

Comunicação, Mídias e Educação

Vanessa Cristina de Abreu Torres Hrenechen
(Organizadora)

/Promotion
/Research
/Business
/Development
/Engineering
/Manufacturing
/Planning

Atena
Editora
Ano 2019

Vanessa Cristina de Abreu Torres Hrenechen

(Organizadora)

Comunicação, Mídias e Educação

Atena Editora
2019

2019 by Atena Editora
Copyright © Atena Editora
Copyright do Texto © 2019 Os Autores
Copyright da Edição © 2019 Atena Editora
Editora Executiva: Profª Drª Antonella Carvalho de Oliveira
Diagramação: Karine de Lima
Edição de Arte: Lorena Prestes
Revisão: Os Autores

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores. Permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Conselho Editorial

Ciências Humanas e Sociais Aplicadas

Prof. Dr. Álvaro Augusto de Borba Barreto – Universidade Federal de Pelotas
Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná
Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília
Prof. Dr. Constantino Ribeiro de Oliveira Junior – Universidade Estadual de Ponta Grossa
Profª Drª Cristina Gaio – Universidade de Lisboa
Prof. Dr. Deyvison de Lima Oliveira – Universidade Federal de Rondônia
Prof. Dr. Gilmei Fleck – Universidade Estadual do Oeste do Paraná
Profª Drª Ivone Goulart Lopes – Istituto Internazionale delle Figlie de Maria Ausiliatrice
Profª Drª Juliane Sant’Ana Bento – Universidade Federal do Rio Grande do Sul
Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense
Profª Drª Lina Maria Gonçalves – Universidade Federal do Tocantins
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Profª Drª Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa
Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará
Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande
Prof. Dr. Willian Douglas Guilherme – Universidade Federal do Tocantins

Ciências Agrárias e Multidisciplinar

Prof. Dr. Alan Mario Zuffo – Universidade Federal de Mato Grosso do Sul
Prof. Dr. Alexandre Igor Azevedo Pereira – Instituto Federal Goiano
Profª Drª Daiane Garabeli Trojan – Universidade Norte do Paraná
Prof. Dr. Darllan Collins da Cunha e Silva – Universidade Estadual Paulista
Prof. Dr. Fábio Steiner – Universidade Estadual de Mato Grosso do Sul
Profª Drª Girlene Santos de Souza – Universidade Federal do Recôncavo da Bahia
Prof. Dr. Jorge González Aguilera – Universidade Federal de Mato Grosso do Sul
Prof. Dr. Ronilson Freitas de Souza – Universidade do Estado do Pará
Prof. Dr. Valdemar Antonio Paffaro Junior – Universidade Federal de Alfenas

Ciências Biológicas e da Saúde

Prof. Dr. Gianfábio Pimentel Franco – Universidade Federal de Santa Maria
Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás
Prof.^a Dr.^a Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina
Prof. Dr. José Max Barbosa de Oliveira Junior – Universidade Federal do Oeste do Pará
Prof.^a Dr.^a Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof.^a Dr.^a Raissa Rachel Salustriano da Silva Matos – Universidade Federal do Maranhão
Prof.^a Dr.^a Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa
Prof.^a Dr.^a Vanessa Bordin Viera – Universidade Federal de Campina Grande

Ciências Exatas e da Terra e Engenharias

Prof. Dr. Adélio Alcino Sampaio Castro Machado – Universidade do Porto
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná
Prof. Dr. Fabrício Menezes Ramos – Instituto Federal do Pará
Prof.^a Dr.^a Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista

Conselho Técnico Científico

Prof. Msc. Abrãao Carvalho Nogueira – Universidade Federal do Espírito Santo
Prof.^a Dr.^a Andreza Lopes – Instituto de Pesquisa e Desenvolvimento Acadêmico
Prof. Msc. Carlos Antônio dos Santos – Universidade Federal Rural do Rio de Janeiro
Prof.^a Msc. Jaqueline Oliveira Rezende – Universidade Federal de Uberlândia
Prof. Msc. Leonardo Tullio – Universidade Estadual de Ponta Grossa
Prof. Dr. Welleson Feitosa Gazel – Universidade Paulista
Prof. Msc. André Flávio Gonçalves Silva – Universidade Federal do Maranhão
Prof.^a Msc. Renata Luciane Polsaque Young Blood – UniSecal
Prof. Msc. Daniel da Silva Miranda – Universidade Federal do Pará

Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)	
C741	Comunicação, mídias e educação [recurso eletrônico] / Organizadora Vanessa Cristina de Abreu Torres Hrenechen. – Ponta Grossa, PR: Atena Editora, 2019. Formato: PDF Requisitos de sistema: Adobe Acrobat Reader. Modo de acesso: World Wide Web. Inclui bibliografia ISBN 978-85-7247-344-6 DOI 10.22533/at.ed.446192205 1. Aprendizagem. 2. Comunicação – Pesquisa – Brasil. 3. Comunicação na educação. I. Hrenechen, Vanessa Cristina de Abreu Torres. CDD 371.1022
Elaborado por Maurício Amormino Júnior – CRB6/2422	

Atena Editora
Ponta Grossa – Paraná - Brasil
www.atenaeditora.com.br
contato@atenaeditora.com.br

APRESENTAÇÃO

Essa obra reúne um conjunto de pesquisas sobre as novas tecnologias e técnicas aplicadas à comunicação. O compilado de artigos traz contribuições relevantes para a comunidade científica e profissionais da área.

O e-book, composto por 36 artigos, apresenta diálogos contemporâneos e reflexões sobre o papel da comunicação nos mais diversos âmbitos. Estudos analisam o uso das novas mídias na educação e avaliam a convergência dos meios na partilha de informações e aprendizagem em conjunto. Pesquisas também retratam o consumo midiático, culturas comunicacionais e as manifestações no espaço urbano.

Há artigos sobre o ambiente *comunicacional* digital e o impacto das novas tecnologias na sociedade. Autores também discutem as discrepâncias entre as visões de mundo dos jornalistas e dos usuários de redes sociais e o papel dos meios de comunicação na representação da realidade. O volume traz pesquisadores de peso que compartilham conhecimento e estimulam novos estudos na área da comunicação.

Vanessa Cristina de Abreu Torres Hrenechen

SUMÁRIO

CAPÍTULO 1	1
OS PRIMEIROS PASSOS DO MUSEU DE GEOCIÊNCIAS DA UNIVERSIDADE FEDERAL DE RORAIMA (MUGEO): HISTÓRICO E ACERVO	
Lena Simone Barata Souza Ezequias Nogueira Guimarães	
DOI 10.22533/at.ed.4461922051	
CAPÍTULO 2	16
CARTOGRAFÍA DIGITAL INTERACTIVA DE LO PATRIMONIAL: DEL RELATO AL “DATO” Y VICEVERSA	
Liliana Fracasso David Aperador Francisco Cabanzo	
DOI 10.22533/at.ed.4461922052	
CAPÍTULO 3	33
A UTILIZAÇÃO DE MAQUETES E IMAGENS TÁTEIS COMO IMPULSIONADORAS DO APRENDIZADO PARA CEGOS E PESSOAS COM BAIXA VISÃO NAS GEOCIÊNCIAS	
Loruama Geovanna Guedes Vardiero Rodson Abreu Marques Tamires Costa Velasco Matheus Gomes Fanelli Jeruza Lacerda Benincá Barbosa Sandro Lúcio Mauri Ferreira	
DOI 10.22533/at.ed.4461922053	
CAPÍTULO 4	45
REPRESENTAÇÃO DA PESSOA COM DEFICIÊNCIA NA TV: UMA ANÁLISE DA SÉRIE “SOBRE RODAS” COM O PARATLETA FERNANDO FERNANDES	
Antonio Janiel Ienerich da Silva Henrique Alexander Grazzi Keske	
DOI 10.22533/at.ed.4461922054	
CAPÍTULO 5	62
ASPECTOS EPISTEMOLÓGICOS DA EXPERIÊNCIA NARRATIVIZADA: AS REDES SOCIAIS COMO LUGAR DE FALA PARA SUJEITOS QUE CONVIVEM COM O AUTISMO	
Igor Lucas Ries	
DOI 10.22533/at.ed.4461922055	
CAPÍTULO 6	74
DISCURSO CIENTÍFICO E DISCURSO ACADÊMICO: SOBRE UM POSSÍVEL GESTO POLISSÊMICO DE LEITURA	
Bianca Queda Costa Solange Maria Leda Gallo	
DOI 10.22533/at.ed.4461922056	

CAPÍTULO 7	78
PARSER E LEITURA AUTOMATIZADA DE CURRÍCULOS DA PLATAFORMA LATTES PARA EXTRAÇÃO DE INDICADORES ACADÊMICOS E TECNOLÓGICOS	
Fernando Sarturi Prass Franklin Matheus Boijink Alexandre de Oliveira Zamberlan	
DOI 10.22533/at.ed.4461922057	
CAPÍTULO 8	96
ANOTAÇÕES SEMÂNTICAS EM REPOSITÓRIOS ACADÊMICOS:UM ESTUDO DE CASO COM O RI UFBA	
Aline Meira Rocha Lais do Nascimento Salvador Marlo Vieira dos Santos e Souza	
DOI 10.22533/at.ed.4461922058	
CAPÍTULO 9	113
CONTEÚDO AUDIOVISUAL DO CURSO DE PEDAGOGIA SEMIPRESENCIAL DA UNESP/UNIVESP	
Dayra Émile Guedes Martínez José Luís Bizelli	
DOI 10.22533/at.ed.4461922059	
CAPÍTULO 10	120
EDUCAÇÃO A DISTÂNCIA: APRENDIZAGEM EM REDE	
Daiane de Lourdes Alves Ângela Cutolo	
DOI 10.22533/at.ed.44619220510	
CAPÍTULO 11	132
DESAFIOS DA TUTORIA EM EAD E ESTRATÉGIAS DE MEDIAÇÃO PEDAGÓGICA: UM ESTUDO DE CASO	
Tamara de Lima Lorayne de Freitas Santos	
DOI 10.22533/at.ed.44619220511	
CAPÍTULO 12	143
CONSTRUÇÃO COLABORATIVA DE CONHECIMENTO – VIVENCIANDO EXPERIÊNCIAS COM A METODOLOGIA ATIVA	
Reyla Rodrigues Ribeiro Levy Silva Ribeiro Bruno Bernardes de Menezes Raquel Aparecida Souza	
DOI 10.22533/at.ed.44619220512	

CAPÍTULO 13	154
MATHQUIZ: UM JOGO EDUCATIVO PARA DISPOSITIVOS MÓVEIS	
José Marcelo Silva Santiago Monck Charles Nunes De Albuquerque Francisco Ranulfo Freitas Martins Junior Fernanda Kécia De Almeida Yuri Soares De Oliveira	
DOI 10.22533/at.ed.44619220513	
CAPÍTULO 14	165
A MÍDIA COMO VERTENTE INTERDISCIPLINAR DA EDUCAÇÃO INCLUSIVA DO ADOLESCENTE EM LIBERDADE ASSISTIDA	
Sebastião Jacinto dos Santos João Clemente de Souza Neto Marcos Júlio Sergi	
DOI 10.22533/at.ed.44619220514	
CAPÍTULO 15	180
EDUCAÇÃO VISUAL: DESENVOLVIMENTO GRÁFICO DE FASCÍCULOS COM CONTEÚDO DIDÁTICO	
Caroline de Cerqueira Medeiros Fabiola Arantes de Moraes	
DOI 10.22533/at.ed.44619220515	
CAPÍTULO 16	194
CULTURA VISUAL E IDENTIDADE DOS ALUNOS DO CAP-UERJ	
Christiane de Faria Pereira Arcuri	
DOI 10.22533/at.ed.44619220516	
CAPÍTULO 17	205
JUVENTUDES INTERIORANAS: ESTUDANTES DE PUBLICIDADE E SUAS MANEIRAS DE COMUNICAR	
Renata Valeria Calixto de Toledo	
DOI 10.22533/at.ed.44619220517	
CAPÍTULO 18	215
FARTURA TRAZ ALEGRIA! O FUNK OSTENTAÇÃO E AS SUBJETIVIDADES JOVENS	
Juliana Ribeiro de Vargas	
DOI 10.22533/at.ed.44619220518	
CAPÍTULO 19	227
REPRESENTATIVIDADE E GÊNERO NAS PRODUÇÕES MUDIÁTICAS: DILEMAS E APROXIMAÇÕES	
Ariana Grzegozeski Schneider Márcio Giusti Trevisol	
DOI 10.22533/at.ed.44619220519	
CAPÍTULO 20	238
A AUTOACEITAÇÃO DA HOMOSSEXUALIDADE A PARTIR DE UM CASO REAL	
Bruno Filipe Griebeler	
DOI 10.22533/at.ed.44619220520	

CAPÍTULO 21	254
A PERFORMANCE ENQUANTO FLUXO DE COMUNICAÇÃO NA MODA	
Antonio Cimadevila Ione Maria Bentz	
DOI 10.22533/at.ed.44619220521	
CAPÍTULO 22	266
A MIDDLEWARE PERSPECTIVE FOR INTEGRATING GINGA-NCL APPLICATIONS WITH THE INTERNET OF THINGS	
Danne Makleyston Gomes Pereira Francisco José da Silva e Silva Carlos de Salles Soares Neto Álan Lívio Vasconcelos Guedes	
DOI 10.22533/at.ed.44619220522	
CAPÍTULO 23	280
UMA ABORDAGEM PARA O DESENVOLVIMENTO E ANÁLISE DE DESEMPENHO DO RECONHECIMENTO OFF-LINE DE VOZ CONTÍNUO	
Lucas Debatin Aluizio Haendchen Filho Rudimar Luís Scaranto Dazzi	
DOI 10.22533/at.ed.44619220523	
CAPÍTULO 24	297
INVESTIGAÇÃO ONTOLÓGICA DA OBRA DE ARTE DIGITAL: LINGUAGEM UBÍQUA, MODELO DE DOMÍNIO E PROGRAMAÇÃO VOLTADA PARA AS ARTES VISUAIS	
Teófilo Augusto da Silva Claudio de Castro Coutinho Filho Carlos Tiago Machel da Silva	
DOI 10.22533/at.ed.44619220524	
CAPÍTULO 25	306
A INFLUÊNCIA DA TRIDIMENSIONALIDADE NA NARRATIVA ANIMADA: <i>FROZEN</i> E O USO DA ESTEREOSCOPIA	
Paula Poiet Sampedro Danilo César Granatto Leonardo Antonio de Andrade Antonio Henrique Garcia Vieira Carolina Lourenço Reimberg de Andrade Felipe Contartesi	
DOI 10.22533/at.ed.44619220525	
CAPÍTULO 26	317
UMA NARRATIVA PROCEDURAL DENTRO DO UNIVERSO FICCIONAL DA DC COMICS	
Leonardo Antonio de Andrade Felipe Contartesi Antonio Henrique Garcia Vieira Carolina Lourenço Reimberg de Andrade Paula Poiet Sampedro Danilo César Granatto	
DOI 10.22533/at.ed.44619220526	

CAPÍTULO 27	332
FINAL FANTASY XV: A NOVA APOSTA MULTIPLATAFORMA DA FRANQUIA	
Maria Tereza Batista Borges	
Mirna Tonus	
DOI 10.22533/at.ed.44619220527	
CAPÍTULO 28	339
PROCESSOS DE SUBJETIVAÇÃO EM JOGOS VIRTUAIS: UM ESTUDO SOBRE CORPO E ESTRATÉGIA NO JOGO <i>LEAGUE OF LEGENDS</i>	
Cíntia Oliveira Demaria	
Márcia Stengel	
Valéria Freire de Andrade	
DOI 10.22533/at.ed.44619220528	
CAPÍTULO 29	352
GAMEPÓLITAN: UMA ANÁLISE DAS OPORTUNIDADES DE COMUNICAÇÃO, UTILIZANDO-SE DO E-SPORT COMO FERRAMENTA DE ENGAJAMENTO	
Luana Britto Silva Vieira	
Marta Cardoso de Andrade	
DOI 10.22533/at.ed.44619220529	
CAPÍTULO 30	368
MÍDIAS DIGITAIS E O SITE DO COMITÊ OLÍMPICO DO BRASIL	
Carlos Augusto Tavares Junior	
DOI 10.22533/at.ed.44619220530	
CAPÍTULO 31	410
HOMOGENEIDADE E ENDOGENIA NOS INTERESSES DE JORNALISTAS DESCONECTAM VALOR NOTÍCIA E POPULAÇÃO	
Ana Maria Brambilla	
DOI 10.22533/at.ed.44619220531	
CAPÍTULO 32	425
O ENQUADRAMENTO DO <i>IMPEACHMENT</i> DA PRESIDENTE DILMA ROUSSEFF (PT) NAS REVISTAS <i>VEJA</i> E <i>CARTA CAPITAL</i>	
Carla Montuori Fernandes	
Eduardo Matidios Pereira	
DOI 10.22533/at.ed.44619220532	
CAPÍTULO 33	437
PARTICIPAÇÃO E MÍDIA: UM DEBATE SOBRE A HEGEMONIA DISCURSIVA DO CAPITALISMO	
Michele Luciane Blind de Moraes	
Tulainy Parisotto	
DOI 10.22533/at.ed.44619220533	
CAPÍTULO 34	449
REPRESENTAÇÕES SOBRE A AMAZÔNIA BRASILEIRA: UM ESTUDO SOBRE O DOCUMENTÁRIO <i>O ACRE EXISTE</i>	
Daya de Kassia Pinheiro Campos	
Francielle Maria Modesto Mendes	
DOI 10.22533/at.ed.44619220534	

CAPÍTULO 35 459

PARÂMETROS DE PRODUÇÃO DE CONTEÚDO RADIOFÔNICO SOBRE SAÚDE PARA CRIANÇAS DE SEIS A DEZ ANOS

Diana Diniz de Jesus

Daniela Pereira Bochembuzo

DOI 10.22533/at.ed.44619220535

CAPÍTULO 36 473

SOCIEDADE CIVIL ATIVA NA MEDIAÇÃO DAS RELAÇÕES DO MERCADO PUBLICITÁRIO COM O PÚBLICO INFANTIL

Marcos José Zablonky

Natally Navarro Encinas Ferreira

DOI 10.22533/at.ed.44619220536

SOBRE A ORGANIZADORA..... 490

ANOTAÇÕES SEMÂNTICAS EM REPOSITÓRIOS ACADÊMICOS: UM ESTUDO DE CASO COM O RI UFBA

Aline Meira Rocha

Universidade Federal da Bahia (UFBA)
Salvador - Bahia

Lais do Nascimento Salvador

Universidade Federal da Bahia (UFBA)
Salvador - Bahia

Marlo Vieira dos Santos e Souza

Universidade Federal da Bahia (UFBA)
Salvador - Bahia

RESUMO: A Web atual é sintática o que dificulta o seu processamento por máquinas e agentes de software. A Web Semântica com suas tecnologias podem dar significado para o conteúdo, além de facilitar seu processamento por agentes de software. O RI UFBA está na Web Sintática e a recuperação dos itens depositados nem sempre é feita de maneira eficiente. As informações sobre os itens depositados são armazenadas em seus metadados, mas a validação destes é feita de maneira manual e nem sempre os termos escolhidos ajudam nessa descrição. A anotação semântica de forma semi-automática certamente ajudaria aos bibliotecários na validação dos depósitos das publicações, que pode otimizar a recuperação dos itens no RI UFBA. Além disso, um experimento com a classificação das publicações também será realizada. A proposta principal é fornecer uma

ferramenta que auxilie aos bibliotecários na validação dos metadados para sugerir anotar os itens do RI UFBA tornando-o um repositório semântico.

PALAVRAS-CHAVE: web semântica, processamento de linguagem natural, repositórios institucionais, classificação de documentos, aprendizado de máquina

ABSTRACT: The current Web is syntactic and it difficult the processing by machines and software agents. The Semantic Web with its technologies can bring more meaning to web content, as well to facilitate its processing by software agents. UFBA's IR is on the Syntactic Web and the recovery of deposited items isn't always done efficiently. Information about the deposited items is stored in their metadata, but their validation is done manually and, not always, the chosen terms are able to help with the description process. The automatic semantic annotation method would certainly help librarians during the validation of the publication's deposit that can enable the recovery of UFBA's IR items. In addition, an experiment with the document classification of the publications will also be. The main proposal is to provide a tool to assist librarians in validation metadata that will be suggested to annotate the UFBA's IR items making it a semantic repository.

KEYWORDS: semantic web, natural language

1 | INTRODUÇÃO

A Web Sintática não permite inferir conhecimento pois seu foco não é no processamento das informações e sim em sua apresentação. Com isso a Web Semântica, uma extensão da Web atual, surge como alternativa para atribuir semântica às informações, permitindo que tanto pessoas quanto máquinas possam processá-las, trabalhando assim em cooperação (BERNERS-LEE, 2001). Para dar significado aos dados, a Web Semântica faz uso de ontologias.

As ontologias permitem que informações representadas por um mesmo conceito mas que utilizem diferentes identificadores sejam relacionadas, interligando esses termos entre si (BERNERS-LEE, 2001).

É possível extrair termos de documentos depositados em um repositório acadêmico, associá-los a uma ontologia e com isso gerar a anotação semântica desses documentos. Para isso, podem ser utilizadas diversas técnicas da Inteligência Artificial, incluindo o PLN (Processamento de Linguagem Natural - Natural Language Processing) e o Aprendizado de Máquina (Machine Learning).

Neste artigo são apresentados os resultados parciais e os próximos passos de um projeto de dissertação de mestrado, com o objetivo de desenvolver uma solução para anotar semanticamente de forma semi-automática os itens de um RI (Repositório Institucional).

Esse artigo está dividido como segue: na seção 2 é abordada a fundamentação teórica e são discutidos os trabalhos relacionados. A seção 3 trata da caracterização do problema. Enquanto a seção 4 aborda as contribuições esperadas. Na seção 5 podem ser vistos a metodologia e o estado atual do trabalho. Já a seção 6 contém a descrição e avaliação dos resultados obtidos até o momento. Por fim, na seção 7 tem-se a conclusão e próximos passos.

2 | FUNDAMENTAÇÃO TEÓRICA E TRABALHOS RELACIONADOS

2.1 Web Semântica

A Web Semântica, uma extensão da Web atual, permite adicionar significado ao conteúdo, possibilitando que tanto pessoas quanto máquinas possam interpretá-los (BERNERS-LEE, 2001). Para isso, a Web Semântica faz uso de ontologias.

O termo ontologia tem origem na Filosofia, tendo como base uma teoria sobre a natureza da existência (GUARINO, 2009). Segundo (GRUBER, 1993), uma ontologia é definida como “uma explícita especificação de uma conceitualização”. (GUARINO, 2009) afirma que o que existe é aquilo que pode ser representado, onde a ontologia é um mundo a ser modelado. Para representar as ontologias é utilizada a linguagem

OWL (Web Ontology Language), que atualmente está em sua versão OWL 2 (W3C, 2012).

A Web Semântica, através da Web de Dados, estabelece links entre dados de diversas fontes. A Web de Dados trata tudo como um recurso (pessoas, coisas, páginas web), que são identificados através de uma URI (Uniform Resource Identifier). Com isso, conteúdos que possuam relação, são ligados por tuplas RDF (Resource Description Framework) formando um imenso banco de dados global (BIZER, 2009).

Uma tripla RDF consiste em um conjunto de triplas no formato sujeito - predicado - objeto, conforme ilustra a figura a seguir. Assim, é possível fazer uma afirmação do tipo Item - pertence a - Comunidade.

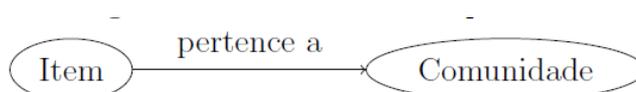


Figura 1.1: Exemplo Tripla RDF.

Fonte: autoria própria

As Anotações Semânticas possibilitam que termos extraídos de um documento quando associados a uma ontologia (ou a conceitos dela) descrevam melhor esse documento (OREN, 2006) fazendo com seus metadados sejam enriquecidos semanticamente, facilitando a sua recuperação.

Essas Anotações podem ser de 4 (quatro) tipos: (1) manual (quando a anotação é realizada manualmente); (2) semi-automática (os termos são extraídos de forma automática e depois são sugeridos a um especialista para serem anotados); (3) automática (os termos são extraídos e anotados de forma automática) e (4) híbrida (quando ocorre mais de um tipo de anotação ao mesmo tempo). Essas Anotações podem ser tanto intrusivas (quando é feita no próprio documento) quanto não-intrusivas (quando é feita num arquivo à parte).

2.2 Técnicas Utilizadas

Para extrair os termos para anotação podem ser utilizadas diversas técnicas da Inteligência Artificial, incluindo o PLN (Processamento de Linguagem Natural - Natural Language Processing) e o Aprendizado de Máquina (Machine Learning).

O PLN realiza o processamento textual de forma que computador possa entender a linguagem humana. Com isso, constrói sistemas que foquem em processamento textual de modo que o computador possa entender textos escritos em linguagem humana e se comunique de forma automática.

O primeiro passo no pré-processamento de um texto é realizar a tokenização que consiste em dividir um texto em pedaços menores, ou seja, em palavras. Por exemplo, a sentença “Estou bem mas não tenho certeza se vou conseguir viajar amanhã”, quando tokenizada, pode ser representada por: ‘Estou’, ‘bem’, ‘mas’, ‘não’, ‘tenho’,

‘certeza’, ‘se’, ‘vou’, ‘conseguir’, ‘viajar’, ‘amanhã’.

Em seguida, pode ser realizada a normalização que envolve tarefas como correção de acentuação, eliminação de caracteres de pontuação, eliminação de caracteres numéricos e remoção de palavras com poucos caracteres.

No tratamento do texto, também é realizada a remoção de stopwords, que são palavras que não são relevantes em um texto e que por não possuírem valor semântico, podem ser descartadas. As stopwords são específicas para cada idioma.

A Bag of Words (BOW) ou “saco de palavras”, representam a matriz de frequência das palavras em um conjunto de textos. Como os algoritmos de aprendizado de máquina não trabalham com informações textuais, é necessário que essas informações sejam convertidos em uma representação numérica. Para a representação da Bag of Words podem ser utilizados valores booleanos com o objetivo de marcar a ocorrência ou não das palavras, onde a presença é representada pelo valor 1 e a ausência representada por 0 ou pode ser informada a quantidade de vezes que as palavras mais frequentes de um texto aparecem nele.

	termo1	termo2	...	termoN
documento1	1	1		0
documento2	0	1		1
...				
documentoN	0	1		0

Tabela 1.1: Bag of Words.

Fonte: autoria própria

Na Tabela 1.1, é ilustrada uma Bag of Words, onde as linhas representam os textos (ou documentos) e as colunas representam as palavras extraídas desses textos.

O Aprendizado de Máquina permite que um sistema aprenda através de um conjunto de dados treinado anteriormente. Assim, um conjunto de dados pode ser treinado e aprender a fazer uma determinada tarefa, como classificar documentos, análise de spams, prever tendências, etc.

No Aprendizado Supervisionado existe uma base de conhecimento anterior. Onde esse modelo pode ser utilizado para prever a classe de dados de conjunto de testes futuros.

Dentre as técnicas existentes, podemos citar a classificação que é uma tarefa preditiva que avalia a probabilidade de eventos futuros. Para isso leva em consideração alguns padrões para classificar algo, como por exemplo, se um e-mail é spam ou não.

2.3 Repositórios Institucionais

Por sua vez, um RI serve para armazenar as produções científicas e acadêmicas geradas por uma instituição, possibilitando que seus pesquisadores possam divulgar as pesquisas realizadas (GOMES, 2010) e aumentar sua visibilidade (LEITE, 2012).

Isso é importante pois é necessário que seja feita uma divulgação dos resultados dessas pesquisas, como retorno aos investimentos feitos em ciência (SAYÃO, 2009).

Atualmente, existem diversos softwares utilizados como solução de armazenamento para RI, sendo a mais utilizado pelas instituições de ensino e pesquisa em todo o mundo, a ferramenta DSpace, que foi desenvolvida em uma parceria entre o MIT (Massachusetts Institute of Technology) e a HP (Hewlett-Packard) [18].

O DSpace se organiza em comunidades, sub-comunidades e coleções, seguindo a estrutura organizacional das instituições. Cada comunidade pode possuir sub-comunidades ou não e inúmeras coleções que são compostas por itens que permitem armazenar arquivos em diversos formatos (SHINTAKU, 2010).

Com o DSpace é possível realizar o auto-arquivamento, assim o próprio pesquisador pode realizar o depósito de sua publicação enquanto um bibliotecário fica responsável pela validação dos metadados informados (ROSA, 2011). Só depois dessa validação é que cada publicação fica disponível para os usuários finais.

O DSpace utiliza o padrão de metadados (dados sobre os dados) Dublin Core (DC) (SHINTAKU, 2010) que é composto por 15 (quinze) elementos (HILDEMANN, 2008), conforme pode ser visto na tabela 1.2:

Elemento Descritivo	Descrição do Valor
Title	Nome dado ao recurso
Creator	Entidade responsável por fazer o conteúdo do recurso.
Subject	Tema do conteúdo do recurso.
Description	Relato do conteúdo do recurso.
Publisher	Entidade responsável por tornar o recurso disponível.
Contributor	Entidade responsável por qualquer contribuição para o conteúdo do recurso.
Date	Data associada a um evento no ciclo de vida do recurso.
Type	Natureza ou gênero do conteúdo do recurso.
Format	Expressão física ou digital do recurso.
Identifier	Referência ambígua ao recurso dentro de um determinado contexto.
Source	Referência a um recurso do qual o presente recurso é derivado.
Language	Língua do conteúdo intelectual do recurso.
Relation	Referência a um recurso relacionado.
Coverage	Extensão ou escopo do conteúdo do recurso.
Rights	Informações sobre os direitos detidos em e sobre o recurso.

Tabela 1.2: Esquema padrão Dublin Core.

Fonte: (HILDEMANN, 2008), traduzido

Os metadados servem para identificar e descrever um determinado documento, tornando mais fácil sua recuperação. A depender do tipo de documento armazenado pode-se utilizar um conjunto de metadados específico (LEITE et al., 2012).

2.4 Trabalhos Relacionados

Diversos trabalhos já foram publicados na área de Web Semântica com enfoque em anotação semântica e extração de keywords. A Tabela 2.2 apresenta uma síntese dos trabalhos relacionados, ressaltando suas propostas.

Trabalho	Proposta
Automatic annotation of tagged content using predefined semantic concepts (Manzato; Goularte, 2012)	Anotação semântica automática de conteúdo multimídia
Exposing scholarly information as linked open data: RDFizing DSpace contents (Konstantinou, 2014)	Anotação semântica em repositório <i>DSpace</i>
Indexação semi-automática de publicações através de técnicas de mineração de texto (2016)	Mineração de texto com extração de termos relevantes para auxiliar na indexação de documentos
Extração Automática de Termos Candidatos às Ontologias: um Estudo de Caso no Domínio da Hemoterapia (Mendonça, 2012)	Extração automática de termos candidatos às ontologias na área médica
Extração Automática de Palavras-Chaves de Texto da Língua Portuguesa (Dias, 2005)	Implementação em língua portuguesa do algoritmo KEA utilizado para extração de palavras-chaves
Descoberta Automática de Palavras-chave Para Classificação de Textos (Leite, 2016)	Avalia uma técnica que atribui um peso nas palavras-chaves de acordo com sua frequência no texto.
Aplicação de Mineração de Textos na Indicação de Palavras-chave em Artigos Científicos (Cassenote, 2016)	Analisa a eficiência da adaptação do algoritmo KEA implementada por (Dias, 2005)

Tabela 2.2: Síntese dos trabalhos relacionados

Fonte: autoria própria

(MANZATO; GOULARTE, 2012) apresentam uma proposta de uma técnica automática de anotação semântica para conteúdos multimídia. As limitações das técnicas automáticas apresentam uma alta dependência de inferências de bases de conhecimento e as abordagens manuais consomem tempo e apresentam dados incompletos. Diversos serviços possibilitam que os usuários possam adicionar tags para anotar o conteúdo produzido. Com isso, foi proposto um anotador onde é recebido um conjunto de tags que são filtradas gerando uma folksonomia (indexação de informação) de tags. Essas tags foram enriquecidas para posterior anotação seguindo 3 estratégias apresentadas no artigo. Por fim, o conteúdo foi anotado com os conceitos obtidos dessas estratégias.

O trabalho (KONSTANTINO et al., 2014) descreve como um repositório no Dspace pode ser anotado semanticamente como um repositório semântico. Baseado no modelo de dados do DSpace, foi feito um mapeamento dos metadados dos conteúdos desse repositório (mapeamento R2RML), gerando um grafo RDF. Considera somente conteúdo armazenado em banco de dados.

Já o artigo (MENDONCA et al., 2012) aborda a extração automática de informações na área médica da hemoterapia utilizando ferramenta de PLN. Adotou-se para a extração de termos candidatos a ferramenta Scketch Engine, a construção de um corpus no domínio de sangue com a anotação POS (Part-of-Speech) Tagger (onde cada termo é etiquetado de acordo com seu uso na frase) e o cálculo das

frequências dos termos candidatos à ontologia. Considerou-se somente os prefixos dos termos pois normalmente estes representam sua semântica. Após esse cálculo, os termos foram agrupados de acordo com a classe semântica correspondente. Em seguida, foi gerada a taxonomia dessas classes semânticas e foi realizada a validação da ontologia por um especialista na área.

O trabalho (DIAS, 2005) ressaltam que as palavras-chaves servem para filtrar conteúdo. Foi implementada uma nova versão do algoritmo KEA para extração de palavras-chaves em língua portuguesa. Uma nova lista de stopwords também foi gerada. Para avaliação do algoritmo foi realizado um experimento com dissertações e teses de diversas áreas, comparando documentos inalterados e documentos com as palavras-chaves extraídas e obteve-se um resultado compatível com os obtidos na aplicação original com a língua inglesa.

O artigo (CASSENTE, 2016) analisa a eficiência do algoritmo KEA. Para tanto foram utilizados artigos científicos das mais diversas áreas de conhecimento, descartando seus resumos e palavras-chaves a fim de obter as palavras-chaves com a aplicação do algoritmo. Para o experimento foi utilizada a adaptação do algoritmo KEA para a língua portuguesa implementada por (Dias, 2005). Sendo realizado um comparativo entre as palavras-chaves indicadas pelos autores e as extraídas pelo algoritmo.

O trabalho (LEITE, 2016) realça a importância das palavras-chaves, porém ressalta que atribuí-las manualmente fica inviável. Com isso, são avaliadas várias técnicas de extração de palavras-chaves e um experimento utilizando a base de dados do site www.preadly.com. Com o auxílio da linguagem *Alchemy*, que possui uma API Java para PLN, através do método *Keyword Extraction*, que atribui um peso de acordo com a frequência das palavras-chaves encontradas no texto.

Esse trabalho diferencia-se por fornecer uma ferramenta que auxilie aos bibliotecários na validação dos metadados de publicações armazenadas em um RI e classificar os documentos nas comunidades/coleções de acordo com seu conteúdo.

3 | CARACTERIZAÇÃO DO PROBLEMA

Repositórios acadêmicos possibilitam o armazenamento das produções científicas de universidades e centros de pesquisa. Eles permitem o auto-arquivamento mas por mais que os metadados de cada publicação sejam preenchidos pelo próprio pesquisador, nem sempre os termos escolhidos para descrevê-la são os mais relevantes e adequados. Além disso, a falta de conhecimento multidisciplinar por parte dos bibliotecários, impacta na validação precisa desses metadados.

Extrair termos relevantes desses conteúdos pode ajudar a descrever melhor esses documentos, contribuindo assim para sua recuperação na busca. O ideal seria que esses termos fossem sugeridos durante a validação dos metadados por parte dos bibliotecários, realizando assim a anotação semântica de maneira semi-automática, já que os termos seriam extraídos dos documentos de forma automática e depois

passariam pela validação de um especialista. .

4 | CONTRIBUIÇÕES ESPERADAS

Nessa proposta espera-se fornecer uma ferramenta que auxilie aos bibliotecários na validação dos metadados dos itens de um RI. A anotação semântica semi-automática auxiliaria o trabalho dos bibliotecários durante a validação dos depósitos das produções e ajudaria na recuperação das informações, otimizando a busca de itens pesquisados no RI.

Para isso, serão sugeridos aos bibliotecários termos relevantes extraídos de textos em formato pdf (Portable Document Format) e a classificação dos documentos nas comunidades/coleções de acordo com seu conteúdo e metadados.

O modelo proposto se baseia nos dados de entrada, no caso arquivos em formato pdf de itens do RI, inicia-se o processo, conforme mostra a Figura 2.1, realizando a extração de texto através de técnicas de Processamento de Linguagem Natural, onde será realizada a extração de tokens relevantes.

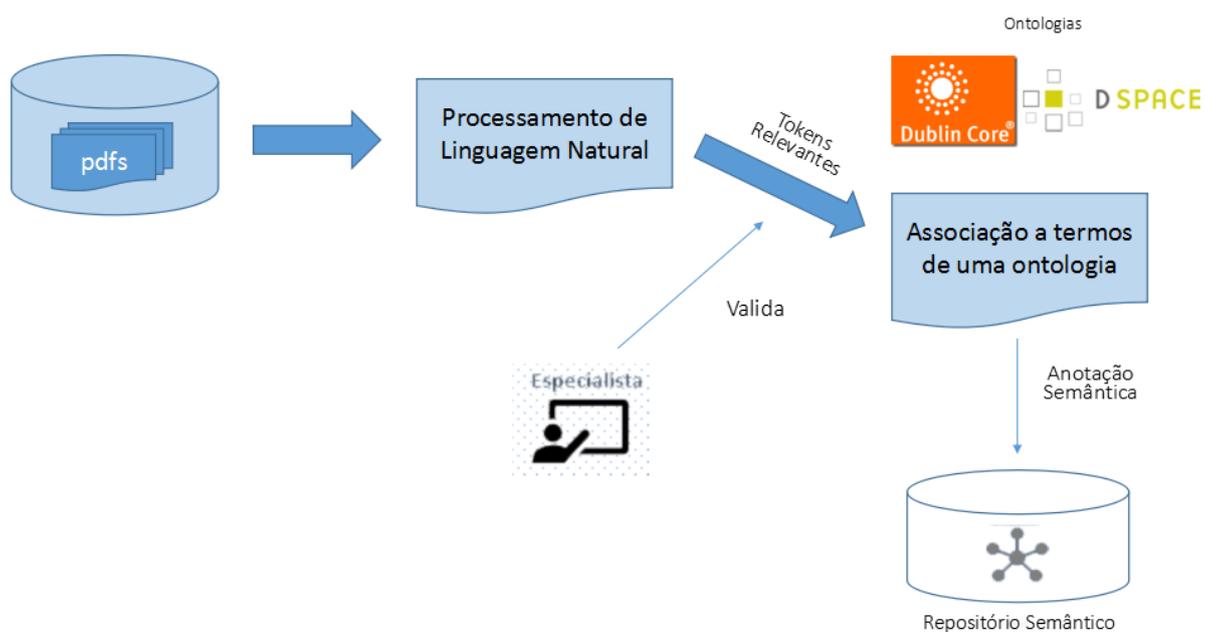


Figura 4.1 - Modelo Proposto.

Fonte: autoria própria

Em seguida, ocorre a extração de keywords, com a identificação de tokens relevantes que foram extraídos dos documentos. Durante a validação dos metadados serão sugeridos ao especialista, no caso o bibliotecário, os tokens relevantes, de modo que estes sejam associados a termos das ontologias DSpace e Dublin Core que servirá para realizar a anotação semântica dos itens no repositório, gerando por fim um repositório semântico.

Ocorrerá também a classificação textual que permitirá classificar melhor os itens depositados na comunidade/coleção correta, levando-se em consideração que um

pesquisador pode ter mais de um vínculo com a instituição e acontecer dele informar erroneamente a comunidade ou coleção onde o item deve ser depositado.

A classificação textual, como pode ser conferida na Figura 4.2, receberá a matriz Bag of Words que representa a frequência de palavras em um conjunto de textos. A Bag of Words e vetor de metadados, no caso do RI, o vetor de comunidades e coleções transformada em dados numéricos já que os algoritmos de Aprendizado de Máquina não trabalham com texto serão utilizados para treinar classificadores implementados com os algoritmos Naive Bayes, SVM e C4.5 para realizar a classificação textual. Por fim, serão realizadas predições e os classificadores implementados serão avaliados pelas métricas escolhidas.

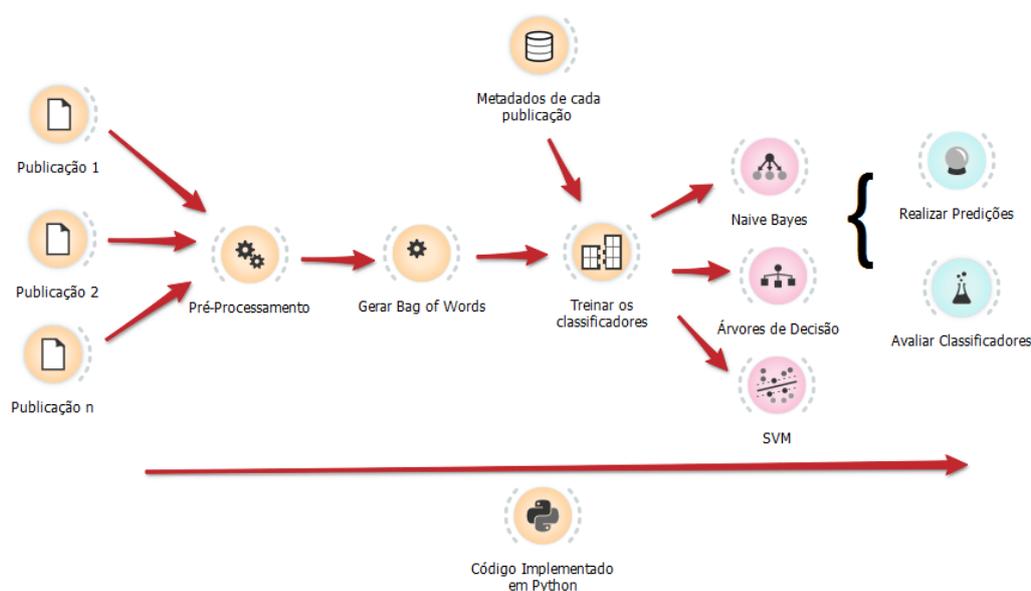


Figura 4.2 - Protótipo da Classificação Textual.

Fonte: autoria própria

Para atingir o objetivo geral deste projeto de mestrado, foram definidos os seguintes objetivos específicos: (1) classificar os documentos nas comunidades/coleções a partir de termos relevantes extraídos; (2) avaliar os métodos de anotação semântica; (3) anotar os itens de um RI utilizando as ontologias Dspace e Dublin Core e (4) realizar um estudo de caso com o RI UFBA.

5 | ESTADO ATUAL DO TRABALHO

5.1 Metodologia

Para atingir os objetivos propostos deste trabalho, foram definidos os seguintes passos metodológicos: (1) realização de levantamento bibliográfico; (2) estudo da estrutura de um repositório acadêmico; (3) pesquisa das ontologias Dspace e Dublin Core; (4) reuniões com equipes envolvidas com a gestão do RI UFBA; (5) montagem da base de dados para realização dos experimentos; (6) avaliação dos classificadores;

(7) realização de um experimento com o pessoal do SIBI/UFBA; (8) geração do repositório semântico; (9) análise e registro dos resultados obtidos.

5.2 Atividades em Andamento

Considerando-se os passos metodológicos informados na seção anterior, até o momento já foram realizados os seguintes passos (1), (2), (3), (4) e (5). Ressalta-se que os passos (1) e (4) correspondem a atividades iterativas a serem realizadas no decorrer desse trabalho.

Como experimento inicial, realizou-se a extração de textos em um conjunto de arquivos pdf, escolhidos aleatoriamente do RI UFBA, com o auxílio da linguagem de programação Python e uma de suas bibliotecas para extração de texto, a pdfminer, em sua versão .six. Essa biblioteca permite extrair textos de arquivos em formato pdf e os converte para txt (arquivo de texto). Nos arquivos extraídos, notou-se que as imagens foram desconsideradas e que os textos apresentados em tabelas foram conservados, porém perdeu-se sua formatação.

Para realizar o processamento destes textos, foi utilizada a biblioteca NLTK (Natural Language Toolkit) também codificada em Python. Essa biblioteca permite executar as tarefas de tokenização (quebrar o texto em tokens, ou seja, em palavras) de sentenças, extração de prefixos ou sufixos, etiquetagem (definição da classe gramatical de uma uma palavra), identificação de entidades nomeadas (como pessoas, empresas ou instituições), remoção de stopwords (palavras pouco relevantes no texto), dentre outras tarefas.

Também foram corrigidas possíveis inconsistências, como corrigir problemas de acentuação, remoção de caracteres numéricos e de pontuação, remoção de tokens com menos de 3 caracteres.

6 | DESCRIÇÃO E AVALIAÇÃO DOS RESULTADOS

Depois de identificar o problema e definir os objetivos deste trabalho, buscou-se mapear os passos metodológicos a serem seguidos. Com isso, para seguir o cronograma estipulado, algumas atividades já se encontram concluídas.

6.1 Realização de levantamento bibliográfico

Realizou-se uma pesquisa em diversas bases de dados por trabalhos desenvolvidos na área de pesquisa. A princípio, foi feita uma pesquisa por trabalhos na área de Ontologias e Anotações Semânticas que tenham como objeto de estudo, os Repositórios Institucionais (ou Acadêmicos). Também foram pesquisados trabalhos com a extração de texto utilizando técnicas de PLN.

Estes trabalhos servem de apoio para a fundamentação teórica deste projeto, bem como para verificar o que está sendo pesquisado e desenvolvido nessas áreas,

além de identificar as lacunas de pesquisa.

6.2 Estudo da estrutura de um repositório acadêmico

Estudou-se a estrutura de um repositório institucional acadêmico, como ele é organizado, mapeando suas principais características e funcionalidades. Verificou-se como foi feita a instalação do RI UFBA, a fim de verificar se foi feita alguma alteração na instalação padrão, como customizações, adição de plugins, etc. Foram verificadas quais as configurações do ambiente onde o RI UFBA se encontra instalado atualmente.

6.2.1 A instalação do RI UFBA.

A versão do DSpace instalada atualmente no RI UFBA é a 3.2, hospedada num servidor Linux Debian, com banco de dados PostgreSQL 8.3, Apache 2.2.16, Tomcat 7.0.47, Java versão JDK (Java SE Development Kit) 1.5, com pacotes Java-Ant e Maven 2 instalados, utilizando as estatísticas da Universidade do Minho.

A UFBA (Universidade Federal da Bahia) armazena uma parte de sua produção científica e acadêmica em seu repositório institucional, o RI UFBA. O RI UFBA foi implantado pela EDUFBA (Editora da UFBA) em 2010 e pode ser acessado no endereço <https://repositorio.ufba.br> (ROSA, 2011).

Suas comunidades, sub-comunidades e coleções estão organizadas de acordo com a estrutura organizacional da UFBA e os depósitos são realizados através do auto-arquivamento, tendo os metadados validados pelos bibliotecários ligados à comunidade onde a publicação foi depositada.

6.3 Pesquisa das ontologias DSpace e Dublin Core

Realizou-se uma pesquisa com a finalidade de verificar se já existiam ontologias publicadas sobre o DSpace e o padrão de metadados Dublin Core. Essa busca incluiu desde repositórios de ontologias a trabalhos publicados sobre o tema.

As ontologias encontradas foram criadas pelo grupo de pesquisa Semantic Web and Interoperability da Universidade de Patras [9] [10]. A Figura 6.1 ilustra a taxonomia da ontologia DSpace encontrada:

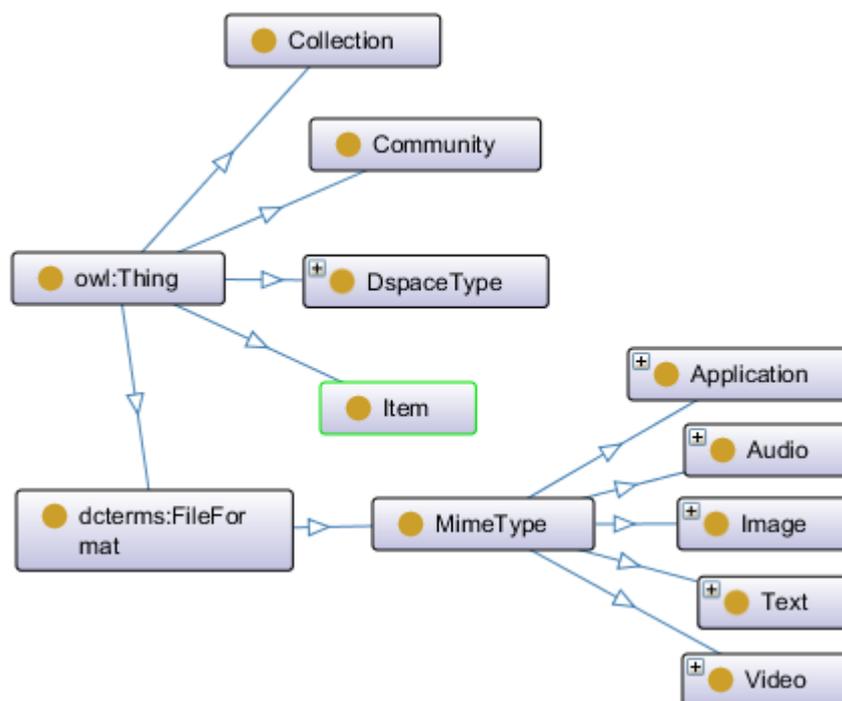


Figura 6.1: Ontologia DSpace.

Fonte: extraída do Protégé

Conforme ilustrado na Figura 6.1, as seguintes classes foram identificadas:

- Collection: representando as coleções
- Community: representando as comunidades, refletindo a divisão hierárquica existente dentro das instituições.
- DspaceType: representando os tipos de documentos que podem ser depositados
- dcterms:FileFormat (com a sub-classe MimeType): representando os formatos dos arquivos suportados. A sub-classe MimeType possui uma partição de 5 (cinco) sub-classes: Application, Audio, Image, Text e Video, referindo-se aos tipos de arquivos que podem ser depositados
- Item: representando as publicações depositadas

As propriedades existentes entre essas classes são descritas com maiores detalhes na Tabela 6.1 abaixo:

Classes	Propriedades
Collection	(not (Item)) and (dcterms:hasPart some Item) Sub-ClassOf Collection / disjoint with item
Community	disjoint with item
MimeType	SubclassOf of dcTerms:FileFormat has subclasses Application, Audio, Image, Text and Video
Item	(not (Item)) and (dcterms:hasPart some Item) Sub-ClassOf Collection / disjoint with Community and Collection

Tabela 6.1: Classes Ontologia DSpace.

Fonte: autoria própria

Por exemplo, a propriedade (not (Item)) and (dcterms:hasPart some Item) SubClassOf Collection - presente nas classes Item e Collection - define que um item é parte de uma coleção menos item dele mesmo e que coleções só podem conter itens.

Enquanto a ontologia dspace-ont.owl possui as classes, representando a estrutura de um repositório institucional, a ontologia dc-ont.owl apresenta as propriedades que ligam essas classes, conforme mostra a Tabela 6.2 a seguir:

Propriedade	Inversa
dcterms:hasFormat	dcterms:isFormatOf
dcterms:hasPart	dcterms:isPartOf
dcterms:hasVersion	dcterms:isVersionOf
dcterms:isReferencedbBy	dcterms:references
dcterms:isReplaceBy	dcterms:replaces
dcterms:isRequiredBy	dcterms:requires

Tabela 6.2: Propriedades DCTerms.

Fonte: autoria própria

Assim, um Item faz parte de uma Collection (dcterms:isPartOf) e o inverso, onde uma Collection contém um determinado Item(dcterms:hasPart). Já a Tabela 6.3, mostra a relação entre as propriedades simétricas da ontologia Dublin Core:

Propriedade	Simétrica
dcterms: relation	dcterms: relation

Tabela 6.3: Relação Simétrica - ontologia Dublin Core.

Fonte: autoria própria

Nesse caso, a propriedade dcterms: relation é simétrica a si mesma com a

finalidade de descobrir relações entre os itens. A ligação entre as duas ontologias é feita através dos dcterms ou QDC (Qualified Dublin Core), qualificadores que representam as relações existentes e servem para refinar os elementos Dublin Core.

6.4 Reuniões com equipes envolvidas com a gestão do RI UFBA

Periodicamente são realizadas reuniões com as equipes responsáveis pelo RI UFBA. A gestão do RI UFBA atualmente é realizada pelo Núcleo Científico-SIBI/UFBA (Sistema Universitário de Bibliotecas) e a CRI-STI/UFBA (Coordenação de Redes e Infraestrutura - Superintendência de Tecnologia da Informação) é responsável pela infraestrutura do ambiente que hospeda o RI UFBA.

6.5 Montagem da base de dados para realização dos experimentos

A base de dados para realização dos experimentos será composta do dump do banco de dados em PostgreSQL para a extração dos metadados e de downloads de itens em formato pdf conforme informado na seção a seguir..

6.6 Avaliação dos classificadores

Com a finalidade de classificar os documentos a partir de seus metadados foi realizada uma análise de quais metadados seriam escolhidos para a realização deste experimento. Optou-se por selecionar as informações de comunidade e coleções dos itens depositados no RI..

Deve-se considerar que as comunidades podem conter sub-comunidades, que geralmente representam as unidades acadêmicas da instituição. As sub-comunidades geralmente representam as divisões existentes em cada uma dessas unidades, como cursos, departamentos e programas de pós-graduação.

Como a estrutura comunidade + sub-comunidade segue uma estrutura hierárquica, para a implementação dos classificadores será utilizado um classificador hierárquico (hierarchical multi-label classification), com isso um item só é associado a um rótulo se estiver também associado ao rótulo-pai. Para tanto, deve ser treinado um classificador para o 1º nível e classificadores para prever o 2º nível, de cada nível. Em seguida deve-se prever os itens de 1º nível e os itens de 2º nível, de cada nível.

Já as coleções geralmente representam os tipos de publicações armazenadas no RI, como artigos, trabalhos de conclusão de curso, dissertações e teses.

Além disso, foi realizada uma análise de quais comunidades seriam utilizadas durante os experimentos. Pela natureza multidisciplinar do RI UFBA também levou-se em consideração a escolha de coleções de mais de curso e que possuíssem no mínimo 10 itens. Finalizada esta análise, chegou-se ao conjunto de documentos representado pela Tabela 6.4, ilustrada a seguir:

Comunidade	Sub-Comunidade	Coleção	Quantidade de Itens
Faculdade de Educação	Programa de Pós-Graduação em Educação (PPGE)	Dissertações de Mestrado (PPGE)	19
Faculdade de Educação	Programa de Pós-Graduação Multidisciplinar e Multiinstitucional em Difusão do Conhecimento	Teses de Doutorado (Pós Difusão do Conhecimento)	19
Instituto de Matemática e Estatística	Programa de Pós-Graduação em Ciência da Computação	Dissertações de Mestrado (PG-COMP)	11
Instituto de Matemática e Estatística	Programa de Pós-Graduação em Matemática	Dissertações de Mestrado (PG-MAT)	37
Escola Politécnica	Mestrado em Meio Ambiente, Água e Saneamento (MAASA)	Dissertações de Mestrado (MA-ASA)	22
Escola Politécnica	Programa de Pós-Graduação em Engenharia Industrial (PEI)	Dissertações de Mestrado (PEI)	30
Escola Politécnica	Programa de Pós-Graduação em Engenharia Química (PPEQ)	Dissertações de Mestrado (PPEQ)	25

Tabela 6.4: Documentos utilizados na Classificação Textual.

Fonte: Autoria própria

7 | CONCLUSÕES E PRÓXIMOS PASSOS

Esse trabalho descreveu uma proposta que tem como objetivo fornecer uma ferramenta que auxilie aos bibliotecários na validação dos metadados e auxiliar na classificação dos documentos a partir de seus metadados, realizando um Estudo de Caso com o RI UFBA.

A anotação semântica de itens de um RI pode favorecer a recuperação das informações, otimizando a busca de itens pesquisados num RI. Essa anotação sendo feita de maneira semi-automática poderá auxiliar o trabalho dos bibliotecários na validação dos metadados de cada item depositado no RI.

Como próximos passos, será finalizada a implementação dos classificadores. Também será realizada uma avaliação dos classificadores, dos métodos de anotação semântica e realização de um experimento com a anotação dos itens pela equipe de bibliotecários do SIBI.

Por fim, será feita uma análise e registro dos resultados obtidos com a geração de um repositório semântico.

REFERÊNCIAS

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. **The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities.** Scientific American, v. 284, n. 5, p. 1-5, 2001. Disponível em: http://csis.pace.edu/marchese/CS835/Lec9/112_SemWeb.pdf

BIZER, C.; HEATH, T.; BERNERS-LEE, T. **Linked data-the story so far. Semantic services, interoperability and web applications: emerging concepts**, p. 205-227, 2009. Disponível em: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

CASSENOTE, M. R. S; SECCO, A; ANTONIAZZI, R. L.; CHICON, P. M. M. **Aplicação de Mineração de Textos na Indicação de Palavras-chave em Artigos Científicos.** 2016. Universidade da Cruz Alta. Disponível em: revistaeletronica.unicruz.edu.br/index.php/computacao/article/download/4013/735

DIAS, M. A. L. MALHEIROS, M. G. **Extração Automática de Palavras-chave de Textos da Língua Portuguesa**. Centro Universitário UNIVARTES. 2005. Disponível em: <https://www.researchgate.net/publication/242564258>

GOMES, M. J.; ROSA, F. G. **Repositórios Institucionais: democratizando o acesso ao conhecimento**. Brasil: EDUFBA, 2010. Disponível em: <https://repositorio.ufba.br/ri/bitstream/ri/616/3/Repositoriosn%20institucionais.pdf>

GRUBER, T. R. et al. **A translation approach to portable ontology specifications. Knowledge acquisition**, Academic Press, v. 5, n. 2, p. 199-220, 1993. Disponível em: <http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>

GUARINO, N.; OBERLE, D.; STAAB, S. **What is an Ontology?** In: Handbook on ontologies. Springer, 2009. p. 1-17. Disponível em: http://iaoa.org/isc2012/docs/Guarino2009_What_is_an_Ontology.pdf

HILLMANN, D. **Using dublin core. Dublin Core Metadata Initiative**, 2008. Disponível em: <http://dublincore.org/documents/usageguide/>

KONSTANTINOOU, N. et al. **Exposing scholarly information as linked open data: RDFizing DSpace contents**. The Electronic Library, Emerald Group Publishing Limited, v. 32, n. 6, p. 834-851, 2014.

KOUTSOMITROPOULOS, D. A. et al. **Knowledge Management and Acquisition in Digital Repositories - A Semantic Web Perspective**. In: CITESEER. KMIS. 2009. p. 117-122. Disponível em: <http://www.hpclab.ceid.upatras.gr/viograka/kotsomit/pubs/kmis09.pdf>

KOUTSOMITROPOULOS, D. A. et al. **Semantic Metadata Interoperability and Inference-Based Querying in Digital Repositories**. Greece, 2009. Disponível em: <http://www.hpclab.ceid.upatras.gr/viograka/kotsomit/pubs/jitr09.pdf>

KOUTSOMITROPOULOS, D. A. et al. **Semantic Web enabled digital repositories**. Greece, 2010. Disponível em: http://web.simmons.edu/weigle/INTERO_METADATA_LANGU/Koutsomitropoulos_SemanticWeb.pdf

KOUTSOMITROPOULOS, D. A.; SOLOMOU, G. D.; KALOU, A. K. **Navigating among Educational Resources in the Web of Linked Data**. In: SPRINGER. International Conference on Web Reasoning and Rule Systems. 2014. p. 185-190. Disponível em: http://rd.springer.com/chapter/10.1007/978-3-319-11113-1_14

LEITE, F. et al. **RI - Repositórios Institucionais: Boas Práticas para a Construção de Repositórios Institucionais da Produção Científica**. Brasil: IBICT, 2012. Disponível em: <http://livroaberto.ibict.br/handle/1/703>

LEITE, E. M. **Descoberta Automática de Palavras-chave Para Classificação de Textos**. Trabalho final de Graduação. Universidade Federal do Estado do Rio de Janeiro - UNIRIO. 2016. Disponível em: bsi.uniriotec.br/tcc/textos/201601Leite.pdf

MANZATO, M. G.; GOULARTE, R. **Automatic annotation of tagged content using predefined semantic concepts**. In: ACM. Proceedings of the 18th Brazilian symposium on Multimedia and the web. [S.l.], 2012. p. 237-244.

MENDONÇA, F. et al. **Extração automática de termos candidatos as ontologias: um estudo de caso no domínio da hemoterapia**. ONTOBRAS-MOST, v. 938, p. 170-175, 2012.

ROSA, F.; MEIRELLES, R. F.; PALACIOS, M. **Repositório Institucional da Universidade Federal da Bahia: implantação e acompanhamento**. 2011. Disponível em: <https://repositorio.ufba.br/ri/bitstream/ri/1590/1/5603.pdf>

SAYÃO, L. et al. **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Brasil: EDUFBA, 2009. Disponível em: https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.pdf

SHINTAKU, M.; MEIRELLES, R. F. **Manual do DSpace: administração de repositórios**. EDUFBA, 2010. Disponível em: [https://repositorio.ufba.br/ri/bitstream/ri/769/1/Manualn%20don%20Dspace\(2\).pdf](https://repositorio.ufba.br/ri/bitstream/ri/769/1/Manualn%20don%20Dspace(2).pdf)

W3C. **OWL 2 Web Ontology Language**. 2012. Disponível em: <https://www.w3.org/TR/owl2-overview/>

SOBRE A ORGANIZADORA

Vanessa Cristina de Abreu Torres Hrenechen: Graduada em Comunicação Social/Jornalismo (UEPG); mestre em Crítica de Mídia (UEPG). Tem 10 anos de experiência em assessoria de imprensa.

Atualmente é proprietária de agência de publicidade que presta serviços na área de marketing e comunicação empresarial.

Agência Brasileira do ISBN
ISBN 978-85-7247-344-6

