



# Dendro Keywords: Apoio à descrição de dados através da extração e visualização de palavras-chave

Cláudio Monteiro, Carla Teixeira Lopes, João Rocha da Silva  
INESC TEC/Faculdade de Engenharia da Universidade do Porto

# Dendro

Ambiente colaborativo onde o investigador pode armazenar e descrever os seus dados.

Foca-se nos estados iniciais da investigação devido à sua importância.

Permite aos investigadores escolherem os seus próprios conjuntos de descritores a partir de ontologias existentes.

# Dendro

## Folder



## Selection



- Up to biodata
- meta
- scala\_2014.jpg
- observations.xls
- ontTraffic.pdf

## Information

## Change log

Description progress

0%

✓ SAVE

↶ UNDO

📄 COPY FROM PARENT

🔒 IN MANUAL MODE

🗑️ CLEAR

## Description

B

I

H



🔍 PREVIEW

Studies of the endangered species of felid living mainly in the Iberian Peninsula in southwestern Europe. They are categorized as endangered by many institutions, including the International Union for the Conservation of Nature

## Coverage



## Descriptor Selection

ALL

AUTO

## Find descriptors

Type to find descriptors...

🔍 DEPICTION



A depiction of some thing. 🧑🧑

🔍 DESCRIPTION



An account of the resource. 🧑🧑

🔍 DATE



A point or period of time associated with an event in the lifecycle of the resource. 🧑🧑

🔍 COVERAGE



The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant. 🧑🧑

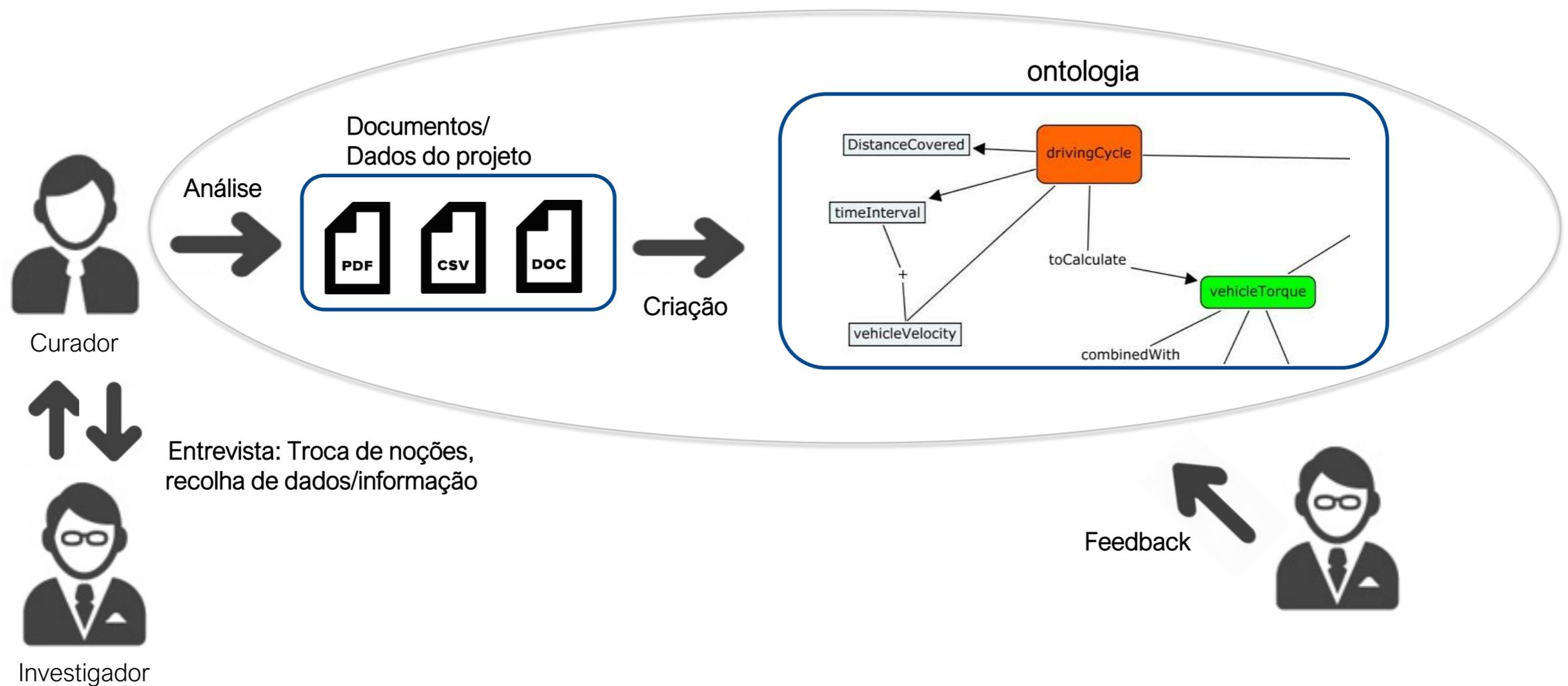
🔍 SUBJECT



# Ontologias existentes no Dendro

Domínio	Tipo de dados	
Fratura de Materiais	Experimental	
Produção de Hidrogénio		
Química Analítica	Experimental	Observacional
Gravimetria		
Oceanografia		
Biodiversidade	Observacional	
Ciências Sociais	Observacional	
Corte e Empacotamento	Simulação	
Simulação de Veículos		
Dinâmica de Fluídos		

# Contexto atual



# Motivação

O processo de criação de ontologias é dispendioso, tanto em termos de tempo, como financeiramente.

Muitas organizações não possuem capacidade para criar ontologias.

Existe um grande interesse em automatizar o processo.

# Ontology Learning

Processo de identificação de elementos, tais como, termos e conceitos de forma a criar ou manter uma ontologia.

Aplicação de métodos e técnicas baseadas em Information Retrieval, Machine Learning e Data Mining.

Processo semiautomático devido à necessidade de intervenção humana.

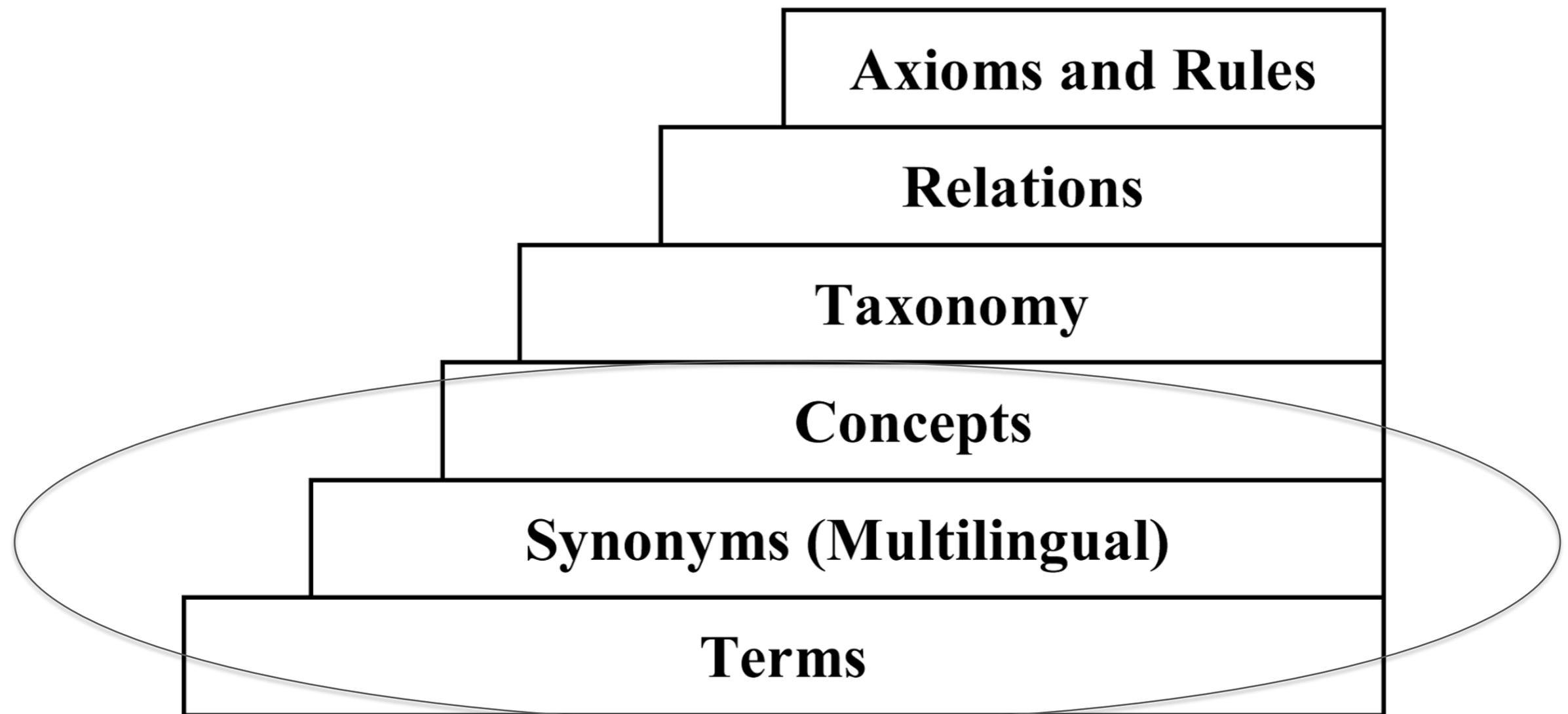
# Ontology Learning

Permite poupar tempo.

Não são necessários conhecimentos específicos do domínio.

Permite encontrar conceitos que possam ter escapado no processo manual.

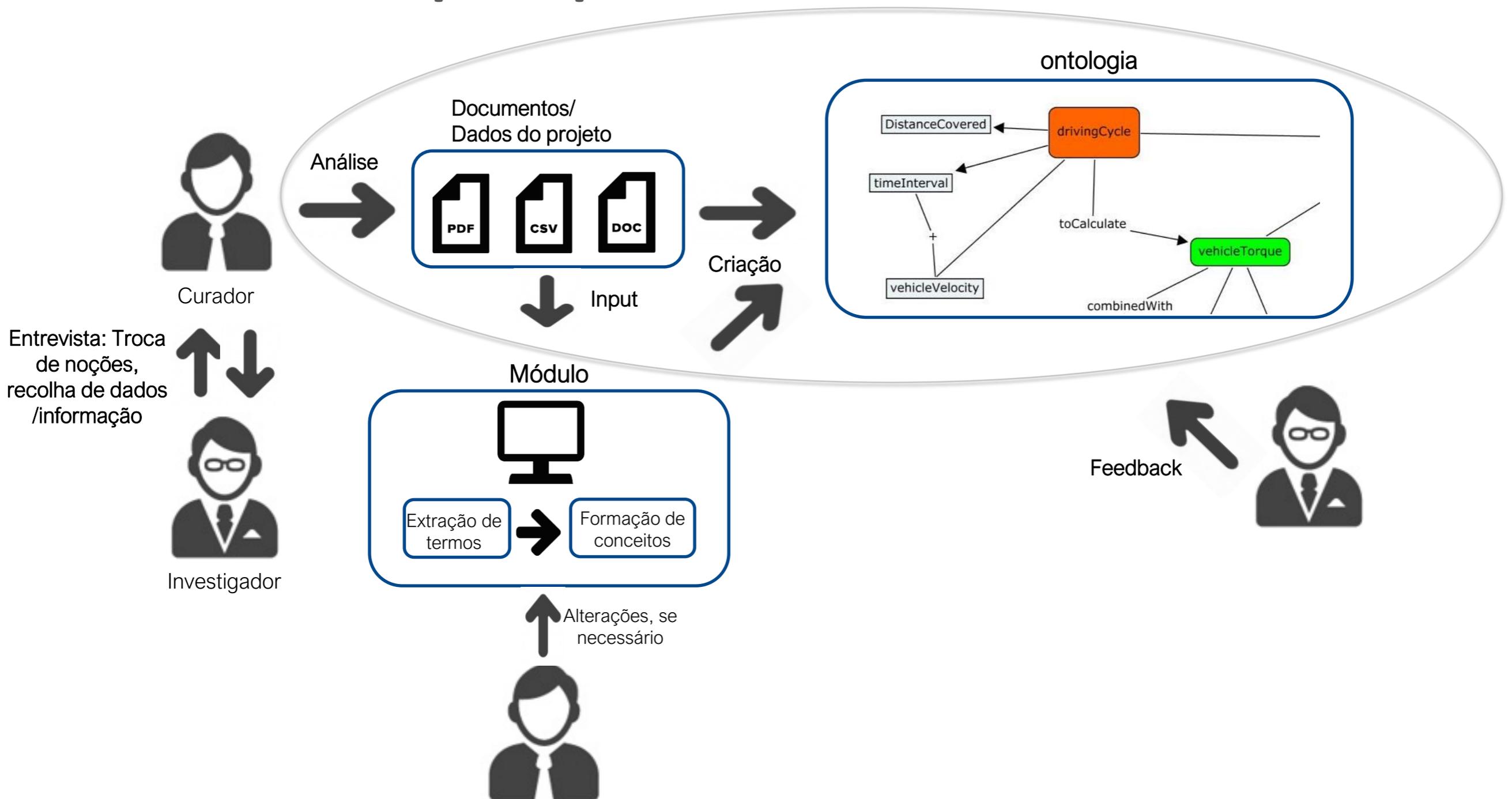
# Ontology Learning Layer Cake



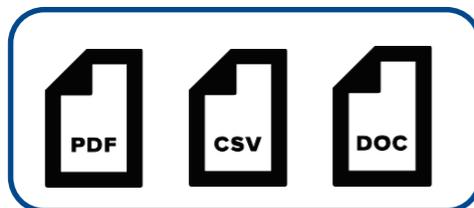
# Ontology Learning

Sistema	Output		Technique	
	Termos	Conceitos	Statistic-based	Linguistic-based
ASIUM(2000)	X			Sentence parsing, Syntactic structure analysis, Subcategorization frames
		X		
Text-to-Onto(2000)	X			Part-of-speech tagging, Sentence parsing, Syntactic structure analysis
		X	Co-occurrence analysis	Concepts from domain lexicon
TextStorm/Clouds(2001)	X			Part-of-speech tagging using WordNet, Syntactic structure analysis, Anaphora resolution
SYNDIKATE(2001)	X			Syntactic structure analysis, Anaphora resolution
		X		Use of semantic templates and domain knowledge
OntoLearn(2002)	X		Relevance analysis	Part-of-speech tagging, Sentence parsing
		X		Concepts and glossary from WordNet
CRTCOL(2005)	X	X	Relevance analysis	Part-of-speech tagging, Sentence parsing, Use of domain lexicon, Word sense disambiguation
OntoGain(2010)	X	X		Part-of-speech-tagging, Shallow parsing, Relevance analysis

# Método proposto



# Método proposto



Módulo

Pré-processamento de texto

1. Stop word removal
2. Tokenization
3. Lemmatization
4. Part-of-Speech Tagging



Extração de termos

- Single term extraction
- Multi-word term extraction



Criação de conceitos

- Clustering



Extração de sinônimos

- WordNet
- Word Similarity Measure



# Artigos relevantes

Antonio Lieto. “Manually vs semiautomatic domain specific ontology building”

Wilson Wong. “Ontology Learning from Text: A Look Back and into the Future”

Roberto Navigli, Paola Velardi, Aldo Gangemi. “Ontology Learning and Its application to Automated Terminology Translation”

Alexander Maedche, Steffen Staab. “Ontology Learning”

Philipp Cimiano, Johanna Volker, Rudi Studer. “A Description of the State-of-the-Art, Applications, Challenges and Trends for Ontology Learning from Text”

Chris Biemann. “Ontology Learning from Text: A Survey of Methods”



Website:

<http://dendro.fe.up.pt/>

Repositório:

<https://github.com/feup-infolab-rdm>

Demonstração:

<http://dendro.fe.up.pt/demo>