

**Universidade Estadual Paulista
Faculdade de Filosofia e Ciências
Campus de Marília
Programa de Pós-Graduação em Ciência da Informação**

Luciana Maria Vieira Pöttker

**Arquitetura para recuperação de objetos de aprendizagem –
uma abordagem baseada em agentes inteligentes e *relevance
feedback***

**Marília – SP
2017**



Luciana Maria Vieira Pöttker

**Arquitetura para recuperação de objetos de aprendizagem –
uma abordagem baseada em agentes inteligentes e *relevance
feedback***

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação como requisito parcial para a obtenção do título de Doutora em Ciência da Informação – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (UNESP), Campus de Marília.

Área de Concentração: Informação, Tecnologia e Conhecimento

Linha de Pesquisa: Informação e Tecnologia

Orientador: Prof. Dr. Edberto Ferneda

Financiamento: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

**Marília – SP
2017**

Pöttker, Luciana Maria Vieira.

P871a Arquitetura para recuperação de objetos de aprendizagem – uma abordagem baseada em agentes inteligentes e *relevance feedback* / Luciana Maria Vieira Pöttker. – Marília, 2017.
203 f. ; 30 cm.

Orientador: Edberto Ferneda.

Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, 2017.

Bibliografia: f. 176-188

1. Recuperação da informação. 2. Objetos de aprendizagem. 3. Agentes inteligentes (Software). 4. Metadados. I. Título.

CDD 029.7

Luciana Maria Vieira Pöttker

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação como requisito parcial para a obtenção do título de Doutora em Ciência da Informação – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (UNESP), Campus de Marília.

Área de Concentração: Informação, Tecnologia e Conhecimento

Linha de Pesquisa: Informação e Tecnologia

BANCA EXAMINADORA

Dr. Edberto Ferneda – Orientador

Universidade Estadual Paulista (UNESP – Campus de Marília)

Dr. Elvis Fusco

Centro Universitário Eurípides de Marília (UNIVEM)

Dr. Walter Moreira

Universidade Estadual Paulista (UNESP - Campus de Marília)

Dra. Ana Carolina Simionato

Universidade Federal de São Carlos (UFSCar)

Dra. Rachel Cristina Vesu Alves

Universidade Estadual Paulista (UNESP - Campus de Marília)

Marília/SP, 16 de março de 2017.

Aos meus amados filhos

Christiano e Lucas

Agradecimentos

A **Deus**, por ter me dado forças e me conduzido pelo caminho certo, nos momentos mais diversos que passei no decorrer desta tese.

A minha família, especialmente aos meus filhos **Christiano e Lucas** e ao meu marido **Walmir** pelo apoio, paciência, carinho e compreensão em todos os momentos e, em especial, em minhas ausências.

Aos meus pais, **Vilmar e Guiomar**, que mesmo distantes sempre acreditaram em mim e me incentivaram a lutar pelos meus sonhos.

Ao meu orientador, **Prof. Dr. Edberto Ferneda**, pela sua amizade e por acreditar, incentivar e apoiar meu trabalho em todos os momentos dessa caminhada.

Ao **Prof. Dr. José Antonio Moreira González**, por me receber em Madrid e na Universidade Carlos III de Madrid – UC3M – Espanha, o que permitiu ao trabalho um amadurecimento epistemológico, teórico e metodológico.

Aos professores **Dr. Elvis Fusco e Dr. Walter Moreira**, pelas contribuições e valiosas sugestões oferecidas no exame de qualificação que contribuíram para o amadurecimento e finalização desta tese.

As professoras **Dra. Ana Carolina Simionato e Dra. Rachel Cristina Vesu Alves**, membros da banca avaliadora, por aceitarem fazer parte da banca de defesa e

pela leitura minuciosa e crítica sobre meu trabalho, que possibilitou seu aprimoramento e a visualização de perspectivas futuras.

A Prof. **Dra. Maria Cláudia Cabrini Gracio**, coordenadora Programa de Pós-Graduação em Ciência da Informação, pelo apoio incondicional ao meu estágio de doutorado-sanduíche.

Aos **docentes do Programa de Pós-Graduação em Ciência da Informação**, pelos conhecimentos adquiridos ao longo dessa trajetória.

À **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES**, pelo financiamento desta pesquisa no exterior.

Aos **colegas e amigos do Programa de Pós-Graduação em Ciência da Informação da UNESP – Campus Marília**, vocês foram muito importante nesta caminhada, me ajudaram a conhecer e entender a Ciência da Informação, sua interdisciplinaridade e contribuições para a ciência.

Aos **colegas e amigos de “despacho” de doutorado-sanduíche na UC3M** pelo aprendizado, companheirismo, convivência, troca de experiências e também pelas “*bromas*”. Vocês foram fundamentais para realização desta estância.

Aos **funcionários do Programa de Pós-Graduação em Ciência da Informação**, pela atenção e competência demonstrada ao longo dessa caminhada.

Ao **Instituto Federal do Paraná – IFPR** pelo incentivo institucional à minha formação e ao meu afastamento.

Aos **colegas do Instituto Federal do Paraná – IFPR** pelo apoio, colaboração e incentivo decorrentes do afastamento de minhas atividades.

Por fim, agradeço a todas as pessoas que, direta ou indiretamente, contribuíram para a realização desta Tese.

A todos vocês, que fizeram parte da realização deste sonho, muito obrigada!

*“Uma mente que se abre para uma nova ideia,
jamais voltará ao seu tamanho original.”*

Albert Einstein

PÖTTKER, Luciana Maria Vieira. **Arquitetura para recuperação de objetos de aprendizagem – uma abordagem baseada em agentes inteligentes e *relevance feedback*. 203 f.** Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação – Faculdade de Filosofia e Ciências da Universidade Estadual Paulista – UNESP, Marília, 2017.

Resumo

Desde a sua criação, a *Web* tem crescido a um ritmo sem precedentes, situação esta que exigiu mudanças na forma como a sociedade busca e tem acesso à informação. O incremento informacional foi observado em todas as áreas do conhecimento e, desde então, problemas relacionados à recuperação de informação são investigados. No escopo dessa investigação, são pesquisados objetos de aprendizagem que estejam descritos, em um padrão de metadados educacional, e mantidos em repositórios específicos para este fim. Todos os problemas e dificuldades relacionados à recuperação de informação são refletidos no contexto particular dos objetos de aprendizagem. Devido à natureza (multimídia) dos objetos de aprendizagem, a complexidade em recuperá-los se torna mais perceptível. Nesta tese, propõe-se um modelo de arquitetura para recuperação de objetos de aprendizagem baseado em uma integração de tecnologias de sistemas de recuperação de informação, metadados, *relevance feedback* e agentes inteligentes. O propósito fundamental da arquitetura para recuperação de objetos de aprendizagem é unificar a representação desses recursos educacionais que são disponibilizados em diferentes repositórios e permitir que o usuário realize buscas qualificadas para localizar os objetos de aprendizagem mais adequados para sua necessidade de informação. Esta pesquisa é classificada como qualitativa e de natureza aplicada, uma vez que se relaciona com o problema prático de recuperação de objetos de aprendizagem disponíveis em repositórios da *Web*. O principal diferencial desta proposta foi de valorizar a inferência do usuário no processo de recuperação de informação, por meio do processo de *relevance feedback*. Neste processo, o usuário estabelece um diálogo com o sistema de recuperação de informação realizando refinamentos nos resultados que lhe foram retornados. Como esse processo é cíclico, ele pode ser executado até que o usuário esteja satisfeito com os resultados que lhe foram retornados. Conclui-se que um sistema de recuperação de informação é mais eficiente quando amplia seu escopo de recuperação a partir de diferentes fontes de dados e permite a inferência do usuário no julgamento da informação que lhe foi retornada.

Palavras-chave: Recuperação de Informação, Objetos de Aprendizagem, Metadados, Agentes Inteligentes, Realimentação por relevância.

PÖTTKER, Luciana Maria Vieira. **Architecture for learning object retrieval – intelligent agents and relevance feedback based approach. 203 f.** Thesis (PhD Degree in Information Science) – Faculdade de Filosofia e Ciências da Universidade Estadual Paulista – UNESP, Marília, 2017.

Abstract

The *Web* has been growing in a record speed since its creation and, therefore, such prospect has demanded changes in the way society seeks for and accesses information. Informational increment was evident in all fields of knowledge and since then, the relevant information retrieval issues have been investigated. In the scope of this investigation, we find researches in learning objects classified into an educational metadata pattern and kept in a specific repository. All the problems and complications related to such retrieval reflect in the learning objects particular context. The complexity in retrieving these learning objects becomes evident given their (multimedia) nature. Here, we suggest an architecture model to retrieve the aforementioned objects that is based on a combination of information retrieval system, metadata, relevance feedback, and intelligent agents. The main purpose of this architecture model is to unify the representation of these educational resources – that are available in a heterogeneous repository – and allow users to perform efficient searches in order to find the most suitable learning objects to their information needs. This is a qualitative and applied research once it relates to the practical problem of learning objects retrieval available on the *Web*. The main difference of this suggestion was to value – via relevance feedback – the importance of the user's inference in the process of such retrieval, in which the user establishes a dialog with the information retrieval system as to enhance the obtained results, and thus – being a cyclical process – it can be executed until he is pleased them. The conclusion is that an information retrieval system more efficient when its scope is enlarged from the different sources of data and allows the inference of the user when judging what he was presented with.

Keywords: Information Retrieval, Learning Objects, Metadata, Intelligent Agents, Relevance Feedback.

Lista de Figuras

Figura 1 – Relação do objeto de aprendizagem com os quatro fatores chave.....	44
Figura 2– Exemplo de objeto de aprendizagem com interface WIMP	46
Figura 3 – Exemplo de objeto de aprendizagem com interface post-WIMP	46
Figura 4 – Modelos de organização de um ROA	53
Figura 5 - Escopo do padrão ARIADNE com os elementos agrupados por categoria de metadados.	66
Figura 6 - Escopo do padrão <i>Dublin Core</i> (DC) com os elementos agrupados por categoria de metadados.....	69
Figura 7 – Fragmento de um arquivo RDF – <i>Dublin Core</i>	72
Figura 8 – Estrutura de arquivos do pacote de conteúdos	74
Figura 9 – Visão geral do padrão ISO MLR – Draft versão 2014.....	76
Figura 10 – Escopo do padrão LOM com os elementos agrupados por categoria de metadados	78
Figura 11 – Fragmento de código de um arquivo LOM.....	79
Figura 12 – Hierarquia de classes do padrão OBAA.....	81
Figura 13 – Fragmento de código do Metadado OBAA	81

Figura 14 – Fragmento de código do objeto de aprendizagem “Outras Infâncias - CINTED/UFRGS - Metadados” descrito com o metadados OBAA.....	82
Figura 15 – Visão abstrata de um agente e sua interação com o ambiente.	96
Figura 16 – Agentes interagindo com o ambiente por meio de sensores e atuadores	97
Figura 17 – Propriedades de um agente.....	100
Figura 18 – Classificação de agentes.....	105
Figura 19 – Visão geral de um sistema baseado em agentes.....	108
Figura 20 – Estágios da evolução da recuperação de informação.....	113
Figura 21 – Elementos de um SRI e suas inter-relações.	118
Figura 22 – Taxonomia dos modelos de recuperação de informação	120
Figura 23 – Categorização dos modelos de recuperação de informação com suas respectivas interdependências.....	120
Figura 24 – Representação gráfica da função de ranqueamento $R(q_i, d_j)$	123
Figura 25 – Resultado de uma busca de acordo com cada operador booleano.	124
Figura 26 – Representação de documentos na forma de vetores sobre os vértices da unidade hiper-cúbica.....	126
Figura 27 – Representação vetorial de um documento em um espaço vetorial tridimensional	127
Figura 28 – Subconjuntos de documentos após a execução de uma busca	131
Figura 29 – Diferentes áreas da recuperação de informação multimídia – MMIR	133
Figura 30 – Elementos constituintes de um motor de busca	135
Figura 31 – Visão geral da arquitetura para recuperação de objetos de aprendizagem	152
Figura 32 – Visão detalhada do Módulo 1 – Agente buscador e persistência.....	153
Figura 33 – Visão detalhada do Módulo 2 – Sistema de Recuperação de Informação	154
Figura 34 – Representação simplificada do modelo de recuperação de informação....	155
Figura 35 – Interface do Agente Buscador.....	158
Figura 36 – Interface – Configurações gerais.....	159

Figura 37 – Fluxograma de execução do <i>crawler</i>	161
Figura 38 – Visão geral da interface do <i>crawler</i>	162
Figura 39 – Visão geral do <i>crawler</i> realizando a extração das URLs	164
Figura 40 – Esquema básico do correlacionamento do arq_ROA.....	165
Figura 41 – Modelo XML <i>Schema</i> para o arquivo XML.	166
Figura 42 – Extração de metadados para o XML correspondente.	166
Figura 43 – Fluxograma do processo de <i>relevance feedback</i>	171

Lista de Quadros

Quadro 1 – Distribuição das etapas da pesquisa.....	33
Quadro 2 – Variações do termo Objeto de Aprendizagem.....	40
Quadro 3 – Características intrínsecas dos objetos de aprendizagem	44
Quadro 4 – Relação das características recomendadas para o desenvolvimento de um objeto de aprendizagem	48
Quadro 5 – Exemplos de repositórios de objetos de aprendizagem nacionais e internacionais, com respectivos padrões de metadados.....	53
Quadro 6 – Diferentes tipos de metadados e suas funções.....	61
Quadro 7 – Elementos de dados do <i>Dublin Core</i>	70
Quadro 8 – Síntese dos padrões de metadados.....	82
Quadro 9 – Correlação inicial entre os padrões de metadados DC e LOM.....	85
Quadro 10 - Visão geral do ROA – BIOE.....	86
Quadro 11 – Visão geral do ROA LUME	87
Quadro 12 – Refinamento na descrição dos objetos de aprendizagem padrão <i>Dublin Core</i>	88

Quadro 13 – Tipos de agentes e seus respectivos PEAS	101
Quadro 14 – Propriedades de um ambiente de tarefas	102
Quadro 15 – Quadro de teóricos que contribuíram com o desenvolvimento da área de RI	113
Quadro 16 – Principais conceitos sobre recuperação de informação compilados da literatura.....	114
Quadro 17 – Classificação dos modelos de recuperação de informação.....	121
Quadro 18 – Recuperação de informação na <i>Web</i> – Um problema desafiador.....	134
Quadro 19 – Tipos de relevância para um SRI.....	141
Quadro 20 – Detalhamento geral do cenário suportado pelo agente buscador de OA .	160

Listas de Tabelas

Tabela 1 – Descrição dos conteúdos e propósitos de cada arquivo do AICC	67
Tabela 2 – Elementos de refinamento do <i>Dublin Core</i>	71
Tabela 3 – Esquema Básico LOM v.1.0	77
Tabela 4 – Equivalência entre os padrões LOM e suas derivações.....	91
Tabela 5 – Mapeamento entre os padrões de metadados educacionais DC e LOM.....	92
Tabela 6 – Framework teórico para recuperação de informação – nas áreas de Ciência da Computação e Ciência da Informação.	112
Tabela 7 – Propriedades de recuperação de dados versus recuperação de informação. 116	
Tabela 8 – Diferentes tipos de motores de busca	136
Tabela 9 – Medidas para avaliação de um sistema de recuperação de informação.....	139
Tabela 10 – Definição de pesos aos elementos de metadados.	157
Tabela 11 – Características do ambiente PEAS em que atuam os agentes inteligentes 159	
Tabela 12 – Matriz binária de frequência dos termos nos elementos de metadados....	168
Tabela 13 – Matriz binária de frequência dos termos nos elementos de metadados e seu grau de similaridade.....	169
Tabela 14 – Matriz binária de incidência ranqueada.	169

Lista de Abreviaturas e Siglas

ACL	Agent Communication Language
ACM-DL	Association for Computing Machinery – Digital Library
AI	Agente Inteligente
AICC	Aviation Industry CBT Committee
AMS	Agent Management System
ANSI	American National Standards Institute
API	Application Programming Interface
ARIADNE	Alliance of Remote Instructional Authoring and Distribution Network for Europe
BD	Banco de Dados
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BIOE	Banco Internacional de objetos de Aprendizagem
BIR	Binary Independent Retrieval

BRAPCI	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
CAM	Content Aggregation Model
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CAREO	Campus Alberta Repository of Educational Objects
CBT	Computer-Based Training
CC	Ciência da Computação
CDU	Classificação Decimal Universal
CI	Ciência da Informação
CESTA	Coletânea de Entidades de Suporte ao Uso da Tecnologia na Aprendizagem
CMI	Computer-managed Instruction
DC	Dublin Core
DCMI	Dublin Core Metadata Initiative
DeVOA	Desenvolvimento e Validação de Objetos de Aprendizagem
FINEP	Financiadora de Estudos e Projetos
FIPA	Foundation for Intelligent Physical Agents
FIPA-ACL	Foundation for Intelligent Physical Agents – Agent Communication Language
FUNTEL	Fundo para o Desenvolvimento Tecnológico das Telecomunicações
GPL	General Public License

HTML	Hipertext Markup Language
HTTP	HyperText Transfer Protocol
IA	Inteligência Artificial
IBM	International Business Machines
IBTF	Instituto Brasileiro de Educação e Tecnologia de Formação a Distância
IDF	Inverse Document Frequency
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronic Engineers
IFLA	International Federation of Library Associations and Institutions
IFPR	Instituto Federal do Paraná
IIOB	Internet Inter-ORB Protocol
IMS LRM	Instructional Management Systems - Learning Resource Metadata
ISO	International Organization for Standardization
ISO MLR	ISO Metadata for Learning Resources
JADE	Java Agent Development Framework
JTC	Joint Technical Committee
KIF	Knowledge Interchange Format
KQML	Knowledge Query and Manipulation Language
LD	Learning Design
LGPL	Lesser General Public License

LMS	Learning Management Systems
LO	Learning Object
LOM	Learning Object Metadata
LOR	Learning Object Repository
LRM	Learning Resource Metadata
LTSC	Learning Technology Standards Committee
MAS	Multi-agents Systems
MCT	Ministério da Ciência e Tecnologia
MEC	Ministério da Educação
MEMEX	Memory Extension
MERLOT	Multimedia Education Resource for Learning and Online Teaching
MOODLE	Modular Object-Oriented Dynamic Learning Environment
MLR	Metadata for Learning Resources
MMIR	Multimedia Information Retrieval
NISO	National Information Standards Organization
OA	Objeto de Aprendizagem
OBAA	Objeto de Aprendizagem Baseado em Agentes
OWL	Web Ontology Language
PEAS	Performance, Environment, Actuators, Sensors
PRP	Probability Ranking Principle

RBU	Repertório Bibliográfico Universal
RDF	Resource Description Framework
RF	Relevance Feedback
RELPE	Rede Latino-americana de Portais Educacionais
RGB	Red, green, blue
RI	Recuperação de Informação
RIVED	Rede Interativa Virtual de Educação
RNP	Rede Nacional de Ensino e Pesquisa
ROA	Repositório de Objetos de Aprendizagem
RTE	Run-time Environment
S.A.B.I.O.	Sistema de Almacenamiento y Búsqueda de Información Organizada
SCIELO	Science Direct, Scientific Electronic Library Online
SCORM	Sharable Content Object Reference Model
SEED	Secretaria de Educação a Distância
SGBD	Sistema Gerenciador de Banco de Dados
SMA	Sistemas Multi-Agentes
SMART	System for the Manipulation and Retrieval of Text
SMTP	Simple Mail Transfer Protocol
SN	Sequencing and Navigation
SRI	Sistemas de Recuperação de Informação

TCP/IP	Transmission Control Protocol/Internet Protocol
TF	Term Frequency
TI	Tecnologia da Informação
TIC	Tecnologias de Informação e Comunicação
TV	Televisão
UC3M	Universidad Carlos III de Madrid
UFRGS	Universidade Federal do Rio Grande do Sul
UML	Unified Modeling Language
UNESP	Universidade Estadual Paulista “Júlio de Mesquita Filho”
UNISINOS	Universidade do Vale dos Sinos
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium
W3C-DTF	W3C Date and Time Formats
WBT	Web Based Training
WIMP	Windows, Icons, Menus, Pointers
WOS	Web of Science
XML	Extensible Markup Language

Sumário

1. INTRODUÇÃO	26
1.1 Delimitação do tema	29
1.2 Problema de Pesquisa	29
1.3 Hipóteses	31
1.4 Tese	31
1.5 Objetivos	31
1.5.1 Objetivo Geral	31
1.5.2 Objetivos Específicos	31
1.6 Procedimentos Metodológicos	32
1.7 Estrutura do Trabalho	35
2. OBJETOS DE APRENDIZAGEM	37
2.1 Repositórios de objetos de aprendizagem – ROA	49
3. METADADOS DE OBJETOS DE APRENDIZAGEM	57
3.1 Padrões de Metadados Educacionais	63
3.1.1 ARIADNE educational metadata recommendation	64
3.1.2 Aviation Industry CBT Committee – AICC	66
3.1.3 <i>Dublin Core</i> – DC	68

3.1.4 Instructional Management Systems - Learning Resource Metadata – IMS LRM -----	73
3.1.5 ISO Metadata for Learning Resources – ISO MLR-----	74
3.1.6 Learning Object Metadata – LOM-----	76
3.1.7 Objetos de Aprendizagem Baseado em Agentes – OBAA-----	79
3.1.8 Síntese dos padrões de metadados educacionais-----	82
3.2 Correlação dos padrões de metadados educacionais-----	84
4. AGENTES INTELIGENTES -----	95
4.1 Classificação dos agentes inteligentes -----	103
4.2 Sistemas Multi-agentes – SMA -----	106
4.3 Estratégias de Comunicação entre Agentes-----	108
5. RECUPERAÇÃO DE INFORMAÇÃO-----	111
5.1 Recuperação de Informação: Modelos Clássicos -----	119
5.1.1 Modelo Booleano -----	124
5.1.2 Modelo Espaço Vetorial -----	125
5.1.3 Modelo Probabilístico-----	128
5.2 Recuperação de Informação Multimídia-----	132
5.3 Recuperação de Informação na Web-----	133
5.3.1 Tipos de Motores de Busca-----	135
5.3.2 Tipos de interface de busca de informação -----	137
5.4 Avaliação dos Sistemas de Recuperação de Informação -----	139
6. RELEVANCE FEEDBACK (RF) -----	143
6.1 Tipos de <i>Relevance Feedback</i> -----	146
6.2 Avaliação de <i>Relevance Feedback</i> -----	148
7. ARQUITETURA PARA RECUPERAÇÃO DE OBJETOS DE APRENDIZAGEM- 151	
7.1 Representação do processo de recuperação de informação -----	154
7.1.1 Representação dos documentos-----	155
7.1.2 Representação da expressão de busca -----	156

7.1.3 Função de busca-----	156
7.2 Módulo 1 – Agente Buscador e Persistência-----	157
7.2.1 Agente Buscador-----	158
7.2.2 <i>Web Crawler</i> -----	161
7.2.3 Persistência -----	164
7.3 Módulo 2 – Sistema de Recuperação de Informação -----	167
8. CONSIDERAÇÕES FINAIS-----	173
REFERÊNCIAS-----	177
APÊNDICES -----	190
APÊNDICE A – MODELOS DE SISTEMA – DIAGRAMAS DE CASOS DE USO ----	191
APÊNDICE B – PSEUDOCÓDIGO – INICIALIZAÇÃO DO AGENTE BUSCADOR -	194
APÊNDICE C – PSEUDOCÓDIGO - <i>LOOP CRAWLER</i> -----	195
APÊNDICE D – FRAGMENTOS DE ARQUIVOS (URLS) RECUPERADOS PELO <i>WEB CRAWLER</i> -----	197
ANEXOS -----	200
ANEXO A – MAPA MENTAL DO PADRÃO OBAA-----	201
ANEXO B – EXEMPLO DE REGISTRO NO PADRÃO OBAA -----	203

1. Introdução

Os objetos de aprendizagem (OA) são recursos educacionais idealizados para potencializar as estratégias de ensino e aprendizagem. Esses recursos, frequentemente disponibilizados na *Web*, são construídos em diversos formatos e podem ser combinados para serem aplicados em diferentes contextos de aprendizagem. Conforme a IEEE LTSC LOM (2002) os objetos de aprendizagem são qualquer tipo de recurso (digital ou não digital) que podem ser utilizados, reutilizados ou referenciados durante todo o processo de ensino e aprendizagem apoiado por computador.

Esses recursos educacionais são mantidos em repositórios de objetos de aprendizagem (ROA, ou em inglês *Learning Object Repository* – LOR) que são ambientes digitais que têm como propósito disponibilizar recursos informacionais descritos em um padrão de metadados específico a fim de facilitar a busca e o acesso a esses objetos, bem como para garantir sua interoperabilidade e reuso. Nesse contexto, os metadados são utilizados para descrever e manter acessíveis os objetos de aprendizagem para que possam ser recuperados e reutilizados em outros contextos (SILVA; CAFÉ; CATAPAN, 2010).

Um repositório de objetos de aprendizagem pode ser criado por uma iniciativa isolada ou envolver um consórcio de instituições acadêmicas. Possui sua própria infraestrutura e organização, que contempla desde a adoção de um padrão de metadados educacional para a descrição dos recursos informacionais até um motor de busca próprio

para viabilizar a recuperação de informação por meio de uma interface de busca. Essa situação limita a recuperação de informação ao seu próprio *corpus*¹ de recursos educacionais, uma vez que os resultados estarão limitados pela base de dados suportada pelo ROA.

Nas últimas décadas, com o advento da *Web* e a incorporação das tecnologias de informação e comunicação (TIC) na esfera educacional, torna-se difícil dimensionar a quantidade de repositórios e/ou objetos de aprendizagem existentes. Recuperar essas informações utilizando os motores de busca de uso geral (como por exemplo, o Google), os metabuscadores ou os recursos de busca específicos de cada repositório são um fator limitante ao acesso desses recursos informacionais, uma vez que o conjunto de resultados alcançados pode ser muito amplo, no caso dos buscadores de uso geral ou dos metabuscadores, ou muito restrito, no caso das buscas em um único ROA.

Esse cenário propõe uma reflexão a respeito da recuperação de informação (RI) na *Web* como um problema desafiador para a sociedade da informação, pois a maneira como a sociedade busca e tem acesso à informação foi alterada abruptamente pela forma como ocorre o incremento informacional, bem como pela dinamicidade e volatilidade das informações nesse ambiente.

Todos os problemas e dificuldades relacionados à recuperação de informação na *Web* refletem-se no contexto particular dos objetos de aprendizagem, acrescido da complexidade relacionada às características inerentes a esses tipos de objetos digitais.

De acordo com Baeza-Yates e Ribeiro-Neto (2013, p. 4), o principal propósito de um sistema de recuperação de informação (SRI) é auxiliar os usuários a localizar informações de seu interesse, maximizando a proporção de satisfação do usuário e minimizando seu esforço.

No escopo de um sistema de recuperação de informação, normalmente o usuário desconhece o *corpus* a ser recuperado e nem sempre consegue formular uma expressão de busca que represente adequadamente sua necessidade de informação. Entretanto, ao

¹ *Corpus* – é um repertório ou um conjunto de itens sobre um tema específico.

avaliar os itens retornados se torna mais fácil identificar o que é ou não inerente a sua busca naquele momento.

Quando se trata da recuperação de objetos de aprendizagem descritos por um padrão de metadados, parte-se do princípio de que a descrição desses recursos pode facilitar a inferência do usuário na análise de relevância das características dos itens que lhe foram retornados. Como os objetos de aprendizagem possuem diversas especificidades em sua descrição é possível estabelecer um diálogo do usuário com o SRI para que os resultados sejam melhor classificados. Esse processo dialógico permite que o usuário realize refinamentos sobre os resultados que lhe foram retornados, por meio de um processo denominado de *relevance feedback*² (RF), para aprimorar os resultados recuperados sem que seja necessário que o usuário refaça sua expressão de busca inicial.

O *relevance feedback* é um processo que permite obter informações complementares acerca da necessidade de informação do usuário por meio da seleção de itens relevantes recuperados de um *corpus* para melhorar a qualidade e o ranqueamento dos resultados obtidos pelo motor de recuperação de informação. Este processo tem o propósito de incrementar o desempenho do SRI por meio da inferência do usuário no refinamento dos resultados de uma busca.

Por isso, acredita-se que se o usuário realizar inferências no SRI indicando, a princípio, quais objetos de aprendizagem recuperados são mais relevantes, em sua concepção, seja possível obter resultados melhores e mais precisos para sua necessidade de informação.

Dessa forma, esse trabalho propõe explorar técnicas de recuperação de informação para o domínio dos objetos de aprendizagem. Serão utilizados agentes inteligentes para buscar e descobrir metadados de objetos de aprendizagem em ROA, bem como será proposto uma arquitetura de recuperação de informação com suporte ao *relevance feedback*.

² *Relevance feedback* ou realimentação de relevância ou *feedback* de relevância ou retroalimentação por relevância – Por uma diversidade de traduções observadas na literatura, nesta investigação, optou-se em manter o termo original em inglês – *relevance feedback*.

Portanto, a proposição desta investigação se justifica pela necessidade atual de melhorar a eficácia dos sistemas de recuperação de objetos de aprendizagem no que tange ao desempenho dos SRI para garantir que esses recursos possam ser recuperados de forma mais precisa, bem como para que o usuário possa inferir no julgamento da informação que ele acredita ser relevante naquele momento. Assim sendo, surgem resultados que contribuem para a área de Ciência da Informação e para a sociedade da informação no que se refere à melhoria da eficácia dos sistemas de recuperação de informação e no grau de satisfação do usuário.

1.1 Delimitação do tema

Ao longo dos últimos anos, a quantidade de informação disponibilizada na *Web* vem se multiplicando em grandes proporções e o desafio é como recuperá-las de forma eficiente e relevante. Seguindo esta tendência, os objetos de aprendizagem têm sido disponibilizados em grandes escalas na *Web*, o que desperta o interesse em investigar novas metodologias para recuperar esses recursos informacionais.

Dessa forma, esta pesquisa investiga a recuperação de objetos de aprendizagem disponíveis em repositórios de objetos de aprendizagem que adotam um padrão de metadados para descrevê-los.

Devido à natureza (multimídia) dos objetos de aprendizagem, a complexidade em recuperá-los se torna mais perceptível. Propor métodos alternativos para recuperação de cada formato (texto, imagem, vídeo, áudio, simulação etc.) é complexo, pois cada tipo (formato) de OA possui suas especificidades para a implementação de um SRI. Portanto, compreende-se a necessidade de apresentar uma solução mais eficiente para recuperar informação nesse domínio do conhecimento.

1.2 Problema de Pesquisa

O problema desta pesquisa é visualizado a partir do seguinte questionamento: *é possível a inferência do usuário, por meio do processo de relevance feedback, contribuir na obtenção de resultados mais precisos de um sistema de recuperação de objetos de aprendizagem?*

Os objetos de aprendizagem possuem uma variedade muito grande de formatos e estrutura de dados, aspecto esse que dificulta o processo de recuperação de informação eficiente na *Web*. Os objetos de aprendizagem devem ser concebidos para atender alguns princípios, tais como: o reuso, a interoperabilidade, a modularidade e a customização para justificar o seu tempo de desenvolvimento e o custo envolvido para criá-lo. Nesse sentido, a W3C Brasil (2014) recomenda a adoção de metadados com o intuito de descrever a sintaxe e os métodos comuns que favorecem a interoperabilidade entre recursos heterogêneos.

Portanto, essa pesquisa explora repositórios de objetos de aprendizagem que tenham seu acervo descrito por meio de um padrão de metadados educacional. A descrição dos objetos de aprendizagem permite que o conteúdo educacional seja representado com maiores detalhes por meio de refinamentos de elementos ou esquemas de codificação que facilitam a localização dos objetos de aprendizagem e possibilitam a distribuição dos conteúdos de aprendizagem às pessoas certas no tempo certo.

Assim, percebe-se que os metadados são mais úteis quando os recursos a serem recuperados não são apenas textuais, mas sim uma combinação de elementos textuais, gráficos e audiovisuais dentro de uma única coleção ou apresentação. Corroborando com esta afirmação, Silva e Souza (2013) enfatizam que os metadados podem ser utilizados tanto para descrever características técnicas de baixo nível (data da criação, por exemplo) quanto para representar atributos de alto nível (grau de dificuldade do recurso educacional, por exemplo).

A partir da análise do estado da arte da tecnologia de objetos de aprendizagem, identificou-se sua expressividade e também as limitações existentes para atingir os objetivos de reutilização, interoperabilidade e melhorar a eficiência nas buscas. Com base na análise dos processos de elaboração, armazenamento e recuperação de objetos de aprendizagem, percebeu-se a possibilidade de propor um modelo de recuperação de informação para esse universo em estudo. O modelo proposto deve permitir a interação do usuário com o motor de recuperação de informação para obter resultados mais relevantes a partir de objetos de aprendizagem recuperados de repositórios que possuem diferentes padrões arquiteturais e que adotam padrões de metadados distintos, por meio de uma abordagem baseada em *relevance feedback*.

1.3 Hipóteses

Diante do problema apresentado, as hipóteses desta pesquisa partem do entendimento de que a comunicação do usuário no processo de recuperação de informação, por meio da realização de refinamentos sucessivos (*relevance feedback*), incrementa a satisfação do usuário, bem como a unificação dos padrões de metadados educacionais, por meio da proposição de um modelo de correlação entre eles, amplia o escopo das buscas em um sistema de recuperação de informação.

1.4 Tese

A tese (ou premissa) desta pesquisa é comprovar que um sistema de recuperação de objetos de aprendizagem é mais eficiente à medida que incrementa seu *corpus* (com base em diferentes fontes de dados – ROA) e permite a interação do usuário no julgamento da relevância da informação que lhe foi retornada.

1.5 Objetivos

Os objetivos que norteiam o desenvolvimento da pesquisa são apresentados nas subseções a seguir.

1.5.1 Objetivo Geral

Com o propósito de contribuir com a área de Ciência da Informação e com a sociedade da informação que tem a necessidade de recuperar informações de maneira rápida e precisa, esta investigação busca, *por um lado, propor a unificação da representação dos objetos de aprendizagem disponibilizados em diferentes repositórios e, por outro lado, permitir ao usuário realizar buscas qualificadas para encontrar os recursos educacionais adequados para sua necessidade de informação.*

1.5.2 Objetivos Específicos

Para compreender e cumprir o objetivo geral dessa pesquisa, os objetivos específicos a serem abordados são os seguintes:

- Analisar as principais arquiteturas utilizadas no desenvolvimento de repositórios de objetos de aprendizagem, bem como identificar os padrões de metadados adotados;
- Propor um modelo de correlação entre os elementos dos padrões de metadados educacionais;
- Investigar como a abordagem de agentes inteligentes pode contribuir com a recuperação de objetos de aprendizagem;
- Analisar os modelos de recuperação de informação e identificar qual melhor se aplica a esta investigação;
- Propor um modelo de *crawler*³ baseado em agentes inteligentes capaz de descobrir e recuperar metadados de objetos de aprendizagem armazenados em repositórios;
- Apresentar uma arquitetura de sistema de recuperação de informação com suporte à abordagem de *relevance feedback*.

1.6 Procedimentos Metodológicos

Para o desenvolvimento desta pesquisa, a metodologia utilizada consistiu em uma análise exploratória, descritiva e aplicada sobre o tema de recuperação de objetos de aprendizagem, o que permitiu a construção de um embasamento teórico sobre os princípios, os padrões e os instrumentos que orientam o desenvolvimento dos sistemas de recuperação de informação, auxiliando na compreensão do problema proposto e na concepção dos requisitos para a proposição da arquitetura para recuperação de objetos de aprendizagem baseado no princípio de *relevance feedback* e agentes inteligentes.

Segundo Köcher (2002, p. 126) os estudos exploratórios permitem incrementar o processo de investigação da natureza de um determinado fenômeno, bem como possibilitam identificar as diversas variáveis que se pretende explorar.

Com base na abordagem do problema, essa pesquisa pode ser classificada como qualitativa e de natureza aplicada, uma vez que se relaciona com um problema prático de recuperação de objetos de aprendizagem disponíveis em repositórios da *Web*.

³ *Crawler* – é um *software* que navega pela *Web*, de forma metódica e automatizada, para descobrir informação. No escopo desta tese, o *crawler* baseado em agentes inteligentes atuará para descobrir e recuperar metadados de novos objetos de aprendizagem disponibilizados nos ROA.

Deste modo, a descrição dos procedimentos metodológicos necessários para a realização desta pesquisa foi detalhada, conforme as etapas constantes no Quadro 1.

Quadro 1 – Distribuição das etapas da pesquisa

ETAPAS (PROCEDIMENTOS METODOLÓGICOS)	DESCRIÇÃO DAS ATIVIDADES
Etapa 1 – Revisão da literatura	<p>O levantamento bibliográfico foi elaborado tanto em nível nacional quanto internacional, em fontes: primárias (livros, capítulos de livros, periódicos, anais de congressos e teses e/ou dissertações) e secundárias (<i>Association for Computing Machinery – Digital Library (ACM-DL)</i>, Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), Google Acadêmico, Periódicos Capes, <i>Science Direct</i>, <i>Scientific Electronic Library Online (SCIELO)</i>, <i>IEEE Xplore Digital Library</i>, <i>Web of Science (WoS)</i>). Não foi imposta uma limitação cronológica para as buscas com a intenção de observar a evolução das investigações correlatas a esta pesquisa.</p> <p>Foram utilizadas como estratégia de busca, as expressões (e suas composições com operadores booleanos) nos seguintes idiomas:</p> <p>Português: ‘objetos de aprendizagem’; ‘repositórios de objetos de aprendizagem’; ‘padrões de metadados educacionais’; ‘recuperação de informação’; ‘modelos clássicos de recuperação de informação’; ‘sistema de recuperação de informação’; relevância; realimentação por relevância; ‘avaliação de desempenho’; ‘interface de busca’; ‘persistência de metadados’; ‘representação de metadados’; ‘motores de busca’; ‘rastreador’; ‘agentes inteligentes’; ‘comunicação entre agentes’.</p> <p>Inglês: ‘learning objects’; ‘learning object repository’; ‘educational metadata standard’; ‘information retrieval’; ‘information retrieval classic models’; ‘information retrieval system’; relevance; ‘relevance feedback’; ‘performance evaluation’; ‘search interface’; ‘metadata persistence’; ‘metadata representation’; ‘search engines’; ‘crawler’; ‘intelligent agents’; ‘agent communication’.</p> <p>Espanhol: ‘objetos de aprendizaje’; ‘repositorios de objetos de aprendizaje’; ‘normas de metadatos educativos’; ‘recuperación de la información’; ‘modelos clásicos de recuperación de la información’; ‘sistema de recuperación de la información’; pertinencia; ‘relevancia de realimentación’; ‘evaluación de desempeño’; ‘interfaz de búsqueda’; ‘metadatos persistentes’; ‘representación de metadatos’; ‘motores de búsqueda’; ‘rastreador’; ‘agentes inteligentes’; ‘comunicación entre agentes’.</p> <p>Os materiais obtidos foram analisados segundo critérios de pertinência ao tema proposto, com o intuito de compor o embasamento teórico sólido e criar aportes para elaboração da pesquisa.</p>
Etapa 2 – Analisar os repositórios de objetos de aprendizagem e identificar sua arquitetura e padrão de metadados adotado	<p>Inicialmente, foi produzido um estudo sobre a origem e as características que compreendem o domínio dos objetos de aprendizagem. Em seguida, foi realizada uma análise sobre as principais arquiteturas e padrões utilizados na construção de repositórios dos objetos de aprendizagem para compreender como os recursos educacionais estão sendo armazenados. Foram analisados sete repositórios, sendo:</p> <ul style="list-style-type: none"> ▪ Três internacionais – ARIADNE (<i>Alliance of Remote Instructional Authoring and Distribution Network for Europe</i>); CAREO (<i>Campus Alberta Repository of Educational Objects</i>); MERLOT II (<i>Multimedia Education Resource for Learning and</i>

	<p><i>Online Teaching</i>);</p> <ul style="list-style-type: none"> ▪ Quatro nacionais – BIOE (Banco Internacional de Objetos de Aprendizagem); CESTA (Coletânea de Entidades de Suporte ao Uso da Tecnologia na Aprendizagem); LUME e Portal OBAA. <p>Nesta análise, identificou-se o padrão de metadados adotado e a arquitetura suportada por cada repositório de objetos de aprendizagem, conforme descrito no Capítulo 2. Em paralelo a este estudo, foram pesquisados quais os padrões de metadados educacionais que estão sendo adotados para possibilitar a representação e descrição dos objetos de aprendizagem nos repositórios, conforme descrito no Capítulo 3.</p>
<p>Etapa 3 – Propor um modelo de correlação entre os padrões de metadados educacionais</p>	<p>A partir da análise dos padrões de metadados utilizados para descrever objetos de aprendizagem foi possível identificar que a maioria deles está baseada em dois padrões de metadados: o <i>Dublin Core</i> (DC) e o <i>Learning Object Metadata</i> (LOM), conforme apresentado na seção 2.2. Dessa forma, foi proposto um modelo de mapeamento entre os padrões de metadados educacionais DC e LOM, tendo como base os 15 elementos básicos do DC. Essa correlação foi obtida pela análise do nível de similaridade entre os elementos dos padrões investigados e por uma amostragem de objetos de aprendizagem recuperados dos ROA, conforme detalhado na seção 3.2 desta tese.</p>
<p>Etapa 4 – Concepção da Arquitetura para Recuperação de Objetos de Aprendizagem</p>	<p>A arquitetura para recuperação objetos de aprendizagem é constituída por dois módulos:</p> <p>Módulo 1 – Agente Buscador e Persistência – é responsável pela descoberta de repositórios de objetos de aprendizagem e tem a função de mapear e atualizar automaticamente os metadados dos novos objetos de aprendizagem descobertos pelo agente inteligente. Este módulo também realiza a identificação do padrão de metadados do ROA, bem como executa a correlação entre os metadados para serem persistidos por meio de tecnologias relacionais. Os princípios dos agentes inteligentes são especificados no Capítulo 4 desta tese.</p> <p>Módulo 2 – Sistema de Recuperação de Informação: é composto de duas partes:</p> <ul style="list-style-type: none"> ▪ <u>Interface de busca</u> – permite a interação do usuário com o Motor de RI da seguinte forma: o usuário especifica sua necessidade de informação por meio de uma expressão de busca, de forma análoga como ocorre nos motores de busca de uso geral da <i>Web</i>. Após visualizar os resultados retornados pelo Motor de RI, o usuário seleciona os itens que ele julgar como relevante e submete novamente ao motor de RI para refinar sua busca e apresentar um novo resultado. Este processo é denominado <i>relevance feedback</i> e pode ser executado quantas vezes o usuário julgar necessário. Para acessar o objeto de aprendizagem selecionado, basta o usuário clicar no <i>link</i> e será redirecionado ao repositório que mantém o recurso educacional. ▪ <u>Motor de RI</u> - interage diretamente com a expressão de busca informada pelo usuário, trata a informação e realiza uma busca na base de metadados, ranqueia os itens recuperados pela presença dos termos nos elementos de metadados (de acordo com similaridade entre a expressão de busca e os registros de metadados) e envia os itens recuperados ao usuário. O motor de RI está apto a receber os <i>feedbacks</i> do usuário por meio da seleção dos itens que o usuário julgou como relevantes, naquele momento, para executar uma nova busca e ranquear novamente o <i>corpus</i> apresentado ao usuário. Este processo é denominado de <i>relevance feedback</i> e pode ser executado inúmeras vezes até que o usuário esteja satisfeito com o resultado apresentado.

	<p>Os fundamentos da recuperação de informação e de <i>relevance feedback</i> são apresentados nos capítulos 5 e 6, respectivamente.</p> <p>A descrição detalhada de cada um dos módulos e tecnologias utilizadas nesta arquitetura é apresentada no Capítulo 7.</p>
--	--

Fonte: elaborado pela autora.

1.7 Estrutura do Trabalho

Esta tese está estruturada em oito capítulos. O primeiro Capítulo exposto nesta **Introdução** apontou o delineamento da investigação, identificando o tema, o problema, as hipóteses, a tese, os objetivos e os procedimentos metodológicos adotados para abordar a temática desta pesquisa. Os demais capítulos que compõem esta tese serão apresentados na seguinte sequência.

O capítulo 2 apresenta os **Objetos de Aprendizagem** como recursos educacionais, bem como explica sua origem, conceitos e características. Ainda, são apresentados os repositórios de objetos de aprendizagem como um ambiente organizado para armazenar esses recursos informacionais de tal forma que permita o compartilhamento e a reutilização dos objetos de aprendizagem.

O capítulo 3, intitulado **Metadados de Objetos de Aprendizagem** apresenta a importância da descrição dos objetos de aprendizagem por meio um padrão de metadados educacional. Justifica-se essa necessidade devido à natureza diversificada de formatos que são usados para o desenvolvimento dos objetos de aprendizagem, assim a descrição desses recursos por meio de um padrão de metadados facilita a busca e a recuperação desses recursos na *Web*.

O capítulo 4, denominado **Agentes Inteligentes**, resgata esta área dentro da inteligência artificial que é responsável por simular o comportamento humano por meio de agentes de software. A abordagem de agentes inteligentes foi empregada nessa pesquisa para propiciar um processo de busca e recuperação de objetos de aprendizagem mais preciso e eficiente nos repositórios de objetos de aprendizagem.

O Capítulo 5, **Recuperação de Informação**, apresenta uma visão geral da história da recuperação de informação, importância e evolução dessas pesquisas para as áreas de Ciência da Informação e Ciência da Computação. Esse capítulo apresenta os métodos e técnicas investigados para propor um modelo eficiente de recuperação de informação na *Web*.

O capítulo 6, **Relevance Feedback (RF)**, demonstra os princípios dos sistemas de recuperação de informação que permitem a interação do usuário no processo de recuperação de informação relevante. São explicadas as técnicas para aplicação de *relevance feedback* em um sistema de recuperação de informação.

O capítulo 7, **Arquitetura para Recuperação de Objetos de Aprendizagem**, demonstra os modelos, módulos, funcionamento e as tecnologias abordadas na solução proposta nesta tese.

Por fim, as **Considerações Finais**, apresentadas no oitavo capítulo, demonstram as conclusões resultantes desse estudo e ainda são propostas melhorias, limitações e possíveis caminhos para aperfeiçoar o desempenho e a eficiência dos sistemas de recuperação de informação na *Web*.

2. Objetos de Aprendizagem

Os primeiros relatos sobre os objetos de aprendizagem datam do início da década de 1990, quando Wayne Hodgins os apresentou como se fossem blocos de Lego⁴. Essa analogia apresentada pelo autor compara os objetos de aprendizagem com os blocos de Lego e os apresenta como se fossem pequenos elementos que podem ser agrupados e reagrupados para produzirem experiências pedagógicas mais significativas (HODGINS, 2000, p. 27).

Questionando a metáfora apresentada por Hodgins, Wiley (2001) apresenta a analogia do átomo/molécula que compara o objeto de aprendizagem com uma pequena unidade (átomo/molécula) que tem a capacidade de se (re) agregar a outras unidades, desde que determinadas características sejam preservadas em sua estrutura interna, ou seja, o objeto de aprendizagem tem que ter um nível de granularidade⁵ mínimo para que possa ser reutilizado. Neste contexto, o autor salienta a relevância do *design* no processo de desenvolvimento dos objetos de aprendizagem para permitir o reuso e o intercâmbio desses recursos em diferentes contextos de aprendizagem.

⁴ Lego – é um brinquedo infantil composto por pequenos blocos encaixáveis a outros blocos que permitem compor novas estruturas, uma vez que podem ser desmontados e reutilizados. Possuem a capacidade de serem reusados, em diferentes combinações, explorando a criatividade para construir novos projetos, fomentando a imaginação e a inovação.

⁵ Granularidade – é a possibilidade de apresentar um conteúdo em unidades menores (módulos) para que possam ser reutilizados em diferentes contextos de aprendizagem.

Após analisar as analogias do “Lego” e do “Átomo”, Paquette e Rosca (2002) discordaram da forma simplista como estas foram apresentadas e propuseram uma analogia de abordagem orgânica para comparar os objetos de aprendizagem, pois consideram que para ocorrer à agregação dos componentes para construir novos recursos é necessário analisar também a reação, o contexto e os atores envolvidos no processo (de aprendizagem) para garantir que a reutilização de fato ocorra. Dessa forma, os autores enfatizam que a fisiologia e a dinâmica do objeto de aprendizagem também devem ser consideradas antes de ser simplesmente empregado em novos contextos de aprendizagem.

Anos mais tarde, novas analogias foram sendo apresentadas para reforçar os princípios dos objetos de aprendizagem. Masie (2003) compara os objetos de aprendizagem com materiais pré-fabricados, delineados e planejados para serem utilizados em diferentes tipos de construções. Desse modo, esses materiais podem ser combinados de diversas formas na construção de casas e edifícios de todos os portes. Assim, pode-se inferir que cada unidade de objeto de aprendizagem deve ser reutilizada para compor novas unidades de um conteúdo educacional ou curso.

A partir das analogias apresentadas ao longo dos anos, entende-se que o potencial pedagógico desses recursos está em reutilizá-los, ou seja, o ideal é que esses materiais possuam um alto grau de modularidade⁶ de tal forma que possibilitem a combinação com outros recursos gerando novas experiências de aprendizagem. Com base na compreensão da forma como os objetos de aprendizagem podem ser interligados, o processo de reutilização poderá ocorrer mais naturalmente.

Para compreender a real dimensão deste domínio, a seguir serão apresentados alguns conceitos formais sobre os objetos de aprendizagem encontrados na literatura e adotados pela comunidade científica.

Inicialmente, Wiley (2001) define “objetos de aprendizagem” como elementos instrucionais baseados no paradigma da orientação a objetos da Ciência da Computação que valorizam a criação de componentes (objetos) que podem ser reutilizados em múltiplos contextos de aprendizagem. Sendo assim, parte-se do princípio de que os

⁶ Modularidade – sinônimo para granularidade

designers tenham condições de produzir pequenos componentes pedagógicos (unidades de ensino) que sejam reutilizados para gerar novas experiências de aprendizagem, adaptadas às necessidades de aprendizagem de cada educando.

Na concepção de Ip, Morrison e Currie (2001), um objeto de aprendizagem é percebido como uma estrutura que deve ter no mínimo quatro subcomponentes que são: o conteúdo, as funções pedagógicas, os objetivos de aprendizagem e, por último, serem visuais e permitirem sua experimentação (*look and feel*). Esses componentes indicados pelo autor garantem que o objeto de aprendizagem cumpra seu papel de facilitar do processo de ensino e aprendizagem.

Para Weller, Pegler e Mason (2003) o objeto de aprendizagem é uma parte digital do material que se dirige a um tópico claramente identificável ou resultado da aprendizagem, desde que tenha potencial de reutilização em diferentes contextos pedagógicos. Nesta mesma linha de pensamento, Sosteric e Hesemeier (2004, p. 40) descrevem um objeto de aprendizagem como um arquivo em formato digital (imagem, filme, áudio e etc.) aplicado em contextos pedagógicos que possuem, internamente ou por meio de associações, sugestões sobre o contexto adequado para explorar o objeto em questão.

Após a análise das várias definições de objetos de aprendizagem investigadas, observa-se que não existe a preocupação formal com a descrição adequada do recurso em um padrão de metadados educacional para fins de recuperação desses recursos informacionais. Portanto, acredita-se que com a descrição desses recursos é possível garantir a busca, a interoperabilidade e o intercâmbio dos objetos de aprendizagem.

Por isso, no contexto desta investigação, um objeto de aprendizagem é definido como um recurso educacional reutilizável e granular, suportado por algum tipo de tecnologia e descrito por um padrão de metadados para proporcionar sua futura recuperação.

Com o passar do tempo, outras nomenclaturas e definições foram sendo apresentadas para refinar e conceituar esses recursos informacionais. Com base em uma avaliação na literatura da área foram constadas as seguintes terminologias (em ordem cronológica de origem) para se referir a objetos de aprendizagem, conforme pode ser observado no Quadro 2.

Quadro 2 – Variações do termo Objeto de Aprendizagem

CONCEITO	DESCRIÇÃO
Objetos de Conhecimento – <i>Knowledge objects</i> (MERRILL, 1998)	Um objeto de conhecimento é uma forma precisa para descrever conteúdo ou conhecimento a ser ensinado por meio de uma base de dados (base de conhecimento) de recursos informacionais (texto, áudio, vídeo, gráficos etc.) para que uma determinada estratégia de instrução possa ser usada para diferentes conteúdos.
Objeto de Mídia – <i>Media objects</i> (SOUTH; MONSON, 2000)	Objeto de mídia é o termo designado para objetos de aprendizagem com propósitos instrucionais. Estes objetos vão desde mapas e gráficos até demonstrações em vídeo e simulações interativas.
Objetos Instrucionais – <i>Instructional objects</i> (GIBBONS; NELSON; RICHARDS, 2000)	É qualquer elemento instrucional que pode ser combinado de forma independente e temporariamente em uma nova composição para criar outro evento instrucional
Unidades de Estudo – <i>Study units</i> (KOPER, 2001)	Unidade de estudo é o elemento principal de um meta-modelo que compreende todo o processo de aprendizagem. É considerada como a menor unidade de ensino que deve satisfazer a um ou mais objetivos de aprendizagem correlatos.
Objetos de Conteúdo Compartilhável – <i>Sharable Content Object</i> (ADL, 2002)	É um objeto de aprendizagem que tem o potencial de se comunicar com o ambiente de <i>e-learning</i> ⁷ em tempo de execução. É considerado o menor nível de granularidade de um recurso de aprendizagem.
Objetos de Aprendizagem Personalizados - <i>Customized learning objects</i> (GUNAWARDENA; ADAMCHIK, 2003)	Um objeto de aprendizagem personalizado é considerado como um módulo integrado que contém o conteúdo principal, exemplos de código, perguntas de revisão e outros materiais suplementares. Pode incluir também metadados e relações semânticas com outros objetos de aprendizagem personalizados.
Objetos de Aprendizagem Reusáveis – <i>Reusable learning objects</i> (MUZIO; HEINS; MUNDELL, 2002; BOYLE, 2003; POLSANI, 2003)	É composto por dois ou mais objetos de aprendizagem independentes que estão interligados para criar uma nova composição para fornecer riqueza pedagógica e uma base significativa para a reutilização em múltiplos contextos instrucionais.
Objetos Educacionais – <i>Educational objects</i> (TAROUCO; FABRE; TAMUSIUNAS, 2003)	Objetos educacionais são quaisquer recursos, suplementares ao processo de aprendizagem, que podem ser reusados. Geralmente aplicam-se a materiais educacionais projetados e construídos concomitantemente com outros para maximizar as situações de aprendizagem na qual o recurso pode ser utilizado.
Recursos Educacionais – <i>Educational resources</i> (MCGREAL, 2004)	Recursos educacionais são materiais que podem ser utilizados na aprendizagem apoiada pela tecnologia. Com descrições de metadados apropriados, eles podem se tornar unidades modulares que podem ser reunidas em conjuntos para formar aulas e cursos.

⁷ *E-learning* – representa um ambiente de aprendizagem, cuja distribuição de conteúdos multimídias, interação social e apoio na aprendizagem são suportados pela *Web* que propicia um meio de comunicação (síncrono ou assíncrono).

Objetos Inteligentes de Aprendizagem – <i>Intelligent learning objects</i> (SILVEIRA et al., 2004)	É uma combinação de agentes inteligentes com objetos de aprendizagem que permitem gerar experiências baseadas em adaptabilidade, interatividade e interoperabilidade.
Objetos de Aprendizagem Interativos – <i>Interactive learning objects</i> (LIM; LEE; RICHARDS, 2006)	São objetos de aprendizagem que auxiliam na construção do conhecimento por meio da incorporação de características como: interação entre os atores envolvidos no processo de ensino e aprendizagem, adaptação e (re) utilização em diferentes contextos.
Objetos de Aprendizagem Genéricos – <i>Generic learning objects</i> (KOOHANG; HARMAN, 2007; ALLEN; MUGISA, 2010)	É considerado como um objeto de aprendizagem atômico que não deve impor regras para a composição de novos recursos, ou seja, quanto mais genérico for o objeto de aprendizagem mais (re) utilizado ele será.
Objetos de Aprendizagem Móveis – <i>Mobile learning objects</i> (AYALA; CASTILLO, 2008)	Objetos de aprendizagem móveis são recursos informacionais desenvolvidos para atingir um objetivo instrutivo por meio da (re) usabilidade dos objetos de aprendizagem utilizados no processo de ensino e aprendizagem. Devem estar disponíveis em qualquer hora e lugar e contemplar o acesso por meio dos dispositivos móveis.
Objeto de Aprendizagem Inteligente – <i>Smart learning objects</i> (STUIKYS, 2015)	Objetos de aprendizagem inteligentes são apresentados como uma evolução do conceito original com suporte a funções inteligentes de forma transparente para o usuário (professor, estudante). O autor apresenta esse conceito como uma tendência que acompanha o desenvolvimento da revolução tecnológica da “internet das coisas” (<i>Internet of Things</i>).

Fonte: elaborado pela autora.

Os conceitos apresentados no Quadro 2 estão interligados direta, ou indiretamente, ao conceito central de objetos de aprendizagem. Assim, com o intuito de alinhar os conceitos inerentes aos objetos de aprendizagem, optou-se por manter nesta investigação o termo padrão “objetos de aprendizagem”, pois apesar da divergência terminológica-conceitual, sua origem induz a compreensão de que esses recursos são desenvolvidos para facilitar o processo de aprendizagem dos educandos.

O domínio de objetos de aprendizagem é volumoso, pois não há como dimensionar a quantidade desses recursos que estão espalhados por repositórios ou iniciativas individuais na *Web*. Esse domínio ainda apresenta outras características relacionadas à dinamicidade e volatilidade desses recursos uma vez que novos objetos de aprendizagem podem ser adicionados, excluídos e modificados a qualquer tempo. Diante deste contexto, o processo de recuperação desses recursos se torna uma tarefa complexa visto que o acervo a ser recuperado é distribuído e composto por uma multiplicidade de formatos muito ampla.

Os objetos de aprendizagem são compostos por conteúdos elaborados a partir de diferentes tipos de formatos e elementos que visam complementar o processo de ensino e aprendizagem. Sua composição pode ser categorizada por meio de sua estrutura interna, que conforme Montilva, Rojas e Duarte (2011) pode ser:

- **Simple**: são objetos de aprendizagem cuja estrutura interna não permite a agregação de outros objetos. Seu conteúdo é formado unicamente por uma combinação de unidades de informação multimídia (textos, gráficos, imagens, áudio, vídeos etc.) e outros elementos complementares (atividades de aprendizagem, exercícios, autoavaliações);
- **Compostos**: são objetos de aprendizagem cuja estrutura interna permite a associar dois ou mais objetos simples, ou seja, permite a integração de vários objetos de aprendizagem que visam complementar uma determinada temática.

Percebe-se que a categorização indicada está diretamente relacionada com os princípios de granularidade e a (re) usabilidade do objeto de aprendizagem que foram amplamente apontadas nas definições apresentadas. Nesta mesma direção, Oton Tortosa (2006) ressalta que:

Os objetos ou unidades de aprendizagem são peças chave para a construção do material docente de forma que os conteúdos educativos se fragmentam em unidades modulares independentes que podem ser utilizados em ambientes distintos e aplicações diferentes. (tradução nossa)

Essa fragmentação ou unidade modular apresentada pelo autor reforça a importância de desenvolver os objetos de aprendizagem pensando em aplicá-los em diferentes situações educacionais. Longmire (2001) aponta outros princípios a serem considerados para concepção de um objeto de aprendizagem, que incluem:

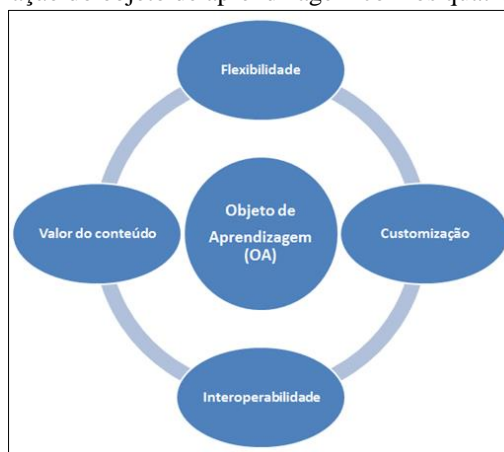
- **Flexibilidade**: o material educacional é desenhado e construído para ser usado em múltiplos contextos, devido a sua facilidade de atualização e gestão de conteúdo, para que possam ser reutilizados e recuperados através da adoção de metadados, por exemplo;
- **Facilidade de atualização, buscas e gerenciamento de conteúdo**: *Tags* de metadados permitem uma rápida atualização, pesquisa e gestão de

conteúdos, filtrando e selecionando apenas o conteúdo relevante para um determinado fim;

- **Customização:** possibilidade de modificar a sequência de conteúdos para que eles possam ser utilizados em diferentes situações, com diversas metodologias de forma personalizada, sendo assim adaptado para qualquer propósito formativo do educando;
- **Modularidade:** maximiza o potencial de personalização do conteúdo, por permitir a entrega e recombinação do material no nível de granularidade desejado;
- **Interoperabilidade:** abordagem de um objeto permite que as organizações possam definir as especificações relativas à concepção, desenvolvimento e apresentação de objetos de aprendizagem com base nas necessidades organizacionais, mantendo a interoperabilidade com outros sistemas e contextos de aprendizagem;
- **Adaptabilidade:** a capacidade de se adaptar aos diferentes estilos de aprendizagem dos alunos facilita a aprendizagem baseada em competências. Assim, quanto mais granular for um objeto de aprendizagem, melhor a possibilidade de esse objeto ser adaptado, combinado e aplicado em propósitos educacionais distintos;
- **Valor do conteúdo:** a possibilidade de reutilização dos objetos de aprendizagem possibilita que estes agreguem diversos tipos de conhecimento aumentando assim, o seu valor informacional.

Dos argumentos destacados por Longmire (2001), quatro deles são considerados como fatores chave, pois colaboram para a disseminação dos objetos de aprendizagem no âmbito educacional, conforme Figura 1.

Figura 1 – Relação do objeto de aprendizagem com os quatro fatores chave.



Fonte: SANTARÉM SEGUNDO e PÖTTKER (2015, p. 275).

Além dos princípios apresentados, foram identificadas na literatura outras características essenciais para o desenvolvimento dos objetos de aprendizagem. Essas características foram compiladas, baseadas em Santarém Segundo e Pöttker (2015, p. 275-276) e são apresentadas no Quadro 3.

Quadro 3 – Características intrínsecas dos objetos de aprendizagem

CARACTERÍSTICAS	DESCRIÇÃO
Acessibilidade	Os objetos de aprendizagem devem ser concebidos para atender as diferentes necessidades da pluralidade de usuários e precisa ser inclusivo.
Colaboração	Os objetos de aprendizagem devem incentivar a colaboração mútua, para fomentar a aprendizagem coletiva, interativa e comunicativa.
Durabilidade	Garantir a reutilização do objeto, independente da mudança de tecnologia (<i>hardware</i> ou <i>software</i>) ou atualização do recurso.
Granularidade	O conteúdo do objeto de aprendizagem deve ser apresentado em pequenas unidades (módulos) para permitir a maximização de seu uso e a reutilização desses recursos. Assim, quanto maior a granularidade do objeto, maior será sua flexibilidade na reutilização.
Indexação e Busca	Os objetos de aprendizagem precisam ser descritos em um padrão de metadados educacional para permitir que esses recursos sejam indexados e recuperados de forma relevante.
Interação	Explorar experiências de interação tanto <i>WIMP</i> ⁸ (<i>Windows, Icons, Menus, Pointers</i>) quanto <i>post-WIMP</i> ⁹ . Permite que o usuário possa interagir com o conteúdo, de alguma forma, para que o aprendizado se torne mais motivador.

⁸ *WIMP* (*Windows, Icons, Menus, Pointers*) – Modelo de interação predominante no desenvolvimento dos objetos de aprendizagem que exploram recursos das interfaces gráficas.

⁹ *Post-WIMP* – Permite a interação por meio de dispositivo não-padrão de interação, como por exemplo: reconhecimento de gestos e voz.

Reutilização	Permitir que o objeto de aprendizagem seja utilizado para diferentes propósitos educacionais, ou seja, para que possa servir como base ou complemento a outros recursos.
Usabilidade	É a facilidade com que se pode usar um objeto de aprendizagem a fim de realizar uma tarefa específica com efetividade, eficiência e satisfação.

Fonte: elaborado pela autora.

Apesar dos princípios e das características intrínsecas apontadas pela literatura serem aceitas amplamente elas não são um consenso entre todos os autores. Um exemplo dessa discordância está na característica da granularidade, na qual alguns autores desconsideram essa característica de forma isolada, ou seja, independente do contexto de criação do objeto de aprendizagem (LONGMIRE (2001); DUFRESNE; SENTENI; RICHARDS (2002)) enquanto outros reforçam uma relação muito próxima entre granularidade e a reutilização, pois consideram que quanto mais granular for um objeto de aprendizagem, maior a probabilidade de reutilização desses recursos educacionais (QUINN (2000); SOUTH; MONSON, (2000); WILEY (2001)).

Segundo Audino e Nascimento (2010), os objetos de aprendizagem são recursos importantes no processo de ensino e aprendizagem, pois fornecem a capacidade de simular e animar fenômenos e/ou comportamentos.

Neste contexto, são apresentados dois exemplos de objetos de aprendizagem (Figura 2 e Figura 3), que possuem o mesmo propósito pedagógico de simular o aprendizado do Sistema Aditivo de Cores¹⁰. Embora, ambos os recursos tenham o mesmo propósito, cada um deles utiliza uma forma distinta de interface com o usuário. Essas distinções são visualizadas e explicadas a seguir.

A Figura 2 apresenta o objeto de aprendizagem em uma interface WIMP, na qual a interação do usuário com a simulação ocorre por intermédio de um software computacional. Neste experimento o usuário simula a composição das cores por meio da seleção da intensidade (ou da ausência da cor) de cada uma das cores luz (vermelho, verde e azul).

¹⁰ Sistema Aditivo de Cores (RGB – *red, green, blue*) – é um sistema aplicado para demonstrar como ocorre a composição das cores luz em dispositivos eletrônicos.

Figura 2– Exemplo de objeto de aprendizagem com interface WIMP



Fonte: Objeto de Aprendizagem desenvolvido pelo Projeto ACESSA Física - Instituto Brasileiro de Educação e Tecnologia de Formação a Distância - IBTF¹¹

A Figura 3 demonstra o objeto de aprendizagem com uma interface *post-WIMP* que demonstra o reconhecimento de gestos na interação do usuário com o recurso educacional. Neste objeto de aprendizagem, o usuário tem a opção de ligar/desligar a cor selecionada para compor a cor resultante no frasco maior. É possível observar que o processo computacional é transparente ao usuário, pois ele não se relaciona diretamente com o computador, mas sim com sua atenção focado para o objeto em investigação, o que favorece o seu desenvolvimento cognitivo.

Figura 3 – Exemplo de objeto de aprendizagem com interface post-WIMP



Fonte: Objeto de Aprendizagem desenvolvido pelo Grupo Projeto DeVOA¹² do IFPR – Câmpus Londrina

Estes dois exemplos de objetos de aprendizagem apresentados demonstram que um mesmo conteúdo pedagógico pode ser ministrado de inúmeras formas e que por

¹¹ Projeto Condigital MEC/MCT. Disponível em:

<http://objetoseducacionais2.mec.gov.br/handle/mec/16806>. Acesso em: 04 jul. 2014.

¹² Projeto DeVOA (Desenvolvimento e Validação de Objetos de Aprendizagem) é realizado no IFPR – Campus Londrina e tem como propósito desenvolver objetos de aprendizagem colaborativos e interativos por meio de interfaces post-WIMP voltados para a área educacional e da saúde. Disponível em: <http://www.projeto devoa.com.br/>. Acesso em: 23 mar. 2015.

meio dos objetos de aprendizagem é possível estimular a cognição do educando na construção de seu próprio conhecimento.

Os objetos de aprendizagem são categorizados por Willey (2001) como:

- **Fundamental** (*Fundamental*) – é um recurso digital individual que não é combinado com nenhum outro tipo de recurso, como por exemplo, uma imagem ou uma animação;
- **Combinado-fechado** (*Combined-closed*) – é um recurso que reúne um pequeno número de recursos digitais combinados durante a criação do mesmo. As partes constituintes deste recurso não são desmembráveis, ou seja, não podem ser reutilizadas, como por exemplo, um videoclipe ou um texto com imagens ou outros elementos;
- **Combinado-aberto** (*Combined-open*) – é um recurso que reúne um grande número de recursos digitais que são combinados, em tempo de execução, durante a solicitação do objeto de aprendizagem. As partes que constituem o objeto de aprendizagem são acessíveis individualmente, seguindo o princípio da reutilização, como por exemplo, uma página *Web* com recursos multimídia;
- **Gerador de Apresentação** (*Generative presentation*) – é um recurso construído com lógica e estrutura para combinar ou gerar o objeto de aprendizagem de baixo nível (fundamental e combinado-fechado). Podem ser construídos do zero ou a partir de objetos de aprendizagem preexistentes. Possuem alto grau de reusabilidade no mesmo contexto educacional, mas são poucos reutilizáveis em outros contextos. Sua principal utilização ocorre em apresentações, exercícios e testes, como por exemplo, uma aplicação Java;
- **Gerador de Instrução** (*Generative-instructional*) – é um recurso que apresenta lógica e estrutura para combinar vários tipos de objetos de aprendizagem (fundamental, combinado-fechado e gerador de apresentação) e para avaliar as interações dos educandos com estas contribuições, buscando suportar diferentes abordagens pedagógicas, com um alto grau de granularidade.

O Quadro 4 demonstra as categorias indicadas por Willey (2001) e suas relações com as características recomendadas para a criação de um objeto de aprendizagem.

Quadro 4 – Relação das características recomendadas para o desenvolvimento de um objeto de aprendizagem

Características	Fundamental	Combinado-fechado	Combinado-aberto	Gerador de apresentação	Gerador de Instrução
Número de elementos combinados	Um	Poucos	Muitos	Poucos – Muitos	Poucos – Muitos
Tipos de objetos contidos	Único	Único, Combinação fechada	Todas	Único, Combinação fechada	Único, Combinação fechada, Gerador de apresentação
Componentes de objetos reutilizáveis	(Não aplicável)	Não	Sim	Sim / Não	Sim / Não
Funções comuns	Demonstração, exibição	Prática ou Instrução pré-desenvolvida	Prática e/ou Instrução pré-desenvolvida	Demonstração, exibição	Instrução gerada pelo computador e/ou prática
Dependência de objeto extra	Não	Não	Sim	Sim / Não	Sim
Tipo de lógica contida no objeto	(Não aplicável)	Nenhuma ou respostas baseadas em um gabarito.	Nenhuma ou instrução de domínio específico e estratégias de avaliação	Estratégias de domínio específico de apresentação	Apresentação independente de domínio, instrucional e estratégias de avaliação
Potencial para reutilização intercontextual	Alta	Média	Baixa	Alta	Alta
Potencial para reutilização intracontextual	Baixa	Baixa	Média	Alta	Alta

Fonte: adaptado de Willey (2001), tradução nossa.

A partir da análise do Quadro 4, pode-se observar que cada uma das categorias mencionadas Wiley (2001) propicia o desenvolvimento de um tipo de recurso que engloba desde o número de elementos que podem ser combinados, se existe dependência ou não de outros recursos até o nível do potencial de reutilização que estes objetos de aprendizagem terão. Estas informações são importantes uma vez que orientam o desenvolvedor desses recursos a aplicar de forma correta as diretrizes utilizadas na criação, categorização e disponibilização do OA.

Atualmente, existe um grande número de objetos de aprendizagem disponíveis na *Web*, armazenados em sites particulares ou em repositórios específicos para este fim. Na seção 2.2 é apresentado um quadro amostral destes repositórios (Quadro 5). Assim, o potencial dinâmico e a natureza multimídia dos objetos de aprendizagem dificulta o processo de recuperação eficiente desses recursos educacionais por meio de um motor de busca de uso geral na *Web* e até mesmo pelos motores de busca de cada ROA que se limita a recuperar apenas os recursos que compõe seu acervo.

Todavia, cabe ressaltar que o custo associado ao desenvolvimento de um OA, assim como o tempo de planejamento e construção desses recursos é muito elevado. A demanda por recursos humanos e materiais exige investimentos significativos nessa área. Por isso, recomenda-se que esses recursos sejam reutilizados em diferentes contextos e adaptáveis aos diferentes estilos de aprendizagem, como uma estratégia de tornar viável seu desenvolvimento e reduzir os custos de produção.

2.1 Repositórios de objetos de aprendizagem – ROA

Os repositórios digitais são ambientes *on-line* que armazenam, mantêm, gerenciam, fornecem e preservam acesso apropriado a recursos informacionais, normalmente descritos por um padrão de metadados específico. Para Santarém Segundo (2010, p. 154), os repositórios digitais, independente se são temáticos ou institucionais, possuem características comuns como, por exemplo: a facilidade de acesso à informação e a interoperabilidade entre os recursos informacionais.

Os repositórios de objetos de aprendizagem são iniciativas que seguem o mesmo propósito dos repositórios digitais só que são desenvolvidos para fins educacionais. Entretanto, os autores Tarouco, Rodrigues e Schmitt (2013) destacam que os ROA se diferenciam de outros tipos de repositórios digitais, uma vez que reúnem objetos de aprendizagem de diferentes naturezas que tem como principal característica a reutilização.

Nesses ambientes, os objetos de aprendizagem são classificados, descritos, preservados e disponibilizados para que sejam reusados em diferentes cenários. Os ROA são mantidos por iniciativas privadas ou governamentais e podem ser temáticos (referem-se a uma única área de conhecimento) ou institucionais (compreendem várias áreas do conhecimento).

Segundo Afonso et al. (2011), os repositórios de objetos de aprendizagem são sistemas de informação que têm como propósito facilitar a interoperabilidade e reutilização de recursos educacionais, por meio da disponibilização de um acervo dinâmico e que contribui para a realização de diversas práticas educacionais. Assim, a finalidade principal dos ROA é permitir que os usuários localizem os recursos

educacionais de diferentes conteúdos, níveis de granularidade, qualidades e formatos em um mesmo ambiente (MOHAN, 2007, p. 527).

De acordo com Arms et al. (2002), o objetivo da interoperabilidade é a construção de serviços coerentes para os usuários, a partir de componentes que são tecnicamente diferentes e administradas por organizações distintas. Isso requer acordos para cooperar em três aspectos:

- Técnico: abrange acordo em aspectos que permitem que os objetos de aprendizagem possam ser trocados, incluindo formatos de arquivos, protocolos etc.;
- Conteúdo: inclui acordos em termos de conteúdo e metadados, o que compreende a conformidade semântica na interpretação da informação;
- Organizacional: contempla acordos para regras básicas de acesso, autenticação, direitos autorais, pagamentos (se for o caso), entre outros.

Esses aspectos de cooperação demonstram que devem existir protocolos comuns de especificação e comunicação entre os ROA. Também é importante destacar que existe um conjunto básico de funções que os ROA devem suportar, conforme mencionado por ADL (2002):

- **Pesquisar/Encontrar** – capacidade de localizar o objeto de aprendizagem solicitado pelo usuário. Isso inclui a capacidade de navegação;
- **Solicitar** – um objeto de aprendizagem que tenha sido localizado;
- **Recuperar** – receber um objeto de aprendizagem que tenha sido requisitado;
- **Enviar** – entregar a um ROA o objeto de aprendizagem para ser armazenado;
- **Armazenar** – colocar o objeto de aprendizagem em um registro de dados com um identificador único que permita que ele seja localizado;
- **Coletar** – obter metadados de objetos de aprendizagem armazenados em outros repositórios por meio de buscas federadas;
- **Publicar** – fornecer metadados para outros repositórios.

Essas funções garantem acesso aos objetos de aprendizagem e outros ativos em um ambiente seguro e confiável. Dessa forma, os ROA atuam como um catálogo digital que visam facilitar a busca de recursos educacionais e que devem permitir: o armazenamento, o controle de versões e publicações, a busca dos recursos, o controle de acesso e a avaliação dos objetos de aprendizagem (KOOHANG; HARMAN, 2007).

Para que a interoperabilidade seja atingida nos repositórios de objetos de aprendizagem é necessário utilizar uma das seguintes técnicas: buscas federadas¹³ (do inglês, *federated search*); recuperação por meio de metadados do objeto de aprendizagem; e por meio de uma coleção de objetos de aprendizagem (ARMS et al., 2002; MASSART; DUNG, 2004; GOLDREI; KAY; KUMMERFELD, 2005).

Nesta tese, optou-se pela técnica de recuperação por meio de metadados do objeto de aprendizagem pelo fato de que se pretende recuperar objetos de aprendizagem armazenados em repositórios com arquitetura e padrões de metadados distintos e não apenas de uma coleção ou federação de repositórios de objetos de aprendizagem.

Na concepção de Downes (2002), existem dois tipos de repositórios de objetos de aprendizagem que são:

- Os que mantêm os objetos de aprendizagem e seus respectivos metadados armazenados em um mesmo local;
- Os que contêm apenas os metadados dos objetos de aprendizagem, pois os recursos educacionais estão armazenados em outro lugar e a partir das informações dos metadados é possível localizá-los.

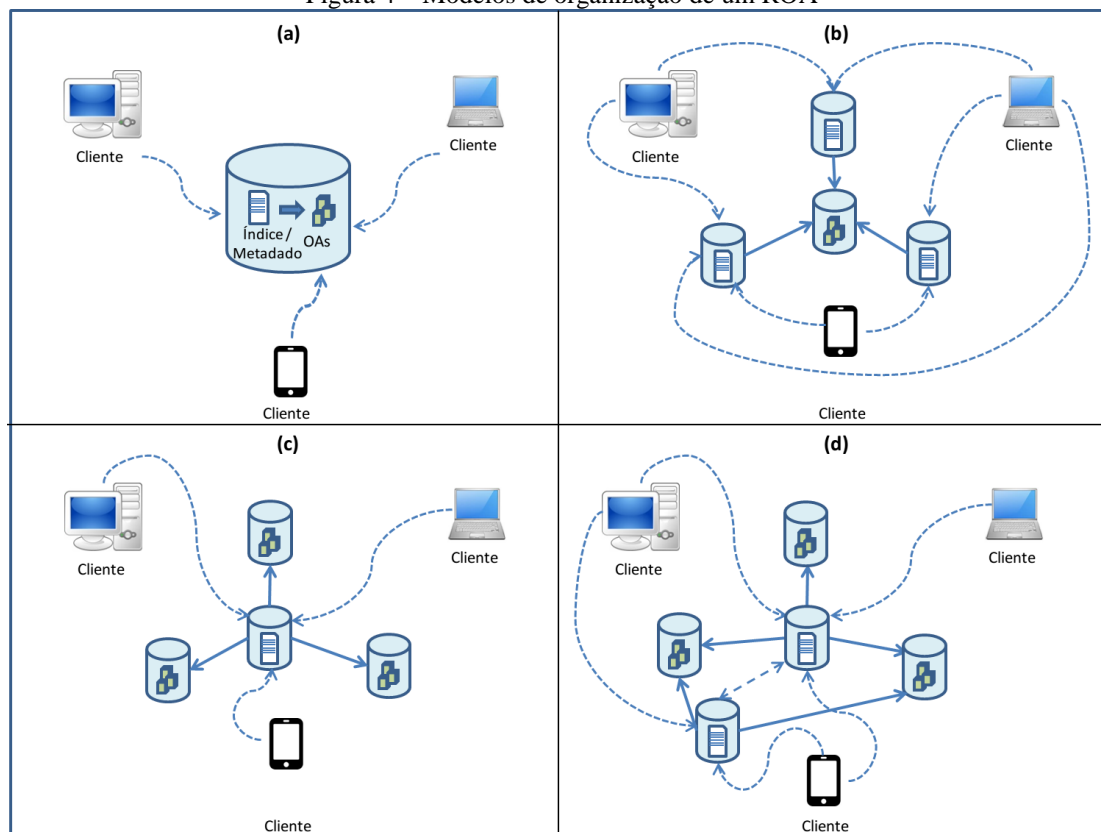
A partir desta concepção, constatou-se a necessidade de reconhecer a forma como os ROA estão organizados, em termos de arquitetura, para compreender como os metadados e os mecanismos de indexação são disponibilizados, ou seja, se estes recursos e seus metadados estão armazenados de forma centralizada ou distribuída na *Web*. Neste sentido, Silveira, Omar e Mustaro (2007, p. 134-138) apresentam quatro modelos de arquitetura para organização de um ROA, conforme apresentado a seguir:

¹³ Buscas federadas – são consideradas como uma tecnologia de recuperação de informação que permite realizar buscas simultâneas em múltiplos motores de busca. O resultado de uma busca federada agrega os resultados recebidos das diferentes fontes. É importante destacar que as federações de ROA buscam os recursos de repositórios de suas instituições parceiras e não de qualquer ROA que esteja disponível na *Web*.

- **Objetos de aprendizagem centralizados + Indexação/metadados centralizados** – A coleção de objetos de aprendizagem e os mecanismos de indexação e metadados estão localizados em uma mesma infraestrutura de servidor centralizado que é acessado por todos os usuários (Figura 4 (a)). Nesta arquitetura são observados altos custos de processamento e a escalabilidade fica comprometida, bem como o desempenho para recuperar os OA não é o ideal;
- **Objetos de aprendizagem centralizados + Indexação/metadados distribuídos** – Distribui as informações de metadados e indexação e mantém os objetos de aprendizagem centralizados (Figura 4 (b)). Assim, é possível ter diferentes padrões de metadados para referenciar o mesmo recurso educacional e torna-se mais fácil manter os índices distribuídos ordenados por assunto, o que é muito útil em repositórios institucionais. Exige uma boa infraestrutura de servidores e a indexação pode ser realizada por meio de diferentes estratégias;
- **Objetos de aprendizagem distribuídos + Indexação/metadados centralizada** – Neste modelo de arquitetura, ao invés de índices distribuídos, os próprios objetos de aprendizagem são armazenados em diferentes repositórios, mantendo a indexação centralizada (Figura 4 (c)). Esta arquitetura é comum para alguns ROA que atuam como “espelho” para garantir a redundância no acesso dos recursos;
- **Objetos de aprendizagem distribuídos + Indexação/metadados distribuída** – Nesta arquitetura, também denominada “totalmente distribuída”, é considerado qualquer número de repositórios de índices/metadados que se referem a múltiplos repositórios de objetos de aprendizagem (Figura 4 (d)). Esta arquitetura permite que a indexação seja realizada por meio de diferentes estratégias e índices distintos de objetos de aprendizagem que podem ser baseados em diversos padrões de metadados.

Esses quatro modelos são ilustrados na Figura 4 e demonstram o funcionamento de cada uma das metodologias propostas pelos autores citados para o desenvolvimento de um ROA.

Figura 4 – Modelos de organização de um ROA



Fonte: elaborado pela autora.

Atualmente existem várias iniciativas de ROA tanto em âmbito nacional quanto internacional. Fundamentado na proposta de Silveira, Omar e Mustaro (2007, p. 134-138), realizou-se um estudo amostral em sete repositórios de objetos de aprendizagem para identificar as características, assim como o padrão de metadados e a arquitetura adotada por cada ROA. O Quadro 5 sintetiza as informações identificadas após a análise dos repositórios, bem como destaca sua descrição e padrão de metadados adotado pelo ROA. Cabe ressaltar que todos os repositórios mencionados no Quadro 5 são de acesso aberto. Os padrões de metadados identificados no Quadro 5 são explicados na seção 3.1 desta tese.

Quadro 5 – Exemplos de repositórios de objetos de aprendizagem nacionais e internacionais, com respectivos padrões de metadados

REPOSITÓRIO	DESCRIÇÃO	PADRÃO DE METADADOS	ARQUITETURA
ARIADNE - Alliance of Remote Instructional Authoring and Distribution Network for	É um repositório que contém objetos de aprendizagem multidisciplinares (tanto remotos quanto locais) que abrange vários níveis de ensino. É uma instituição sem fins lucrativos que possui uma infraestrutura de tecnologia baseada em padrões	LOM com suporte aos padrões: <i>Dublin Core</i> (DC) e ISO/MLR (ISO Metadata for Learning)	OA distribuídos + indexação centralizada Figura 2 (c)

Europe	abertos e escaláveis. Foi criado em 1996 por um grupo de interesse europeu e atualmente foi expandido para uma rede de membros de interesse global. Conta com cerca de um milhão de recursos educacionais, na maioria das vezes disponíveis pelos seus metadados. A licença de uso de cada recurso é individual, sendo na sua grande maioria gratuito. É uma iniciativa internacional. Disponível em: www.ariadne-eu.org/index.php . Acesso em: 22 out. 2015.	Resources)	
BIOE – Banco Internacional de Objetos de Aprendizagem	É um repositório que contém objetos de aprendizagem de acesso livre, em vários formatos e para diversos níveis de ensino e áreas de conhecimento. Foi iniciado em 2007. É uma iniciativa nacional. Foi idealizado pelo Ministério da Educação (MEC) em parceria com o Ministério da Ciência e Tecnologia (MCT), a Rede Latino-americana de Portais Educacionais (RELPE) e algumas universidades brasileiras ¹⁴ . Este repositório permite o acesso aos metadados do objeto de aprendizagem por meio de um link de acesso quando o objeto é recuperado. Possui um acervo de 26.264 objetos de aprendizagem divididos em seis categorias de ensino. Disponível em: objetoseducacionais2.mec.gov.br/ . Acesso em: 22 out. 2015.	<i>Dublin Core</i> (DC)	OA centralizados + indexação / metadados centralizados Figura 4 (a)
CAREO – Campus Alberta Repository of Educational Objects	É um repositório centralizado de objetos de aprendizagem multidisciplinares dos professores da Universidade de Alberta no Canadá. É considerado independente, com acesso local e remoto aos recursos educacionais. Seu acervo conta com mais de 4.000 recursos educacionais, em várias áreas de conhecimento, com acesso aberto e gratuito. É uma iniciativa internacional. Disponível em: www.careo.org/ . Acesso em: 22 out. 2015.	LOM	OA distribuídos + indexação distribuída Figura 4 (d)
CESTA – Coletânea de Entidades de Suporte ao Uso da Tecnologia na	Foi criado para organizar e descrever os objetos de aprendizagem desenvolvidos pelos pesquisadores da equipe do Programa de Pós-Graduação em Informática na	Compatível com o padrão LOM	OA distribuídos + indexação centralizada Figura 4 (c)

¹⁴ Relação das Universidades Brasileiras que colaboram com o desenvolvimento do BIOE. Disponível em: objetoseducacionais2.mec.gov.br/comite-viewer. Acesso em: 16 jul. 2015.

Aprendizagem	Educação – PGIE/UFRGS e do CINTED - Centro Interdisciplinar de Novas Tecnologias na Educação. É uma iniciativa nacional. Disponível em: www.cinted.ufrgs.br/CESTA/ . Acesso em: 22 out. 2015.		
LUME – Repositório Digital da Universidade Federal do Rio Grande do Sul	O Lume é um repositório de objetos de aprendizagem produzidos no âmbito da UFRGS. Lume é um nome próprio que faz referência à manifestação de conhecimento, saber e luz. Este repositório também armazena outros documentos que, por sua área de abrangência e/ou pelo seu caráter histórico são preservados e difundidos neste espaço. É uma iniciativa nacional que conta com um acervo de 149.645 objetos de aprendizagem. Disponível em: www.lume.ufrgs.br . Acesso em: 22 out. 2015.	Dublin Core (DC)	OA centralizados + indexação / metadados centralizados Figura 4 (a)
MERLOT II – Multimedia Education Resource for Learning and Online Teaching	É considerado como uma coleção de curadoria livre, aberta e <i>on-line</i> . Seu acervo didático é multidisciplinar e contempla várias áreas do conhecimento. É um repositório centralizado que contém apenas os metadados para os objetos de aprendizagem armazenados em locais remotos. É uma iniciativa internacional e é mantido por um consórcio de várias entidades dos Estados Unidos e Canadá, com acesso aberto e gratuito. Disponível em: https://www.merlot.org/merlot/index.htm . Acesso em: 22 out. 2015.	LOM	OA distribuídos + indexação centralizada Figura 4 (c)
Portal OBAA	É um repositório disponibilizado no portal da Comunidade OBAA como uma referência aos objetos de aprendizagem que estão sendo portados e/ou desenvolvidos utilizando o esquema de metadados do OBAA. É uma iniciativa nacional mantida pela Universidade Federal do Rio Grande do Sul (UFRGS) e pela Universidade do Vale dos Sinos (UNISINOS) em parceria com a FINEP, FUNTTEL, MEC, RNP e entre outras entidades. Disponível em: repositorio.portalobaa.org/ . Acesso em: 22 out. 2015.	OBAA compatível com LOM	OA centralizados + indexação / metadados centralizados Figura 4 (a)

Fonte: elaborado pela autora.

Após a análise dos ROA, constatou-se que a grande maioria dos repositórios é idealizada a partir de demandas comuns de grupos de pesquisadores que formam consórcios engajados na construção de uma plataforma que atenda as necessidades educacionais globais e que permita o compartilhamento e a reutilização dos objetos de aprendizagem.

Portanto, percebe-se que os repositórios de objetos de aprendizagem estão em consonância com uma concepção de ensino aberto e colaborativo que incentiva a interoperabilidade e a reutilização dos recursos educacionais propiciando sua aplicação em tempos e contextos diferentes. Entretanto, observa-se que cada iniciativa é independente e, em poucos casos ocorre uma convergência que permite o intercâmbio desses recursos educacionais.

Neste capítulo foi apresentada uma visão geral a respeito dos objetos de aprendizagem. Os repositórios de objetos de aprendizagem foram demonstrados como uma alternativa para o armazenamento, descrição, preservação e busca dos recursos educacionais. Foram citadas as funções predominantes e a infraestrutura recomendada para o desenvolvimento dos repositórios, bem como as principais iniciativas, tanto em âmbito nacional quanto internacional, de repositórios de objetos de aprendizagem.

O próximo capítulo destaca a importância da adoção de um padrão de metadados educacional para a descrição dos objetos de aprendizagem. Ao final do próximo capítulo será apresentado um modelo de correlação entre os padrões de metadados para unificar a forma de representá-los e auxiliar no processo de recuperação dos objetos de aprendizagem.

3. Metadados de Objetos de Aprendizagem

O desenvolvimento de metadados¹⁵ está situado no contexto da história dos bancos de dados (BD, SGBD – Sistema Gerenciador de Banco de Dados, dados geoespaciais etc.) e, em paralelo, com o advento da *Web* e outras disciplinas. O uso de metadados tem-se mostrado de extrema importância para a sociedade da informação, uma vez que com eles é possível descrever, gerenciar, preservar e recuperar diferentes domínios de informações com a intenção de torná-los mais interoperáveis para facilitar o processo de busca e recuperação de informação.

O termo metadados é utilizado de várias formas, em diferentes áreas, para as áreas de Ciência da Informação e Ciência da Computação é considerado como uma descrição de dados que podem ser utilizados para identificar e descrever características comuns entre diferentes recursos informacionais com a finalidade de facilitar a interoperabilidade (intercâmbio de informações) e a descoberta na *Web* (ALVES, 2010, p. 59).

¹⁵ Metadados – seu prefixo “*meta*” é uma palavra grega que significa “além de”, dessa forma, pode-se interpretar a palavra metadados como um acréscimo de informação a um dado, a um recurso informacional (IFLA, 1998).

De acordo com Weibel (1995), os metadados são considerados como “dados sobre dados”, ou seja, são dados associados a recursos informacionais que descrevem outros dados. Essa definição é apresentada de forma generalizada e restringe-se ao contexto mais sintático da informação. Assim, a IFLA (1998) complementa a definição, destacando que “o termo se refere a qualquer informação utilizada para a identificação, descrição e localização de recursos”.

Com o advento da *Web* e à proporção em que seu volume informacional é incrementado, percebe-se a necessidade de propor novas formas ou métodos de tratamento da informação para representá-las e recuperá-las de forma mais eficiente. Os metadados possuem um lugar de destaque neste cenário e para melhor compreender essa dimensão são apresentadas algumas definições mais detalhadas sobre a abrangência e a importância do termo neste escopo.

Milstead e Feldman (1999) acrescentam que

[...] metadados são dados que descrevem propriedades de um recurso para diversos propósitos, como o contexto em que recurso se insere, sua qualidade, suas condições de uso, sua identificação, suas estratégias de preservação etc..

Conforme Gilliland-Swetland (1998), os metadados são “dados sobre indexação e catalogação criados para ordenar e tornar mais acessível à informação”. A IEEE LTSC LOM (2002), afirma que os metadados facilitam o compartilhamento e o intercâmbio de objetos de aprendizagem, permitindo o desenvolvimento de catálogos e inventários, levando em conta a diversidade de contextos nos quais os objetos e seus metadados podem ser reutilizados. Portanto, os componentes de metadados – a semântica, as regras de conteúdo e a sintaxe – são construtores fundamentais para manipular os recursos disponíveis na *Web* (CAPLAN, 2003).

Complementando essa visão, Méndez Rodríguez (2002, p. 21) afirma que os metadados são a informação estruturada sobre a informação distribuída. Essa definição apresentada pela autora exemplifica bem o cenário heterogêneo que compõe a *Web* e os desafios de representar a informação para que ela possa ser recuperada de maneira eficiente.

Segundo Meadow et al. (2007, p. 43) os metadados são a descrição da organização ou conteúdo de *corpus* documental, tal como: um registro ou um banco de

dados, ou seja, refere-se ao dado sobre o dado ou a informação sobre a informação. Assim, um registro de metadados tipicamente consiste de um conjunto de elementos (campos) que descrevem, em detalhes, o conteúdo e as propriedades de um determinado recurso informacional, como por exemplo: a URL de acesso ao recurso; o tipo de formato; o direito de propriedade intelectual; a data de criação, entre outros. Essa definição permite visualizar que a descrição de um objeto de aprendizagem, que é desenvolvido em diferentes formatos, é possível e pode facilitar o processo de recuperação de informação.

Segundo Alves (2005, p. 115)

[...] os metadados são conjuntos de atributos, mais especificamente dados referenciais, que representam o conteúdo informacional de um recurso que pode estar em meio eletrônico ou não. Já os formatos de metadados, também chamados de padrões de metadados, são estruturas padronizadas para a representação do conteúdo informacional que será representado pelo conjunto de dados-atributos (metadados).

Baseada nas conceituações apresentadas é possível inferir que não existe uma única definição universal, pois todas enfatizam a relevância da descrição dos recursos informacionais com a finalidade de melhorar os resultados das buscas e conseqüentemente possibilitar o processamento computacional, uma vez que os metadados são compreensíveis tanto por humanos quanto por máquinas.

As autoras Zeng e Qin (2008, p. 08) consideram que, no contexto atual, os metadados englobam as seguintes vantagens:

- Usuários podem navegar do registro de metadados até o recurso na *Web*;
- Os registros de metadados indicam a localização física do recurso informacional;
- Possibilidade de processamento automático de certos tipos de metadados minimizando possíveis custos com o processamento manual.

Haynes (2004, p. 11-12) define cinco argumentos para justificar a importância do uso de metadados e esses são resumidos como segue:

- Melhoram o desempenho de sistemas de recuperação de informação, pois estabelecem um contexto para descritores individuais e permitem

conhecer o funcionamento dos metadados para definir um mecanismo de indexação mais preciso;

- Viabilizam um caminho para o gerenciamento de objetos digitais eletrônicos, uma vez que fornecem um formato comum para o gerenciamento e manipulação do recurso;
- Podem auxiliar na determinação da autenticidade e integridade do dado;
- Permitem a interoperabilidade e o intercâmbio de dados em diferentes plataformas computacionais;
- Facilitam o acesso às informações de conteúdo da *Web*.

Diante do volume informacional heterogêneo e pouco estruturado disponível na *Web*, os metadados são utilizados para descrever e possibilitar uma localização mais eficiente dos recursos informacionais, bem como para garantir sua interoperabilidade e melhorar o desempenho dos sistemas de recuperação de informação.

Conforme sugere Day (2001) existem sete propósitos que devem ser seguidos para descrever um modelo geral de metadados:

- 1) Descrição do recurso;
- 2) Descoberta do recurso;
- 3) Administração e gerenciamento de recursos;
- 4) Registros de direitos de propriedade intelectual;
- 5) Documentação de ambiente de software e hardware;
- 6) Gerenciamento de preservação de recursos digitais;
- 7) Fornecimento de informações sobre contexto e autenticidade.

Segundo a NISO (2004), os metadados são informações estruturadas que descrevem, explicam, localizam e tornam mais fácil recuperar, usar ou gerenciar um recurso informacional. Neste sentido, são apresentadas três categorias que classificam os metadados como:

- Descritivos: descrevem um recurso com o propósito de descoberta ou identificação. Nesta categoria são incluídos elementos como título, autor, resumo e palavras-chave;
- Estruturais: indicam como objetos compostos são linkados, para garantir o princípio da granularidade;

- Administrativos: fornecem informações para auxiliar no gerenciamento de um recurso. Nesta categoria são incluídos elementos como data da criação e tipo de arquivo. Existem vários subconjuntos de dados administrativos; dois deles, às vezes, são listados separadamente como tipos metadados:
 - Metadados para gerenciamento de direitos, que tratam dos direitos de propriedade intelectual;
 - Metadados para preservação, que contêm informações necessárias ao arquivamento e à preservação de um determinado recurso.

Sob o olhar de Gilliland (2008), os metadados agregam semântica às informações com a finalidade de propiciar a busca eficiente aos recursos informacionais. Para tal, a autora classifica os metadados, de acordo com seu propósito, em cinco categorias, conforme pode ser observado no Quadro 6.

Quadro 6 – Diferentes tipos de metadados e suas funções

TIPO	DEFINIÇÃO	EXEMPLOS
Administrativo	Metadado usado no gerenciamento e administração de recursos informacionais.	<ul style="list-style-type: none"> ▪ Aquisição de informação ▪ Registro de direitos e reprodução ▪ Documentação dos requisitos legais de acesso ▪ Informação de localização ▪ Critérios de seleção para digitalização ▪ Controle de versão
Descritivo	Metadado utilizado para descrever ou identificar recursos informacionais	<ul style="list-style-type: none"> ▪ Registros de catalogação ▪ Auxílio para encontrar a informação ▪ Índices especializados ▪ Utilização de hiperlinks entre recursos ▪ Anotações pelos usuários
De preservação	Metadado relacionado à preservação de recursos informacionais.	<ul style="list-style-type: none"> ▪ Documentação da condição física do gerenciamento de preservação dos recursos ▪ Documentação das ações tomadas para preservar as versões físicas e digitais dos recursos. Exemplo: atualização e migração de dados

Técnico	Metadado relacionado ao funcionamento do sistema e ao comportamento dos metadados.	<ul style="list-style-type: none"> ▪ Documentação de hardware e software ▪ Informação digitalizada. Exemplo: formatos, índices de compressão, rotinas de escalonamento.
De uso	Metadado relacionado com o nível e tipo de uso dos recursos informacionais.	<ul style="list-style-type: none"> ▪ Registros de exibição ▪ Registros de usuários e utilização ▪ Reuso de conteúdo e informação de multiversão.

Fonte: adaptada de Gilliland (2008, p. 9, tradução nossa).

As categorias apresentadas pela autora demonstram a dimensão, em termos de descrição, que um recurso pode atingir uma vez que é possível detalhá-lo em distintos níveis de especificação durante o processo de sua descrição. Quanto mais criterioso for o processo de descrição, melhor será sua futura localização.

Segundo Prothman (2000), os metadados são divididos em dois eixos distintos: a) os metadados que descrevem a própria entidade de dados ou o objeto de informação; b) aqueles que descrevem informações sobre seu armazenamento e uso. A partir dos dois eixos observados, o autor apresenta cinco categorias de classificação:

- Metadados de acesso – são os dados que permitem consultas, navegação e recuperação de informação e como os dados são estruturados logicamente;
- Metadados semânticos – são os dados mantidos para dar propósito/significado aos dados armazenados e como os mesmos podem ser usados;
- Metadados qualitativos – são os dados mantidos para permitir/explicar análise qualitativa dos dados por dados estatísticos;
- Metadados de transferência – estes dados explicam como os dados podem ser transferidos de/entre as aplicações;
- Metadados de armazenamento – são aqueles que explicam como e onde os dados do sistema são armazenados dentro de um sistema ou repositório.

Essa categorização demonstra a preocupação do autor em dimensionar a descrição recurso informacional, desde o seu armazenamento até a forma de acesso e interoperabilidade desses recursos. Méndez Rodríguez (2002, p. 141-159) classifica os padrões de metadados de acordo com seu propósito, e os separa em dois grupos:

- Gerais: são esquemas de metadados destinados à representação e descrição de documentos eletrônicos, independentes de sua temática ou finalidade informativa, com a intenção de fornecer dados de diversos tipos sobre o recurso;
- Específicos: são esquemas de metadados desenvolvidos para um domínio específico para atender as necessidades de diferentes comunidades e ou áreas de aplicação, como por exemplo, os padrões de metadados educacionais destinados para a descrição de objetos de aprendizagem.

López Guzmán (2005) enfatiza a utilidade da adoção de metadados para descrever recursos informacionais, especialmente quando estes não são textuais, ou seja, quando seu conteúdo não pode ser indexado automaticamente pelos motores de busca, como os recursos multimídia e os de áudio, por exemplo.

Para proporcionar uma recuperação de informação mais eficiente, torna-se necessária a adoção de metadados que são utilizados para representar a informação de forma estruturada. Portanto, com o propósito de possibilitar a descoberta, reutilização e facilitar a interoperabilidade entre os objetos de aprendizagem foram definidos padrões de metadados que são úteis na descrição desses recursos educacionais para viabilizar o processo de recuperação de informação. A seguir serão apresentados os padrões de metadados educacionais adotados em âmbito mundial.

3.1 Padrões de Metadados Educacionais

Um padrão de metadados é um modelo de dados que tem sua própria sintaxe e semântica, com características muito similares a uma linguagem de marcação como XML (*eXtensible Markup Language*) ou ao RDF (*Resource Description Framework*), por exemplo. Diante deste contexto, percebe-se que ocorre uma convergência de diferentes comunidades em desenvolver seu próprio padrão de metadados, com a

intenção de descrever de forma mais completa possível seu recurso informacional para facilitar o processo de recuperação de informação.

Conforme Chowdhury (2010, p. 61) as funções primárias de um padrão de metadados são facilitar: a identificação, a localização, a recuperação, a manipulação e o uso objetos digitais na *Web*.

Diante desse contexto, percebe-se que existem várias iniciativas, como ISO (*International Organization for Standardization*), IEEE (*Institute of Electrical and Electronic Engineers*), LTSC (*Learning Technology Standards Committee*) e DCMI (*Dublin Core Metadata Initiative*) que têm reunido esforços no desenvolvimento de padrões de metadados educacionais com a intenção de garantir a interoperabilidade dos objetos de aprendizagem e permitir uma recuperação mais relevante desses recursos. Nas subseções a seguir, serão apresentados os padrões de metadados reconhecidos mundialmente para a área educacional.

3.1.1 ARIADNE educational metadata recommendation

A *ARIADNE educational metadata recommendation* foi criada com o intuito de explorar e desenvolver os resultados obtidos com os projetos europeus anteriores (*ARIADNE I* e *ARIADNE II*), os quais permitiram desenvolver ferramentas e metodologias para produção, gestão e reutilização de elementos pedagógicos assistidos por computador. Atualmente, o núcleo base de sua infraestrutura corresponde a uma rede distribuída de repositórios de aprendizagem, de mesmo nome¹⁶ (*ARIADNE*, 2015).

O esquema de metadados *ARIADNE* foi orientado para ser empregado no contexto educacional e procurava resolver dois tipos de problemas que surgiam quando um sistema de metadados é utilizado, como:

- Indexação (criação de metadados) que deve ser o mais fácil possível;
- A exploração dos metadados pelo utilizador na procura de material pedagógico relevante deve ser o mais fácil e eficiente possível.

¹⁶ *ARIADNE* Repositório de Objetos de Aprendizagem – Disponível em: www.ariadne-eu.org/index.php. Acesso em: 22 out. 2014.

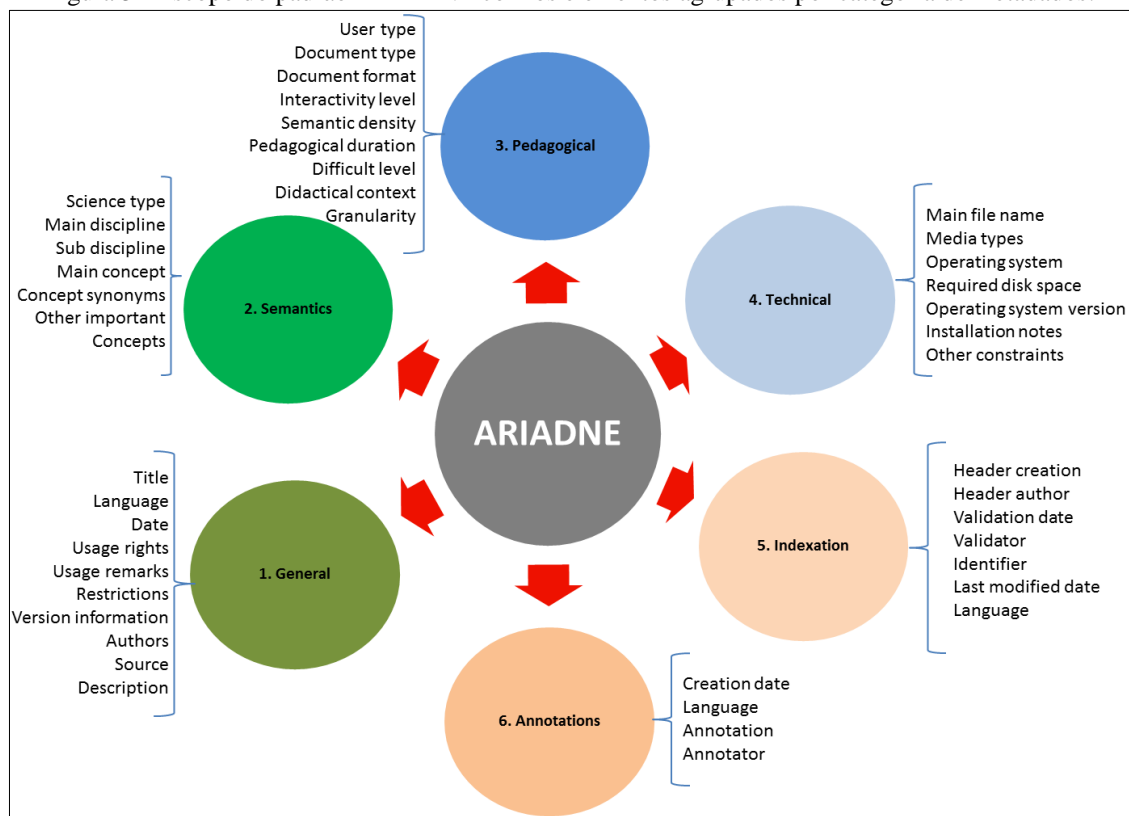
Os elementos do padrão de metadados ARIADNE foram agrupados em seis categorias distintas, de acordo com a sua função:

- Geral (*General*): informações que descrevem as características gerais do objeto de aprendizagem, como por exemplo, o título, o idioma, a data etc.;
- Semântica (*Semantics*): agrupa os elementos que descrevem o conteúdo semântico do objeto de aprendizagem, incluindo a área científica, disciplina principal, subdisciplina etc.;
- Pedagógica (*Pedagogical*): descreve as características pedagógicas e educacionais do objeto de aprendizagem, por exemplo, a densidade semântica, nível de interatividade, nível de dificuldade, granularidade etc.;
- Técnica (*Technical*): agrupa os elementos que descrevem os requisitos técnicos e características do objeto de aprendizagem, incluindo a versão do sistema operacional, espaço mínimo em disco, tipos de mídia etc.;
- Indexação (*Indexation*): descreve os próprios metadados, incluindo o seu identificador, autor, data da última alteração etc.;
- Anotação (*Annotations*): agrupa os elementos que descrevem notas de pessoas ou organizações sobre o objeto de aprendizagem, incluindo data de criação, conteúdo, idioma etc.

A Figura 5 apresenta a lista completa dos elementos (SANTOS, 2006) que compõem o esquema de metadados ARIADNE, agrupados por categoria. A distribuição dos 44 elementos das seis categorias que compõe o padrão ARIADNE possibilita observar que os recursos educacionais podem ser descritos com um nível de refinamento que engloba desde as características técnicas do recurso até as informações semânticas e pedagógicas que estão relacionadas com o objeto de aprendizagem.

Embora o projeto ARIADNE (como padrão de metadados educacional) não esteja funcional, ele foi responsável pelo desenvolvimento de um dos primeiros esquemas de metadados específicos para objetos de aprendizagem e serviu como base para o desenvolvimento do padrão LOM (ver subseção 3.1.4), alguns anos mais tarde.

Figura 5 - Escopo do padrão ARIADNE com os elementos agrupados por categoria de metadados.



Fonte: elaborada pela autora.

3.1.2 Aviation Industry CBT Committee – AICC

O padrão de metadados *Aviation Industry CBT (Computer-Based Training) Committee – AICC* foi desenvolvido para indústria de aviação americana, em 1988, por um grupo internacional para fornecer base tecnológica para formação de profissionais nessa área. Eles desenvolvem e publicam uma variedade de recomendações de instrução gerenciadas por computador (*Computer-managed instruction – CMI*). As duas perspectivas que contemplam o AICC são relacionadas às especificações:

- Técnicas de conteúdo no servidor de comunicação;
- Do servidor para disponibilidade desses conteúdos.

A missão do AICC é fornecer e promover informações, orientações e normas que resultem na aplicação efetiva da CBT e WBT (*Web Based Training*). Os objetivos do padrão AICC são:

- Auxiliar os operadores de avião na elaboração de orientações que promovam a aplicação econômica e eficaz de treinamento baseado em computador (CBT);

- Desenvolver diretrizes que permitam a interoperabilidade dos padrões de metadados;
- Disponibilizar um fórum aberto para a discussão de CBT e outras tecnologias de formação.

Segundo Oton Tortosa (2006), as principais contribuições do AICC para a área educacional são:

- **Estruturação de cursos:** tem o objetivo de definir uma estrutura para os cursos e simplificar o processo de intercâmbio de um sistema para outro. Neste sentido, o AICC contribuiu para a padronização de estruturas de cursos com um documento de orientações para interoperabilidade;
- **Ambiente de execução:** O AICC propõe a separação de CBT (que se refere ao conteúdo) e CMI (que se refere ao gerenciamento das instruções). O AICC propõe dois mecanismos, um baseado no protocolo HTTP e outro via API (*Application Programming Interface*) que permite a comunicação entre o LMS e o conteúdo.

As diretrizes do AICC abordam três aspectos de interoperabilidade com os sistemas CMI:

- A comunicação entre um sistema CMI e uma lição;
- O intercâmbio de um curso entre diferentes sistemas CMI;
- O armazenar dos dados de avaliação da lição.

O AICC não define nenhum componente (elemento) para seu padrão, apenas identifica sete arquivos (alguns opcionais) que podem ser usados para descrever a estrutura e conteúdo de um curso. Também define cinco níveis de complexidade que determinam o número de arquivos que serão requeridos em cada arquivo (BIANCO, De MARSICO; TEMPERINI, 2005). A Tabela 1 apresenta a descrição dos arquivos que compõem o AICC.

Tabela 1 – Descrição dos conteúdos e propósitos de cada arquivo do AICC

ARQUIVO	DESCRIÇÃO
Arquivo de descrição de curso (<i>Course Description File</i>)	Informação geral sobre o curso, incluindo uma descrição textual e uma estruturação geral do curso - número e tipo de elementos.

Tabela de unidade endereçável (<i>Assignable Unit Table</i>)	Informações sobre as unidades endereçáveis no curso. Cada unidade tem seu próprio registro (ou linha na tabela). As informações incluem: nome da unidade, identificação e contagem de controle para essa unidade.
Tabela de descritores (<i>Descriptor Table</i>)	Lista completa de cada elemento do curso, incluindo: unidades, blocos, objetivos e objetivos complexos. É usado como o arquivo de referência cruzada básico mostrando a correspondência dos sistemas gerais com a identificação do usuário para cada elemento.
Tabela de estrutura do curso (<i>Course Structure Table</i>)	Os dados básicos na estrutura do curso incluem todas as unidades e blocos do curso, apresentando como eles são organizados.
Arquivo de relacionamento de objetivo (<i>Objectives Relationships File</i>)	Os objetivos têm relacionamentos complexos e variáveis com outros objetivos do curso. É um arquivo opcional.
Lista de pré-requisitos (<i>Prerequisite Listing</i>)	É um arquivo que relaciona a lista de pré-requisitos em três níveis de complexidade: <ul style="list-style-type: none"> ▪ Pré-requisito simples de uma unidade ou bloco; ▪ Pré-requisito na forma de sentença lógica (“E” e “OU”); ▪ Identificação de pré-requisitos para cada modo (revisão, navegação, normal) da lição.
Requerimentos completos (<i>Completion Requirements</i>)	É um arquivo designado para permitir a especificação explícita de um bloco ou objetivo. É um arquivo essencial.

Fonte: adaptada de (BIANCO, De MARSICO e TEMPERINI, 2005), tradução nossa.

Os arquivos que compõem o metadados AICC apresentam a dimensão e a complexidade suportada pelo padrão para a descrição de recursos informacionais. Ressalta-se que o padrão AICC por ter sido desenvolvido para um fim específico (treinamentos na área de aviação) pressupõe que os arquivos garantem a interoperabilidade e a comunicação entre o ambiente de *e-learning* e o conteúdo educacional.

3.1.3 *Dublin Core* – DC

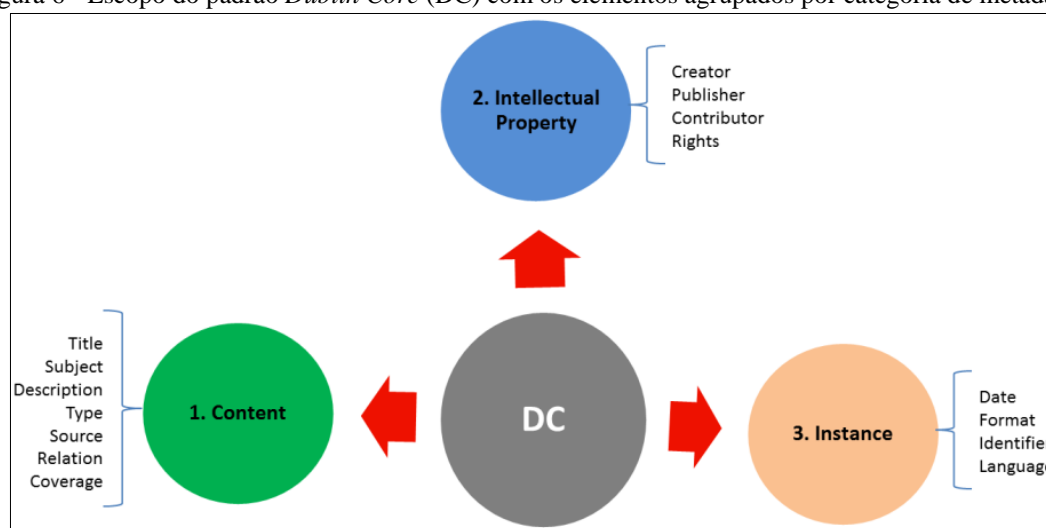
O *Dublin Core* (DC) é um padrão de metadados bastante difundido na *Web*. Consiste de um conjunto de elementos descritivos com a finalidade de facilitar a descrição de recursos digitais espalhados na *Web*. Foi desenvolvido a partir de uma iniciativa da DCMI, em 1995, com o objetivo de garantir a descrição e identificação mínima do recurso para localização e recuperação, bem como os princípios de simplicidade, extensibilidade, modularidade e interoperabilidade semântica (DCMI, 2013).

A partir do desenvolvimento do padrão de metadados DC, o termo metadados começou a ser utilizado para denominar a representação de recursos informacionais na *Web* (MÉNDEZ RODRÍGUEZ, 2002; ZENG; QIN, 2008).

Seus objetivos são: desenvolver metadados para a recuperação de informação na *Web* para diversos domínios de informação; definir padrão de interoperabilidade para conjunto de metadados; e disponibilizar conjunto de metadados específicos para uma determinada comunidade.

A DCMI é uma organização que tem por objetivo desenvolver vocabulários especializados (*metadata vocabularies*) para descrever os recursos que são recuperados em sistemas de recuperação de informação e para promover a adoção e a interoperabilidade de um conjunto de metadados. O DC, em 2003, tornou-se padrão para metadados ANSI (ANSI/NISO Z39.85) e norma ISO 15836. O DC é composto, originalmente, por 15 elementos que podem ser divididos em três categorias, conforme pode ser visualizado na Figura 6.

Figura 6 - Escopo do padrão *Dublin Core* (DC) com os elementos agrupados por categoria de metadados



Fonte: elaborado pela autora.

O Quadro 7 apresenta os elementos suportados pelo padrão DC, agrupados em suas três categorias e com a respectiva descrição do elemento. Cabe ressaltar que o padrão DC não é exclusivo para a área educacional, uma vez que é utilizado para a descrição de qualquer recurso informacional.

Quadro 7 – Elementos de dados do *Dublin Core*

GRUPO	ELEMENTO	DESCRIÇÃO
Conteúdo	dc.title	Nome do recurso.
	dc.subject	Tópico que descreve o conteúdo do recurso.
	dc.description	Sobre o conteúdo do recurso.
	dc.type	A natureza ou gênero do conteúdo do recurso.
	dc.source	Uma referência para um recurso a partir do qual o presente recurso é derivado.
	dc.relation	Uma referência para um recurso relacionado.
	dc.coverage	A extensão ou escopo do conteúdo de um recurso.
Propriedade intelectual	dc.creator	Quem é o responsável primário para a criação do conteúdo do recurso
	dc.publisher	Quem é o responsável por tornar o recurso disponível.
	dc.contributor	Quem faz contribuições para o conteúdo do recurso.
	dc.rights	Informações sobre os direitos assegurados sobre o uso recurso
Instanciação	dc.date	Data associada ao recurso.
	dc.format	A manifestação física ou digital do recurso.
	dc.identifier	Uma referência única para um recurso dentro de um contexto.
	dc.language	Uma linguagem do conteúdo intelectual do recurso.

Fonte: adaptada de Chowdhury (2010, p. 66), tradução nossa.

Além dos 15 elementos básicos do DC, também é possível utilizar o padrão de metadados DC em nível qualificado, ou seja, permite que os elementos de DC sejam refinados e apresentem valores padronizados por esquemas de codificação externos, a fim de tornar a descrição mais precisa. Os seus qualificadores são organizados em duas classes:

- Elementos de refinamento – destinados a restringir o significado de um elemento, conforme pode ser visualizado na Tabela 2;
- Esquemas de codificação – conjunto de sistemas e normas que auxiliam na interpretação do valor de um item, como por exemplo: URI (*Uniform*

Resource Identifier) e W3C-DTF¹⁷ (*W3C Date and Time Formats*). São considerados como vocabulários controlados ou anotações formais.

Tabela 2 – Elementos de refinamento do *Dublin Core*

ELEMENTO (DC)	ELEMENTO DE REFINAMENTO
dc.title	Alternative
dc.subject	
dc.description	Table of Contents Abstract
dc.type	
dc.source	
dc.relation	Is Version of Has Version Is Replaced by Replaces Is Required by Requires Is Part of Has Part Is Referenced by References Is Format of Has Format of Conforms to
dc.coverage	Spatial Temporal
dc.creator	
dc.publisher	
dc.contributor	
dc.rights	Access Rights License
dc.date	Created Valid Available Issued Modified

¹⁷ W3C-DTF - Recomendação W3C - Disponível em: <https://www.w3.org/TR/NOTE-datetime>. Acesso em: 05 set. 2015.

	Date Copyrighted Date Submitted Date Accepted
dc.format	Extent Medium
dc.identifier	Bibliographic Citation
dc.language	
dc.audience	Mediator Education Level
dc.rights holder	

Fonte: adaptada de Rouyet e Martín (2004).

Os elementos do padrão DC são todos opcionais e repetíveis. Qualquer elemento pode ser modificado por um qualificador. O núcleo do conjunto de elementos pode ser estendido com outros elementos, conforme necessário, para um determinado domínio, bem como a integração em iniciativas mais abrangentes como o XML e RDF que são tecnologias pertencentes à *Web Semântica*. A Figura 7 ilustra um fragmento de arquivo RDF utilizando o padrão de metadados *Dublin Core*.

Figura 7 – Fragmento de um arquivo RDF – *Dublin Core*

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <rdf:Description rdf:about="http://media.example.com/audio/guide.ra">

    <dc:creator>Rose Bush</dc:creator>
    <dc:title>A Guide to Growing Roses</dc:title>
    <dc:description>Describes process for planting and nurturing different kinds of rose bushes.</dc:description>
    <dc:date>2001-01-20</dc:date>

  </rdf:Description>
</rdf:RDF>
```

Fonte: disponível em: dublincore.org/documents/usageguide/. Acesso em: 04 set. 2015.

Assim, o padrão de metadados DC pode ser utilizado para várias finalidades, desde uma simples descrição de recursos (vídeo, som, imagem, texto e páginas *web*) até uma combinação de vocabulários de diferentes padrões de metadados, para proporcionar interoperabilidade. Esse padrão se caracteriza pela sua estrutura e flexibilidade, podendo ser usado para descrever qualquer tipo de recurso informacional.

3.1.4 Instructional Management Systems - Learning Resource Metadata – IMS LRM

O padrão IMS-LRM¹⁸ é mantido por uma associação americana internacional (EDUCAUSE, 2015), composta por um conjunto de universidades e também organizações comerciais e governamentais que têm como foco desenvolver especificações baseadas em XML para recursos educacionais. Foi um projeto iniciado no ano de 1997, com o propósito de introduzir o uso e a administração de tecnologias de informação aplicadas a pesquisas de ensino/aprendizagem. Seus objetivos são:

- Definir normas específicas para garantir a interoperabilidade entre aplicações e serviços em *e-learning*;
- Suportar normas internacionais para aplicação em produtos e serviços.

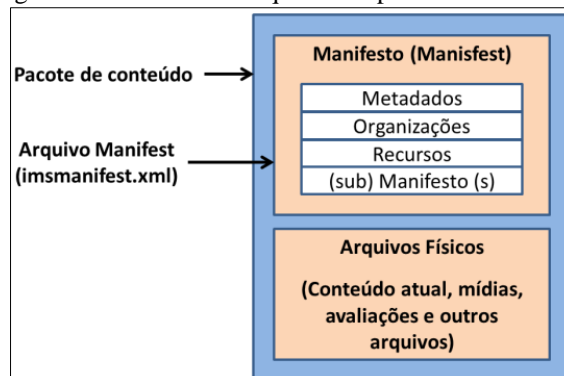
Esse padrão de metadados foi baseado nos padrões LOM e DC e consiste nas seguintes subpartes:

- Metadado (*Meta-data*): o núcleo de elementos é usado para descrever o material de aprendizagem;
- Pacote de Conteúdo (*Content Packing*): descrição da estrutura de agregação dos recursos educacionais dentro de cursos ou parte deles. A Figura 8 apresenta a estrutura desse pacote;
- Questões e testes de interoperabilidade (*Question & Test Interoperability*): um conjunto de normas que descreve como compartilhar testes simples e avaliações completas e dados. Permite representar diferentes tipos de questões, definições e computação dos resultados. Os principais tipos de questões suportadas são: respostas simples, de múltipla escolha, verdadeiro ou falso e completar lacunas;
- Projeto de Aprendizagem (*Learning Design – LD*): é a especificação de uma linguagem para modelagem de unidades de estudo. Seu propósito é fornecer um tipo de abstração para a descrição de processos de aprendizagem com diferentes finalidades pedagógicas;

¹⁸ IMS LRM – Site oficial disponível em: <https://www.imsglobal.org/> – Acesso em: 27 jul. 2015.

- Sequenciamento simples (*Simple Sequencing*): é uma especificação que descreve como organizar e apresentar os objetos de aprendizagem aos aprendizes.

Figura 8 – Estrutura de arquivos do pacote de conteúdos



Fonte: adaptada de IMS GLOBAL, 2015.

3.1.5 ISO Metadata for Learning Resources – ISO MLR

O padrão de metadados *ISO Metadata for Learning Resources* (ISO MLR – ISO/IEC 19788-4) foi desenvolvido por meio de uma iniciativa ISO/IEC JTC 1/SC 36 baseada no padrão LOM e *Dublin Core* com o intuito de aprimorar os parâmetros de descrição, incorporando capacidade multilíngue e aspectos de acessibilidade (CURRIER, 2008).

O objetivo do padrão é definir regras para descrição de metadados que facilitem a busca e recuperação de objetos de aprendizagem (ISO, 2015). O padrão consiste em onze partes que são apresentadas a seguir (ISO MLR, 2014):

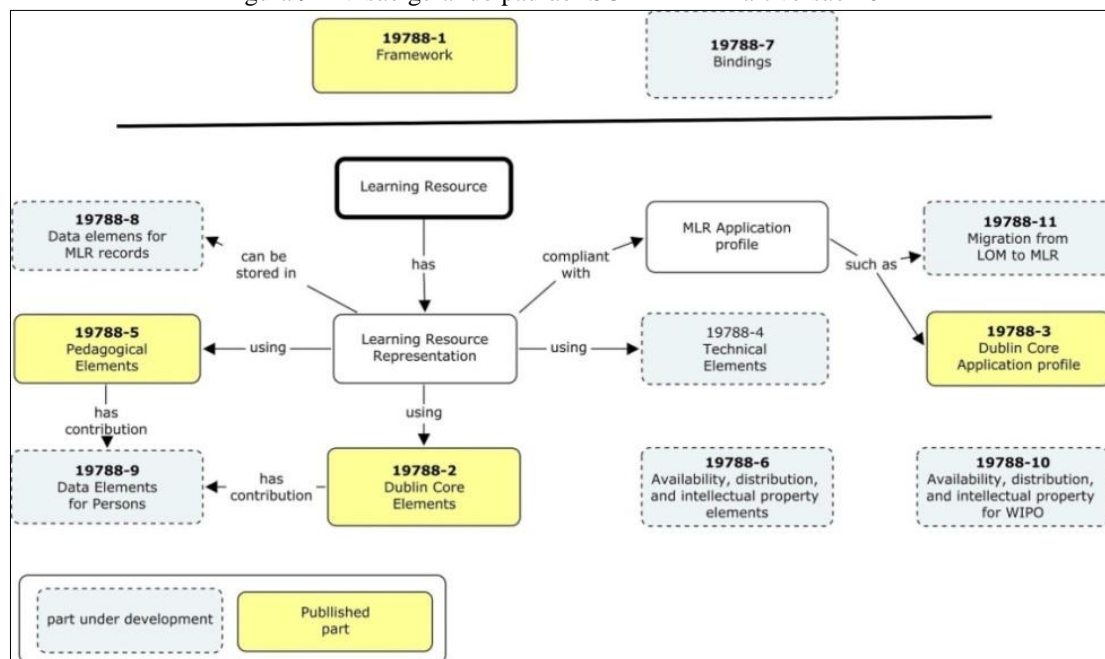
- Parte 1 – Estrutura (*Framework*) – fornece princípios, regras e estruturas para a especificação da descrição de um recurso de aprendizagem;
- Parte 2 – Elementos DC (*Dublin Core elements*) – permite o uso dos 15 elementos de dados do padrão *Dublin Core*;
- Parte 3 – Perfil de aplicação básico (*Basic application profile*) – possibilita a adição de restrições para alguns elementos de dados. Promove a interoperabilidade entre repositórios de objetos de aprendizagem;
- Parte 4 – Elementos Técnicos (*Technical elements*) – fornece informações sob a forma de elementos de metadados sobre as condições

relativas às exigências técnicas, de localização e informações de tamanho do recurso educacional;

- Parte 5 – Elementos educacionais (*Educational elements*) – facilita a descrição de elementos de dados pedagógicos que auxiliam na verificação do progresso da aprendizagem;
- Parte 6 – Disponibilização, distribuição e elementos de propriedade intelectual (*Availability, distribution and intellectual property elements*);
- Parte 7 – Associações (*Bindings*) – permite realizar mapeamentos e especificações RDF;
- Parte 8 – Elementos de dados para registros MLR (*Data elements for MLR records*) – possibilita o armazenamento de suporte da descrição de objetos de aprendizagem e o intercâmbio dos recursos educacionais;
- Parte 9 – Elementos de dados para pessoas (*Data elements for people*) – fornece elementos de dados para a descrição de pessoas (ou grupo de pessoas) que estão relacionados com a descrição de um recurso de aprendizagem;
- Parte 10 – Perfil de aplicação para acesso, distribuição e propriedade intelectual (*Application profile for access, distribution and intellectual property (WIPO compliant) elements*);
- Parte 11 – Migração de LOM para MLR (*Migration from LOM to MLR*) – permite converter, por meio da definição de regras e heurísticas, um arquivo LOM para um conjunto de elementos de dados MLR.

O *draft* da última versão foi publicado em novembro de 2014. A Figura 3.5 apresenta uma visão geral do ISO MLR. A partir dessa data, as partes 1 a 5 foram liberadas como normas internacionais. As outras partes ainda estão em fase de revisão e desenvolvimento.

Figura 9 – Visão geral do padrão ISO MLR – Draft versão 2014



Fonte: Disponível em: https://commons.wikimedia.org/wiki/File:MLR_structure_2013-07.jpg.jpg. Acesso em: 05 set. 2015.

3.1.6 Learning Object Metadata – LOM

O padrão LOM foi definido pela IEEE, uma organização internacional que desenvolve padrões e recomendações técnicas para várias áreas de conhecimento, como por exemplo: engenharia, computação, biomédica, entre outras. Em 2002, a IEEE LTSC reconheceu o padrão IEEE 1484.12.2 LOM com o intuito de descrever e classificar objetos de aprendizagem, para que os mesmos possam ser descritos e localizados por motores de busca para serem utilizados e reutilizados em diferentes contextos de aprendizagem (IEEE LTSC LOM, 2002).

Conforme Bianco, De Marsico e Temperini (2005), os objetivos do padrão LOM são:

- Permitir que discentes e docentes possam pesquisar, avaliar, adquirir e utilizar objetos de aprendizagem;
- Habilitar o compartilhamento e a troca de objetos de aprendizagem por meio de quaisquer sistemas de aprendizagem suportados por tecnologia;
- Permitir o desenvolvimento de objetos de aprendizagem em unidades (granularidade) de forma que possam ser combinadas e decompostas em maneiras significativas;

- Possibilitar que agentes computacionais possam compor automaticamente e dinamicamente aulas personalizadas para um discente individual;
- Ativar – quando desejado, a documentação e reconhecimento da conclusão dos objetivos de aprendizagem e desempenho existentes ou novos associados com objetos de aprendizagem;
- Estimular a reutilização dos objetos de aprendizagem em todas as suas formas de distribuição;
- Possibilitar a padronização independente do próprio conteúdo de aprendizagem;
- Fornecer aos pesquisadores um padrão que facilite o compartilhamento dos objetos de aprendizagem;
- Definir um padrão simples e extensível a vários domínios, de modo a ser mais facilmente adaptado e aplicado;
- Apoiar a segurança e autenticação necessárias para a distribuição e uso dos objetos de aprendizagem.

De acordo com a IEEE LTSC LOM (2002), a estrutura básica do padrão LOM para descrição de objetos de aprendizagem, ocorre por meio de campos que estão agrupados por categorias e que podem ser recuperados por qualquer sistema de classificação. Os elementos *Base Schema* do LOM v.1.0, dividem-se em nove categorias, que são descritos a seguir, na Tabela 3.

Tabela 3 – Esquema Básico LOM v.1.0

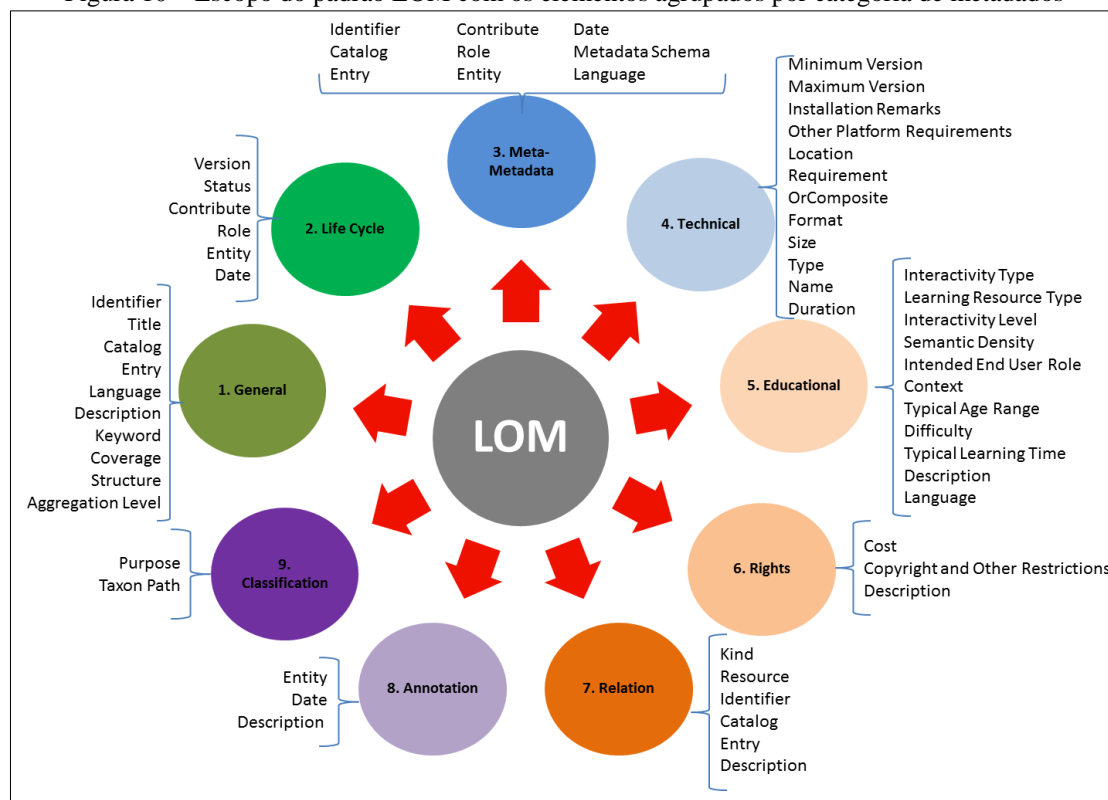
CATEGORIA	DESCRIÇÃO
Geral (<i>General</i>)	Informações gerais que descrevem o objeto de aprendizagem.
Ciclo de Vida (<i>Lyfe Cicle</i>)	Características relacionadas ao histórico, estado atual do objeto de aprendizagem e como este objeto foi afetado durante sua evolução.
Metadados (Meta – Metadata)	Informações sobre a própria instância do metadado. Não está relacionado com conteúdo do objeto de aprendizagem.
Técnica (<i>Technical</i>)	Características técnicas dos objetos de aprendizagem.
Educacional (<i>Educational</i>)	Características educacionais e pedagógicas do objeto de aprendizagem.
Direitos (<i>Rights</i>)	Direitos de propriedade intelectual e condições de uso para o objeto de aprendizagem.

Relação (<i>Relation</i>)	Características que definem a relação entre o objeto de aprendizagem e outros objetos de aprendizagem correlacionados.
Anotação (<i>Annotation</i>)	Comentários sobre o objeto de aprendizagem e prevê informações de quando e por quem foram criados os comentários
Classificação (<i>Classification</i>)	Descreve o objeto de aprendizagem em relação a um sistema de classificação.

Fonte: disponível em: <http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf> - Acesso em: 16 dez. 2013.

A estrutura básica do padrão de metadados educacional LOM V1.0 é apresentada na Figura 10, com a divisão dos elementos agrupados nas nove categorias pertencentes ao padrão.

Figura 10 – Escopo do padrão LOM com os elementos agrupados por categoria de metadados



Fonte: elaborado pela autora.

O padrão LOM especifica um cabeçalho que fornece informações sobre o objeto de aprendizagem. Os elementos que compõem esse cabeçalho são os metadados que relacionados ao OA. Assim, esse padrão não interfere no conteúdo ou nas regras dos objetos de aprendizagem, uma vez que apenas agrupa metadados. A Figura 11 ilustra um fragmento de código de um arquivo LOM.

Figura 11 – Fragmento de código de um arquivo LOM

```

...
- <lom xmlns="http://www.imsglobal.org/xsd/imsmd_v1p2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.imsglobal.org/xsd/imsmd_v1p2_imsmd_v1p2p2.xsd">
- <general>
- <title>
  <langstring xml:lang="pt">Título do objeto de Aprendizagem </langstring>
</title>
<language>pt</language>
- <description>
<langstring xml:lang="pt">Descrição textual do conteúdo do objeto educacional</langstring>
</description>
- <keyword>
  <langstring xml:lang="pt">Palavras-chave</langstring>
</keyword>
- <technical>
  <format>Exemplo: video/mpeg</format>
  <size> em bytes</size>
  <location type="URI">URL do objeto aprendizagem</location>
</technical>
- <educational>
- <interactivitytype>
- <value>
  <langstring xml:lang="x-none">pode ser Active, Expositive ou Mixed</langstring>
</value>
...

```

Fonte: elaborado pela autora.

Dessa forma, evidencia-se que o padrão LOM tem por objetivo facilitar a indexação, descrição e busca dos objetos de aprendizagem. Esta norma descreve instâncias de metadados para objetos de aprendizagem que representam as características relevantes de como os objetos de aprendizagem podem ser usados, reusados e categorizados.

3.1.7 Objetos de Aprendizagem Baseado em Agentes – OBAA

O padrão de metadados OBAA é um projeto nacional para objetos de aprendizagem compatível com o padrão LOM que permite a interoperabilidade desses objetos nas plataformas *Web*, TV Digital e para dispositivos móveis, além de suportar requisitos de acessibilidade para pessoas com necessidades especiais e registrar informações educacionais específicas do contexto brasileiro. Também especifica um modelo básico para a sintaxe e a semântica dos metadados, por meio da especificação de uma ontologia OWL (*Web Ontology Language*) para os mesmos (VICCARI et al., 2009).

Segundo Viccari et al. (2010), o projeto de “Objetos de Aprendizagem baseados em Agentes” (OBAA) foi desenvolvido pela Universidade Federal do Rio Grande do Sul (UFRGS) em parceria com a Universidade do Vale dos Sinos (UNISINOS) com o

objetivo de estabelecer uma especificação padronizada para os requisitos técnicos e funcionais de uma plataforma de produção, edição e distribuição de conteúdos digitais interativos, como os objetos de aprendizagem.

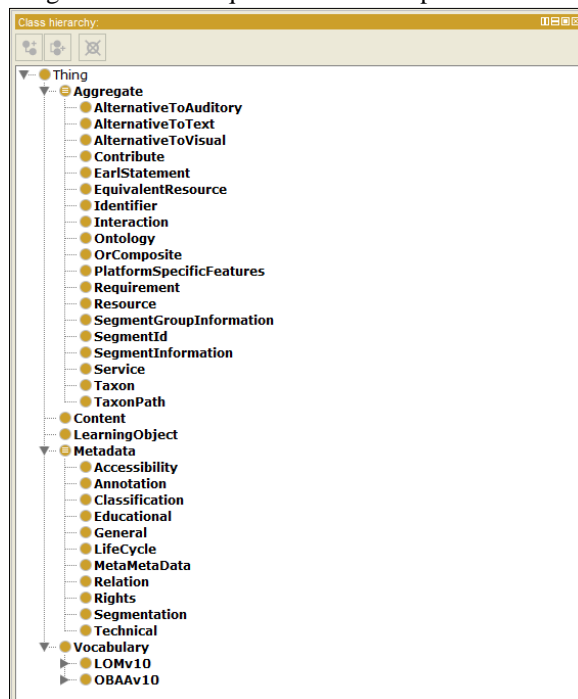
De acordo com Gluz e Viccari (2011) a definição de um padrão para objetos de aprendizagem é um ponto crítico para a implantação de qualquer plataforma educacional digital. A proposta de padrão de metadados OBAA (VICCARI et al., 2009) foi definida com base no padrão LOM (IEEE LTSC LOM, 2002), incluindo suporte para:

- Adaptabilidade;
- Compatibilidade com os padrões de metadados existentes;
- Acessibilidade;
- Independência tecnológica.

O padrão OBAA é uma extensão do padrão LOM, ou seja, o conjunto completo de metadados do OBAA é formado por todas as categorias do LOM, com mais alguns conjuntos de metadados adicionais que complementam as categorias Técnica (*Technical*) e Educacional (*Educational*); além de duas categorias novas relativas a aspectos de acessibilidade e segmentação. O mapa mental¹⁹ do padrão OBAA (Anexo A) apresenta os elementos originais do padrão LOM, bem como os elementos que foram inseridos no padrão OBAA, o que totaliza 108 elementos. A estrutura hierárquica das classes que compõem o padrão OBAA pode ser observada na Figura 12.

¹⁹ Mapa mental do padrão OBAA. Disponível em: <http://www.portalobaa.org/padrao-obaa/OBAA-v2.mm/view>. Acesso em: 16 jan. 2014.

Figura 12 – Hierarquia de classes do padrão OBAA



Fonte: Padrão OBAA – Gerado pelo software Protégé

Essa visualização foi gerada pelo software Protégé (Protégé, 2006) que é um ambiente integrado para modelagem e aquisição de conhecimento para a construção de ontologias que são parte integrante das tecnologias que compõem a *Web Semântica*. Esse software possibilita, também, a geração do código de extensão .owl para a hierarquia de classes proposta. A Figura 13 apresenta um fragmento de código .owl do metadado OBAA versão 221.

Figura 13 – Fragmento de código do Metadado OBAA

```

...
<!-- http://obaa.unisinos.br/obaa22.owl#hasPerception -->
<owl:ObjectProperty rdf:about="http://obaa.unisinos.br/obaa22.owl#hasPerception">
  <rdfs:comment>OBAA (pt-BR): Especifica a forma sensorial pela qual o aluno receberá as informações do objeto de aprendizagem.
  Vocabulário: visual (visual), auditive (auditiva), mixed (mista), other (outra)
</rdfs:comment>
</owl:ObjectProperty>

<!-- http://obaa.unisinos.br/obaa22.owl#hasPlatformSpecificFeatures -->
<owl:ObjectProperty rdf:about="http://obaa.unisinos.br/obaa22.owl#hasPlatformSpecificFeatures">
  <rdfs:comment rdf:datatype="&xsd:string">OBAA (pt-BR): Esta propriedade aponta para o conjunto de características técnicas das
  plataformas para a qual o Objeto de Aprendizagem foi previsto. Deverá ser criado um registro deste conjunto de metadados
  para cada plataforma suportada pelo OA e cujas informações técnicas diferem das informações técnicas já descritas no item 4 (Technical),
  ou seja, apenas quando mídias diferentes forem disponibilizadas para cada plataforma.
</rdfs:comment>
  <rdfs:subPropertyOf rdf:resource="&owl;topObjectProperty"/>
</owl:ObjectProperty>
...

```

Fonte: Padrão OBAA

A Figura 14 ilustra um fragmento de código do objeto de aprendizagem “Outras Infâncias - CINTED/UFRGS - Metadados” descrito com o metadados OBAA.

Figura 14 – Fragmento de código do objeto de aprendizagem “Outras Infâncias - CINTED/UFRGS - Metadados” descrito com o metadados OBAA

```

Outras Infâncias - CINTED/UFRGS - Metadados

1: Informações Gerais do Objeto
1.3: Idioma: Português do Brasil
1.4: Descrição: Destina-se a discussão sobre as infâncias
1.5: Palavras chave: Infância Trabalhadora, Infância Executiva, Infância em Situação de Risco, Infância em Situação de Rua

2: Ciclo de Vida
2.1: Versão: 1.0
2.2: Status: Final
2.3: Contribuições Para o Objeto
2.3.1: Tipo de contribuição: Criação
2.3.2: Entidade que contribuiu: GEIN - Grupo de Estudos de Educação Infantil
2.3.2: Entidade que contribuiu: NUTED - Núcleo de Tecnologia Digital aplicada à Educação
2.3.3: Data da contribuição: Não documentada

4: Informações Técnicas Gerais Sobre o Objeto
4.1: Formato: Texto XHTML, CSS, Imagens JPEG
4.2: Tamanho (Mega bytes): 1
4.3: Localização: http://gia.inf.ufrgs.br/OBAA/storage/outrasinfancias/index.jsf

4.4: Requisitos Técnicos Para Funcionalidade do Objeto

4.4.1: Opção 1:
4.4.1.1: Tipo: Web Browser
4.4.1.2: Nome: Internet Explorer
4.4.1.3: Versão mínima: 7.0

4.4.1: Opção 2:
4.4.1.1: Tipo: Web Browser
4.4.1.2: Nome: Mozilla Firefox
4.4.1.3: Versão mínima: 3.0

```

Fonte: disponível em: www.portalobaa.org/padrao-obaa/exemplos-de-oa-convertidos-para-o-obaa/objetos-interoperaveis/outras-infancias-cinted-ufrgs-metadados - Acesso em: 16 jan. 2014.

3.1.8 Síntese dos padrões de metadados educacionais

Nesta seção é apresentada uma síntese dos padrões de metadados investigados no escopo desta tese. Conforme pode ser observado ao longo deste capítulo, os objetos de aprendizagem podem ser descritos por qualquer uma das taxonomias apresentadas: ARIADNE, AICC, DC, LOM etc., pois todos os padrões cumprem com sua finalidade de descrever os recursos e facilitar a busca para que eles possam ser intercambiados e localizados de acordo com a demanda de informação do usuário.

O Quadro 8 apresenta uma síntese das principais características suportadas por cada padrão de metadados para elucidar as principais particularidades de cada padrão investigado.

Quadro 8 – Síntese dos padrões de metadados

METADADOS	COMPATIBILIDADE	IDIOMA	ELEMENTOS	PADRÃO
Ariadne	Serviu de base para o LOM, embora atualmente não seja mais adotado como padrão de metadados, caracteriza-se como repositório de objetos de aprendizagem.	Inglês	44 elementos divididos em seis categorias (<i>General, Semantics, Pedagogical, Technical, Indexation, Annotations</i>)	Aberto / Internacional
AICC	Específico para capacitação na área de aviação. Propõe uma	Inglês	Não define nenhuma categoria/ elemento para seu padrão.	Aberto / Internacional

	estruturação de cursos e do ambiente de execução (<i>e-learning</i>) para treinamentos baseados em computador e na <i>Web</i> .		Identifica sete arquivos (<i>Course Description File, Assignable Unit Table, Descriptor Table, Course Structure Table, Objectives Relationships File, Prerequisite Listing, Completion Requirements</i>) que são usados para descrever a estrutura e conteúdo de um curso.	
DC	Descrição de qualquer tipo de recurso informacional. Garante os princípios extensibilidade, modularidade e interoperabilidade semântica.	Inglês	15 elementos divididos em três categorias (<i>Content, Intellectual Property, Instance</i>)	Aberto / Internacional
IMS-LRM	Baseado em DC e LOM	Inglês	Composto por 5 partes: (<i>Meta-data, Content Packing, Question & Test Interoperability, Learning Design e Simple Sequencing</i>)	Aberto / Internacional
ISO-MLR	Baseado em DC e LOM	Inglês	Consiste de 11 partes (<i>Framework, Dublin Core elements, Basic application profile, Technical elements, Educational elements, 'Availability, distribution and intellectual property elements', Bindings, Data elements for MLR records, Data elements for people, Application profile for access, distribution and intellectual property (WIPO compliant) elements, Migration from LOM to MLR</i>)	Aberto / Internacional
LOM	Baseado em Ariadne	Inglês	76 elementos divididos em nove categorias (<i>General, Lyfe Cicle, Meta – Metadata, Technical, Educational, Rights, Relation, Annotation, Classification</i>)	Aberto / Internacional
OBA	Compatível com LOM	Inglês /	Contempla todas as	Aberto /

		Português	categorias do LOM e adiciona novos elementos para as categorias technical e educational, totalizando 108 elementos.	Nacional
--	--	-----------	---	----------

Fonte: elaborado pela autora.

Percebe-se que todos os padrões de metadados possibilitam a descrição dos objetos de aprendizagem em seus respectivos repositórios, independente do nível de descrição que cada padrão suporta. Um fator que deve ser destacado é que os padrões DC e LOM são os que possuem uma maior compatibilidade com os demais e são os mais utilizados para a descrição dos objetos de aprendizagem nos ROA, conforme estudo apresentado na seção 2.2 desta tese.

No sentido de obter uma correlação entre os padrões de metadados educacionais, as organizações e grupos de pesquisa colaboram para a criação de um padrão universal. Enquanto essa situação não se concretiza, esta investigação propõe, na seção a seguir, um mapeamento entre os principais padrões de metadados para permitir uma unificação da representação dos objetos de aprendizagem disponibilizados nos repositórios.

3.2 Correlação dos padrões de metadados educacionais

O estabelecimento de uma correlação entre os metadados educacionais é importante, pois permite determinar uma associação entre os padrões para facilitar o processo recuperação de informação relevante de fontes de dados heterogêneas.

Assim, a partir da análise dos ROA apresentados no Quadro 6 (ver seção 2.2), observou-se que os padrões de metadados mais comumente adotados nos repositórios de objetos de aprendizagem foram: DC, LOM, ISO/MLR, OBAA. Destaca-se que o padrão ISO/MLR é baseado no *Dublin Core*, enquanto que o padrão OBAA é compatível com o LOM. Diante do exposto, optou-se em adotar um padrão de mapeamento entre os padrões de metadados DC (Figura 6) e LOM (Figura 10) que será utilizado como base para recuperação dos objetos de aprendizagem descritos por diferentes padrões de metadados e mantidos em repositórios específicos.

Assim sendo, a correlação foi estabelecida tendo como base os quinze elementos padrão *Dublin Core*, pois como enfatizam Silva e Souza (2013) este é o padrão adotado

como suplemento de métodos já consolidados para pesquisa de metadados para *Web*, uma vez que permite a descrição de qualquer tipo de recurso informacional (textual e multimídia). O Quadro 9 apresenta uma correlação inicial entre os padrões DC e LOM.

Quadro 9 – Correlação inicial entre os padrões de metadados DC e LOM

<i>DUBLIN CORE (DC)</i>	LOM
dc.title	general.title
dc.subject	general.keyword
dc.description	general.description
dc.type	educational.learningresourcetype
dc.source	technical.location
dc.relation	relation.resource
dc.coverage	general.coverage
dc.creator	lifecycle.entity
dc.publisher	lifecycle.role
dc.contributor	lifecycle.contribute
dc.rigths	rights.description
dc.date	lifecycle.date
dc.format	technical.format
dc.identifier	general.identifier
dc.language	general.language

Fonte: elaborado pela autora.

Essa correlação foi obtida pela análise inicial do nível de similaridade entre os elementos dos padrões investigados. Percebeu-se que é possível estabelecer uma relação direta com todos os elementos do padrão DC. Em seguida, identificou-se que seis das nove categorias pertencentes ao padrão LOM foram contempladas nessa correlação.

Após essa correlação inicial, realizou-se uma verificação na forma como os objetos de aprendizagem estão sendo descritos nos repositórios de objetos de aprendizagem com a intenção de verificar possíveis equivalências e persistências adicionais que estão sendo aplicadas no processo de descrição.

O primeiro ROA avaliado foi o BIOE – Banco Internacional de Objetos de Aprendizagem que adota o padrão *Dublin Core*. O repositório BIOE apresenta seu

acervo dividido em seis categorias principais, que por sua vez se subdividem em outras, conforme pode ser observado no Quadro 10. Neste ROA fazem parte do acervo objetos de aprendizagem nas línguas: portuguesa (*pt*), inglesa (*en*) e espanhola (*es*). Cabe ressaltar que as últimas inserções no repositório datam do ano de 2015, conforme estatísticas fornecidas pelo repositório.

Quadro 10 - Visão geral do ROA – BIOE

CATEGORIA	SUBCATEGORIAS (Nº DE OAS)	TOTAL DA CATEGORIA
Educação Infantil	Arte Visual (30) Linguagem oral e escrita (255) Matemática (89) Movimento (32) Música (7) Natureza e sociedade (438)	851
Ensino Fundamental	Séries Finais (3440) Séries Iniciais (1628)	5068
Ensino Médio	Artes (79) Biologia (1590) Educação Física (23) Filosofia (91) Física (2192) Geografia (216) História (234) Língua Estrangeira (974) Língua Portuguesa (757) Literatura (399) Matemática (1814) Química (1723) Sociologia (197)	10289
Educação Profissional	Ambiente, Saúde e Segurança (107) Apoio Escolar (8) Controle e Processos Industriais (26) Educação Profissional: Hospitalidade e Lazer (1) Gestão e Negócios (4) Informação e Comunicação (316) Infraestrutura (0) Produção Alimentícia (32) Produção Cultural e Design (0) Produção Industrial (0) Recursos Naturais (29)	523

Educação Superior	Ciências Agrárias (1045) Ciências Biológicas (1962) Ciências da Saúde (405) Ciências Exatas e da Terra (3321) Ciências Humanas (1190) Ciências Sociais Aplicadas (172) Engenharias (144) Linguística, Letras e Artes (892) Multidisciplinar (75)	9206
Modalidades de Ensino	Educação de Jovens e Adultos (312) Educação Escolar Indígena (15)	327
Total de objetos de aprendizagem no repositório BIOE:		26.264

Fonte: elaboração própria, com base nos dados do repositório BIOE em 22 mar. 2016.

Do universo de objetos de aprendizagem constantes em cada categoria, foi selecionado o primeiro recurso de cada subcategoria para avaliação da descrição do OA. Assim, foram analisados os metadados de 40 recursos educacionais e com isso foi possível verificar a consistência de descrição dos objetos de aprendizagem com o padrão metadados.

O próximo repositório avaliado foi o ROA LUME. O Quadro 11 apresenta uma visão geral do seu acervo. Dessa forma, no escopo deste repositório foram avaliados 27 recursos educacionais.

Quadro 11 – Visão geral do ROA LUME

CATEGORIA	SUBCATEGORIAS (Nº DE OAS)	TOTAL DA CATEGORIA
Archives	CEME - Center for Memory of Sport (5881) Institute of Physics (258) UFRGS Legislation (3748) UFRGS' Museum (474) Office of Cultural Heritage (2871) UFRGS TV (811)	14043
UFRGS' Events	Teaching and Science Popularization Fair - FEPOP (139) Scientific Initiation Fair (366) Technological Innovation Fair - FINOVA (383) Education Meeting (2959) Scientific Initiation Meeting (47040)	50887

Scientific Production	Journal Articles (13622) Books and Books Chapters (389) Intellectual Property (239) Conference Proceedings, Abstracts and Papers (27020)	41270
Educational Resources	Agricultural Sciences (3) Biological Sciences (5) Health Sciences (11) Exact and Earth Sciences (4) Humanities (1) Applied and Social Sciences (4) Linguistics, Language and Arts (2)	30
Theses and Dissertations	UFRGS' Theses and Dissertations (25105) Non-UFRGS' Theses and Dissertations (240)	25345
Academic and Technical Works	Technical and Research Reports (78) Final Papers - Specialization Courses (4193) Final Papers - Undergraduate Courses (13799)	18070
Total de objetos de aprendizagem no repositório LUME:		149.645

Fonte: elaborado pela autora.

Ressalta-se que o procedimento de busca no ROA LUME foi mais complexo, pois não existe uma relação dos objetos constantes em cada categoria, como observado no repositório BIOE. Assim, utilizou-se como critérios de busca palavras-chave aleatórias, mas relacionadas com a temática de cada categoria para selecionar o primeiro recurso recuperado em cada categoria e conseqüentemente realizar a análise dos OA.

O Quadro 12 apresenta uma equivalência para os elementos de refinamento do padrão de metadados *Dublin Core*, com base na descrição dos objetos de aprendizagem do acervo dos ROA: BIOE e LUME. Observou-se que o repositório LUME utiliza alguns elementos de refinamento próprios que estão destacados no Quadro 13 pelo símbolo (*).

Quadro 12 – Refinamento na descrição dos objetos de aprendizagem padrão *Dublin Core*

DC	DC – ELEMENTO DE REFINAMENTO
dc.title	dc.title. Alternative
dc.subject	dc.subject.keyword dc.subject.category
dc.description	dc.description.abstract dc.description.conservation

	dc.description.origem* dc.description.origin dc.description.soundtrack dc.description.tableofcontents dc.description.vinculoufrgs*
dc.type	dc.type.qualificationlevel dc.type.qualificationname dc.type.work
dc.source	
dc.relation	dc.relation.isversionof dc.relation.isreferencedby dc.relation.requires
dc.coverage	dc.coverage.culture dc.coverage.temporal
dc.creator	
dc.publisher	dc.publisher.place
dc.contributor	dc.contributor.advisor dc.contributor.author dc.contributor.directorjournalistic dc.contributor.imagecapture dc.contributor.interviewee dc.contributor.role dc.contributor.screenwriter dc.contributor.technicaldirector
dc.rights	dc.rights.license
dc.date	dc.date.accessioned dc.date.available dc.date.created dc.date.issued dc.date.submitted
dc.format	dc.format.color dc.format.duration dc.format.extent dc.format.medium dc.format.mimetype dc.format.original dc.format.support

dc.identifier	dc.identifier.doi dc.identifier.uri
dc.language	dc.language.iso
dc.audience	dc.title.alternative
dc.rights holder	dc.subject.keyword dc.subject.category

Fonte: elaborado pela autora.

O mesmo procedimento adotado para obter a consistência do padrão DC pelos ROA foi realizado para o padrão LOM. Para tanto, foram analisados os repositórios: Portal OBAA, CESTA e MERLOT II.

O repositório do Portal OBAA possui um acervo de 59 objetos de aprendizagem. Cabe ressaltar que todos os registros de metadados (no formato completo) desse acervo estão vazios. Quando visualizado os metadados no formato simples, os elementos de metadados que são visualizados são: Título, Autor, URI e Data. O Anexo B apresenta um exemplo de registro de metadados no formato simplificado (resumido) e completo. Portanto, para este repositório foram avaliados os cinco exemplos disponíveis no site no projeto OBBA²⁰.

Já o repositório CESTA disponibiliza duas versões de seu repositório: a antiga e a atual. Para esse repositório em específico não foi avaliado o perfil de descrição dos OA, uma vez que o ROA disponibiliza um modelo denominado *CESTAcore*²¹ com todos os elementos de refinamento do padrão assim, estabeleceu-se a compatibilidade com o padrão LOM a partir desse modelo.

O repositório Merlot II apresenta seu acervo dividido em nove categorias principais (*Academic Support Services, Arts, Business, Education, Humanities, Mathematics and Statistics, Science and Technology, Social Sciences, Workforce Development*). Para esse repositório, utilizou-se o mesmo procedimento de seleção dos objetos de aprendizagem que foi aplicado ao ROA LUME, pois estes repositórios possuem características semelhantes, sendo analisados nove objetos de aprendizagem (o

²⁰ Exemplos de objetos de aprendizagem descritos no padrão de metadados OBAA. Disponível em: <http://www.portalobaa.org/padrao-obaa/exemplos-de-oa-convertidos-para-o-obaa>. Acesso em: 16 jan. 2014.

²¹ CESTAcore – Disponível em: <<http://www.cinted.ufrgs.br/CESTA/CESTAcore.html>>. Acesso em: 22 mar. 2016.

primeiro recuperado de cada categoria principal). Cabe destacar que o motor de busca desse repositório permite que o usuário selecione sua língua nativa para a busca no repositório e também existe a possibilidade de construção de objeto de aprendizagem por meio da ferramenta *ContentBuilder*. Apresenta uma indicação do que os outros usuários buscaram juntamente com o recurso localizado, como sugestão de visualização.

Nesse sentido, após a avaliação dos objetos de aprendizagem dos repositórios que adotam padrão LOM como base, realizou-se uma equivalência com os padrões MERLOT II, *CESTAcCore* e OBBA, conforme pode ser observado na Tabela 4. Como o padrão OBAA é nacional é possível descrever os objetos de aprendizagem tanto na língua inglesa quanto na língua portuguesa, por isso foi estabelecida a equivalência em ambas as línguas nas duas últimas colunas da Tabela 4.

Tabela 4 – Equivalência entre os padrões LOM e suas derivações

LOM	MERLOT II	CESTACORE	OBAA	
general.title	Title	Título	obaa.general.title	Geral.Título
general.keyword	Keywords	Palavras-chave	obaa.general.keyword	Geral. “Palavras chave”
general.description	Description	Descrição	obaa.general.description	Geral.Descrição
educational.learning_resourcetype	Material Type	Recurso de aprendizagem	obaa.educational.learningresourcetype	Educacional. “Tipo de Recurso Educacional”
technical.location	URL	Localização	obaa.technical.Location	“Informações Técnicas Gerais Sobre o Objeto”. Localização
relation.resource			obaa.relation.resource	Relação.Recurso
general.coverage			obaa.general.coverage	Geral.Cobertura
lifecycle.entity	Author	Entidades que contribuíram	obaa.lifecycle.entity	“Ciclo de Vida”.Entidade
lifecycle.role		Status	obaa.lifecycle.role	“Ciclo de Vida”.Função
lifecycle.contribute	Submitter	Tipo de contribuição	obaa.lifecycle.contribute	“Ciclo de Vida”. Contribuição
rights.description	Copyright	Direito Autoral	obaa.rights.description	Direitos. Descrição
lifecycle.date	Date added Date	Data	obaa.lifecycle.date	“Ciclo de Vida”.Data

	Modified			
technical.format	Technical Format	Formato	obaa.technical.format	Técnico.Formato
general.identifier		Identificador	obaa.general.identifier	Geral. Identificador
general.language	Language	Idioma	obaa.general.language	Geral.Idioma

Fonte: elaborado pela autora.

Após essa correlação inicial, realizou-se uma verificação na forma como os objetos de aprendizagem estão sendo descritos nos repositórios de objetos de aprendizagem com a intenção de verificar possíveis equivalências e persistências que estão sendo aplicadas no processo de descrição desses recursos.

A partir da análise da amostra de metadados do acervo de objetos de aprendizagem dos repositórios investigados, apresenta-se o esquema de mapeamento relacional entre os padrões de metadados *Dublin Core* e LOM com suas respectivas equivalências com os demais padrões de metadados adotados pelos repositórios investigados, conforme pode ser visualizado na Tabela 5. O símbolo de equivalência (\leftrightarrow) que aparece na Tabela 5 indica uma associação direta entre os elementos dos repositórios que adotam o padrão de metadados LOM.

Com a proposição do mapeamento entre os padrões de metadados *Dublin Core* e LOM é possível melhorar a interoperabilidade entre os repositórios de objetos de aprendizagem para garantir um melhor refinamento nas buscas realizadas nesses ambientes.

Tabela 5 – Mapeamento entre os padrões de metadados educacionais DC e LOM

DC	DC – REFINAMENTOS	LOM	LOM – REFINAMENTOS
dc.title	dc.title.alternative	general.title	obaa.general.title \leftrightarrow Title Geral.Título \leftrightarrow Título
dc.subject	dc.subject.keyword dc.subject.category	general.keyword	obaa.general.keyword \leftrightarrow Keywords Geral.“Palavras chave” \leftrightarrow Palavras-chave
dc.description	dc.description.abstract dc.description.origem dc.description.origin dc.description.vinculoufrgs	general.description	obaa.general.description \leftrightarrow Description Geral.Descrição \leftrightarrow Descrição

dc.type	dc.type.qualificationlevel dc.type.qualificationname dc.type.work	educational.learningresource type	obaa.educational.learningresource type ↔ Material Type Educatonal. “Tipo de recurso Educatonal” ↔ Recurso de aprendizagem
dc.source		technical.location	obaa.technical.location ↔ “Informações Técnicas Gerais Sobre o Objeto”. Localização ↔ URL
dc.relation	dc.relation.isreferencedby dc.relation.requires	relation.resource	obaa.relation.resource ↔ Relação.Recurso
dc.coverage	dc.coverage.culture dc.coverage.temporal	general.coverage	obaa.general.coverage ↔ Geral.Cobertura
dc.creator		lifecycle.entity	obaa.lifecycle.entity “Ciclo de Vida”. Entidade ↔ Entidades que contribuíram Author
dc.publisher	dc.publisher.place	lifecycle.role	obaa.lifecycle.role ↔ “Ciclo de Vida”.Função Status
dc.contributor	dc.contributor.advisor dc.contributor.author dc.contributor.role dc.contributor.technicaldirector	lifecycle.contribute	obaa.lifecycle.contribute ↔ “Ciclo de Vida”.Contribuição Submitter Tipo de Contribuição
dc.rights	dc.rights.license	rights.description	obaa.rights.description ↔ Direitos.Descrição Copyright ↔ Direito Autoral
dc.date	dc.date.accessioned dc.date.available dc.date.created dc.date.issued dc.date.submitted	lifecycle.date	obaa.lifecycle.date ↔ “Ciclo de Vida”.Data ↔ Data Date Added Date Modified
dc.format	dc.format.duration dc.format.extent dc.format.original dc.format.support	technical.format	obaa.technical.format ↔ Technical Format ↔ Técnico.Formato ↔ Formato
dc.identifier	dc.identifier.doi dc.identifier.uri	general.identifier	obaa.general.identifier ↔ Geral.Identificador ↔ Identificador

dc.language	dc.language.iso	general.language	obaa.general.language ↔ Language ↔ Geral.Idioma ↔ Idioma
-------------	-----------------	------------------	--

Fonte: Elaborado pela autora.

Os padrões de metadados são fundamentais para a descrição de objetos de aprendizagem, pois devido à natureza diversificada desses recursos educacionais se torna impossível recuperá-los em escala global uma vez que cada formato de recurso exigiria uma forma distinta de recuperação. Portanto, pode-se afirmar que os metadados exercem um papel fundamental na organização e localização dos objetos de aprendizagem, com a intenção de torná-los acessíveis.

Neste capítulo, foram destacadas as particularidades de cada padrão de metadados educacional, bem como foi realizado um estudo para unificar os padrões de metadados e propor um modelo de correlação entre eles. No capítulo a seguir, são apresentados os agentes inteligentes que tem a função, nesta tese, de colaborar com a recuperação dos metadados de objetos de aprendizagem disponíveis em repositórios específicos para este fim.

4. Agentes Inteligentes

Os agentes inteligentes são inseridos nesta investigação com o intuito de potencializar a busca por informações estruturadas (metadados) em diferentes repositórios de objetos de aprendizagem, para conseqüentemente melhorar a eficácia de um sistema de recuperação de informação.

No escopo da inteligência artificial (IA), os agentes inteligentes (ou racionais, como também são denominados) são considerados como sistemas computacionais que simulam o comportamento humano agindo em conformidade com determinadas situações e estímulos, ou seja, pode-se defini-los como uma entidade (semi ou completamente autônoma) que atua coerentemente de acordo com suas percepções do exterior e de seu estado de conhecimento para escolher as ações em função de suas percepções para atingir suas metas.

Russel e Norvig (2004, p. 36) afirmam que o conceito de agente inteligente está diretamente relacionado ao conceito de racionalidade que depende de quatro fatores:

- Critério de **desempenho** que define o grau de sucesso (medidas de desempenho);
- **Conhecimento** preliminar do agente sobre o ambiente (memória perceptiva);
- **Ações** que o agente pode executar (conhecimento do ambiente);

- Sequência de **percepções** do agente até o momento (o que ele pode executar).

Estes fatores conduzem a uma compreensão formal do funcionamento do agente racional, que é explicada da seguinte forma:

Para cada seqüência de percepções possível, um agente racional deve selecionar uma ação que se espera venha a maximizar sua medida de desempenho, dada a evidência fornecida pela seqüência de percepções e por qualquer conhecimento interno do agente. (RUSSEL; NORVIG, 2004, p. 36).

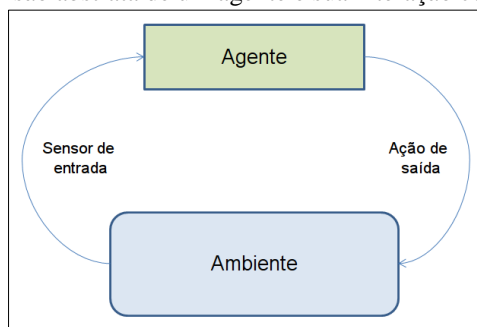
Entretanto, Russel e Norvig (2004, p. 37) destacam que a racionalidade não está relacionada com a perfeição do agente no ambiente, pois a racionalidade tende a maximizar o desempenho esperado do agente, enquanto que a perfeição (ou onisciência²²) tende a maximizar seu desempenho real, o que na prática é impossível de ser atingido.

Segundo Wooldridge (2002, p. 15, tradução nossa):

[...] um agente é um sistema computacional que está situado em algum ambiente e que é capaz de realizar ações autônomas neste ambiente para alcançar seus objetivos de projeto.

A Figura 15 apresenta uma visão abstrata de um agente, na qual se observa que a ação de saída gerada pelo agente, pode afetar o ambiente. Este é um processo cíclico no qual o agente tem uma entrada sensorial do ambiente e produz como saídas às ações que o afetam.

Figura 15 – Visão abstrata de um agente e sua interação com o ambiente.



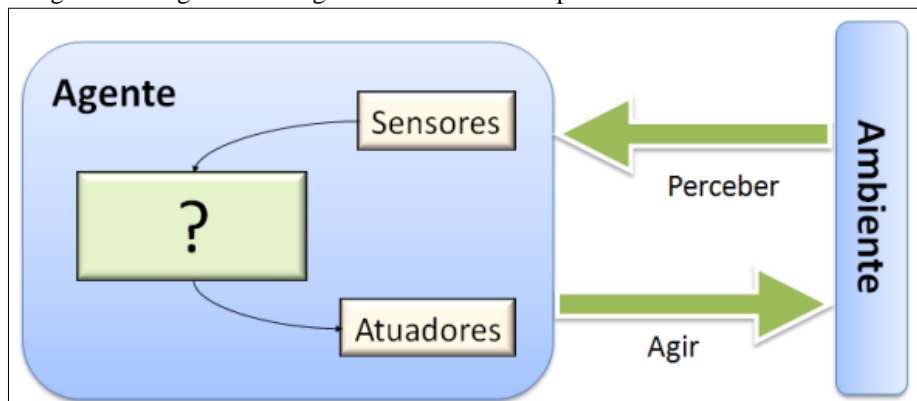
Fonte: elaboração própria, com base em Wooldridge (2002, p. 16).

²² Onisciência é a capacidade de saber tudo incessantemente. Um agente onisciente deve saber o resultado real de suas ações e pode agir de acordo com ele; porém, a onisciência é impossível na prática, pois não é possível saber tudo *a priori*. (RUSSEL e NORVIG, 2004, p. 37).

Com base na Figura 15, pode-se inferir que a saída de ação gerada pelo agente para afetar seu ambiente e como a maioria dos domínios possui uma complexidade razoável, um agente não terá controle total sobre o ambiente em que atua. Dessa forma, ele terá controle parcial, na melhor das hipóteses, à medida que as ações podem influenciar o ambiente. Portanto, do ponto de vista do agente uma mesma ação pode ser realizada duas ou mais vezes em circunstâncias aparentemente idênticas, mas que podem ter efeitos totalmente distintos, ou seja, ele pode deixar de agir conforme o esperado para o ambiente em questão.

De acordo com Russel e Norvig (2004, p. 33), “um agente é tudo o que pode ser considerado capaz de perceber seu ambiente por meio de sensores e de agir sobre esse ambiente por intermédio de atuadores”. Assim, os agentes têm a função de interagir com o ambiente de tal forma que tenham capacidade de reunir informações sobre o meio em que estão inseridos, para que estejam aptos a tomar decisões baseados nestas informações e iniciar execuções específicas conforme a ação percebida neste ambiente. Deste modo, esses agentes percebem o ambiente por meio de sensores e agem por meio de atuadores, conforme pode ser observado na Figura 16.

Figura 16 – Agentes interagindo com o ambiente por meio de sensores e atuadores



Fonte: adaptada de Russel e Norvig (2004, p. 34).

Na concepção de Hayes-Roth (1995), os agentes inteligentes realizam continuamente três funções:

- Percebem as condições dinâmicas do ambiente;
- Agem para alterar as condições do ambiente;
- Raciocinam para interpretar as percepções, resolver problemas, realizar inferências e determinar ações a serem executadas.

De acordo com Huhns e Singh (1997) a autonomia dos agentes pode ser classificada em cinco tipos diferentes:

- **Absoluta:** o agente possui controle completo sobre suas percepções, raciocínio, ações e é pouco previsível;
- **Social:** o agente conhece os outros agentes presentes no sistema e é sociável exercendo, no entanto, a sua autonomia em certas circunstâncias;
- **De interface:** em grande parte dos sistemas, onde a autonomia absoluta é impossível, a autonomia máxima possível para o agente é a autonomia relacionada com sua forma de interface com o exterior;
- **De execução:** a liberdade que o agente possui na execução de ações no ambiente;
- **De projeto:** é o grau de autonomia dos projetistas do agente na sua construção, pois quanto maior for à autonomia de projeto, maior poderá ser a heterogeneidade dos agentes.

A autonomia dos agentes inteligentes reflete diretamente no comportamento que o agente vai ter no ambiente. A seguir são apresentadas algumas características desejáveis a qualquer agente de software (WOOLDRIDGE, 2002, p.23):

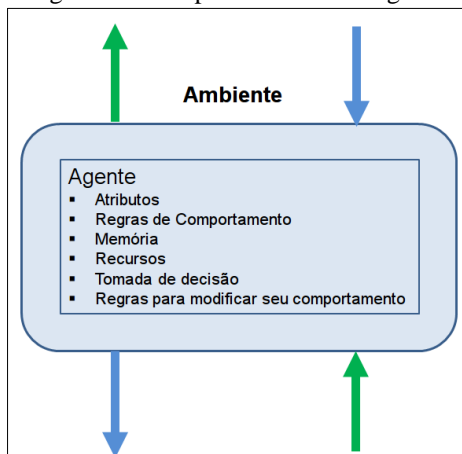
- **Reatividade** – os agentes percebem o seu entorno e podem reagir a várias formas de estímulos;
- **Autonomia** – os agentes devem apresentar a capacidade de operar sem a intervenção humana e de outros agentes e devem ter algum tipo de controle sobre suas ações e seu estado interno;
- **Comunicabilidade** – os agentes podem agir em sociedade com humanos e com outros agentes. A comunicação entre agentes ocorre por meio da troca de mensagens, usando linguagens de comunicação entre agentes, denominadas de ACL (*Agent Communication Language*);
- **Proatividade** – os agentes não atuam simplesmente em resposta ao seu entorno, mas também devem ser capazes de tomar decisão para se adaptarem as mudanças e se ajustarem a metas definidas;
- **Confiabilidade** – para a aceitação do agente é necessário um alto nível de confiança de que o agente pode representar exatamente o usuário;

- **Adaptabilidade** – é a capacidade do agente de modificar seu comportamento em função de seu aprendizado com experiências anteriores;
- **Racionalidade** – capacidade do agente de raciocinar sobre os dados obtidos para obter a melhor solução possível;
- **Continuidade temporal** – o agente é um processo executado continuamente ao longo do tempo, ou seja, é persistente;
- **Flexibilidade** – reside na habilidade dos agentes de escolher dinamicamente as ações e a sequência de execução das mesmas, em resposta a um estado do ambiente;
- **Aprendizagem** – o agente deve ser capaz de aprender baseado em experiências prévias adquiridas ao reagir a mudanças em seu ambiente.

Macal e North (2006) evidenciam a importância das características dos agentes inteligentes e apresentam as seguintes considerações para a modelagem dos agentes (Figura 17):

- **Ser identificável** – como uma entidade discreta que possui um conjunto de características e regras que regem seu comportamento para tomada de decisão. Devem ser autossuficientes;
- **Estar situado** – reside em um ambiente que interage junto com outros agentes. Os agentes têm protocolos para comunicação com outros agentes e têm a capacidade de reconhecer e distinguir características de outros agentes;
- **Ser orientado a objetivo** – tem metas a serem atingidas com relação ao seu comportamento. Um agente pode comparar o resultado do seu comportamento em relação aos seus objetivos;
- **Ser autônomo e autodirigido** – um agente pode funcionar de forma independente em seu ambiente e nas suas relações com outros agentes, pelo menos, ao longo de um número limitado de situações que são de interesse;
- **Ser flexível** – ter a capacidade de aprender e adaptar seu comportamento com base na experiência. Isto requer alguma forma de memória. Um agente pode ter regras que modificam as suas regras de comportamento.

Figura 17 – Propriedades de um agente.



Fonte: adaptada de Macal e North (2006).

A Figura 17 ilustra as considerações citadas pelos autores, elenca as propriedades que contribuem para o comportamento dos agentes e que estão diretamente relacionadas com as características do ambiente em que o agente vai atuar para solucionar o problema que lhe foi proposto.

A partir das caracterizações analisadas na literatura, observou-se que existe conformidade em pelo menos três aspectos: autonomia (citada por todos os autores), adaptabilidade (em função de seu aprendizado anterior para alcançar os objetos) e a racionalidade (para auxiliar no processo de tomada de decisão).

Portanto, no escopo desta tese, um agente será considerado como uma entidade autônoma, adaptável, racional e também pró-ativa para que o agente possa se adaptar as mudanças de um ambiente tão heterogêneo quanto são os repositórios de objetos de aprendizagem na *Web*.

Neste sentido, percebe-se que os agentes inteligentes surgem como uma tecnologia que aliada às técnicas de recuperação de informação têm potencial de proporcionar uma busca mais relevante e, conseqüentemente, apresentam um melhor desempenho nos resultados apresentados aos usuários com intuito de atender de forma mais relevante sua necessidade de informação.

Luck e D’Inverno (2001) propõe uma hierarquia de quatro níveis para representar um ambiente de software suportado por agentes inteligentes que é composta por: entidades, objetos, agentes e agentes autônomos. A ideia básica desta hierarquia é modelar um ambiente que corresponde a entidades, algumas das quais são objetos.

Desse conjunto de objetos, alguns são agentes, e destes agentes, alguns são agentes autônomos. Percebe-se que os agentes são entidades autônomas que são capazes de exercer sua escolha sobre suas ações e interações. Os agentes não podem ser chamados diretamente de objetos. Entretanto, eles podem ser construídos utilizando tecnologia de objetos.

O ambiente, no qual atuam os agentes inteligentes, pode ser classificado de acordo com seu: desempenho (*Performance*), ambiente (*Environment*), atuadores (*Actuators*) e sensores (*Sensors*). Essa classificação é reconhecida pelo acrônimo PEAS e permite a caracterização do agente de acordo com a função que ele executará. Assim, os ambientes de tarefas são fundamentalmente os “problemas” para os quais os agentes racionais são as “soluções”. Portanto, o tipo de ambiente interfere diretamente no projeto apropriado para o programa do agente. O Quadro 13 exemplifica alguns tipos de agentes e suas respectivas peculiaridades com o ambiente PEAS.

Quadro 13 – Tipos de agentes e seus respectivos PEAS

TIPO DE AGENTE	PERFORMANCE (DESEMPENHO)	ENVIRONMENT (AMBIENTE)	ACTUATORS (ATUADORES)	SENSORS (SENSORS)
Controlador de refinaria	Maximizar pureza, rendimento e segurança	Refinaria e operadores	Válvulas, bombas, aquecedores e mostradores	Sensores de temperatura, pressão e produtos químicos
Motorista de táxi	Viagem segura, rápida, em conformidade com a Lei e confortável. Maximizar os lucros	Estradas, outros tipos de tráfego, pedestres e clientes	Direção, acelerador, freio, sinal, buzina e visor.	Câmeras, sonar, velocímetro, GPS, Hodômetro, sensores de motor e teclado.
Robô de seleção de peças	Porcentagem de peças em bandejas corretas	Correia transportadora com peças e bandejas	Braço e mãos articulados	Câmera, sensores angulares articulados
Sistema de diagnóstico médico	Paciente saudável, minimizar custos e processos judiciais	Paciente, hospital e equipe	Exibir perguntas, testes, diagnósticos, tratamentos e indicações	Entrada pelo teclado para sintomas, descobertas, respostas do paciente

Fonte: adaptada de Russel e Norvig (2004, p. 40).

Os exemplos do Quadro 13 demonstram a quantidade de cenários reais que podem ser resolvidos com o uso de agentes inteligentes. Essa complexidade de ambientes heterogêneos exige soluções específicas para cada necessidade.

As propriedades do ambiente são classificadas em seis dimensões que representam a variedade de problemas que podem ser solucionados com o uso dos agentes inteligentes aplicados em ambientes heterogêneos. O Quadro 14 apresenta essas dimensões com suas respectivas características (RUSSEL; NORVIG, 2004, p. 41-42).

Quadro 14 – Propriedades de um ambiente de tarefas

PROPRIEDADE	CARACTERÍSTICAS
Completamente observável x parcialmente observável	Os sensores do agente detectam todos os aspectos que são relevantes para a escolha de sua ação; por sua vez a relevância depende da medida de desempenho. Por essa razão o agente não precisa manter qualquer estado interno sobre o mundo. A presença de ruídos e de sensores imprecisos (ou ausentes) pode tornar esse ambiente parcialmente observável.
Determinístico x estocástico	Um ambiente determinístico é aquele completamente determinado pelo estado atual e pela ação executada pelo agente, caso contrário, ele é considerado estocástico. Em princípio, um agente não precisa se preocupar com a incerteza em um ambiente completamente observável e determinístico. Porém se o ambiente for parcialmente observável, ele poderá parecer estocástico. Por isso um ambiente deve ser classificado como determinístico ou estocástico sempre do ponto de vista do agente. Entretanto, se o ambiente for determinístico exceto pela ação de outros agentes, ele será considerado como estratégico.
Episódico x sequencial (ou paralelo)	A experiência do agente é dividida em episódios atômicos, ou seja, cada episódio consiste na percepção do agente e, em seguida, na execução de uma única ação. Assim, o episódio seguinte não depende das ações de episódios anteriores. Já no ambiente sequencial, a decisão atual pode afetar todas as decisões futuras. Portanto, ambientes episódicos são muito mais simples que ambientes sequenciais, pois o agente não precisa pensar à frente.
Estático x dinâmico	Se o ambiente puder se alterar enquanto um agente está deliberando, ele é considerado como um ambiente é dinâmico para esse agente; caso contrário, ele é estático. Ambientes estáticos são fáceis de manipular, pois o agente não precisa observar o mundo enquanto está decidindo, nem se preocupar com a passagem do tempo. Se o próprio ambiente não mudar com a passagem do tempo, mas o nível de desempenho do agente se alterar, o ambiente é considerado como semidinâmico.
Discreto x contínuo	A distinção entre discreto e contínuo pode se aplicar ao estado do ambiente, ao modo como o tempo é tratado e ainda às percepções e ações do agente. Assim, no ambiente discreto existe um número limitado e claramente definido de percepções, ações e estados enquanto que no ambiente contínuo existe um número possivelmente infinito de percepções, ações e estados.
Agente único x multi-agentes	Quando um único agente opera no ambiente. Quando vários agentes interagem em um mesmo ambiente é denominado de multi-agentes, também conhecido como sistemas multi-agentes.

Fonte: elaborado pela autora.

Após análise das dimensões propostas no Quadro 14, torna-se claro que a compreensão dessas categorias permite determinar a aplicabilidade do agente em sua implementação que vai desde a forma de acesso ao ambiente até o número de agentes suportado por ele. Percebe-se que o cenário mais crítico a ser resolvido por agentes inteligentes é ambiente composto pelas seguintes propriedades: observável, estocástico, sequencial, dinâmico, contínuo e multi-agentes por serem as dimensões mais complexas de serem atingidas.

4.1 Classificação dos agentes inteligentes

Os agentes inteligentes agem de acordo com os objetivos delimitados e os estímulos recebidos durante sua existência temporal. Dessa forma, os agentes inteligentes podem ser classificados de acordo com seu comportamento, da seguinte forma (RUSSEL; NORVIG, 2004, p. 46-53):

- **Agente reativo simples:** são os agentes mais simples que selecionam suas ações baseadas apenas na percepção atual. Seu funcionamento se baseia em regras de condição-ação (*if* <condição> *then* <ação>). Apresenta como vantagem uma representação inteligível, modular e eficiente para as regras de condição-ação. Uma desvantagem desse tipo de agente é que a tabela de regras condição-ação pode ser tornar excessiva para resolver problemas complexos e possui pouca autonomia. Indicado para ambientes completamente observáveis, episódicos e pequenos. Possui um reflexo imprescindível para ambientes dinâmicos;
- **Agente reativo baseado em modelos:** são os agentes mais complexos que os reativos simples. São baseados em modelos de representação do “mundo” em estados possíveis, ou seja, o agente deve manter um estado interno que dependa do histórico de percepções e reflita os aspectos não observados no estado atual. Mundo, neste caso, refere-se à descrição completa do ambiente em que um agente se encontra. Indicado para ambientes parcialmente observáveis, determinísticos e pequenos;
- **Agente baseado em objetivos:** são agentes que expandem a capacidade dos agentes baseados em modelo, por meio de seus objetivos que descrevem as situações desejáveis. A seleção das ações baseadas em

objetivos pode ser obtida de duas formas: direta (quando o resultado de uma única ação atinge o objetivo) e complexa (quando são necessárias longas sequências de ações para atingir o objetivo). Para encontrar os objetivos do agente são utilizados algoritmos de busca e planejamento, pois a tomada de decisão implica na consideração de futuro, o que não ocorre com os agentes reativos simples (regras de condição-ação). De acordo com Steiner (1996), um agente baseado em objetivos representa o conhecimento da maneira como o mundo pode ser transformado por meio da execução de ações. Assim, pode-se afirmar que estes agentes são mais flexíveis pelo fato de que eles podem inferir em reações no ambiente;

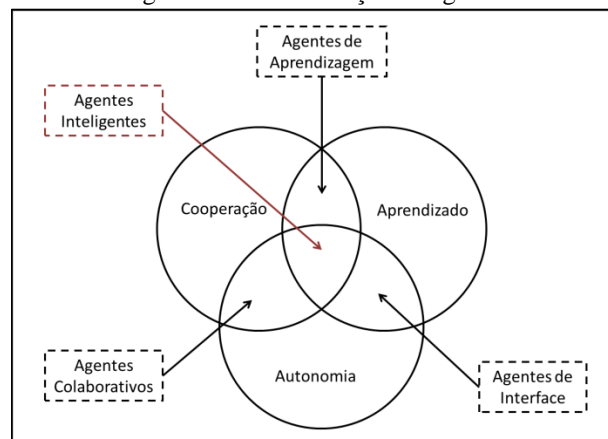
- **Agente baseado na utilidade:** são agentes que buscam definir seu grau de satisfação com os estados, ou seja, se um estado do mundo é mais desejável que outro, então ele terá uma maior utilidade para o agente. Uma função de utilidade mapeia um estado (ou uma sequência de estados) em um número real que descreve o grau de satisfação com este estado. São indicados para projetos no qual a tomada de decisão opera com a incerteza. Não tem restrições relacionadas com ambientes e apresenta a desvantagem de não ser adaptável;
- **Agente com aprendizagem:** são agentes que podem atuar em ambientes inicialmente desconhecidos e podem se tornar mais eficientes do que seu conhecimento prévio poderia permitir. Podem ser divididos em quatro componentes conceituais: elemento de aprendizado (responsável pela execução de aperfeiçoamento); elemento de desempenho (responsável pela seleção das ações externas); crítico (responsável por realimentar o elemento de aprendizagem, informando o *status* de seu funcionamento e por determinar de que maneira o elemento de desempenho precisa ser modificado para funcionar melhor no futuro) e, por último, gerador de problemas (responsável por sugerir ações que levarão a experiências novas e informativas). Não tem restrições relacionadas com ambientes e apresenta a vantagem de ser adaptável.

Esses cinco tipos de agentes demonstram as diferentes maneiras de comportamento e como estes executam suas ações diante de uma sequência de percepções no ambiente ao qual atuam.

Segundo Nwana (1996) a tipologia de agentes identifica outras seis dimensões de classificação, que são:

- **Mobilidade** – os agentes podem estar estáticos ou serem móveis (neste caso, podem estar residente em sua máquina de origem ou temporariamente em outra máquina);
- **Raciocínio** – presença (ou não) de um modelo de raciocínio simbólico, como deliberativo ou reativo;
- **Função do agente** – de acordo com sua função principal, como por exemplo, agente de busca de informação na *Web* ou como agente de interface para facilitar a interação homem-computador;
- **Característica híbrida** – combinam duas ou mais abordagens em um único agente;
- **Atributos ideais e primários** – como autonomia, cooperação, aprendizagem. A partir dessas características, são derivados quatro tipos de agentes: colaborativo, de aprendizagem, de interface e inteligentes (conforme Figura 18).
- **Atributos secundários** – tais como versatilidade, benevolência, veracidade, confiabilidade, continuidade temporal, habilidade de falhar e qualidades mentais e emocionais (BRADSHAW et. al, 1997).

Figura 18 – Classificação de agentes



Fonte: adaptada de Nwana (1996)

Cabe destacar que na proposta apresentada por Nwana, estas dimensões não podem ser interpretadas de forma isolada, mas sim como um complemento aos tipos de agentes. Na prática, os agentes colaborativos tem mais ênfase na cooperação e na autonomia do que os agentes com capacidade de aprendizagem, o que não exclui a possibilidade dos primeiros desenvolverem capacidades de aprendizagem.

Wooldridge e Jennings (1995, p. 119) classificam os agentes sob dois pontos de vista:

- **Noção fraca de agentes** – contempla a maior parte dos agentes. Inclui propriedades como: autonomia, comunicação, reatividade, pró-atividade etc.
- **Noção forte de agentes** – pressupõe a adição de propriedades e, normalmente, precisa ser complementada por conceitos aplicados a ações humanas, tais como crenças, desejos, intenções e obrigações. Portanto, a habilidade de perceber e aprender sobre um problema para resolvê-lo, substitui a ação e intervenção humana.

No contexto desta investigação, serão utilizados agentes autônomos, com capacidade de aprendizagem e adaptação para possibilitar a localização e a recuperação dos metadados de objetos de aprendizagem armazenados em repositórios. De acordo com as classificações apresentadas nesta seção, os agentes implementados nesta arquitetura são categorizados como agentes inteligentes, conforme Figura 18 de Nwana (1996) e segundo Russel e Norvig (2004) como agentes com aprendizagem. De acordo com a concepção de Wooldridge e Jennings (1996, p. 119) possuem noção fraca de agente.

4.2 Sistemas Multi-agentes – SMA

Os Sistemas Multi-agentes – SMA (*Multi-agent Systems* – MAS) são sistemas que tem a capacidade de estudar o comportamento de um conjunto de agentes autônomos com o objetivo de solucionar um problema que está além de suas capacidades individuais (JENNINGS, 1996).

De acordo com Vlassis (2003), um sistema multi-agente é definido como um grupo de agentes que podem potencialmente interagir e compartilhar um mesmo

ambiente. Do ponto de vista de cada agente a mera presença de múltiplos agentes faz com que o ambiente pareça ser dinâmico. Assim, seus principais elementos são: o agente e o ambiente.

Os SMA têm como principal desafio controlar a comunicação entre os agentes que formam uma comunidade de agentes que interagem e convivem mutuamente para alcançar seus objetivos comuns e individuais (SILVEIRA; GOMES, 2003).

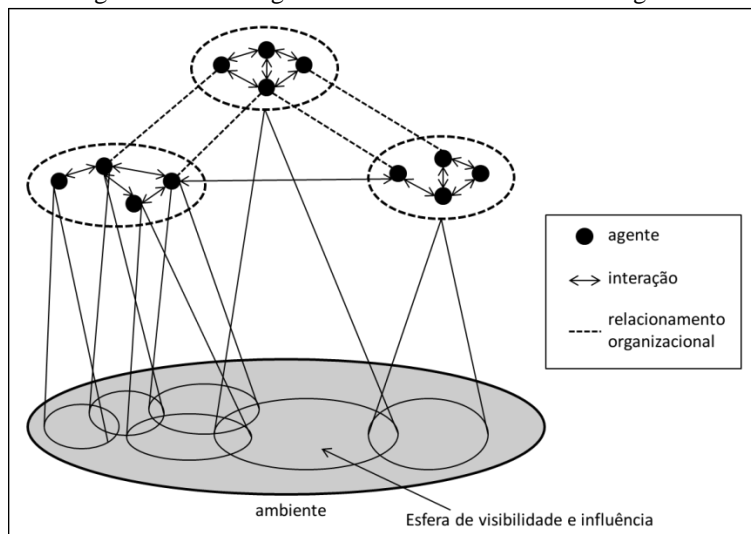
De acordo com Weiss (1999, p.3) um sistema multi-agente tem as seguintes propriedades:

- Cada agente tem apenas uma visão limitada de todo o problema a ser resolvido, ou seja, cada agente tem de resolver um subproblema;
- As capacidades dos agentes de resolver problemas são limitadas;
- O sistema é distribuído;
- A solução dos subproblemas ocorre de forma assíncrona.

Para solucionar problemas complexos de forma distribuída, os agentes têm que se comprometer com mecanismos de coordenação e interação (FISCHER; RUSS; VIERKE, 1998). Estes mecanismos podem variar de simples intercâmbios de informação a negociações de atividades interdependentes quando for necessário. Na maioria dos casos, os agentes agem para alcançar objetivos, seja em nome de indivíduos ou como parte de alguma iniciativa que visa resolver um problema mais complexo.

A Figura 19 descreve um sistema multi-agentes, no qual vários agentes interagem para resolver um problema distribuído. Percebe-se que as ações de vários agentes podem interferir no ambiente, fato este que conseqüentemente aumenta a demanda por coordenação em um sistema multi-agente.

Figura 19 – Visão geral de um sistema baseado em agentes



Fonte: adaptado de Jennings (2000).

Portanto, infere-se que um sistema multi-agente pode ser visualizado como um conjunto de agentes (agrupados, ou não) que se comunicam por meio de interações de alto nível e com um relacionamento organizacional para garantir a atuação, em conjunto, dos agentes de forma coordenada na resolução de um problema.

Um sistema multi-agente sempre deve considerar dois pontos: as características da infraestrutura de comunicação (abstrações de baixo nível) a serem usadas e as estratégias de coordenação (abstrações de alto nível) das atividades dos agentes. Constata-se que comunicação, coordenação e colaboração são propriedades desejáveis para sistemas multi-agentes e estes são indicados para ambientes distribuídos e heterogêneos como é o caso da *Web*.

4.3 Estratégias de Comunicação entre Agentes

As estratégias de comunicação envolvem negociações, autenticações e acordos sobre as ações a serem executadas por um grupo de agentes ou por um agente em específico. A interação entre os agentes é uma característica fundamental para que eles atinjam seus objetivos e está vinculada a algumas formas de comunicação entre eles (VLASSIS, 2003), bem como é necessário que exista um sincronismo entre os agentes (WOOLDRIDGE, 2002).

A comunicação entre agentes estimula a autonomia e incentiva a existência de sociedades de agentes capazes de solucionar problemas mais complexos. Em particular,

a comunicação entre agentes torna possível que programas escritos por diferentes linguagens possam se comunicar por meio de interfaces comuns. Para que isso seja possível, os agentes precisam de uma linguagem de comunicação comum que esteja mais preocupada com a troca de informações do que com seu conteúdo. A comunicação é geralmente dividida em três subseções, que são:

- Protocolos de interação: refere-se à estratégia de alto nível seguida pelos agentes de software que governam suas interações com outros agentes;
- Linguagens de comunicação: é o meio no qual as atitudes sobre o conteúdo da mensagem de intercâmbio são comunicados;
- Protocolos de transporte: são os mecanismos de transporte que usam a linguagem de comunicação. Esses protocolos incluem TCP/IP, SMTP, HTTP, IIOP etc.

Uma das estratégias de comunicação entre os agentes ocorre por meio de uma linguagem de comunicação foi que denominada de ACL (*Agent Communication Language*) e foi dividida nas seguintes partes (NECHES et al., 1991):

- Vocabulário: consiste em um dicionário que engloba todos os conceitos utilizados no domínio dos agentes. Também é conhecido como ontologia;
- Linguagem interna – KIF (*Knowledge Interchange Format*): trata-se de linguagem de programação baseada em lógica de primeira ordem, para codificação de dados simples, regras, restrições e expressões. Também conhecida como formatos para troca de conhecimento;
- Linguagem externa – KQML (*Knowledge Query and Manipulation Language*): é uma camada linguística que encapsula estruturas KIF, para comunicação mais eficiente. Também denominada de linguagem para manipulação e busca de conhecimento.

A linguagem de comunicação ACL permite que os agentes possam se comunicar entre si para a realização de suas tarefas. Dessa forma, uma mensagem ACL é uma expressão KQML na qual os argumentos são sentenças no formato KIF compostas a partir de termos pertencentes ao vocabulário ACL. Uma ACL é estruturada em três camadas: um agente comum de comunicação; um formato comum para a mensagem de comunicação e uma ontologia compartilhada.

Outra estratégia de comunicação é proposta pela FIPA (*Foundation for Intelligent Physical Agents*) que é uma organização de padrões que promove o uso de agentes baseado em tecnologia e interoperabilidade das suas normas com outras tecnologias. Especificações FIPA representam um conjunto de normas que se destinam a promover a interação de agentes heterogêneos e os serviços que eles podem representar (FIPA, 2015).

A FIPA-ACL (*Foundation for Intelligent Physical Agents – Agent Communication Language*) é uma linguagem de comunicação de agente associada a uma arquitetura aberta do padrão FIPA. É muito semelhante ao KQML, pois suporta qualquer especificação de linguagem de conteúdo e ontologia. Entretanto, existem algumas diferenças fundamentais entre FIPA-ACL e KQML, como por exemplo:

- Os modelos semânticos de FIPA-ACL não permitem que um agente manipule diretamente a base de conhecimento virtual de outro agente, função esta suportada pela KQML;
- A arquitetura FIPA possui uma especificação denominada AMS (*Agent Management System*) que descreve os serviços que gerenciam as comunidades de agentes, ou seja, o FIPA-ACL não gerencia explicitamente as comunidades de agentes como é o caso da KQML.

A comunicação é fundamental para que os agentes possam interagir com o ambiente controlado por um sistema multi-agentes uma vez que o sucesso na resolução do problema está diretamente relacionado com o papel de cada agente.

Neste capítulo foram apresentados os princípios dos agentes inteligentes que foram utilizados nesta investigação para auxiliar no processo de busca dos metadados de objetos de aprendizagem nos repositórios. No capítulo a seguir, são demonstrados os modelos de recuperação de informação, assim como os fundamentos da recuperação multimídia e na *Web*. São detalhadas as principais diferenças entre os modelos de recuperação de informação, com a intenção de identificar o modelo ideal para a arquitetura de recuperação de objetos de aprendizagem proposta nesta tese.

5. Recuperação de Informação

A área de Recuperação de Informação é considerada como sendo uma disciplina interdisciplinar, compreendida tanto pela Ciência da Informação quanto pela Ciência da Computação e tem como premissa básica disponibilizar o acesso facilitado às informações de acordo com o interesse do usuário.

Sob o olhar da Ciência da Informação, a recuperação de informação é investigada por meio de uma interpretação mais centrada no usuário, na maneira como as pessoas compreendem, buscam e utilizam a informação. Na perspectiva da Ciência da Computação, a recuperação de informação é pesquisada através do desenvolvimento de algoritmos e técnicas de recuperação de informação que visam aumentar o desempenho dos sistemas de recuperação de informação.

Nesta investigação, parte-se do enfoque da recuperação de informação sob o olhar da Ciência da Informação, que se preocupa com o grau de satisfação do usuário na busca e recuperação de informação que lhe interessa. Entretanto, a visão computacional não pode ser ignorada uma vez que fornece as estratégias para tratar, representar e recuperar grandes volumes informacionais. Portanto, as duas perspectivas serão considerados no desenvolvimento desta tese.

Ainda sob o ponto de vista interdisciplinar, Khapre e Basha (2012) propuseram um *framework* conceitual baseado em três perspectivas que são o usuário, a informação e a tecnologia para demonstrar a convergência da recuperação de informação no âmbito

da Ciência da Informação, da Ciência da Computação e da transversalidade entre essas duas grandes áreas de conhecimento, conforme pode ser observado na Tabela 6.

Tabela 6 – Framework teórico para recuperação de informação – nas áreas de Ciência da Computação e Ciência da Informação.

PERSPECTIVA DE PESQUISA	CIÊNCIA DA INFORMAÇÃO	CIÊNCIA DA COMPUTAÇÃO	ÁREAS TRANSVERSAIS
Informação	<ul style="list-style-type: none"> ▪ Informações multicamada ▪ Relação hierárquica de informação ▪ Informações úteis 	<ul style="list-style-type: none"> ▪ Representação da informação ▪ Ranqueamento da informação ▪ Similaridade de texto 	<ul style="list-style-type: none"> ▪ Relevância
Usuário	<ul style="list-style-type: none"> ▪ Princípio do menor esforço ▪ Processo iterativo de recuperação 	<ul style="list-style-type: none"> ▪ Fornecimento de informação 	<ul style="list-style-type: none"> ▪ Princípio da Incerteza ▪ Interações
Tecnologia	<ul style="list-style-type: none"> ▪ Preferências de canais 	<ul style="list-style-type: none"> ▪ Consulta ▪ Viés técnico ▪ Visão do MEMEX²³ 	<ul style="list-style-type: none"> ▪ Acesso à informação

Fonte: Khapre e Basha (2012, tradução nossa).

Fundamentada na análise das três perspectivas do *framework* conceitual (Tabela 6) percebe-se que a preocupação da Ciência da Informação está mais relacionada com o comportamento das buscas do usuário para localizar a informação de seu interesse. Enquanto que a área de Ciência da Computação se empenha em extrair informação relevante a partir de múltiplas fontes de informação para atender o usuário. Cabe destacar que as ambas as Ciências, preocupam-se com o acesso rápido e preciso da informação relevante.

Um breve resumo da evolução dos estudos relacionados à área de recuperação de informação, com seus principais percussores e suas respectivas contribuições para o avanço da área de recuperação de informação são compilados no Quadro 15.

²³ MEMEX – *Memory Extension* – foi idealizada por Bush (1945) como um dispositivo com capacidade de armazenar grande quantidade de informações, como livros, registros e artigos, com o objetivo de facilitar o processo de organização, indexação e recuperação do conhecimento humano.

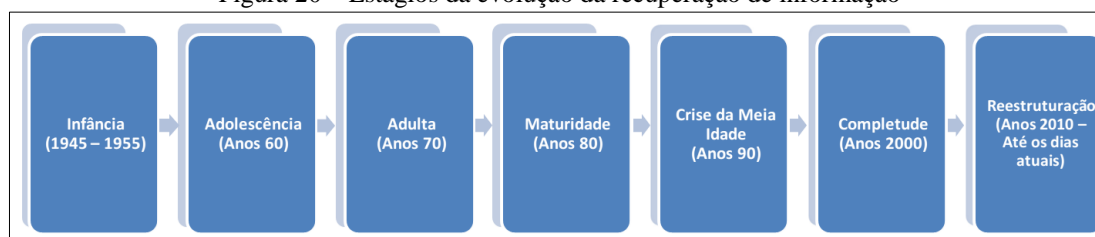
Quadro 15 – Quadro de teóricos que contribuíram com o desenvolvimento da área de RI

AUTOR(ES)	ESTUDO(S)/EVENTO(S)/CONTRIBUIÇÕES
Paul Otlet	É considerado o pai da documentação. Em seus estudos propôs mecanismos para organizar informações para a posterior recuperação do conteúdo, como por exemplo: Em 1895 – Repertório Bibliográfico Universal (RBU). Em 1904, em parceria com Henri La Fontaine criam o sistema de Classificação Decimal Universal (CDU).
Vannevar Bush	Preocupou-se com o problema da explosão informacional, em 1945. Idealizou o MEMEX como uma máquina capaz de funcionar como uma extensão da memória humana e armazenar conhecimento.
Hans Peter Luhn	Em 1947, engenheiro e pesquisador da IBM projetou um sistema mecanizado, com base em cartões perfurados, para procurar compostos químicos. Seus estudos possibilitaram realizar experimentos na área de indexação automática
Calvin Mooers	Em 1951, cunhou o termo “ <i>information retrieval</i> ” e elencou as situações problemas abordadas por esta nova disciplina.
Gerard Salton	Nas décadas de 60 e 70 dedicou seus estudos ao desenvolvimento modelo espaço-vetorial (ou simplesmente vetorial) para ser utilizado num SRI denominado SMART (<i>System for the Manipulation and Retrieval of Text</i>).
Melvin Earl Maron e John Lary Kuhns	Em 1960, apresentam os princípios do modelo probabilístico em recuperação de informação, por meio da discussão de temas sobre relevância e indexação probabilística. Esse modelo foi formalmente definido, anos mais tarde por Robertson e Jones (1976).
Michael Lesk	Em 1996, apresenta os estágios da evolução da área de recuperação de informação (Figura 20) traçando uma analogia com os estágios de evolução do homem, conforme proposto por Shakespeare, em 1599.

Fonte: elaborado pela autora.

Constata-se neste quadro que os estudos que sustentam o desenvolvimento da área de recuperação de informação são a base para as pesquisas desenvolvidas atualmente e continuam sendo fundamentais para o avanço das pesquisas nesta área.

Figura 20 – Estágios da evolução da recuperação de informação



Fonte: elaborado pela autora.

Os estágios da evolução da RI (LESK, 1996) são representados por meio de uma analogia aos sete estágios da evolução do homem propostos por Shakespeare (1599), conforme pode ser visualizado na Figura 20. A ideia do autor foi demonstrar que a área de RI perpassa por vários momentos críticos desde o momento de sua criação quando Vannevar Bush se depara com o desafio da explosão informacional ao final da II Guerra Mundial até o incremento informacional vivenciado nos dias de hoje com o advento da *Web*.

De acordo com Salton (1968) a recuperação de informação é uma área de pesquisa que tem por premissa básica a estrutura, análise, organização, armazenamento, recuperação e busca de informação relevante. Essa proposição do autor se torna ainda mais pertinente com a explosão informacional da *Web* iniciada na década de 1990.

Conforme Ferneda (2003, p. 14) no âmbito da Ciência da Informação o termo:

“Recuperação de Informação” significa, para uns, a operação pela qual se seleciona documentos, a partir do acervo, em função da demanda do usuário. Para outros, “Recuperação de Informação” consiste no fornecimento, a partir de uma demanda definida pelo usuário, dos elementos de informação documentária correspondentes. O termo pode ainda ser empregado para designar a operação que fornece uma resposta mais ou menos elaborada a uma demanda, e esta resposta é convertida num produto cujo formato é acordado com o usuário (bibliografia, nota de síntese, etc.). Há ainda autores que conceituam a recuperação de informação de forma muito mais ampla, ao subordinar à mesma o tratamento da informação (catalogação, indexação, classificação).

A literatura que aborda os conceitos da RI permite apresentar uma compilação das principais definições pertinentes à área que são apresentados no Quadro 16.

Quadro 16 – Principais conceitos sobre recuperação de informação compilados da literatura

AUTOR(ES)	DEFINIÇÃO
Mooers (1951)	A recuperação de informação é um processo de comunicação. É um meio que os usuários de um sistema ou serviço de informação podem encontrar documentos, registros, imagens gráficas ou registros de sons que satisfaçam sua necessidade ou interesse.
Rijsbergen (1979)	A recuperação de informação está relacionada com a recuperação de documentos sejam provavelmente relevantes para a necessidade de informação do usuário, expressa por meio de uma busca.
Salton e McGill (1983)	A recuperação de informação está relacionada com a representação, armazenamento, organização e acesso aos itens de informação.

Blair (1990)	Recuperação de informação é um processo fundamental de comunicação, na qual usuários com necessidade de informação descrevem seu critério de busca sobre uma coleção na qual será efetuada a busca que atenda a solicitação do usuário.
Ingwersen (1992)	A recuperação de informação está relacionada com os processos vinculados com a representação, armazenamento, busca e identificação de informação relevante para a necessidade de informação de um usuário humano. Tem por objetivo investigar e compreender os processos de recuperação de informação para desenhar, construir e avaliar SRI para que o processo de comunicação entre homens e máquinas seja eficaz e facilitado.
Spink e Saracevic (1993)	“Um processo de alta complexidade envolvendo numerosos fatores e variáveis além de decisões e o entrelaçamento dos subprocessos inter-relacionados com a busca. Uma das chaves para um dos subprocessos é a da seleção de termos para a estratégia de busca que por sua vez é influenciada por outros fatores, particularmente os relacionados com os resultados. Sintetizam essa questão com uma pergunta para as investigações: que termos de busca devem ser selecionados para um determinado tema que represente efetivamente o problema de informação do usuário?”
Meadow et al. (2007)	A recuperação de informação é um processo de comunicação. É o meio pelo qual os usuários de um sistema ou serviço de informação solicitam informação ao sistema de recuperação de informação.
Manning, Raghavan e Schütze (2008, p.1)	Recuperação de Informação consiste em encontrar material (geralmente documentos) de natureza não estruturada (usualmente texto) que satisfaça a necessidade de informação a partir de uma vasta coleção (normalmente armazenada em computadores).
Baeza-Yates e Ribeiro-Neto (2013)	A recuperação de informação é responsável pela representação, armazenamento, organização e acesso a itens de informação, como documentos textuais e multimídia, páginas <i>web</i> , registros estruturados e semiestruturados, com o objetivo de fornecer aos usuários facilidade de acesso as informações de seu interesse.

Fonte: elaborado pela autora.

A partir da análise dos conceitos apresentados pelos precursores da RI, torna-se explícito que não existe uma unanimidade sobre essa conceituação, mas em todos os conceitos é possível observar a preocupação que vai desde a representação da informação até a precisão do processo de recuperação de informação.

Meadow et al. (2007, p. 3) seguem a mesma linha de que a recuperação de informação é um processo de comunicação. Esse processo de comunicação inclui o usuário, que formula sua expressão de busca por meio de uma interface e perpassa por todo processamento computacional pertinente ao processo de recuperação de informação.

De acordo com Oliván (2008, p. 23) esse processo de comunicação pode envolver diferentes elementos, tais como:

- Comunicação usuário-intermediário: inicia quando um usuário necessita uma informação e solicita a um intermediário que traduzirá a solicitação (linguagem natural) em uma estratégia de busca;
- Comunicação com o buscador (usuário ou intermediário) – SRI: acontece quando a pessoa que vai realizar a busca se conecta, local ou remotamente, ao sistema que contém a informação, estabelecendo um diálogo interativo por meio de uma interface desenvolvida para este fim.

Neste contexto, Rijsbergen (1979) afirma que existem algumas diferenças fundamentais entre recuperação de informação (por meio de um SRI, em linguagem natural) e recuperação de dados (por meio de um SGBD, em linguagem artificial) que interferem diretamente no resultado apresentado ao usuário. A Tabela 7 apresenta uma compilação das principais distinções entre esses dois tipos de recuperação.

Tabela 7 – Propriedades de recuperação de dados versus recuperação de informação.

PROPRIEDADE	RECUPERAÇÃO DE DADOS	RECUPERAÇÃO DE INFORMAÇÃO
Correspondência	Exata	Parcial, melhor
Inferência	Dedução	Indução
Modelo	Determinista	Probabilística
Classificação	Monotética ²⁴	Politética ²⁵
Linguagem de consulta	Artificial	Natural
Especificação de consulta	Completa	Incompleta
Itens recuperados	Coincidência exata	Relevante
Resposta ao erro	Sensível	Insensível

Fonte: adaptada de Rijsbergen (1979, tradução nossa).

Após a análise da Tabela 7, percebe-se que a recuperação de informação é mais complexa do que a recuperação de dados e essa complexidade pode ser observada em diferentes níveis, de acordo com: a necessidade de informação do usuário, a precisão da

²⁴ Monotética – As classes são definidas por objetos que possuem os atributos suficientes para pertencer a uma mesma classe.

²⁵ Politética – Cada indivíduo de uma classe possui apenas uma parte dos atributos característicos de todos os membros de determinada classe.

expressão de busca utilizada pelo usuário e volume informacional a ser recuperado. Dessa forma, na recuperação de dados os resultados devem ser precisos e retornar resultados exatos para as consultas, enquanto que na recuperação de informação os resultados podem ser tão imprecisos e longos quanto insignificante para a necessidade de informação do usuário.

Portanto, a recuperação de informação pode ser encarada como um sistema comunicacional que busca identificar em um *corpus* quais itens (a princípio) que atendem à necessidade de informação do usuário.

Na visão de Salton e McGill (1983, p. 7), os sistemas de recuperação de informação podem ser conceituados como um sistema que trata da representação, do armazenamento, da organização e do acesso aos itens de informação. Harter (1986, p. 2) complementa essa definição quando afirma que um SRI é um dispositivo que se interpõe entre os usuários e o *corpus* a ser recuperado.

Chowdhury (2010, p. 4) apresenta uma visão conceitual de um SRI que pode suportar um ou mais tipos diferentes de documentos (textuais ou multimídia). Nessa perspectiva todos os documentos devem ser processados para criar um índice que deve ser buscado no processo de recuperação de informação, conforme pode ser visualizado na Figura 21 que ilustra o escopo que um SRI deve suportar.

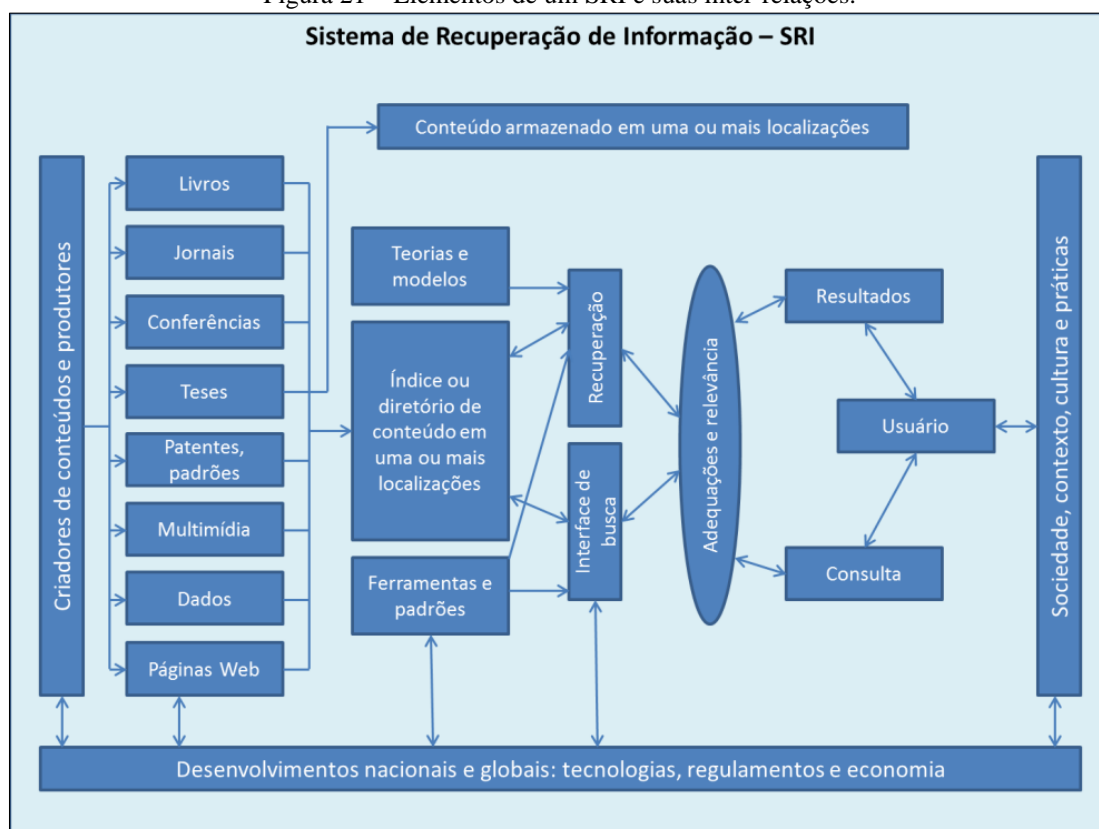
A fim de satisfazer a necessidade de informação do usuário, o SRI deve encontrar a melhor maneira de interpretar a necessidade de informação e recuperar todos os documentos que são relevantes para a consulta ao mesmo tempo em que procura obter o menor número de documentos não relevantes quanto possível.

Conforme Chowdhury (2010, p. 6) um SRI tem por finalidade permitir que usuários encontrem informação relevante de uma coleção organizada de documentos, pois um SRI trabalha com várias fontes de informação, por um lado, e requerimentos do usuário, por outro. Assim, ele deve ter as seguintes funções:

- Analisar os conteúdos das fontes de informação, bem como consultas do usuário;
- Combinar estes para recuperar esses itens que são relevantes;
- Identificar a informação relevante para as áreas de interesse de uma comunidade de usuários;

- Analisar o conteúdo de documentos;
- Representar o conteúdo de fontes analisadas para que as futuras buscas dos usuários sejam satisfeitas;
- Analisar a consulta do usuário e representá-la de tal forma que seja encontrado na base de dados;
- Combinar a sentença de busca com a base de dados armazenada, visando determinar a similaridade da consulta com os dados armazenados;
- Recuperar informação relevante, por meio de uma variedade de ferramentas e técnicas que são empregadas para determinar a relevância dos itens recuperados e seu ranqueamento;
- Fazer mudanças constantes em todos os aspectos do sistema, levando em consideração a rápida evolução das TIC relacionados com a mudança dos padrões da sociedade, usuários e suas expectativas e necessidades de informação.

Figura 21 – Elementos de um SRI e suas inter-relações.



Fonte: adaptada de Chowdhury (2010, p.4), tradução nossa.

A partir de então, a área de recuperação de informação cresceu, em importância, particularmente devido ao aumento da disponibilidade de informação em formato digital

(RIJSBERGEN, 1979; WITTEN; MOFFAT; BELL, 1999; GARFIELD, 2001). Assim, o processo de recuperação de informação relevante na *Web*, tornou-se complexo devido ao crescimento exponencial de seu volume de dados (tanto textuais quanto multimídia) e a falta de estruturação dos documentos, bem como pelo seu comportamento dinâmico e volátil.

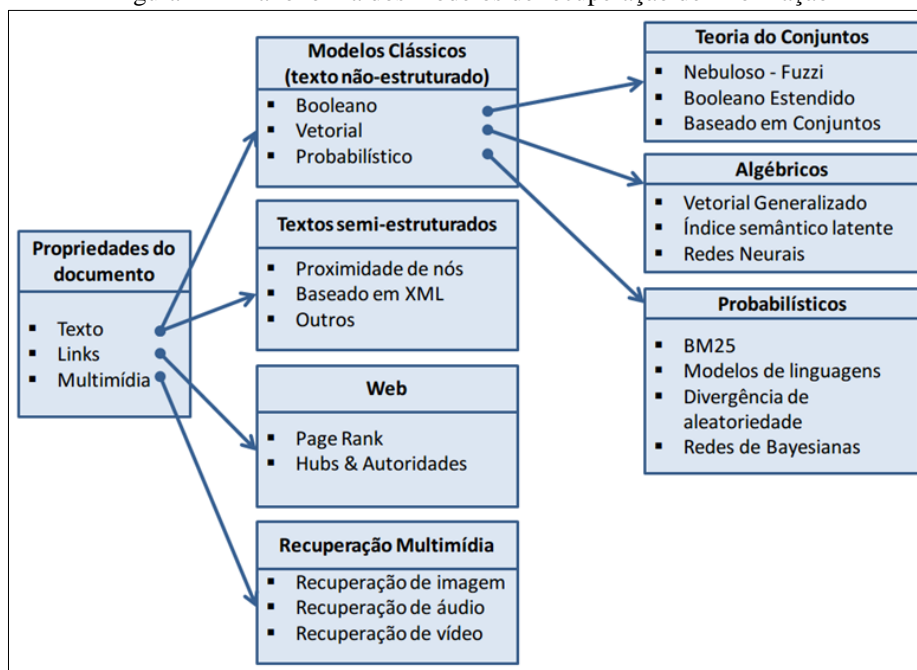
5.1 Recuperação de Informação: Modelos Clássicos

Os primeiros modelos de RI datam das décadas de 60 e 70 quando surgem os estudos dos modelos clássicos (booleano, espaço-vetorial e probabilístico) que foram aprimorados na década de 80. Os princípios desses modelos continuam sendo utilizados nos SRI atuais e motores de busca da *Web* e todos tem um ponto em comum que é a dependência da expressão de busca que é fornecida pelo usuário.

Normalmente, os SRI utilizam termos indexados para recuperar o conteúdo de um *corpus* documental a partir de uma expressão de busca inicial. Todo sistema de recuperação de informação precisa de um modelo computacional que é responsável por determinar como será o funcionamento do sistema de recuperação de informação.

Nesse contexto, os autores Baeza-Yates e Ribeiro-Neto (2013, p. 24) apresentaram esses modelos por meio de uma taxonomia (Figura 22) que representam os modelos de recuperação de informação textual, *links* e os métodos de recuperação de objetos multimídia. Nesta taxonomia os autores apresentam as diferentes formas de resolver a recuperação de informação de acordo com as características informacionais a serem recuperadas.

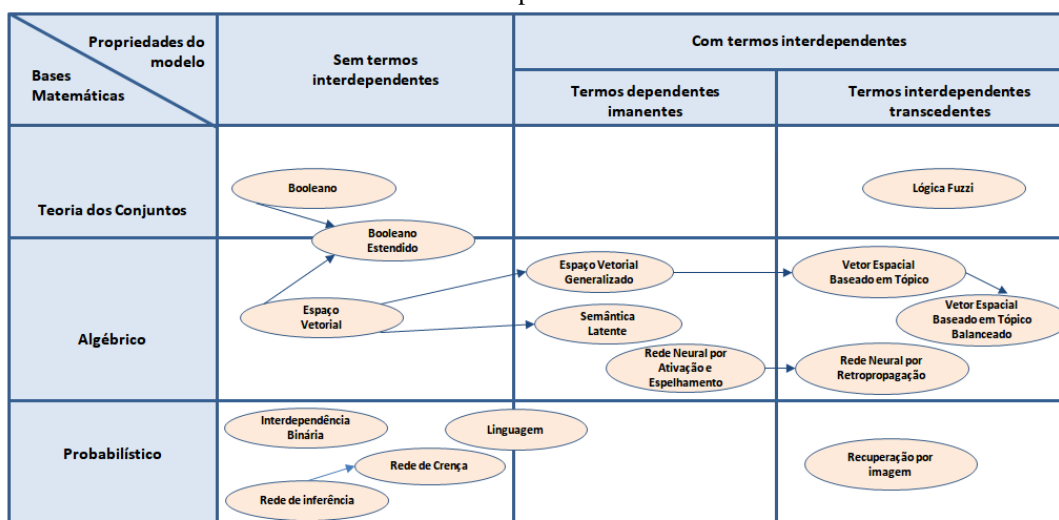
Figura 22 – Taxonomia dos modelos de recuperação de informação



Fonte: Baeza-Yates e Ribeiro-Neto (2013, p. 24).

A taxonomia apresentada permitiu observar que os modelos clássicos de RI têm seus fundamentos fortemente amparados em bases matemáticas sólidas. Neste sentido, Kuropka (2004) propõe uma categorização das propriedades dos modelos de recuperação de informação e os fundamentos matemáticos que as sustentam, conforme pode ser observado na Figura 23.

Figura 23 – Categorização dos modelos de recuperação de informação com suas respectivas interdependências



Fonte: adaptada de KUROPKA (2004, tradução nossa).

A partir da análise da categorização apresentada na Figura 23, observa-se que as bases matemáticas são originárias: dos princípios da Teoria dos Conjuntos (booleano);

dos modelos algébricos (vetores, matrizes e tuplas) e dos modelos probabilísticos (inferências). Com relação às propriedades do modelo são consideradas as seguintes associações:

- Sem termos interdependentes – não consideram o relacionamento entre os termos;
- Termos dependentes imanentes – permite uma representação dos relacionamentos entre os termos que pode ser direta ou indireta, oriunda da co-ocorrência de termos;
- Termos interdependentes transcendententes – permite a relação entre os termos, sem determinar como o relacionamento é estabelecido. São dependentes de fontes externas (humanas, heurísticas etc.) para estabelecer uma associação entre os termos.

Frakes e Baeza-Yates (1992, p. 10) apresentam um modelo de classificação dos modelos de recuperação de informação, conforme Quadro 17. Esse modelo apresenta, de forma breve, as principais características de cada modelo. Neste contexto, percebe-se que a eficiência de execução de um SRI está diretamente relacionada tanto pelas características lógicas quanto físicas suportadas por cada modelo conceitual.

Quadro 17 – Classificação dos modelos de recuperação de informação

MODELO CONCEITUAL	ESTRUTURA DE ARQUIVO	OPERAÇÕES DE CONSULTA	OPERAÇÕES COM TERMOS	OPERAÇÕES COM DOCUMENTOS	ARQUITETURA DE HARDWARE
Booleano	Arquivos lineares	<i>Feedback</i>	Stem	Parse	vonNeumann
Booleano Estendido	Arquivos invertidos	Parse	Peso	Exibição	Paralela
Probabilístico	Assinaturas	Booleana	Tesouro	Cluster	Específica para RI
String de busca	Árvores Pat	Cluster	Stoplist	Ranqueamento	Discos Óticos
Espaço-vetorial	Grafos e Hashing		Truncagem	Ordenar Máscara de campo Atribuir IDs	Discos Magnéticos

Fonte: adaptado de Frakes e Baeza-Yates (1992, p. 10, tradução nossa).

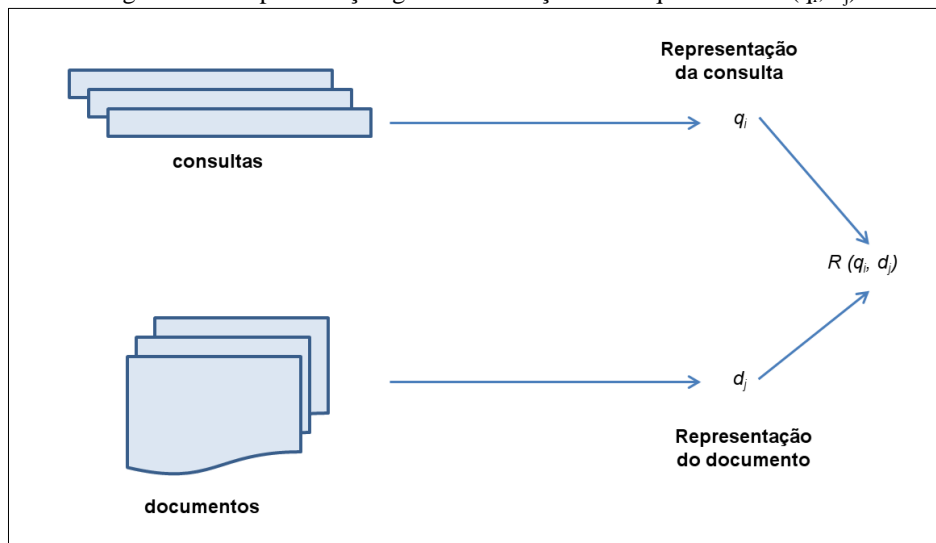
Segundo FERNEDA (2012, p.20), um modelo de recuperação de informação é a especificação formal de três elementos: a representação dos documentos, uma expressão de busca e a função de busca. Assim, o autor enfatiza que a eficiência de um SRI está diretamente relacionada ao modelo de recuperação de informação adotado que influencia diretamente em seu modo de operação.

Baeza-Yates e Ribeiro-Neto (2013, p. 22) defendem que as premissas fundamentais de um modelo de RI podem ser caracterizadas por meio de uma quádrupla, conforme representada a seguir:

$$[\mathbf{D}, \mathbf{Q}, F, R(q_i, d_j)]$$

onde:

1. \mathbf{D} é um conjunto composto por visões lógicas (ou representações) dos documentos da coleção.
2. \mathbf{Q} é um conjunto composto de visões lógicas (ou representações) das necessidades de informação dos usuários. Essas representações são denominadas de consultas.
3. F é um *framework* para modelar as representações dos documentos, das consultas e de seus relacionamentos, por meio conjuntos e relações Booleanas, vetores e operações de álgebra linear, espaços amostrais e distribuições de probabilidade.
4. $R(q_i, d_j)$ é uma função de ranqueamento que atribui um número real à representação da consulta q_i que pertence a \mathbf{Q} e a representação de um documento d_j que pertence a \mathbf{D} . Esse ranqueamento permite definir um ordenamento entre os documentos em relação à consulta q_i . A função de ranqueamento é representada graficamente, conforme Figura 24.

Figura 24 – Representação gráfica da função de ranqueamento $R(q_i, d_j)$ 

Fonte: Baeza-Yates e Ribeiro-Neto (2013, p. 23).

Cabe salientar que a recuperação de informação precisa lidar com um problema essencial que é presumir quais documentos recuperados serão considerados como relevantes ou não pelos critérios de busca adotados pelo usuário. Esse é um problema complexo, pois o grau de incerteza e imprecisão são fatores a serem considerados nesse processo e dependem do modelo de recuperação de informação adotado. Uma forma de apresentar os itens relevantes para o usuário é por meio de um sistema de ranqueamento que será explicado ao longo da seção 5.1.1 de acordo com cada modelo clássico de recuperação de informação, pois cada modelo trata esta questão de uma forma distinta.

Portanto, a complexidade da modelagem em recuperação de informação é percebida pela necessidade de ranqueamento do resultado de uma consulta que pode ser dividido em duas etapas, conforme ressaltam Baeza-Yates e Ribeiro-Neto (2013, p. 21-22):

- Elaboração de um *framework* lógico para representação de documentos e consultas – normalmente baseado em conjuntos, vetores ou distribuições de probabilidades;
- Definição de uma função de ranqueamento que tem como premissa computar o grau de similaridade (*sim*) de cada documento com a consulta efetuada – diretamente relacionada à ordenação dos documentos recuperados.

Os modelos clássicos de recuperação de informação são: booleano, espaço vetorial e probabilístico. A seguir, cada um dos modelos será descrito com suas respectivas características de funcionamento e desempenho.

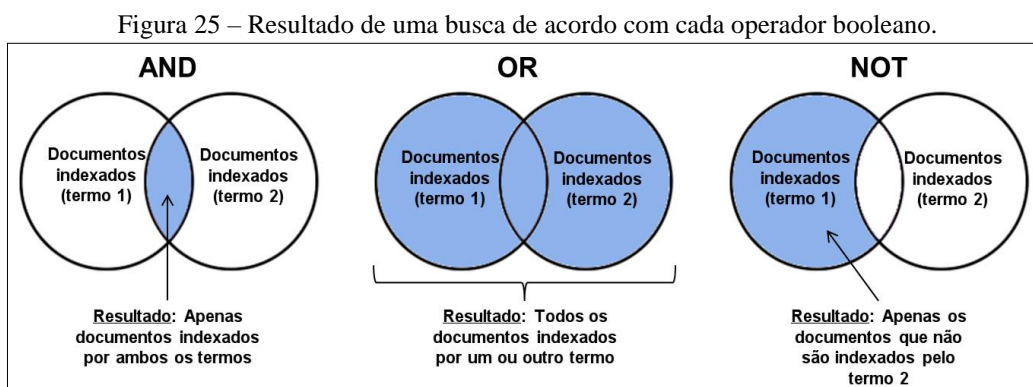
5.1.1 Modelo Booleano

O modelo de recuperação de informação booleano tem seus fundamentos baseado na Teoria dos Conjuntos e na Álgebra Booleana. Parte do princípio de que os termos de indexação estão presentes ou ausentes no documento a ser recuperado, ou seja, as frequências são binárias (1 indica a presença e 0 indica a ausência do termo) na matriz de termos do documento.

É um modelo para a recuperação de informação em que se pode representar qualquer consulta na forma de uma expressão booleana, ou seja, por meio de termos que são combinados com os operadores:

- Conjunção – AND (E) – utilizado para restringir uma busca;
- Disjunção – OR (OU) – utilizado para ampliar uma busca;
- Negação – NOT (NÃO) – restringe uma busca, excluindo os termos irrelevantes.

A Figura 25 apresenta o resultado de uma busca que simula a utilização dos três operadores e seus respectivos resultados. A área colorida na Figura 25 representa o retorno da busca realizada no SRI.



Fonte: elaborada pela autora.

Esse modelo considera cada documento como relevante ou não relevante a uma determinada consulta realizada. O formalismo claro e a simplicidade na utilização de

pesos binários para os termos de indexação são as principais vantagens do modelo booleano.

Uma desvantagem observada neste modelo é que não existe um sistema de ranqueamento, o que faz com que muitos ou poucos documentos sejam recuperados. Outro inconveniente é que os usuários podem estranhar a forma de construir expressões de busca baseadas em expressões booleanas. Desta maneira, alguns autores consideram que o modelo booleano é muito mais utilizado para recuperação de dados do que para recuperação de informação (MANNING; RAGHAVAN; SCHÜTZE (2008, p. 5) e BAEZA-YATES; RIBEIRO-NETO (2013, p. 31)).

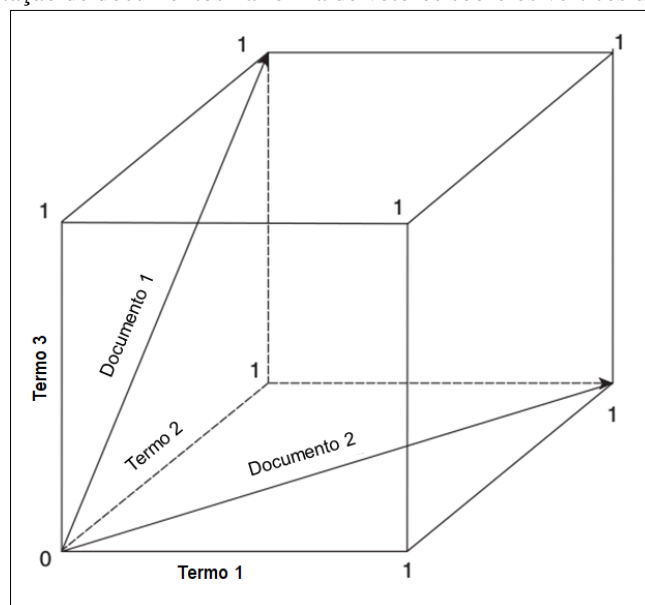
5.1.2 Modelo Espaço Vetorial

O modelo espaço vetorial, ou simplesmente modelo vetorial, foi proposto por Salton (1971) e utiliza o sistema de atribuição de pesos para definir o ranqueamento dos documentos recuperados. Foi proposto para eliminar algumas limitações observadas no modelo Booleano.

De acordo com Meadow et al. (2007, p. 242), o modelo vetorial trabalha com o quadrante positivo de um sistema de coordenadas cartesianas como um espaço, nos quais os eixos representam os termos do vocabulário de indexação.

O modelo vetorial, conforme pode ser observado na Figura 26, assume um espaço n -dimensional, no qual uma dimensão representa uma palavra ou termo possível de ser recuperado. Isto significa que, em geral, n terá um valor na casa dos milhares ou centenas de milhares (número de palavras de uma língua natural). Cada texto está representado como um vetor neste espaço. Cada componente do vetor (distância ao longo de um eixo) ou é 0 ou 1 (depende se o termo correspondente está ausente ou presente no texto, respectivamente). Na sua forma mais simples, cada vetor de texto termina num ponto no espaço vetorial dentro da unidade hipercúbica com n dimensões. Uma vez que cada termo possível ou está presente ou ausente, todos os vetores devem atingir o ponto em um vértice do hipercubo (MEADOW et al., 2007, p. 64).

Figura 26 – Representação de documentos na forma de vetores sobre os vértices da unidade hiper-cúbica.



Fonte: Meadow et al. (2007, p. 242).

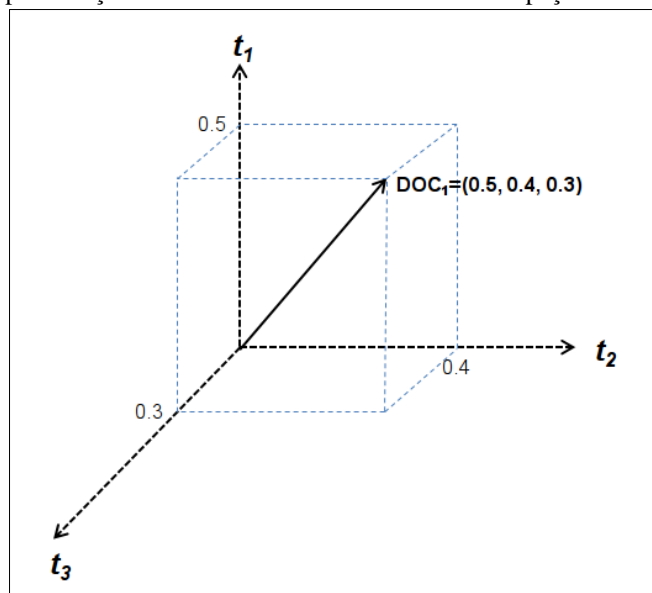
Na concepção de Peña, Baeza-Yates e Rodríguez Muñoz (2002, p. 218), o modelo vetorial é representado no sistema por duas listas ordenadas numericamente que correspondem tanto aos pesos atribuídos ao documento (em função da soma dos valores atribuídos aos termos de indexação que representam o documento), quanto aos valores da pergunta do usuário (em função dos pesos atribuídos aos termos de indexação da pergunta).

Neste modelo, os documentos são representados por vetores nos quais cada elemento representa o peso (ou a relevância) de determinado termo para o documento. São atribuídos pesos positivos e não binários aos termos de indexação e aos documentos para computar o grau de similaridade entre eles. Cada peso do elemento deve ser normalizado para assumir valores entre 0 e 1, sendo que os pesos mais próximo de 1 indicam uma maior importância desse termo para descrever o documento (FERNEDA, 2012, p. 31).

A Figura 27 representa graficamente um documento $DOC_1=(0.5, 0.4, 0.3)$ em um espaço tridimensional, ou seja, que considera o documento com três termos de indexação. A similaridade entre a expressão de busca e os documentos pode ser obtida por meio de uma função de correlação, como por exemplo: cosseno do ângulo entre os dois vetores (que é o mais utilizado), coeficiente de Jacquard, coeficiente de Pearson,

entre outros. A seleção da função de correlação depende de quem constrói o SRI e não está relacionada com o modelo vetorial propriamente dito.

Figura 27 – Representação vetorial de um documento em um espaço vetorial tridimensional



Fonte: adaptada de Ferneda (2012, p.32)

No modelo vetorial, um documento (d_j) e uma consulta do usuário (q) são representados como vetores com t dimensões, que é representado pela expressão (BAEZA-YATES; RIBEIRO-NETO, 2013, p.46):

$$\begin{aligned} \text{sim}(d_j, q) &= \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} \\ &= \frac{\sum_{i=1}^t (w_{i,j} \times w_{i,q})}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}} \end{aligned}$$

onde:

1. $\vec{d}_j \cdot \vec{q}$ é o produto interno dos dois vetores.
2. $|\vec{d}_j|$ e $|\vec{q}|$ são as normas dos vetores dos documentos e da consulta. O fator $|\vec{q}|$ não afeta o ranqueamento (ordenação dos documentos) porque ele é o mesmo para todos os documentos. O fator $|\vec{d}_j|$ faz a normalização pelo tamanho do documento.
3. $w_{i,j} \geq 0$ e $w_{i,q} \geq 0$, a expressão $\text{sim}(d_j, q)$ varia entre 0 e 1.

Dessa forma, o usuário do SRI recebe como resposta um conjunto ordenado (ranqueado de acordo com o nível de similaridade) de documentos relacionados com a expressão de busca original. Entretanto, Witten, Moffat e Bell (1999) esclarecem que não é necessário que todos os termos da expressão de busca estejam presentes no documento, bem como nem todos os termos terão o mesmo peso para determinar a relevância ou não do documento.

5.1.3 Modelo Probabilístico

O modelo probabilístico foi proposto por Maron e Kuhns (1960) com o propósito de recuperar informação por meio dos fundamentos da probabilidade para estimar a relevância e considerar o grau de incerteza no julgamento entre um *corpus* documental e a expressão de busca do usuário.

Este modelo descreve o processo de recuperação de informação sob o ponto de vista probabilístico, isto é, especifica a probabilidade de relevância de um determinado documento para a consulta que foi realizada por um usuário. Este modelo permite que o usuário possa inferir sob os resultados que lhe foram apresentados, para que ele possa julgar a relevância dos itens recuperados pelo SRI. Acredita-se que por meio das interações do usuário seja possível alcançar, gradativamente, resultados mais relevantes para a necessidade de informação do usuário.

Serviu como base para os estudos de outros pesquisadores, como foi o caso de Robertson e Jones (1976) quando propuseram o modelo *Binary Independent Retrieval* (BIR) – Recuperação Independente Binária. Fundamenta-se na representação binária dos documentos, tal como no modelo de recuperação booleano, indicando a presença ou não dos termos (com os valores 1 ou 0, respectivamente). A diferença reside no método estatístico e nas premissas segundo as quais o seu funcionamento é estabelecido, por meio das seguintes declarações:

- A partir da expressão de busca do usuário, os documentos na coleção são classificados em dois grupos: o subconjunto de documentos relevantes e o subconjunto de documentos não relevantes;
- Existe uma resposta ideal do sistema, constituída pelo conjunto de documentos relevantes recuperados, que é denominado de conjunto de relevantes recuperados.

- Existe uma consulta ideal, que é aquela que fornece um conjunto de itens relevantes para o usuário. Cabe destacar que é difícil formular a consulta ideal, pois o usuário não tem como conhecer todo o *corpus* a ser recuperado.

Robertson (1977) apresenta a hipótese que fundamenta o princípio de ranqueamento probabilístico (*Probability Ranking Principle – PRP*) da seguinte forma:

Dada uma consulta de usuário q e um documento d_j da coleção, o modelo probabilístico tenta estimar a probabilidade do usuário achar o documento d_j interessante (isto é, relevante). O modelo supõe que essa probabilidade de relevância depende apenas das representações da consulta e do documento, ou seja, das informações disponíveis ao sistema. Além disso, o sistema supõe que exista um conjunto de todos os documentos que o usuário prefira como conjunto resposta para a consulta q . Tal conjunto de resposta ideal é chamado de R e deve maximizar a propriedade de relevância para o usuário. Os documentos no conjunto R são previstos como *relevantes* à consulta; já os documentos que não estão nesse conjunto são previstos como *não relevantes*. (BAEZA-YATES; RIBEIRO-NETO, 2013, p.48)

Esta hipótese se baseia na ordenação automática dos itens recuperados no *corpus* por ordem decrescente de probabilidade de relevância para o usuário. Baeza-Yates e Ribeiro-Neto (2013, p. 49) destacam que esta hipótese apresenta alguns problemas relacionados a variáveis externas ao sistema de recuperação de informação uma vez que, do ponto de vista do usuário, a resposta obtida pode não ser a ideal. Outra observação é que o princípio não deixa claro como são computadas as probabilidades de relevância.

Baeza-Yates e Ribeiro-Neto (2013, p. 49) apresentam as seguintes considerações para compreender o cômputo das probabilidades neste modelo: dada uma consulta q , o modelo probabilístico atribui a cada documento d_j , como medida de similaridade com a consulta, a razão $P(d_j \text{ relevante a } q) / P(d_j \text{ não relevante a } q)$ que calcula a probabilidade do documento d_j ser relevante à consulta q . Considerando-se a probabilidade de relevância como score para o ranqueamento, isso minimiza a probabilidade de um julgamento errôneo (FUHR (1992) e RIJSBERGEN (1979)).

Baeza-Yates e Ribeiro-Neto (2013, p. 49), explicam o cálculo dessa probabilidade da seguinte forma: um documento d_j é representado por um vetor de

pesos binários que indicam a presença ou a ausência de termos de indexação como segue:

$$\vec{d}_j = (w_{1,j}, w_{2,j}, \dots, w_{t,j})$$

onde $w_{1,j} = 1$ se o termo k_i ocorre no documento d_j e $w_{1,j} = 0$ caso contrário. Seja R um conjunto de documentos estimado (ou inicialmente considerado) como relevantes para o usuário para a consulta q . Seja \bar{R} o complemento de R (ou seja, o conjunto de documentos não relevantes). $P(R/\vec{d}_j, q)$ é a probabilidade de que o documento d_j com a representação \vec{d}_j seja relevante para a consulta q . Além disso, $P(\bar{R}/\vec{d}_j, q)$ é a probabilidade de que o documento d_j não seja relevante para a consulta q . A similaridade $sim(d_{j,q})$ entre o documento d_j e a consulta q é definida pela razão

$$sim(d_{j,q}) = \frac{P(R/\vec{d}_j, q)}{P(\bar{R}/\vec{d}_j, q)}$$

utilizando a regra de Bayes, tem-se:

$$sim(d_{j,q}) = \frac{P(\vec{d}_j/R, q) \times P(R, q)}{P(\vec{d}_j/\bar{R}, q) \times P(\bar{R}, q)} = \frac{P(\vec{d}_j/R, q) \times P(R/q)}{P(\vec{d}_j/\bar{R}, q) \times P(\bar{R}/q)}$$

onde $P(\vec{d}_j/R, q)$ é a probabilidade que um documento aleatoriamente selecionado do conjunto R de documentos relevantes (para a consulta q) terá a representação \vec{d}_j . Além disso, $P(R/q)$ é a probabilidade que um documento que tenha sido selecionado aleatoriamente a partir do *corpus* seja considerado como relevante para a consulta q . Os significados atribuídos a $P(\vec{d}_j/R, q)$ e $P(\bar{R}/q)$ são análogos e complementares.

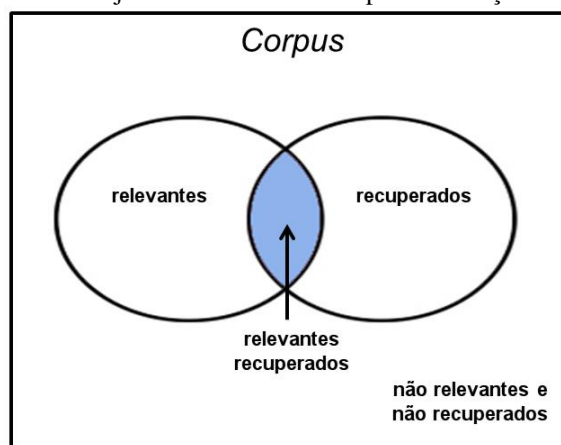
Assim, como $P(R/q)$ e $P(\bar{R}/q)$ são os mesmos para todos os documentos do *corpus*, pode-se inferir que:

$$sim(d_{j,q}) \sim \frac{P(\vec{d}_j/R, q)}{P(\vec{d}_j/\bar{R}, q)}$$

A partir de uma expressão de busca, pode-se dividir o *corpus* (que é composto por N itens) em quatro subconjuntos distintos. A Figura 28 ilustra os subconjuntos (recuperados, relevantes, relevantes e recuperados, não relevantes e não recuperados)

possíveis de serem obtidos, após a execução de uma expressão de busca sob um determinado *corpus*.

Figura 28 – Subconjuntos de documentos após a execução de uma busca



Fonte: adaptada de Ferneda (2012, p.47).

Analisando a Figura 28, pode-se perceber que o subconjunto de documentos relevantes recuperados é o resultado da intersecção entre os conjuntos “relevantes” e “recuperados”. Esta seria a área com a resposta ideal para ser apresentada ao usuário.

O modelo probabilístico, diferentemente do modelo espaço vetorial, não determina um peso (pré-fixado) para os termos indexados dos documentos e das expressões de busca do usuário, ou seja, o ranqueamento é realizado pelo princípio da ordenação probabilística (calcula-se o peso dinamicamente a partir da expressão de busca e o *corpus* documental).

Ferneda (2013, p. 51) explica o funcionamento do modelo probabilístico da seguinte forma:

A partir de uma expressão de busca, composta por um ou mais termos, o usuário expressa sua necessidade de informação e a submete ao sistema. Por meio de cálculos de probabilidade o sistema calcula, para cada documento do corpus, um valor numérico (similaridade), que representa a provável relevância do documento para a consulta. Esse valor é utilizado para ordenar os resultados da busca. Tendo um primeiro conjunto de documentos, o usuário pode marca alguns deles que considera verdadeiramente relevantes para a sua necessidade. O conjunto de documentos marcados pode ser então submetido ao sistema, permitindo fornecer resultados mais precisos. Esse processo, denominado *relevance feedback*, pode ser repetido até que o usuário se sinta satisfeito com os resultados.

Segundo Croft, Callan e Lafferty (2001) e Fuhr (1992), o modelo probabilístico é muito eficaz para sistemas de recuperação de informação, apesar de sua complexidade de implementação, os resultados são muito eficientes, além possibilitar que o usuário realize o *relevance feedback* (ver capítulo 6).

Ferneda (2013, p. 52) conclui que os modelos clássicos de RI:

[...] compartilham um mesmo paradigma de representação no qual os termos extraídos dos documentos e das buscas são suficientes para se efetivar o processo de recuperação de informação, considerando-o um sistema fechado no qual o significado dos elementos lexicais é dado pelas suas inter-relações no interior de um *corpus* documental.

5.2 Recuperação de Informação Multimídia

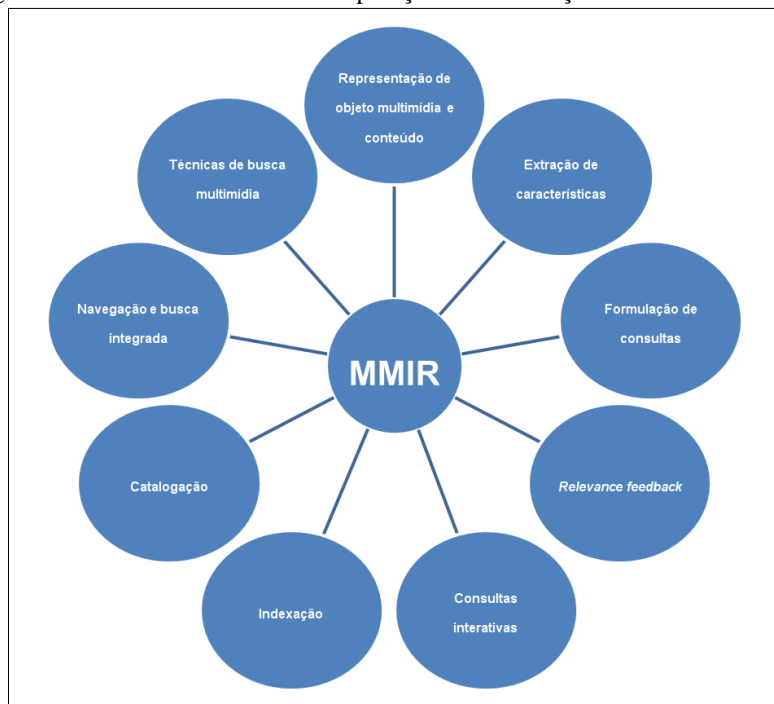
A recuperação da informação multimídia (*Multimedia Information Retrieval* – MMIR) e na *Web* se difere da clássica recuperação de texto, principalmente pela diversidade de formatos encontrados nesse tipo de informação, bem como pela variedade de linguagens, com alfabetos e *scripts* (por exemplo, Indiano), os quais podem ser muito extensos (como por exemplo, o Chinês).

Segundo Rüger (2009, p. 4), um sistema de recuperação de informação multimídia deve realizar o processamento de uma busca para encontrar documentos multimídia, por meio de um motor de busca, que corresponda à expressão de busca inicial do usuário.

Baeza-Yates e Ribeiro-Neto (2011, p. 588) enfatizam que a tarefa de um sistema de recuperação de informação multimídia é recuperar dados de texto, imagem, vídeo e som relacionados ao interesse do usuário e ranqueá-lo de acordo com o grau de interesse do usuário.

Destaca-se que devem ser aplicadas diferentes estratégias de recuperação de informação e ranqueamento, considerando as especificidades de cada formato a ser recuperado. Os autores também enfatizam que a MMIR engloba diferentes áreas (Figura 29) que vai desde a semântica contida de dados multimídia até as técnicas de busca, indexação, extração e navegação.

Figura 29 – Diferentes áreas da recuperação de informação multimídia – MMIR



Fonte: elaborado pela autora.

Dessa forma, para o domínio de recuperação de informação multimídia, a descrição dos recursos por meio de um padrão de metadados é muito importante devido aos diferentes tipos de formatos utilizados para a criação da informação. Cabe ressaltar que a informação multimídia é parte integrante da *Web*.

5.3 Recuperação de Informação na *Web*

A recuperação de informação na *Web* é um processo que requer uma atenção especial, pois seu acervo é distribuído e composto por uma multiplicidade de formatos muito grande. Aliada a essa situação existe a dificuldade do usuário em formular sua expressão de busca de maneira adequada. Assim, torna-se necessário desenvolver soluções que possibilitem uma recuperação de informação mais eficiente e significativa para o usuário do SRI na *Web*.

De acordo com Baeza-Yates e Ribeiro-Neto (2011, p. 449-450) a recuperação de informação na *Web* é encarada como um problema desafiador que pode ser dividido em duas classes: centrado nos dados e centrado no usuário e sua interação com o sistema de recuperação de informação. O Quadro 18 apresenta uma visão geral dessa classificação proposta pelos autores na qual, percebe-se que os dados são problemas intrínsecos da

Web e que cabe ao usuário conceber uma expressão de busca eficaz para que os resultados retornados satisfaçam sua necessidade de informação.

Quadro 18 – Recuperação de informação na *Web* – Um problema desafiador

RECUPERAÇÃO DE INFORMAÇÃO NA WEB	
CENTRADA NOS DADOS	Dados distribuídos – parte da natureza intrínseca da <i>Web</i> , os dados espalham-se por uma grande quantidade de computadores, servidores e plataformas.
	Dados voláteis – a dinamicidade da <i>Web</i> faz com que os dados possam ser adicionados, alterados e removidos a qualquer tempo.
	Volume de dados – dimensionar a quantidade de dados é complicado, pois a escalabilidade é alta.
	Dados não-estruturados e redundantes – a ausência de estruturação dos dados dificulta o processo de recuperação na <i>Web</i> .
	Qualidade dos dados – os dados podem ser imprecisos, fragmentados, incompletos e inconsistentes.
	Dados heterogêneos – os dados possuem diversos tipos de formatos e línguas.
CENTRADA NO USUÁRIO	Expressão de busca – necessidade de elaborar uma expressão de busca (em linguagem natural) que indique sua real necessidade de informação.
	Interpretação dos resultados – os resultados retornados são de fato relevantes para a necessidade de informação do usuário

Fonte: elaborado pela autora.

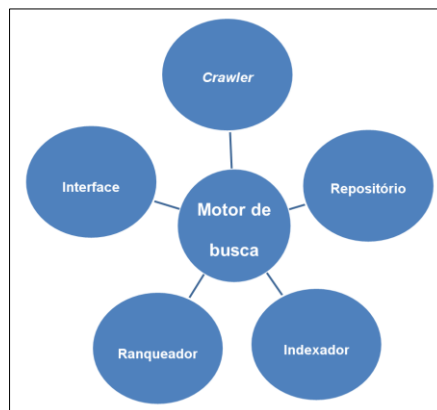
A recuperação de informação relevante na *Web* é uma tarefa complexa que exige abordar diferentes tipos de formatos e estruturas de dados além de ter que manter um sistema de comunicação eficiente com o usuário para que a informação retornada seja relevante. À medida que a quantidade de informação na *Web* é incrementada, os motores de busca precisam se adaptar para gerir toda diversidade desse volume informacional. Assim, destaca-se a importância dos motores de busca na *Web* que, além de outras funções precisam estabelecer esse processo comunicacional com o usuário.

Baeza-Yates e Ribeiro-Neto (2013, p. 100) enfatizam que os primeiros motores de busca da *Web* funcionavam como se fossem simples modelos de RI e precisaram ser repensados a partir do momento que as seguintes características passaram a ser observadas:

- Coleções compostas por páginas *Web* e não documentos;
- Páginas *Web* precisavam ser coletadas;
- Coleções muito maiores e dinâmicas.

Nesse sentido, observam-se esforços constantes para aprimorar o desempenho dos sistemas de recuperação de informação para que os resultados apresentados aos usuários fossem os mais relevantes possíveis. Normalmente, um motor de busca é composto por cinco elementos, conforme pode ser visualizado na Figura 30.

Figura 30 – Elementos constituintes de um motor de busca



Fonte: elaborada pela autora.

Com base nesses elementos, o motor de busca permite rastrear periodicamente a informação que está disponível na *Web* ou em repositórios específicos para que seja indexada e disponibilizada ao usuário quando ele desejar realizar uma busca.

5.3.1 Tipos de Motores de Busca

Os motores de busca são instrumentos de pesquisa projetados para auxiliar o usuário na busca de informação na *Web* por meio da utilização de expressões ou palavras-chave.

O termo "motor de busca" é geralmente usado para descrever todos os tipos de mecanismos de busca, que podem ser categorizados em: gerais (como por exemplo, Google²⁶) que localizam e indexam o máximo de conteúdos da *web*, fazendo o mínimo de restrições possíveis; específicos (como por exemplo, repositório de objetos de aprendizagem BIOE²⁷) que localizam as informações indexadas em um *corpus* limitado ao seu escopo e os meta-buscadores (como por exemplo, Dogpile²⁸) que localizam os

²⁶ Disponível em: <<http://www.google.com>>. Acesso em: 23 ago. 2016.

²⁷ Disponível em: <<http://objetoseducacionais2.mec.gov.br/>>. Acesso em: 23 ago. 2016.

²⁸ Disponível em: <<http://www.dogpile.com>>. Acesso em: 23 ago. 2016.

resultados retornados por outros motores de busca e incorporam-os em um mesmo resultado a ser apresentado ao usuário.

A forma de atuação de cada motor de busca pode ser classificada em um dos seguintes tipos:

- Baseada em *crawler* – também conhecido como *Web Crawler*, criam suas listas automaticamente a partir de um software que rastreia a *Web* na busca das informações desejadas para indexá-las e construir sua base de pesquisa. As alterações de página da *Web* podem ser capturadas dinamicamente por mecanismos de pesquisa com base em rastreadores e afetarão a forma como essas páginas da *Web* serão ranqueadas nos resultados da pesquisa;
- Diretório mantido por humanos – dependem de editores humanos para criar suas listas. Normalmente, os *webmasters* enviam uma breve descrição para o diretório de seus sites ou os editores escrevem para os sites que revisam, e essas descrições manualmente editadas formarão a base de pesquisa. Portanto, as alterações feitas em páginas da *Web* individuais não terão efeito direto sobre como essas páginas serão ranqueadas nos resultados da pesquisa.
- Meta-buscadores – é um sistema que busca informações em outros motores de busca (normalmente os mais conhecidos, como o Google e o Bing) e apresenta uma combinação dos melhores resultados retornados por cada buscador. É considerado como um buscador de buscadores.

A Tabela 8, baseada em Zhang (2004), apresenta uma relação de motores de busca, com seus respectivos tipos de funcionamento.

Tabela 8 – Diferentes tipos de motores de busca

MOTOR DE BUSCA	TIPOLOGIA	DISPONÍVEL EM:
Aol	Fornecer resultados de busca baseado no <i>crawler</i> do Google	http://search.aol.com/aol/webhome
Ask	Fornecer resultados de busca baseado no <i>crawler</i> do Teoma	http://www.ask.com/
Bing	Baseado em <i>crawler</i>	http://www.bing.com/
Dmoz	Diretório mantido por humanos	http://www.dmoz.org/

Dogpile	Metabuscador	http://www.dogpile.com/
Excite	Metabuscador	http://msxml.excite.com/
Google	Baseado em <i>crawler</i>	http://www.google.com
Hotbot	Fornece resultados de busca baseado no <i>crawler</i> do Google e do Teoma	http://www.hotbot.com/
Kartoo	Metabuscador	http://www.kartoo.com/
LookSmart	Diretório mantido por humanos	http://www.looksmart.com/
Mamma	Metabuscador	https://www.mamma.com/
NetscapeSearch	Fornece resultados de busca baseado no <i>crawler</i> do Google	http://isp.netscape.com/
Teoma	Baseado em <i>crawler</i>	http://www.teoma.com/
Yahoo	Baseado em <i>crawler</i>	https://search.yahoo.com/

Fonte: elaborada pela autora.

No escopo desta investigação, será proposto um motor de busca baseado em *crawler* para recuperar metadados de objetos de aprendizagem de repositórios específicos para este fim. Este *web crawler* será baseado em agentes inteligentes e foi denominado de Agente Buscador, conforme relatado no Capítulo 7.

5.3.2 Tipos de interface de busca de informação

A visualização da informação aplicada à área de recuperação da informação viabiliza explorar as capacidades cognitivas humanas e busca reduzir o esforço do usuário na interação com o SRI. Neste sentido, as interfaces de busca são consideradas como aliadas da visualização da informação recuperada, pois permitem que o usuário visualize os dados oriundos de um sistema de recuperação de informação. Zhang (2008, p.5) destaca dois paradigmas que refletem o comportamento do usuário na busca de informação que são:

- *Browsing* – é um mecanismo importante para explorar e descobrir informação. Neste paradigma, o usuário precisa ter uma noção clara sobre sua necessidade de informação, pois sua busca tem que partir de uma página conhecida e prossegue por meio da navegação entre os *links* estabelecidos pela página inicial;

- *Query searching* – é considerado como uma tarefa mais complexa que envolve a articulação de uma necessidade de informação dinâmica dentro de um grupo lógico de palavras-chave relevantes. Neste paradigma, o usuário traduz sua necessidade de informação por meio de uma expressão de busca que é submetida ao SRI que examina em sua base os itens que provavelmente correspondem à solicitação realizada.

De acordo com Cove e Walsh (1988) o browsing pode ser classificado em três categorias:

- Browsing de busca (*Search browsing*) – quando se tem ciência do que deseja pesquisar;
- Browsing de propósito geral (*General purpose browsing*) – quando a pesquisa é realizada em uma base para buscar fontes que as informações de interesse;
- Browsing imprevisto (*Serendipity browsing*) – quando a pesquisa é realizada inesperadamente, de forma desestruturada e não direcionada.

Conforme Zhang (2008, p. 15), a visualização aplicada ao processo de recuperação da informação pode transformar a maneira de encontrar informações em uma tarefa simples e intuitiva, uma vez que facilita a interação do usuário com SRI e não exige conhecimentos técnicos complexos nem grandes esforços cognitivos.

Zhang (2008, p. 16) considera que a aplicação de técnicas de visualização na recuperação da informação pode ser orientada por:

- *Query searching and browsing* (consulta-navegação) – o processo interação é iniciado por meio de uma expressão de busca inicial que é submetida ao SRI que retornará um conjunto de resultados em uma interface de navegação (visualização) pela qual o usuário continuará a refinar sua busca até que sua necessidade de informação tenha sido atendida;
- *Browsing and query searching* (navegação-consulta) – o usuário entra em contato com um conjunto de dados por meio de uma interface de navegação. Em seguida, os usuários submetem suas consultas para o

ambiente de visualização e os resultados correspondentes são apresentados dentro de contextos de apresentação visual;

- *Browsing only* (apenas-navegação) – o usuário possui apenas a visão global do conteúdo do SRI, no qual ele navega em busca de informações mais específicas.

Ressalta-se que somente *query searching* não foi categorizada como uma técnica isolada porque é considerada como um paradigma de recuperação de informação tradicional e, portanto, não requer um espaço visual para sua compreensão. Nesta investigação, optou-se pela técnica de visualização *Query searching and browsing* uma vez que, percebe-se a necessidade de interação do usuário na seleção dos itens que ele julga como relevantes.

5.4 Avaliação dos Sistemas de Recuperação de Informação

A avaliação de um sistema de recuperação de informação é julgada como um procedimento importante para os resultados retornados por um SRI. Meadow (1993) em sua obra “*Text Information Retrieval Systems*” elencou algumas medidas baseadas em relevância, processo e resultado que podem ser utilizadas para avaliação de um sistema de recuperação de informação. Essas medidas foram compiladas por Martínez Méndez (2002, p. 71-72) e são apresentadas na Tabela 9.

Tabela 9 – Medidas para avaliação de um sistema de recuperação de informação

CRITÉRIO	DEFINIÇÃO	BASEADA EM:		
		RELEVÂNCIA	PROCESSO	RESULTADO
Precisão	É a razão entre o número de registros relevantes recuperados e o número total de registros recuperados.	X		X
Exaustividade	É a razão entre o número de documentos relevantes recuperados e o total de documentos relevantes.	X		X
Medida de Efetividade E-P	Média da efetividade em pares de valores de exaustividade e	X		X

	precisão.			
Seleção	Mede quantos documentos têm na base de dados, o grau de sobreposição com outras bases relacionadas e o que se espera da base de dados antes da busca.		X	
Conteúdo	Tipo de documentos da base de dados, temática dos documentos e frequência de atualização.		X	
Tradução de uma consulta	Verifica se o usuário pode realizar a consulta diretamente ou precisa de auxílio		X	
Erros no estabelecimento da consulta	Média de erros sintáticos na escrita da busca que proporcionam a recuperação de um conjunto de dados vazio ou errôneo		X	
Tempo médio de realização da busca	Tempo médio de realização de uma estratégia de busca		X	
Dificuldade na realização da busca	Na relação acima devem ser adicionados os problemas que os usuários iniciantes podem encontrar		X	
Número de comandos precisos em uma busca	Média de instruções necessárias para realizar uma busca		X	
Custo da busca	Custos diretos e indiretos na realização da busca		X	
Número de documentos recuperados	Extensão do resultado de uma busca		X	
Número de documentos revisados pelo usuário	Média dos documentos que os usuários estão dispostos a revisar		X	
Medida de satisfação do	Medidas que pretendem medir a reação dos usuários			X

usuário	diante do resultado de uma busca			
---------	----------------------------------	--	--	--

Fonte: adaptada de Martínez Méndez (2002, p. 71-72, tradução nossa).

Observando-se os parâmetros apontados na Tabela 9, percebe-se que os critérios baseados em relevância e nos resultados estão diretamente relacionados com a arquitetura proposta nesta investigação.

Segundo Saracevic (1996, p. 57), ao longo de muitos anos a eficácia dos SRI foi computada apenas em termos de relevância e utilidade da informação. Entretanto, com a dimensão da explosão informacional propiciada pela *Web* outros critérios (qualidade, seletividade, veracidade, síntese, e/ou impacto da informação) estão sendo como parâmetros para avaliar os SRI. Dessa forma, Saracevic (2007) sumariza alguns tipos de relevância, conforme pode ser visualizado no Quadro 19.

Quadro 19 – Tipos de relevância para um SRI.

TIPO DE RELEVÂNCIA	DESCRIÇÃO
De sistema ou algorítmica	Estabelece a relação entre uma consulta e os objetos de informação (<i>corpus</i>) recuperados ou não, por um determinado procedimento ou algoritmo.
Temática	Estabelece a relação entre o tema solicitado em uma consulta e o tema coberto no <i>corpus</i> recuperado.
Cognitiva ou pertinência	Estabelece a relação entre o estado de conhecimento e a necessidade de informação cognitiva de um usuário e o <i>corpus</i> recuperado.
Situacional ou de utilidade	Estabelece a relação entre a situação, tarefa ou problema e o <i>corpus</i> recuperado.
Afetiva ou motivacional	Estabelece a relação entre as intenções, objetivos e motivações de um usuário e o <i>corpus</i> recuperado.

Fonte: adaptada de Saracevic (2007, tradução nossa).

Normalmente, os sistemas de recuperação de informação são avaliados através de consultas em *corpus* de referência do objeto em estudo, pois os documentos devem ser ranqueados para serem apresentados para os usuários de acordo com o critério de busca que foi formulado pelo próprio usuário.

Duas métricas mais comumente utilizadas para avaliar a eficiência dos resultados obtidos pelo sistema de recuperação de informação são:

- Precisão (*precision*): é a proporção entre o número de itens relevantes recuperados em uma consulta sobre o total de itens retornados como resposta a uma busca, conforme expressão a seguir:

$$P = \frac{|{\{itens\ relevantes\}} \cap {\{total\ de\ itens\ recuperados\}}|}{|{\{total\ de\ itens\ recuperados\}}|}$$

- Revocação (*recall*): é a proporção entre a quantidade de itens relevantes recuperados em uma consulta sobre o número total de itens relevantes na coleção, conforme expressão a seguir:

$$R = \frac{|{\{itens\ relevantes\}} \cap {\{total\ de\ itens\ recuperados\}}|}{|{\{total\ de\ itens\ relevantes\ na\ coleção\}}|}$$

Assim, quanto mais exaustivamente um sistema de recuperação de informação indexar seu *corpus*, maior será sua revocação (número de documentos recuperados) na busca e, inversamente proporcional, a sua precisão será menor. Dessa forma, pode-se considerar que relevância é a característica central de um SRI e é o que distingue Sistemas de Recuperação de Informação de Sistemas de Recuperação de Dados.

Neste capítulo foi apresentado o estado da arte referente à recuperação de informação com ênfase na *Web*, assim como o desempenho dos principais modelos de recuperação de informação. Do mesmo modo, foram demonstradas as principais diferenças entre tipos de sistemas de recuperação de informação e quais as medidas de desempenho que são adotadas para avaliá-los.

No próximo capítulo, o processo de *relevance feedback* é descrito para compreender como ele será inserido na arquitetura para recuperar objetos de aprendizagem. O processo de *relevance feedback* permite a interação do usuário no julgamento da informação que lhe foi retornada, por meio de refinamentos sucessivos, até que o resultado retornado pelo sistema de recuperação de informação seja significativo para sua necessidade de informação.

6. Relevance Feedback (RF)

A relevância é um dos conceitos centrais de um sistema de recuperação de informação e está diretamente relacionada com o ponto de vista do usuário na seleção da informação desejada naquele momento. Neste contexto, para avaliar a relevância da informação recuperada por um SRI é necessário que ocorra a participação do usuário no julgamento das informações recuperadas.

O *corpus* a ser recuperado pelo SRI, geralmente, é desconhecido pelo usuário. Neste sentido, considera-se que as informações descritas durante o processo de descrição do objeto de aprendizagem, quando apresentadas ao usuário, podem auxiliá-lo no julgamento dos itens que tenham uma maior probabilidade de serem relevantes para sua busca.

Desta maneira, acredita-se que o processo de *relevance feedback* quando aplicado ao domínio dos objetos de aprendizagem pode facilitar a compreensão do usuário de sua necessidade de informação. Com esse processo é possível estabelecer um diálogo do usuário com o sistema de recuperação de informação que visa melhorar o ranqueamento dos itens retornados sem que ele precise alterar sua expressão de busca inicial. Este processo permite selecionar os itens julgados como relevantes e realizar uma nova submissão ao SRI para que um novo ranqueamento seja realizado com base na inferência do usuário.

Segundo Rocchio (1971), *relevance feedback* faz parte do domínio de recuperação de informação e é definido como um processo no qual as consultas de um usuário de um SRI são incrementadas automaticamente a partir de consultas iniciais formuladas pelo usuário com base em julgamentos de relevância para os documentos recuperados de um *corpus*. Contribuindo com esta mesma visão Buckley e Salton (1995), afirmam que *relevance feedback* é um processo automático de refinamento a uma busca (inicial), que utiliza informações fornecidas pelo usuário sobre a relevância dos documentos previamente recuperados buscando apresentar resultados cada vez mais apurados ao usuário.

Na opinião de Salton e Buckley (1997) as vantagens de RF incluem:

- Proteger os usuários dos detalhes do processo de reformulação da consulta;
- Dividir a operação de busca em uma sequência de pequenas etapas de pesquisa, projetadas para abordar gradualmente o assunto desejado;
- Proporcionar um processo de alteração de busca controlada, desenhado para enfatizar alguns termos e enfatizar outros, como exigido em ambientes de pesquisa específicos.

O *relevance feedback* tem por objetivo aumentar a relevância e o desempenho do processo de recuperação de informação de acordo com a necessidade de informação do usuário. Conforme Lavrenko e Croft (2001) e Zhai e Lafferty (2001), esta técnica têm se mostrado eficiente no incremento da precisão dos sistemas de recuperação de informação.

Para Manning, Raghavan e Schütze (2008, p. 178), a ideia do *relevance feedback* é envolver o usuário no processo de recuperação de informação com a intenção de melhorar o conjunto de resultados apresentados. Nesse processo, o usuário emite seu *feedback* sobre o conjunto de documentos recuperados apresentados. O procedimento básico para realização de *relevance feedback* é descrito pelos autores da seguinte forma:

- O usuário realiza uma busca;
- O sistema retorna um conjunto (inicial) de resultados;

- O usuário analisa e marca (seleciona) alguns documentos como relevantes (ou não relevantes) e submete ao sistema de recuperação de informação;
- O sistema processa a melhor representação da informação necessária baseada no *feedback* do usuário;
- O sistema apresenta um novo conjunto de itens ranqueados baseado na inferência do usuário.

O processo de *relevance feedback* pode passar por uma ou mais iterações²⁹ do tipo descrito acima. O processo explora a ideia de que pode ser difícil de formular uma boa consulta quando você não conhece bem o assunto, mas pode ser mais fácil julgar sobre os documentos recuperados, e por isso faz sentido realizar um refinamento iterativo. Em tal cenário, *relevance feedback* também pode ser eficaz no acompanhamento evolutivo da necessidade de informações de um usuário: a busca de alguns documentos pode levar o usuário a um refinamento de sua compreensão das informações que está buscando.

Geralmente, os esquemas de *relevance feedback* são considerados em dois tipos diferentes de ações: as que alteram a consulta originalmente emitida pelo usuário (possivelmente considerando múltiplos pontos de busca) e; as que modificam a medida de similaridade usada para classificar e recuperar dados em um sistema de recuperação de informação (DOULAMIS; DOULAMIS, 2006). Nesta tese, optou-se em modificar a medida de similaridade e o ranqueamento dos itens com base na identificação dos elementos (metadados) comuns que compõe a seleção do usuário, como por exemplo o tipo de formato do objeto de aprendizagem ou o autor.

Assim, no escopo desta investigação, por meio de *relevance feedback* será possível realizar uma nova busca em conformidade com a realimentação de informações fornecidas pelo usuário, no qual ele informará, quais dos objetos de aprendizagem foram considerados por ele como relevantes, harmonizando assim o resultado da busca de acordo com a necessidade de informação do usuário.

²⁹ Iteração – sinônimo de recursivo, ou seja, pode ser executado inúmeras vezes.

De acordo com Hersh (2009, p. 205), *relevance feedback* possibilita inserir novos documentos ao resultado de uma busca com base em sua similaridade em relação aos documentos que foram considerados como relevantes pelo usuário, além de proporcionar que ocorra uma redefinição dos pesos dos documentos relevantes recuperados.

Ressalta-se que o processo de *relevance feedback* é bem interessante, porém está diretamente relacionado com a espontaneidade do usuário em informar seu posicionamento em relação aos documentos que ele julga como relevantes para realizar uma nova submissão ao SRI. Se o usuário não indicar os documentos que ele julgar como relevantes, essa funcionalidade não será aplicada. Williamson (1978) sugere que a maior parte do mérito para o processo de RF é devido às decisões individuais do usuário, e não devido à capacidade do SRI em fazer uma contribuição adicional.

Outra consideração importante relacionada ao processo de *relevance feedback* é que ele não pode ser aplicado na formulação de consulta inicial do usuário, uma vez que não terá nenhum resultado disponível para que a relevância seja indicada. Portanto, apesar de todos os benefícios deste processo é preciso ter conhecimento sobre suas limitações, caso o usuário se recuse a fornecer sua opinião ao sistema de recuperação de informação.

6.1 Tipos de *Relevance Feedback*

Comumente as técnicas de *relevance feedback* são realizadas em tempo real para fornecer resultados mais precisos com vistas a atender a necessidade de informação do usuário em um determinado momento.

A classificação dos tipos de *relevance feedback* é muito relativa e varia de acordo com a concepção de cada autor. De acordo com Baeza-Yates e Ribeiro-Neto (2013) existem duas abordagens básicas para um ciclo RF que são:

- Explícito: quando o usuário interage diretamente com o processo de recuperação de informação por meio da identificação dos documentos relevantes para aumentar o peso desses documentos em uma nova formulação de consulta. Espera-se que os documentos mais relevantes sejam ranqueados e apresentados prioritariamente ao usuário.

Experimentos iniciais foram realizados com o modelo vetorial (sistema Smart) (SALTON, 1971) e mais tarde com o modelo probabilístico (ROBERTSON e JONES, 1976).

- Implícito: quando a informação para a reformulação da consulta é derivada implicitamente pelo sistema, sem a intervenção direta do usuário. As duas abordagens básicas para essa categoria são: obter a informação de realimentação dos documentos do topo do *ranking* no conjunto de resultados, que é denominada de análise local; e obter a informação de realimentação a partir de fontes externas (tesauros, relação de termos extraídos dos documentos etc.), que é chamada de análise global.

Na concepção de Manning, Raghavan e Schütze (2008, p.177) esta abordagem é classificada da seguinte forma:

- *Pseudo (ou blind) relevance feedback*: fornece um método local para a análise automática. Esse método automatiza a parte manual de *relevance feedback*, de modo que o usuário obtém um melhor desempenho de recuperação sem que haja interação direta. O método consiste em fazer a recuperação normal e encontrar um conjunto inicial de documentos mais relevantes para, em seguida, assumir que o topo k de documentos ranqueados são relevantes e, finalmente, fazer um *relevance feedback* antes dessa suposição. Ruthven e Lalmas (2003) concluem que essa técnica funciona bem para 'boas' consultas iniciais – aquelas que foram eficientes em recuperar documentos relevantes – e mal para consultas iniciais 'ruins' – aquelas que não foram eficientes em recuperar documentos relevantes. Existem duas soluções possíveis para este problema: ou melhorar a classificação inicial, de modo que ocorra uma maior probabilidade de documentos relevantes sendo usados para modificar a consulta, ou melhorar a detecção de características relevantes, ou seja, desenvolver melhores técnicas de *relevance feedback*;
- *Relevance feedback* indireto (ou global): ocorre quando são usadas fontes indiretas de evidência ao invés de *feedback* explícito na relevância como

base para o RF. Isso é muitas vezes chamado *feedback* implícito que é menos confiável do que o *feedback* explícito, mas é mais útil do que o *pseudo relevance feedback*, que não contém nenhuma evidência de decisões do usuário. Além disso, enquanto os usuários são muitas vezes relutantes em fornecer *feedbacks* explícitos, é mais fácil coletar *feedbacks* implícitos em grandes quantidades, para um sistema volumoso, como é o caso dos motores de busca da *Web*.

Após a avaliação das diferentes classificações propostas na literatura, optou-se em adotar o *relevance feedback* explícito (Baeza-Yates; Ribeiro-Neto, 2013) pelo fato de que nesta investigação a opinião do usuário no julgamento da informação recuperada pode contribuir para que o processo de recuperação de objetos de aprendizagem seja mais relevante.

6.2 Avaliação de *Relevance Feedback*

A avaliação de um SRI com suporte ao *relevance feedback* pode ser realizada de várias formas. O ideal é que se possa medir e comparar o desempenho do sistema com e sem o uso de *relevance feedback*.

Uma abordagem simples para a avaliação de RF seriam as medidas de revocação (*recall*) e precisão (*precision*) (ver seção 5.4) antes e depois do RF e comparar ambos. Esta técnica foi proposta por Chang, Cirillo e Ranzon (1971), porém foi identificado que esta não é uma boa medida de melhoria, pois como o desempenho pode aumentar simplesmente porque os documentos classificados como relevantes pelo usuário serão ranqueados prioritariamente no conjunto de recuperação de *feedback* e os documentos irrelevantes consequentemente descartados. De acordo com Orengo (2004), isso é conhecido como o "efeito de classificação" e leva a melhorias dramáticas mascarando a vantagem real proporcionada pelo processo RF.

Uma medida da eficácia de um sistema de *relevance feedback* deve refletir o número de documentos relevantes recentemente recuperados devido ao processo de *feedback*. Alguns métodos têm sido propostos com o objetivo de medir a real melhoria proporcionada por RF. De acordo com Orengo (2004), as abordagens mais conhecidas são descritas a seguir:

- Congelamento completo (*Full Freezing*): Esse método, usado pelo sistema SMART (SALTON, 1971), congela as fileiras de todos os documentos apresentados ao usuário nas iterações de realimentação e atribui ao primeiro documento recuperado pela i -ésima iteração um *rank* $iN + 1$ (onde N são os documentos são apresentados ao usuário para a seleção do *feedback*). Essa método fornece resultados que são enganosamente baixos, uma vez que os *rankings* superiores são congelados, portanto, nenhum documento recém-recuperado pode alcançar uma classificação mais elevada do que os documentos recuperados em iterações anteriores.
- Congelamento modificado (*Modified Freezing*): Este método, proposto por Ide (1971), difere do congelamento completo, em que documentos não relevantes classificados abaixo do último documento relevante não são congelados, portanto, o número de documentos congelados em cada iteração varia. Esta técnica fornece resultados mais realistas do que o método anterior. No entanto, ele só pode ser usado para *feedback* positivo, ou seja, os documentos julgados como irrelevantes não têm suas fileiras congeladas, o que lhes permite serem usados novamente para modificar a consulta.
- Método de coleta residual (*residual collection method*): Esta é a abordagem mais utilizada para a avaliação de RF. Neste método, os documentos que foram julgados pelo usuário são excluídos da coleta e das avaliações de relevância. A segunda lista classificada conterá apenas documentos que não foram julgados como relevantes. Este método fornece uma avaliação imparcial de RF, no entanto, a eliminação de documentos altamente relevantes que estavam presentes na execução inicial irá reduzir números de revocação e de precisão. Portanto, não é possível comparar diretamente os resultados deste método com os resultados calculados usando a técnica de Chang, Cirillo e Ranzon (1971).

Neste método de coleta residual, a avaliação dos resultados compara apenas as coleções residuais, ou seja, a execução inicial é refeita menos os documentos mostrados anteriormente ao usuário. Este método fornece uma avaliação imparcial e mais realista

do *feedback*. No entanto, uma vez que os documentos relevantes altamente classificados foram removidos da coleta residual, os números de precisão e revocação são geralmente inferiores aos dos métodos de avaliação padrão e não podem ser diretamente comparados com o desempenho medido pelo método de avaliação padrão.

Portanto, o processo de *relevance feedback* é bastante eficiente para refinar a busca do usuário. No entanto, para que este processo ocorra de forma satisfatória é necessário que o *corpus* a ser recuperado esteja bem persistido e que o usuário esteja disposto a interagir com o sistema de recuperação de informação. No capítulo a seguir será apresentada a arquitetura proposta nesta tese para recuperação de objetos de aprendizagem.

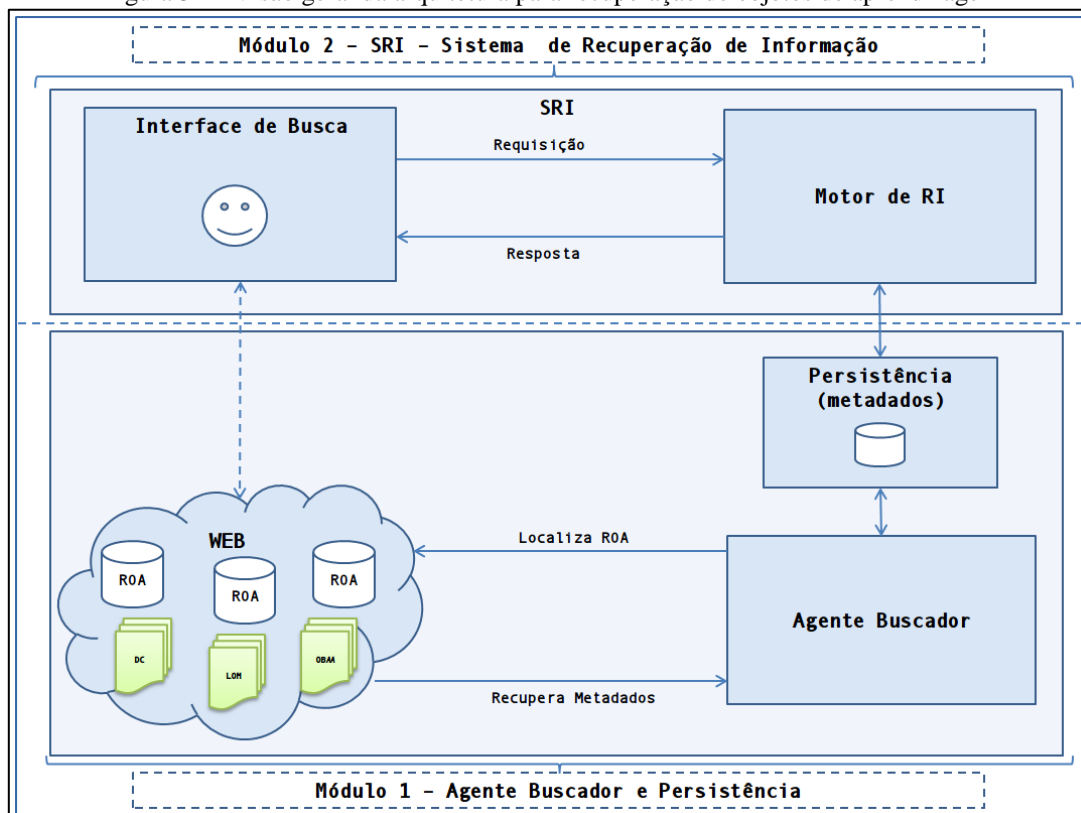
7. Arquitetura para Recuperação de Objetos de Aprendizagem

O propósito fundamental da arquitetura para recuperação de objetos de aprendizagem é unificar a representação desses recursos educacionais que são descritos e disponibilizados em repositórios específicos para este fim para permitir que o usuário realize buscas qualificadas e localize os objetos de aprendizagem mais adequados para sua necessidade de informação.

Portanto, para atingir este propósito a arquitetura para recuperação objetos de aprendizagem foi fundamentada com base em uma integração de tecnologias de sistemas de recuperação de informação, metadados, *relevance feedback* e agentes inteligentes.

Para compreender o funcionamento desta arquitetura é apresentado o modelo desenvolvido ao longo desta tese, conforme pode ser visualizado na Figura 31. A arquitetura foi projetada, em dois módulos: o primeiro é denominado “Agente Buscador e Persistência” dos metadados de objetos de aprendizagem que foram recuperados a partir dos repositórios. O segundo módulo é o “Sistema de Recuperação de Informação” que engloba a interface com o usuário e o motor de recuperação de informação. Este módulo é responsável pela comunicação do usuário com o SRI.

Figura 31 – Visão geral da arquitetura para recuperação de objetos de aprendizagem

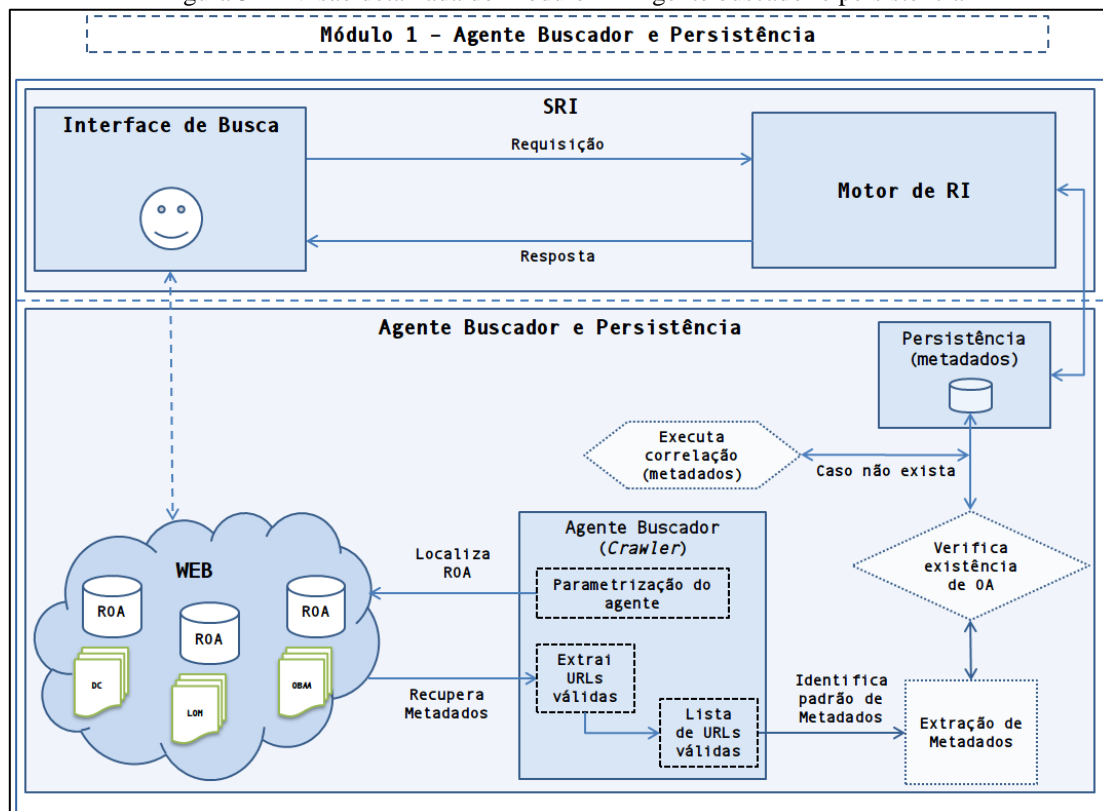


Fonte: elaborado pela autora.

Observa-se, na Figura 31, que os dois módulos que compõem a arquitetura interagem por meio dos metadados que foram persistidos quando da recuperação pelo agente buscador. Nesta arquitetura, pode-se considerar que o módulo 2 é executado *on-line*, ou seja, no momento em que o usuário solicitar a busca, enquanto o módulo 1 pode ser executado *off-line*, isto é, em um tempo distinto ao da demanda do usuário.

O módulo 1 tem como objetivo buscar repositórios de objetos de aprendizagem e localizar os recursos informacionais por meio da identificação de seu padrão de metadados. Atua, em um primeiro momento, como um *crawler* inteligente que é parametrizado para descobrir, buscar e extrair informações sobre os metadados de objetos de aprendizagem nos repositórios da *Web*. A Figura 32 apresenta a arquitetura proposta para esse módulo, cuja descrição completa é apresentada na seção 7.2.

Figura 32 – Visão detalhada do Módulo 1 – Agente buscador e persistência



Fonte: elaborado pela autora.

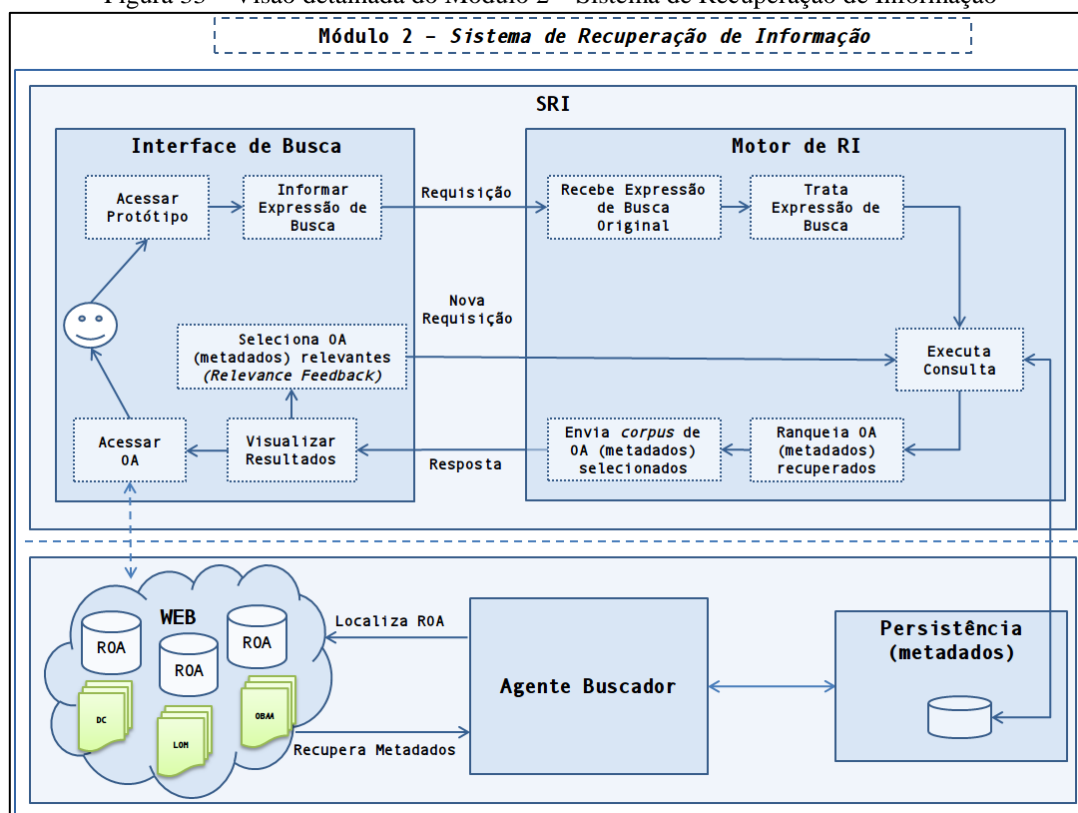
O Módulo 2 (Figura 33) refere-se ao processo de recuperação de informação. O usuário acessa a interface de recuperação de objetos de aprendizagem e traduz sua necessidade de informação por meio de uma expressão de busca. O usuário submete sua requisição ao motor de recuperação de informação que, por sua vez, retorna um conjunto de itens recuperados e ranqueados.

O usuário pode analisar e avaliar os itens recebidos e selecionar (a partir das informações que lhe são apresentadas, como por exemplo: título, palavras-chave, autor, formato etc.) os que lhe parecem como mais relevantes naquele momento. A partir desta seleção, o usuário faz uma nova requisição ao motor de recuperação de informação baseada nos itens que ele julgou como relevantes. Este processo dialógico entre o usuário e o sistema de recuperação de informação é denominado de *relevance feedback*.

O motor de RI recebe a informação, calcula uma nova representação de sua necessidade de informação e retorna novos resultados baseado em melhores taxas de recuperação. Esse processo é cíclico e pode ser executado inúmeras vezes até que o usuário esteja satisfeito com a resposta que lhe foi retornada. A Figura 33 apresenta uma

visão detalhada das fases desse processo. A descrição completa do funcionamento do módulo de Sistema de Recuperação de Informação é apresentada na seção 7.3.

Figura 33 – Visão detalhada do Módulo 2 – Sistema de Recuperação de Informação



Fonte: elaborado pela autora.

Para compreender o escopo geral desta arquitetura são apresentados alguns modelos de sistema, representados através dos diagramas de casos de uso, que tem o potencial de descrever o cenário proposto nesta investigação com suas respectivas funcionalidades, conforme Apêndice A. Na seção a seguir será indicado como ocorre à representação do processo de recuperação de informação nesta arquitetura.

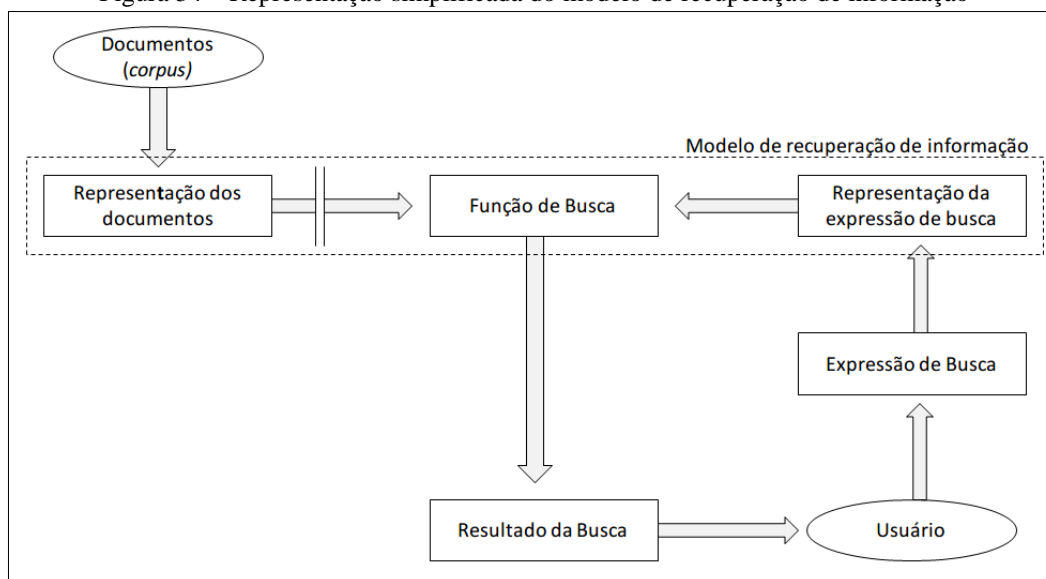
7.1 Representação do processo de recuperação de informação

O processo de recuperação de informação tem por premissa básica descobrir itens de um *corpus* para apresentar ao usuário os resultados que mais se aproximam de sua necessidade de informação que deve ser formalizada por meio de uma função de busca.

A representação do processo de recuperação de informação proposto nesta tese é baseado no modelo recomendado por Ferneda (2012, p. 14), conforme Figura 34. Neste

modelo são consideradas as seguintes etapas: representação dos documentos, a representação da expressão de busca e a função de busca.

Figura 34 – Representação simplificada do modelo de recuperação de informação



Fonte: Ferneda (2012, p. 14)

Nas subseções a seguir, serão detalhados os elementos que compõem o modelo de recuperação de informação no escopo desta investigação.

7.1.1 Representação dos documentos

No caso desta investigação o *corpus* documental é composto pelos registros de metadados que foram recuperados dos repositórios de objetos de aprendizagem. Como cada repositório possui seu próprio padrão de metadados foi necessário estabelecer uma correlação entre esses padrões, conforme foi descrito da subseção 3.2 desta tese. A unificação obtida com a correlação dos metadados permite ampliar o escopo das buscas do usuário.

Os metadados são recuperados por meio de um *crawler* parametrizado que extrai as URLs válidas (aquelas que apresentam informações de metadados) em um arquivo de log. A partir da lista de URLs válidas, retornada pelo *crawler*, identifica-se o padrão de metadados e os metadados são extraídos. Verifica-se a existência do objeto de aprendizagem, caso o mesmo não esteja persistido, executa-se a correlação e armazena as informações dos metadados. Os metadados são persistidos em um modelo abstrato de metadados padrão XML.

7.1.2 Representação da expressão de busca

Para que o usuário se comunique com o sistema de recuperação de informação é necessário que ele acesse uma interface de busca e demonstre sua necessidade de informação por meio da inserção de uma expressão de busca que será submetida ao motor de recuperação de informação para localizar os itens que sejam do provável interesse do usuário.

A representação da expressão de busca está relacionada com a forma como o usuário comunica sua necessidade informacional. Essa necessidade informacional deve ser expressa de forma similar ao da representação dos documentos que compõe o *corpus*, que neste caso ambos são expressos em linguagem natural. Quanto mais homogênea for a expressão de busca com os documentos do *corpus* melhor serão os resultados retornados ao usuário.

Assim sendo, o usuário inicia o processo de comunicação no momento em que informa uma expressão de busca inicial para que o sistema de recuperação de informação processe a requisição e devolva um conjunto de itens recuperados com base na similaridade entre a expressão de busca informada e os registros de metadados.

A partir da análise dos itens recuperados o usuário pode informar ao SRI quais itens são relevantes, em sua concepção, e o sistema de recuperação de informação recalcula os valores de relevância presumida e apresenta um novo ranqueamento com base no *feedback* do usuário com vistas a adequar os resultados à necessidade de informação do usuário.

7.1.3 Função de busca

A função de busca está situada no centro do processo de recuperação de informação. A função de busca compara o grau de similaridade entre a expressão de busca informada pelo usuário e os elementos de metadados de objetos de aprendizagem que foram recuperados dos repositórios.

Para ranquear os itens recuperados pelo grau de similaridade, nesta tese, optou-se em atribuir pesos aos elementos de metadados, conforme pode ser observado na Tabela 10. Os pesos foram distribuídos a partir de uma análise de relevância dos elementos de metadados e foi atribuído um peso maior aos três primeiros elementos,

pois estes estão diretamente relacionados com a descrição do conteúdo informacional do objeto de aprendizagem, fator este que contribui para a recuperação mais relevante da informação, uma vez que existe a necessidade de ranqueamento dos metadados de objetos de aprendizagem.

Tabela 10 – Definição de pesos aos elementos de metadados.

ELEMENTO DE METADADOS	PESO
arq_ROA:title	1
arq_ROA:subject	0.8
arq_ROA:description	0.6
arq_ROA:types	0.2
arq_ROA:source	0.4
arq_ROA:relation	0.2
arq_ROA:coverage	0.4
arq_ROA:creator	0.4
arq_ROA:publisher	0.4
arq_ROA:contributor	0.4
arq_ROA:rights	0.2
arq_ROA:date	0.2
arq_ROA:identifier	0.0
arq_ROA:language	0.4

Fonte: elaborado pela autora.

Com base nos pesos estabelecidos é calculado o grau de relevância para cada item recuperado. Desta maneira, é possível ranquear os itens recuperados em ordem decrescente de uma relevância presumida. A função de similaridade adotada nesta investigação é explicada na seção 7.3.

7.2 Módulo 1 – Agente Buscador e Persistência

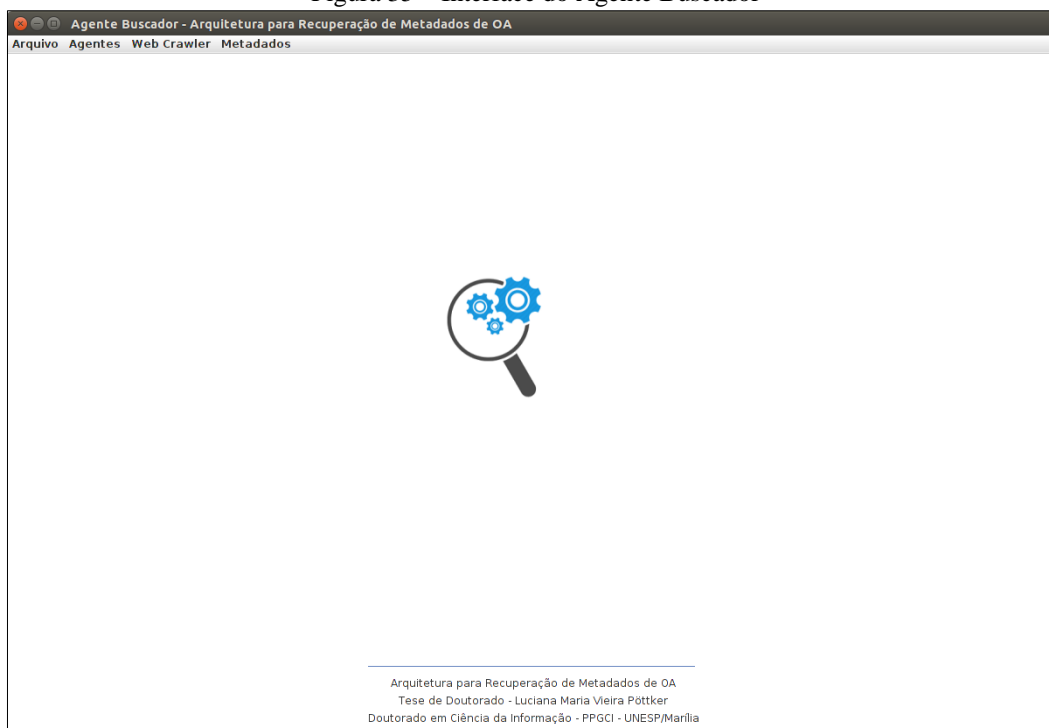
Nesta seção são apresentados os fundamentos do módulo 1 da arquitetura proposta para recuperação de objetos de aprendizagem que está dividida em duas partes: a primeira que descreve o funcionamento do agente buscador e a segunda que explica como os metadados são persistidos.

7.2.1 Agente Buscador

O módulo do agente buscador compreende em uma integração de recursos desenvolvidos para recuperar, extrair e correlacionar os metadados de objetos de aprendizagem disponíveis em repositórios.

A Figura 35 apresenta a interface inicial deste módulo que foi elaborado no ambiente integrado de desenvolvimento Eclipse, em linguagem Java. Para programar os agentes foi utilizado o framework JADE³⁰ (*Java Agent Development Framework*). A comunicação entre os agentes ocorre por meio da linguagem de comunicação FIPA-ACL (ver seção 4.3).

Figura 35 – Interface do Agente Buscador



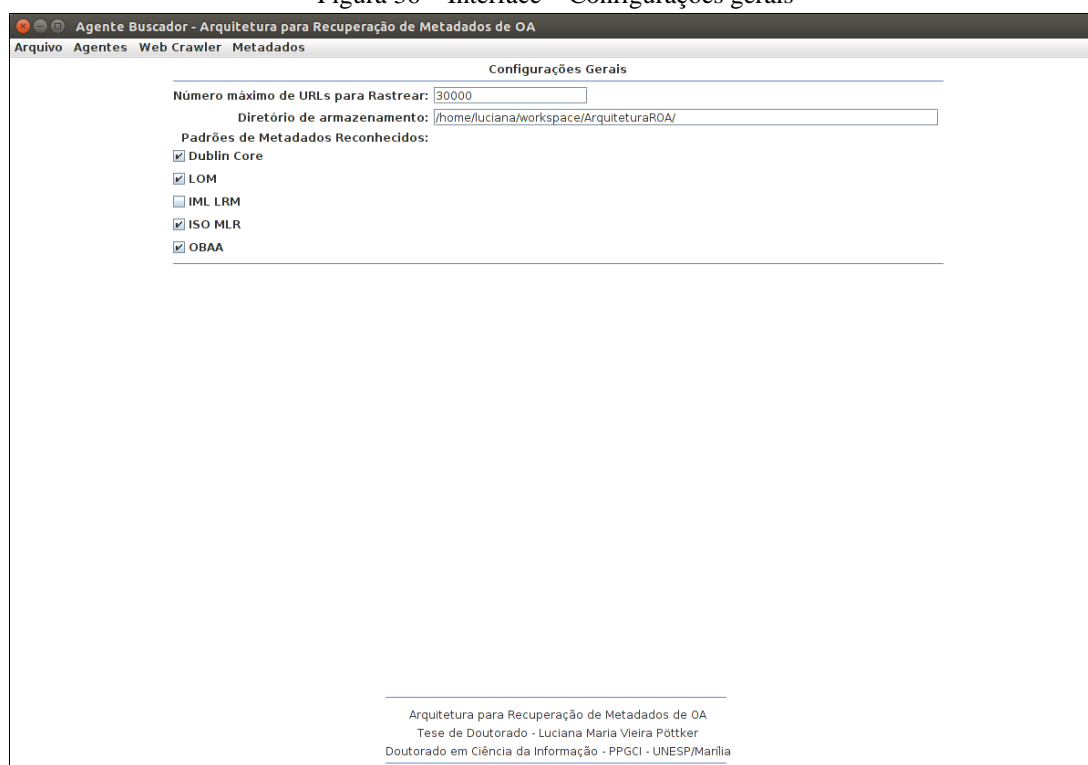
Fonte: elaborado pela autora.

A interface apresenta quatro menus que fornecem o acesso às funcionalidades deste módulo. O primeiro menu denominado de “Arquivo” apresenta as opções de configurações gerais, bem como permite o gerenciamento das informações onde o arquivo será armazenado e, também podem ser redefinidos os ajustes de número

³⁰ Jade – é uma biblioteca de classes Java que fornece o ambiente para o desenvolvimento dos agentes inteligentes seguindo as recomendações da FIPA. Desenvolvido originalmente pela Telecom Italia Lab, em 2000. Esta biblioteca é livre e é distribuída sob a licença General Public License (GNU) – Lesser General Public License (LGPL). Disponível em: <<http://jade.tilab.com/>>. Acesso em: 23 mar. de 2015.

máximo de URLs a serem rastreadas. A Figura 36 ilustra a interface de configurações gerais. Outra opção deste menu é a de finalização da aplicação.

Figura 36 – Interface – Configurações gerais



Fonte: elaborado pela autora.

O menu seguinte é o denominado de “Agentes” e apresenta as opções de controle dos agentes inteligentes, que incluem ações de: inicialização, parametrização e finalização dos agentes.

Para descrever o ambiente no qual os agentes devem desempenhar suas funções são utilizadas as propriedades indicadas no Capítulo 4. Portanto, para o escopo desta investigação o ambiente é classificado de acordo com seu: desempenho, ambiente, atuadores e sensores, conforme apresentado na Tabela 11.

Tabela 11 – Características do ambiente PEAS em que atuam os agentes inteligentes

TIPO DE AGENTE	PERFORMANCE (DESEMPENHO)	ENVIRONMENT (AMBIENTE)	ACTUATORS (ATUADORES)	SENSORS (SENSORES)
Agente buscador de OA	Maximizar a busca de objetos de aprendizagem em repositórios	ROA (<i>Web</i>)	Identificar inclusões, alterações e exclusões de objetos de aprendizagem	Identificar padrão de metadados e extrair link
Agente	Descobrir novos	<i>Web</i>	Identificar o	Reconhecer

descobridor de ROA	repositórios de objetos de aprendizagem para incrementar alcance de OA recuperados.		padrão de metadados adotado pelo repositório	novos ROA descritos por um padrão de metadados educacional
--------------------	---	--	--	--

Fonte: elaboração própria.

Com base no conhecimento dos ambientes suportados pelos agentes é possível apresentar uma visão geral das especificações e objetivos dos cenários suportados pelos agentes, conforme pode ser observado no Quadro 20.

Quadro 20 – Detalhamento geral do cenário suportado pelo agente buscador de OA

CENÁRIO	SEQUÊNCIA	TIPO	DESCRIÇÃO
INICIALIZAR	1	Percepção	Inicializar agente
	2	Objetivo	Preparar agente para receber parâmetros
	3	Ação	Exibir interface de parametrização
PARAMETRIZAR	1	Percepção	Identificar interface criada
	2	Objetivo	Coletar parâmetros
	3	Objetivo	Configurar arquivo de log para receber URLs
	4	Ação	Iniciar rastreamento
FINALIZAR	1	Percepção	Coleta concluída
	2	Objetivo	Armazenar links.
	3	Objetivo	Fechar arquivo de log
	4	Ação	Finalizar agentes

Fonte: elaborado pela autora.

Inicialmente, o “Agente buscador de OA” é parametrizado com informações referentes aos padrões de metadados, ao repositório e a quantidade máxima de URLs a serem recuperadas. O Apêndice B apresenta um pseudocódigo referente à inicialização deste agente. Após esta parametrização, inicia-se a busca pelas URLs que contêm os metadados de objetos de aprendizagem. Essa busca é realizada pelo cálculo de similaridade entre as informações de metadados e o conteúdo da URL.

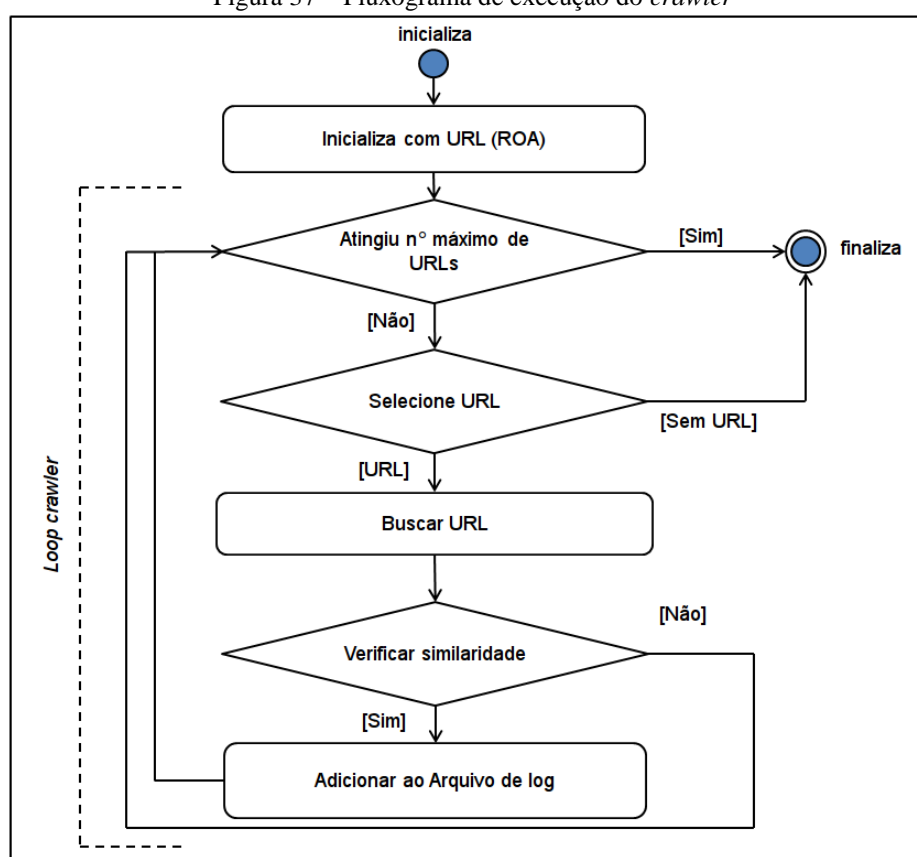
O “Agente descobridor de ROA” é parametrizado com informações relacionadas aos repositórios descritos a partir de um padrão de metadados. A partir daí ele busca por novos repositórios na *Web* baseado em sua capacidade de aprendizagem e adaptação para possibilitar a localização e a identificação de novos ROA.

Os menus “*Web Crawler*” e Metadados são descritos nas subseções 7.2.2 e 7.2.3, respectivamente.

7.2.2 *Web Crawler*

O *crawler*, proposto nesta tese, foi concebido para atuar em repositórios de objetos de aprendizagem e recuperar URLs que contém metadados de OA para sua futura recuperação em um SRI. A Figura 37 apresenta um fluxograma simplificado de cada segmento de execução do *crawler*.

Figura 37 – Fluxograma de execução do *crawler*



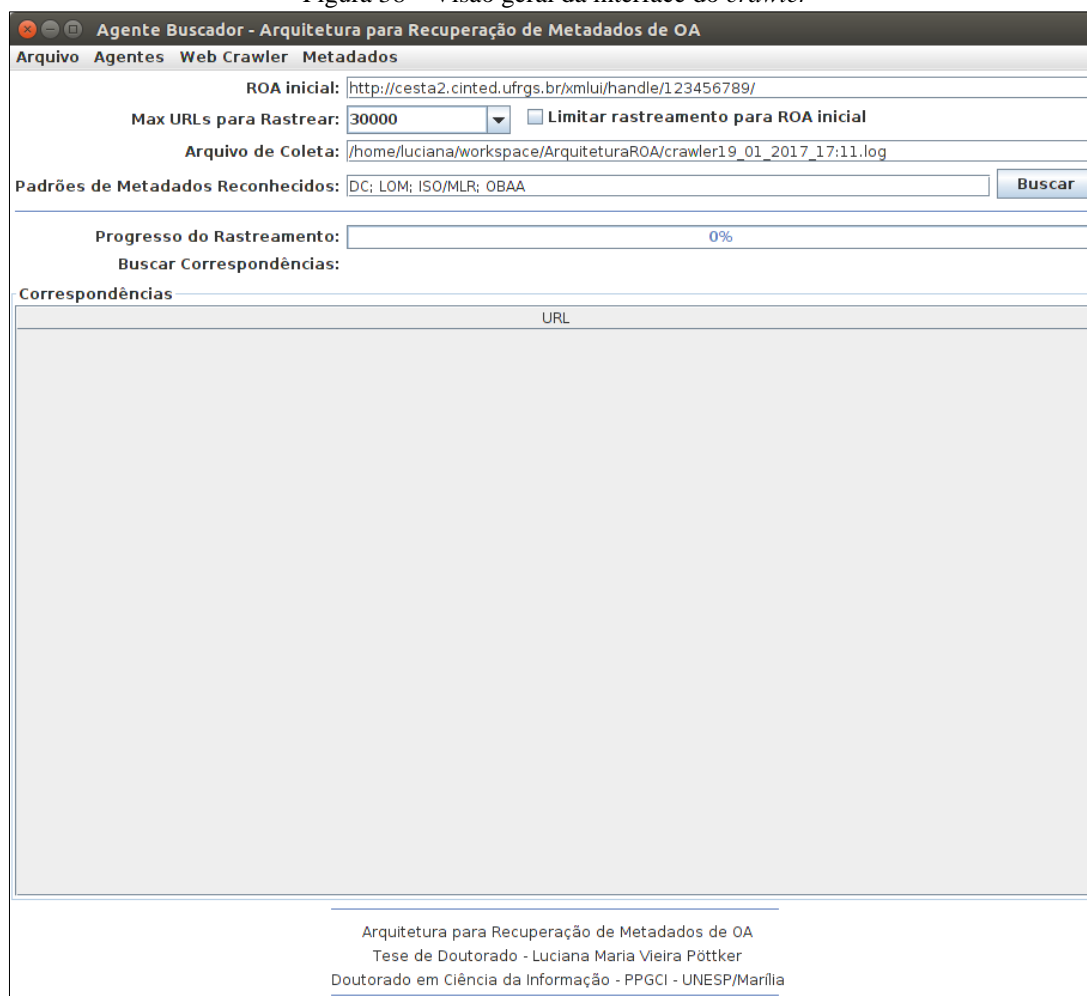
Fonte: elaborado pela autora.

A execução do *crawler* é iniciada após a parametrização dos agentes. A primeira verificação a ser realizada é descobrir se o número máximo de URLs a ser rastreada ainda não foi atingido, caso tenha sido atingido o *crawler* é finalizado. Se ainda existirem URLs a serem rastreadas, a mesma é localizada e verifica-se se existe similaridade entre os elementos de metadados e a URL, caso exista a URL é adicionada ao arquivo de log, caso contrário volta à verificação da ocorrência do número máximo de URLs. O *crawler* permanece em *looping* até que o número máximo de URLs seja

atingido ou não existiram mais URLs a serem rastreadas. O apêndice C apresenta o pseudocódigo referente ao *loop* de execução do *crawler*.

A Figura 38 apresenta a visão geral da interface em execução do *crawler*. Conforme explicado anteriormente, sua parametrização é realizada por meio dos agentes inteligentes.

Figura 38 – Visão geral da interface do *crawler*



Fonte: elaborado pela autora.

Como os ROA estão na *Web* existem características pertinentes a este ambiente que tornam o rastreamento mais complexo, como por exemplo: o grande volume de dados, a taxa de variação rápida e a geração dinâmica de páginas. Estas características combinadas produzem uma grande variedade de possíveis páginas rastreáveis. Portanto, o comportamento de um *crawler* deve ser composto por uma combinação de políticas, que incluem (CASTILLO, 2005):

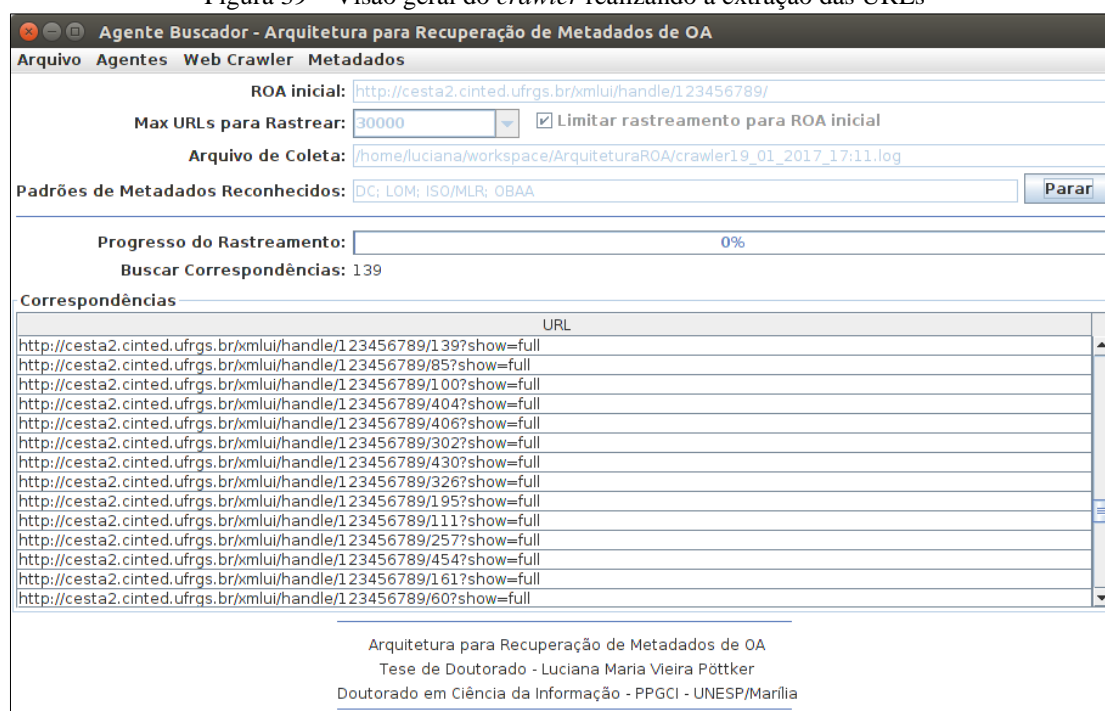
- Política de Seleção – no caso desta investigação serão informados, inicialmente, repositórios para que os metadados de objetos de aprendizagem sejam recuperados. Após essa seleção inicial o “Agente Descobridor de ROA” será parametrizado para buscar por novos repositórios. Esta é uma tarefa que possui um alto nível de dificuldade, já que deve funcionar com informações parciais e o conjunto de páginas é desconhecido durante o *crawling*;
- Política de Revisitação – A dinamicidade e volatilidade do ambiente *Web* requer a definição de uma política de revisitação para garantir que os metadados de objetos de aprendizagem que foram recuperados estejam sempre disponíveis. Portanto, os repositórios devem ser revistados em intervalos regulares para verificar alterações e manter o sistema atualizado. Para isso são definidos dois tipos de política de revisitação (CHO; GARCIA-MOLINA, 2003):
 - Política uniforme – revisita todas as páginas com a mesma frequência.
 - Política proporcional – revisita mais as páginas que são atualizadas com uma maior frequência.
- Política de cordialidade – Os *crawlers* podem obter informações de forma muito rápida e para evitar o *overloading*³¹ nos repositórios é necessário que o *crawler* seja organizado de tal forma que seja configurado um parâmetro do tipo “*crawl-delay*” para indicar o tempo mínimo (em segundos) entre requisições. Não existe um consenso entre esse tempo, Cho e Garcia-Molina (2003) definem o tempo de 10 segundos entre as requisições enquanto que Baeza-Yates e Castillo (2002) configuram o tempo de 15 segundos. Nesta investigação optou-se em adotar a recomendação de Cho e Garcia-Molina (2003) de 10 segundos (10000 milissegundos) para evitar a sobrecarga nos ROA. Esse parâmetro é dinâmico e pode ser alterado conforme a necessidade de rastreamento.

³¹ Sobrecarga nos servidores

- Política de Paralelização – a partir do momento em que o número de ROA aumentar é necessário coordenar o *crawler* de forma distribuída para que possa rodar múltiplos processos em paralelo com o objetivo de maximizar a taxa de *download* e minimizar custos da paralelização.

As políticas são implementadas diretamente via código fonte. Um fragmento do retorno do *crawler*, após a parametrização dos agentes para o ROA Cesta³² é apresentado na Figura 39. Os arquivos obtidos após a execução do *crawler* são usados como base para a extração dos metadados. No Apêndice D são apresentados fragmentos dos arquivos log de URLs recuperados para diferentes tipos de repositórios de objetos de aprendizagem.

Figura 39 – Visão geral do *crawler* realizando a extração das URLs



Fonte: elaborado pela autora.

7.2.3 Persistência

A extração dos metadados é realizada por meio de um *parsing*³³ sobre o HTML recuperado. Para auxiliar no processo de extração foi utilizado o HTML *parser* JSOUP (JSOUP, 2016) que é uma biblioteca Java para manipulação e extração de informação

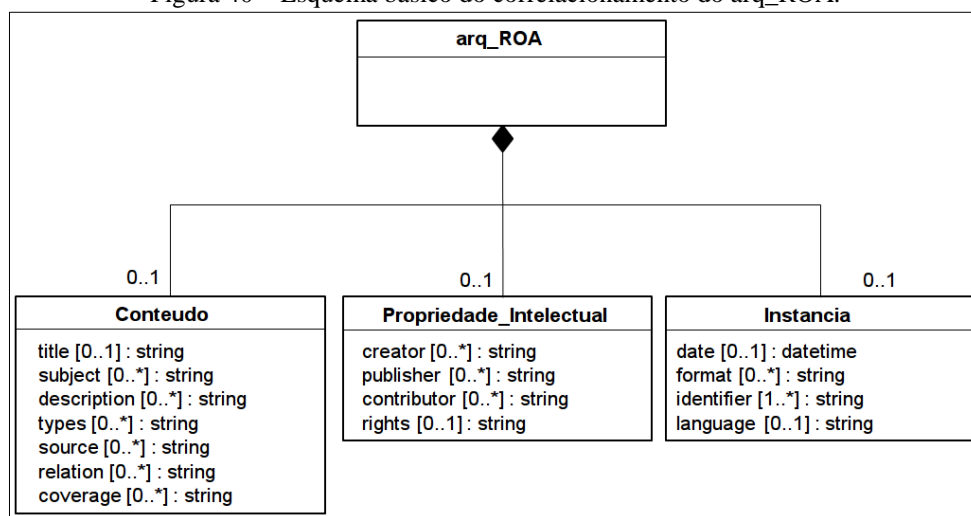
³² ROA Cesta – Disponível em: <<http://cesta2.cinted.ufrgs.br/xmlui>>. Acesso em: 25 set. 2016.

³³ Análise sintática.

no formato HTML. Optou-se em extrair os metadados diretamente do HTML, uma vez que nem todos os repositórios disponibilizam seus metadados em formato XML.

Nesta etapa ocorre a correlação entre os padrões de metadados que são extraídos para o padrão proposto nesta investigação e que foi denominado de *arq_ROA*. A Figura 40 ilustra o esquema base do resultado do correlacionamento entre os padrões de metadados (ver seção 3.2). Ainda é possível observar a cardinalidade de cada elemento e entre as entidades do modelo. O símbolo \blacklozenge representa que a arquitetura *arq_ROA* é composta por três classes: “Conteúdo”, “Propriedade_Intelectual” e “Instancia”.

Figura 40 – Esquema básico do correlacionamento do *arq_ROA*.



Fonte: elaborado pela autora.

Os metadados extraídos são persistidos, inicialmente, em um modelo de metadados (padrão XML) que possuem propriedades e valores associados a cada elemento da correlação de metadados. A Figura 41 apresenta um fragmento da descrição do modelo XML *Schema* para representação da gramática formal do arquivo XML. A XML *Schema* permite definir os elementos, os tipos de dados e as possíveis restrições que possam ocorrer para determinado elemento, como é caso das declarações *minOccurs* (define o número mínimo de ocorrências) e *maxOccurs* (define o número máximo de ocorrências) para o elemento.

Figura 41 – Modelo XML Schema para o arquivo XML.

```

<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
  targetNamespace="http://www.w3schools.com"
  xmlns="http://www.w3schools.com"
  elementFormDefault="qualified">
  <xs:element name="arq_ROA">
    </xs:complexType>
    <xs:sequence>
      <xs:element name="title" type="xs:string" minOccurs="0" />
      <xs:element name="subject" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="description" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="types" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="relation" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="source" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="coverage" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="creator" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="publisher" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="contributor" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="rights" type="xs:string" minOccurs="0"/>
      <xs:element name="date" type="datetime" minOccurs="0"/>
      <xs:element name="format" type="xs:string" minOccurs="0"/>
      <xs:element name="language" type="xs:string" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="identifier" type="xs:string" minOccurs="0"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
</xs:schema>

```

Fonte: elaborado pela autora.

A Figura 42 apresenta um exemplo de extração dos metadados de objetos de aprendizagem para o arquivo XML correspondente. A imagem da esquerda apresenta a captura de tela da descrição dos metadados do objeto de aprendizagem, enquanto que a imagem da direita representa a extração para o XML correspondente.

Figura 42 – Extração de metadados para o XML correspondente.

dc.audience.mediator	Secretaria de Educação Básica (SEB/MEC)	pt_BR	<arq_ROA:title> Tudo se Transforma - Alquimia </arq_ROA:title>
dc.description.abstract	Compreender o conceito de alquimia. Relacionar conceitos de Química e alquimia	pt_BR	<arq_ROA:subject> Alquimia </arq_ROA:subject>
dc.description2	Conceituar alquimia	pt_BR	<arq_ROA:subject> Hermetismo </arq_ROA:subject>
dc.title	Tudo se Transforma - Alquimia	pt_BR	<arq_ROA:subject> Química moderna </arq_ROA:subject>
dc.language	pt	pt_BR	<arq_ROA:subject> Química moderna </arq_ROA:subject>
dc.location.country	br	pt_BR	<arq_ROA:subject> Química moderna </arq_ROA:subject>
dc.date.created	2011-07-04		<arq_ROA:subject> Química moderna </arq_ROA:subject>
dc.audience.educationlevel	Ensino Médio	pt_BR	<arq_ROA:subject> Química moderna </arq_ROA:subject>
dc.description.tableofcontents	Ensino Médio::Química	pt_BR	<arq_ROA:subject> Química moderna </arq_ROA:subject>
dc.type	Video	pt_BR	<arq_ROA:description> Compreender o conceito de alquimia. Relacionar conceitos de Química e alquimia </arq_ROA:description>
dc.subject.category	Educação Básica::Ensino Médio::Química::Modelos de constituição: substâncias, transformações químicas	pt_BR	<arq_ROA:description> Compreender o conceito de alquimia. Relacionar conceitos de Química e alquimia </arq_ROA:description>
dc.rightsholder	MEC	pt_BR	<arq_ROA:types> Video </arq_ROA:types>
dc.rights.license	Termo de cessão dado pelo autor ou seu representante diretamente ao Ministério da Educação - MEC que permite o uso do recurso para distribuição, tradução, edição, excetuando-se o uso comercial	pt_BR	<arq_ROA:types> Video </arq_ROA:types>
dc.date.submitted	2011-07-04T19:05:03Z	*	<arq_ROA:relation> video.html </arq_ROA:relation>
dc.date.accessioned	2011-07-04T19:05:03Z		<arq_ROA:relation> video.html </arq_ROA:relation>
dc.date.available	2011-07-04T19:05:03Z		<arq_ROA:contributor> Projeto Condigital MEC - MCT </arq_ROA:contributor>
dc.date.issued	2011-07-04T19:05:03Z		<arq_ROA:contributor> Projeto Condigital MEC - MCT </arq_ROA:contributor>
dc.identifier.uri	http://objetoseducacionais2.mec.gov.br/handle/mec/18538		<arq_ROA:rights> Termo de cessão dado pelo autor ou seu representante diretamente ao Ministério da Educação - MEC que permite o uso do recurso para distribuição, tradução, edição, excetuando-se o uso comercial </arq_ROA:rights>

Fonte: elaborado pela autora.

Após a extração dos metadados para o arquivo XML, será realizado um mapeamento dos elementos para uma estrutura de dados relacional. Os processos

referentes à extração, correlação e geração do XML rodam em segundo plano (*background*).

7.3 Módulo 2 – Sistema de Recuperação de Informação

O sistema de recuperação de informação tem a função de estabelecer a comunicação com o usuário e buscar os metadados de objetos de aprendizagem que satisfazem a sua necessidade de informação.

O motor de recuperação de informação recebe a solicitação do usuário e verifica a probabilidade de um determinado registro de metadados conter ou não a informação solicitada pelo usuário. Assim, a arquitetura para recuperar objetos de aprendizagem se baseia no modelo de recuperação de informação probabilístico, pois parte-se do entendimento de que no início do processo de recuperação de informação o *corpus* de documentos relevantes é desconhecido, portanto a probabilidade da expressão de busca informada pelo usuário estar presente ou não nos itens a serem recuperados é a mesma.

O modelo probabilístico realiza a busca de informação baseado nos fundamentos da probabilidade para ponderar a similaridade e considerar o grau de incerteza no julgamento entre um *corpus* documental e a expressão de busca do usuário. Este modelo precisa de uma hipótese inicial para estabelecer os documentos relevantes e posteriormente ranqueá-los. No caso desta investigação, também é necessário definir uma hipótese inicial que é obtida pela ocorrência da expressão de busca em alguma parte do registro de metadados. Após essa seleção inicial é preciso determinar as propriedades que definem conjunto de documentos como relevantes.

Para exemplificar essa situação, supõe-se que os registros de metadados podem ser representados por uma matriz de incidência binária cuja presença ou a ausência da expressão de busca em alguns elementos do metadados é designado por 1 ou 0, respectivamente. A Tabela 12 apresenta um exemplo fictício de um conjunto de itens recuperado com 10 registros de metadados de OA e sua frequência de incidência nos três elementos de metadados para realizar o ranqueamento, pois estes são relacionados com o conteúdo do recurso. As frequências 1 e 0 indicam a presença ou ausência do termo no referido elemento de metadados. Assim, o processo de formação da matriz

ocorre por meio a ocorrência da expressão de busca do usuário no elemento de metadados correspondente.

Tabela 12 – Matriz binária de frequência dos termos nos elementos de metadados.

	TITLE	SUBJECT	DESCRIPTION
OA ₁	1	1	1
OA ₂	0	0	1
OA ₃	1	1	0
OA ₄	1	0	1
OA ₅	1	1	0
OA ₆	0	1	1
OA ₇	1	1	1
OA ₈	0	1	0
OA ₉	1	1	0
OA ₁₀	1	1	1

Fonte: elaborado pela autora.

Para realizar o ranqueamento dos itens recuperados, propõe-se a equação a seguir para o cálculo de similaridade entre os itens recuperados e a expressão de busca do usuário:

$$sim(expr_busca) = \sum_{i=1} (w * f_i)$$

onde:

1. w é o peso atribuído ao elemento de metadados, conforme Tabela 10;
2. f_i é a frequência de ocorrência da expressão de busca.

O grau de similaridade entre a expressão de busca e os registros de metadados é obtido pelo somatório do produto entre o peso atribuído ao elemento de metadados e a frequência de ocorrência da expressão de busca. A Tabela 13 apresenta a simulação de ranqueamento considerando um conjunto de 10 itens recuperados de um *corpus*.

Tabela 13 – Matriz binária de frequência dos termos nos elementos de metadados e seu grau de similaridade

	TITLE (PESO 1)	SUBJECT (PESO 0.8)	DESCRIPTION (PESO 0.6)	SIMILARIDADE
OA₁	1	1	1	2,4
OA₂	0	0	1	0,6
OA₃	1	0	0	1,0
OA₄	1	0	1	1,6
OA₅	1	1	0	1,8
OA₆	0	1	1	1,4
OA₇	1	1	1	2,4
OA₈	0	1	0	0,8
OA₉	1	1	0	1,8
OA₁₀	1	1	1	2,4

Fonte: elaborado pela autora.

Os resultados são apresentados ao usuário em ordem decrescente dos pesos calculados que representam a média aritmética ponderada calculada de acordo com a incidência de frequência dos elementos na matriz binária, conforme pode ser visualizado na Tabela 14.

Tabela 14 – Matriz binária de incidência ranqueada.

	TITLE (PESO 1)	SUBJECT (PESO 0.8)	DESCRIPTION (PESO 0.6)	SIMILARIDADE
OA₁	1	1	1	2,4
OA₇	1	1	1	2,4
OA₁₀	1	1	1	2,4
OA₅	1	1	0	1,8
OA₉	1	1	0	1,8
OA₄	1	0	1	1,6
OA₆	0	1	1	1,4
OA₃	1	0	0	1,0

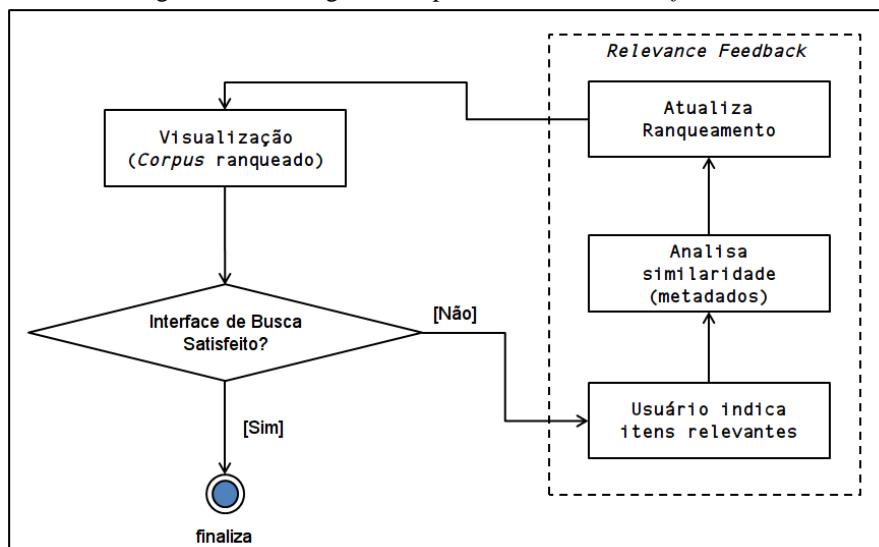
OA₈	0	1	0	0,8
OA₂	0	0	1	0,6

Fonte: elaborado pela autora.

Normalmente, o usuário desconhece o *corpus* a ser recuperado e elaborar uma expressão de busca significativa para sua necessidade de informação pode ser complexo. Por isso, recomenda-se que ocorra um refinamento no resultado apresentado por meio da seleção dos itens que o usuário julgar como relevantes. Dessa forma, os itens podem ser ranqueados novamente para que os resultados sejam mais adequados para a necessidade de informação do usuário. Essa é outra justificativa para a indicação de uma arquitetura baseada no modelo probabilístico nesta arquitetura, pois esta suporta a técnica de *relevance feedback*.

A técnica de *relevance feedback* é importante para esta arquitetura, pois permite que ocorra um processo de diálogo entre o usuário e o SRI. Quando se trata da recuperação de objetos de aprendizagem descritos em um padrão de metadados, parte-se do princípio de que a descrição desses recursos pode facilitar a inferência do usuário na análise de relevância dos itens que lhe foram retornados para aprimorar os resultados recuperados sem que seja necessário que o usuário refaça sua expressão de busca. Com base no conjunto inicial de itens retornados ao usuário é possível que os resultados sejam melhorados, gradativamente, por meio de interações com o usuário.

A Figura 43 apresenta o fluxograma de execução do processo de *relevance feedback*. Assim sendo, após o ranqueamento inicial apresentado ao usuário é possível que ele indique quais os itens que, em sua opinião, são relevantes, para que o motor de recuperação de informação possa realizar um novo ranqueamento com base na inferência do usuário. A interface de busca deve permitir a seleção dos itens que no momento parecem ser mais relevantes para necessidade de informação do usuário para que uma nova requisição seja enviada ao SRI.

Figura 43 – Fluxograma do processo de *relevance feedback*

Fonte: elaborado pela autora.

O *relevance feedback* para o modelo probabilístico clássico considera o conjunto de documentos positivos (relevantes) para recalculer a probabilidade de encontrar outros documentos relevantes. Dessa forma, o esquema geral para executar RF é o seguinte:

1. Apresente um conjunto de documentos ranqueados por meio do cálculo de similaridade proposto nesta seção: $sim(expr_busca) = \sum_{i=1} (w * f_i)$;
2. Obtenha o *feedback* do usuário;
3. Identifique similaridades nos elementos de metadados (por exemplo: o tipo de formato: animação, vídeo, áudio ou o alcance (*coverage*) do objeto de aprendizagem: educação infantil, ensino superior etc.);
4. Gere uma nova matriz de incidência binária incluindo os campos similares com suas frequências;
5. Apresente um novo ranqueamento com as informações *relevance feedback* obtidas;
6. Repita os passos 2, 3, 4 e 5 até que o usuário esteja satisfeito.

Este processo é cíclico e pode ser realizado até que a necessidade de informação do usuário tenha sido satisfeita. Evidencia-se que nesta arquitetura deve ser utilizado o *relevance feedback* explícito, conforme explicado na subseção 6.1, pois considera-se nesta arquitetura apenas o julgamento do usuário e não as inferências internas do sistema de recuperação de informação.

Ressalta-se que o processo de *relevance feedback* não pode ser aplicado na formulação da expressão de busca inicial do usuário, uma vez que não existe nenhum resultado disponível para indicar a relevância. O sucesso deste processo está diretamente relacionado com a espontaneidade do usuário em informar seu posicionamento em relação ao conjunto de itens que lhe foi apresentado.

De acordo com Jin e Hauptmann (2001), essa interação do usuário com o SRI possibilita uma recuperação de informação mais eficiente. Os trabalhos de Lavrenko e Croft (2001), Robertson e Jones (1976), Rocchio (1971), Salton e Buckley (1990) e Zhai e Lafferty (2001) comprovaram que o processo de RF é eficiente e incrementa precisão do processo de recuperação de informação. Assim, esta técnica visa melhorar a efetividade na recuperação de informação baseado na avaliação de relevância dos documentos pelo usuário.

Portanto, o principal objetivo desta arquitetura foi de disponibilizar um modelo capaz de permitir que um usuário obtenha um *corpus* de objetos de aprendizagem (por meio de metadados) mais significativo para sua necessidade de informação e que a partir da interação do usuário com o sistema de recuperação de informação seja possível melhorar o retorno das buscas para que o usuário possa acessar o recurso a partir do repositório no qual ele está sendo mantido.

8. Considerações Finais

Percebeu-se, no decorrer desta investigação, que os desafios inerentes à recuperação de informação relevante na *Web* são constantes e incrementados pelas características intrínsecas deste ambiente, tais como a dinamicidade e a volatilidade das informações. No âmbito educacional estas características também são observadas e acrescidas da complexidade e das peculiaridades dos objetos de aprendizagem que são idealizados para atender a diferentes objetivos pedagógicos e desenvolvidos em diversos tipos de formatos.

Neste capítulo são apresentadas as considerações finais desta investigação, bem como são apontadas as principais contribuições e limitações que foram observadas ao longo do desenvolvimento desta tese. Por fim, conclui-se este capítulo indicando os trabalhos derivados desta tese que podem ser realizados futuramente.

Foram analisadas as principais arquiteturas utilizadas no desenvolvimento de repositórios de objetos de aprendizagem, bem como foram identificados os padrões de metadados adotados por cada iniciativa investigada. Destaca-se que os padrões de metadados são fundamentais para a descrição de objetos de aprendizagem, pois devido à diversidade de formatos desses recursos educacionais se torna difícil recuperá-los em escala global. Portanto, pode-se afirmar que os metadados exercem um papel primordial na descrição e busca desses recursos educacionais, com a intenção de torná-los mais acessíveis. Para tanto foi apresentado um mapeamento entre os padrões de metadados

educacionais: *Dublin Core* e LOM que sugere uma melhora na interoperabilidade entre objetos de aprendizagem mantidos por diferentes repositórios. A proposta de mapeamento resultante da análise dos elementos dos esquemas de metadados permitiu estabelecer uma relação direta com todos os elementos do padrão *Dublin Core*.

Os agentes inteligentes desempenharam um papel fundamental na descoberta de informação disponível na *Web* e potencializaram a busca por informações estruturadas (metadados) em diferentes repositórios de objetos de aprendizagem. O modelo proposto de *crawler* baseado em agentes inteligentes foi capaz de descobrir, recuperar e mapear objetos de aprendizagem armazenados em repositórios que descrevem seus recursos por diferentes padrões de metadados.

Foram investigados os conceitos e modelos de recuperação de informação para identificar qual melhor se aplica ao domínio de problema aqui investigados. Identificou-se que dos três modelos clássicos de recuperação de informação (booleano, espaço vetorial e probabilístico), o que mais se mostrou adequado para atender esta proposta foi baseado no modelo probabilístico que realiza a busca de informação baseado nos fundamentos da probabilidade para ponderar a similaridade e considerar o grau de incerteza no julgamento entre um *corpus* documental e a expressão de busca do usuário.

Este modelo considera como hipótese inicial a presença ou ausência dos termos nos metadados de objetos de aprendizagem. Para realizar o ranqueamento foram atribuídos pesos aos elementos de metadados que servem de parâmetros para calcular o grau de relevância para cada item recuperado. Os resultados ranqueados são apresentados em ordem decrescente ao usuário.

O modelo booleano foi desconsiderado, pois não apresenta um sistema de ranqueamento eficaz, o que faz com que muitos ou poucos objetos de aprendizagem sejam recuperados. O modelo espaço vetorial não se aplica pelo fato de que os metadados não apresentam informações textuais suficientes para valorar os termos de indexação. Portanto, a arquitetura para recuperar objetos de aprendizagem foi baseada no modelo probabilístico também por suportar a técnica de *relevance feedback*.

A arquitetura para recuperação de metadados de objetos de aprendizagem permitiu unificar recursos educacionais descritos em diferentes padrões de metadados e mantidos por diferentes repositórios por meio da proposição de um modelo de

correlação entre os padrões de metadados educacionais o que possibilita uma maior interoperabilidade, reutilização e compartilhamento desses recursos educacionais de forma tangível ao usuário.

O principal diferencial desta proposta foi de valorizar a inferência do usuário no processo de recuperação de informação, por meio do processo de *relevance feedback*. Neste processo, o usuário estabelece um diálogo com o sistema de recuperação de informação realizando refinamentos nos resultados que lhe foram retornados. Como esse processo é cíclico, ele pode ser executado até que o usuário esteja satisfeito com os resultados que lhe foram retornados.

Ressalta-se que o processo de *relevance feedback* depende exclusivamente da participação do usuário, caso ele não queira emitir seu parecer terá como resultado a apenas a probabilidade gerada pelo SRI. Observou-se em trabalhos correlatos de Warpechowski, Souto e Oliveira (2006), Guzmán, Sánchez e Torres (2006), Gonçalves (2007), Fuentes et al. (2010) e Vian e Silveira (2011) que nenhuma das propostas aborda a questão da inferência do usuário no julgamento da informação que lhe foi retornada. Caso o usuário não fique satisfeito com o resultado ele deve refazer sua expressão de busca, ou seja, não se mantém um processo dialógico do usuário com o sistema de recuperação de informação para um refinamento dos resultados apresentados ao usuário.

As contribuições desta pesquisa foram observadas tanto para a área de Ciência da Informação quanto para a sociedade da informação que tem a necessidade de recuperar a informação que está disponível na *Web* de forma rápida e precisa. Entre as principais contribuições do presente trabalho, podem-se destacar:

- Compreensão, em termos arquiteturais, da forma como os recursos educacionais são mantidos nos repositórios de objetos de aprendizagem e identificação dos padrões de metadados adotados por cada iniciativa;
- Proposição de um modelo de correlação entre os principais padrões de metadados educacionais;
- Apresentação de um *crawler* baseado em agentes inteligentes capaz de descobrir, recuperar e mapear objetos de aprendizagem armazenados em repositórios que adotam diferentes padrões de metadados;

- Indicação de um modelo de arquitetura para recuperação de objetos de aprendizagem com suporte ao *relevance feedback*.

As limitações desta tese estão diretamente relacionadas com a ausência de precisão na descrição dos objetos de aprendizagem e ocorrem principalmente pela falta de padronização e de ferramentas que facilitem a recuperação de objetos de aprendizagem.

Observou-se que os repositórios de objetos de aprendizagem que adotam padrões de metadados para a descrição dos recursos educacionais nem sempre descrevem esses recursos de forma exaustiva, ou seja, apesar dos padrões conterem diversos elementos de descrição, muito deles são ignorados durante este processo.

Outro fator limitador desta investigação é que a arquitetura proposta nesta tese não foi implementada na íntegra, apenas o módulo referente ao agente buscador foi desenvolvido. Esta limitação possibilita visualizar perspectivas futuras de continuidade desta investigação, com o desenvolvimento de um protótipo para validar a arquitetura proposta. A partir do protótipo é possível propor um sistema de recomendação de objetos de aprendizagem, baseado no conceito de relevância do usuário. Outra perspectiva inerente a esta investigação é expandir a quantidade de padrões de metadados suportado no modelo de correlação para ampliar o escopo de busca de objetos de aprendizagem.

Portanto, comprova-se a tese de que um sistema de recuperação de informação é mais eficiente quando amplia seu escopo de recuperação a partir de diferentes fontes de dados (ROA) e permite a inferência do usuário no julgamento da informação que lhe foi retornada. Assim, acredita-se que um sistema de recuperação de informação retorna resultados mais relevantes quando possibilita estabelecer um processo dialógico entre o usuário e o sistema de recuperação de informação.

Referências

ADL. Emerging and Enabling Technologies for the design of Learning Object Repositories Report. **Advanced Distributed Learning Initiative**, 2002. Disponível em: <<http://xml.coverpages.org/ADLRepositoryTIR.pdf>>. Acesso em: 27 nov. 2014.

AFONSO, M. C. L.; EIRÃO, T. G.; MELO, J. H. M.; ASSUNÇÃO, J. S.; LEITE, S. V. Banco Internacional de Objetos Educacionais (BIOE): tratamento da informação em um repositório educacional digital. **Perspect. Ciênc. Inf.**, Belo Horizonte, v. 16, n. 3, p. 148-158, set. 2011.

ALLEN, C.A.; MUGISA, E.K. Improving learning object reuse through OOD: a theory of learning objects. **J Object Technol** 9:1–22, 2010.

ALVES, R. C. V. **Web Semântica**: uma análise focada no uso de metadados. 2005, 180f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia Ciências, Universidade Estadual Paulista, Marília, 2010.

ARIADNE. The Alliance of Remote Instructional Authoring and Distribution Networks for Europe. Disponível em: <<http://www.ariadne-eu.org>>. Acesso em: 24 ago. 2015.

ARMS, W. Y.; HILLMANN, D.; LAGOZE, C.; KRAFFT, D.; SAYLOR, J.; TERRIZZI, C.; van de SOMPEL, H. A spectrum of interoperability: the site for science prototype for the NSDL. **D-Lib magazine**, v.8, 1, janeiro. 2002. Disponível em: <<http://www.dlib.org/dlib/january02/arms/01arms.html>>. Acesso em: 23 ago. 2015.

AUDINO, D.F.; NASCIMENTO, R.S. Objetos de Aprendizagem – Diálogo entre Conceitos e uma Proposição Aplicada à Educação. **Revista Contemporânea de Educação**, vol.5, n.10, 2010, p. 128-148.

AYALA, G.; CASTILLO, S. Towards computational models for mobile learning objects. In: **5th IEEE international conference on wireless, mobile, and ubiquitous technology in education**, Beijing, p. 23–26, 2008.

BAEZA-YATES, R.; CASTILLO, C. Balancing volume, quality and freshness in web crawling. In **Soft Computing Systems – Design, Management and Applications**, Santiago, Chile. IOS Press Amsterdam, p. 565-572, 2002.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval: The concepts and technology behind search**. Addison-Wesley Professional. 2ª Edition, 2011. 944p.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação: conceitos e tecnologia das máquinas de busca**. 2. Ed. – Porto Alegre: Bookman, 2013.

BIANCO, A. M.; DE MARSICO, M.; TEMPERINI, M. **Standards for e-learning**. Project QUIS: (Quality, Interoperability and Standards in e-learning). The TISIP Research Foundation, Trondheim, 2005. ISBN 82-8055-018-6.

BLAIR, D. C. **Language and Representation in Information Retrieval**. Elsevier Science Publishers, 1990.

BOYLE, T. Design principles for authoring dynamic, reusable learning objects. **Aust J Educ Technol**. 19, p. 46–58, 2003.

BRADSHAW, J. M.; DUTFIELD, S.; BENOIT, P.; WOOLLEY, J. KAoS: Towards and Industrial Strength Open Agent Architecture. In: Bradshaw, J. (Ed.), **Software Agents**. Cambridge, MA: MIT Press/AAAI Press. 1997. Disponível em: <<http://agents.umbc.edu/introduction/01-Bradshaw.pdf>>. Acesso em: 02 nov. 2016.

BUCKLEY, C.; SALTON, G. Optimization of Relevance Feedback Weights. In: **Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval**. Washington: USA. 1995. p. 9-13.

BUSH, V. As we may think. **Atlantic Monthly**, v. 176, n. 1, p. 101-108, 1945. Disponível em: <<http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>>. Acesso em: 02 jun. 2014.

CAPLAN, P. **Metadata fundamentals for all librarians**. Chicago: American Library Association, 2003.

- CASTILLO, C. Effective web crawling. **SIGIR Forum**, v. 39, p. 55–56, 2005. Disponível em: <<http://doi.acm.org/10.1145/1067268.1067287>>. Acesso em: 22 set. 2016.
- CHANG, Y. K.; CIRILLO, C.; RANZON, J. Evaluation of Feedback Retrieval Using Modified Freezing, Residual Collection, and Test and Control Groups. In: G. Salton (Ed.), **The SMART Retrieval System**. Experiments in Automatic Document Processing. Englewood Cliffs, New Jersey: Prentice Hall, 1971. p. 335-370
- CHO, J.; GARCIA-MOLINA, H. Effective page refresh policies for web crawlers. **ACM Transactions on Database Systems**, 28(4), 2003.
- CHOWDHURY, G. G. **Introduction to Modern Information Retrieval**. 3rd ed, Facet Publishin, 2010.
- COVE, J. F., WALSH, B. C. Online Text Retrieval via Browsing. **Information Processing & Management**. v. 24. n. 1. p. 31-37, 1988.
- CROFT, B.; CALLAN, J.; LAFFERTY, J. **Language Modeling and Information Retrieval**. Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, p. 98-112, 2001.
- CURRIER, S. **Metadata for Learning Resources: An Update on Standards Activity for 2008**. Ariadne, 2008. Disponível em: <<http://www.ariadne.ac.uk/issue55/currier/>>. Acesso em: 05 set. 2015.
- DAY, M. **Metadata in a Nutshell**. Information Europe, 6 (2), 11, 2001.
- DCMI**. Dublin Core Metadata Initiative. Disponível em: <<http://dublincore.org/>>. Acesso em: 11 dez. 2013.
- DOULAMIS, N.; DOULAMIS, A. Evaluation of relevance feedback schemes in content-based in retrieval systems. **Signal Processing: Image Communication**, v. 21, n. 4, p. 334 – 357, 2006. ISSN 0923-5965. Disponível em: <<http://www.sciencedirect.com/-/science/article/pii/S0923596505001360>>. Acesso em: 18 out. 2016.
- DOWNES, S. Design and reusability of learning objects in an academic context: a new economy of education? Moncton: **National Research Council**, 2002. Disponível em: <www.downes.ca/files/milan.doc>. Acesso em: 30 jul. 2015.
- DUFRESNE, A.; SENTENI, A.; RICHARDS, G. La contextualisation des banques de ressources: barrières et clés. **Canadian Journal of Learning Technology**, v.28, n.3, p. 27-42, 2002.

DUTRA, R.L.S.D.; TAROUÇO, L.M.R. Objetos de Aprendizagem: Uma comparação entre SCORM e IMS Learning Design. **RENOTE. Revista Novas Tecnologias na Educação**, Porto Alegre - RS - Brasil, v. 4, n.1, p. 1-10, 2006.

EDUCAUSE. 2015. Disponível em: <<http://www.educause.edu/>>. Acesso em: 04 set. 2015.

FERNEDA, E. **Recuperação de Informação**: análise sobre a contribuição da Ciência da Informação para a Ciência da Computação. 2003, 137f. Tese (Doutorado em Ciência da Comunicação) – Escola de Comunicações e Artes da Universidade de São Paulo, São Paulo, 2003.

_____. **Introdução aos Modelos Computacionais de Recuperação de Informação**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2012.

_____. **Ontologia como recurso de padronização terminológica em um Sistema de Recuperação de Informação**. 2013. 109 f. Relatório (Estágio Pós-Doutoral), Ciência da Informação, Universidade Federal da Paraíba, João Pessoa. 2013.

FISCHER, K.; RUSS, C.; VIERKE, G. **Decision Theory and Coordination in Multiagent Systems**. Research Report RR-98-02, Deutsches Forschungszentrum fuerKuenstliche Intelligenz GmbH, October 1998.

FIPA. Foundation for Intelligent Physical Agents. Disponível em: <<http://www.fipa.org/>>. Acesso em: 15 dez. 2015.

FRAKES, W. B.; BAEZA-YATES, R. **Information Retrieval: Data Structures & Algorithms**. New Jersey: Prentice-Hall, 1992.

FUENTES, M. L. M.; ARTEAGA, J.M.; RODRÍGUEZ, F. A.; VANDERDONKT, J.; OREY, M. **MIRROS: Intermediary Model to Recovery Learning Objects**. In: **Computación y Sistemas**. v. 13 N. 4, 2010, p. 373-384.

FUHR, N. Probabilistic models in information retrieval. **The Computer Journal**, v.35, nº.3, p.243-255, 1992. Disponível em: <<http://twiki.di.uniroma1.it/pub/Estrinfo/Materiale/probabilisticIR.pdf>>. Acesso em: 15 set. 2016.

GARFIELD, E. A Retrospective and Prospective View of Information Retrieval and Artificial Intelligence in the 21st Century. **Journal of The American Society for Information Science and Technology**. 52(1), 2001.

GIBBONS, A. S.; NELSON, J.; RICHARDS, R. The nature and origin of instructional objects. In D. A. Wiley (Ed.), **The instructional use of learning objects**. Bloomington, IN: **Association for Educational Communications and Technology**, 2000. Disponível em: <<http://www.reusability.org/read/chapters/gibbons.doc>>. Acesso em: 28 nov. 2014.

- GILLILAND, A. J. **Setting the stage**: In Introduction to metadata 3.0. Edited by Murtha Baca. J. Paul Getty Trust. 2008.
- GILLILAND-SWETLAND, A. J. **La definición de los metadatos**. In: BACA, M. (Ed.). Introducción a los metadatos vías a la información digital. Traducido al español por Marisol Jacas-Santoll. Los Angeles, CA: J. Paul Getty Trust, 1998. p. 1-9.
- GLUZ, J. C.; VICCARI, R. M. Uma Ontologia OWL para Metadados IEEE-LOM, Dublin-Core e OBAA. In: **XXII Simpósio Brasileiro de Informática na Educação (SBIE 2011)**, 2011, Aracaju, Brasil. Anais do Simpósio Brasileiro de Informática na Educação, 2011. v. 1. p. 204-213.
- GOLDREI, S., KAY, J.; KUMMERFELD, B. Exploiting user models to automate the harvesting of metadata for Learning Objects. **Proceedings of 2005 Conf. on Adaptive and intelligent Web-based educational systems (AIWBES'05)**, Málaga, España, 2005. Disponível em: <<http://www.lcc.uma.es/~eva/waswbe05/papers/goldrei.pdf>>. Acesso em: 23 ago. 2015.
- GONÇALVES, V. M. B. **A Web Semântica no Contexto Educativo**: Um sistema para a recuperação de objectos de aprendizagem baseado nas tecnologias para a Web Semântica, para o e-Learning e para os agentes. Tese (Engenharia Electrotécnica e de Computadores) – Universidade do Porto – Portugal, 2007.
- GUNAWARDENA, A.; ADAMCHIK, V. A customized learning objects approach to teaching programming. **ACM SIGCSE Bull** 35(3):264, 2003. Disponível em: <<http://www.cs.cmu.edu/~adamchik/PLO/ITiCSE-03.pdf>>. Acesso em: 20 set. 2014.
- GUZMÁN, J.; SÁNCHEZ, A.; TORRES, D. Multi-Agent System for Digital Documents Recuperation. **Revista Colombiana de Tecnologías de Avanzada**. Volumen 1. 2006. ISSN 1692-7257.
- HARTER, S. P. **Online Information Retrieval**: Concepts, principles and techniques. Orlando: Academic Press. Inc., 1986. 259 p.
- HAYES-ROTH, B. An Architecture for Adaptive Intelligent System. **Artificial Intelligence**. Special Issue on Agents and Interactivity, 72, p. 329-365; 1995.
- HAYNES, D. **Metadata for Information Management and Retrieval**. London: Facet Publishing, 2004. 186p.
- HERSH, W. R. **Information Retrieval**: A Health and Biomedical Perspective. 3. ed. New York, NY: Springer-Verlag, 2009.
- HODGINS, W. H. **Into the future**: a vision paper. Hameenlinna, Finland: Comission on Technology and Adult Learning, fev. 2000. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.8864&rep=rep1&type=pdf>>. Acesso em: 14 ago. 2015.

HUHNS, M. N.; SINGH, M. P. **Agents and multiagents systems**: Themes, approaches and challenges. Readings in Agents, 1997. p. 1-13.

IDE, E. New Experiments In Relevance Feedback. In G. Salton (Ed.) **The SMART Retrieval System**. Experiments in Automatic Document Processing. Englewood Cliffs, New Jersey: Prentice Hall, p. 337-354, 1971.

IEEE LTSC LOM. IEEE Learning Technology Standards Committee (LTSC) – Standard for Learning Object Metadata (LOM). 2002. Disponível em: <http://grouper.ieee.org/groups/ltsc/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf>. Acesso em: 25 ago. 2015.

IFLA – International Federation of Library Associations and Institutions. 1998. Disponível em: <<http://www.ifla.org/archive/udt/op/udtop8/udtop8.htm>>. Acesso em: 28 set. 2015.

IMS GLOBAL. IMS Global Learning Consortium. 2015. Disponível em: <<http://www.imsglobal.org/>>. Acesso em: 06 set. 2015.

IP, A.; MORRISON, I.; CURRIE, M. What is a learning object, technically? In: **WebNet**, pp 580–586. 2001. Disponível em: <[http://users.tpg.com.au/adslfrcf/lo/learningObject\(WebNet2001\).pdf](http://users.tpg.com.au/adslfrcf/lo/learningObject(WebNet2001).pdf)>. Acesso em: 28 jul. 2015.

ISO. ISO International Organization for Standardization, 2015. Disponível em: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=45392&development=on>. Acesso em: 05 set. 2015.

ISO MLR. ISO/IEC 19788. 2014. Disponível em: <<https://www.iso.org/obp/ui/#iso:std:iso-iec:19788:-4:ed-1:v1:en>>. Acesso em: 05 set. 2015.

JENNINGS, N. R. On Agent-based Software Engineering. **Artificial Intelligence**, v. 117, n. 277-296, 2000. Disponível em: <<http://eprints.soton.ac.uk/253741/1/aij2000.pdf>>. Acesso em: 23 out. 2015.

JENNINGS, N. R. **Coordination Techniques for Distributed Artificial Intelligence**. In: O'HARE, G.M.P.; JENNINGS, N.R. (Eds.). Foundations of distributed artificial intelligence. New York: John Wiley & Sons, 1996. p.187-21 Disponível em: <<https://pdfs.semanticscholar.org/e3e6/175b5c003b9dcd17165452769159df163143.pdf>>. Acesso em: 23 out. 2015.

JIN, R.; HAUPTMANN, A.G. **A New Probabilistic Model for Title Generation**. Carnegie Mellon University, Pittsburgh, USA, 2001.

JSOUP. Java HTML Parser. Disponível em: <<http://jsoup.org/>>. Acesso em: 30 set. 2016

KHAPRE, S.C.; BASHA, M.S.S. A Theoretical Paradigm of Information Retrieval in Information Science and Computer Science. **IJCSI - International Journal of Computer Science Issues**, 9, p.232 – 240, 2012.

KÖCHER, J. C. **Fundamentos da metodologia científica: teoria da ciência e prática da pesquisa**. 20 ed. Petrópolis: Vozes, 2002.

KOOHANG, A.; HARMAN, K. **Learning objects and instructional design**. Institute Informing Science, Santa Rosa, 2007.

KOPER, R. **Modeling Units of Study from a Pedagogical Perspective: the pedagogical meta-model behind EML**. Preprint, 2001. Disponível em: <<http://dspace.ou.nl/bitstream/1820/36/1/Pedagogical%20metamodel%20behind%20EMLv2.pdf>>. Acesso em: 16 jul. 2015.

KUROPKA, D. **Modelle zur Repräsentation natürlichsprachlicher Dokumente: Ontologiebasiertes Information-Filtering und Retrieval mit relationalen Datenbanken**. Seiten: Erscheinungsjahr, 2004.

LAVRENKO, V.; CROFT, W. B. Relevance based language models. In: **Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '01**, New York, NY, USA: ACM, 2001. Disponível em: <<http://doi.acm.org/10.1145/383952.383972>>. Acesso em: 02 out. 2016.

LESK, M. **The seven ages of information retrieval**. Ottawa: IFLA, 1996. Disponível em: <<http://www.ftsm.ukm.my/ss/TP6082/seven%20ages%20ir.pdf>>. Acesso em: 01 jul. 2015.

LIM, C.P.; LEE, S.L.; RICHARDS, C. Developing interactive learning objects for a computing mathematics models. **Int J E-Learn** 5(2):221–244, 2006.

LONGMIRE, W. **A Primer On Learning Objects**. American Society for Training & Development. Virginia. USA. 2001.

LÓPEZ GUZMÁN, C. **Los Repositorios de Objetos de Aprendizaje como soporte a un entorno e-learning**. 2005. Tese (Doutorado) – Universidad de Salamanca, Salamanca, 2005. Disponível em: <http://gedos.usal.es/jspui/bitstream/10366/56649/1/DIA_Repositoriosobjetos.pdf>. Acesso em: 15 jun. 2015.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**. Cambridge University Press. 2008.

MARTÍNEZ MÉNDEZ, F. J. **Propuesta y desarrollo de un modelo para la evaluación de la recuperación en Internet**. 2002. 283 f. Tese (Doutorado) – Facultad de Ciencias de la Documentación, Universidad de Murcia, Murcia, 2002.

MASIE. The Masie Center – “Learning and Technology e-lab & think-Tank. Making Sense of Learning Specification & Standards: a Decision Maker’s Guide to their Adoption”, 2003. Disponível em: <<http://www.masieweb.com>>. Acesso em: 14 jul. 2014.

MASSART, D.; DUNG, L.T. Federated Search of Learning Object Repositories: The CeLeBraTe Approach. **Intl. Conf. RIVF'04**. Fevereiro Hanoi – Vietnã, 2004. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=D705E87FA1FB71EE01C64F97149F4E9D?doi=10.1.1.58.5126&rep=rep1&type=pdf>>. Acesso em: 23 ago. 2015.

MCGREAL, R. **Online Education Using Learning Objects**. Open and Distance Learning Series. London, Routledge/Falmer, 2004.

MEADOW, C. T. **Text Information retrieval Systems**. San Diego: Academic Press, 1993.

MEADOW, C.T.; BOYCE, B.R.; KRAFT, D.H.; BARRY, C. **Text Information Retrieval Systems**. 3rded. London UK: Elsevier, 2007.

MÉNDEZ RODRÍGUEZ, E. **Metadados y recuperación de información**: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón: Trea, 2002. 429 p.

MERRILL, M. D. Knowledge objects. **CBT Solutions** (March/April), 1-11, 1998. Disponível em: <<http://mdavidmerrill.com/Papers/KnowledgeObjects.PDF>>. Acesso em: 10 fev. 2015.

MILSTEAD, J.; FELDMAN, S. Metadata: cataloging by any other name. **On Line Magazine**, New Haven, v.23, n.1, 1999.

MOHAN, P. Learning Objects Repositories. In: KOOHANG, A.; HARMAN, K. **Learning Objects**: theory, praxis, issues, and trends. Santa Rosa: Informing Science Press, p.527-548, 2007.

MONTILVA, J.; ROJAS, M.; DUARTE, A. O. RDOA-WS: repositorio distribuido de objetos de aprendizaje soportado con servicios web. In: **Avances en Sistemas e Informática**; Vol. 8, núm. 2, p. 183-190, 2011. Disponível em: <<http://www.bdigital.unal.edu.co/28851/>>. Acesso em: 23 jun. 2014.

MOOERS, C. **Zatocoding applied to mechanical organization of knowledge**. American Documentation, v. 2, n. 1, p. 20-32, 1951.

MUZIO, J. A.; HEINS, T.; MUNDELL, R. Experiences with reusable E-learning objects Form theory to practice. In: **Internet and Higher Education** 5, p. 21-34, 2002. Disponível em: <<http://www.qou.edu/arabic/researchProgram/eLearningResearchs/experiences.pdf>>. Acesso em: 16 jul. 2015.

- NASH, S. Learning objects, learning object repositories, and learning theory: preliminary best practices for online courses. **Int J Knowl Learn Objects** 1:217–228, 2005.
- NECHES, R.; FIKES, R.; FININ, T.; GRUBER, T.; PATIL, R.; SENATOR, T.; SWARTOUT, W.R. Enabling Technology For Knowledge Sharing. **AI Magazine**, Volume 12, No. 3, Fall, 1991.
- NISO. **Understanding Metadata**. Bethesda, MD: NISO Press, 2004. Disponível em: <<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>>. Acesso em: 14 ago. 2014.
- NWANA, H. **Software Agents: An Overview**. Knowledge Engineering Review. Cambridge University Press v.3, p.1-40, 1996.
- OLIVÁN, J. A. S. **Recuperación de la Información**. 1ª ed. Buenos Aires – Argentina: Alfagrama, 2008.
- ORENGO, V. M. **Assessing relevance using automatically translated documents for cross-language information retrieval**. Tese de doutorado. Universidade Middlesex, Londres, 2004.
- OTON TORTOSA, S. **Propuesta de una Arquitectura Software Basada en Servicios para la Implementación de Repositorios de Objetos de Aprendizaje Distribuidos**. 2006. Tese (Doutorado em Ciência da Computação). Escuela Técnica Superior de Ingeniería Informática, Universidad de Alcalá. Alcalá de Henares, Espanha. 2006.
- PAQUETTE, G.; ROSCA, I. Organic aggregation of knowledge objects in educational systems. In: **Canadian Journal of Learning and Technology**. Volume 28(3) Fall / automne, 2002. Disponível em: <<http://www.cjlt.ca/index.php/cjlt/article/view/111/104>>. Acesso em: 06 ago. 2015.
- PEÑA, R.; BAEZA-YATES, R.; RODRÍGUEZ MUÑOZ, J. V. **Gestión digital de la información: de bits a bibliotecas digitales y a la web**. Madrid, 2002.
- POLSANI, P.R. Use and abuse of reusable learning objects. **J Digit Inf** 3. 2003. Disponível em: <<http://journals.tdl.org/jodi/article/view/89/88>>. Acesso em: 2 jul. 2015.
- PROTÉGÉ. **The Protégé Ontology Editor and Knowledge Acquisition System**. 2006. Disponível em: <<http://protege.stanford.edu>>. Acesso em: 26 de julho de 2015.
- PROTHMAN, B. Metadata: Managing needles in the proverbial haystacks. **IEEE potentials**, Volume 19, Issue 1, p. 20-23, 2000.
- QUINN, C. **Learning objects and instruction components**. 2000. Disponível em: <http://ifets.ieee.org/discussions/discuss_feb2000.html>. Acesso em: 23 mai. 2015.

RIJSBERGEN, C.J. **Information retrieval**. London: Butterworths, 1979. Disponível em: <<http://www.dcs.gla.ac.uk/Keith/Preface.html>>. Acesso em: 15 ago. 2014.

ROBERTSON, S.E. The probability ranking principle in IR. **Journal of Documentation**, p.294-304, 1977.

ROBERTSON, S.E.; JONES, K.S. Relevance weighting of search terms. **Journal of the American Society for Information Science**, v. 27, n. 3, p.129-146, 1976.

ROCCHIO, J. J. Relevance feedback in information retrieval. In G. Salton (Ed.), **The Smart Retrieval System – Experiments in Automatic Document Processing**. Englewood Cliffs, NJ, USA: Prentice Hall, Inc. p. 313 – 323, 1971.

ROUYET, J. I.; MARTÍN, V. **A comparative study of the metadata in SCORM and Dublin Core**, 2004. Disponível em: <http://www.cc.uah.es/spdece/papers/Rouyet_Final.pdf>. Acesso em: 04 set. 2015.

RÜGER, S. **Multimedia Information Retrieval**. Synthesis Lectures on Information Concepts, Retrieval, and Services, vol. 1, n. 1, 2009.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial**. Segunda Edição, Ed. Campus, 2004.

RUTHVEN, I.; LALMAS, M. A survey on the use of relevance feedback for information access systems. **The Knowledge Engineering Review**, n.18, v.2, 2003.

SALTON, G. **Automatic Information Organization and Retrieval**. New York: McGraw-Hill, 1968.

_____. **The SMART Retrieval System: experiments in automatic document processing**. Englewood Cliffs, NJ: Prentice Hall, 1971.

SALTON, G.; BUCKLEY, C. Improving Retrieval Performance by Relevance Feedback. In K. Sparck Jones & P. Willet (Eds.), **Readings in Information Retrieval**. San Francisco, California: Morgan Kauffmann, 1997.

SALTON, G.; MCGILL, J.M. **Introduction to Modern Information Retrieval**. New York, McGraw-Hill, 1983.

SANTARÉM SEGUNDO, J. E. **Representação iterativa: um modelo para repositórios digitais**. 2010. 224 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2010.

SANTARÉM SEGUNDO, J. E.; POTTKER, L. M. V. Repositórios de objetos educacionais de aprendizagem como fonte de recurso para instrumentalização do ensino. In: Maria Regina Momesso; Filomena Elaine Paiva Assolini; Luzmara Ferreira

Curcino; Fabiane Verardi Burlamaque; Glória Maria V. Palma. (Org.). **Das práticas do ler e escrever: ao universo das linguagens, códigos e tecnologias**. Ied. Porto Alegre: Cirkula, v. 1, p. 269-288, 2015.

SANTOS, O. A. **Proposta de serviços para suporte à personalização de eAprendizagem**. 2006. Tese (Doutoramento em Ciências e Tecnologia da Comunicação). Aveiro, Portugal: Universidade de Aveiro. 2006.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p.41-62, jan./jun. 1996.

SARACEVIC, T. Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part III: Behavior and Effects on Relevance. **Journal of the American Society for Information Science and Technology**, v. 58, n.13, p.212-214, 2007.

SHAKESPEARE, W. **As You Like It**, Act 2, Scene 7, 143-166. 1599.

SILVA, D. L.; SOUZA, R. R. Representação de documentos multimídia: dos metadados às anotações semânticas. In: **XIV Encontro Nacional de Pesquisa em Ciência da Informação**, Florianópolis. ENANCIB, 2013.

SILVEIRA, R.; GOMES, E. FIPA Compliant Pedagogical Agents in Distributed Intelligent Learning Environments, **IADIS International Conference, e-SOCIETY 2003**, Lisboa, 2003.

SILVEIRA, R. A.; GOMES, E. R.; PINTO, V. H.; VICCARI, R. M. **Thinking in Agents and Learning Objects: The Intelligent Learning Object approach**. Berlin Heidelberg: Springer-Verlag, 2004.

SILVEIRA, I. F., OMAR, N., MUSTARO, P.N. Architecture of Learning Objects Repositories. In: HARMAN, K., KOOHANG, A. (Eds.). **Learning Objects: Standards, Metadata, Repositories, and LCMS**. Santa Rosa, California: Informing Science Press, p. 131-155, 2007.

SOSTERIC, M.; HESEMEIER, S. A first step toward a theory of learning objects. In: R. McGreal (Ed.), **Online education using learning objects**. London: Routledge/Falmer, 2004.

SOUTH, J. B.; MONSON, D. W. A university-wide system for creating, capturing, and delivering learning objects. In: WILEY, David A (Ed.). **The instructional use of learning objects**. Bloomington: AECT, 2000. Disponível em: <<http://www.reusability.org/read/chapters/south.doc>>. Acesso em: 30 jul. 2015.

SPINK, A.; SARACEVIC, T. Dynamics of search term selection during mediated online searching. In: **ASIS ANNUAL MEETING**, 56th, 1993, Columbus. Proceedings... New York, v. 30, p. 63-72, 1993.

STEINER, D.D. **IMAGINE**: An Integrated Environment for Constructing Distributed Artificial Intelligence Systems. In: O'HARE, G.M.P.; JENNINGS, N.R. (Eds.). Foundations of distributed artificial intelligence. New York: John Wiley & Sons, 1996.

STUIKYS, V. **Smart Learning Objects for Smart Education in Computer Science**: Theory, Methodology and Robot-Based Implementation. Springer International Publishing, 2015.

TAROUCO, L. M. R.; FABRE, M.C. J. M.; TAMUSIUNAS, F. R. Reusabilidade de objetos educacionais. **RENOTE**: Revista Novas Tecnologias da Educação, Porto Alegre, v. 1 n. 1, p. 1-11, fev. 2003. Disponível em: <<http://seer.ufrgs.br/index.php/renote/article/view/13628/7697>>. Acesso em: 16 jul. 2015.

TAROUCO, L. M. R.; RODRIGUES, A. P.; SCHMITT, M. A. R. Integração do MOODLE com repositórios abertos. **Perspect. Ciênc. Inf.**, Belo Horizonte, v. 18, n. 1, p. 66-85, Mar. 2013. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362013000100006&lng=en&nrm=iso>. Acesso em: 16 jul. 2015.

VIAN, J.; SILVEIRA, R. A. Multiagent System for Indexing and Retrieving Learning Objects. In: **9th International Conference on Practical Applications of Agents and Multi-Agent Systems**, Salamanca. Highlights in Practical Applications of Agents and Multiagent Systems. New York: Springer, 2011.

VICCARI, R. M.; BEZ, M. R.; SILVA, J. M. C.; RIBEIRO, A. M.; GLUZ, J. C.; SANTOS, É. R. Proposta de Padrão de Objetos de Aprendizagem Baseados em Agentes (OBAA). In: **Latin American Conference on Learning Objects**, 2010, São Paulo. Latin American Conference on Learning Objects, 2010.

VICCARI, R.; GLUZ, J.; SANTOS, E.; et al. **Projeto OBAA Relatório Técnico RTOBAA01** – Proposta de Padrão para Metadados de Objetos de Aprendizagem Multiplataforma. UFRGS/CINTED, 2009. Disponível em: <<http://www.portalobaa.org/obaac/padraoobaa/relatoriostecnicos/RTOBAA01.pdf/view>>. Acesso em: 28 nov. 2013.

VLASSIS, N. **A concise introduction to multiagent systems and distributed AI**. Informatics Institute, University of Amsterdam, September. 2003. Disponível em: <http://db.sis.pitt.edu/infsci3350/resources/cimasdai.pdf>. Acesso em: 02 nov. 2016.

W3C Brasil – **World Wide Web Consortium - Escritório Brasil**. Disponível em: <<http://www.w3c.br/Padroes/WebSemantica>>. Acesso em: 10 ago. 2014.

WARPECHOWSKI, M.; SOUTO, M. A. M.; OLIVEIRA, J. P. M. Techniques for Metadata Retrieval of Learning Objects. In: **International Workshop on Applications of Semantic Web technologies for E-Learning (SW-EL)**, 2006, Dublin. Proceedings of Workshops held at the IV International Conference on Adaptive Hypermedia ad

Adaptive Web-based Systems (AH 2006). Dublin: National College of Ireland, v. 1. p. 96-101, 2006.

WEIBEL, S. Metadata: The Foundations of Resource Description. **D-Lib Magazine**, July 1995. Disponível em: <<http://www.dlib.org/dlib/July95/07weibel.html>>. Acesso em: 14 jul. 2014.

WELLER, M.; PEGLER, C.; MASON, R. **Putting the pieces together**: What working with learning objects means for the educator. 2003.

WEISS, G. **Multiagent Systems**: A Modern Approach to Distributed Artificial Intelligence, Cambridge / Massachusetts: MIT Press, 1999.

WILEY, D. **Connecting learning objects to instructional design theory**: A definition, a metaphor, and a taxonomy. The Instructional Use of Learning Objects. Wiley, D. (ed.) 2001. Disponível em: <<http://www.reusability.org/read/chapters/wiley.doc>>. Acesso em: 15 ago. 2014.

WILLIAMSON, R. Does Relevance Feedback Improve Documents Retrieval Performance? In: **1st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval**: ACM, 1978.

WITTEN, I.; MOFFAT, A.; BELL, T. C. **Managing Gigabytes**: compressing and indexing documents and images. Morgan Kaufmann Publishers, Inc. Second Edition, 1999. 544p.

WOOLDRIDGE, M. J. **An Introduction to Multiagents Systems**. John Wiley & Sons Chichester, England. 2002.

WOOLDRIDGE, M.; JENNINGS, N. R. **Intelligent Agents**: Theory and Practice. Knowledge Engineering Review, 10(2), 1995.

ZHANG, J. **Visualization for Information Retrieval**. New York, NY: Springer. 2008.

ZHANG, Y. **A Comparison of Search Engines For Finding Resources**, 2004. Disponível em: <<http://www.yuanlei.com/studies/articles/is567-searchengine/page2.htm>>. Acesso em: 22 set. 2016.

ZENG, M. L.; QIN, J. **Metadata**. New York: Neal-Schuman Publishers, 2008.

ZHAI, C.; LAFFERTY, J. D. Model-based feedback in the language modeling approach to information retrieval. In: **Proceedings of CIKM '01**, 2001, p. 403–410. Disponível em: <<http://sifaka.cs.uiuc.edu/czhai/pub/cikm2001-fb.pdf>>. Acesso em: 20 out. 2016.

Apêndices

Apêndice A – Modelos de Sistema – Diagramas de Casos de Uso

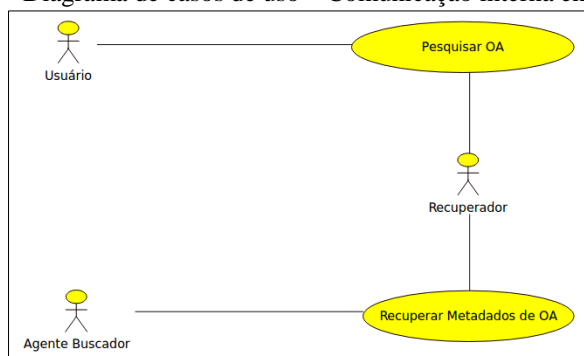
Os diagramas de casos de uso são os diagramas mais gerais e informais da UML (*Unified Modeling Language*). A UML é uma linguagem de modelagem unificada usada com o propósito geral de permitir uma visualização gráfica dos artefatos de software e pode ser utilizada em todos os domínios da aplicação. Nesta investigação foram utilizados diagramas de casos de uso para demonstrar a interação entre os atores envolvidos na arquitetura para recuperação objetos de aprendizagem e suas respectivas funcionalidades.

Os diagramas de casos de uso consistem em um modelo de usuário que permite identificar o domínio e as funcionalidades do sistema, por meio da representação gráfica dos atores (e suas interações), atividades e relações.

O primeiro diagrama de casos de uso ilustra a comunicação interna entre os atores envolvidos na arquitetura, conforme pode ser visualizado na Figura A.1. O ator Usuário pode pesquisar objetos de aprendizagem por meio de uma interface de busca que por sua vez se comunica com o ator Recuperador que trata a informação recebida, executa o motor de RI, localiza e ranqueia os itens de metadados recuperados. O ator

Usuário pode realizar o *relevance feedback* sobre os itens que foram retornados, enquanto que o ator Recuperador recebe e trata essa nova solicitação do Usuário. O ator Agente Recuperador tem a função de buscar e recuperar metadados de OA dos repositórios para serem persistidos em outro momento.

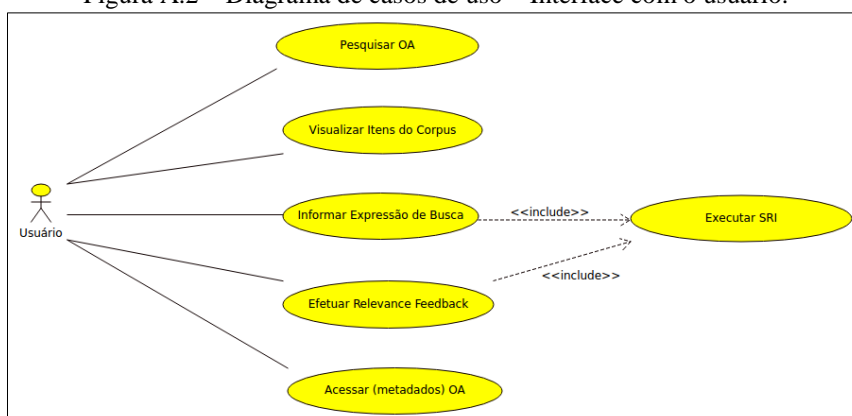
Figura A.1 – Diagrama de casos de uso – Comunicação interna entre os atores.



Fonte: elaborado pela autora.

O diagrama de casos de uso “Interface com o usuário” (Figura A.2) tem a função de viabilizar a comunicação do usuário com o SRI. Esta comunicação vai desde o momento em que o usuário declara sua necessidade de informação, por meio de uma expressão de busca até o momento em que o usuário interage com o sistema informando seu *feedback* na seleção dos itens que mais lhe parecem relevantes. O relacionamento do tipo `<<include>>` indica que o caso de uso incluído será sempre executado quando da execução do outro caso de uso, ou seja, a direção do relacionamento é do caso de uso que está incluindo para o caso de uso incluído.

Figura A.2 – Diagrama de casos de uso – Interface com o usuário.

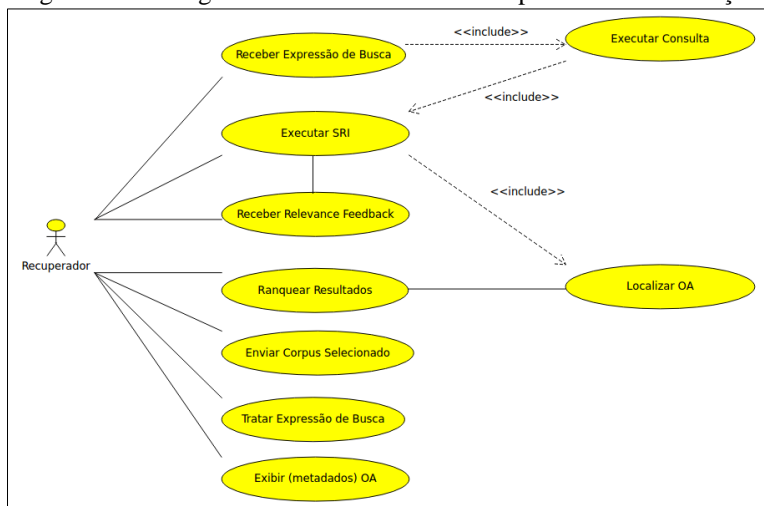


Fonte: elaborado pela autora.

O diagrama de casos de uso “Recuperador de Informação” (Figura A.3) ilustra as ações deste ator que tem a função de receber a expressão de busca do usuário, tratá-la

e submetê-la ao motor de RI que irá localizar os itens que atendem a solicitação e ranqueá-los com o intuito de retornar um conjunto de itens relevante ao usuário, que pode continuar o processo comunicação com SRI selecionado os itens que ele julga como relevantes, por meio do mecanismo de *relevance feedback*.

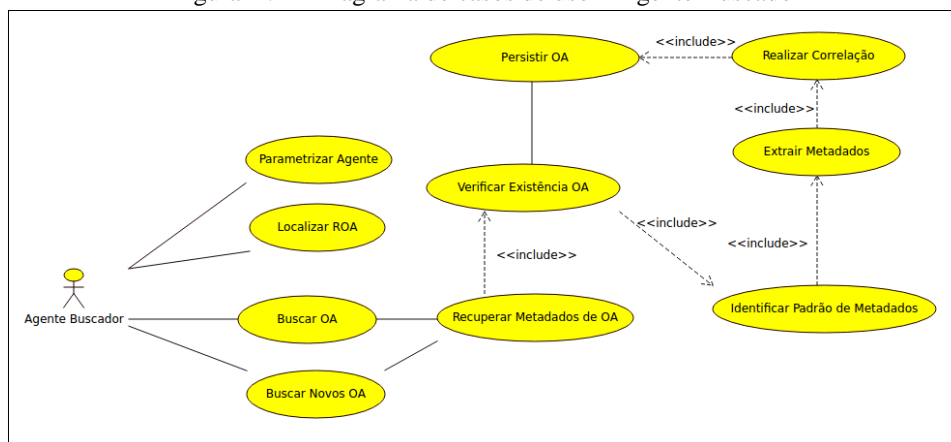
Figura A.3 – Diagrama de casos de uso – Recuperador de informação.



Fonte: elaborado pela autora.

O diagrama de casos de uso “Agente Buscador” (Figura A.4) apresenta as atividades realizadas por este ator que tem a função de localizar os ROA e buscar por (novos) metadados de OA. A partir da localização desses recursos educacionais, os metadados são identificados e verifica-se se já estão persistidos ou não. Caso ainda não estejam, serão executados os casos de uso responsáveis por extrair os metadados, correlacioná-los e persisti-los.

Figura A.4 – Diagrama de casos de uso – Agente Buscador



Fonte: elaborado pela autora.

Apêndice B – Pseudocódigo – Inicialização do Agente Buscador

O fragmento de código a seguir corresponde à inicialização do agente buscador de objetos de aprendizagem.

```
agente_busca_OA(info_metadados, ROA, Max_URLs)
{
    starting_urls := search_engine(ROA);
    for agent (1..#INIT_URLs)
    {
        initialize(agent, info_metadados, ROA, Max_URLs);
        situate(agent, starting_urls);
    }
}
```

Apêndice C – Pseudocódigo - *Loop Crawler*

O fragmento de código a seguir corresponde à execução do looping do *crawler* na busca pelas URLs que contêm os metadados de objetos de aprendizagem.

```

// Loop crawler
while (crawler && Lista_Crawler.size() > 0)
{
    // Verifica se o número máximo de URL foi alcançado, caso a restrição tenha sido especificada.
    if (Max_URLs != -1)
    {
        // Se atingir o número máximo de URLs, finalizar.
        if (Lista_Rastreados.size() == Max_URLs)
        {
            break;
        }
    }
    // Coloca a URL na lista.
    String url = (String) Lista_Crawler.iterator().next();
    // Converte a string url para um object URL.
    URL URL_Verificada = verifyUrl(url);
    // Atualiza as estatísticas do crawler.
    Atualiza_Estatistica_Crawler(url, Lista_Rastreados.size(), Lista_Crawler.size(), URL);
    // Adiciona a página na lista do crawler.
    Lista_Rastreados.add(url);
    // Realiza o download da página
    String Cont_Pagina = downloadPage(URL_Verificada);
    // Se o download foi bem sucedido, recupera todos os links e verifica a similaridade.
    if (Cont_Pagina != null && Cont_Pagina.length() > 0)
    {
        // Recupera a lista de links válidos das páginas
        ArrayList links = RecuperaLinks(URL_Verificada, Cont_Pagina, Lista_Rastreados, Limite_Host);

        // Adiciona os links para o Crawler.
        Lista_Crawler.addAll(links);

        // Verifica a similaridade, e em caso afirmativo, adiciona.
        if (searchStringMatches(Cont_Pagina, info_metadados))
        {
            addMatch(url);
        }
    }
    // Atualiza as estatísticas do crawler
    Atualiza_Estatistica_Crawler(url, Lista_Rastreados.size(), Lista_Crawler.size(),Max_URLs);
}
}

```

Apêndice D – Fragmentos de arquivos (URLs) recuperados pelo *Web Crawler*

ROA – BIOE – Data da coleta: 04/01/2017

...

<http://objetoseducacionais2.mec.gov.br/handle/mec/18779>

<http://objetoseducacionais2.mec.gov.br/handle/mec/19004>

<http://objetoseducacionais2.mec.gov.br/handle/mec/15726>

<http://objetoseducacionais2.mec.gov.br/handle/mec/3733>

<http://objetoseducacionais2.mec.gov.br/handle/mec/6270>

<http://objetoseducacionais2.mec.gov.br/handle/mec/20508>

<http://objetoseducacionais2.mec.gov.br/handle/mec/16242>

<http://objetoseducacionais2.mec.gov.br/handle/mec/10603>

<http://objetoseducacionais2.mec.gov.br/handle/mec/15335>

<http://objetoseducacionais2.mec.gov.br/handle/mec/10855>

<http://objetoseducacionais2.mec.gov.br/handle/mec/12217>

<http://objetoseducacionais2.mec.gov.br/handle/mec/17611>

<http://objetoseducacionais2.mec.gov.br/handle/mec/5304>

<http://objetoseducacionais2.mec.gov.br/handle/mec/8617>

<http://objetoseducacionais2.mec.gov.br/handle/mec/17562>

...

ROA – CESTA – Data da coleta:04/01/2017

...

<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/492?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/292?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/482?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/443?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/410?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/303?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/254?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/149?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/140?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/211?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/290?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/139?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/85?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/100?show=full>
<http://cesta2.cinted.ufrgs.br/xmlui/handle/123456789/404?show=full>

...

ROA – LUME – Data da coleta:04/01/2017

...

<http://www.lume.ufrgs.br/handle/10183/7989>
<http://www.lume.ufrgs.br/handle/10183/990>
<http://www.lume.ufrgs.br/handle/10183/1001>
<http://www.lume.ufrgs.br/handle/10183/991>
<http://www.lume.ufrgs.br/handle/10183/1006>
<http://www.lume.ufrgs.br/handle/10183/992>
<http://www.lume.ufrgs.br/handle/10183/993>
<http://www.lume.ufrgs.br/handle/10183/994>
<http://www.lume.ufrgs.br/handle/10183/996>
<http://www.lume.ufrgs.br/handle/10183/39696>
<http://www.lume.ufrgs.br/handle/10183/694>
<http://www.lume.ufrgs.br/handle/10183/39699>
<http://www.lume.ufrgs.br/handle/10183/97893>
<http://www.lume.ufrgs.br/handle/10183/997>
<http://www.lume.ufrgs.br/handle/10183/1002>

...

ROA – MERLOT – Data da coleta: 04/01/2017

...

<https://www.merlot.org/merlot/viewMaterial.htm?id=3140>
<https://www.merlot.org/merlot/viewMaterial.htm?id=6885>
<https://www.merlot.org/merlot/viewMaterial.htm?id=14696>
<https://www.merlot.org/merlot/viewMaterial.htm?id=25796>
<https://www.merlot.org/merlot/viewMaterial.htm?id=23189>
<https://www.merlot.org/merlot/viewMaterial.htm?id=81895>
<https://www.merlot.org/merlot/viewMaterial.htm?id=81896>
<https://www.merlot.org/merlot/viewMaterial.htm?id=81897>
<https://www.merlot.org/merlot/viewMaterial.htm?id=731750>
<https://www.merlot.org/merlot/viewMaterial.htm?id=1016646>
<https://www.merlot.org/merlot/viewMaterial.htm?id=1058659>
<https://www.merlot.org/merlot/viewMaterial.htm?id=1128990>
<https://www.merlot.org/merlot/viewMaterial.htm?id=1131540>
<https://www.merlot.org/merlot/viewMaterial.htm?id=1148556>
<https://www.merlot.org/merlot/viewMaterial.htm?id=3551198>

...

Anexos

Anexo A – Mapa Mental do Padrão OBAA



Anexo B – Exemplo de registro no padrão OBAA

OBAA MILOS - Mozilla Firefox

repositorio.portalobaa.org/handle/obaa/29

Logar

Página Inicial > Demonstração OBAA > Apresentações > Ver Item

OBAA MILOS

Registro de metadados resumido

Apresentar o registro completo

Título: OBAA MILOS
Autor: Gluz, João
URI: <http://repositorio.portalobaa.org/handle/obaa/29>
Data: 2012-05-17

Arquivos neste Item

Arquivos	Tamanho	Formato	Visualizar	Description
OBAA-FINEP-MEC-MC MILOS.ppt	250kb	Microsoft PowerPoint	Visualizar/ Abrir	Apresentação sobre o OBAA

Este item aparece na(s) seguinte(s) Coleção(ões)

- [Apresentações](#)

[Apresentar o registro completo](#)

Buscar no repositório

Visualizar

- Todo o repositório
 - o [Comunidades e Coleções](#)
 - o [Pela data de envio](#)
 - o [Autor](#)
 - o [Título](#)
 - o [Assunto](#)
- Esta Coleção
 - o [Pela data de envio](#)
 - o [Autor](#)
 - o [Título](#)
 - o [Assunto](#)

Minha conta

- [Entrar](#)
- [Cadastrar](#)

Este portal está usando o Manakin, uma nova facilidade criada pela Biblioteca da Universidade do Texas A&M University. A interface pode ser extensivamente modificada pelos "Aspectos" e "Temas" baseado em XSL. Para maiores informações visite: <http://di.tamu.edu> e <http://dspace.org>

Contate-nos | Envie uma mensagem para os administradores do repositório

Digital Initiatives Research & Technology

OBAA MILOS - Mozilla Firefox

repositorio.portalobaa.org/handle/obaa/29?show=full

Logar

Página Inicial > Demonstração OBAA > Apresentações > Ver Item

OBAA MILOS

Registro de metadados completo – Vazio!

Apresentar o registro simples

```

dc.contributor.author
dc.date.accessioned
dc.date.available
dc.date.issued
dc.identifier.uri
dc.title
obaa.general.title
obaa.general.language

obaa.general.description

obaa.general.keyword
obaa.general.keyword
obaa.general.structure
obaa.general.aggregationlevel
obaa.lifecycle.version
obaa.lifecycle.status
obaa.lifecycle.contribute.role
obaa.lifecycle.contribute.role
obaa.lifecycle.contribute.entity
    
```

Buscar no repositório

Visualizar

- Todo o repositório
 - o [Comunidades e Coleções](#)
 - o [Pela data de envio](#)
 - o [Autor](#)
 - o [Título](#)
 - o [Assunto](#)
- Esta Coleção
 - o [Pela data de envio](#)
 - o [Autor](#)
 - o [Título](#)
 - o [Assunto](#)

Minha conta

- [Entrar](#)
- [Cadastrar](#)