

# INTELIGÊNCIA ARTIFICIAL E FERRAMENTAS DA WEB SEMÂNTICA APLICADAS A RECUPERAÇÃO DA INFORMAÇÃO: UM MODELO CONCEITUAL COM FOCO NA LINGUAGEM NATURAL

## ARTIFICIAL INTELLIGENCE AND SEMANTIC WEB TOOLS APPLIED TO INFORMATION RETRIEVAL: A CONCEPTUAL MODEL WITH A FOCUS ON NATURAL LANGUAGE

Caio Saraiva Coneglian<sup>a</sup>

José Eduardo Santarem Segundo<sup>b</sup>

### RESUMO

**Objetivo:** o artigo apresenta como objetivo a proposição de um modelo de recuperação da informação que redesenha este campo de estudos, a partir da aproximação da linguagem computacional com a linguagem natural, utilizando os princípios da representação da informação, para que o significado e o contexto dos dados estejam explícitos para o processo da busca; para tanto, aproxima-se e relaciona-se aos processos de Inteligência Artificial, processamento de linguagem natural e às ferramentas da Web Semântica. **Metodologia:** utilizando o método quadripolar, sendo eles: polo epistemológico, polo teórico, polo técnico e polo morfológico. Ademais, a pesquisa foi exploratória, tendo um caráter aplicado. **Resultados:** Como resultados, criou-se este modelo de recuperação da informação, pautado no contexto semântico e na aplicação da Inteligência Artificial, capaz de tornar a linguagem natural a base do processo, e considerando o contexto e o significado dos termos para os usuários. **Conclusões:** aponta-se que o presente trabalho realiza uma importante aproximação entre a Ciência da Informação e a Inteligência Artificial, trazendo para seu escopo, em especial no âmbito da recuperação da informação, aplicações reais de como este segundo campo de estudos pode aprimorar a área como um todo.

**Descritores:** Recuperação da informação. Web Semântica. Inteligência artificial.

---

<sup>a</sup> Doutor em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (UNESP). E-mail: caio.coneglian@gmail.com

<sup>b</sup> Doutor em Ciência da Informação pela Universidade Estadual Paulista (UNESP). Docente do Departamento de Educação, Informação e Comunicação da Universidade de São Paulo (USP). E-mail: santarem@usp.br

Processamento de linguagem natural. Ontologia.

## 1 INTRODUÇÃO

A Ciência da Informação e a Ciência da Computação possuem uma série de relações interdisciplinares, capazes de, quando aplicadas em conjuntos, aprimorar diversos campos de pesquisa, e efetivamente, melhorar a qualidade das Tecnologias da Informação e Comunicação. Adicionalmente, campos como a Recuperação da Informação, que tem como uma das suas principais características a interdisciplinaridade, pode ser um fértil terreno para a aplicação de novas tecnologias, considerando todas as técnicas e conceitos da Ciência da Informação.

Neste contexto, uma das áreas que mais tem se destacado, em especial dentro da Computação, é a Inteligência Artificial (IA). Nos últimos anos, verifica-se uma série de aplicações com grande impacto sendo desenvolvidas por pesquisadores e empresas e utilizadas pelas pessoas em seu cotidiano. A IA tem revolucionado diversas áreas e setores da economia, podendo, quando aplicada junto a outras técnicas e teorias, aprimorar a Ciência da Informação.

Dentro da Inteligência Artificial, um subcampo, chamado de Processamento de Linguagem Natural (PLN), tem se destacado pela capacidade de permitir o entendimento do modo como as pessoas se comunicam, seja de forma oral ou escrita, pelos instrumentos computacionais. Essa área, quando apoiada por técnicas de aprendizagem de máquina, está tornando o relacionamento entre humanos e máquinas muito mais simples e eficiente.

Em um contexto distinto, outro campo de estudo que possui uma série de relações com a Computação é a Web Semântica. Tal campo tem evoluído de forma significativa nas últimas décadas, se posicionando, por meio de suas teorias e ferramentas, como um foco de estudos para aprimorar a compreensão pelas agentes computacionais, da linguagem humana.

Ambos os campos citados, Processamento de Linguagem Natural e Web Semântica, se destacam por abordagens distintas para a resolução de problemas. Todavia, quando analisados sob o prisma da aproximação entre

peças e computadores para o processo de entendimento da linguagem, identifica-se uma clara relação e possibilidades de pesquisas. Em especial, no âmbito da Recuperação da Informação, verifica-se um conjunto de possibilidades de aplicações conjuntas entre IA, PLN e Web Semântica, visando tornar mais eficiente o processo de recuperação.

Nesse sentido, esta pesquisa se vincula às dificuldades durante o processo de recuperação da informação, no que tange, especialmente, à contextualização das informações pelos mecanismos computacionais, bem como, ao desafio existente de realizar o processo de busca e recuperação utilizando linguagem natural.

Desta forma, a pesquisa apresenta como objetivo geral a proposição de um modelo conceitual de recuperação da informação, a partir da aproximação da linguagem computacional com a linguagem natural, utilizando os princípios da representação da informação, para que o significado e o contexto dos dados estejam explícitos para o processo da busca. Para tal, aproximam-se e relacionam-se os processos de Inteligência Artificial, processamento de linguagem natural e as ferramentas da Web Semântica.

## 2 TRABALHOS RELACIONADOS

Há muitas pesquisas que propuseram modelos utilizando elementos de Recuperação da Informação com o apoio de linguagem natural, em especial aplicando em *chatbots*. Contrapondo esse cenário, destaca-se que há uma quantidade bem menor de trabalhos que adotam tal cenário, utilizando ontologias e bases de dados utilizando os padrões do *linked data*.

O Quadro 1 apresenta uma síntese dos trabalhos relacionados encontrados. Para isso os trabalhos foram divididos em 4 tipos, que refletem o aspecto principal de cada publicação. Os tipos são: 1) Question Answering, modelo de recuperação da informação do estilo de pergunta e resposta, e *Linked Data*: esses trabalhos têm um enfoque maior no uso de dados, seguindo formatos de *Linked Data* para apoiar sistemas de *Question Answering*; 2) Ontologia e QA: trata de como as ontologias podem auxiliar no tratamento da

linguagem nos sistemas de *Question Answering*; 3) Assistente Virtual: trabalhos que exploram o conceito e criam assistentes virtuais, e que têm em sua base o processamento de linguagem natural como essência; e 4) *Linked Data*, QA e ontologia: trabalhos que relacionam o *Linked Data* e as ontologias no âmbito do *Question Answering*.

**Quadro 1 – Comparação de trabalhos relacionados**

<b>Tipo</b>	<b>Trabalho</b>	<b>Adoção <i>Linked Data</i></b>	<b>Uso de ferramenta s da Web Semântica</b>	<b>Ontologi as em todos os processo s do Q.A.</b>	<b>PLN para compre en- são da semânti ca</b>
QA e <i>Linked Data</i>	Lopez <i>et al.</i> (2013)	Sim	Sim	Não	Parcial
QA e <i>Linked Data</i>	Walter (2012)	Sim	Sim	Não	Parcial
Ontologia e QA	Asiaee <i>et al.</i> (2015)	Não	Sim	Não	Parcial
Ontologia e QA	Kumar e Zayaraz (2015)	Não	Sim	Não	Parcial
Assistente virtual	Bernardo e Santanché (2017)	Não	Não	Não	Sim
Assistente virtual	Costa, Campelo e Campos (2018)	Não	Não	Não	Sim
Ontologia e QA	Abdi, Idris, Ahmad (2018)	Não	Sim	Não	Parcial
Ontologia e QA	Ferrandéz <i>et al.</i> (2009)	Não	Sim	Não	Parcial
Ontologia e QA	Xie <i>et al.</i> (2015)	Não	Sim	Não	Parcial
<i>Linked Data</i> , QA e ontologia	Bouziane <i>et al.</i> (2018)	Sim	Sim	Não	Parcial
QA e <i>Linked Data</i>	Cabaleiro, Peñas e Manandhar (2017),	Sim	Parcial	Não	Parcial

**Fonte:** elaborado pelos autores

Todos os trabalhos trazem importantes contribuições para a área de QA e PLN, porém nenhuma das pesquisas aborda, sob a ótica da Ciência da Informação, como as ontologias podem ser utilizadas durante todo o processo de recuperação da informação sendo, em função disso, limitado o uso de ontologias como um auxílio para interpretar termos desconhecidos. Nesse sentido, o presente trabalho busca inserir as ontologias como o centro da recuperação da informação apoiada por técnicas de Inteligência Artificial e Processamento de Linguagem Natural, sendo utilizadas para contextualizar os

termos escolhidos pelos usuários, os termos recuperados do *Linked Data* e a construção das respostas.

No âmbito da Ciência da Informação, os trabalhos mencionados não estão focados na conceituação e no uso de recuperação e representação da informação, enquanto este demonstra como a definição do modelo proposto, considerando os aspectos e os modelos da Ciência da Informação, pode contribuir para aprimorar o processo da recuperação da informação, com o auxílio de Inteligência Artificial e processamento de linguagem natural.

### 3 METODOLOGIA

Para o desenvolvimento desta pesquisa foi utilizado o método quadripolar, proposto por Bruyne, Herman e Schoutheete em 1974. Essa proposta divide em quatro polos, os métodos de uma pesquisa: polo epistemológico, polo teórico, polo técnico e polo morfológico.

No polo epistemológico foram tratadas as interseções e relações nas questões epistemológicas relativas à Ciência da Informação, para que assim esta pesquisa possa se embasar cientificamente e delimitar com clareza a problemática e as hipóteses a serem fornecidas. Nesse sentido, o presente trabalho parte do paradigma pós-custodial, informacional e científico apontado por Silva (2006, p. 158) que o define como um paradigma: “[...] emergente porque está a surgir no dealbar, em curso na era da informação e nos meandros de uma conjuntura de transição bastante híbrida, complexa e sujeita a um ritmo de inovação tecnológica e científica quase vertiginoso (a sociedade da informação, em rede etc.)”. A pesquisa se encontra nesse paradigma, pois a complexidade dos ambientes informacionais digitais, no que tange a favorecer uma recuperação da informação mais próxima da linguagem natural, contempla as questões elencadas pelo autor.

Complementarmente, o objeto da pesquisa é a informação como coisa, definido por Buckland (1991). Utiliza-se esse conceito, pois será tratado apenas da informação registrada, que está colocada em algum tipo de suporte e é tangível originalmente. Ainda no polo epistemológico, esta pesquisa está

posicionada no âmbito das Tecnologias de Informação e Comunicação, ao tratar das questões relativas aos estudos de tecnologias dentro da Ciência da Informação, em especial nas pesquisas de informação e tecnologia.

No que tange ao polo teórico, as hipóteses e os objetivos desta pesquisa foram delineados a partir de uma investigação teórica acerca das temáticas do *Linked Data*, das ontologias, da recuperação da informação, da Inteligência Artificial e da Web Semântica, para que a partir dessa abordagem possa ser desenvolvido este trabalho.

O polo técnico reúne as tarefas que possibilitaram a realização da pesquisa, que permitirá validar ou não as hipóteses propostas inicialmente. Nesse sentido, esta pesquisa apresenta-se como um estudo exploratório bibliográfico, pois busca, na literatura e em documentos, embasamento para a construção de um modelo conceitual.

Por fim, no polo morfológico são apresentados os resultados obtidos com a pesquisa, que são em suma: o modelo conceitual, contemplando as características e os princípios teóricos que irão relacionar ontologias, Recuperação da Informação, Inteligência Artificial, Processamento de Linguagem Natural, *Linked Data*.

#### **4 INTELIGÊNCIA ARTIFICIAL E PROCESSAMENTO DE LINGUAGEM NATURAL**

A Inteligência Artificial vem sendo foco de estudos de diversas áreas do conhecimento nos últimos anos. Esse campo de estudos que nasce vinculado à área de Ciência da Computação teve uma grande evolução, em especial pelo avanço tecnológico das últimas décadas, o que levou à criação de diversas tecnologias para o dia a dia das pessoas, tecnologias que usam a Inteligência Artificial e são nela embasadas.

A Inteligência Artificial é definida em quatro categorias principais: sistemas que pensam como humanos; sistemas que agem como humanos, sistemas que pensam de forma racional e sistemas que agem de forma racional. Verifica-se que a abordagem centrada nos humanos possui um caráter mais empírico,

enquanto a abordagem racional está mais focada nas questões matemáticas e ligadas à engenharia (RUSSELL; NORVIG, 2016).

Há diversas áreas e subcampos da Inteligência Artificial que buscam atender propósitos e objetivos diferentes. Cada uma dessas áreas se destaca por algoritmos e técnicas distintos, contribuindo para a evolução da área de diversas formas.

Dentro da área de Inteligência Artificial, encontra-se a área de processamento de linguagem natural, conceito fundamental para o desenvolvimento teórico deste trabalho. O processamento de linguagem natural será utilizado posteriormente, para a realização do modelo e para a integração que será proposta entre Web Semântica e processamento de linguagem natural, além de ser importante para explicar o modo de recuperação da informação que será descrito neste trabalho.

A área de processamento de linguagem natural (PLN) se encontra posicionada dentro da área de Inteligência Artificial, por fazer uso dos mecanismos desse segundo campo de estudo, para realizar a interpretação e a compreensão dos textos.

Vieira e Lopes (2010, p. 184) apontam que: “Processamento de Linguagem Natural (PLN) é uma área de Ciência da Computação que estuda o desenvolvimento de programas de computador que analisam, reconhecem e/ou geram textos em linguagens humanas, ou linguagens naturais”.

Essa definição insere a área de PLN dentro da Ciência da Computação, tendo, porém, uma série de interdisciplinares, em especial com a linguística, pela necessidade de compreensão e tratamento da linguagem natural e humana.

Há diversas abordagens para apresentar o processamento de linguagem natural, que pode ser classificado por suas formas e modelos. Uma abordagem bastante tradicional diz respeito aos níveis de PLN, que demonstram qual a capacidade de processamento, levando em consideração os aspectos que os humanos utilizam para compreender um texto. Destaca-se que o processo de compreensão é bastante dinâmico, havendo uma grande interação entre esses níveis de processamento. Assim, quanto maior for a capacidade e o nível

alcançado por um sistema de processamento de linguagem natural, melhor serão os resultados obtidos.

Nesse sentido, destacam-se os seguintes níveis de PLN: fonologia, morfologia, léxico, sintático, semântico, discursivo e pragmático (LIDDY, 2021).

A existência dos diversos níveis não implica que um sistema de processamento de linguagem natural necessite implementar todos eles. Na verdade, a maioria dos sistemas de PLN opta por implementar os níveis mais baixos, pela própria facilidade, e por serem mais assertivos. Além disso, nem sempre é necessário que todos os níveis sejam implementados para que se atinja o objetivo de uma determinada aplicação.

Nesse sentido, um dos pontos que levam ao desenvolvimento de aplicações que utilizam basicamente os níveis mais baixos da linguagem natural está na maior facilidade de tratar as unidades menores de análise, como morfemas e palavras, enquanto os níveis mais altos, devem tratar sentenças e o texto como um todo, o que envolve relações mais complexas.

Uma importante corrente do desenvolvimento de soluções de recuperação da informação começou a utilizar os conceitos de processamento de linguagem natural (PLN) para desenvolver ferramentas e tecnologias para a realização de buscas. Assim, essas ferramentas buscam analisar de forma automática a linguagem natural humana utilizando algoritmos computacionais para encontrar o que o usuário deseja. A principal técnica expoente desse cenário é o *Question Answering* (QA), que são sistemas que têm como objetivo fornecer informações precisas e diretas respondendo a uma pergunta construída por um usuário.

As técnicas de QA utilizam algoritmos avançados que permitem uma compreensão da linguagem natural, em que um usuário expressa por meio de uma pergunta em linguagem natural uma informação que ele deseja (ALMANSA, 2016). Um exemplo seria o usuário questionar: “Quais livros que tratam de Web Semântica estão disponíveis para empréstimo em uma biblioteca?”, necessitando que um sistema de QA compreenda esta questão, respondendo exatamente o que o usuário deseja, fornecendo, por exemplo, uma lista com os livros disponíveis sobre essa temática.

Em um outro contexto, a área de aprendizagem dentro da Inteligência Artificial é um dos campos mais estudados e aplicados atualmente. Existem diversos sistemas que estão inserindo tais tecnologias no dia a dia das pessoas, criando uma série de facilidades para diversas funções e se tornando um elemento indispensável para as organizações. Exemplos como assistentes virtuais de bancos e operadores de telemarketing virtual utilizam outros elementos da Inteligência Artificial, como o processamento de linguagem natural, mas tem no aprendizado de máquina a sua principal função: a capacidade de “aprender”.

Uma definição de *machine learning* é dada por Jordan e Mitchell (2015, p. 255, tradução nossa):

O aprendizado de máquina é uma disciplina focada em duas questões inter-relacionadas: Como construir sistemas de computador que melhoram automaticamente com a experiência? e Quais são as leis fundamentais da teoria estatística da informação computacional que governam todos os sistemas de aprendizagem, incluindo computadores, seres humanos e organizações? O estudo do aprendizado de máquina é importante tanto para abordar essas questões científicas e de engenharia fundamentais, quanto para o software de computador altamente prático que ele produziu e utilizou em vários aplicativos.

A definição apresentada coloca dois importantes elementos para apontar o que é *machine learning*. O primeiro trata do aspecto da máquina conseguir evoluir a partir das experiências e do uso que vai ocorrendo. Esse primeiro ponto é, atualmente, o mais conhecido das aplicações de aprendizagem de máquinas, e demonstra um dos princípios dessa área, que está em, a partir do uso, fazer acontecer o aprimoramento da aplicação.

O segundo aspecto dessa definição está na questão estatística, em que o uso dessa ciência contribui para compreender as relações entre computadores, seres humanos e organizações. O uso de estatística é essencial na área de *machine learning*, pois o computador, por ter todo o seu funcionamento baseado em leis matemáticas, tem que usar da estatística para conseguir aprender e traçar possibilidades visando dar uma resposta satisfatória a um determinado problema humano.

Um terceiro aspecto trazido pela definição apresentada contrapõe dois importantes elementos dessa área, que diz respeito ao objetivo de se utilizar e de se aplicar *machine learning*. O primeiro trata da aplicação em questões científicas e de engenharia, que utilizam os algoritmos para fazer avançar a ciência e conseguir grandes inovações. O outro aspecto é o caráter prático, que vem sendo aplicado em soluções do cotidiano das pessoas, que pode ser visto em soluções utilizadas por milhões de usuários, como assistentes virtuais de dispositivos computacionais.

A partir da compreensão da IA, PLN e *machine learning*, verifica-se que a Web Semântica, por meio de suas teorias e ferramentas, possui uma série de relações entre esses distintos campos de estudos que podem ser exploradas. Desta forma, na sequência apresenta-se a conceituação acerca da Web Semântica.

## 5 WEB SEMÂNTICA

A Web Semântica vem se desenvolvendo nos últimos anos, tornando-se um elemento fundamental em diversos nichos de pesquisa, ao fornecer um arcabouço conceitual importante para o desenvolvimento da Web, um conjunto de ferramentas que contribuem para diversos contextos da Ciência da Informação e da Ciência da Computação e uma série de aplicações que estão aprimorando a forma como os usuários utilizam a Web e como os dados são nela publicados.

Nessa seara, Berners-Lee, Hendler e Lassila (2001) começam a dar forma a Web Semântica como uma proposta para permitir que os conteúdos informacionais fossem providos com mais significados e mais informações descritivas. Assim, a proposta da Web Semântica traria: “[...] estrutura para o conteúdo significativo das páginas da Web, criando um ambiente no qual os agentes de software que transitam de uma página para outra possam realizar prontamente tarefas sofisticadas para os usuários” (BERNERS-LEE; HENDLER; LASSILA, 2001, não paginado).

Essa citação demonstra que o objetivo da Web Semântica sempre foi facilitar as tarefas realizadas pelos usuários, em especial pela dificuldade existente naquele momento para o usuário realizar até mesmo tarefas simples. Além disso, destaca-se que a estrutura em que as informações estavam disponibilizadas na Web anteriormente a sua proposta, não permitiria aos mecanismos computacionais expandirem o modo como os usuários navegam e utilizam a Web.

Vale destacar que, os conceitos e as ferramentas da Web Semântica estão sendo desenvolvidas e aprimoradas significativamente desde a sua proposta inicial em 2001. Atualmente, as ferramentas da Web Semântica estão sendo utilizadas em diversos âmbitos, não estando restritas ao domínio da Web. Aponta-se que as áreas de recuperação da informação, organização da informação e Inteligência Artificial estão utilizando e contribuindo no aprimoramento das principais ferramentas da Web Semântica.

Primeiramente, vale ressaltar que o RDF é tratado como um modelo de representação de recursos e, por vezes, é considerado apenas uma tecnologia da Web Semântica, pelo seu uso nas mais diversas ferramentas que estão dentro do escopo desse campo de estudo. No entanto, o conceito que embasa a representação de recursos utilizando o modelo RDF tem extrapolado o próprio RDF, uma vez que há uma série de aplicações, que tem em sua base os conceitos da Web Semântica, que utilizam o princípio do RDF, sem utilizar o modelo em si.

No que tange à tecnologia do RDF, esta é uma das mais utilizadas na Web Semântica, pois os dados publicados seguindo os princípios dessa proposta são sempre disponibilizados nesse modelo. No entanto, o conceito do RDF pode ser efetivado por meio de diversos tipos de ferramentas, como XML, JSON, Turtle, entre outros. Destaca-se que isso ocorre devido à evolução tecnológica e a necessidades diversas que existem nas aplicações, que exigem formatos distintos.

Outro elemento central da Web Semântica, ontologia, está se tornando importantes elementos da Web Semântica, para possibilitar informações capazes de contextualizar um cenário, bem como para fornecer informações que

possam ser compreendidas, computacionalmente, no que diz respeito ao sentido que determinados termos podem ter. Nesse contexto, desde o início da Web Semântica, as ontologias estão se tornando cada vez mais populares e utilizadas em diversos cenários, havendo diversas iniciativas que buscam construir ontologias para fornecer uma descrição formalizada de um domínio.

Em uma visão aplicada à Web Semântica e à Ciência da Informação, Santarem Segundo e Coneglian (2015, p. 227) apontam que, do ponto de vista de aplicação tecnológica, “[...] entendem-se as ontologias como artefatos computacionais que descrevem um domínio do conhecimento de forma estruturada, através de classes, propriedades, relações, restrições, axiomas e instâncias”.

Além disso, há diversas linguagens para a construção de ontologias, que possuem mais ou menos expressividade. Nesse contexto, devido a importância das ontologias para a Web Semântica, era necessário que fosse criada uma linguagem de ontologias capaz de efetivamente contextualizar e expressar significados, com um alto nível de semântica formal. Assim, a W3C propôs a *Web Ontology Language* (OWL).

Vale ressaltar, que há uma série de outras tecnologias e ferramentas que são parte da Web Semântica, e são capazes de apoiar o desenvolvimento de aplicações para as mais diversas áreas do conhecimento. No entanto, RDF e ontologias são a base para a construção do modelo proposto neste artigo, que será apresentado na sequência.

## 6 MODELO

Partindo da união entre os conceitos e as ferramentas da Web Semântica, os princípios da Inteligência Artificial, bem como o processamento de linguagem natural e a recuperação da informação, este trabalho apresenta um modelo conceitual de Recuperação da Informação.

A base do modelo está na proposta de recuperação da informação, inspirado nos conceitos do *Question Answering*, em que o usuário irá ter acesso e encontrar as informações de que ele necessita por meio do uso de linguagem

natural. Nesse sentido, o processo de recuperação da informação que tem como base o processamento de linguagem natural está estruturado no modo como a pergunta é compreendida, frente ao modo como a resposta será estruturada e apresentada ao usuário.

No entanto, este modelo busca inserir nesse processo de recuperação da informação, o uso mais efetivo de compreensão textual com a aplicação de técnicas de Inteligência Artificial. A ideia está em tornar a compreensão do que o usuário está querendo mais clara, bem como permitir que o tratamento e a recuperação das informações sejam mais eficientes. Isso será possível pela utilização de estruturas e técnicas da Inteligência Artificial que possibilite a classificação e a compreensão dos textos.

Cabe uma primeira ressalva ao uso do termo compreensão, no que tange à Inteligência Artificial. Quando se utiliza compreensão, nesse cenário, tem-se como perspectiva apresentar que a IA será utilizada visando a mostrar, em históricos recentes e estatísticos, o que aquela palavra pode significar, e com que termos aquele conceito está vinculado. Isso se faz necessário uma vez que é fundamental verificar, de uma forma automatizada, qual é o provável significado daquele termo; além disso, busca-se verificar quais conceitos podem estar vinculados a ele.

Adicionalmente, insere-se neste modelo a principal contribuição ao tratamento utilizando PLN, que está na aplicação dos conceitos e das ferramentas da Web Semântica, visando a tornar o processo efetivamente semântico, considerando o significado e o contexto no qual os termos se encontram. As ferramentas da Web Semântica e Inteligência Artificial serão adotadas buscando diminuir a lacuna (*gap*) existente entre o modo como as tecnologias computacionais tratam conceitos, em que o significado não é claro e não está definido efetivamente para o seu tratamento computacional.

Nesse contexto, ainda que a aplicação de técnicas de Inteligência Artificial possa dar uma visão geral, utilizando como premissas os usos anteriores de termos, existe a falta de uma compreensão de quais relacionamentos e significados os termos efetivamente possuem. A aplicação das ferramentas da Web Semântica juntamente com o uso de Inteligência Artificial e do

processamento de linguagem natural podem tornar o processo mais efetivo, e se aproximando, de fato, de uma compreensão daquilo que uma pessoa está dizendo (ou escrevendo).

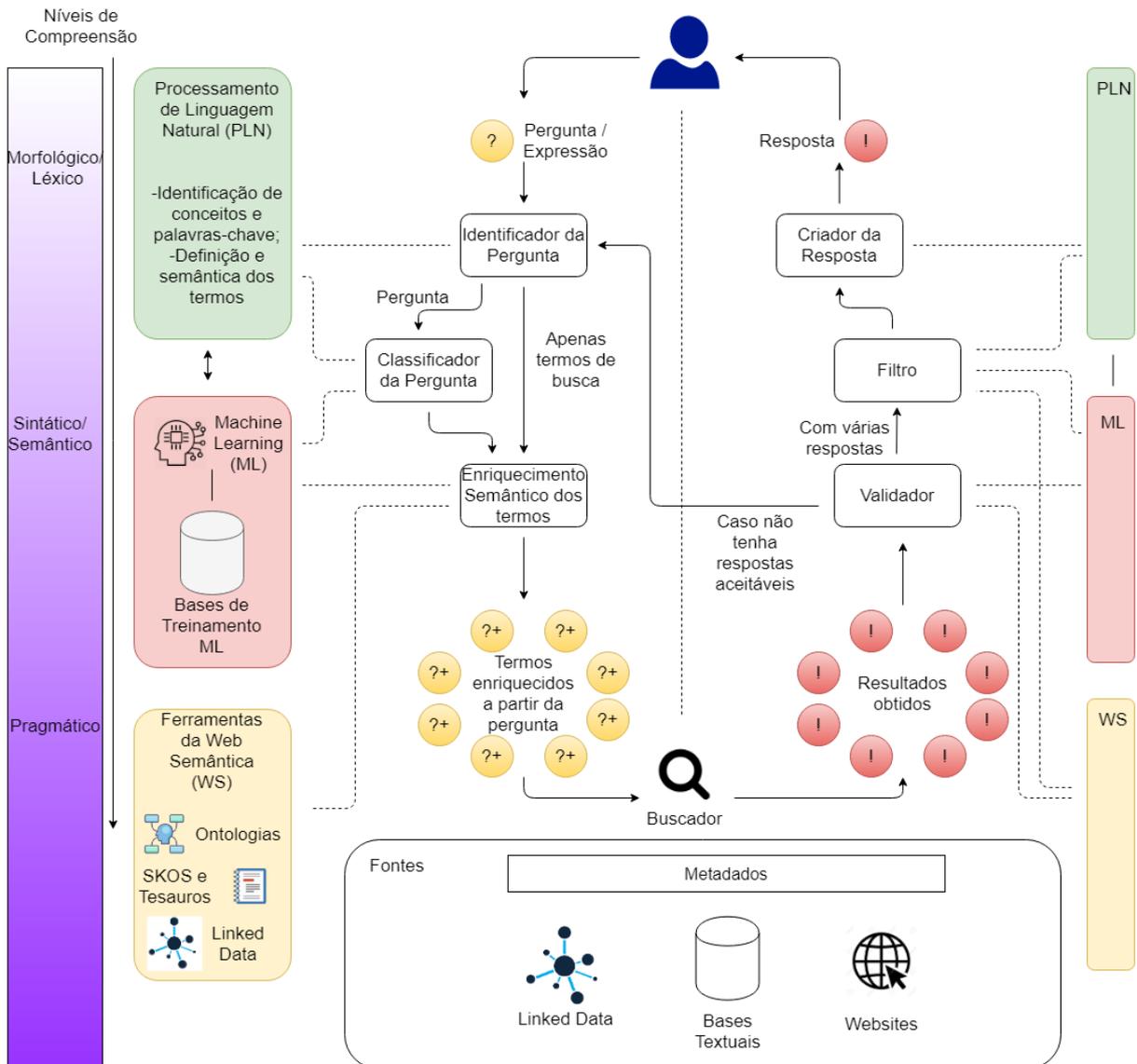
Por fim, uma característica central e determinante para tornar o processo de fato efetivo e eficiente está no tratamento e no uso de fontes informacionais que possuem um nível de semântica formal elevado. Nesse contexto, o uso de dados estruturados em formato de *Linked Data*, que utilizam modelos de dados e vocabulários expressivos e que estão publicados e interligados a outros modelos de dados, torna a recuperação da informação mais precisa. Destaca-se que, ainda que a qualidade dos dados publicados no *Linked Data* não seja boa, o uso de dados publicados nesse formato permite que as ontologias que apoiam o modelo tenham uma maior aderência aos dados, permitindo que os resultados alcançados sejam melhores.

Isso ocorre pelo fato de que, ao utilizar dados que estão estruturados com base em ontologias, por exemplo, é possível ter uma compreensão mais clara do que as informações ali contidas significam. O princípio está em que, ao compreender o que uma pergunta significa, um sistema desenvolvido a partir do modelo proposto, teria como resultado grafos que traduziriam aquilo que o usuário deseja saber; com tais grafos, o acesso e a recuperação em fontes informacionais de *Linked Data* seria natural e, por consequência, facilitado. Sendo assim, a recuperação da informação consideraria uma série de aspectos semânticos para recuperar as informações que foram buscadas.

Vale destacar que o modelo proposto é aplicado apenas à ambientes informacionais digitais, pela necessidade da exploração do documento para a construção dos resultados e das respostas. Assim, não é possível aplicar tal modelo em ambientes analógicos, como em acervos de bibliotecas tradicionais.

Depois dessa breve explicação do modelo, apresenta-se a seguir, na Figura 1, o modelo proposto.

**Figura 1 - Modelo conceitual de recuperação da informação**



Fonte: elaborado pelos autores

O modelo apresentado na Figura 1 está dividido em quatro partes principais: a primeira (superior central) trata da parte referente a todo o processo de recuperação da informação, e a segunda (inferior central) demonstra as fontes informacionais que serão utilizadas para possibilitar o processo de recuperação da informação. Nas duas laterais, encontram-se três importantes camadas: processamento de linguagem natural, *machine learning* e ferramentas da Web Semântica, responsáveis por apoiar o processo da recuperação da informação. E na lateral esquerda, está presente uma escala de nível de

compreensão, utilizada para demonstrar o aprofundamento nos níveis linguísticos.

As três camadas das laterais serão chamadas de camadas de suporte, pois elas que viabilizam o modelo por meio de suas tecnologias e algoritmos. A primeira camada é a de processamento de linguagem natural, que é responsável por fazer o entendimento e identificação dos termos, definir as categorias de tais termos, apoiar a montagem da resposta e apoiar a classificação da pergunta.

Essa camada está fortemente vinculada à camada de *machine learning*, responsável por utilizar algoritmos, como o de *topic modeling*, como uma técnica de aprendizado não supervisionado, para auxiliar o processo de identificação dos termos e palavras-chave. Utilizam-se também técnicas de aprendizado supervisionado de classificação, que buscam classificar os conteúdos em categorias definidas. Com ambas as técnicas de *machine learning*, é possível considerar que, além disso, essa camada auxilia no processo de enriquecimento e validação, uma vez que as bases de treinamento do algoritmo irão promover uma constante evolução em comparação com outros textos e resultados.

A terceira camada é a de ferramentas da Web Semântica, focada, especialmente, no enriquecimento da pergunta e no apoio do filtro e da validação da resposta. Em suma, as ferramentas da Web Semântica participam das etapas que necessitam aumentar o nível de semântica formal, tornando os resultados mais aprimorados. As principais ferramentas da Web Semântica utilizadas são ontologias, modelos de dados em RDF e SKOS, regras e inferências com o SWRL e o SPARQL, para promover a recuperação de dados.

Quanto ao processo do modelo, que se encontra na parte central da Figura 1, considere-se que esse processo está dividido em duas partes, pergunta e resposta, trazendo, assim, um conceito oriundo do *Question Answering*, em que o processo é tratado dessa forma: o usuário realiza a pergunta, o sistema trata e compreende essa pergunta, busca-se o resultado, e o sistema reage em formato de resposta, sendo essa a principal aproximação entre o QA e o modelo proposto.

Essa primeira parte do modelo é composta basicamente por sete módulos: identificador da pergunta, classificador da pergunta, enriquecimento

semântico da pergunta, buscador, validador, filtro e criador da resposta, apresentados a seguir.

O primeiro módulo é o **identificador da pergunta** que tem como objetivo verificar se o que o usuário buscou é de fato uma pergunta, ou apenas termos de uma busca normal. Esse processo é importante, pois, a partir disso, o modelo irá realizar ou o tratamento daquela pergunta, direcionando para o módulo do classificador da pergunta, ou se for apenas termos de pergunta, já direciona diretamente para o enriquecimento semântico da pergunta.

No caso de pergunta, a próxima etapa é o **classificador da pergunta**, que irá receber a pergunta feita pelo usuário e irá realizar um processo de identificação e análise de, primeiramente, qual é o tipo de pergunta que o usuário está fazendo e, mais adiante, verificação de quais são os termos e pontos-chave que compõem essa pergunta. Para isso o classificador de perguntas irá se relacionar com os algoritmos de *machine learning*, para que, por meio do aprendizado supervisionado com o uso da técnica de classificação, seja possível categorizar conceitos e termos junto ao módulo de processamento de linguagem natural, que visa a compreender como tais termos estão sendo utilizados.

Destaca-se que, nessa primeira etapa, o foco não está em compreender ou em obter o significado dos termos, mas, sim, em ter uma visão do que o usuário está querendo ao fazer aquela pergunta. Nesse sentido, essa etapa tem como objetivo apenas entender a pergunta, levantando os termos principais que o usuário deseja.

Posteriormente, após essa classificação, que fornecerá os termos-chave e a identificação de qual é o tipo de pergunta que o usuário está fazendo, a próxima etapa está no **enriquecimento semântico dos termos**. Esse elemento do modelo recebe a análise prévia obtida pelo classificador, com os termos-chave levantados, para, a partir daí, obter o significado e as relações que existem entre esses termos e outros, buscando ter uma visão mais clara dos conceitos que circulam e se relacionam aos inseridos pelo usuário.

O enriquecimento semântico está vinculado a dois elementos principais, as ontologias e os vocabulários, oriundos das ferramentas da Web Semântica, e aos classificadores de texto, oriundos da camada de *machine learning*.

Ontologias serão utilizadas nesse processo para possibilitar que o enriquecimento semântico leve em consideração o significado e o contexto em que os termos obtidos estão. Como relatado anteriormente, uma ontologia é capaz de expressar uma série de relações e propriedades que expressam semanticamente como um termo está posicionado frente a um determinado cenário.

O uso da ontologia é essencial nesse processo, uma vez que o desenvolvimento de ontologias é realizado considerando aspectos linguísticos e da biblioteconomia, que são capazes de expressar um domínio com clareza. No âmbito deste modelo, as principais funções das ontologias são:

- Definição de sinônimos;
- Definição de termos relacionados;
- Definição de termos genéricos e específicos;
- Definição de atributos que são relacionados a conceitos;

Adicionalmente, o enriquecimento semântico utiliza um dos principais conceitos propostos neste trabalho, que está na união entre os princípios da Web Semântica com a Inteligência Artificial, por meio de técnicas de processamento de linguagem natural com processos de compreensão e classificação dos termos. Nesse contexto, o processo realizado pelo enriquecimento semântico será aprimorado com elementos classificadores de texto, que buscam, por meio de aprendizagem de máquina, com o uso de algoritmos de *topic modelling*, obter informações e classificações sobre os termos e conceitos previamente classificados pela ontologia.

Assim, a classificação ocorrerá em duas etapas, permitindo que o nível de semântica formal seja mais elevado, o que será apresentado com detalhes nas subseções seguintes, ao mesmo tempo que aumenta a probabilidade de ter pelo menos algumas informações daquele termo, seja via ontologia, seja via classificador de texto. Além disso, o uso desses classificadores é interessante, pois o uso de aprendizado de máquina tornará o processo mais aprimorado ao longo do tempo e, por outro lado, permitirá obter características das relações das ontologias.

Ao final desse processo, aquela pergunta ou termo inicial estará expandido, ampliado e enriquecido, o que é representado pelas bolinhas com o “?+” na Figura 1.

A etapa seguinte está no **buscador**, que tem como função encontrar as fontes informacionais capazes de atender as necessidades informacionais dos usuários. Nesse sentido, o buscador terá três tipos principais de fontes de informação: dados publicados em formatos de *Linked Data*, fontes textuais disponíveis na Web e bases de dados estruturadas diversas.

O buscador tem uma importante função, uma vez que todos os tratamentos realizados nas etapas anteriores terão como resultado, em um primeiro momento, o modo como a busca é realizada nas fontes informacionais. Assim, o buscador terá que realizar tratamento para o *Linked Data*, de como que a busca deverá ser realizada em bases RDF, utilizando linguagem SPARQL. A busca em bases de *Linked Data* é a mais completa, uma vez que possibilita utilizar as relações encontradas nas ontologias, além de permitir buscas nos elementos descritivos e expandir a sua compreensão. Isso é possível, pois OWL e RDF utilizam a mesma sintaxe, e a busca realizada em SPARQL pode utilizar todos os tipos de relações que foram identificadas na ontologia.

No que tange ao buscador, salienta-se que as fontes obtidas com o *Linked Data* serão essenciais para se ter um resultado mais satisfatório, pois as buscas realizadas nas bases do *Linked Data* terão uma grande gama de relações e propriedades que foram obtidas a partir da ontologia, e que podem ser localizadas por meio do SPARQL. Assim, os resultados obtidos com as fontes informacionais do *Linked Data* irão auxiliar o processo de busca nas outras fontes, como os textos e as bases de dados.

Vale destacar que as buscas são realizadas utilizando tanto as informações descritivas, por meio dos metadados, quanto os próprios documentos, que permitem um melhor processo de recuperação da informação. Por tal motivo, apresenta-se no modelo a questão dos metadados, que são utilizados neste processo.

Dessa forma, serão realizadas buscas que serão aprimoradas, visando a obter resultados que sejam semanticamente aderentes àquilo que os usuários

estão buscando. Isso é fundamental, pois o processo de recuperação de informação é essencial e deve ser realizado utilizando os princípios da Web Semântica e as informações extraídas das ontologias.

O módulo que aparece na sequência é o **validador**, que tem como função identificar se o processo realizado foi satisfatório. O princípio desse módulo está em utilizar as informações contextuais das ferramentas da Web Semântica e realizar uma análise das respostas com o *machine learning*, para verificar se os resultados foram satisfatórios. Caso não tenha nenhum resultado satisfatório, o processo retorna para a fase de identificação da pergunta, para que seja todo refeito. No caso de haver ao menos uma resposta, o processo segue para a etapa do filtro.

Na etapa seguinte, a do **filtro**, as respostas potenciais encontradas passam por mais um refinamento, trazendo os elementos semânticos e de classificação, que usam como base princípios de aprendizagem de máquina, para aprimorar e definir as melhores respostas para o usuário. Com a aplicação do algoritmo supervisionado de classificação, é possível obter categorias e informações mais claras das informações obtidas, auxiliando na definição daquela que pode ser mais adequada para ser a resposta dada ao usuário. Esse refinamento será apresentado nas seções seguintes. O princípio desta etapa está no uso dos significados e do contexto em que os termos se encontram para definir o que será dado como resposta para o usuário ao final do processo.

Esse processo está vinculado aos conceitos do *Question Answering*, em que o usuário irá receber uma resposta exata daquilo que está buscando. Assim, dentre os documentos e dados obtidos no buscador, deve-se realizar um processo em que é encontrado exatamente qual é a resposta daquilo que o usuário está buscando.

Esse processo é feito centrado em técnicas de processamento de linguagem natural, com o uso de palavras-chave e da proximidade entre os termos, como discutido anteriormente. No entanto, esse processo também utilizará as relações identificadas na ontologia, que serão apresentadas com detalhes nas subseções seguintes, para aprimorar os resultados alcançados, além de oferecer uma melhor compreensão do que o usuário deseja, e das

informações encontradas. As relações das ontologias fornecerão outros elementos que podem ser utilizados para a identificação de informações dentro dos documentos e das informações encontradas.

O uso das ontologias nesse processo ocorrerá para que seja possível compreender o significado dos termos, bem como as relações existentes, para que, ao identificar quais das respostas potenciais satisfazem a busca, comparar com tais relações e termos e conceitos que fornecem um nível de semântica formal e de contexto da busca realizada.

Finalmente, a última etapa do processo é o **criador da resposta**, que irá finalizar o processo fornecendo para o usuário o que foi encontrado de uma forma aderente àquilo que ele buscou. O criador de respostas irá se embasar também em bases de dados e significados para apresentar uma resposta mais aderente ao contexto do usuário, considerando os significados e sinônimos encontrados, dos termos utilizados na busca do usuário.

Essa etapa também contempla a interface de resposta que o usuário vai receber, tendo a possibilidade de retroalimentar a busca, para aprimorar, caso o usuário não tenha as suas necessidades informacionais contempladas. Esse processo interativo e iterativo aprimora os resultados alcançados, melhorando os algoritmos de aprendizagem de máquinas adotados, ao mesmo tempo que se aproxima de dar melhores resultados aos usuários.

Por fim, a última informação contida no modelo trata dos níveis de compreensão que se atinge com o modelo. Esses níveis são embasados no processamento de linguagem natural, e vai se aprofundando a partir dos módulos e do tratamento que é dado aos termos. Nesse sentido, a primeira etapa tem apenas uma compreensão morfológica e léxica, visto que se tem apenas um reconhecimento nocional de elementos da escrita ou da fala. Na sequência, os níveis sintático e semântico são utilizados para a identificação da pergunta, até a sua classificação, atingindo o enriquecimento, juntamente com o nível pragmático, em que PLN, *machine learning* e ferramentas da Web Semântica são utilizadas para atingir um nível mais aprofundado do nível de compreensão.

Outro aspecto importante a ser destacado está nos objetivos do modelo. Em suma, tem-se três objetivos principais: realizar a compreensão da pergunta

e da resposta que será dada. Tal compreensão é responsável pelo entendimento da pergunta, além de trazer a necessidade de uma interface que irá criar a resposta que atende as necessidades informacionais dos usuários. Esse objetivo tem como foco principal o processamento de linguagem natural para permitir que a pergunta seja compreendida de forma eficiente.

O segundo objetivo está vinculado à semântica formal. O foco está em obter qual é o significado daquelas informações trazidas a partir do tratamento da pergunta, relacionando os conceitos que as técnicas de Inteligência Artificial estão encontrando, juntamente com a semântica relacionada às ferramentas da Web Semântica. Esse mesmo ponto é abordado no que se refere à definição do que será dado como resposta para os usuários, realizando os relacionamentos das informações que lhes serão apresentadas.

Finalmente, o terceiro objetivo é a recuperação da informação, em que serão levantadas as informações que atendem às necessidades informacionais dos usuários, levando em consideração os aspectos da análise semântica realizada, tanto no aspecto da IA, quanto das ferramentas da Web Semântica. Devido a tal objetivo, realiza-se a busca das informações e de tratamento das respostas obtidas, visando a passar para as fases seguintes do processo.

## **7 CONSIDERAÇÕES FINAIS**

As relações interdisciplinares estão permitindo uma real evolução em diversos campos de pesquisa, bem como, está trazendo reais inovações para o cotidiano das pessoas. Em especial, a aplicação de Tecnologias da Informação e Comunicação neste processo, está criando revoluções nas mais diversas áreas do conhecimento. A Inteligência Artificial, por exemplo, está presente em uma parcela significativa das soluções propostas e criadas atualmente.

Desta forma, trazer a Inteligência Artificial, junto com os seus subcampos, como Processamento de Linguagem Natural e Aprendizado de Máquinas, traz contribuições para a Ciência da Informação. Primordialmente, como proposto por esta pesquisa, a Recuperação da Informação pode ser fortemente impactada, com a criação de novos modelos e formas de se relacionar com os usuários.

Vale destacar ainda, que a Web Semântica, por meio de suas teorias e ferramentas, é capaz de tornar a compreensão da semântica formal, por parte dos algoritmos computacionais, muito mais elevada, permitindo assim, a criação de tecnologias mais eficientes no relacionamento com humanos.

Desta forma, o modelo proposto neste artigo, se mostra, ao vincular a Inteligência Artificial com a Web Semântica no âmbito da Recuperação da Informação, como uma forma de relacionar a Ciência da Informação com esses emergentes campos de estudos. Além disso, demonstra um caminho para uma série de relações interdisciplinares que podem ser realizadas.

O modelo proposto, quando implementado, parcial ou totalmente, pode se tornar uma ferramenta integrável a tecnologias como *chatbot* ou assistentes pessoais, bem como com as próprias ferramentas de recuperação da informação. Compreende-se, que esse modelo conceitual demonstra um conjunto de vínculos entre as áreas tratadas, mas se mostra, como uma alternativa de entendermos a Recuperação da Informação em um cenário, em que a Inteligência Artificial está presente em grande parte das aplicações propostas.

Portanto, esta pesquisa valida as relações interdisciplinares entre Ciência da Informação e Ciência da Computação, a partir da proposta de um novo modelo de recuperação da informação, pautado na Inteligência Artificial e seus diversos campos, e a Web Semântica.

## REFERÊNCIAS

ABDI, A.; IDRIS, N.; AHMAD, Z. Q. An ontology-based question answering system in the physics domain. **Soft Computing**, [S.l.], v. 22, n. 1, p. 213-230, 2018.

ALMANSA, L. F. **Um framework de Question Answering nos domínios de epigenética e de imagens citológicas de tireoide**. 2016. 88 f. Dissertação (Mestrado em Ciências) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2016.

ASIAEE, A. H.; MINNING, T.; DOSHI, P.; TARLETON, R. L. A framework for ontology-based Question Answering with application to parasite immunology. **Journal of biomedical semantics**, [S.l.], v. 6, n. 1, p. 1, 2015.

BERNARDO, R. O.; SANTACHÉ, A. **Aplicação de chatbots no desenvolvimento de jogos em saúde**. 2017. 12 f. Relatório técnico (Projeto Final de Graduação) – Instituto de Computação, Universidade Estadual de Campinas, Campinas, 2017. Disponível em: <http://www.ic.unicamp.br/~reltech/PFG/2017/PFG-17-22.pdf>. Acesso em: 27 ago. 2021.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific american**, [S.l.], v. 284, n. 5, p. 28-37, 2001.

BOUZIANE, A.; BOUCHIHA, D.; DOUMI, N.; MALKI, M. Toward an arabic question answering system over linked data. **Jordanian Journal of Computers and Information Technology (JJCIT)**, [S.l.], v. 4, n. 02, 2018.

BUCKLAND, M. K. A informação como coisa. **Journal of the American Society for Information Science**, v. 45, n. 5, p. 351-360, 1991

CABALEIRO, B.; PEÑAS, A.; MANANDHAR, S. Grounding proposition stores for question answering over linked data. **Knowledge-Based Systems**, [S.l.], v. 128, p. 34-42, 2017.

COSTA, E.; CAMPELO, C.; SAMPAIO, L. Classificação automática de questões. Problema de matemática para aplicações do pensamento computacional na educação. *In*: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 7., 2018. **Anais [...]**. [S.l.]: WCBIE, 2018. p. 569-578.

FERRÁNDEZ, O.; IZQUIERDO, R.; FERRÁNDEZ, S.; VICEDO, J. L. Addressing ontology-based question answering with collections of user queries. **Information Processing & Management**, [S.l.], v. 45, n. 2, p. 175-188, 2009.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. **Science**, [S.l.], v. 349, n. 6245, p. 255-260, 2015. Disponível em: <https://cs.uwaterloo.ca/~y328yu/mycourses/480-2018/readings/JordanMitchell.pdf>. Acesso em: 31 mar. 2021.

KUMAR, G. S.; ZAYARAZ, G. Concept relation extraction using Naïve Bayes classifier for ontology-based Question Answering systems. **Journal of King Saud University-Computer and Information Sciences**, [S.l.], v. 27, n. 1, p. 13-24, 2015. Disponível em: [http://ac.els-cdn.com/S1319157814000020/1-s2.0-S1319157814000020-main.pdf?\\_tid=897d5a8a-6701-11e6-944e-00000aab0f26&am](http://ac.els-cdn.com/S1319157814000020/1-s2.0-S1319157814000020-main.pdf?_tid=897d5a8a-6701-11e6-944e-00000aab0f26&am)

acdnat=1471716878\_680a3eddb48f7f42cc5ef92ed05ac125. Acesso em: 17 ago. 2021.

LIDDY, E. D. Natural language processing. *In: Encyclopedia of Library and Information Science*, 2nd ed. NY: Marcel Decker, Inc. 2001. Disponível em: <https://surface.syr.edu/cgi/viewcontent.cgi?referer=http://scholar.google.com.br/&httpsredir=1&article=1019&context=cnlp>. Acesso em: 5 abr. 2021.

LOPEZ, V.; UNGER, C.; CIMIANO, P.; MOTTA, E. Evaluating Question Answering over Linked Data. *Web Semantics: Science, Services and Agents on the World Wide Web*, [S.l.], v. 21, p. 3-13, 2013. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S157082681300022X>. Acesso em: 14 mar. 2021.

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence: a modern approach**. Malaysia: Pearson Education Limited, 2016.

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Tecnologias da Web Semântica aplicadas a organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados. *In: GUIMARÃES, J. A. C.; DODEBEI, V. (org.). Organização do Conhecimento e Diversidade Cultural*. 1. ed. Marília: Fundepe, 2015. v. 3, p. 224-233. Disponível em: <http://isko-brasil.org.br/wp-content/uploads/2015/09/Organiza%C3%A7%C3%A3o-do-Conhecimento-e-Diversidade-Cultural-ISKO-BRASIL-2015.pdf>. Acesso em: 15 abr. 2021.

SILVA, A. M. da. **A Informação: da compreensão do fenômeno e construção do objeto científico**. Santa Maria da Feira: Edições Afrontamento, 2006.

VIEIRA, R.; LOPES, L. Processamento de linguagem natural e o tratamento computacional de linguagens científicas. **EM CORPORA**, p. 183-201, 2010. Disponível: [https://s3.amazonaws.com/academia.edu.documents/33579040/linguagensespecializadasemcorpora.pdf?response-content-disposition=inline%3B%20filename%3DLINGUAGENS\\_ESPECIALIZADAS.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20190705%2Fus-east-1%2Fs3%2Faws4\\_request&X-Amz-Date=20190705T174023Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=3b8d5c040735319be304f7cd4b8c23a3a6ee1f42715e40a4781c08c07cc6b7ef#page=184](https://s3.amazonaws.com/academia.edu.documents/33579040/linguagensespecializadasemcorpora.pdf?response-content-disposition=inline%3B%20filename%3DLINGUAGENS_ESPECIALIZADAS.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20190705%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20190705T174023Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=3b8d5c040735319be304f7cd4b8c23a3a6ee1f42715e40a4781c08c07cc6b7ef#page=184). Acesso em: 15 abr. 2021.

WALTER, S.; UNGER, C.; CIMIANO, P.; BAR, D. Evaluation of a layered approach to Question Answering over Linked Data. *In: The Semantic Web – ISWC 2012*. Springer Berlin Heidelberg, p. 362-374, 2012. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-642-35173-0\\_25](https://link.springer.com/chapter/10.1007/978-3-642-35173-0_25). Acesso em: 14 mar. 2020.

XIE, X.; SONG, W.; LIU, L.; DU, C.; WANG, H. Research and implementation of automatic question answering system based on ontology. *In: The 27th Chinese Control and Decision Conference (2015 CCDC)*. IEEE, 2015. p. 1366-1370.

## ARTIFICIAL INTELLIGENCE AND SEMANTIC WEB TOOLS APPLIED TO INFORMATION RETRIEVAL: A CONCEPTUAL MODEL WITH A FOCUS ON NATURAL LANGUAGE

### ABSTRACT

**Objective:** the present research has as objective the proposition of a model of Information Retrieval, that redraws this field of studies, from the approximation of the computational language with the natural language, using the principles of the representation of the information so that the meaning and the context of the data are explicit for the search process, for this purpose the Artificial Intelligence, Natural Language Processing and the Semantic Web tools are related and related. **Methodology:** using the quadripolar method, namely: epistemological pole, theoretical pole, technical pole and morphological pole. Furthermore, the research was exploratory, having an applied character. **Results:** As results, this Information Retrieval model was created, based on the semantic context and the application of Artificial Intelligence, capable of making natural language the basis of the process, and considering the context and the meaning of the terms for users. **Conclusions:** it is pointed out that the present work makes an important approximation between Information Science and Artificial Intelligence, bringing to its scope, especially in the scope of Information Retrieval, real applications of how this second field of studies can improve the area as a whole.

**Descriptors:** Information Retrieval. Semantic Web. Artificial intelligence. Natural Language Processing. Ontology.

## INTELIGENCIA ARTIFICIAL Y HERRAMIENTAS WEB SEMÁNTICAS APLICADAS A LA RECUPERACIÓN DE INFORMACIÓN: UN MODELO CONCEPTUAL ENFOCADO EN EL LENGUAJE NATURAL

### RESUMEN

**Objetivo:** el artículo tiene como objetivo proponer un modelo de recuperación de información que rediseñe este campo de estudios, a partir de la aproximación del lenguaje computacional con el lenguaje natural, utilizando los principios de representación de la información, de manera que el significado y contexto de los datos sea explícito para el proceso de búsqueda; para ello, se acerca y se relaciona con los procesos de Inteligencia Artificial, el procesamiento del lenguaje natural y las herramientas de la Web Semántica. **Metodología:** utilizando el método cuadripolar, a

saber: polo epistemológico, polo teórico, polo técnico y polo morfológico. Además, la investigación fue exploratoria, de carácter aplicado. **Resultados:** Como resultado, se creó este modelo de recuperación de información, basado en el contexto semántico y la aplicación de la Inteligencia Artificial, capaz de hacer del lenguaje natural la base del proceso, y considerando el contexto y significado de los términos para los usuarios. **Conclusiones:** Se señala que el presente trabajo realiza una importante aproximación entre las Ciencias de la Información y la Inteligencia Artificial, trayendo a su alcance, especialmente en el campo de la recuperación de información, aplicaciones reales de cómo este segundo campo de estudios puede mejorar el área en su conjunto.

**Descriptores:** Recuperación de información. Web semántica. Inteligencia artificial. Procesamiento natural del lenguaje. Ontología.

**Recebido em:** 10.12.2021

**Aceito em:** 30.04.2022