



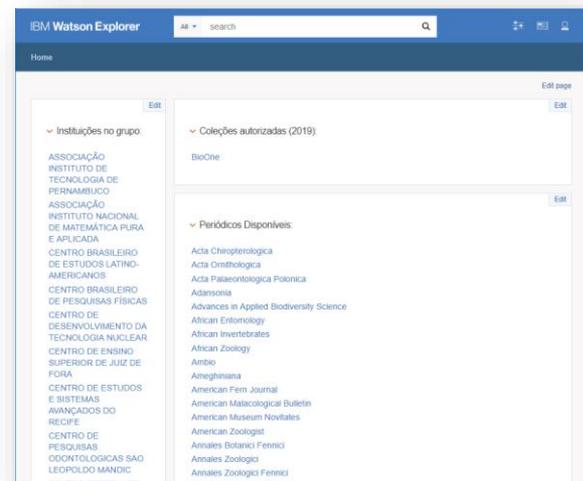
# Machine Learning para Construção de uma Base de Conhecimento: Usando IBM Watson no Portal de Periódicos

Cássio Couto

Analista de Dados – Portal de Periódicos – CAPES/RNP



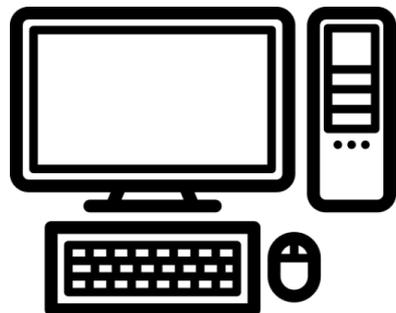
## Watson Explorer



Machine Learning? Computação Cognitiva?

**Como máquinas são capazes de aprender e pensar?**

## O Teste de Turing



---

Interface



## O Teste de Turing



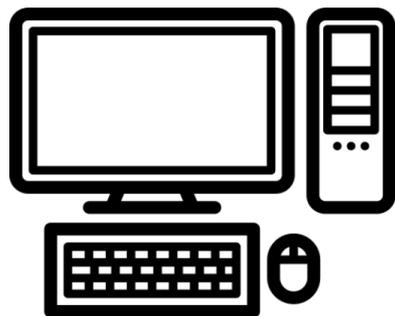
The screenshot shows a BBC News article from June 9, 2014. The article is in the 'Technology' section and is titled 'Computer AI passes Turing test in 'world first''. The author is Eugene Goostman, described as 'THE WEIRDEST CREATURE IN THE WORLD'. The article features a 3D-rendered character of a young boy with glasses. A text box on the page reads: 'A computer program called Eugene Goostman, which simulates a 13-year-old Ukrainian boy, is said to have passed the Turing test at an event organised by the University of Reading.' The byline is 'VLADIMIR VESELOV' and the caption at the bottom is 'Eugene Goostman simulates a 13-year-old Ukrainian boy'.

Será que Machine Learning vai tomar o lugar do ser humano no futuro?

A computer program called Eugene Goostman, which simulates a 13-year-old Ukrainian boy, is said to have passed the Turing test at an event organised by the University of Reading.

## A Sala Chinesa (The Chinese Room) - John Searle

Sala A



Sala B



Código do algoritmo

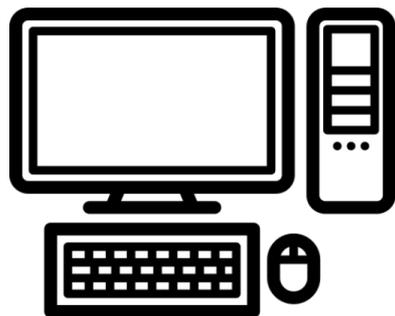


Eu não apenas traduzo, mas também compreendo chinês! Além disso, eu passo no teste de Turing!

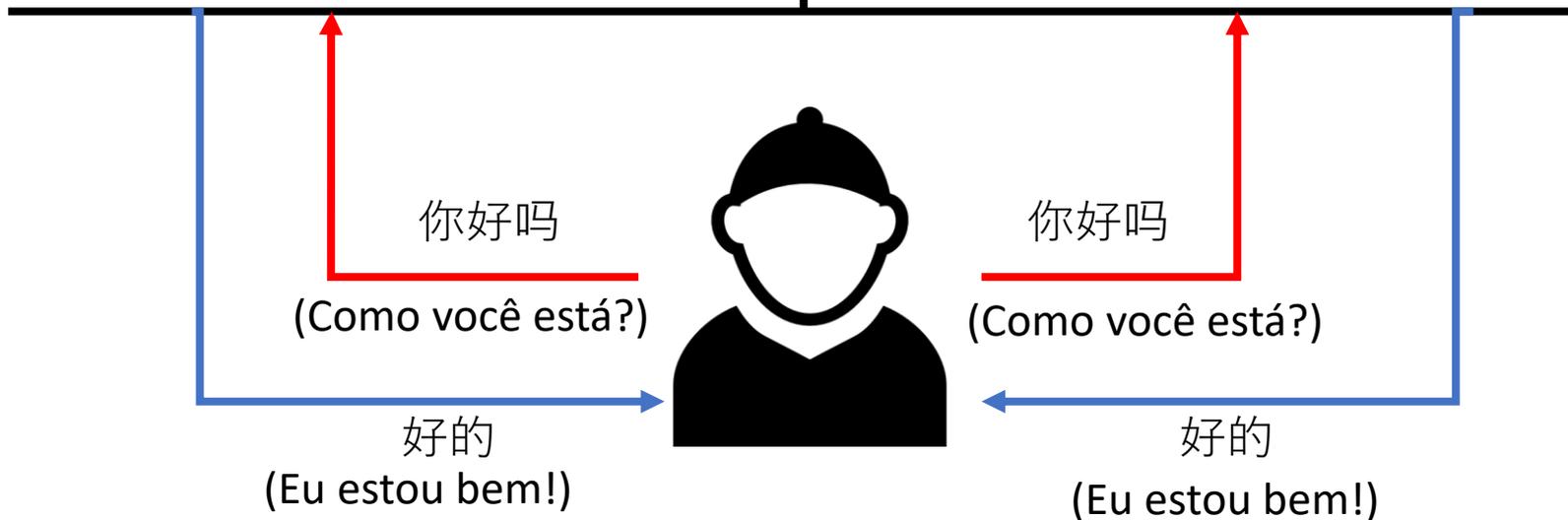
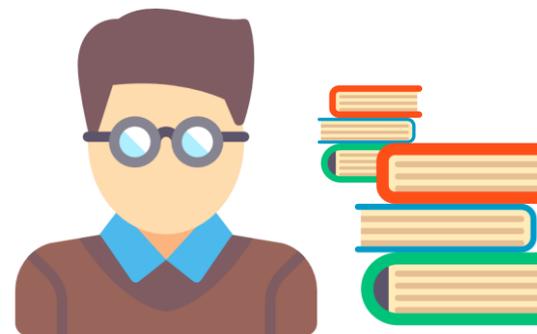
Eu não sei uma palavra de chinês. Apenas estou executando o algoritmo do programa da Sala A sem questionar nada.

## A Sala Chinesa (The Chinese Room) – John Searle

Sala A



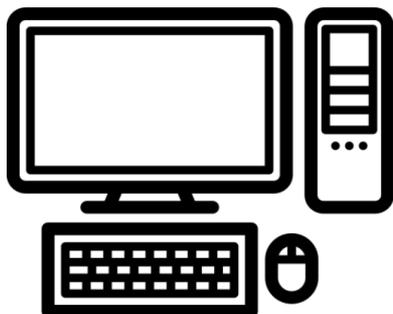
Sala B



## A Sala Chinesa (The Chinese Room) -



Eu continuo sem saber chinês, mas consigo imitar uma pessoa respondendo em chinês

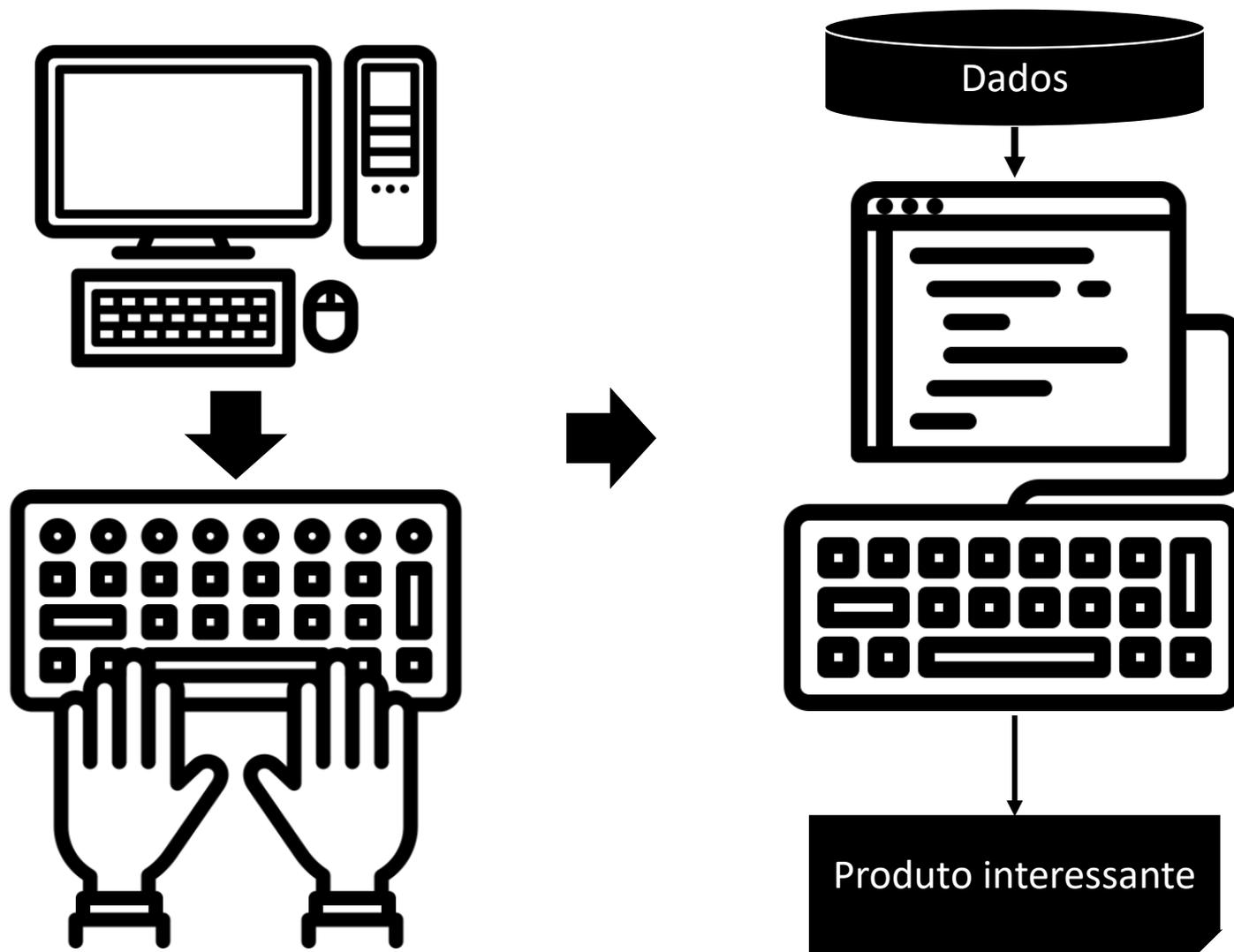


Por consequência, o programa não **compreende** chinês, mas sim **simula** o processo de compreensão e fala.

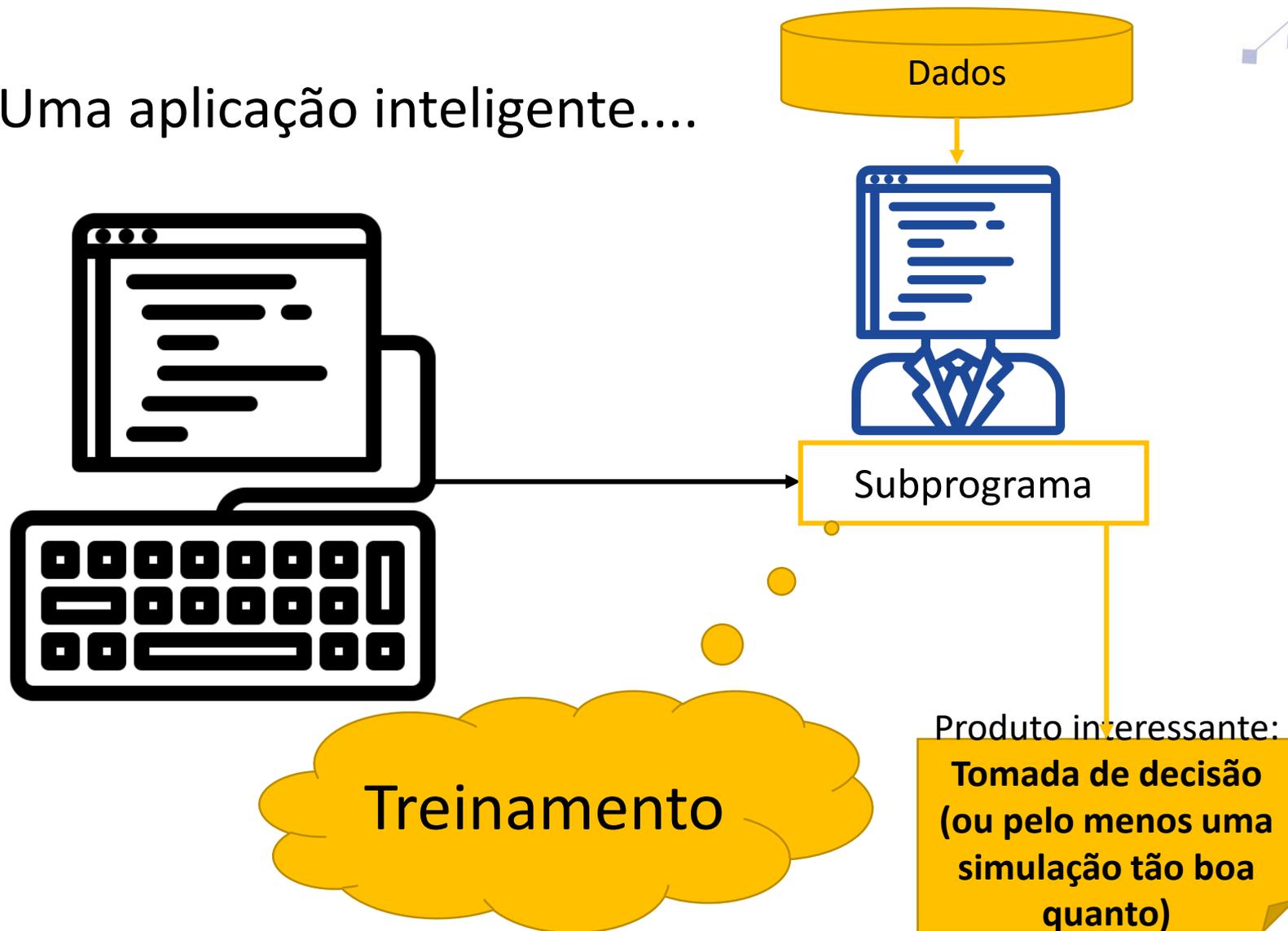
## Por que isso é importante para compreender Machine Learning?

Enquanto aplicações cognitivas são excelentes em simular ações humanas em tempo extraordinário e livres de certas condições

Elas são incapazes de compreensão e intenção



Uma aplicação inteligente....



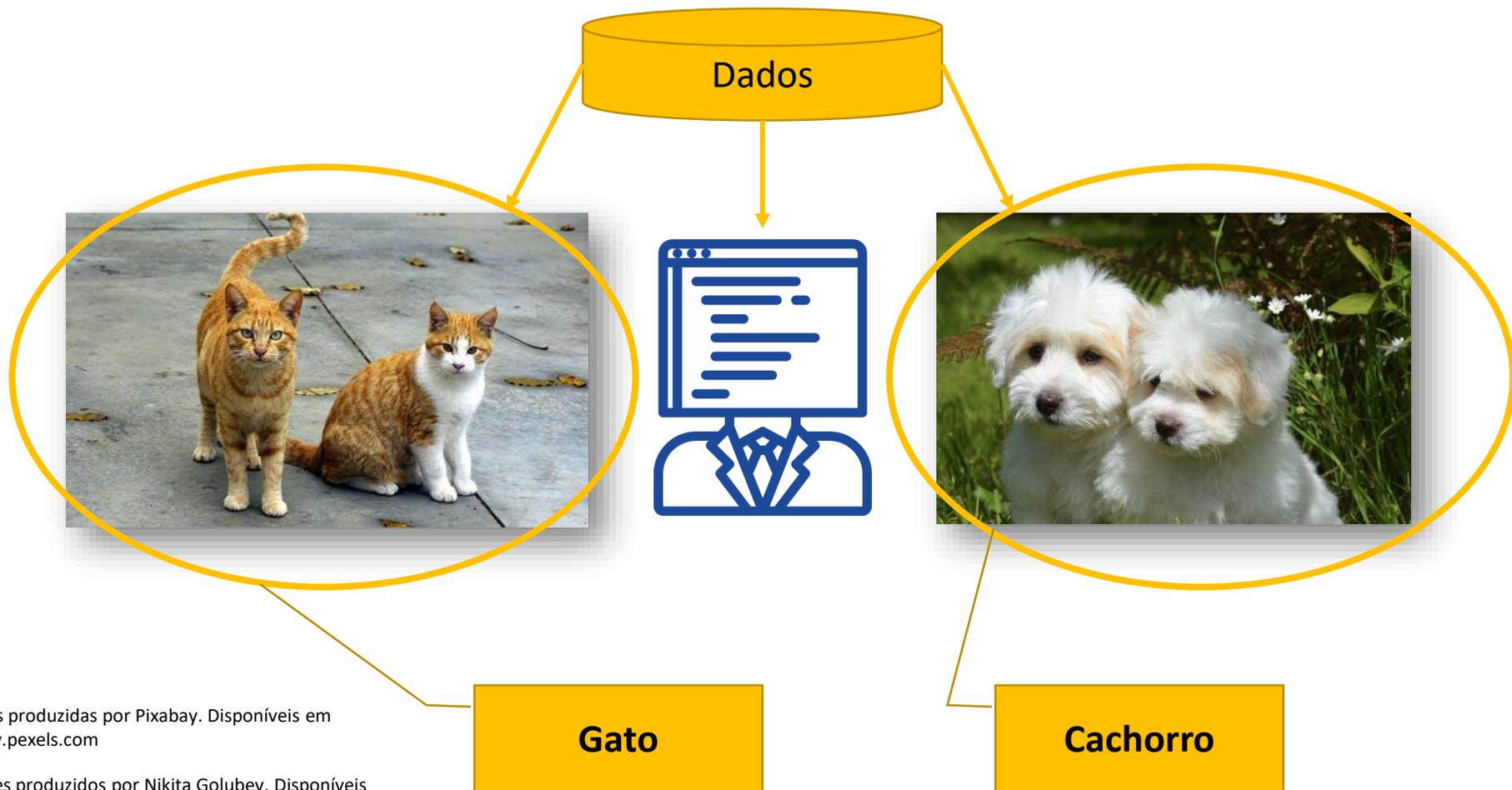
Que treinamento é esse?

Aprendizado Supervisionado

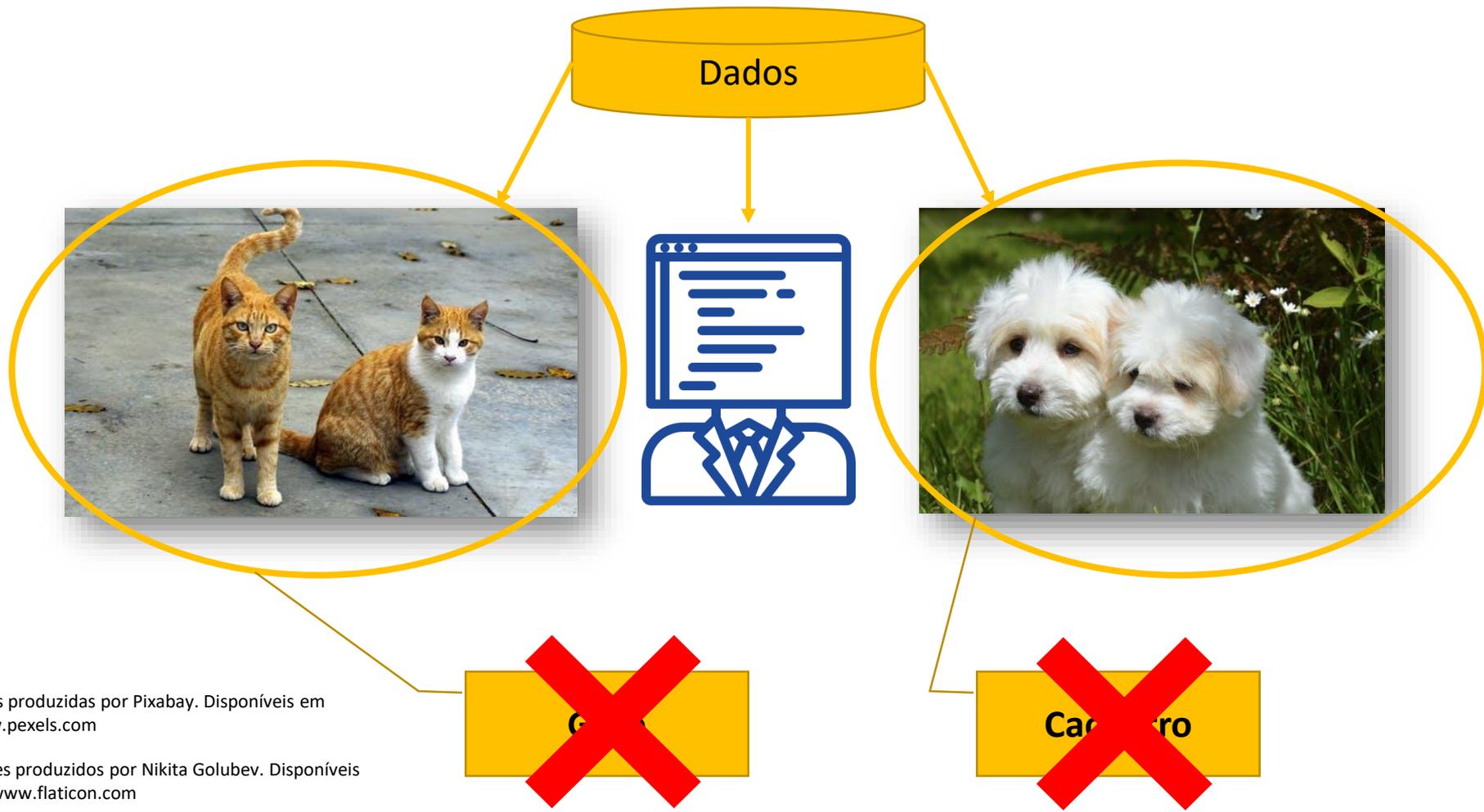


Aprendizado Não-Supervisionado

## Aprendizado Supervisionado



## Aprendizado Não-Supervisionado



## Aprendizado Supervisionado

Treinamento



**Gato**

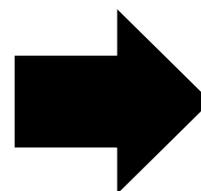
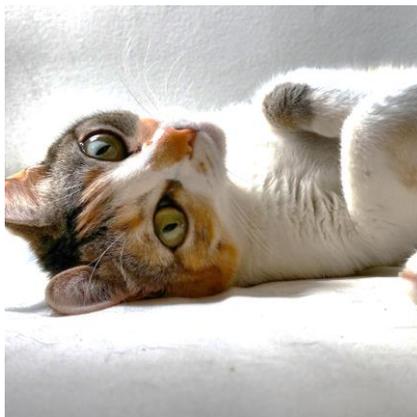


**Cachorro**

É preciso expor o programa a uma quantidade significativa de amostras!

## Aprendizado Supervisionado

Pré-  
processamento



- Equalizar pixels
- Remover fundo
- Procurar posições dos pixels dos olhos, do focinho e das orelhas



E no caso de um texto em linguagem natural?

The argument and thought-experiment now generally known as the Chinese Room Argument was first published in a paper in 1980 by American philosopher John Searle (1932- ). It has become one of the best-known arguments in recent philosophy.

## Processamento de Linguagem Natural

Raízes etimológicas

The argument and thought-experiment now generally known as the Chinese Room Argument was first published in a paper in 1980 by American philosopher John Searle (1932- ). It has become one of the best-known arguments in recent philosophy.

Segmentação em orações e palavras

Verificação de sinônimos

Extração de Entidades Nomeadas

John Searle;  
América; China...

Classificação Automática

Texto sobre Filosofia (60% certeza)

Análise de sentimentos

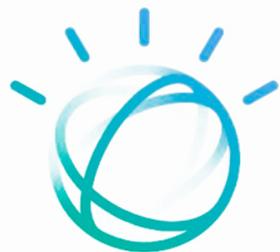
Discurso com tom positivo ("one of the best-known")

Deteccção de Idioma/  
Tradução Automática

"O argumento e experimento mental..."

Ok, sabemos que o aprendizado são algoritmos essencialmente

**E o tal do Watson? O que ele faz?  
Como ele funciona?**



## Watson Explorer



### Apresentador:

- Há na América Latina uma equipe de futebol que foi a única a ser pentacampeã de um campeonato muito famoso.

### Concorrentes:

De que país é a **seleção canarinho**?

Quantas copas do mundo a **equipe de futebol brasileira** participou?

Quem é o **time brasileiro de futebol**?



Watson Explorer



- Há na América Latina uma equipe de futebol que foi a única a ser pentacampeã de um campeonato muito famoso.

Entidades de interesse





Watson Explorer



Entidades de  
interesse

**Equipes de futebol na América Latina:**

Flamengo, Fluminense, Juventus...

Equipe de Futebol Brasileiro, Equipe de Futebol Paraguaia...  
Remo, Bahia, Fortaleza...

**Campeonatos famosos:**

Olímpiadas, Copa do Mundo de Futebol...

Miss Universo, Miss Brasil...

Olímpiadas de Inverno, Roland-Garros, etc

**Pentacampeões mundiais:**

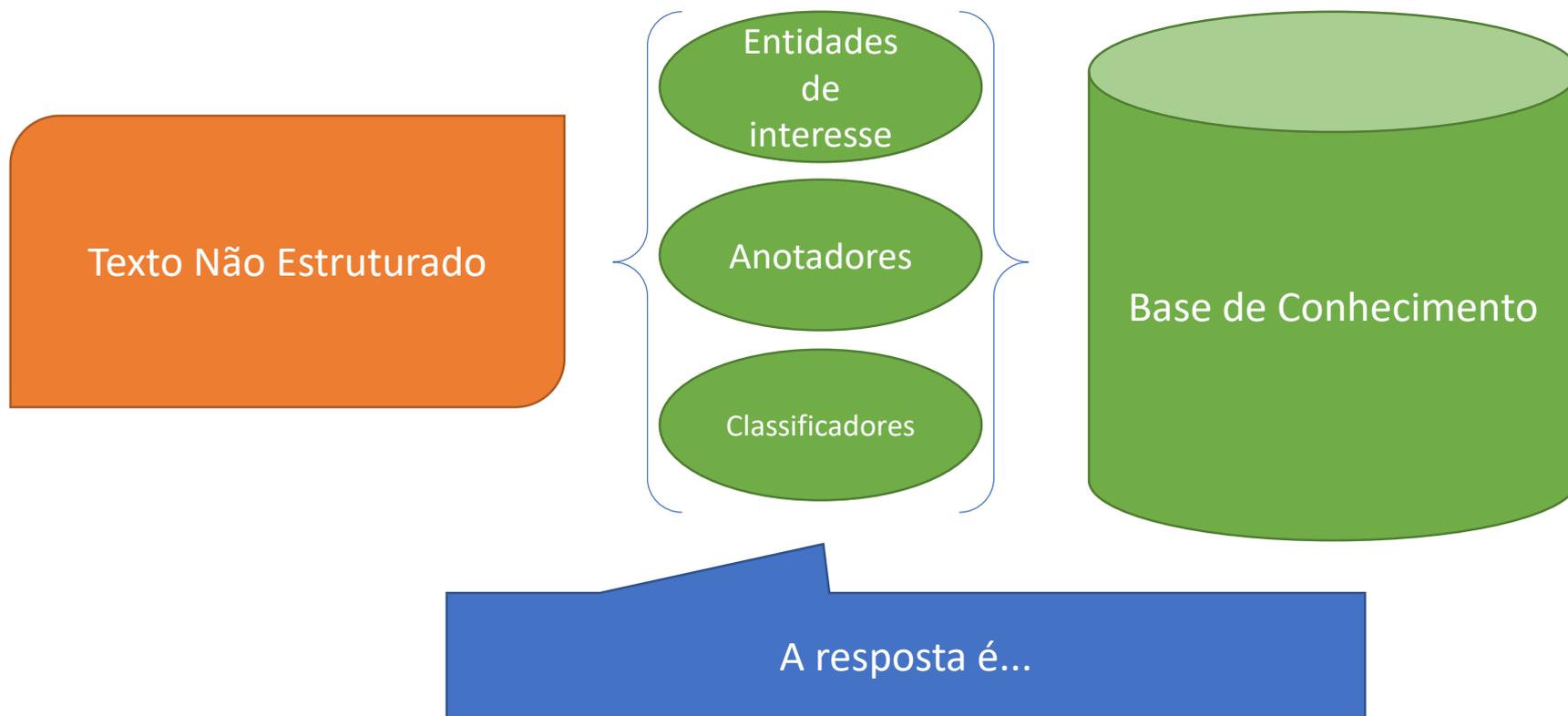
Equipe de Futebol Brasileiro (Copa)  
Maya Gabeira (surfista)  
José Marquez (ciclista português)  
Santos (torneio Rio-São Paulo)



Ele deve estar  
falando disso!



## Watson Explorer



# Pipeline UIMA

Tudo isso pertence a Unstructured Information Management Application

- + Complexo
- Indicado pra grandes volumes de dados
- Várias atividades intermediárias



A decorative graphic element in the bottom-left corner of the slide. It consists of a grid of thin, light-colored lines that form a series of overlapping squares. The grid is oriented diagonally, with the lines running from the bottom-left towards the top-right. The squares in the grid are filled with a light, semi-transparent color, creating a subtle pattern against the yellow background.

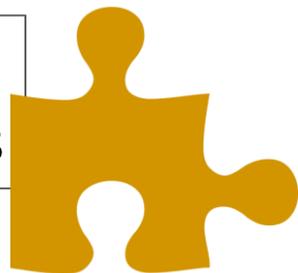
**Como isso está sendo aplicado no Portal de Periódicos?**

## O conteúdo do Portal de Periódicos

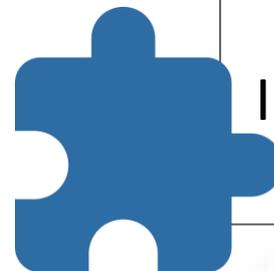
48.966 periódicos  
50+ milhões de artigos  
330.373 livros  
525 bases  
1+ milhão de teses (2018)

Grande quantidade de  
conteúdo textual **não-  
estruturado**

Sistema de Autorizações



Sistema de Indexação e Busca de Conteúdo



Watson Explorer

Teses e Dissertações



Sistema de Gestão Integrado



Plataforma Lattes



## Dissertação de mestrado



**Célia Ghedini Ralha**  
 Bolsista de Produtividade em Pesquisa do CNPq - Nível 2  
 Endereço para acessar este CV: <http://lattes.cnpq.br/5632722847264046>  
 Última atualização do currículo em 18/02/2019

Célia Ghedini Ralha é PhD pela Universidade de Leeds, Inglaterra (1996) e Departamento de Ciência da Computação da Universidade de Brasília (UnB), onde chefe do departamento durante suas gestões subsequentes (2006-2010). Colaborou

Currículo

[15] Célia G Ralha, Carolina G Abreu, Cássio GC Coelho, Alexandre Zaghetto, Bruno Macchiavello, and Ricardo B Machado. A multi-agent model system for land-use change simulation. *Environmental Modelling & Software*, 42:30–46, 2013.

[16] Cássio G. C. Coelho, Carolina G. Abreu, Rafael M. Ramos, Aldo H. D. Mendes, George Teodoro, and Célia G. Ralha. Mase-bdi: agent-based simulator for environmental land change with efficient and parallel auto-tuning. *Applied Intelligence*, 45(3):904–922, Oct 2016.

[17] Barbara A Kitchenham, David Budgen, and O Pearl Brereton. The value of mapping studies—a participant-observer case study. In *EASE*, volume 10, pages 25–33, 2010.

[18] Paolo Bresciani, Anna Perini, D. L. G. ... John Mylopoulos. Tropos: methodology. *Autonomous A*, 236, 2004.

[19] Jaelson Castro, Manuel K requirements-driven informat project. *Information systems*, 27

[20] S. J. Russell and P. Norvig. *A* Prentice-Hall, USA, 3<sup>rd</sup> edição

[21] Michael Wooldridge and Nich ory and practice. *Knowledgea*

Dados linkados



**Environmental Modelling & Software**  
 ELSEVIER  
 Volume 42, April 2013, Pages 30-46

## A multi-agent model system for land-use change simulation

Célia G. Ralha <sup>a</sup>, Carolina G. Abreu <sup>a, b</sup>, Cássio G.C. Coelho <sup>a</sup>, Alexandre Zaghetto <sup>a</sup>, Bruno Macchiavello <sup>a</sup>, Ricardo B. Machado <sup>c</sup>

Artigo em periódico assinado no Portal

## Índice do Portal



Environmental Modelling & Software  
 ELSEVIER Volume 42, April 2013, Pages 30-46

A multi-agent model system for land-use change simulation

Célia G. Ralha <sup>a</sup> ✉, Carolina G. Abreu <sup>a, b</sup>, Cássio G.C. Coelho <sup>a</sup>, Alexandre Zaghetto <sup>a</sup>, Bruno Macchiavello <sup>a</sup>, Ricardo B. Machado <sup>c</sup>

## Classificadores

Célia Ghedini Ralha possui

- Código Docente/Programa/IES
- Código Currículo Lattes
- Códigos RG e CPF



Célia Ghedini Ralha deve ser **Autora Nacional**



## Dicionários

Célia Ghedini Ralha é equivalente a

RALHA, Célia Ghedini; Ralha, Célia Ghedini; Ghedini Ralha, Célia; Ralha, Célia; RALHA, C. G.; RALHA, CÉLIA G.; Célia G. Ralha; GHEDINI, Celia; C. GHEDINI; Celia Ralha; Ralha, Celia; Celia G. Ralha; Ralha, Celia G.; G. RALHA, CELIA; Ralha, Celia Ghedini



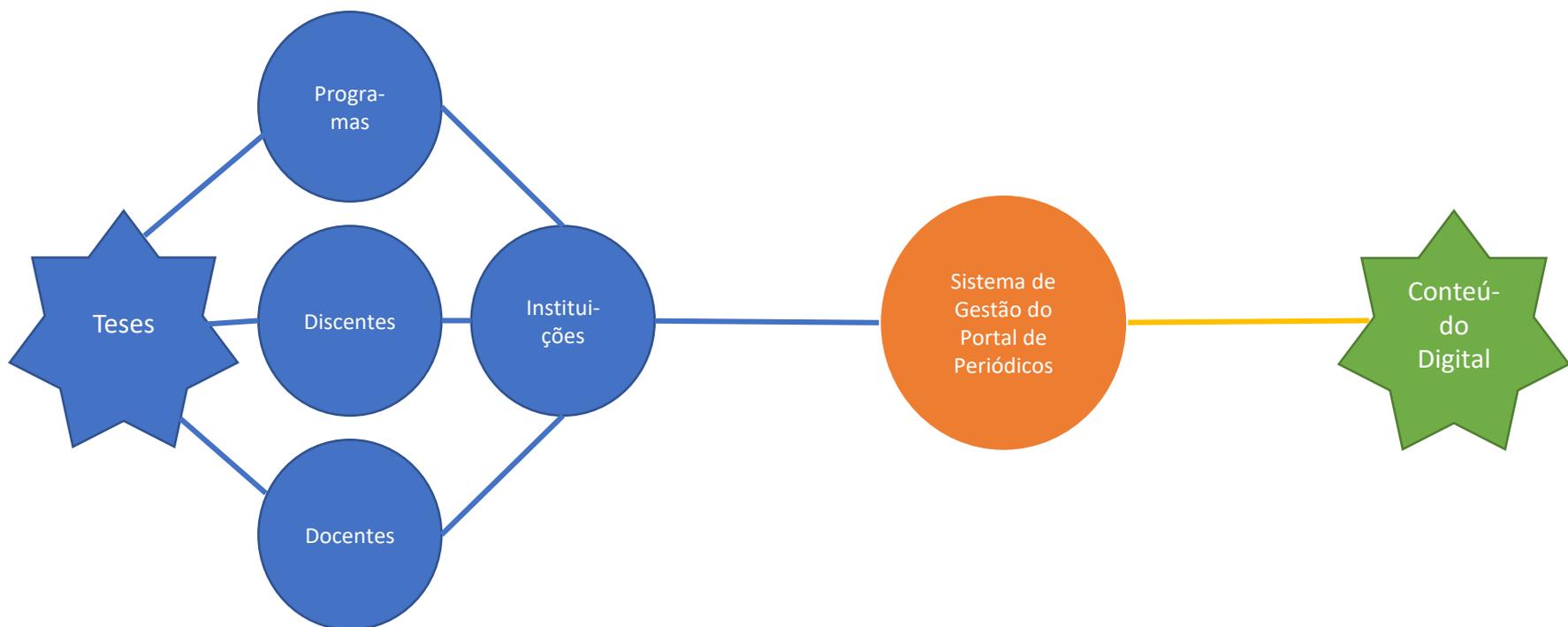
## Anotadores

Célia Ghedini Ralha é

- Docente do programa INFORMÁTICA/UNB
- Docente do programa ENGENHARIA ELÉTRICA/UNB
- Docente do programa ENGENHARIA DE SISTEMAS ELETRÔNICOS E DE AUTOMAÇÃO/UNB

Plataforma Lattes

Sistema de Gestão PPG



## Legenda



Dado não-estruturado



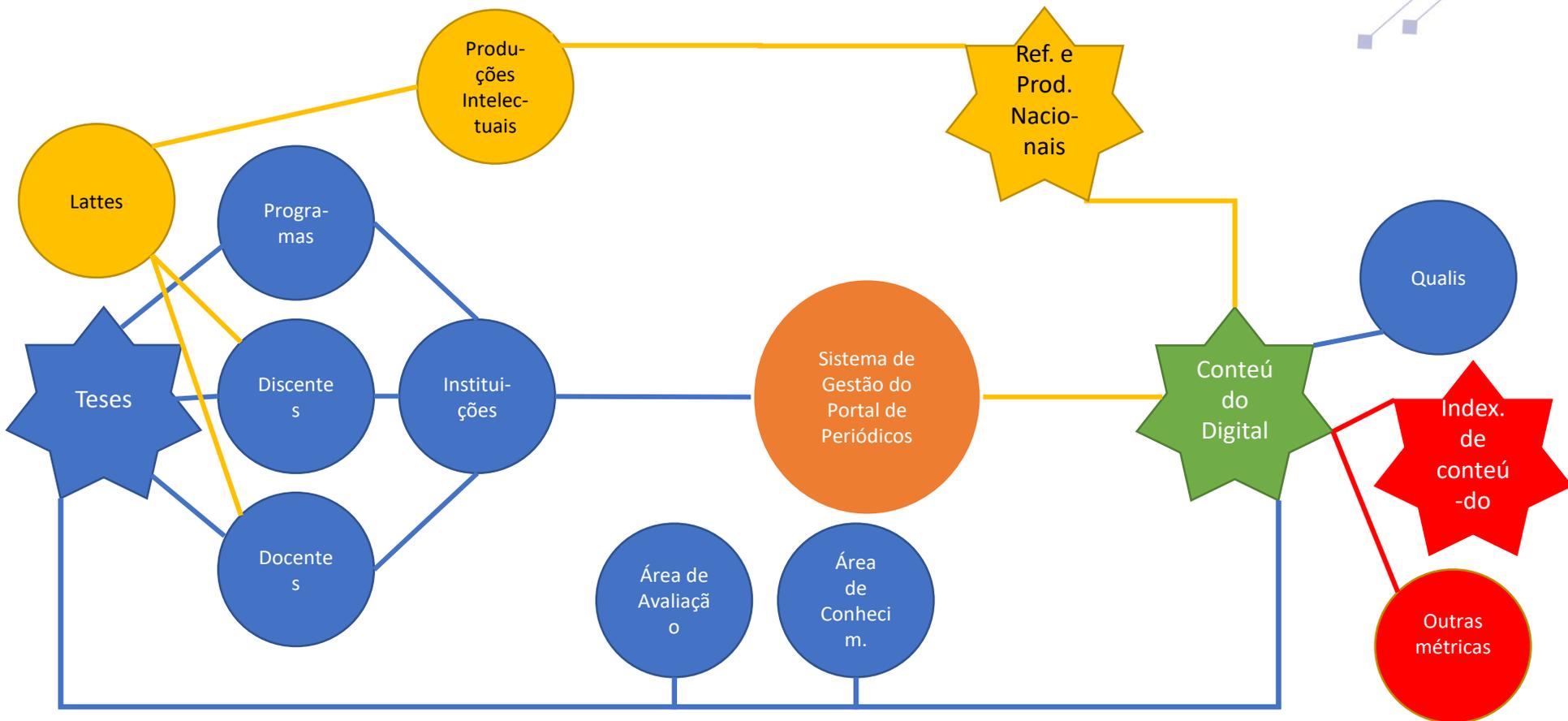
Dado estruturado



Junção difusa



Sucupira  
Portal de Periódicos  
Sistema de Indexação



## Legenda



Dado não-estruturado



Dado estruturado



Junção difusa



Sucupira



Portal de Periódicos



Sistema de Indexação



Sistemas Externos



Lattes

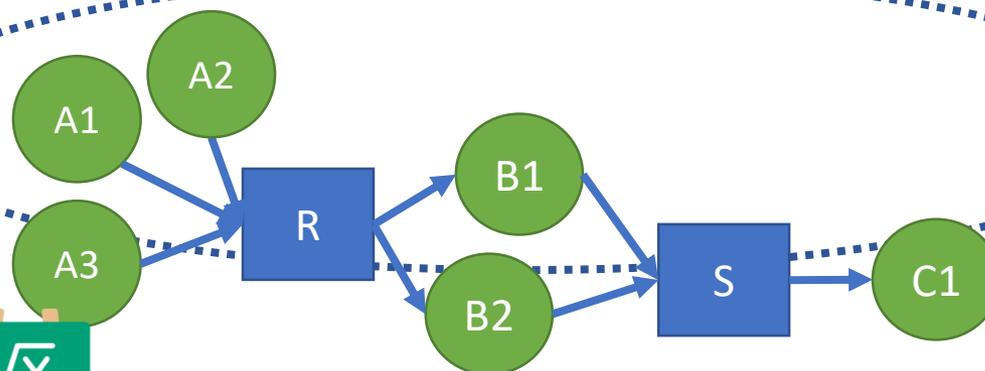
## Utilização do Watson

Base de Conhecimento



Descobrir e melhorar:

- Aproveitar o conteúdo
- Gerenciar os contratos
- Aperfeiçoar a eficiência do Portal



Classificadores

Classe X

Classe Y

Classe Z



Cássio Couto

[cassio.coelho@rnp.br](mailto:cassio.coelho@rnp.br)



MINISTÉRIO DA  
DEFESA

MINISTÉRIO DA  
CIDADANIA

MINISTÉRIO DA  
SAÚDE

MINISTÉRIO DA  
EDUCAÇÃO

MINISTÉRIO DA  
CIÊNCIA, TECNOLOGIA,  
INOVAÇÕES E COMUNICAÇÕES

