: object
        score: 0.6289
        label: "positive"
      retrieved_url:
        "https://pt.wikipedia.org/wiki/Chatterbot"
      language: "pt"
```



# NLU + Node-RED

- Recuperando **Entidades** (vamos limitar em 10):

📖 Extract the following features:

<input type="checkbox"/> Categories	<input type="checkbox"/> Keywords
<input type="checkbox"/> Concepts	<input type="checkbox"/> Metadata
<input type="checkbox"/> Document Emotion	<input type="checkbox"/> Relations
<input type="checkbox"/> Document Sentiment	<input type="checkbox"/> Semantic Roles
<input checked="" type="checkbox"/> Entities	<input type="checkbox"/> Syntax
<input type="checkbox"/> Entity Emotion	
<input type="checkbox"/> Entity Sentiment	

Maximum Entities

10

Limit Text Characters


0

```
msg.features : Object
  ▾ object
    ▸ usage: object
      retrieved_url:
        "https://pt.wikipedia.org/wiki/Chatterbot"
      language: "pt"
    ▾ entities: array[10]
      ▸ 0: object
      ▸ 1: object
      ▸ 2: object
      ▸ 3: object
      ▸ 4: object
      ▸ 5: object
      ▸ 6: object
      ▸ 7: object
      ▸ 8: object
      ▸ 9: object
```

```
▾ entities: array[10]
  ▾ 0: object
    type: "Organization"
    text: "IBM"
    relevance: 0.949058
    ▸ disambiguation: object
      count: 3
      confidence: 0.999944
  ▾ 1: object
    type: "Person"
    text: "Índice"
    relevance: 0.586148
    count: 1
    confidence: 0.362646
  ▾ 2: object
    type: "Person"
    text: "Michael Mauldin"
    relevance: 0.570978
    count: 1
    confidence: 0.998299
```

# NLU + Node-RED

- Se selecionarmos Entidades de Emoção e Sentimento obtemos:

 Extract the following features:

<input type="checkbox"/> Categories	<input type="checkbox"/> Keywords
<input type="checkbox"/> Concepts	<input type="checkbox"/> Metadata
<input type="checkbox"/> Document Emotion	<input type="checkbox"/> Relations
<input type="checkbox"/> Document Sentiment	<input type="checkbox"/> Semantic Roles
<input checked="" type="checkbox"/> Entities	<input type="checkbox"/> Syntax
<input checked="" type="checkbox"/> Entity Emotion	
<input checked="" type="checkbox"/> Entity Sentiment	

Maximum Entities

10

Limit Text Characters


0

```
▼ entities: array[10]
  ▼ 0: object
    type: "Organization"
    text: "IBM"
    ▼ sentiment: object
      score: 0
      label: "neutral"
      relevance: 0.949058
      ► disambiguation: object
        count: 3
        confidence: 0.999944
  ▼ 1: object
    type: "Person"
    text: "Índice"
    ▼ sentiment: object
      score: 0
      label: "neutral"
      relevance: 0.586148
      count: 1
      confidence: 0.362646
  ▼ 2: object
    type: "Person"
    text: "Michael Mauldin"
```


- Perceba que ele não captura Emoção na entidades por causa da limitação da língua suportada;

# NLU + Node-RED

- **Keywords** (podemos pegar emoção e sentimentos associados também):

 Extract the following features:


<input type="checkbox"/> Categories	<input checked="" type="checkbox"/> Keywords
<input type="checkbox"/> Concepts	<input checked="" type="checkbox"/> Keyword Emotion
<input type="checkbox"/> Document Emotion	<input checked="" type="checkbox"/> Keyword Sentiment
<input type="checkbox"/> Document Sentiment	<input type="checkbox"/> Maximum Keywords
<input type="checkbox"/> Entities	<input type="checkbox"/> Metadata
	<input type="checkbox"/> Relations
	<input type="checkbox"/> Semantic Roles
	<input type="checkbox"/> Syntax

10 

```
msg.features : Object
▼ object
  ▶ usage: object
    retrieved_url:
      "https://pt.wikipedia.org
      /wiki/Chatterbot"
    language: "pt"
  ▼ keywords: array[10]
    ▼ 0: object
      text: "grandes benefícios dos
      chatbots"
      ▼ sentiment: object
        score: 0.949571
        label: "positive"
        relevance: 0.563139
        count: 1
    ▼ 1: object
      text: "conferência do Facebook"
      ▶ sentiment: object
        relevance: 0.547331
        count: 1
    ▼ 2: object
      text: "notificações de
      empresas"
      ▶ sentiment: object
        relevance: 0.544742
        count: 1
```

# NLU + Node-RED

- **Metadados:** informações nos campos de metadados da páginas HTML.

 Extract the following features:

<input type="checkbox"/> Categories	<input type="checkbox"/> Keywords
<input type="checkbox"/> Concepts	<input checked="" type="checkbox"/> Metadata
<input type="checkbox"/> Document Emotion	<input type="checkbox"/> Relations
<input type="checkbox"/> Document Sentiment	<input type="checkbox"/> Semantic Roles
<input type="checkbox"/> Entities	<input type="checkbox"/> Syntax

Limit Text Characters

```
msg.features : Object
```

```
▼ object
```

```
► usage: object
```

```
  retrieved_url: "https://pt.wikipedia.org  
/wiki/Chatterbot"
```

```
▼ metadata: object
```

```
  title: "Chatterbot - Wikipédia, a  
enciclopédia livre"
```

```
  publication_date: ""
```

```
  image: "https://upload.wikimedia.org  
/wikipedia/commons  
/9/98/GNU_Emacs_ELIZA_example.png"
```


```
  feeds: array[0]
```

```
  authors: array[0]
```

```
  language: "pt"
```

# NLU + Node-RED

- **Relações** : reconhecer quando duas entidades estão relacionadas e identificar o tipo de relação entre elas.

 Extract the following features:

<input type="checkbox"/> Categories	<input type="checkbox"/> Keywords
<input type="checkbox"/> Concepts	<input type="checkbox"/> Metadata
<input type="checkbox"/> Document Emotion	<input checked="" type="checkbox"/> Relations
<input type="checkbox"/> Document Sentiment	<input type="checkbox"/> Semantic Roles
<input type="checkbox"/> Entities	<input type="checkbox"/> Syntax

Limit Text Characters

```
msg.features : Object
▼ object
  ▶ usage: object
    retrieved_url: "https://pt.wikipedia.org/wiki/Chatterbot"
  ▼ relations: array[9]
    ▼ 0: object
      type: "ownerOf"
      sentence: "[2] Chatbots em aplicativos de mensagens [ editar | editar código-fonte ] Durante a F8, conferência do Facebook voltada para desenvolvedores, a rede social anunciou o lançamento da sua loja de robôs."
      score: 0.424528
    ▼ arguments: array[2]
      ▼ 0: object
        text: "sua"
        ▶ location: array[2]
        ▶ entities: array[1]
      ▼ 1: object
        type: "locatedAt"
        sentence: "Em 2017, o WhatsApp em parceria com algumas empresas, inclusive brasileiras, começou a testar o envio de mensagens para
```

# NLU + Node-RED

- Função **semântica** :

📄 Extract the following features:

<input type="checkbox"/> Categories	<input type="checkbox"/> Keywords
<input type="checkbox"/> Concepts	<input type="checkbox"/> Metadata
<input type="checkbox"/> Document Emotion	<input type="checkbox"/> Relations
<input type="checkbox"/> Document Sentiment	<input checked="" type="checkbox"/> Semantic Roles
<input type="checkbox"/> Entities	<input type="checkbox"/> Semantic Entities
	<input type="checkbox"/> Semantic Keywords

Maximum Semantic Roles

10

☐ Syntax

## Funções de semântica

Analise sentenças no formato assunto-ação-objeto e identifique entidades e palavras-chave que sejam sujeitos ou objetos de uma ação. Por exemplo:

### Entrada

Texto: "Em 2011, o Watson competiu no Jeopardy!"

### Resposta

Sujeito: Watson

Ação: competiu

Objeto: no Jeopardy

- Infelizmente o NLU não reconhece relações semânticas para textos em português;



09/05/2021 12:02:04 node: NLU

msg : string[29]

"unsupported text language: pt"

09/05/2021 12:02:04 node: NLU

msg : string[29]

"unsupported text language: pt"

# NLU + Node-RED

## • Sintaxe:

📖 Extract the following features:

- |   |  |
|---|--|
| <input type="checkbox"/> Categories         | <input type="checkbox"/> Keywords                    |
| <input type="checkbox"/> Concepts           | <input type="checkbox"/> Metadata                    |
| <input type="checkbox"/> Document Emotion   | <input type="checkbox"/> Relations                   |
| <input type="checkbox"/> Document Sentiment | <input type="checkbox"/> Semantic Roles              |
| <input type="checkbox"/> Entities           | <input checked="" type="checkbox"/> Syntax           |
|   | <input checked="" type="checkbox"/> Syntax Sentences |
|   | <input checked="" type="checkbox"/> Lemma Tokens     |
|   | <input checked="" type="checkbox"/> Part of Speech   |

```
▼ object
  ▶ usage: object
  ▼ syntax: object
    ▶ tokens: array[2190]
    ▶ sentences: array[101]
  retrieved_url: "https://pt.wikipedia.org/wiki/Chatterbot"
  language: "pt"
```

```
▼ syntax: object
  ▶ tokens: array[2190]
  ▼ sentences: array[101]
    ▼ [0 ... 9]
      ▼ 0: object
        text: "Chatbot (ou chatterbot) é um programa de computador que tenta simular um ser humano na conversação com as pessoas."
        ▶ location: array[2]
      ▶ 1: object
```



# Bases de Conhecimento Disponíveis

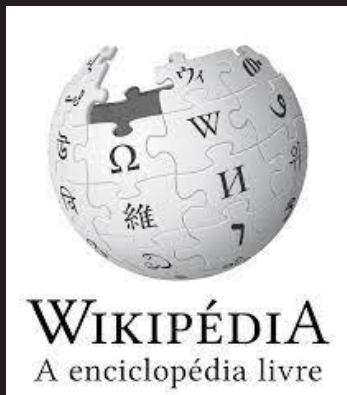
Exemplo de utilização do Reddit

# Bases de Conhecimento

- Na Internet existem várias bases de conhecimento que podem ser utilizadas para deixar um bot mais inteligente e/ou para treinar algoritmos de IA;
- Sites como portais de notícias e enciclopédias podem ser utilizados; às vezes é possível integrar buscadores como Google e Bing;
- Fóruns e redes sociais também são bases de dados riquíssimas e de difícil manipulação (lembre-se da necessidade de processar dados não estruturados codificados em língua natural);

# Bases de Conhecimento

- Já mostramos como buscar em páginas da Wikipédia através da URL. Agora vamos tentar extrair dados de redes sociais e fóruns. De vez de buscar utilizando a URL vamos utilizar a própria API desses sites. Apesar de ser possível utilizar todos eles, alguns necessitam de um cadastro de desenvolvedor que pode demorar. Por isso vamos optar por explorar o Reddit.



# Criando aplicativo no Reddit

- Entre no site do Reddit e crie uma conta <https://www.reddit.com/>



- Entre no seu email cadastrado e redefina a senha (password) e seu nome de usuário (username). Você vai precisar de ambos;
- Para criar o aplicativo, vá em <https://ssl.reddit.com/prefs/apps/>

# Criando aplicativo no Reddit

- Uma vez na página, observe a parte de criação. Escolha script para uso pessoal e dê um nome para o aplicativo. Na aba de redirect url você pode escrever <http://localhost:8080>. Clique em creat app.

**create application**

Please [read the API usage guidelines](#) before creating your application. After creating, you will be required to [register](#) for production API use.

**name**

☐ web app      A web based application

☐ installed app      An app intended for installation, such as on a mobile phone

☒ script      Script for personal use. Will only have access to the developers accounts

**description**

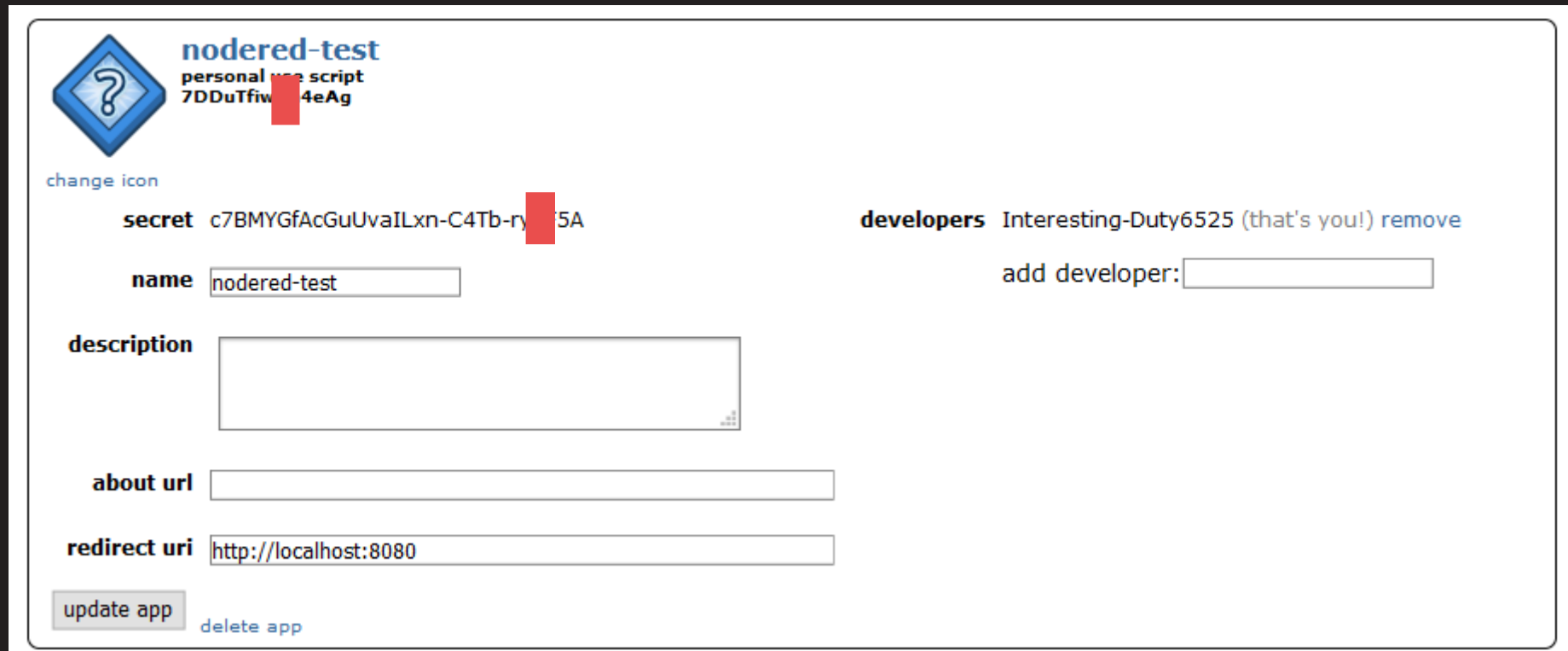
**about url**

**redirect uri**

application created

# Criando aplicativo no Reddit

- O resultado será:



The screenshot shows the Reddit application management page for an app named 'nodered-test'. The app is categorized as 'personal use script' with a blue diamond icon containing a question mark. The secret key is 'c7BMYGfAcGuUvaILxn-C4Tb-ny5A'. The developers list includes 'Interesting-Duty6525 (that's you!)'. The form includes fields for the app name, description, about URL, and redirect URI (set to 'http://localhost:8080'). There are buttons for 'update app' and 'delete app', and a link to 'change icon'.

**nodered-test**  
personal use script  
7DDuTfiw4eAg

[change icon](#)

**secret** c7BMYGfAcGuUvaILxn-C4Tb-ny5A

**developers** Interesting-Duty6525 (that's you!) [remove](#)

**name**

**add developer:**

**description**

**about url**

**redirect uri**

[delete app](#)

# Criando aplicativo no Reddit

- Atenção as informações:

**client ID**

**client secret**



The screenshot shows the Reddit application management interface. At the top, the application name 'nodered-test' is displayed next to a blue diamond icon with a question mark. Below this, the 'client ID' (7DDuTfiw4eAg) and 'client secret' (secret c7BMYGfAcGuUvaILxn-C4Tb-ry5A) are shown in red boxes with red arrows pointing to them from labels above. The 'name' field contains 'nodered-test'. The 'description' field is empty. The 'about url' field is empty. The 'redirect uri' field contains 'http://localhost:8080'. On the right, the 'developers' section lists 'Interesting-Duty6525 (that's you!)' with a 'remove' link. Below this is an 'add developer:' input field. At the bottom left, there are 'update app' and 'delete app' buttons.

change icon

**secret** c7BMYGfAcGuUvaILxn-C4Tb-ry5A

**developers** Interesting-Duty6525 (that's you!) [remove](#)

add developer:

**name**

**description**

**about url**

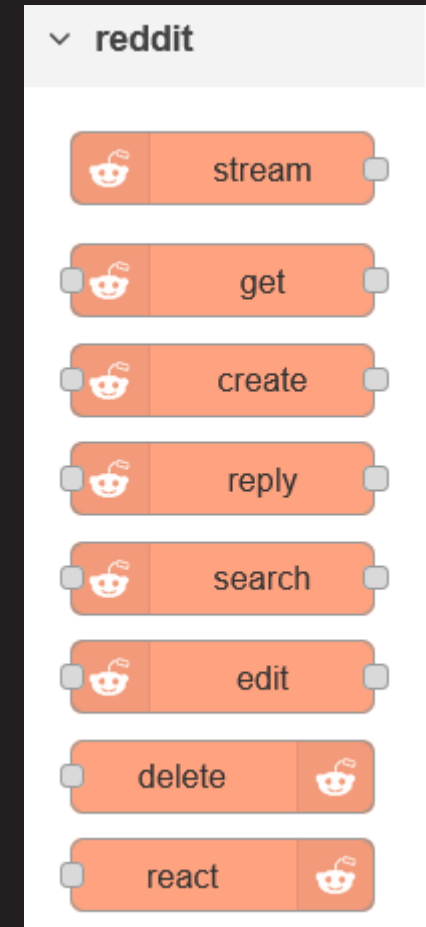
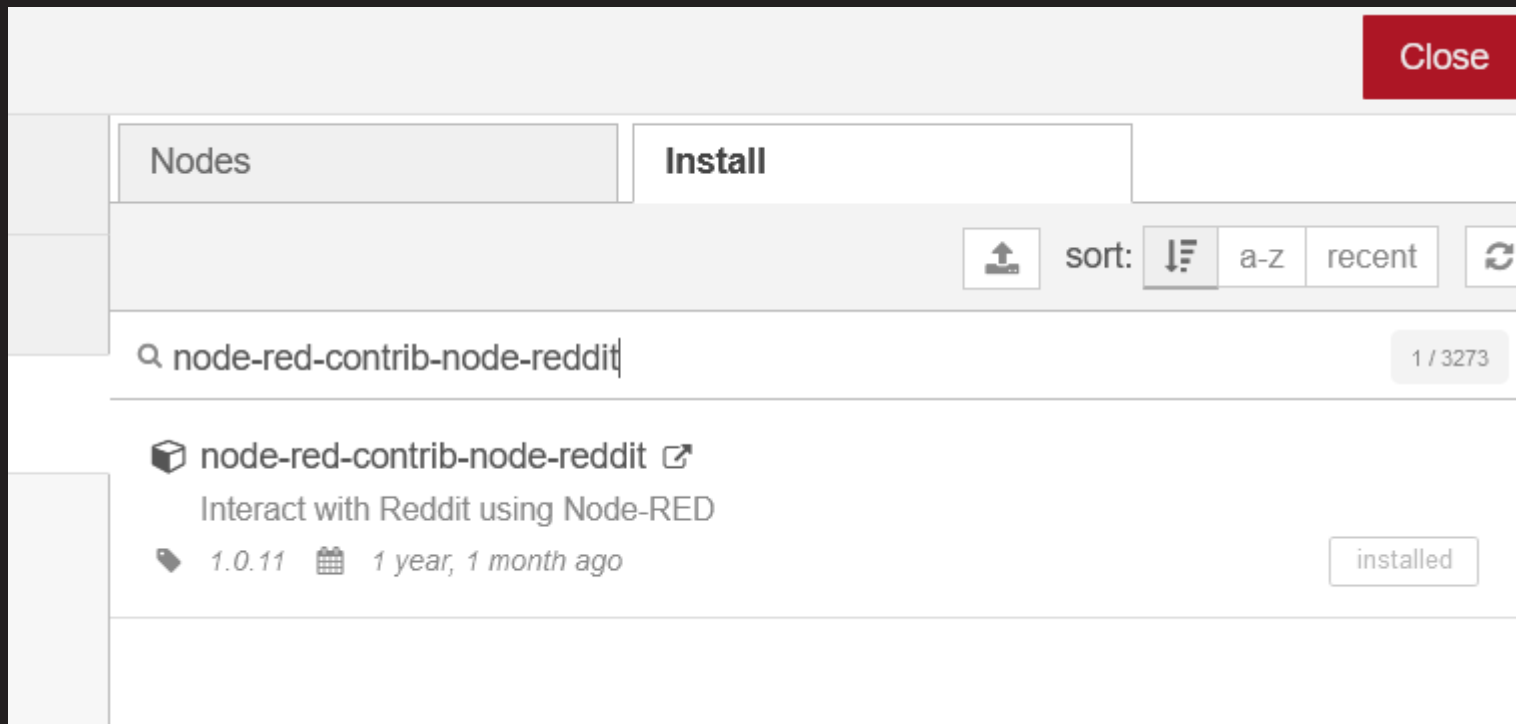
**redirect uri**

[delete app](#)



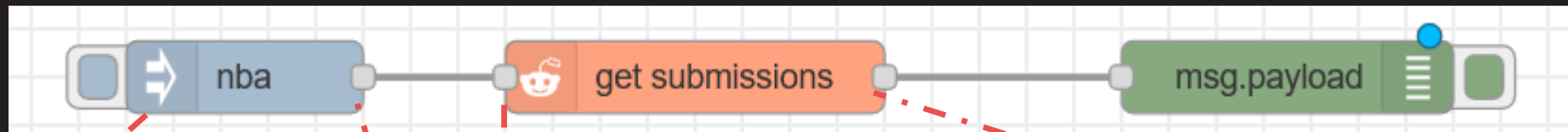
# Instalando nós Reddit no Node-RED

- Instale o node-red-contrib-node-reddit



# Criando fluxo Node-RED + Reddit

- Vamos usar o nó get para pegar informações sobre NBA:



Preencha as credenciais corretamente

**Edit inject node**

Delete

**Properties**

Name: Name

msg. payload = a\_z nba

**Edit get node**

Delete Cancel Done

**Properties**

credentials: BotBuscaReddit

name: name

type: submission

source: subreddit

subreddit: {{payload}}

sort: hot

limit: 3

**BotBuscaReddit**

name: BotBuscaReddit

client ID: \_\_PWRD\_\_

client secret: \_\_PWRD\_\_

user agent: nodered-test

authentication: username/password

username: ferreirahds

password: .....

# Criando fluxo Node-RED + Reddit

- Dê Deploy e injete a mensagem de NBA. Esse será o subreddit pesquisados. A mensagem de retorno será algo como:

▼ *object*

```
approved_at_utc: null
subreddit: "nba"
▶ selftext: "After beating the
clippers today, the Knicks now
have a better record than the
Lakers.␣␣Hats off to the
Knicks, and what Thibs has
done with this roster. They've
been the best story all year.
␣␣If someone told you two
years ago that the Knicks
would have a better record
than the Lakers you'd never
believe them in a million
years lol.␣␣How far do you
have the Knicks going in the
playoffs?"
```

```
author_fullname: "t2_9xyw8uqe"
saved: false
mod_reason_title: null
gilded: 2
clicked: false
title: "The New York Knicks
now have a better record than
the defending champion Los
Angeles Lakers"
link_flair_richtext: array[0]
subreddit_name_prefixed:
"r/nba"
```

```
hidden: false
pwls: 6
link_flair_css_class: null
downs: 0
thumbnail_height: null
top_awarded_type: "ACTIVE"
hide_score: false
name: "t3_n8omj9"
quarantine: false
link_flair_text_color: "dark"
upvote_ratio: 0.93
author_flair_background_color:
"#edeff1" 
```

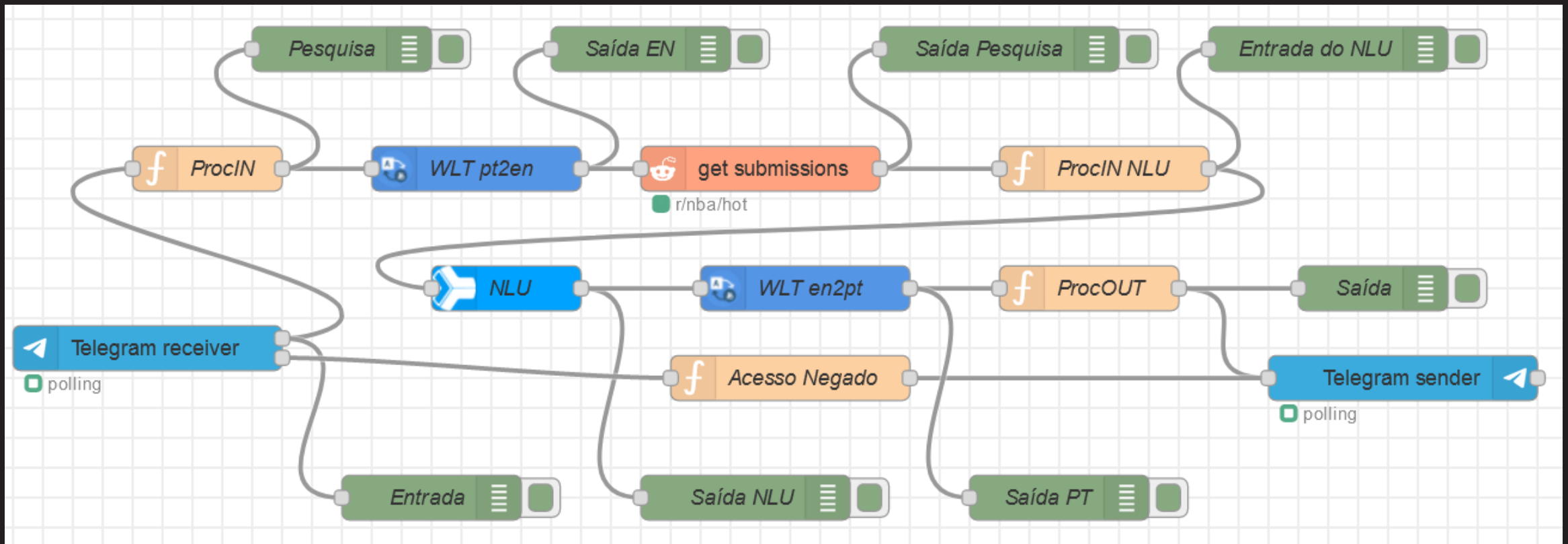
...

# Deixando o bot mais inteligente

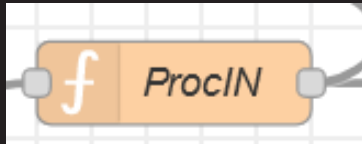
Exemplo de uso do NLU

# Telegram + Translator + Reddit + NLU

- Vamos fazer um fluxo para buscar informações no fórum Reddit:



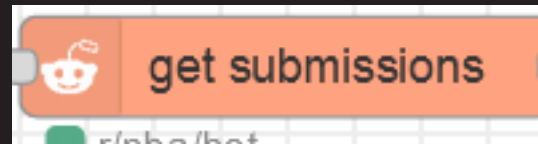
# Telegram + Translator + Reddit + NLU



Name: ProcIN

Setup On Start

```
1 msg.params={
2   id : msg.payload.chatId
3 };
4 msg.payload = msg.payload.content;
5 return msg;
```



Properties

credentials: BotBuscaReddit

name: name

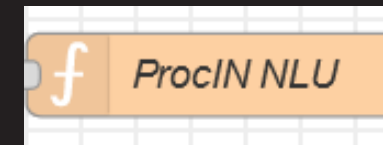
type: submission

source: subreddit

subreddit: {{payload}}

sort: hot

limit: 3




Properties

Name: ProcIN NLU


Setup On Start On Me

```
1 msg.autor = {
2   nome : msg.payload.author_fullname,
3 };
4 msg.payload = msg.payload.title
5 return msg;
```

# Telegram + Translator + Reddit + NLU

 Extract the following features:

<input checked="" type="checkbox"/> Categories	<input checked="" type="checkbox"/> Keywords
Limit Categories 3	<input checked="" type="checkbox"/> Keyword Emotion
<input checked="" type="checkbox"/> Concepts	<input checked="" type="checkbox"/> Keyword Sentiment
Maximum Concepts 5	Maximum Keywords 10
<input type="checkbox"/> Document Emotion	<input type="checkbox"/> Metadata
<input checked="" type="checkbox"/> Document Sentiment	<input type="checkbox"/> Relations
Sentiment Targets <input type="text"/>	<input type="checkbox"/> Semantic Roles
<input checked="" type="checkbox"/> Entities	<input checked="" type="checkbox"/> Syntax
<input checked="" type="checkbox"/> Entity Emotion	<input checked="" type="checkbox"/> Syntax Sentences
<input checked="" type="checkbox"/> Entity Sentiment	<input checked="" type="checkbox"/> Lemma Tokens
Maximum Entities 5	<input checked="" type="checkbox"/> Part of Speech
Limit Text Characters 0	



Name ProcOUT


Setup On Start On Message On Stop


```
1 //Pega todos os conceitos
2 let imax = msg.features.concepts.length;
3 let conceitos = '';
4 for (let i = 0; i < imax; i++){
5     conceitos = msg.features.concepts[i].text+', '+conceitos;
6 }
7
8 //Pega o sentimento
9 let sentimento = msg.features.sentiment.document.label;
10
11 //Pega as categorias
12 imax = msg.features.categories.length;
13 let categorias = '';
14 for (let i = 0; i < imax; i++){
15     categorias = msg.features.categories[i].label+', '+categorias;
16 }
17
18 //Compoem a resposta
19 let r1='Segundo '+msg.autor.nome+' há informações sobre '+conceitos;
20 let r2=' que se inserem nas seguintes categorias '+categorias+'.\n';
21 let r3 = 'O assunto é predominantemente '+sentimento+'.\n';
22 let r4 = 'O assunto é: '+msg.payload;
23
24 //Prepara o pacote para o Telegram
25 msg.payload = {
26     chatId : msg.params.id,
27     type: 'message',
28     content: r1+r2+r3+r4
29 }
30 return msg;
```



# Telegram + Translator + Reddit + NLU


- Para testar, envie uma mensagem no Telegram escrito notícias, nba ou futebol;


 **Henrique**  
notícias

 **Vendas**  
Segundo t2\_3brt0yhi há informações sobre que se inserem nas seguintes categorias /society/crime/personal offense, /law, govt and politics/law enforcement, /law, govt and politics, .  
O assunto é predominantemente neutral.  
O assunto é: Policiais não deveriam ter disparado na casa de Breonna Taylor, diz relatório

Segundo t2\_8lpchk0e há informações sobre Ramayana, que se inserem nas seguintes categorias /religion and spirituality/buddhism, /law, govt and politics, /style and fashion/body art, .  
O assunto é predominantemente neutral.  
O assunto é: Poeta de Mianmar morre após detenção, corpo devolvido com órgãos removidos

Segundo t2\_dyrrz há informações sobre Heavy metal music, que se inserem nas seguintes categorias /sports/go kart, /family and parenting/parenting teens, /society/dating, .  
O assunto é predominantemente negative.  
O assunto é: Mulher sufocada por ex-namorado o atira até a morte em N. Harris Co.

 **Henrique**  
nba

 **Vendas**  
Segundo t2\_9xyw8uqe há informações sobre New York, New York Knicks, National Basketball Association, New York City, Los Angeles Lakers, que se inserem nas seguintes categorias /sports/basketball, .  
O assunto é predominantemente positive.  
O assunto é: The New York Knicks agora tem um recorde melhor do que o campeão de defesa Los Angeles Lakers

Segundo t2\_6vjwa há informações sobre 1942, 1916, 1966, May 8, 2008, que se inserem nas seguintes categorias /technology and computing/hardware/computer components/chips and processors, /technology and computing/software/graphics software, /technology and computing/operating systems/windows, .  
O assunto é predominantemente neutral.  
O assunto é: [SÉRIO NEXT DAY THREAD] Post-Game Discussion (Maio de 08, 2021)

Segundo t2\_6vjwa há informações sobre Mathematics, Applied mathematics, que se inserem nas seguintes categorias /technology and computing/operating systems/windows, /technology and computing/software/graphics software, /technology and computing/hardware/computer components/chips and processors, .  
O assunto é predominantemente neutral.  
O assunto é: Sunday Stats Thread + Game Thread Index



# Telegram + Translator + Reddit + NLU

- **Importante:** o reddit é um fórum de internet. Ele pode ser útil mas as informações nele devem ser tomadas com bastante cautela e criticismo; **Cuidado com Fakenews!**
- O nós do Reddit também podem ser usados para criar um bot dentro do fórum, que pode postar e comentar em postagens;
- O NLU é uma ferramenta bastante poderosa, mas sua utilidade só fica clara quando se necessita extrair conhecimento das bases de dados consultadas, o que pode envolver projetos mais sofisticados;

**Agora é com você!**

# Exercícios

1. Explore os campos de resposta do Reddit e do NLU para extrair conhecimento; Tente compor mensagem diferente ou fluxos que usem informação do NLU para filtrar o tipo de resposta que é passado para o usuário do chat. Muito pode ser feito!

# Estudo Complementar

Dicas, links e livros

# Estudo Complementar

1. Documentação NLU: <https://cloud.ibm.com/docs/natural-language-understanding?topic=natural-language-understanding-getting-started>
2. NLU API: <https://cloud.ibm.com/apidocs/natural-language-understanding>
3. Node-RED + Reddit: <https://flows.nodered.org/node/node-red-contrib-node-reddit>
4. Node-RED + Reddit 2: <https://discourse.nodered.org/t/beginner-friendly-access-to-the-reddit-api-using-node-red/27127>
5. Node-RED + Reddit 3: [https://www.reddit.com/r/redditdev/comments/azjhxm/beginnerfriendly\\_access\\_to\\_the\\_reddit\\_api\\_using/](https://www.reddit.com/r/redditdev/comments/azjhxm/beginnerfriendly_access_to_the_reddit_api_using/)
6. Node-RED + Twitter (dev): <https://flows.nodered.org/node/node-red-contrib-twitter>
7. Node-RED + Twitter (pessoal): <https://flows.nodered.org/node/node-red-node-twitter>

# Próximos Passos

O que veremos na próxima aula

# Na próxima aula...

- Salvando informações;

# OBRIGADO

FIAP

Copyright © 2022 | Slides criados por Daniel Gomes, Henrique Ferreira e Miguel Bozer, adaptados do material do Prof. Andrey Masiero e Marcelo Grave – FIAP

Todos os direitos reservados. A reprodução ou divulgação total ou parcial deste documento é expressamente proibida sem o consentimento formal, por escrito, do professor/autor.