

Proxério Manoel Felisberto

**GESTÃO DOS METADADOS CONTIDOS NOS  
IDENTIFICADORES DE OBJETOS DE BIBLIOTECONOMIA**

Dissertação submetida ao Programa de Pós-Graduação em Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Tecnologias da Informação e Comunicação.

Orientador: Prof. Dr. Roderval Marcelino

Coorientador: Prof. Dr. João Bosco da Mota Alves

Araranguá  
2016

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Felisberto, Proxério Manoel  
Gestão dos Metadados Contidos nos Identificadores de  
Objetos de Biblioteconomia / Proxério Manoel Felisberto ;  
orientador, Roderval Marcelino ; coorientador, João Bosco  
da Mota Alves. - Araranguá, SC, 2016.  
150 p.

Dissertação (mestrado) - Universidade Federal de Santa  
Catarina, Campus Araranguá. Programa de Pós-Graduação em  
Tecnologias da Informação e Comunicação.

Inclui referências

1. Tecnologias da Informação e Comunicação. 2. Recuperação  
da Informação. 3. Digital Object Identifier. 4. Metadados.  
I. Marcelino, Roderval. II. Alves, João Bosco da Mota.  
III. Universidade Federal de Santa Catarina. Programa de  
Pós-Graduação em Tecnologias da Informação e Comunicação. IV.  
Título.

Proxério Manoel Felisberto

**GESTÃO DOS METADADOS CONTIDOS NOS  
IDENTIFICADORES DE OBJETOS DE BIBLIOTECONOMIA**

Esta Dissertação foi julgada adequada para obtenção do Título de “Mestre”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina.

Araranguá, 12 de maio de 2016.

---

Profa. Simone Meister Sommer Bilessimo, Dra.  
Coordenadora do Curso

**Banca Examinadora:**

---

Prof. Roderval Marcelino, Dr.  
Orientador  
Universidade Federal de Santa Catarina

---

Prof. Vinícius Medina Kern, Dr.  
Universidade Federal de Santa Catarina (Videoconferência)

---

Prof. Vilson Gruber, Dr.  
Universidade Federal de Santa Catarina

---

Prof. Alexandre Leopoldo Gonçalves, Dr.  
Universidade Federal de Santa Catarina



Este trabalho é dedicado a todos que de alguma maneira colaboraram para o êxito desta pesquisa.



## AGRADECIMENTOS

A caminhada ao lado de pessoas tão especiais torna-se leve e prazerosa, mesmo através de rochas escarpadas. Agradeço à Providência Divina e a todos quantos ela permitiu estar comigo nesta jornada me auxiliando e contribuindo para seu êxito, em especial, minha esposa Iraci Zanivan Felisberto e meu filho Thiago Zanivan Felisberto.

A minha família pela dedicação, compreensão, apoio e incentivo sempre incondicionais.

Aos colegas remanescentes da graduação e aos novos conhecidos pelos ensinamentos e pelos debates em um ambiente de sã camaradagem rumo a construção de novos conhecimentos.

A todos que direta ou indiretamente prestaram sua contribuição e desinteressadamente contribuíram com este empreendimento.

Aos professores do PPGTIC pela forma amistosa, organizada e profissional com que conduziram suas disciplinas, os meus agradecimentos e o meu respeito.

Agradeço aos integrantes da banca examinadora por terem aceitado o convite e pelas valiosas observações para o aprimoramento do presente trabalho.

Ao coorientador, Prof. Dr. João Bosco da Mota Alves, pela sabedoria contida nos conselhos e pela convivência preexistentes ao curso os meus agradecimentos e minha admiração.

Especialmente, ao orientador, Prof. Dr. Roderval Marcelino, que tive o privilégio de conhecer durante o Curso de Tecnologia em Automação Industrial, realizado na SATC, que esteve comigo nos Estágios I e II e no Trabalho de Conclusão de Curso na graduação em Tecnologias da Informação e Comunicação, realizado na UFSC, que foi o inspirador deste trabalho e que sempre me orientou com precisão, paciência, sabedoria e amizade, os meus agradecimentos, respeito, admiração e minha amizade.



“O que sabemos é uma gota; o que ignoramos é um oceano.”

(Sir Isaac Newton)



## RESUMO

O intuito do presente trabalho é analisar o impacto que o gerenciamento dos metadados contidos em identificadores digitais de objetos, utilizados na Biblioteconomia e apoiados pelas Tecnologias da Informação e Comunicação, pode causar sobre a usabilidade do Mecanismo Online para Referências e sobre a satisfação dos usuários ao fazer uso do mesmo. Para a realização desta pesquisa, um protótipo que recupera os metadados vinculados a um determinado *Digital Object Identifier* junto aos repositórios das Autoridades de Registro da *International DOI Foundation* foi desenvolvido e integrado ao Mecanismo Online para Referências, com o objetivo de auxiliar seus usuários preenchendo automaticamente campos do formulário de geração de referências a artigos de periódico. Um questionário para levantar o grau de utilidade percebida frente as funcionalidades do protótipo permaneceu disponível durante cinco meses para ser respondido pelos usuários que fizeram uso destas. O referido questionário é constituído por dez questões fechadas e uma questão aberta. Uma escala de Likert de seis níveis foi adotada para quantificar as opções de resposta das questões fechadas. A estrutura deste relatório comporta 8 (oito) capítulos, a saber: Introdução; Recuperação da Informação; Identificação de Recursos de Informação; Tecnologias Computacionais Utilizadas; Procedimentos Metodológicos; Desenvolvimento do Protótipo; Resultados; e, Considerações Finais. Da análise dos dados coletados, conclui-se que a gestão aqui abordada impacta positivamente na usabilidade do sistema e no grau de utilidade percebida e de satisfação do usuário, em concordância com 93,9% dos usuários amostrados.

**Palavras-chave:** Tecnologias da Informação e Comunicação. Recuperação da Informação. *Digital Object Identifier*. Metadados.



## ABSTRACT

The purpose of this study is to analyze the impact that the digital object identifiers contained metadata, used in Librarianship and supported by Information and Communication Technologies, can have on Mecanismo Online para Referências' quality and on the perceived user satisfaction while using the application. In order to validate this research, a prototype that retrieves metadata linked to a particular Digital Object Identifier from the Registration Authorities of the International DOI Foundation repositories was developed and integrated into Mecanismo Online para Referências to assist its users by automatically filling form fields used to generate references to journal papers. A questionnaire to determine the perceived usefulness degree of the prototype's functionalities remained available during five months to be answered by users that used it. The questionnaire consists of ten closed questions and one open question. A six levels' Likert scale was adopted to quantify the response options of closed questions. The structure of this report contains eight (8) chapters, namely: Introduction; Information Retrieval; Information Resources Identification; Used Computer Technologies; Methodological Procedures; Prototype Development; Results; Final Considerations. From the analysis of the sampled data and concludes that the proposed functionality impacts positively the application's quality and the perceived degree of usefulness and user satisfaction, as agreed by 93.9% of the sampled users.

**Keywords:** Information and Communication Technologies. Information Retrieval. Digital Object Identifier. Metadata.



## LISTA DE FIGURAS

Figura 1 – Sequência de ações desenvolvidas para a realização da dissertação. ....	35
Figura 2 – Ciclo da Informação. ....	38
Figura 3 – Gráfico Conceitual da Representação Documentária. ....	44
Figura 4 – Exemplo de registro no padrão MARC 21. ....	69
Figura 5 – Exemplo de prefixo e sufixo de um DOI® para periódicos UFSC. ....	79
Figura 6 – Exemplo de referência bibliográfica utilizando um enlace ativo. ....	80
Figura 7 – Diagrama da sintaxe de uma coleção de pares de string/valor. ....	94
Figura 8 – Diagrama da sintaxe de uma lista ordenada. ....	95
Figura 9 – Tela do MORE referente ao formulário de contato. ....	106
Figura 10 – Tela do MORE referente ao formulário de referência a artigo de revistas. ....	107
Figura 11 – Diagrama de Cadastro de Referência a Artigo de Periódico. ....	108
Figura 12 – Diagrama de Máquina de Estado de Cadastro de Referência a Artigo de Periódico. ....	109
Figura 13 – Tela do MORE de inclusão e gerenciamento de coleções. ....	110
Figura 14 – Tela de gerenciamento de coleções pessoais. ....	111
Figura 15 – Exemplo de objeto JSON. ....	112
Figura 16 – Formulário parcialmente preenchido com o uso de um nome DOI. ....	113
Figura 17 – Extrato do relatório do <i>Google Analytics</i> , referente a utilização do sistema MORE durante o mês de novembro de 2015. ....	113
Figura 18 – Estrutura de chave de dados para recuperação da informação. ....	115
Figura 19 – Histograma das Questões de Medida da Utilidade Percebida. ....	119
Figura 20 – Histograma das Questões de Medida da Satisfação do Usuário. ....	120
Figura 21 – Histograma do Somatório dos Itens de Likert. ....	121
Figura 22 – Gráfico comparativo dos dados consolidados para Utilidade Percebida, Satisfação do Usuário e Somatório dos Itens de Likert. ....	121



## LISTA DE QUADROS

Quadro 1 – Exemplo de registro no padrão <i>Dublin Core</i> . .....	68
Quadro 2 – Exemplo de código XML para registro de nomes DOI na CrossRef. ....	79



## LISTA DE TABELAS

Tabela 1 – Compilação dos Identificadores de Objeto (recomposta de Park et al. (2008) e NISO (2006)).....	71
Tabela 2 – Dados estatísticos individualizados por questão. ....	118
Tabela 3 – Coeficiente de Correlação de Pearson entre as questões do questionário. ....	122
Tabela 4 – Intensidade e direção dos valores do coeficiente de correlação. ....	123
Tabela 5 – Coeficiente de Correlação de Pearson entre as medidas de Utilidade Percebida e Satisfação do Usuário. ....	123



## LISTA DE ABREVIATURAS E SIGLAS

ABNT – Associação Brasileira de Normas Técnicas  
AJAX – *Asynchronous JavaScript and XML*  
API – *Application Programming Interface*  
AR – Autoridades de Registro  
ARK – *Archival Resource Key*  
CDD – Classificação Universal de Dewey  
CDU – Código de Classificação Universal  
CI – Ciência da Informação  
cID – *Content ID*  
CIDF – *Content ID Forum*  
CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico  
CNRI – *Corporation for National Research Initiatives*  
CRUD – *Create, Read, Update, Delete*  
CURL – *Client URL Request Library*  
DC – *Dublin Core*  
DCMI - *Dublin Core Metadata Initiative*  
DOI<sup>®</sup> – *Digital Object Identifier*  
GAD – Gerenciamento de Ativos Digitais  
HTTP – *Hypertext Transfer Protocol*  
ICE – Integração de Conteúdo Empresarial  
IDF – *International DOI Foundation*  
IETF – *Internet Engineering Task Force*  
ISAN – *International Standard Audiovisual Number*  
ISBN – *International Standard Book Number*  
ISNI – *International Standard Name Identifier*  
ISRC – *International Standard Recording Code*  
ISSN – *International Standard Serial Number*  
JSON – *JavaScript Object Notation*  
LC – *Library of Congress*  
LCCN – *Library of Congress Control Number*  
MARC – *MAchine-Readable Cataloguing*  
MIME – *Multipurpose Internet Mail Extensions*  
MORE – Mecanismo Online para Referências  
NBNs – *National Bibliography Numbers*  
NCA – *National Computerization Agency*  
NISO – *National Information Standard Organization*  
PIR – *Personal Information Retrieval*

PPGTIC – Programa de Pós-Graduação em Tecnologias da Informação e Comunicação  
PURL – *Persistent URL*  
REST – *Representational State Transfer*  
RI – Recuperação da Informação  
SRI – Sistemas de Recuperação da Informação  
SeTIC – Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação  
TDI – Tratamento Descritivo da Informação  
TI – Tecnologia da Informação  
TICs – Tecnologias da Informação e Comunicação  
UCI – *Universal Content Identifier*  
UEM – Universidade Estadual de Maringá  
URI – *Uniform Resource Identifier*  
URL – *Uniform Resource Locator*  
URN – *Uniform Resource Name*  
XML – *eXtensible Markup Language*  
www – *World Wide Web*

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>25</b>
1.1	APRESENTAÇÃO DO PROBLEMA.....	25
1.2	OBJETIVOS .....	29
1.2.1	Objetivo Geral.....	29
1.2.2	Objetivos Específicos .....	29
1.3	JUSTIFICATIVA .....	30
1.4	ESCOPO DO TRABALHO.....	31
1.5	ADERÊNCIA AO PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO.....	32
1.6	METODOLOGIA.....	33
1.7	ESTRUTURA DO TRABALHO.....	34
<b>2</b>	<b>RECUPERAÇÃO DA INFORMAÇÃO</b> .....	<b>37</b>
2.1	INFORMAÇÃO .....	37
2.2	CIÊNCIA DA INFORMAÇÃO.....	39
2.3	REPRESENTAÇÃO DOCUMENTÁRIA.....	41
2.4	RECUPERAÇÃO DA INFORMAÇÃO PROPRIAMENTE DITA .....	51
<b>3</b>	<b>IDENTIFICAÇÃO DE RECURSOS DE INFORMAÇÃO</b> ....	<b>59</b>
3.1	METADADOS .....	59
3.1.1	Categorização dos Metadados .....	63
3.1.2	Padrão de Metadados .....	66
3.2	IDENTIFICADORES DE OBJETOS .....	69
3.3	IDENTIFICADORES DIGITAIS DE OBJETOS.....	73
3.3.1	<i>Universal Content Identifier (UCI)</i> .....	75
3.3.2	<i>Content ID (cID)</i> .....	76
3.3.3	<i>Digital Object Identifier (DOI®)</i> .....	77
<b>4</b>	<b>TECNOLOGIAS COMPUTACIONAIS UTILIZADAS</b> .....	<b>83</b>
4.1	<i>REPRESENTATIONAL STATE TRANSFER (REST)</i> .....	83
4.2	<i>JAVASCRIPT OBJECT NOTATION (JSON)</i> .....	90
4.3	<i>CLIENT URL REQUEST LIBRARY (CURL)</i> .....	95
4.3.1	Biblioteca libcurl.....	96
4.3.2	Curl, a ferramenta de linha de comando .....	97
<b>5</b>	<b>PROCEDIMENTOS METODOLÓGICOS</b> .....	<b>99</b>
<b>6</b>	<b>DESENVOLVIMENTO DO PROTÓTIPO</b> .....	<b>105</b>
<b>7</b>	<b>RESULTADOS</b> .....	<b>117</b>
<b>8</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>127</b>
	<b>REFERÊNCIAS</b> .....	<b>131</b>
	<b>APÊNDICE A – Excerto do Código Fonte do Protótipo</b> .....	<b>143</b>
	<b>APÊNDICE B – Questionário</b> .....	<b>145</b>
	<b>ANEXO A - Exemplo de Registro MARXML</b> .....	<b>149</b>



# 1 INTRODUÇÃO

## 1.1 APRESENTAÇÃO DO PROBLEMA

Historicamente ocorreram duas revoluções envolvendo a área do conhecimento Ciência da Informação e, por consequência uma de suas ramificações, a Biblioteconomia. A primeira refere-se ao aumento considerável de publicações, que se chama de documentária e, a segunda, ao desenvolvimento de tecnologias de informação e comunicação que estão intrinsecamente ligadas à produção e disseminação dessas publicações, chamada de tecnológica (DZIEKANIAK, 2010).

Constata-se que a produção de conhecimento cresce exponencialmente e cabe à Biblioteconomia a tarefa de “reunir, processar e difundir o conteúdo informacional dos documentos”. Sua capacidade de mediar relações interdisciplinares e sua aproximação às Tecnologias da Informação e Comunicação trouxe à Biblioteca a possibilidade de recuperação da informação através destas tecnologias (DZIEKANIAK, 2010, TEIXEIRA; SCHIEL, 1997; TEIXEIRA, 2010; RODRIGUES; CRIPPA, 2011; XAVIER, 2009; TEIXEIRA, 2005; OLIVEIRA; ARAUJO, 2012; PONTES JUNIOR; CARVALHO; AZEVEDO, 2013; SARACEVIC, 1999; DODEBEI, 2014).

Brito, Guedes e Shintaku (2013) ponderam que a *Web*<sup>1</sup> revelou-se uma fonte de informação valiosa na disseminação da informação, em particular, na comunicação científica. No entanto, apesar da importância da internet, algumas questões ainda não foram totalmente elucidadas, enlaces rompidos e alteração de URL ainda são comuns.

Em complemento às considerações acima, Weber (2013) esclarece que “a preservação digital levanta desafios de uma natureza fundamentalmente diferente se comparados com a preservação dos formatos tradicionais”. A autora cita novas formas de acesso, guarda e uso do conteúdo informacional digital, bem como as preocupações sobre a garantia de acesso a estes conteúdos no longo prazo.

---

<sup>1</sup> *Web* é o termo pelo qual se faz referência ao ramo da internet conhecida como rede mundial de computadores, popularizada a partir de 1991, devido a implementação de uma interface gráfica que facilitou o acesso e estendeu seu alcance ao público em geral. Comporta-se como uma aplicação distribuída de hipermídia utilizando-se do protocolo HTTP (*Hypertext Transfer Protocol*) que possibilita a leitura não linear através de enlaces que conduzem à apresentação dos recursos informacionais a eles vinculados.

A fim de fazer frente a este desenvolvimento acelerado, uma grande quantidade de sistemas de informação diferentes e uma grande variedade de documentos criaram um contexto muito complexo, requerendo soluções mais abrangentes que as existentes àquele momento de modo a permitir a identificação e localização dos objetos digitais de modo eficaz (BRITO; GUEDES; SHINTAKU, 2013).

É neste contexto que surgem os identificadores de objetos e com eles ganham notoriedade os metadados, que foi criado por Jack E. Myers nos idos de 1960, para referir-se a “conjuntos de dados”. Na década de 80 o termo metadados surge na literatura de sistemas de gestão de banco de dados para descrever “as características da informação contida nas bases de dados” e, na década de 90, passa a ser utilizado por comunidades específicas que contribuem para o desenvolvimento de padrões facilitando seu intercâmbio entre diferentes sistemas (ALVES; SANTOS, 2013).

Os identificadores receberam ao longo deste período várias classificações e denominações, porém serão abordadas apenas as que, neste momento, são estritamente necessárias ao entendimento do problema. Assim, vários projetos de identificadores foram realizados durante este período, dos quais podemos destacar alguns que se encontram em pleno funcionamento, tais como: ISAN (*International Standard Audiovisual Number*), ISBN (*International Standard Book Number*), ISRC (*International Standard Recording Code*), ISSN (*International Standard Serial Number*), ISNI (*International Standard Name Identifier*) (BRITO; GUEDES; SHINTAKU, 2013).

Entretanto, para suprir as necessidades do ambiente *Web* e de bibliotecas digitais esses identificadores se mostraram deficientes, apesar de serem identificadores persistentes, mas não são identificadores digitais. Keefer (2001) entende que à medida que o número de documentos digitais cresce e a facilidade de navegação entre eles melhora, os usuários se sentem cada vez mais frustrados quando não são capazes de ter acesso eletrônico aos artigos citados por outros artigos. Estima, ainda, o autor que 7% dos usuários vivenciam tal situação.

Identificadores persistentes são aqueles que dispõem de uma estrutura tecnológica com pré-disposição de longevidade temporal e evidenciam as seguintes características: acionabilidade, escopo do padrão, arquitetura e infraestrutura subjacente e estado do padrão. Segundo os ensinamentos de Weber (2013) esta abordagem “consiste num identificador numérico, que vem acompanhado de resolução, metadados e política da instituição que os adota”.

Alguns destes identificadores possuem normas em estágio avançado de desenvolvimento. Podemos incluir neste rol o URN (*Uniform Resource Name*), o PURL (*Persistent URL*), o DOI® (*Digital Object Identifier*), o NBNs (*National Bibliography Numbers*), o ARK (*Archival Resource Key*) e o *Open URL* (TONKIN, 2008).

O protocolo URN (*Uniform Resource Name*) foi desenvolvido pela IETF (*Internet Engineering Task Force*) numa tentativa de identificação persistente de objetos digitais (PARK et al., 2011).

No tocante aos sistemas de identificadores digitais de objetos a NISO (*National Information Standard Organization*) propõe quatro requisitos básicos: “a sintaxe de nomeação; um sistema de informação; metadados; políticas e procedimentos de governança e aplicação” (PASKIN, 2005 apud PARK et al., 2011). São exemplos deste tipo de sistema para identificação digital de objetos: DOI (*Digital Object Identifier*), UCI (*Universal Content Identifier*) na Coreia do Sul e cID (*Content ID*) no Japão.

Segundo Sayão (2010), tratar sobre serviços e sistemas de informação, atualmente, faz-se necessário o envolvimento direto com questões relacionadas aos metadados, argumento este que se coaduna com o de Dziekaniak (2010). O ambiente da biblioteca cumpre seu papel de um esquema formal utilizado no gerenciamento de objetos, digitais e não digitais, ali disponibilizados e, por consequência, metadados estruturados. As buscas por padronização dos metadados fizeram emergir alguns projetos, tais como o MARC 21, da NISO (*National Information Standard Organization*), e o *Dublin Core* da DCMI (*Dublin Core Metadata Initiative*).

No que se refere à catalogação, uma das principais razões para o uso de metadados é a de “facilitar a descoberta de informações relevantes, seja no ambiente da biblioteca, seja no ambiente *Web*”. Quanto mais essencial é o metadado do mundo digital, maior complexidade e abrangência serão conferidas à sua estrutura (SAYÃO, 2010).

Existem várias definições para o termo metadados conforme o contexto de sua utilização e o entendimento dos autores e dos pesquisadores.

Oliveira (2009) entende que uma definição genérica para o termo metadados é tudo o “que pode ser dito sobre qualquer objeto de informação em qualquer nível de agregação”.

A Norma ISO 26324:2012 define que metadados são dados específicos associados com uma referência, com base em um modelo de dados estruturado persistindo a associação com os objetos de qualquer

grau de precisão desejado e granularidade para suportar identificação, descrição e serviços.

E a NISO (2004) conceitua metadados como informações estruturadas que descrevem, explicam, localizam, ou de outra forma torna mais fácil recuperar, usar ou gerenciar um recurso de informação. No ambiente de biblioteca é comumente usado para qualquer esquema formal de descrição de recursos, aplicando-se a qualquer tipo de objeto, digital ou não digital.

Verifica-se, em Tonkin (2008), a falta de um padrão único para a interligação entre os sistemas identificadores, com a agravante que o volume de informação originalmente digital cresce vertiginosamente.

Dentre os sistemas identificadores de objetos, o *Digital Object Identifier* (DOI<sup>®</sup>) *System* da *International DOI Foundation* (IDF) vem ganhando destaque a nível mundial conforme pode ser observado em Ma e Yang (2014). Este sistema oferece estrutura, e dispõe de regras específicas, para gerenciamento dos metadados associados aos nomes DOI atribuídos por suas Autoridades de Registro (AR). A sindicância realizada nas estruturas de metadados informadas pelos seus afiliados contribuem para a correção de eventuais erros. No entanto, em alguns casos, a sintaxe deixa a desejar quanto a forma da escrita para utilização destes metadados na geração de referências bibliográficas de acordo com as normas técnicas da ABNT.

A nível nacional pode-se observar um crescimento no número de afiliados à IDF que estão utilizando o DOI. Dentre estes afiliados cita-se como exemplo:

- O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), que através de sua Plataforma Lattes, vem utilizando identificadores digitais de objetos para auxiliar no preenchimento e na validação de produção científica ali cadastrada; e,.

- A Universidade Federal de Santa Catarina (UFSC) que é pioneira nacional na utilização do DOI<sup>®</sup> em seus periódicos e destaca-se como uma das organizações com o maior número de nomes DOI<sup>®</sup> atribuídos.

No contexto deste trabalho busca-se a aquisição e o gerenciamento dos metadados associados a um nome DOI para posterior utilização daqueles necessários à geração e ao gerenciamento de referências bibliográficas através do Mecanismo Online para Referências (MORE). O MORE nasceu das observações da Sra. Maria Bernadete Martins Alves relativas às dificuldades encontradas pelos usuários da norma técnica referente à elaboração de referências bibliográficas no ambiente da Biblioteca Universitária da UFSC e seu

desenvolvimento ocorreu em 2005. Trata-se de uma ferramenta *web* gratuita destinada ao auxílio de usuários de bibliotecas (pesquisadores, professores e alunos) na tarefa de gerar e gerenciar suas listas de referências bibliográficas, de acordo com a NBR 6023/2002 da ABNT e mantê-las armazenadas em uma base de dados própria (ALVES; MENDES; ALVES, 2006).

Atualmente o MORE foi institucionalizado pela UFSC e encontra-se hospedado nos servidores da Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação (SeTIC) desta Universidade. Conta com mais de 160.000 usuários cadastrados e quase 6.000.000 de referências armazenadas em sua base de dados.

O panorama aqui apresentado contribuiu para o surgimento da seguinte pergunta de pesquisa: **Qual o impacto da gestão dos metadados contidos nos identificadores digitais de objetos, utilizados na Biblioteconomia, sobre a usabilidade do MORE e esta, por sua vez, sobre a utilidade percebida e a satisfação do usuário?**

## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Analisar o impacto da gestão dos metadados contidos nos identificadores digitais de objetos, utilizados na Biblioteconomia, sobre a usabilidade do MORE e esta, por sua vez, sobre a utilidade percebida e a satisfação do usuário.

### 1.2.2 Objetivos Específicos

O levantamento dos objetivos específicos deu-se em função do caminho a ser percorrido para alcançar o proposto no objetivo geral. São eles:

- Estudar cientificamente os sistemas de gestão de metadados de biblioteconomia e seus fundamentos teóricos;
- Desenvolver um protótipo capaz de extrair os metadados armazenados em um identificador digital de objetos e selecionar aqueles utilizados na geração de referência bibliográfica a artigo de periódico;
- Verificar se a integração do protótipo ao sistema em uso altera o comportamento de algum requisito não funcional;

- Levantar os benefícios que a gestão dos metadados contidos nos identificadores digitais de objetos, utilizados em Biblioteconomia, podem trazer aos usuários de bibliotecas e à comunidade em geral.

### 1.3 JUSTIFICATIVA

Segundo nos ensina Park et al. (2011) o identificador mais comum na *Web* é o URL (*Uniform Resource Locator*). É usado para localizar um recurso, ligando este recurso da internet à sua localização na rede ou no sistema de arquivos. Todavia, a remoção deste recurso compromete sua funcionalidade. Também não dispõem de características que possibilitem associar meta informações a respeito do recurso.

Com o intuito de amenizar estes e outros problemas correlatos, além de prover acesso rápido e seguro aos documentos é que surgiram os sistemas de identificadores digitais de objetos, trazendo consigo toda a estrutura tecnológica de suporte, cuja finalidade é dispor de informações sobre o objeto identificado, ou seja, seus metadados.

Ressalta-se, ainda, a contribuição prestada por uma análise que busca verificar, junto aos usuários do MORE, se a utilização das Tecnologias da Informação e Comunicação (TICs) podem impactar sobre a usabilidade do sistema, a utilidade percebida e a satisfação deste usuário.

Entende-se que a abordagem empregada envolvendo a gestão dos metadados do *Digital Object Identifier (DOI<sup>®</sup>)* em auxílio aos usuários de bibliotecas e a verificação do nível de utilidade percebida por estes no momento de sua utilização pressupõe uma contribuição científica baseada na disponibilidade dos resultados alcançados à comunidade acadêmica, uma vez que não foi possível encontrar nas bases de dados científicas ou na *Web* algum trabalho que contemple este assunto sob esta ótica.

O entendimento anterior emergiu das consultas realizadas aos periódicos do Portal de Periódicos da CAPES, ao *Google Scholar* e na *Web*, através dos buscadores *Google* e *Yahoo*. Utilizou-se como termos de busca “*Digital Object Identifier*” e “Identificador Digital de Objetos”, bem como suas variações. Nestes repositórios consultados foram encontrados dois trabalhos: o de Park et al. (2011) e o de Ma e Yang (2014). Entretanto, Park et al. (2011) analisam os fatores de sucesso para a adoção de sistemas de identificadores digitais de objetos através de pesquisa realizada junto aos fornecedores de recursos informacionais digitais, na Coreia do Sul, que utilizam o Sistema *Universal Content*

*Identifier* (UCI) provido pelo governo daquele país. O trabalho de Ma e Yang (2014) advém da preocupação dos pesquisadores chineses com a alta taxa de produção científica e a avaliação da necessidade urgente de um esquema de aquisição ágil de bibliografias. Assim, propuseram um modelo de estrutura que combina serviço de conteúdo *CrossCite* DOI aberto e resolvedores DOI das agências do registro.

Do exposto, um trabalho acadêmico sobre a gestão destas informações contidas nos identificadores digitais de objetos justifica-se pelos benefícios vislumbrados aos usuários de bibliotecas, bem como à comunidade acadêmica em geral, no tocante a economia de tempo, a produtividade, a simplicidade e a praticidade de gerenciamento de suas bibliografias e à diminuição da taxa de erros de digitação que ocasiona a geração de referências bibliográficas incorretas.

#### 1.4 ESCOPO DO TRABALHO

Esta pesquisa tem por objetivo analisar o impacto da gestão dos metadados referentes ao sistema de identificadores digitais de objetos mantido pela *International DOI Foundation* (IDF), mais especificamente aqueles utilizados na identificação de recursos digitais informacionais de interesse da Biblioteconomia, sobre a usabilidade do Mecanismo Online para Referências (MORE) e esta, por sua vez, sobre a utilidade percebida e sobre a satisfação do usuário.

O desenvolvimento do subsistema (protótipo) mostra-se necessário para oferecer o devido suporte ao intercâmbio de metadados entre o MORE e as Autoridades de Registro (AR) da IDF. Fica restrito somente ao DOI<sup>®</sup> e ao formulário de elaboração de referências bibliográficas a artigos de revistas, uma vez que quase a totalidade dos nomes DOI<sup>®</sup>, atualmente atribuídos, identificam este tipo de fonte.

Não se encontra sob o escopo deste trabalho a verificação da veracidade dos metadados recebidos, uma vez que as AR sindicam estes dados antes de validar qualquer atribuição de identificador. Metadados de outros identificadores (tais como: UCI, cID, ISBN, etc) também não são contemplados na implementação do protótipo.

O público alvo desta pesquisa fica restrito aos usuários do MORE que fizerem uso do subsistema através da utilização de um nome DOI<sup>®</sup> para efetuar o preenchimento automático dos campos do formulário.

A presente pesquisa trabalha somente com o serviço de conteúdos abertos disponibilizados pelas Autoridades de Registro da *International DOI Foundation* e com seus resolvedores DOI.

## 1.5 ADERÊNCIA AO PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO

O Programa de Pós-Graduação em Tecnologias da Informação e Comunicação (PPGTIC) tem como área de concentração Tecnologia e Inovação, dividida em três linhas de pesquisa, a saber: Tecnologia Computacional, Tecnologia Educacional e Tecnologia, Gestão e Inovação.

A linha de pesquisa que provocou maior atenção no desenvolvimento deste trabalho foi a de Tecnologia Computacional. De acordo com PPGTIC (2016) o objetivo dos trabalhos “desta linha de pesquisa é desenvolver modelos, técnicas e ferramentas computacionais auxiliando na resolução de problemas de natureza interdisciplinar”. Aplicações nas áreas de educação e gestão são norteadoras para o desenvolvimento destas novas tecnologias computacionais.

A aderência desta dissertação ao PPGTIC dá-se mediante a evidência de características que podem ser observadas ao longo deste trabalho. Inicialmente, evoca-se a interdisciplinaridade fundamentada na Seção 1.6, uma vez que é um dos pilares de sustentação do Programa.

Não obstante as evidências encontradas em todos os capítulos desta dissertação, os capítulos 5 e 6 são esclarecedores quanto ao alinhamento deste trabalho com os objetivos da linha de pesquisa Tecnologias Computacionais. Pode-se observar o desenvolvimento de uma ferramenta computacional com a finalidade de promover o intercâmbio de metadados que serão utilizados na geração de referências bibliográficas.

É de domínio público que os principais utilizadores de referências bibliográficas são os usuários de bibliotecas (alunos, professores, pesquisadores e outros), ou seja, um público voltado à educação, uma vez que estas referências, e respectivas citações, são utilizadas na concessão do mérito ao autor da ideia (ou texto) empregada na fundamentação de pesquisas em desenvolvimento.

No tocante a gestão e inovação percebe-se que ao intercambiar metadados de identificadores digitais de objetos de Biblioteconomia com as Autoridades de Registro (AR) da *International DOI Foundation* (IDF), de forma gratuita, promove-se novas opções para auxiliar na gestão e utilização destes metadados. No entanto, sua utilização de forma gratuita é muito difícil identificar na literatura disponível. Verifica-se que quase toda literatura consultada faz alusão, somente, àqueles serviços fornecidos pelas AR da IDF que são remunerados, a

exceção de Ma e Yang (2014) que utilizam dados abertos no *framework* por eles desenvolvido.

## 1.6 METODOLOGIA

Esta Seção tem por objetivo descrever a metodologia utilizada e classificar a pesquisa sob a perspectiva de 08 (oito) critérios que serão especificados a seguir. Cabe ressaltar que não há um consenso entre os estudiosos do assunto, porém a classificação aqui utilizada aglutina o pensamento de vários autores.

Quanto aos campos e setores do conhecimento, classifica-se como interdisciplinar, uma vez que o conhecimento necessário ao seu desenvolvimento tramita por várias disciplinas e está “baseado numa relação de integração entre as partes constituintes” destas disciplinas acadêmicas. Exemplificando, pode-se citar: a Ciência da Informação, a Biblioteconomia, Ciência da Computação e as Tecnologias da Informação e Comunicação (FARIAS FILHO; ARRUDA FILHO, 2013).

Quanto à utilização de seus resultados, trata-se de uma pesquisa aplicada em virtude da utilização dos resultados obtidos na disponibilização de novas funcionalidades do MORE ao público em geral. Segundo Cervo, Bervian e Silva (2007) “na pesquisa aplicada, o investigador é movido pela necessidade de contribuir para fins práticos mais ou menos imediatos, buscando soluções para problemas concretos.

Quanto a sua abrangência no tempo, enquadra-se como estudos transversais uma vez que se optou pelo método de pesquisa *survey* interseccional para o levantamento dos dados. Esclarecem Farias Filho e Arruda Filho (2013) que estudos transversais são estudos “feitos uma vez e representam apenas um determinado momento”.

Quanto aos seus objetivos, pode-se afirmar que é uma pesquisa, eminentemente, descritiva porque busca descrever o grau de satisfação do usuário em relação a utilização do DOI<sup>®</sup> no momento que ele responde o questionário aplicado. Para Cervo, Bervian e Silva (2007) a “pesquisa descritiva observa, registra, analisa e correlaciona fatos ou fenômenos (variáveis) sem manipulá-los”.

Quanto ao tipo de abordagem, é predominantemente quantitativa, considerando-se que o questionário é composto, basicamente, por questões fechadas e utiliza como resposta uma escala de *Likert* de 6 (seis) níveis. No entendimento de Vieira (2009) a pesquisa quantitativa fundamenta-se em informações de natureza numérica, uma vez que “o

pesquisador busca classificar, ordenar ou medir as variáveis para apresentar estatísticas, comparar grupos ou estabelecer associações”.

Quanto aos procedimentos técnicos, sua fundamentação teórica está baseada na pesquisa bibliográfica que, segundo Pádua (2012), “é a realizada por meio da identificação, localização e compilação dos dados escritos em livros, artigos de revistas especializadas, publicações de órgãos oficiais, bases de dados, etc.” Utiliza-se a experimentação no desenvolvimento, testes e integração do protótipo ao sistema atual. Segundo Cervo, Bervian e Silva (2007) as técnicas de experimentação se fundamentam no princípio do determinismo, onde “nas mesmas circunstâncias, as mesmas causas produzem os mesmos efeitos”. A obtenção dos dados deu-se pelo método do levantamento uma vez que este “envolve a interrogação direta das pessoas, cujo comportamento se deseja conhecer” (FARIAS FILHO; ARRUDA FILHO, 2013).

Quanto ao local de realização, pode-se classificá-la como pesquisa de campo, considerando que a coleta dos dados acontece no ambiente natural do fenômeno estudado, porém, não se pode esquecer que o protótipo foi desenvolvido em um ambiente controlado (laboratório) apresentando, desta forma, um viés de pesquisa experimental (SEVERINO, 2007).

Quanto à procedência dos dados, entende-se que são dados primários porque “tem como base os dados coletados em ‘primeira mão’ pelo pesquisador, de forma original” (FARIAS FILHO; ARRUDA FILHO, 2013).

Para alcançar os objetivos desta pesquisa, o trabalho foi realizado através das ações especificadas nos capítulos 5, 6 e 7 desta dissertação e apresentados de forma esquemática na Figura 1.

## 1.7 ESTRUTURA DO TRABALHO

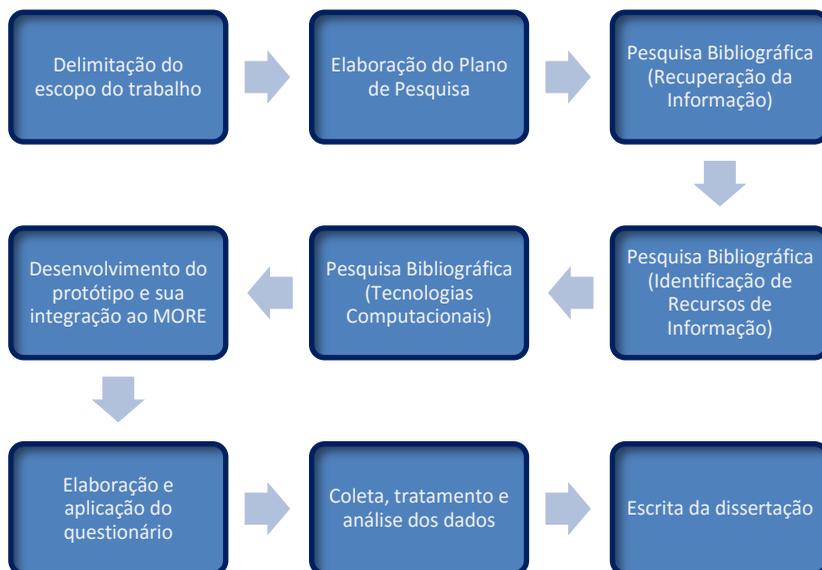
Esta dissertação está organizada em sete capítulos, lista de referências bibliográficas, dois apêndices e um anexo.

O Capítulo 1 faz uma contextualização do tema de forma a emergir a pergunta da pesquisa e seus objetivos, além de apresentar a justificativa, a delimitação, a metodologia e a aderência ao PPGTIC da pesquisa. Finaliza com a estruturação do trabalho.

Os Capítulos 2, 3 e 4 são dedicados a revisão bibliográfica sobre os conhecimentos necessários ao desenvolvimento da pesquisa.

No Capítulo 2 faz um estudo dos conceitos relacionados ao domínio da Recuperação da Informação (RI) e a variação semântica de determinados termos em virtude do contexto em que é empregado, uma

Figura 1 – Sequência de ações desenvolvidas para a realização da dissertação.



Fonte: elaborada pelo autor.

vez que trata-se de uma área do conhecimento eminentemente interdisciplinar.

O Capítulo 3 refere-se a identificação dos recursos de informação de interesse da Biblioteconomia e os esforços realizados no sentido de buscar uma solução aos problemas relativos à organização e representação da informação produzida pela explosão documentária ocorrida a partir de meados do século passado.

O Capítulo 4 dedica-se a abordar as principais tecnologias computacionais empregadas no desenvolvimento, testes e integração do protótipo utilizado.

O Capítulo 5 relata os procedimentos metodológicos empregados na condução das tarefas realizadas no percursos deste trabalho.

No Capítulo 6 traz um detalhamento do desenvolvimento do protótipo.

No Capítulo 7 os dados colhidos através do instrumento de pesquisa são tratados, tabulados e analisados a fim de prover subsídios para responder a pergunta formulada na Seção 1.1.

O Capítulo 8 apresenta as conclusões, baseadas na análise dos dados, e as sugestões para trabalhos futuros, seguido do referencial bibliográfico, apêndices e anexo.

## 2 RECUPERAÇÃO DA INFORMAÇÃO

O presente capítulo é dedicado ao estudo dos conceitos relacionados ao domínio da Recuperação da Informação (RI), bem como busca elucidar, mesmo que com pouca profundidade, a variação semântica de determinados termos em virtude do contexto em que é empregado, uma vez que se trata de uma área do conhecimento eminentemente interdisciplinar, permeando, entre outras áreas do saber, a Ciência da Informação, a Biblioteconomia, a Ciência da Computação e as Tecnologias de Informação e Comunicação.

### 2.1 INFORMAÇÃO

Capurro e Hjørland (2003) exploram as diferentes concepções do conceito de informação de acordo com o entendimento firmado pelos pesquisadores das áreas que o utilizam “dentro de seu próprio contexto e com relação a fenômenos específicos”. Entendem os autores que há uma complexidade crescente na definição do termo, uma vez que este está sendo usado indiscriminadamente e das mais diversas formas. Atribuem ao impacto causado pelas Tecnologias da Informação e Comunicação “sobre as ciências naturais e sociais em particular” uma “noção corriqueira” do significado do termo.

Considera-se que o termo informação é utilizado em dois contextos básicos: como “o ato de moldar a mente” e como “o ato de comunicar conhecimento”.

Numa perspectiva antropológica Capurro (1996) define informação como uma categoria relacionada à ocorrência de mensagens humanas, cujas estruturas estão ligadas ao conceito grego de mensagem e ao discurso filosófico. Nas ciências naturais é entendida como “algo que flui entre um emissor e um receptor”, ou ainda, como a “transmissão de sinais para o conceito da termodinâmica e da evolução biológica” (CAPURRO; HJØRLAND, 2003).

Para as ciências humanas e sociais, a utilização do termo está relacionada a algum dado útil à constituição de conhecimento. Belkin (1978) em suas considerações propõe que o conceito de informação em Ciência da Informação (CI) “deve referir-se, pelo menos, ao domínio específico da CI, como indicado pelos problemas que ela coloca”, o que pressupõe “comunicação humana com objetivo e significado segundo seus próprios requisitos”.

Segundo Saracevic (1999), em um sentido estrito, a informação é “considerada em termos de sinais ou mensagens para decisões que

envolvem pouco ou nenhum processamento cognitivo, ou esse tratamento que pode ser expresso em algoritmos e probabilidades”. E, em um sentido mais amplo, é “tratada como algo que envolve diretamente o processamento cognitivo e compreensão. É o resultado da interação de duas estruturas cognitivas, uma ‘mente’ e (em geral) um ‘texto’. A informação é o que afeta ou altera o estado de uma mente”.

Neste mesmo sentido, Rodrigues e Crippa (2011) consideram que “informação é aceita como aquela que transforma o estado atual de conhecimento de uma pessoa ou coletividade”.

Pontes Junior, Carvalho e Azevedo (2013) argumentam que o conhecimento pode ser entendido “como a alteração cognitiva de um sujeito em contato com a informação, sendo que essa informação pode ser interpretada e é somada com as vivências e experiências desse indivíduo”. Desta forma, é estabelecida uma interdependência entre informação e conhecimento, este se nutre daquela, “que, por sua vez, não tem razão de existir, senão para se transformar em conhecimento”.

Dodebei (2014) apresenta o modelo de transferência da informação, denominado “ciclo da informação”, que reduz a realidade da representação do conhecimento a seis etapas: produção, registro, aquisição, organização, disseminação e assimilação, conforme pode ser visto na Figura 2.

Figura 2 – Ciclo da Informação.



A produção do conhecimento vem acontecendo num sentido horizontal, fruto das abordagens interdisciplinares em que as informações deixam de servir a uma disciplina especificamente, podendo ser extremamente úteis para outras, dependendo do contexto.

O desenvolvimento tecnológico possibilitou uma variedade de suportes para o registro das informações produzidas e estas dispõem da possibilidade de disponibilização das mais variadas formas, entre elas, a impressa (papel) e a digital (*pendrive, online*).

A assimilação das informações diz respeito a utilização desta na construção de novos conhecimentos. A seleção consiste em filtrar os documentos de acordo com os interesses e a necessidade de informações do público alvo. “Assim, a seleção, principalmente, é um processo redutor, e como tal faz uso da comparação como indicador de tomada de decisão para a aquisição”.

A representação documentária busca, através dos mais variados graus de redução semântica, disseminar à sociedade, com economia de recursos, os documentos demandados por sua necessidade de informações. A disseminação da informação está ligada a natureza humana de transferir “informações adquiridas para a geração de novos conhecimentos” (DODEBEI, 2014).

Das argumentações de Van Rijsbergen (1979) pode-se entender que foram desenvolvidas estruturas de informação a fim de abrigarem, “especificamente, uma organização lógica de informações, tais como representantes de documentos, com a finalidade de recuperação da informação”.

## 2.2 CIÊNCIA DA INFORMAÇÃO

A sobrecarga informacional verificada em meados do século passado contribuiu para aumentar consideravelmente a complexidade da tarefa de organizar e representar estes documentos. Neste contexto, reveste-se de grande importância a área do conhecimento coberta pela Ciência da Informação (CI).

A subjetividade dos conceitos trabalhados pela CI contribui para dificultar o estabelecimento de condutas que levem à solução de problemas que se coadunem com os interesses do usuário. “Tais problemas são perceptíveis nos mais variados âmbitos os quais a CI pode vir a atuar – biblioteca, arquivo, centro de documentação, museu, etc.” (TEIXEIRA, 2010).

Neste mesmo sentido, Rodrigues e Crippa (2011) entendem que apesar de um dos principais objetivos da CI ser o de “facilitar o acesso

rápido e eficaz, para os usuários de Sistemas de Recuperação da Informação (SRI), às informações que tais sujeitos julguem ser relevantes”, a subjetividade com que é tratada a definição de informação, conforme visto anteriormente, e o conceito de relevância é traduzido em dificuldades na obtenção plena deste desiderato.

Na citação, a seguir, de Saracevic (1999, tradução do autor) podemos observar a preocupação do autor com a CI:

Ciência da informação tem três características gerais que são o principal motivo de sua evolução e da sua existência. Estas são compartilhadas com diversos campos modernos. Elas também podem ser vistas como áreas problemáticas com as quais a ciência da informação tem de lidar em um nível geral.

a) Em primeiro lugar, a ciência da informação é interdisciplinar por natureza; no entanto, as relações com várias disciplinas estão mudando. A evolução interdisciplinar está longe de terminar.

b) Em segundo lugar, a ciência da informação está inexoravelmente ligada à tecnologia da informação. Um imperativo tecnológico convincente e restringindo a evolução da ciência da informação, assim como a evolução de uma série de outros campos, e, além disso, da sociedade da informação como um todo.

c) Em terceiro lugar, a ciência da informação é, com muitos outros campos, um participante ativo na evolução da sociedade da informação. Ciência da informação tem uma forte dimensão social e humana, acima e além de tecnologia. (SARACEVIC, 1999).

Segundo Carvalho, Lucas e Gonçalves (2010) o fato da Ciência da Informação estudar, pesquisar e propor novos métodos para organizar e representar o conhecimento, presta sua contribuição ao refinamento dos sistemas que usam recursos computacionais no que tange as tecnologias aplicadas na recuperação da informação. Assim, concede ao usuário a possibilidade de colaborar na organização do conhecimento, criando um ambiente favorável aos seguintes questionamentos:

- no tocante a adequação dos processos atuais de catalogação e indexação visando a disponibilidade destas informações para posterior recuperação;

- relativo à linguagem natural e sua aplicação em documentos disponibilizados na rede mundial de computadores (*Web*); e,
- referente a colaboração na construção do conhecimento que ocorre na atualidade e sua importância na sociedade atual.

No entendimento de Saracevic (1999) a CI foi pioneira no desenvolvimento de “processos e Sistemas de Recuperação da Informação (SRI) na década de 1950, e foi definido como seu principal objetivo a recuperação de informações relevantes”.

Considera Belkin (1978) que no momento em que as redes de computadores começavam a influenciar o comportamento social e a permitir maior agilidade da comunicação das informações científicas, os problemas advindos desta nova era implicaram então, pelo menos, no seguinte conjunto de preocupações para a ciência da informação:

- com informações humanas, sistemas de comunicação cognitivas;
- com a relação entre informação e gerador;
- com a relação entre informação e usuários;
- com a idéia de informação desejada;
- com a eficácia das informações e a eficácia da transferência da informação.

Para Teixeira (2010) sistema de informação pode ser definido, no contexto da CI, como um conjunto integrado por “rotinas de processamento de informação, registros informacionais e usuários, com o objetivo de permitir a manipulação da informação contida nos registros pelos usuários para atender às suas necessidades”.

### 2.3 REPRESENTAÇÃO DOCUMENTÁRIA

A explosão documental gerada pelo processo de especialização do saber, ocorrida em meados do século passado estimulou “estudos e experiências sobre a construção de linguagens artificiais que pudessem representar o universo conceitual de campos específicos do conhecimento”. Os pesquisadores e profissionais ligados à representação desses conteúdos informativos passam a estudar essas linguagens, as quais são chamadas de linguagens documentárias, com a finalidade “de recuperação e disseminação da informação, tradução automática, análise de conteúdo, análises literárias, semióticas e linguísticas, realizadas manualmente ou de forma automática”. Entretanto, o viés interdisciplinar da construção dessas linguagens, apesar de serem extremamente eficazes na função de recuperação da informação, apresenta-se como tarefa de custo muito elevado (DODEBEI, 2014).

Segundo Sousa e Araújo Júnior (2013) as linguagens documentárias situam-se no âmbito da representação documentária e são empregadas para representar semanticamente o conteúdo dos documentos de modo a reduzir a utilização de recursos através de uma linguagem artificial, “a fim de recuperar os documentos pertinentes em resposta às consultas que tratam deste conteúdo, que por sua vez é operacionalizada por meio da análise documentária”.

Comentam, ainda, os referidos autores, que sistemas de informação ou unidades de informação frequentemente utilizam as linguagens documentárias “para descrever o conteúdo dos documentos, a fim de prepará-los para o armazenamento e a posterior recuperação das informações que contêm”. Através dessas linguagens documentárias pode-se estabelecer “relações entre os termos e os conteúdos de que tratam os documentos, espelhando suas funções/atividades e conduzindo os usuários dos sistemas de arquivo à recuperação da informação” (SOUSA; ARAÚJO JÚNIOR, 2013).

No entendimento de Strehl (2011) a integração entre os elementos e termos da uma linguagem artificial especializada frente a linguagem dos usuários, também denominada de linguagem natural, é o pressuposto de consistência desta linguagem para representação documentária. Apesar do nível de desenvolvimento alcançado pelos atuais algoritmos de processamento da avaliação automática da relevância dos documentos e, desta maneira, propiciado o aumento de pontos de acesso a um determinado documento, em contrapartida abre maiores probabilidades de acontecerem equívocos nas associações ou nas relações entre os termos das linguagens documentária e natural.

A representação das etapas constituintes do processo de indexação (manual ou automática), além de seus conceitos, através da utilização de linguagem natural nas linguagens documentárias, mostra “que a complexidade dos recursos de representação da informação está intimamente ligada aos requisitos necessários para garantia da qualidade dos sistemas de recuperação (STREHL, 2011).

Segundo Oliveira e Araujo (2012) o fato das linguagens documentárias atuarem no processo de representação, como instrumentos auxiliares, elas viabilizam o estabelecimento de comunicação entre o sistema e o usuário.

Na obra de Dodebei (2014) verifica-se que as linguagens documentárias podem ser concebidas como metarrepresentações ou representações documentárias que, no contexto das representações sociais, atuam como linguagens de comunicação no estabelecimento da relação entre a informação documentária e a necessidade do usuário.

Estas representações documentárias encontram-se “ao lado de outras formas de representação da informação, como resumos, catálogos, bibliografias, índices, inventários, repertórios”.

Ainda, segundo Dodebei (2014), são de competência das linguagens documentárias as seguintes funções:

- organizar o campo conceitual da representação documentária;
- servir de instrumento para a distribuição útil dos livros ou documentos;
- controlar as dispersões léxicas, sintáticas e simbólicas no processo de análise documentária.

A representação do conteúdo de um documento, através de termos apropriados, pelos indexadores é facilitada pelo uso de linguagens documentárias, cuja função primordial, de acordo com Martins e Carvalho (2014), “é organizar a informação especializada e o da indexação é traduzir a linguagem natural do pesquisador em linguagem documentária por meio da representação do conceito”.

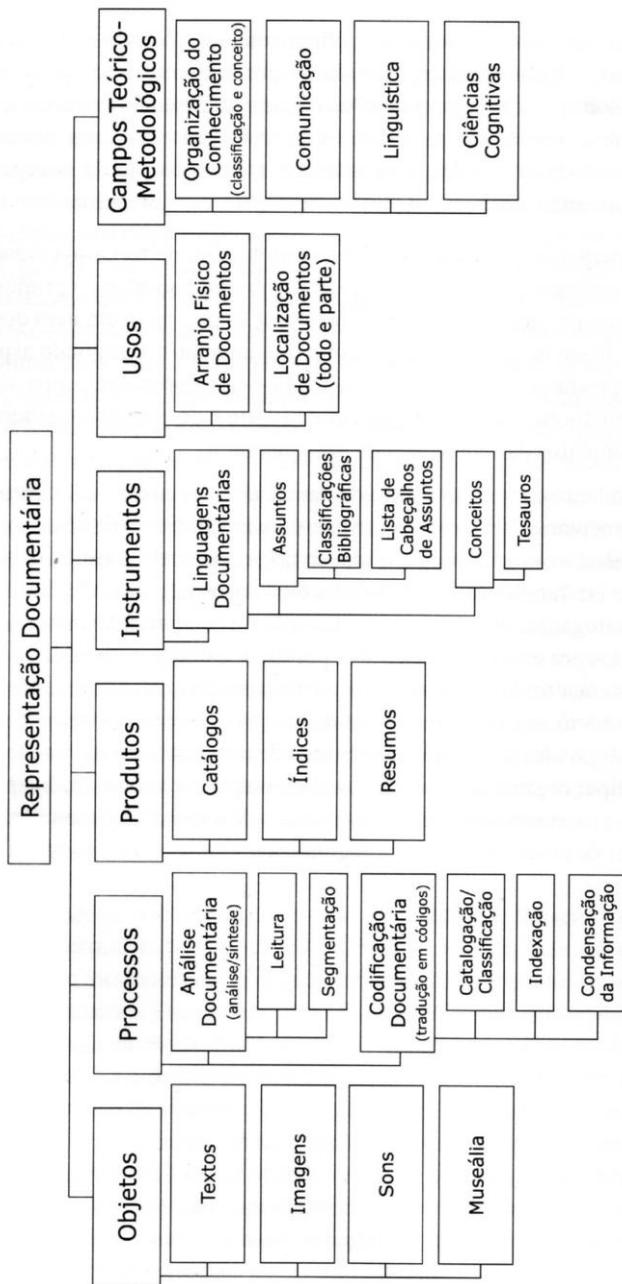
Araújo (2012) faz uma abordagem direcionada ao ambiente de biblioteca onde esses vocabulários controlados são utilizados pelos bibliotecários, como ferramenta, na identificação de termos que representem o conteúdo semântico do documento. Considera que o catálogo eletrônico é o meio pelo qual o usuário recupera a informação que necessita. Assim, “nas duas situações, caracterizadas pelos processos de indexação e de busca e recuperação da informação, a linguagem especializada é a fonte de informação”.

Para Van Rijsbergen (1979) o vocabulário pode, também, ser não controlado, o qual não requer algum tipo de restrição mais rigorosa. No entanto, o vocabulário controlado refere-se a uma lista de termos de índice aprovados que um indexador pode usar e os controles sobre a linguagem pode incluir relações hierárquicas entre os termos de índice. Pode-se, ainda, definir que certos termos só podem ser usados como adjetivos (ou qualificadores), não havendo nenhum limite para qualquer tipo de controle sintático.

A Figura 3 mostra as possibilidades conceituais, não todas, do domínio da representação documentária, para uma visão geral destas relações.

O desenvolvimento, e constante aprimoramento, das Tecnologias da Informação e Comunicação aliado aos recursos computacionais (*hardware* e *software*) disponíveis permitem executar com maior agilidade, simplicidade e eficiência os processos de classificação, indexação, armazenamento e recuperação das informações (TEIXEIRA; SCHIEL, (1997).

Figura 3 – Gráfico Conceitual da Representação Documentária.



Fonte: Dodebei (2014, p. 44).

Na concepção de Araújo Júnior e Tarapanoff (2006) “a análise do documento é uma significativa contribuição para a comunicação e o fluxo da informação em qualquer organização e para qualquer sistema de recuperação da informação”, uma vez que estes documentos em linguagem natural, denominados bases textuais, não estruturam seus conteúdos em um formato padronizado, tal como pode ser observado nas bases de dados.

Na teoria da classificação a tarefa de agrupar (indivíduos e/ou coisas) em grupo segundo suas propriedades semelhantes e/ou dessemelhantes denomina-se classificar. Sousa e Araújo Júnior (2013) consideram

a classificação como uma função matricial, pois é por meio dela que se realiza a avaliação, que em uma etapa posterior, vai possibilitar a gestão dos prazos de guarda e da destinação final (tabela de temporalidade), independentemente da metodologia aplicada e de sua configuração, sempre tem como ponto de partida os conjuntos documentais definidos na classificação.

Classificar o conhecimento da humanidade foi, e tem sido, requisito para organizá-lo visando sua recuperação em tempo futuro. Para acompanhar o incremento da produção do conhecimento foram desenvolvidos métodos, técnicas e sistemas que almejam suprir os anseios referentes à recuperação da informação que é fortemente afetada pela explosão documentária (CARVALHO; LUCAS; GONÇALVES, 2010; TEIXEIRA; SCHIEL, 1997; OLIVEIRA; ARAUJO, 2012).

Na concepção de Ribeiro (2013) a classificação é tida “como uma operação intelectual e técnica, que se traduz numa categorização/sistematização para fins organizativos e numa representação formal tendo em vista a recuperação da informação”. Neste sentido, percebe-se que as bibliotecas são pioneiras na valorização do acesso à informação, cuja assertiva advém da preocupação desta com a criação de dispositivos classificatórios em benefício da representação e da recuperação da informação (TEIXEIRA; SCHIEL, 1997).

Reafirmam Pontes Junior, Carvalho e Azevedo (2013) a necessidade premente, em meados do século passado, de organização da informação e do conhecimento, citando os sistemas de classificação como uma das primeiras tentativas, que se tem registro, no sentido de encontrar uma solução para o problema que se acentuou logo após a segunda guerra mundial. Nesse período, fruto desse esforço, surgem

vários sistemas com a finalidade de contribuir para a organização deste volume de documentos. Dentre estes sistemas pode-se citar o Código de Classificação Universal (CDU) e a Classificação Universal de Dewey (CDD).

Para Okada e Ortega (2009) a efetividade e a eficiência na recuperação de informações estão ligadas a tarefa de classificação, uma vez que “não há habilidade de busca que supere modos inconsistentes de organização da informação”.

No contexto da representação documentária, outro processo de codificação documentária que se pode verificar é a catalogação. Okada e Ortega (2009) a define “como o processo de descrição dos elementos informativos que permitem identificar um documento e de indicação dos pontos de acesso que remetam a estes elementos”. No entendimento das autoras “as regras de catalogação são concebidas a partir de certos princípios que objetivam a identificação de um documento de forma unívoca”. Além disso, se o registro for manipulado através de recursos digitais, se faz necessária a adoção de código de catalogação em formato que facilite sua manutenção, bem como, ofereça economia de recursos computacionais na realização de tarefas com estes registros, tais como a recuperação e a atualização, se for o caso.

Em relação aos pontos de acesso recomenda-se um controle rigoroso a fim de evitar redundâncias e incertezas semânticas na informação, os quais devem ser selecionados, de acordo com as normas em vigor, por profissionais especializados.

Segundo Lancaster (1993) apud Strehl (2011) “o ponto de acesso pode ser definido como o meio pelo qual um item bibliográfico é recuperado, no momento da realização de uma busca”. Este conceito é de fundamental importância no momento de avaliar algum sistema de recuperação da informação, uma vez que a maximização da omissão de itens não úteis (precisão), quanto a identificação dos itens úteis (revocação), dependem da qualidade dos pontos de acesso.

Um fator que contribui para aumentar a complexidade dos processos de organização da informação, para posterior recuperação, é a utilização da *Web*, uma vez que este novo cenário restringe e, até mesmo, abdica da intermediação de profissionais especializados transferindo esta tarefa para um ambiente de construção social onde “a organização da informação é realizada pelos próprios usuários produtores e consumidores da informação” (CARVALHO; LUCAS; GONÇALVES, 2010).

Quando se adentra ao domínio da organização e representação da informação para posterior recuperação precisa-se ter em mente a

importância dos termos “precisão” e “revocação” para este ramo do conhecimento.

Sousa e Araújo Júnior (2013) ressaltam a importância do conceito de precisão para a avaliação da qualidade da recuperação da informação uma vez que precisão “pode ser entendida como a extensão com a qual itens recuperados em uma base de dados são considerados úteis por quem demandou a pesquisa”.

Neste mesmo sentido e ampliando a definição anterior, Araújo Júnior e Tarapanoff (2006) entendem que

a precisão não se dá *per se*, mas no contexto em que operam a revocação, a exaustividade e a especificidade e, sobretudo, tendo como ponto de equilíbrio o usuário que vai definir, em nome da sua necessidade de informação, o que é útil ou inútil dentre toda a informação recuperada.

De acordo com Teixeira (2010), o índice de precisão está relacionado à razão entre os documentos recuperados considerados úteis pelo demandante da informação e o conjunto total de documentos recuperados. E o índice de revocação está relacionado à razão entre os documentos recuperados considerados úteis pelo demandante da informação e o total de documentos do *corpus* que poderiam ser assim considerados por este usuário.

No entendimento de Ramos e Munhoz (2011) o índice de revocação é inversamente proporcional ao índice de relevância dos itens recuperados, uma vez que quanto mais próximo do total de documentos relevantes do *corpus* for a recuperação maior ainda (no sentido de uma relação não linear) será o número de documentos não relevantes que serão recuperados.

No entendimento de Lancaster (2004) verifica-se que o desejo de alguém que faz uma busca é encontrar documentos úteis, e preferencialmente somente estes, para satisfazer sua necessidade de informação, evitando a recuperação de itens inúteis. Entende o autor que revocação é a capacidade de recuperar documentos úteis, e precisão é a capacidade de evitar documentos inúteis.

Verificando a história dos Sistemas de Recuperação da Informação (SRI) encontra-se a busca incessante por maiores índices de precisão e revocação através de técnicas desenvolvidas para examinar documentos, fazer identificação conceitual e traduzir termos para linguagens de indexação.

Em uma tentativa de resolução desse problema, uma outra forma de tradução de conceitos foi desenvolvida, a indexação por atribuição. Por atribuição, os termos que representam os assuntos tratados são oriundos de uma fonte que não é o próprio documento, mas, sim, de uma linguagem documentária, frequentemente, de algum tipo de vocabulário controlado ou tesauro. (STREHL, 2011).

Nas considerações de Martins e Carvalho (2014) é incontestável que a qualidade da indexação pode interferir na eficácia do processo de busca nas bases de dados e que esta complexa tarefa é influenciada pelo conhecimento tácito e pela ontogenia do indexador (ou do arquiteto do software) “para que o documento seja bem representado nas bases de dados”.

Para Oliveira e Araujo (2012) a indexação pode ser entendida como “o processo de compreensão da essência do conteúdo de um documento e representação deste num modo de expressão em linguagem de indexação”.

O foco de um processo de indexação é destacar toda a informação que possa suprir uma necessidade informacional do usuário, com o desiderato de armazená-la para posterior recuperação por quem demandá-la. Assim, verifica-se que as necessidades pessoais determinam que informações serão construídas e estas o farão segundo suas habilidades, inclusive as cognitivas.

A elaboração de critérios de indexação, levando em consideração a realidade do público alvo, contribui para fazer uma previsão da recuperação da informação, melhorando os índices de precisão e recuperação da informação demandada (RAMOS; MUNHOZ, 2011).

O trabalho do especialista em informação, ao qual concerne a responsabilidade de indexador, reveste-se de especial importância porque precisa compreender a demanda para estabelecer a relação com o material armazenado na base de dados a fim de que possa ser oferecido ao usuário uma resposta que supra sua necessidade informacional. No entanto, a qualidade da base de dados e dos termos autorizados no vocabulário influenciam no trabalho de indexação que provê a alimentação desta base. Manter atualizado o conhecimento das necessidades do público alvo para o aperfeiçoamento do processo de indexação para “proporcionar subsídios para a determinação dos

requisitos a serem utilizados no âmbito do gerenciamento estratégico da informação” (ARAÚJO JÚNIOR; TARAPANOFF, 2006).

Ampliando os entendimentos anteriores, Martins e Carvalho (2014) salientam que a eficiência na disseminação da informação mantém estreita relação com a qualidade do trabalho do indexador, bem como, a qualidade das bases de dados, uma vez que estas comportam-se como instrumentos de divulgação científica contribuindo para a evolução da ciência. A ontogenia do indexador reveste-se de vital importância para uma boa representação do documento nas bases de dados. Consideram, ainda, que apesar do processo de indexação ser subjetivo interfere na eficácia do resultado da busca realizada pelos pesquisadores, lembrando que “um documento deve ser bem analisado e representado para que os usuários finais possam se apropriar desses conhecimentos, continuando a produzir informações científicas”.

Van Rijsbergen (1979) argumenta que o indexador esforça-se para antecipar quais os termos do índice um usuário estaria disposto a empregar com a intenção de recuperar o documento cujo conteúdo ele está prestes a descrever. Implicitamente ele está construindo consultas para as quais o documento é relevante. Na indexação automática, presume-se que o texto de um documento ou de uma consulta passa automaticamente pelo mesmo processo de análise, cuja saída será uma representação do conteúdo, e, se o documento é relevante para a consulta, será mostrada por um processo computacional.

Ressalta que para um computador estabelecer a relevância de um documento para uma consulta se faz necessária a construção de um modelo que possa quantificar as decisões de relevância.

A indexação não constitui um fim em si mesma. Define-se de modo muito pragmático a ‘boa indexação’ como a indexação que permite que se recuperem itens de uma base de dados durante buscas para as quais sejam respostas úteis, e que impede que sejam recuperados quando não sejam respostas úteis (LANCASTER, 2004).

Segundo Dodebei (2014) o princípio da indexação da informação e da construção de resumos para facilitar a recuperação da informação contou com as regras de análise documentária criadas para a organização do ‘Biblion’ ou Livro Universal, ou seja, a literatura deve ser analisada, isolando-se quatro categorias gerais: ‘fatos’; ‘interpretação dos fatos’; ‘estatísticas’; ‘fontes’.

Para Barité et al. (2010) a documentação cumpre o papel de sintetizar o conhecimento científico e especializado para que, através de metodologias padronizadas, o profissional da informação possa “reapresentar esse conhecimento em sistemas de organização do conhecimento de diversas naturezas: sistemas de classificação, tesouros, listas de cabeçalhos de assunto, taxonomias, ontologias temáticas, etc”.

Oriunda da Biologia, a taxonomia inicialmente foi utilizada na classificação de seres vivos e, posteriormente, foi introduzida no meio digital para contribuir na organização das informações disponíveis em portais corporativos e de busca baseados em diretórios, uma vez que possuem como elementos-chave uma estrutura hierárquica fundamentada em etiquetas. O fato das taxonomias constituírem-se em “um sistema de classificação de conceitos para ambiente *Web*, não se pode descartar as diferenças que esse ambiente possui em relação aos outros sistemas, como tesouros, a própria classificação e as ontologias” (PONTES JUNIOR; CARVALHO; AZEVEDO, 2013).

A taxonomia, segundo Sousa e Araújo Júnior (2013), “pode ser definida como sistema de classificação que apoia o acesso à informação, permitindo classificar, alocar, recuperar e comunicar informações em um sistema de maneira lógica” cuja estrutura apresenta os termos organizados em camadas, utilizando-se de uma hierarquia simples e flexível para adequar-se às necessidades de classificação e indexação dos documentos do *corpus*.

A Ciência da Informação, conforme Dodebei (2014), começou a fazer uso do termo tesouro a partir da década de 40 do século passado, com mais ênfase nos processos de recuperação da informação, valendo-se da sua capacidade de transportar conceitos e suas relações mútuas. A autora considera que

o emprego dos tesouros nas tarefas de indexação e recuperação de informações tenta resolver o problema da alocação de documentos em classes de assuntos, não só por sua capacidade de controlar o vocabulário, mas porque é um instrumento que relaciona os descritores/termos de forma mais consistente, apresentando uma estrutura sintética simplificada e uma complexa rede de referências cruzadas. Isto permite ao especialista localizar com mais facilidade a palavra-chave para uma busca (DODEBEI, 2014).

O avanço da ciência e da tecnologia configura-se como impulsionador do uso dos tesouros, em virtude da necessidade, cada vez maior, de dispor de artifícios que viabilizem a recuperação da informação especializada. Para Oliveira e Araujo (2012) a fragmentação do conhecimento, fruto de um maior grau de especialização, levou à adoção dos tesouros em outros campos do saber. Destacam como funções dos tesouros o controle de sinônimos e quase sinônimos e a distinção de homógrafos. Entendem, ainda, que os tesouros “são instrumentos imprescindíveis à sistematização e disseminação do conhecimento humano, uma vez que a fragmentação das disciplinas assim o exigem”.

Utilizando-se da linguagem documentária, Pontes Junior, Carvalho e Azevedo (2013) afirmam que, o tesouro organiza as informações e é normalmente empregado na indexação de documentos referentes a assuntos e áreas específicos.

No que concerne às ontologias, Fachin (2009) argumenta que descrevem categorias em um determinado domínio do conhecimento. Sua criação dá-se “a partir do tratamento de cada conteúdo a ser inserido nesse segmento, obedecendo aos critérios linguísticos, semânticos e representativos”. Sua utilização pode ser vinculada ao compartilhamento do conhecimento valendo-se do uso do vocabulário, da semântica e dos relacionamentos similares entre conceitos de um domínio.

## 2.4 RECUPERAÇÃO DA INFORMAÇÃO PROPRIAMENTE DITA

A história da humanidade e a evolução da ciência se confundem com a necessidade de recuperação e disponibilização das informações, que armazenadas, em algum meio físico, possam ser resgatadas de forma precisa, rápida e com certo grau de relevância (FACHIN, 2009).

A contribuição da Biblioteconomia, conforme Oliveira e Araújo (2012), remonta a antiguidade quando as Bibliotecas de Papiro na Alexandria, de Terracota na Babilônia e a de Pergaminho em Pérgamo deram origem aos Sistemas de Informações. Para se adaptarem às necessidades decorrentes de cada época estes sistemas passaram por profundas mudanças até sua concepção atual, de modo a satisfazer as peculiaridades advindas com os ambientes virtuais, onde muitos já consideram o contexto social e o histórico do usuário.

Num entendimento inicial, Recuperação da Informação (RI) é o processo, manual ou automático (digital), pelo qual retorna-se ao solicitante de uma pergunta formulada a informação armazenada que satisfaça sua necessidade informacional.

O principal objetivo de um sistema de recuperação de informação (SRI) é fornecer rapidamente aos usuários a informação que procuram. A complexidade em recuperar somente aqueles documentos que são importantes para o usuário é um dos principais obstáculos a ser contornado pelos Sistemas de Recuperação da Informação (SRI) (FACHIN, 2009; TEIXEIRA, 2010; RODRIGUES; CRIPPA, 2011; WU et al., 2013; )

Segundo Lopes (2002), no contexto da recuperação da informação, “a estratégia de busca pode ser definida como uma técnica ou conjunto de regras para tornar possível o encontro entre uma pergunta formulada e a informação armazenada em uma base de dados”. Ou seja, selecionar e retornar ao interessado um conjunto de documentos (informações) que compõe a resposta à pergunta submetida à apreciação.

Van Rijsbergen (1979) pondera que as distinções que ocorrem entre os diversos tipos de estratégias de busca, frequentemente, podem ser compreendidas através da linguagem de consulta, que é a língua em que a necessidade de informação é expressa, uma vez que a natureza da estratégia de busca pode revelar a natureza da linguagem de consulta. Um exemplo é a pesquisa booleana cuja consulta, normalmente, apresenta termos característicos de combinações lógicas de palavras-chave.

Recuperação da Informação (RI) foi assim definida por Manning, Raghavan e Schütze (2009, tradução nossa):

Recuperação da Informação é encontrar o material (geralmente documentos) de natureza não estruturada (geralmente texto), que satisfaz uma necessidade de informação, dentro de grandes coleções (geralmente armazenados em computadores).

Na comparação com a recuperação de dados verifica-se que a RI proporciona um campo de atuação mais amplo e complexo, uma vez que considera o contexto na atribuição da relevância para recuperar um item em contrapartida a consulta realizada, além de buscar, continuamente, a aproximação da necessidade informacional formulada pelo usuário com os descritores do conjunto de informações obtidas como resultado de uma consulta realizada (TEIXEIRA; 2010).

Para Xavier (2009) a RI é um campo de pesquisa que se preocupa com o processo de busca, a exploração e a descoberta de informação em

bases de dados estruturadas de modo a entregar ao usuário as informações que satisfaçam suas necessidades informacionais, ou seja, “é prover a usuários resultados precisos, relevantes e confiáveis”.

No entendimento de Fachin (2009) o grau de desenvolvimento alcançado pelas Tecnologias da Informação e Comunicação (TICs), nas décadas de 80 e 90, possibilitou o avanço na área da Computação, Biblioteconomia e da Ciência da Informação. A autora, ainda, define RI nestes termos:

recuperação da informação é tornar possível e concreto o encontro entre uma pergunta formulada, a informação armazenada e o retorno positivo ao usuário solicitante, quer de forma manual ou automatizada/digital. Em sistemas digitais, os usuários recuperam informações através de *browsing* (navegação através de *links* de um documento ou de um *site*) ou através de *searching* (consulta a um banco de dados).

A linha do tempo da RI foi fracionada por Saracevic (1999) em três fases distintas. A primeira e a ideia original, emergiu em meados do século XX, onde a recuperação da informação trabalha com a lógica formal no processamento da informação. A segunda, surgiu em seguida, considera a relevância e procura orientar seu processo às necessidades de informação e avaliações dos usuários. A terceira teve lugar somente duas décadas depois e é fundamentada na interação, permitindo o intercâmbio direto e o *feedback* entre sistemas e pessoas envolvidas em processos de RI.

Martins e Lima (2013) consideram a Recuperação da Informação sob três abordagens distintas: fisicista, cognitivista e social. A fisicista, com origem nas décadas de 50 e 60, tinha como única preocupação os métodos e processos puramente técnicos. A abordagem cognitivista, adotada na década de 70, reorienta a concepção dos Sistemas de Recuperação da Informação para que passam a considerar o comportamento do usuário sem levar em conta o seu contexto social. A década de 90 foi acompanhada de uma conscientização para o que se chama de abordagem social “que considera os usuários como seres humanos, compostos não só por racionalidade e estruturas cognitivas, mas também por desejos, emoções, interações sociais e culturais e fatores históricos”.

Apesar da recuperação da informação englobar variados aspectos, Van Rijsbergen (1979) destaca três áreas de pesquisa, consideradas

importantes, uma vez que elas respondem por boa parte do assunto. São elas: a análise de conteúdo, preocupa-se com a representação do conteúdo informacional dos documentos de maneira adequada ao processamento por computador; as estruturas de informação, explora as relações entre os documentos a fim de melhorar a eficiência e a eficácia das estratégias de recuperação; e, avaliação, preocupa-se com a medição da eficácia da recuperação.

O crescimento exponencial do volume de informação levou a RI estabelecer-se como um ramo do conhecimento científico que busca amenizar três problemas dele advindos: representação da informação, especificação da busca da informação e criação de mecanismo para recuperação. Ressalta-se que a RI transita, de forma interdisciplinar, por diversos domínios, desde a Ciência da Informação até Ciência da Computação, dispondo de tarefas e ferramentas de organização e recuperação da informação e conhecimento, como: classificação, tesauros, taxonomia e ontologias, entre outras (PONTES JUNIOR; CARVALHO; AZEVEDO, 2013).

No entendimento de Teixeira (2005) este crescimento acelerado do volume de informações geradas e, por consequência, consumidas conduziram ao emprego da Tecnologia da Informação como uma das principais ferramentas no incremento da qualidade e da produtividade dos Sistemas de Recuperação da Informação (SRI), os quais devem “atender às necessidades específicas dos usuários, permitindo ao máximo o acesso a informações relevantes”.

Conforme Souza (2006) *apud* Araujo (2012), as funções de um SRI são as seguintes:

- representar as informações contidas nos documentos e expressas pelos processos de indexação e descrição dos documentos;
- armazenar e gerir física e/ou logicamente esses documentos e suas representações; e
- recuperar as informações ali contidas e os documentos armazenados no sistema.

Sejam esses sistemas de RI, para a pesquisa em escala *Web* ou a um nível empresarial, o impacto na vida diária é grande, no entanto, raramente é medido. Normalmente, sua avaliação é focada em seu motor de busca, medindo o quão rápido ele pode responder a uma consulta, ou o nível de relevância da informação recuperada.

As especificidades de busca de cada organização tem norteado a construção de mecanismos de recuperação inteligente da informação. Fachin (2009) e Strehl (2011) afirmam que a utilização de agentes inteligentes favorecem o desenvolvimento de SRI que possa atender

estas especificidades de acordo com o público alvo, e que cabe aos “criadores investigar, analisar e utilizar estes recursos”.

Neste sentido, ainda, Weikum et al. (2009) apontam para a convergência, tanto da perspectiva da recuperação da informação quanto da perspectiva dos bancos de dados, na utilização crescente de dados estruturados e semiestruturados. Esclarecem que técnicas de extração de informações linguísticas aplicadas à base de aprendizagem (o universo documental em linguagem natural) tem contribuído para ampliar as fontes textuais com registros estruturados e permitido a realização de consultas, nos moldes realizados pelos bancos de dados, sobre dados originalmente não estruturados.

A *Web 2.0*<sup>2</sup> e sua interatividade proporcionou a possibilidade aos usuários, através dos mais diversos aplicativos, contribuir para o aumento do volume de informações armazenadas, bem como, da quantidade de interessados na sua recuperação. Ghorab et al. (2012) observam que a maioria dos SRI atuais não considera as características do usuário que realiza a consulta, retornando o mesmo conjunto de informações para usuários diferentes que realizem a mesma consulta. Entendem os referidos autores que estes sistemas deveriam considerar as características do usuário no desenvolvimento destes sistemas, os quais são denominados *Personal Information Retrieval* (PIR). Isso pode ser feito mantendo o controle de informações e interesses pessoais do usuário e, em seguida, usar essas informações para personalizar a consulta ou o conjunto de resultados apresentados.

O estudo realizado por Peltonen e Lin (2014) analisa o custo de uma tarefa de recuperação de informação baseada nos metadados a fim de levantar o grau de similaridade entre os documentos recuperados, classificá-los, agrupá-los e mostrá-los através de uma interface gráfica.

Nas considerações de Ramos e Munhoz (2011) um sistema de busca considerado “ideal” deverá dispor de funcionalidades que permitam a recuperação por partes de palavras (como em maq costu para máquina de costura, por exemplo), por sinônimos (chave de luz para interruptor, por exemplo), por parte alternada do todo e pela semântica da expressão fornecida pelo usuário.

---

<sup>2</sup> *Web 2.0* é o termo utilizado para referir-se à segunda geração da *World Wide Web* (*www*) e às suas funcionalidades, tais como: interatividade, interoperabilidade e dinamicidade, entre outras. Possibilitou o surgimento dos ambientes colaborativos e interativos a exemplo das *wikis*, das redes sociais e dos blogs.

Ainda segundo Araujo (2012) a função principal de um SRI “é dispor de informações contidas nos documentos indexados, a partir de uma descrição sintética, objetiva e representativa de seu conteúdo formal e temático”. Teixeira e Schiel (1997) corroboram que estes sistemas integram vários processos, tais como: seleção, aquisição, indexação, busca e recuperação das informações.

Conforme Souza (2006) *apud* Araújo (2012), as funções de um SRI são as seguintes:

- 1) representação do conteúdo informacional existente nos documentos, normalmente através de alguma linguagem documentária que possa descrevê-los;
- 2) gerenciamento e armazenamento desses documentos e de suas representações; e
- 3) recuperação das informações representadas e dos próprios documentos gerenciados pelo sistema.

A relação existente entre SRI e usuários vão além do processo de busca e recuperação da informação, devendo assemelhar-se àquela que ocorre entre este e o bibliotecário durante o processo de indexação.

As evoluções dos SRI ocorridas nas mais diversas situações de tempo, espaço e necessidades organizacionais específicas, favoreceram o desenvolvimento de vários métodos de busca, dentre os quais destacamos o booleano, o espaço vetorial, o probabilístico, o *Clustering* e o *Feedback*. E para suprir as peculiaridades da *Web Semântica*, os modelos difuso, booleano estendido, especial vetorial generalizado, indexação semântica latente, redes neurais e recuperação textual estruturada (FACHIN, 2009).

O Modelo Booleano pode ser entendido como um conjunto de termos de uma consulta ligados por operadores booleanos de disjunção, conjunção ou negação (*‘and’*, *‘not’* ou *‘or’*), recuperando apenas os documentos considerados ‘verdadeiros’ em função da expressão lógica da consulta (CARVALHO; LUCAS; GONÇALVES, 2010).

Segundo os ensinamentos de Manning, Raghavan e Schütze (2009) o Modelo Espaço Vetorial sugere um espaço no qual os termos de um documento e os termos da consulta são vetorizados de modo que o grau de similaridade entre eles seja calculado através de equações matemáticas e o conjunto de resultados ordenado de acordo com este grau de similaridade.

Trata-se de um modelo estatístico e multidimensional, onde cada termo representa uma dimensão e pode ter associado a ele um peso para refletir sua importância, tanto no documento quanto na consulta. Assim, o ângulo entre esses vetores determina o grau de similaridade entre eles,

ou seja, o grau de similaridade é inversamente proporcional ao valor do ângulo formado entre eles.

Manning, Raghavan e Schütze (2009) elucidam, ainda, o fato de que documentos com os mesmos termos podem estar localizados em uma mesma região do espaço vetorial e, conseqüentemente, possuem conteúdos similares.

Convém destacar que na fase de transformação dos dados, considerando as restrições de recursos computacionais e o objetivo principal da RI, estes são armazenados em um índice invertido que contém os termos e a frequência destes no documento e no *corpus* para facilitar sua localização e recuperação.

O Modelo Probabilístico fundamenta-se no princípio da ordenação probabilística para ordenar os documentos recuperados segundo a relevância calculada dinamicamente com fulcro nas consultas dos usuários (FACHIN, 2009).

O Modelo *Clustering* baseia-se na formação de agrupamentos de documentos que apresentam algum aspecto similar e esta classificação ocorre segundo algoritmos específicos do sistema. No entanto, a classificação pode não ser ditada por assuntos ou por áreas do conhecimento (FACHIN, 2009).

O *feedback* é baseado na realimentação da informação a partir da saída, a fim de utilizá-la para melhorar o desempenho na entrada seguinte (VAN RIJSBERGEN, 1979).

As técnicas de avaliação, segundo Wu et al. (2013), se dividem em duas categorias principais: orientada para o sistema, centrada na recuperação eficaz e eficiente do motor de busca de um SRI; e orientada para o utilizador, voltada para satisfação do usuário com o sistema, mede fatores como o nível da ajuda prestada pelo sistema ao usuário no sentido deste alcançar e perceber que está alcançando seus objetivos.

Ao encerrar este capítulo vale ressaltar a importância e o potencial da *Web*, para esta sociedade dita da informação e do conhecimento, como a maior e mais abrangente base de conhecimento do mundo, porém ainda apresenta desafios na exploração dessas potencialidades. Assim, Weikum et al. (2009) colocam como grande desafio a construção de modelos que possam extrair os fatos importantes a partir da *Web* e organizá-los em uma base de conhecimento explícito que captura entidades e relações semânticas entre eles.



### 3 IDENTIFICAÇÃO DE RECURSOS DE INFORMAÇÃO

A sobrecarga informacional ocorrida a partir de meados do século passado, fortalecida pelo emprego das Tecnologias da Informação e Comunicação (TICs), fez emergir alguns problemas relativos à organização e representação da informação contida nesses documentos (TEIXEIRA; SCHIEL, 1997; TEIXEIRA, 2010; RODRIGUES; CRIPPA, 2011; XAVIER, 2009; TEIXEIRA, 2005; OLIVEIRA; ARAUJO, 2012; PONTES JUNIOR; CARVALHO; AZEVEDO, 2013; SARACEVIC, 1999; DODEBEI, 2014).

Este Capítulo busca, apoiado na bibliografia disponível, levantar os esforços realizados no sentido de solucionar o problema supra e a importância dos identificadores de objetos utilizados na área da Biblioteconomia para que esta possa cumprir sua tarefa de “reunir, processar e difundir o conteúdo informacional dos documentos” (DZIEKANIAK, 2010).

A busca de soluções abrangentes que permitam, de modo eficaz, a identificação e localização dos objetos digitais fez surgir os identificadores de objetos e com eles os metadados (BRITO; GUEDES; SHINTAKU, 2013). Complementa Testal (2001) que em virtude do crescimento exponencial da quantidade de informação produzida “muitos projetos de normalização e localização foram desenvolvidos, cada um com suas vantagens e seus inconvenientes, encaminhados numa mesma direção: a identificação, acesso e descrição dos recursos”.

#### 3.1 METADADOS

As TICs tem influenciado o comportamento sociocultural de boa parte da humanidade, bem como vários campos da ciência, principalmente no tocante à produção, ao armazenamento e ao acesso à informação científica. A facilidade de acesso a estas tecnologias contribuiu para uma mudança de paradigma na forma como se dá a comunicação científica.

Para acompanhar este processo evolutivo de tendência ao digital, em preterição ao impresso, e para a Biblioteconomia seguir fazendo o tratamento descritivo da informação (TDI), a fim de bem representá-la, os profissionais da área passaram a adotar, a partir dos anos 90 do século passado, o termo “metadados”. Estes são utilizados nos “campos representativos da descrição de uma obra, principalmente acerca do seu título, autor(es), local de publicação, editora, data de publicação,

assunto(s) de que trata a obra e o seu resumo” (DZIEKANIAK, 2010; ALVES, 2010; ALVES; SANTOS, 2013; 2013a).

Este termo foi inicialmente utilizado pelos pesquisadores envolvidos com as TICs para descrever a arquitetura de suas bases de dados relacionais, no que refere às mais diversas propriedades dos campos de cada entidade relacionamento (tabela) (SAYÃO, 2010; DZIEKANIAK, 2010).

Segundo Sayão (2010) as novas imposições ditadas pelo ambiente digital conduziram ao redesenho da “ideia puramente descritiva de metadados, criando expansões para o seu conceito com o intuito de abrigar novos propósitos e funções”.

Para um entendimento do termo tratamento descritivo da informação (TDI), segue a definição dada por Alves (2010) e corroborada por Alves e Santos (2013):

A expressão tratamento descritivo da informação (TDI) é empregada para designar o processo de descrição na representação de um recurso, ou seja, a descrição, a individualização e a caracterização de um item informacional, na construção do registro bibliográfico que abarca o conjunto de atributos intrínsecos e extrínsecos (metadados) que caracterizam e individualizam um recurso.

Esclarece Alves (2010) que houve uma evolução no TDI em relação aos períodos anteriores ao surgimento dos metadados, uma vez que as TICs passam a ser agentes que determinam as mudanças, deixando de ser, tão somente, “instrumentos facilitadores do processo de catalogação”. O novo contexto proporcionado pelas TICs impactaram na percepção “de uma revisão e aprimoramento das teorias, dos princípios, dos fundamentos e dos instrumentos de representação no domínio bibliográfico, no intuito de serem melhor estabelecidos em meio digital”. Como consequência desta nova realidade pode-se afirmar que as TICs permitiram a realização do TDI no ambiente digital, “consolidando a existência dos metadados do domínio bibliográfico em uma estrutura passível de ser processada automaticamente”.

Uma definição para o termo metadados ainda não é consensual, ocorrendo variações conforme o contexto em que é utilizado. Defini-lo de acordo com a sua sintaxe que significa literalmente “dados sobre dados” conduziria a uma superficialidade (MORI; CARVALHO, 2004), uma vez que é “necessário entender e contextualizar seu significado de

acordo com o domínio onde está sendo utilizado” (ALVES; SANTOS, 2013; 2013a).

Para contribuir na formação de uma melhor convicção sobre o significado do termo metadados apresenta-se, a seguir, algumas definições, salientando que existem muitas, no entanto, para o propósito deste trabalho entende-se que as apresentadas aqui são suficientes.

Segundo a IDF (2015, tradução do autor)

Metadados pode ser definido, no contexto dos identificadores de objetos, como informações sobre um referente identificado, a fim de fornecer aos seres humanos ou máquinas os dados que eles precisam para que possam fazer uso desse referente identificado. Metadados podem incluir nomes, identificadores, descrições, tipos, classificações, locais, horários, medidas, relações e qualquer outro tipo de informação relacionada com um referente.

Nos estudos recentes de Alves e Santos (2013) “para o domínio bibliográfico o termo metadados pode ser entendido como atributos ou dados referenciais que representam um recurso”, cuja principal função “é fornecer uma representação padronizada e inequívoca dos recursos informacionais, com a finalidade de acesso, busca, localização, recuperação, intercâmbio de dados e interoperabilidade entre sistemas”.

Conforme salienta Ikematu (2001) “a finalidade principal dos metadados é documentar e organizar de forma estruturada os dados das organizações, com o objetivo de minimizar duplicação de esforços e facilitar a manutenção dos dados”. Ressalta, ainda, o autor o fato dos metadados permearem todas as funções da Tecnologia da Informação (TI), entretanto elucida que ocorrem falhas nos esforços de metadados porque há uma tendência em considerá-los apenas um recurso técnico.

Metadados são atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação (ALVES, 2010).

Para Mori e Carvalho (2004) “a finalidade principal dos metadados é documentar e organizar, de forma estruturada, os dados das organizações, com o objetivo de minimizar duplicação de esforços e facilitar a manutenção dos dados”.

Segundo Arellano (2004) a criação e a utilização de metadados são fundamentais nas estratégias operacionais visando a preservação digital, pelo fato destas estratégias basearem-se na conservação de *software* e *hardware*, as vezes emulando ou migrando, a fim de “garantir a autenticidade, registrar o gerenciamento de direitos e coleções de dados, e para a interação com recursos de busca”. Os metadados descrevem as características importantes do objeto digital, incluindo sua localização.

Os metadados têm um papel de fundamental importância na organização e no acesso às informações nos sistemas tradicionais, como nas coleções de livros de uma biblioteca ou nos ambientes informacionais baseados em redes de computadores, como é a própria *web*. Entretanto, o conceito de metadado pode ser expandido para apoiar a gestão de objetos digitais, cujo escopo inclui os processos de preservação digital de longo prazo (SAYÃO, 2010).

Metadado é informação estruturada que descreve, explica, localiza, ou de outra forma torna mais fácil a recuperação, o uso, ou o gerenciamento de um recurso de informação (NISO, 2004, tradução nossa). Ainda, segundo a NISO (2004) metadados revestem-se de fundamental importância na garantia da sobrevivência dos recursos e do acesso no futuro. Ressalta que o arquivamento e a preservação necessitam elementos especiais para a manutenção do histórico de um objeto digital (registro da origem e das diversas alterações sofridas ao longo do tempo), “para detalhar suas características físicas, e para documentar o seu comportamento, a fim de imitá-la em tecnologias futuras”.

Os metadados, conforme a NISO (2004), podem ser utilizados para qualquer nível de granularidade na descrição de um recurso. Podem descrever desde uma coleção (por exemplo, uma enciclopédia) até uma pequena parte de um recurso maior (por exemplo um verbete de um dicionário).

Na obra de Alves e Santos (2013) constata-se que os metadados ganham destaque como elementos intrínsecos e de relevada importância

nos sistemas e nos ambientes informacionais digitais, uma vez que abarcam os aspectos tecnológicos e representacionais específicos de domínio, “promovendo a representação, a individualização, o intercâmbio, a interoperabilidade entre sistemas, o acesso e a recuperação de recursos informacionais”.

Salientam as autoras que ocorre, no domínio bibliográfico, a convergência desses dois aspectos. O aspecto tecnológico caracteriza-se pela evolução das ferramentas tecnológicas com emprego dos princípios da engenharia de *software*, no sentido de levantar os requisitos funcionais, elaborar modelos conceituais e aprimorar a arquitetura dos sistemas de informação e das bases de dados, a fim de prover estruturação eficiente dos recursos informacionais em ambientes digitais diversificados. O aspecto representacional caracteriza-se pelo processo evolutivo de padronização na “codificação dos dados e, conseqüentemente, o aprimoramento de teorias, princípios, fundamentos e instrumentos para o tratamento descritivo da informação”.

### **3.1.1 Categorização dos Metadados**

Verifica-se na bibliografia uma diversidade de categorizações para os metadados, as quais ocorrem segundo alguma característica e/ou propriedade apresentada. Este trabalho se propõe a apresentar algumas destas categorizações segundo o entendimento de seus autores.

A categorização proposta pela NISO (2004), e seguida por Sayão (2010), é a que congrega uma parcela significativa dos autores que se dedicam à pesquisa sobre o assunto. Nela os metadados são segmentados em três categorias conceituais: metadados administrativos, metadados descritivos e metadados estruturais.

- Metadados Administrativos – provê informações relativas ao recurso informacional facilitando o gerenciamento deste recurso, tais como: datas de criação e de atualizações, tipo do arquivo, finalidade a que se destina, como foi gerado, restrições de acesso e outras informações técnicas.

- Metadados Descritivos – descrevem um recurso de modo a facilitar sua descoberta e recuperação. São elementos típicos deste tipo de metadados, o título, o autor, o resumo e as palavras-chave, entre outros.

Uma razão importante para a criação de metadados descritivos é facilitar a descoberta de informações relevantes. Além de descoberta de

recursos, os metadados podem ajudar a organizar os recursos eletrônicos, facilitar a interoperabilidade e integração de recursos legados, fornecer identificação digital e suporte de arquivamento e preservação (NISO, 2004).

- Metadados estruturais – contém as informações necessárias à recomposição de recursos compostos. Por exemplo, como as partes de um mosaico devem ser dispostas para facilitar sua reorganização.

Os metadados administrativos podem, ainda, apresentar subdivisões, das quais destaca-se duas: metadados de gestão de direitos e metadados para preservação.

- Metadados de gestão de direitos – contém informações sobre os direitos do autor e sobre direitos patrimoniais.

- Metadados para preservação – disponibilizam informações utilizadas no processo de arquivamento e preservação de um recurso. “Os metadados de preservação podem ser definidos, de uma forma simples e direta, como a informação que apoia e documenta a preservação de longo prazo de materiais digitais” (SAYÃO, 2010).

Neste sentido, Ferreira (2006) considera que os metadados de preservação

tem como objetivo descrever e documentar os processos e atividades relacionadas com a preservação de materiais digitais. Ou seja, a metainformação de preservação é responsável por reunir, junto do material custodiado, informação detalhada sobre a sua proveniência, autenticidade, atividades de preservação, ambiente tecnológico e condicionantes legais.

Complementa Arellano (2004) que os metadados de preservação “são uma forma especializada de administrar metadados que podem ser usados como um meio de estocar a informação técnica que suporta a preservação dos objetos digitais”.

Na categorização de Ikematu (2001) os metadados podem ser técnicos e de negócios.

- Metadados técnicos – são aqueles destinados à descrição dos dados utilizados pelas diversas “ferramentas para armazenar, manipular ou movimentar dados”.

- Metadados de negócio – descrevem os dados destinados ao entendimento, pelo usuário, do contexto do negócio. Salienta o autor

que o uso de metadados na geração de regras de negócio executáveis, pode possibilitar a representação, através de metadados, “de instruções de regra de negócio de acordo com o esquema de classificação que pode ser transformado em sistemas de informação do negócio”.

Na classificação de Mori e Carvalho (2004) os metadados podem ser divididos em dois tipos: estrutural ou semântico. Metadado estrutural é aquele destinado a fornecer informações sobre “a organização e estrutura dos dados gravados”. Metadado semântico disponibiliza informações para o entendimento do significado dos dados providos, bem como de seus relacionamentos semânticos.

Alves e Santos (2013) apresentam uma categorização de tipos de metadados de acordo com o desempenho funcional destes. Esclarecem que a função dos metadados “está relacionada com o propósito que se pretende atingir com sua aplicação em um sistema de informação”. São eles: administrativos, descritivos, preservação, técnicos e uso.

- Administrativos – são metadados relativos às peculiaridades do recurso informacional, de modo a permitir seu gerenciamento e administração. São exemplos de informações contidas neste tipo de metadados: data e local de criação, direitos do autor e direitos patrimoniais, tipo de arquivo, controle de acesso, etc.

- Descritivos – este tipo de metadados cumpre a função de identificação, descrição e representação dos recursos de informações. “Esse tipo fornece informações relacionadas com a catalogação, como título, autor, imprensa, data, resumo, palavras-chave, e ainda a relação dos *hiperlinks* entre os recursos, anotações de usuários, etc”.

- Preservação – destinam-se às informações relativas à longevidade do recurso informacional, no que concerne à sua preservação e conservação. São informações sobre as condições físicas da mídia que armazena o recurso e dos dados armazenados, além das ações tomadas para mantê-lo em bom estado de uso, etc.

- Técnicos – estes metadados disponibilizam informações relativas às funcionalidades dos sistemas e às interações dos metadados. “Esse tipo fornece informações sobre *hardware* e *software*, digitalização, controle do tempo de resposta dos sistemas, autenticidade e segurança dos dados (criptografia e senhas), etc”.

- Uso – são destinados a prover informações vinculadas à utilização dos recursos informacionais. Enquadram-se neste tipo de informação os registros de abertura e fechamento de sessões de usuários, os registros de exibição, controles de acesso, etc.

Destacam as autoras, outros fatores, além das funcionalidades, utilizados para categorizar os metadados:

- Quanto à origem dos metadados: internos, intrínsecos e externos;
- Quanto ao método de criação dos metadados: automáticos e manuais;
- Quanto à natureza dos metadados: não especializados e especializados;
- Quanto ao *status*: estáticos, dinâmicos, de longa duração e de curta duração;
- Quanto à estrutura: estruturados e não estruturados;
- Quanto à semântica: controlados e não controlados;
- Quanto ao nível: nível de coleção e nível de item.

### 3.1.2 Padrão de Metadados

O volume e a diversidade da produção de recursos de informação, em um elevado número de domínios, tornou a construção de representações ainda mais complexa. Salientam Alves e Santos (2013; 2013a) que para contornar os óbices advindos desta complexidade fez-se necessário padronizar metadados e seu esquema de padrão. Em domínios específicos reveste-se de relevada importância a elaboração dos metadados, uma vez que estes devem “refletir as necessidades representacionais de cada domínio e, ao mesmo tempo, estar em consonância com os aspectos tecnológicos” responsáveis por gerenciar as bases de dados e garantir seu uso pelos variados sistemas de informação digitais.

Consideram Mori e Carvalho (2004) que o estabelecimento de padrões é parte importante da tarefa de disseminação da informação, facilitando a interoperabilidade entre os sistemas das organizações.

Em sua tese de doutorado, Alves (2010) define que

os padrões de metadados são estruturas de descrição constituídas por um conjunto predeterminado de metadados metodologicamente construídos e padronizados. O objetivo do padrão de metadados é descrever uma entidade gerando uma representação unívoca e padronizada que possa ser utilizada para recuperação da mesma. Os padrões de metadados estruturados constituem-se como uma tendência atual a ser utilizada na Web, no intuito de oferecer a descrição estruturada dos recursos informacionais.

No entendimento de Mori e Carvalho (2004) os padrões de metadados cumprem a função de “fornecer as definições e formar uma rede para automatizar registros de propriedades e dados cadastrais de uma forma padronizada e consistente”.

As possibilidades proporcionadas pelas TICs conduziram ao desenvolvimento de vários projetos voltados à representação da informação no meio digital. Observa-se que devido a especificidade de alguns domínios um esquema de metadados de propósito geral não satisfaz todas as necessidades próprias daquele domínio, uma vez que são amplos e gerais. Para o escopo desta pesquisa torna-se conveniente citar um padrão de propósito geral amplamente utilizado no ambiente *Web*, o formato *Dublin Core* (DC) e um padrão que se estabeleceu no âmbito da Biblioteconomia, de propósito específico para o domínio bibliográfico, o formato MARC (*MAchine-Readable Cataloguing*) que efetiva e consolida uma integração estratégica entre o processo de catalogação e as TICs (ALVES, 2010; MORI; CARVALHO, 2004).

O padrão de metadados *Dublin Core* (DC) foi criado em 1995, por uma equipe interdisciplinar, com o objetivo precípua de descrever uma ampla quantidade de recursos informacionais digitais disponibilizados na *Web* de forma simples e eficaz.

Sua arquitetura é formada por 15 (quinze) elementos semânticos, não obrigatórios e que podem ser repetidos indefinidamente, se necessário. Neste padrão cada item de metadado utilizado na descrição do recurso segue a sintaxe chave-valor, ou seja, “há duas classes de termos: elementos (nomes) e qualificadores (adjetivos), que podem ser arranjados como um padrão simples de instruções” (MORI; CARVALHO, 2004).

Reafirma Alves (2010) a simplicidade e a eficácia do padrão de metadados DC, na descrição de um grande número de recursos de informação na *Web*, apesar de sua concepção inicial visar a descoberta de recursos naquela rede de computadores.

O Quadro 1 traz um exemplo de registro no padrão *Dublin Core*, onde pode ser observado as características supracitadas relativas ao padrão em tela.

O formato MARC (*MAchine-Readable Cataloguing*), específico para a área de Biblioteconomia e, atualmente, largamente utilizado no domínio bibliográfico, segundo Mori e Carvalho (2004) é fruto de uma adaptação da informação contida em uma da ficha catalográfica para o formato digital de registro bibliográfico (*Cataloging Record*), o qual foi elaborado de acordo com uma estrutura pré-determinada o que permite sua leitura por máquinas (*Machine-Readable*), mais especificamente

computadores. Este formato “é uma maneira de se registrar dados e metadados de maneira que máquinas possam lê-los”.

Quadro 1 – Exemplo de registro no padrão *Dublin Core*.

```
<?xml version="1.0" encoding="UTF-8"?>
<srw_dc:dc xmlns:srw_dc="info:srw/schema/1/dc-schema" xmlns="http://purl.org/dc/elements/1.1/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="info:srw/schema/1/dc-schema
    http://www.loc.gov/standards/sru/resources/dc-schema.xsd">
  <title>The lost symbol : a novel /</title>
  <creator>Brown, Dan, 1964-</creator>
  <type>text</type>
  <type>Suspense fiction. gsafd</type>
  <type>Adventure fiction. gsafd</type>
  <publisher>New York : Doubleday,</publisher>
  <date>c2009.</date>
  <language>eng</language>
  <description>
    Symbolist Robert Langdon returns in this new thriller follow-up to The Da Vinci Code.
  </description>
  <subject>Cryptographers--Fiction.</subject>
  <subject>Freemasonry--Fiction.</subject>
  <coverage>Washington (D.C.)--Fiction.</coverage>
  <identifier>URN:ISBN:9780385504225</identifier>
</srw_dc:dc>
```

Fonte: <<https://lccn.loc.gov/2009464840/dc>>. Acesso em 31 jan. 2016.

A *Library of Congress* (LC) criou o formato MARC, nos anos 60, em virtude da necessidade de dispor de uma estrutura padronizada para a “descrição, o armazenamento e o intercâmbio automatizado de registros bibliográficos e catalográficos de diferentes tipos de recursos informacionais em uma unidade de informação”. Atualmente, após várias unificações e atualizações de sua arquitetura descritiva e de suas regras de representação, o projeto denomina-se MARC 21, consolidando a “integração estratégica dos tradicionais princípios de TDI com as vantagens trazidas pelo uso” das TICs (ALVES; SANTOS, 2013a).

O formato MARC 21, segundo Chan (2007 apud ALVES, 2010), “é um conjunto de padrões desenvolvidos com a finalidade de representar e comunicar de forma legível por máquina, metadados descritivos sobre itens de informação”, mais especificamente os itens bibliográficos, no entanto não se restringe a estes. Na Figura 4 pode-se observar um exemplo de registro no padrão MARC 21, comumente utilizado para intercâmbio entre sistemas de informação.

Atenta às peculiaridades do ambiente *Web* quanto a exigência de maior flexibilidade na sintaxe de codificação para um esquema de metadados que queira se estabelecer neste ambiente, a *Library of Congress* implementou uma versão do MARC 21 na linguagem XML (*eXtensible Markup Language*), denominada MARCXML. A proposição

desta versão do formato MARC 21 em linguagem de marcação visa “solucionar a necessidade de flexibilidade, extensibilidade e modularidade sem, contudo, perder a especificidade de domínio, garantindo que a interoperabilidade entre sistemas não seja prejudicada” (ALVES; SANTOS, 2013a).

Figura 4 – Exemplo de registro no padrão MARC 21.

```
01189cam a2200325 a
45000010009000000500170000900800410002690600450006792500440011295501280015601000170
028402000180030103500240031904000670034304200140041004300120042405000290043608200160
046510000230048124500450050425000120054926000350056130000290059652000920062565000290
07176500026007466510032007726550029008046550030008331592244920141221150918.0090925s2
009 nyua 000 1 eng d a7bcbccopycatd2encipf20gy-gencatlg0 aacquireb2
shelf copiesxAY, 2014-12-21 brg04 2009-09-25 z-processorirg04 2009-09-25 to
CALMaBarcode 00240193838 returned from bindery 2009-09-29;tcopy 2 to CALM a
2009464840 a9780385504225 a(OCOLC)ocn399843314
aUPZcUPZdDovdJz6dDCBdREUDUUCdBTCTAdYDXCPdABGdFN4dDLC alccopycat an-us-
dc00aPS3552.R685434bL67 200900a813/.542221 aBrown, Dan,d1964-14aThe lost symbol :ba
novel /cDan Brown. 1st ed. aNew York :bDoubleday,cc2009. a509 p. :bill. ;c25 cm.
aSymbologist Robert Langdon returns in this new thriller follow-up to The Da Vinci
Code. 0acryptographersvFiction. 0aFreemasonryvFiction. 0aWashington (D.C.)vFiction.
7aSuspense fiction.2gsafd 7aAdventure fiction.2gsafd
```

Fonte: Assumpção e Santos (2015).

No Anexo A encontra-se um exemplo de registro no padrão MARCXML.

Entende-se que os estudos sobre metadados vem contribuindo para o desenvolvimento de projetos para a padronização de esquemas de metadados, quer seja de propósito geral ou de propósito específico, permitindo uma boa representação dos recursos informacionais nos diversos sistemas de informação e/ou no ambiente *Web*.

### 3.2 IDENTIFICADORES DE OBJETOS

A explosão documentária (informacional) ocorrida nas últimas décadas, aliada à necessidade de utilização destas informações *a posteriori*, através de uma recuperação ágil e precisa, contribuiu para aumentar a complexidade da organização dos documentos produzidos.

A este contexto soma-se a facilidade proporcionada pela *Web 2.0* na produção de documentos, uma vez que o usuário deixa de ser mero espectador e passa a ser sujeito atuante nesta produção documental.

Os profissionais ligados às áreas cujos interesses e responsabilidades foram afetados por esta explosão buscaram soluções que permitissem o tratamento descritivo, o armazenamento e, sobretudo, a localização para possibilitar a recuperação.

Uma iniciativa para a solução do problema foi a criação de identificadores de objetos que receberam ao longo do tempo várias classificações e denominações, no entanto, para o escopo desta pesquisa serão abordadas apenas as necessárias ao entendimento do problema.

A Tabela 1 traz uma compilação dos identificadores de objeto (recomposta de Park et al. (2008) e NISO (2006)) adaptada de Park et al. (2011).

No ambiente digital a preservação dos recursos faz emergir desafios de uma natureza fundamentalmente diferente se comparados com a preservação dos formatos tradicionais, face a manutenção do acesso a estes recursos, prevenindo-se contra alterações da tecnologia em uso e falhas na mídia armazenadora dos mesmos (WEBER, 2013).

Uma estratégia para auxiliar no enfrentamento destes desafios foi a adoção de identificadores que dispõem de uma estrutura tecnológica com pré-disposição de longevidade temporal e que evidenciam as características de acionabilidade, de escopo do padrão, de arquitetura e infraestrutura subjacente e de estado do padrão, denominados identificadores persistentes.

Considera Tonkin (2008) que identificadores persistentes são sustentáveis e permitem fazer referência a um objeto digital. Por exemplo “um arquivo ou conjunto de arquivos, como um *ePrint* (artigo, papel ou relatório), uma imagem ou um arquivo de instalação de um software”. Comenta o autor que

os únicos identificadores persistentes interessantes também são persistentemente acionáveis (ou seja, você pode ‘clicá-los’); no entanto, ao contrário de um *hyperlink* simples, identificadores persistentes devem continuar a proporcionar o acesso ao recurso, mesmo quando ele se move para outros servidores ou mesmo a outras organizações. Um objeto digital pode ser movido, removido ou renomeado por muitas razões.

Segundo os ensinamentos de Weber (2013) este tipo “consiste num identificador numérico, que vem acompanhado de resolução, metadados e política da instituição que os adota” e podem ser utilizados no ambiente digital para a proteção dos direitos do autor e dos direitos patrimoniais.

Salienta Tonkin (2008) que o identificador persistente contém metadados com a finalidade de descrever o objeto digital e não a referência de uma instância deste objeto. Por isso, esses identificadores,

Tabela 1 – Compilação dos Identificadores de Objeto (recomposta de Park et al. (2008) e NISO (2006)).

Area	Produto	RFID	Identificadores
Físico			<ul style="list-style-type: none"> <li>● ISO/IEC 15459: ISO/IEC JTC 1/SC 31)</li> <li>● EPC (Electronic Product Code): EPC Global</li> <li>● mCode: Korea Communications Commission</li> <li>● GTIN ou EUN/UPC: GS1: anteriormente conhecido como Uniform Code Council ou UCC</li> </ul>
		Código de Barras	
		Publicações	<ul style="list-style-type: none"> <li>● ISBN (International Standard Book Number): U.S. ISBN Agency</li> <li>● ISRN (International Standard Technical Report Number): ISO TC 46 SC 9</li> <li>● ISSN (International Standard Serial Number): ISDS International Center</li> <li>● ISMN (International Standard Music Number): International ISMN Agency</li> <li>● ISIL (International Standard Identifier for Libraries): Danish Agency for Libraries and Media</li> </ul>
		Biblioteca	
		Humano	<ul style="list-style-type: none"> <li>● MS passport: Microsoft</li> <li>● Electronic ID, Electronic passport: U.S. and Europe</li> <li>● ISO 11784/5: TC23/SC19/WG3</li> </ul>
		Animal, ações	
Digital	Áudio/Vídeo		<ul style="list-style-type: none"> <li>● MPEG 21 DII (Digital Item Identification): ISO/IEC JTC1</li> <li>● DMP (Digital Media Project): DMP General Assembly</li> <li>● ISAN (International Standard Audiovisual Number): ISO/TC 46/SC 9</li> <li>● ISRC (International Standard Recording Code): International ISRC Agency</li> <li>● DOI (Digital Object Identifier): International DOI Foundation</li> <li>● cID (Content ID): Content ID Forum in Japan</li> <li>● UCI (Universal Content Identifier): Ministério de Cultura, Esportes, e Turismo da Korea</li> </ul>
		Propósito Geral	

Fonte: adaptado de Park et al. (2011)

também considerados indiretos, necessitam que um resolvidor especifique e direcione o usuário para uma cópia atualizada do objeto. Isto ocorre porque, ao tomar conhecimento da descrição do objeto, o resolvidor possibilita ao navegador localizar e recuperar uma instância específica deste objeto.

O URL (*Uniform Resource Locator*) é o identificador mais utilizado na *Web*. Destina-se à localização de recursos, exercendo a função de enlace e fornecendo uma identificação resumida desta localização. No entanto, é limitado e tem se mostrado muito frágil porque, ao vincular os recursos a um local de rede e a um arquivo específico, perde sua funcionalidade se estes recursos são removidos e/ou renomeados (PARK et al. 2011).

Ponderam Brito, Guedes e Shintaku (2013) que a *Web* revelou-se uma fonte de informação valiosa na disseminação da informação, em particular, na comunicação científica. No entanto, apesar da importância da internet, algumas questões ainda não foram totalmente elucidadas, enlaces rompidos e alteração de URL ainda são comuns, impossibilitando a identificação do conteúdo digital de forma permanente.

Reafirmam Neumann e Brase (2014) que a utilização de URL não é uma boa opção para identificar um objeto na *Web* “por causa das inúmeras vezes que páginas ou objetos não podem mais ser encontrados (por exemplo ‘Erro 404 - Página não encontrada’) devido ao nome do servidor ou alterações”, uma vez que alterado o local de armazenamento perde-se o objeto. Em consonância com os autores acima, Jacso (2002) já havia anunciado que os “URLs trouxeram sua própria variedade de estilos de citação e recomendações, mas são notórios por sua falta de persistência e confiabilidade”.

Ressalta-se que alguns destes identificadores possuem normas em estágio avançado de desenvolvimento. Pode-se incluir neste rol o URN (*Uniform Resource Name*), o PURL (*Persistent URL*), o DOI (*Digital Object Document*), o NBNs (*National Bibliography Numbers*), o ARK (*Archival Resource Key*) e o *Open URL* (TONKIN, 2008).

O protocolo URN (*Uniform Resource Name*) foi desenvolvido pela IETF (*Internet Engineering Task Force*) numa tentativa de identificação persistente de objetos digitais, independente do local, e projetado para mapear facilmente outros *namespaces*<sup>3</sup>. “O protocolo

---

<sup>3</sup> *Namespace* é uma coleção de identificadores originalmente atribuídos. Isto é, um identificador é sempre atribuído a somente um recurso, vedada sua

URN é utilizado por vários identificadores de objeto para uma variedade de recursos digitais disponíveis na *Web*, incluindo a música, arquivos gráficos, *eBooks* e vídeos” (PARK et al., 2011).

### 3.3 IDENTIFICADORES DIGITAIS DE OBJETOS

A medida que o número de documentos digitais cresce e a facilidade de navegação entre eles melhora, os usuários se sentem cada vez mais frustrados quando os artigos citados por outros artigos encontram-se inacessíveis eletronicamente. Verifica-se a falta de um padrão único para a interligação entre os sistemas identificadores, com a agravante que o volume de informação originalmente digital cresce vertiginosamente.

Nas considerações de Sidman e Davidson (2001) os provedores de conteúdo orientado para o consumidor entendem que as chaves para permanecer num mercado cada vez mais exigente de entrega digital é a proteção de direitos autorais e interoperabilidade entre sistemas de informação. Apesar do conteúdo *on-line* não requerer ações típicas dos recursos físicos, existem uma série de ações análogas, tais como: “cadeia de transações necessárias para facilitar a sua venda, distribuição, sindicância, proteção de direitos autorais, e reuso”.

Fundamentalmente, a segurança de um comércio eletrônico e a garantia dos direitos do autor e dos direitos patrimoniais serão alcançados “em função do próprio desenvolvimento que alcancem os metadados e a descrição e reconhecimento inequívoco dos conteúdos” (TESTAL, 2001).

A edição digital de um periódico (comércio eletrônico de conteúdo digital) “requer gerenciamento de conteúdo com uma variedade de serviços associados para gerenciar o acesso e outros direitos”, além da persistência de identificadores unívocos (PASKIN, 1999).

Estes periódicos *on-line* tem se mostrado uma fonte de informação ágil que permeia o meio científico auxiliando pesquisadores e profissionais que buscam o acesso aos seus conteúdos. Entretanto, requer uma estrutura adjacente para validar estas informações, além de registrá-las em órgãos de controle, manter os metadados e um sistema de

---

atribuição a um recurso diferente do originalmente atribuído. Um único recurso, no entanto, pode ter mais do que um identificador que lhe são atribuídos para fins diferentes (DAIGLE et al., 2002).

resolução, controlar os direitos do autor, verificar plágio e originalidade (DAMASIO, 2013).

Com o intuito de amenizar estes e outros problemas correlatos, além de prover acesso rápido e seguro aos documentos é que surgiram os sistemas de identificadores digitais de objetos, trazendo consigo toda a estrutura tecnológica de suporte, cuja finalidade é dispor de informações sobre o objeto identificado, ou seja, seus metadados.

Segundo Wang (2007) os identificadores de objeto digital e seus metadados associados “são ingredientes fundamentais na construção de uma infraestrutura para suportar um banco de dados que integre editoras de artigos de periódicos e citações com base em um modelo de produção distribuída”.

Em referência aos sistemas de identificadores digitais de objetos a NISO (*National Information Standard Organization*) propõe quatro requisitos básicos: “a sintaxe de nomeação, um sistema de informação, metadados, políticas e procedimentos de governança e aplicação” (PASKIN, 2005 apud PARK et al., 2011).

Os sistemas que cumprem os requisitos acima são conhecidos como sistemas de identificadores acionáveis. São exemplos deste tipo de sistema para identificação digital de objetos: DOI (*Digital Object Identifier*), UCI (*Universal Content Identifier*) na Coreia e CID (*Content ID*) no Japão.

Alguns conceitos relativos aos identificadores são ‘resolução’, ‘singularidade’, ‘interoperabilidade’ e ‘persistência’, cujas definições abaixo seguem o entendimento de Paskin (2011):

- Resolução – é o processo pelo qual um identificador é utilizado como entrada para um serviço que fornecerá como saída um ou mais elementos de informação relacionados à entidade identificada;

- Singularidade – refere-se a característica de um conjunto de caracteres indicar um e somente um recurso informacional (o ‘referente’). Em contrapartida o inverso não é uma consequência lógica, ou seja, um recurso pode ter mais de um identificador. Por exemplo, um livro pode ter um ISBN (*International Standard Book Number*) e também um LCCN (*Library of Congress Control Number*);

- Interoperabilidade – evidencia a possibilidade de usar um identificador em diversos serviços além daquele onde foi originalmente atribuído;

- Persistência – é a propriedade de um identificador referir-se ao mesmo recurso indefinidamente. Assim, a persistência pode ser considerada como a ‘interoperabilidade com o futuro’.

Salienta Paskin (2011) que “a gestão de conteúdos em redes digitais requer um identificador persistente, único (original), resolúvel, e interoperável”.

### 3.3.1 *Universal Content Identifier (UCI)*

O *Universal Content Identifier* (UCI), baseado em um *namespace* URN, foi desenvolvido no âmbito da *National Computerization Agency* (NCA), do governo Coreano, para nomear recursos digitais persistentes, tais como música, vídeos, textos, imagens, *eBooks*, e outros tipos de recursos digitais produzidos ou dirigidos pela NCA.

A NCA é uma organização sem fins lucrativos com a missão de desenvolver e promover a infraestrutura de informação e gerir conteúdos digitais públicos na Coreia e, possivelmente, em todo o mundo (KANG, 2005).

Implantado a partir de janeiro de 2010, em pouco mais de um ano, o UCI já contava com mais de 24 milhões de itens de conteúdo digital registrados em sua estrutura. A este fato soma-se a sua adoção por uma série de empresas na Coreia, incluindo empresas de radiodifusão, organismos públicos de investigação, empresas de jornais e empresas de portal Internet (PARK et al., 2011).

O sistema UCI, gerenciado pela NCA, provê as funcionalidades de registro, resolução, pesquisa e administração, beneficiando o comércio eletrônico com a garantia de que todos os recursos são registrados digitalmente, identificados e resolvidos, independente da localização com base nos URN's.

O identificador UCI consiste em duas partes: código de prefixo e código de conteúdo. O código de conteúdo também é dividido em código de exemplo e código de qualificação, que é opcional.

O sistema UCI garante a unicidade do identificador através de suas autoridades de registro e da própria NCA, de modo que um identificador UCI seja atribuído a no máximo um único recurso durante o processo de registro (KANG, 2005).

A persistência do identificador será respaldada através de um sistema de *backup* e da designação de um outro serviço de registro em substituição a algum desses serviços que por ventura venha a ficar inoperante.

O processo de atribuição do identificador para um recurso digital é emitido pelo organismo de registro, a pedido de um titular e a resolução é controlada pela NCA e seus delegados. A cessão e utilização de identificadores é realizada de acordo com as regras estabelecidas pela

NCA e que seguem os requisitos e sintaxe do URN. Dispõe de um mecanismo de validação através de uma lista de código de prefixo disponível tanto via *on-line* quanto *off-line*, mediante solicitação (KANG, 2005).

### 3.3.2 *Content ID (cID)*

O content ID (cID) foi desenvolvido no Japão e se assemelha ao sistema de identificadores que será abordado na sequência, o *Digital Object Identifier* (DOI<sup>®</sup>). Destina-se à identificação de conteúdos digitais para apoiar o gerenciamento de direitos autorais, proteção e pagamento. Concentra-se principalmente nos conteúdos de vídeo e adota técnicas de marca d'água na proteção de direitos de propriedade intelectual associados ao conteúdo digital (PARK et al., 2011).

Proposto no âmbito do *Content ID Forum* (CIDF) como um identificador único para a atribuição de pacotes de conteúdo digital (que contenha um ou vários itens) destinado à distribuição, cuja identificação do conteúdo continua persistentemente associada ao pacote de conteúdos digitais através da utilização de marca d'água ou outras tecnologias, proporcionando benefícios de longo alcance a todos os participantes da cadeia de distribuição de conteúdo digital.

Segundo o CIDF (2007, tradução nossa) o *Content ID* é um “identificador único atribuído a um pacote de conteúdos digitais para a distribuição e exploração que está ligado ao atribuir exclusivamente informações que descrevem o pacote de conteúdos digitais”.

Dispõe de um descritor de conteúdo distribuído que abriga os metadados e cumpre as funções abaixo, enumeradas por CIDF (2007):

- Permitir a identificação exclusiva do recurso;
- Ativar a autenticação de conteúdo por meio de assinatura XML;
- Identificar sinais de adulteração (alteração ou emenda) com os metadados e/ou recurso por meio de assinatura XML;

A marca d'água digital incorpora metadados em pacotes de conteúdo digital, alterando o conteúdo digital de uma forma que é quase imperceptível em uso normal.

Os metadados de um *Content ID* (ID de um item digital), um identificador único e um conjunto de atributos que descrevem item(s) e pacote(s) de conteúdos digitais (e os termos sob os quais são distribuídos), assim possibilitando o pagamento eficaz de direitos e/ou *royalties* (ou taxa) para o interessado apropriado (CIDF, 2007).

### 3.3.3 *Digital Object Identifier (DOI®)*

O sistema DOI® é um empreendimento da iniciativa conjunta de três associações comerciais da indústria editorial (*International Publishers Association; International Association of Scientific, Technical and Medical Publishers; Association of American Publishers*). Não obstante sua origem na publicação textual, sua concepção foi como uma estrutura genérica para a gestão de identificação de conteúdo em redes digitais, atendendo a tendência rumo à convergência digital e à disponibilidade multimídia. Com a finalidade de desenvolver e gerenciar o sistema DOI® também foi criada, em 1997, a *Internacional DOI® Foundation (IDF)*. A sintaxe do DOI® foi padronizada, em 2000, através da *National Information Standards Organization (NISO)* e o sistema foi padronizado pela Organização Internacional de Normalização como ISO 26324:2012, *Digital Object Identifier System (IDF, 2015)*.

Toda a estrutura tecnológica subjacente do sistema DOI® é dada pelo *Handle System* desenvolvido pela *Corporation for National Research Initiatives (CNRI) (TESTAL, 2001)*. O *handle system* é uma estrutura que permite alças para efetuar a resolução de forma distribuída, utilizando-se de clientes dedicados e clientes comuns, como navegadores *Web* com extensões especiais ou *plugins*, ou sem extensões, passando por vários servidores intermediários (CHANDRAKAR, 2006).

De acordo com ISO (2012) – ISO 26324:2012 – o Sistema *Digital Object Identifier (DOI®)* fornece uma infraestrutura para identificação única e persistente de objetos de qualquer tipo. Ressalta que o significado de DOI® é ‘identificador digital de um objeto’ e não, como pode parecer a uma primeira vista, ‘identificador de um objeto digital’. O sistema foi desenvolvido para fazer uso dos benefícios oferecidos pela internet. Um nome DOI é atribuído, de modo permanente a um objeto para fornecer um enlace de rede persistente e resolúvel para informações atualizadas sobre esse objeto, inclusive onde o objeto, ou informações sobre ele, pode ser encontrado na Internet.

Um nome DOI® é um identificador (não um local) de um recurso disponível em redes digitais. Dispõe de um sistema de identificação persistente e acionável e de intercâmbio de informações interoperáveis, cujo gerenciamento se dá em redes digitais. Sua atribuição pode ser feita a qualquer tipo de recurso – físico, digital ou abstrato – para

compartilhamento ou gerenciamento. Por exemplo a gestão da propriedade intelectual (ISO, 2012).

A descrição do referente é realizada através de metadados. Exige-se, para validar um nome DOI<sup>®</sup> (identificador propriamente dito), um conjunto mínimo de metadados, denominado ‘DOI<sup>®</sup> *Kernel*’, para fins de reconhecimento e interoperabilidade. O reconhecimento é obtido a partir do instante em que os metadados do *Kernel* são suficientes descrever claramente o recurso referido pelo DOI<sup>®</sup>, permitindo ao usuário identificá-lo com precisão (por vários nomes, identificadores e as relações). A interoperabilidade é alcançada quando os elementos de dados ou os seus valores são comuns aos diversos esquemas de metadados (IDF, 2015).

A IDF (2015) “reconhece que a integração automatizada de metadados é a chave para a realização do pleno potencial do DOI<sup>®</sup> como ferramenta para o comércio e cultura digital”.

Retomando a ideia de que o “consumo e utilização de informação é altamente relacionada com a sua recuperação e a maneira como uma unidade de informação pode ser identificada e caracterizada por um SRI”, salientam Poulos, Korfiatis e Bokos (2011) o fato de as Bibliotecas e outras instituições de preservação da informação buscarem o desenvolvimento e a implementação de identificadores persistentes, dentre os quais pode-se destacar, como mais reconhecido, o *Digital Object Identifier* (DOI<sup>®</sup>). Esclarecem, ainda, os autores que “identificadores identificam exclusivamente entidades de diversos tipos, tornando possível o referenciamento preciso e eficaz”.

A característica de acionabilidade é conferida ao DOI<sup>®</sup> pela resolução, cujo processo de internet “associa um DOI a um ou mais valores atuais das informações sobre o objeto identificado. Ele é usado para significar o ato de apresentar um ou mais elementos de informação atual relacionado com o identificador” (CHANDRAKAR, 2006).

A sintaxe do DOI<sup>®</sup> é constituída de em duas partes, separadas por uma barra inclinada (“/”): um prefixo e um sufixo. “O prefixo identifica os editores, que por sua vez atribui o sufixo. O sufixo identifica o trabalho individual ou parte(s) de uma obra”. Cada DOI<sup>®</sup> atribuído vincula-se a um ou mais URL que permite sua precisa e eficaz resolução. O editor é o responsável pela designação do sufixo único, cuja singularidade será validada pela Autoridade de Registro (AR) e pela IDF (JACSO, 2002).

A Figura 5 traz um exemplo de como a Equipe do Portal de Periódicos da UFSC faz a atribuição de um DOI<sup>®</sup> a um artigo de seus periódicos.

Figura 5 – Exemplo de prefixo e sufixo de um DOI® para periódicos UFSC.

**para artigo:**

**10.5007/1518-2924.2010v15n30p1**

↑                    ↑                    ↑                    ↑                    ↑                    ↓

**prefixo            ISSN            ano    volume    número    pag. inicial**

**do periódico**

**(separado por ponto)**

Fonte: Weber (2013, p. 84).

O Quadro 2 mostra um exemplo de código XML, requisitado pela AR CrossRef®, elaborado por um afiliado, para atribuir um nome DOI® a um recurso informacional.

Quadro 2 – Exemplo de código XML para registro de nomes DOI na CrossRef.

---

```

- <doi_batch version="4.3.0" xsi:schemaLocation="http://www.crossref.org/schema/4.3.0 http://www.crossref.org/schema/deposit/crossref4.3.0.xsd">
- <head>
  <doi_batch_id>RevTreina_1372256562</doi_batch_id>
  <timestamp>1372256562</timestamp>
- <depositor>
  <name>Ronnie Silva</name>
  <email_address>ronnie@ibict.br</email_address>
</depositor>
  <registrant>IBICT</registrant>
</head>
- <body>
- <journal>
  - <journal_metadata>
    <full_title>Revista de Treinamento</full_title>
    <issn_media_type="electronic">66668888</issn>
    <issn_media_type="print">12345678</issn>
  </journal_metadata>
  - <journal_issue>
    - <publication_date media_type="online">
      <month>05</month>
      <day>22</day>
      <year>2013</year>
    </publication_date>
    - <journal_volume>
      <volume>2</volume>
    </journal_volume>
    <issue>1</issue>
  - <doi_data>
    <doi>10.5555/revtreina.v2i1</doi>
  - <resource>
    http://10.0.0.191/ojs-2.4.2/index.php/revistatreinamento/issue/view/4
  - <resource>
    </resource>
  </doi_data>

```

---

Fonte: Brito, Guedes e Shintaku (2013, p. 32).

Esclarecem Brito et al. (2015) a vantagem em utilizar um nome DOI<sup>®</sup> antecedido pelo prefixo “<http://dx.doi.org/>” para que este forneça um enlace ativo, facilitando o acesso do leitor ao texto completo, com apenas um clique de mouse. A utilização do link ativo em uma referência facilita a aplicação das métricas de acesso e citação, além de dispensar a indicação do URL do recurso informacional, conforme mostrado na Figura 6.

Figura 6 – Exemplo de referência bibliográfica utilizando um enlace ativo.

FELISBERTO, Proxério Manoel; MARCELINO, Roderval; ALVES, Maria Bernardete Martins; ALVES, João Bosco da Mota; SILVA, Juarez Bento da; Gruber, Vilson. A geração e o gerenciamento de referências bibliográficas com o Mecanismo Online para Referências - MORE. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 20, n. 42, p. 79-92, abr. 2015. <http://dx.doi.org/10.5007/1518-2924.2015v20n42p79>

Fonte: Brito et al. (2015, p. 40).

Segundo Wang (2007) a utilização do DOI<sup>®</sup>, em um ambiente distribuído, facilita a gestão de objetos digitais para os editores e os seus clientes, uma vez que cada nome DOI<sup>®</sup> “está associado a uma série de metadados, um conjunto de informações bibliográficas e comercial no domínio do conteúdo (título, autor, data de publicação, *copyright*, preço, etc.)”.

No trabalho de Rosenblatt (2002) verifica-se que a Integração de Conteúdo Empresarial (ICE<sup>4</sup>) pode conduzir a um novo sistema de Gerenciamento de Ativos Digitais (GAD) e que a “utilização do DOI<sup>®</sup> como um identificador e mecanismo de enlace com a arquitetura do ICE pode trazer vários benefícios” (os oferecidos pelo sistema DOI<sup>®</sup>), evitando “custos e riscos do desenvolvimento e manutenção de ferramentas proprietárias”.

No Brasil, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) utiliza os metadados do DOI<sup>®</sup>, através da Plataforma Lattes, como uma forma de validar a produção científica

---

<sup>4</sup> ICE é uma aproximação para o gerenciamento de ativos digitais para o nível empresarial criando um simples repositório monolítico em favor da adição de uma camada de funcionalidade que integra conteúdo armazenado em vários tipos de sistemas para unidade de nível departamental ou de negócios (ROSENBLATT, 2002).

cadastrada pelos pesquisadores em seus currículos (WEBER, 2013; DAMASIO, 2013).

Neste trabalho utiliza-se as bases de metadados DOI<sup>®</sup> das AR CrossRef<sup>®</sup>, DataCite<sup>®</sup> e mEDTA<sup>®</sup> que nas considerações de Ma e Yang (2014) é uma abordagem universal, pelo fato destas AR deterem ou gerenciarem mais de 95% dos nomes DOI<sup>®</sup> atribuídos até o presente momento. Segundo Damasio (2013), a CrossRef<sup>®</sup> com mais de 77 milhões de enlaces DOI<sup>®</sup> desponta com 96% dos registros mundiais.



## 4 TECNOLOGIAS COMPUTACIONAIS UTILIZADAS

Ao longo deste capítulo procura-se abordar as tecnologias computacionais utilizadas na construção do protótipo que foi integrado ao MORE com a finalidade precípua de viabilizar o resgate dos metadados do identificador informado pelo usuário, a fim de dispor de subsídios para o preenchimento dos campos do formulário utilizado na geração da referência bibliográfica.

Por oportuno, cabe ressaltar que são discutidas aqui apenas aquelas revestidas de suma importância no desenvolvimento do protótipo, uma vez que as aplicadas no sistema, como um todo, foram estudadas detalhadamente em Felisberto (2013). Buscou-se, então, na literatura, a semântica dos seguintes termos: *REpresentational State Transfer* (REST), *JavaScript Object Notation* (JSON) e *Client URL Request Library* (CURL).

### 4.1 REPRESENTATIONAL STATE TRANSFER (REST)

Roy Fielding em sua tese de doutorado referiu-se a um restrito conjunto de operações, com semântica uniforme, para construir uma infraestrutura que pode suportar qualquer tipo de aplicação *Web*, entendendo este estilo arquitetônico como *REpresentational State Transfer* (REST). Assim, REST descreve a *Web* como uma aplicação hipermídia distribuída cujos recursos se comunicam através da troca de representações do estado destes recursos (WEBBER; PARASTATIDIS; ROBINSON, 2010).

Segundo esclarecimentos de Severance (2015), Fielding cunhou o termo REST “para capturar a idéia de que os elementos de aplicações de rede estavam trocando seus estados usando protocolos em toda a rede”.

Um estilo arquitetônico é um conjunto coordenado de restrições arquitetônicas cuja finalidade é delimitar características e/ou funções de elementos arquitetônicos e as relações permitidas entre esses elementos dentro de qualquer arquitetura que está em conformidade com esse estilo (FIELDING, 2000; CA TECHNOLOGIES, 2015).

Para um entendimento mais preciso sobre os termos que gravitam em torno de um estilo arquitetônico utiliza-se as definições encontradas em Fielding (2000), onde:

- arquitetura de software é uma abstração dos elementos em tempo de execução de um sistema de software em alguma fase de sua operação. Um sistema pode ser composto de diversos níveis de abstração e muitas fases da operação, cada um com sua própria

arquitetura de software. Assim, uma arquitetura de software é definida por uma configuração dos elementos de arquitetura – componentes, conectores e dados – limitando suas relações, a fim de conseguir um conjunto desejado de propriedades de arquitetura;

- um componente é uma unidade abstrata de instruções de software e estado interno que proporciona uma transformação de dados através da interface;

- um conector é um mecanismo abstrato que medeia comunicação, coordenação ou cooperação entre os componentes;

- um ponto de referência é um elemento de informação que é transferido a partir de um componente, ou recebido por um componente, através de um conector;

- configuração é a estrutura das relações entre os componentes arquitetônicos, conectores e dados durante um período de tempo de execução do sistema.

Para Webber, Parastatidis e Robinson (2010) “REST introduz um conjunto de princípios que, quando aplicada ao design de sistemas distribuídos, produzem as características desejáveis de escalabilidade, uniformidade, desempenho e encapsulamento”.

Em CA Technologies (2015), REST é entendida como um “conjunto de restrições arquitetônicas que tentam minimizar a latência e as comunicações em rede e, ao mesmo tempo, maximizam a independência e a escalabilidade de implementações de componentes”.

Na visão de Battle e Benson (2008), *Representational State Transfer* (REST) é uma padronização para manipulação de recursos que firmou-se como um padrão de fato para a implementação de serviços em aplicações *Web 2.0*. Considera, ainda, que a simplicidade do REST, juntamente com o seu ajuste natural através de HTTP, tem contribuído para a sua escolha como um método para aplicações *Web 2.0*, no tocante ao gerenciamento de seus dados.

Segundo Castillo et al. (2013) REST “é um método alternativo para a construção de serviços *Web*”. Em um estilo arquitetônico REST, um cliente envia solicitações ao servidor que as processa e retorna respostas para o cliente. Solicitações e respostas representam recursos que podem ser encontrados através de um *Uniform Resource Identifier* (URI).

No trabalho de Griffin e Flanagan (2011) REST é definido como “um estilo de arquitetura híbrida para sistemas distribuídos de hipermídia, que utiliza uma interface uniforme para manipular o estado de um determinado recurso”. Apesar de não ser dependente de qualquer protocolo em particular, normalmente utiliza o HTTP. Mesmo havendo

a possibilidade de criação de um sistema baseado em REST que não seja construído sobre HTTP, as implementações mais práticas do REST são construídas utilizando HTTP.

A contribuição de REST na viabilização e massificação da *Web 2.0* atraiu a atenção e tornou-se objeto de estudo de vários pesquisadores e profissionais. Alguns dos quais estão citados acima e, por julgar oportuno e para corroborar os entendimentos supra, segue os ensinamentos de Rykowski (2015) sobre o assunto. Para o autor, “REST define a arquitetura e as regras para a construção de um sistema de rede que permite altíssima escalabilidade”. Este modelo é baseado em um recurso identificado por *Uniform Resource Identifier* (URI). Uma vez que cada recurso REST é representado por um endereço próprio, não se faz necessário colocá-los todos no mesmo local, permitindo que cada um dos recursos (ou qualquer subconjunto) possa estar localizado em um hospedeiro diferente. Sua simplicidade e economia de recursos permite o fornecimento de implementações REST para hardware e software limitado. Na verdade, os recursos de REST, além da sua própria funcionalidade, necessitam apenas a implementação de um protocolo (o mais utilizado é o HTTP) para operar. Sistemas baseados em REST são escaláveis e podem ser distribuídos livremente.

Na concepção resumida de Fielding (2000), seu idealizador,

REST é um conjunto coordenado de restrições arquitetônicas que tenta minimizar a latência de comunicação e de rede e, ao mesmo tempo maximizar a independência e a escalabilidade das implementações de componentes. Isto é conseguido, colocando restrições sobre a semântica do conector onde outros estilos têm-se centrado sobre a semântica de componentes. REST permite o armazenamento em cache e reutilização de interações, substituíbidade dinâmica de componentes, e processamento de ações por parte de intermediários, respondendo assim às necessidades de um sistema hipermídia distribuído em escala de Internet.

Reafirma o autor supra que o *Representational State Transfer* (REST) é um estilo híbrido de arquitetura para sistemas distribuídos de hipermídia, cuja finalidade é fornecer um conjunto de restrições de arquitetura que, ao ser aplicado como um todo, permite melhorar a escalabilidade das interações de componentes, a generalidade das

interfaces, a implantação independente de componentes e componentes intermediários para reduzir a latência de interação, além de reforçar a segurança e encapsular sistemas legados.

Em Ong et al. (2015) verifica-se que o “núcleo do projeto baseado em REST é um conjunto de operações de transferência do estado universal para qualquer armazenamento de dados e sistema de recuperação”. A comunidade *Web 2.0* adotou um mapeamento informal de operações CRUD (*Create, Read, Update, Delete*) para os comandos fornecidos pelo protocolo HTTP: POST, GET, PUT e DELETE, respectivamente. Estes comandos identificam a operação CRUD particular que está sendo solicitada do recurso identificado pelo URL de controle.

Outros dois comandos frequentemente usados são HEAD que é semelhante ao GET, porém o servidor retorna apenas o cabeçalho da mensagem e OPTIONS que retorna informações sobre as opções de comunicação disponíveis para o recurso identificado pelo URI de solicitação (CA TECHNOLOGIES, 2015).

Por convenção, o termo RESTful é utilizado para designar o conjunto completo de requisitos de uma API, de um serviço ou de uma interface que esteja em consonância com os princípios de design de REST (ONG et al., 2015; GRIFFIN; FLANAGAN, 2011).

Em um projeto RESTful, segundo Ong et al. (2015),

cada objeto é representado como um recurso único e pode ser consultado de maneira uniforme. Um serviço HTTP RESTful expõe um conjunto consistente de semântica que utiliza métodos HTTP (GET, POST, PUT, DELETE, etc.) em conjunto com URIs exclusivos para acessar os recursos subjacentes. Isto permite a criação de uma API, utilizando uma combinação de métodos de HTTP e URIs.

A identificação de REST como um estilo arquitetônico advém de suas seis restrições (princípios), que ao serem aplicadas conjuntamente o caracterizam como um estilo híbrido, uma vez que estas restrições foram, originalmente, empregadas em outros estilos arquitetônicos e, portanto, já conhecidas pelos profissionais da área. São elas: cliente/servidor, sem estado, cache, interface uniforme, sistema em camadas e codificação sob demanda (esta última é opcional).

- **Cliente/Servidor:** o objetivo principal desta restrição é a definição de tarefas (operações) que cada um deles deve executar,

porque ao separar as preocupações da interface do usuário das preocupações de armazenamento de dados, podemos melhorar a portabilidade da interface do usuário através de múltiplas plataformas e melhorar a escalabilidade, simplificando os componentes do servidor. Esta separação de componentes oferece as condições necessárias para a evolução independente, de modo a suportar ambientes que exijam alta escalabilidade (FIELDING, 2000).

Nesta situação o servidor disponibiliza determinadas operações através de serviços que ficam em espera permanente por solicitações dos clientes consumidores das representações ofertadas (CA TECHNOLOGIES, 2015; GRIFFIN; FLANAGAN, 2011).

- **Sem Estado:** para alcançar o desiderato desta restrição, cada solicitação de cliente para servidor deve conter todas as informações necessárias para compreender o pedido, e não pode tirar vantagem de qualquer contexto armazenado no servidor, cabendo somente ao cliente manter o estado da sessão em curso. Desta forma, melhora-se as propriedades de visibilidade, confiabilidade e escalabilidade, entretanto, pode diminuir o desempenho da rede, aumentando os dados repetitivos (sobrecarga por interação) enviados numa série de pedidos, uma vez que os dados não podem ser deixadas no servidor num contexto partilhado (FIELDING, 2000; CA TECHNOLOGIES, 2015; GRIFFIN; FLANAGAN, 2011).

- **Cache:** a contribuição desta restrição é no sentido de melhorar a eficiência da rede, impondo a necessidade de constar, explícita ou implicitamente, no cabeçalho de uma mensagem de resposta a um pedido a classificação daquela resposta como armazenável em cache, não armazenável em cache, ou armazenável temporariamente em cache. Para as respostas armazenáveis em cache ou armazenáveis temporariamente em cache é facultado ao cliente o direito mantê-las em um cache privado para reutilização dos dados para resposta aos pedidos posteriores equivalentes.

A vantagem da adição de cache é que eles possibilitam a eliminação, completa ou parcial, de algumas interações, melhorando a eficiência, escalabilidade e desempenho percebido pelo usuário através da redução da latência média de uma série de interações. Uma observação interessante é que, o pedido da rede mais eficiente é aquele que não utiliza a rede, ou seja, a capacidade de reutilizar um cache de resultados de resposta melhora consideravelmente o desempenho da aplicação (FIELDING, 2000; CA TECHNOLOGIES, 2015; GRIFFIN; FLANAGAN, 2011).

- **Interface Uniforme:** a ênfase em uma interface uniforme entre os componentes é a característica central que distingue o estilo arquitetônico REST de outros estilos baseados em rede. A generalidade da interface do componente simplifica a arquitetura geral do sistema e melhora a visibilidade das interações. Em contrapartida, a eficiência pode ser degradada pelo uso de uma interface uniforme, uma vez que a informação é fornecida de acordo com padrões pré-estabelecidos e não contempla todas as especificidades dos diversos aplicativos que a utilizam. Para obter uma interface uniforme em REST, quatro restrições de interface foram definidas: identificação de recursos; manipulação de recursos através de representações; mensagens autodescritivas; e, hipermídia como o motor do estado do aplicativo (FIELDING, 2000; CA TECHNOLOGIES, 2015; GRIFFIN; FLANAGAN, 2011).

- **Sistema em Camadas:** as restrições do estilo do sistema em camadas permite a composição de uma arquitetura em camadas hierárquicas com a finalidade de limitar a interação do componente à camada com a qual está interagindo. A partir do momento que o conhecimento do sistema fique restrito a uma única camada, sua complexidade geral também é limitada, promovendo sua independência. Assim, as camadas podem ser usadas para encapsular serviços legados e para proteger os novos serviços de clientes legados. Os intermediários podem também ser utilizados para melhorar a escalabilidade do sistema, permitindo o balanceamento da carga de serviços através de múltiplas redes e processadores (FIELDING, 2000; CA TECHNOLOGIES, 2015; GRIFFIN; FLANAGAN, 2011).

- **Codificação Sob Demanda:** esta restrição é opcional e transfere para o cliente (p.ex. um agente de usuário - navegador *Web*) a atualização da lógica, independente da lógica do servidor, através de “código executável fornecido por provedores de serviços para consumidores” (p.ex. *applets* e *scripts*).

Retomando as restrições definidas em REST para uma interface uniforme, segue as considerações de Fielding (2000) e CA Technologies (2015) sobre o assunto:

- **Identificação de recursos:** um conceito chave na concepção e implementação de REST e utilizado para designar a abstração de uma informação é o recurso. Um recurso pode ser qualquer informação: um documento ou imagem, um serviço temporal, uma coleção de outros recursos, um objeto não-virtual, etc.

Para identificar univocamente um recurso particular envolvido em uma interação entre componentes, REST faz uso de identificadores. Uma interface genérica, fornecida por conectores, permitem o acesso e a

manipulação do valor ajustado de um recurso. O proprietário/detentor do recurso que recebe o identificador de recursos, possibilitando sua referência, é responsável por manter a validade semântica do mapeamento ao longo do tempo.

A adoção do HTTP como o protocolo mais utilizado na construção de aplicações RESTful contribuiu para popularizar *Uniform Resource Identifier* (URI), *Uniform Resource Locator* (URL) e *Uniform Resource Name* (URN) como identificadores de recursos. Vale lembrar que um URI é uma cadeia de caracteres utilizada na identificação de um recurso abstrato ou físico e pode ser obtido através de um URL ou um URN. O URL é “usado para identificar o local online de um recurso individual, enquanto o URN é um identificador persistente independente de local”.

- **Manipulação de recursos através de representações:** uma representação é uma sequência de bytes, além dos metadados desta representação para descrever esses bytes. As arquiteturas baseadas em REST se comunicam através da transferência de representações de recursos, cuja latência pode ser afetada tanto pela concepção dos protocolos de comunicação como pelo design dos formatos das representações de dados. Outros nomes comumente usados, mas menos precisos para uma representação incluem: documento, arquivo, entidade, mensagem HTTP, instância ou variante.

Hoje em dia os tipos *Multipurpose Internet Mail Extensions* (MIME) possibilitam a troca de dados pela internet em diferentes formatos: áudio, vídeo, texto, imagens e programas de aplicativos.

- **Mensagens autodescritivas:** a combinação de metadados em cabeçalhos, bem como o corpo de uma mensagem, favorece a criação de uma mensagem autodescritiva. As arquiteturas baseadas no estilo REST esperam que cada solicitação de cliente e resposta de servidor seja uma mensagem autodescritiva (também referenciada como “sem estado” ou “sem contexto”), de modo a conter todas as informações necessárias a conclusão da tarefa. Cada mensagem passada entre o cliente e o servidor pode ter um corpo e metadados.

- **Hipermídia como o motor do estado do aplicativo:** esta restrição de interface uniforme utiliza-se de links e formulários como uma maneira de descrever para o cliente as ações disponíveis que são suportadas atualmente pelo serviço. Um exemplo bem didático é a grande quantidade de funcionalidades suportada pela HTML, e o navegador *Web* comum implementa rotinas para a compreensão das regras para todas elas. A facilidade de reconhecimento dos links e formulários em uma mensagem HTML é evidente, no entanto, as regras

de processamento ou as semânticas que estão associadas a eles podem não ser tão claras.

Neste contexto, hiper-mídia pode ser definida pela presença de informações de controle do aplicativo embutidas dentro da apresentação das informações ou em uma camada acima. A arquitetura baseada em REST oferece aos provedores de serviço a habilidade de enviar informações de controle (links e formulários) a aplicativos de cliente, conhecidos como controles de hiper-mídia.

As considerações desta Seção favorecem o entendimento do modelo utilizado na implementação do protótipo desenvolvido para este trabalho e pelas autoridades de registro do DOI para a troca de mensagens usadas na obtenção dos metadados a serem utilizados na referência bibliográfica.

#### 4.2 JAVASCRIPT OBJECT NOTATION (JSON)

Nos últimos anos, um formato para intercâmbio de dados, deixou de ser um quase segredo, pouco conhecido e raramente usado, para tornar-se a escolha certa para aplicações de dados tradicionais. Trata-se de JSON, que é a sigla para *JavaScript Object Notation*, ou Notação para Objetos em *JavaScript* (SEVERANCE, 2012; SILVA, 2009).

Segundo Severance (2012) JSON “é um formato popular para serialização de dados”. Os programadores usam-no intensivamente para codificar dados para transferência entre um servidor e um aplicativo AJAX, para conectar dois servidores comunicando-se através de serviços *Web*, e em muitos outros cenários semelhantes.

Para Smith (2015) o formato de dados JSON é derivado dos literais da linguagem de programação *JavaScript*, tornando-o um subconjunto desta linguagem. O fato de ser um subconjunto, limita JSON de alguns recursos adicionais que a linguagem *JavaScript* em si já não possui. Embora seja um subconjunto de uma linguagem de programação, JSON não é uma linguagem de programação em si, mas, de fato, um formato de intercâmbio de dados.

JSON é conhecido como a norma de intercâmbio de dados, o que implica que ele pode ser usado como o formato de dados sempre que a troca de dados ocorrer. Esta troca de dados pode acontecer entre agente de usuário e servidor, entre dois servidores ou entre qualquer dois ou mais nós de rede que necessitem intercambiar dados.

Esclarece Silva (2009) que JSON “é uma técnica *JavaScript* que usa sintaxe leve e simples para armazenar dados. Tal como XML, JSON

está estruturado de modo a ser facilmente lido, interpretado e processado por máquinas”.

Ensina Bassett (2015) que “um formato de intercâmbio de dados é um formato de texto usado para trocar dados entre as plataformas”. O mundo precisa de formatos de intercâmbio de dados, como XML e JSON, para trocar dados entre sistemas muito diferentes, uma vez que o “mundo real da tecnologia é constituído de um vasto número de diferentes linguagens, costumes e arquiteturas”. A capacidade de comunicação entre estes sistemas exclusivos é essencial para muitas empresas e organizações. A utilização de um formato único para os dados leva ao emprego de um único tradutor e, por consequência, à economia de tempo e recursos computacionais.

Além de JSON, existem outros formatos de intercâmbio de dados, como *Extensible Markup Language* (XML), ou formatos mais tabulares, como valores separados por vírgulas (CSV), entre outros. A decisão por qual formato de dados aceitar para a comunicação muitas vezes é influenciada e tem a ver com a forma como o formato de dados refere-se aos costumes, língua, e arquitetura.

Apesar de JSON ter nascido de um subconjunto de *JavaScript*, uma vez que é baseado em objeto *JavaScript* de literais, o espírito de um formato de intercâmbio de dados é para ser independente da linguagem. Retirando o termo *JavaScript* (para desvincular da linguagem), teríamos um formato de intercâmbio de dados “notação de objeto”. Onde “objeto” é um conceito de programação comum, muito utilizado na Programação Orientada a Objetos, e é semelhante à forma como se descreveria um objeto do mundo real, e “notação” implica em um sistema de caracteres para representar dados como números ou elementos (BASSETT, 2015).

O objetivo principal de um formato de intercâmbio de dados é a portabilidade, ou seja, a compatibilidade na transferência de informações entre plataformas e sistemas, no entanto, JSON, assim como outros formatos, não cobrem valores atípicos. Pode-se considerar JSON uma linguagem independente que representa os dados de uma forma que se comunica com elementos comuns de muitas linguagens de programação. A forma como os dados são representados, tais como números e palavras, até mesmo as linguagens de programação que não são orientadas a objetos podem aceitar este formato que aplica conceitos de programação universais. Atualmente, encontram-se disponíveis bibliotecas JSON ou suporte nativo para JSON em várias linguagens de programação e sistemas (BASSETT, 2015; SILVA 2009).

Atribui-se a invenção do JSON a Douglas Crockford, um especialista em *JavaScript*, residente na Califórnia, nos Estados Unidos. Contudo, Douglas afirma que ele apenas o descobriu e mostrou quão útil JSON pode ser, dando-lhe um nome e uma gramática, formalizados no RFC 4627. Foi apresentado oficialmente na Conferência XML 2006 realizada em Boston, nos Estados Unidos, de 5 a 7 de dezembro de 2006. Em dezembro de 2009, JSON foi oficialmente reconhecido como um padrão ECMA, ECMA-404, e é agora um aspecto nativo da padronização de ECMAScript-262, 5ª edição (SILVA, 2009; SMITH, 2015).

Crockford documentou a gramática de JSON em <http://json.org>, em 2001, e rapidamente espalhou-se como uma alternativa para o formato de dados XML. Com a adoção generalizada de AJAX (*Asynchronous JavaScript and XML*), a popularidade de JSON começou a subir, as pessoas começaram a notar a sua facilidade de implementação e como ela concorria com a XML. JSON veio para simplificar e agilizar as tarefas em requisições AJAX, oferecendo uma alternativa a *JavaScript* para XML. Devido à sua simplicidade gramatical, o que torna JSON altamente interoperável, é considerado o formato de dados de fato da *Web* (SMITH, 2015; SILVA, 2009).

De acordo com Crockford, sua inovação foi apenas uma de uma longa história de inovações tecnicamente equivalentes, mas sintaticamente diferentes, maneiras “naturais” para representar as estruturas de dados que usam linguagens de programação (SEVERANCE, 2012).

Ainda nas considerações de Severance (2012), “as estruturas mais comuns que usamos na programação são variáveis escalares, listas lineares, e pares de chave-valor”. JSON representa estas estruturas na serialização de modo mais natural e direta, reduzindo o consumo de recursos computacionais entre as estruturas em memória em aplicativos e o formato de serialização. Uma implementação JSON em JavaScript lhe dá uma vantagem distinta sobre outros formatos de serialização, como XML, quando se trabalha com aplicações parcialmente escritas em JavaScript. Pode-se afirmar que o sucesso de JSON dá-se ao fato de que as estruturas de dados, por ele representadas, são exatamente as mesmas estruturas de dados que as linguagens de programação representam.

JSON é baseado em literais de objeto de *JavaScript*. Em programação, a palavra “literal” é um substantivo, um valor que é representado literalmente com dados e é escrito precisamente, uma vez que se destina a ser interpretado. Entretanto, o intercâmbio de dados

com JSON exclui as funções de literais de objeto de JavaScript. A representação de propriedades, utilizando-se os diferentes tipos de dados, é conseguida com pares nome-valor (BASSETT, 2015).

O conceito de pares nome/valor é generalizada na computação e são também referenciados como pares chave/valor, pares atributo/valor, e pares de campo/valor. Em um par nome/valor, primeiro declara-se o nome e depois o valor. Numa estrutura JSON, o valor pode ser uma *string*, um número, um booleano, nulo, uma matriz ou um objeto. O caractere dois pontos (:) é utilizado para separar os nomes e os valores, de modo que o nome é sempre grafado à esquerda e o valor é sempre à direita (BASSETT, 2015).

Parafraseando Silva (2009) “a sintaxe JSON, tal como a da XML, é rígida e não admite nenhum tipo de erro”. Recomenda-se a validação da sintaxe de estruturas em JSON pelo fato de uma eventual falha não retornar nenhuma mensagem de erro, muitas vezes dificultando sua detecção.

Uma abordagem objetiva sobre a sintaxe JSON é encontrada em Bassett (2015) da qual extraiu-se os ensinamentos empregados na implementação do protótipo para obtenção de um objeto JSON que contenha os metadados da fonte bibliográfica identificada por um DOI.

A sintaxe JSON apropriada requer que o nome, do par nome/valor, seja grafado sempre entre aspas duplas e pode ser qualquer *string* válida. Deve-se evitar espaço em branco ou caractere especial (que não seja a-z, A-Z ou 0-9) no nome, a fim de manter a máxima portabilidade, uma vez que na maioria dos casos, é considerada boa prática de programação a utilização dos caracteres do alfabeto (A-Z ou a-z), apenas.

Ao contrário do nome do par nome/valor, o valor nem sempre será grafado entre aspas duplas. Se o nosso valor é um dado tipo *string*, temos de colocá-lo entre aspas e os demais tipos de dados restantes de JSON (*object*, *number*, *boolean*, *null* e *array*) não serão cercados por aspas duplas.

Para torná-lo um objeto, uma estrutura JSON deve estar envolta por um par de chaves. Se um objeto JSON conter vários pares de nome/valor, estes serão separados por uma vírgula.

Quando uma estrutura em JSON é lida por uma máquina, qualquer um dos caracteres abaixo, que não esteja entre aspas duplas (fora de um valor de *string*), será assim interpretado e executado:

- { (chave de abertura) - começa um objeto;
- } (chave de fechamento) - finaliza um objeto;
- [ (colchete de abertura) - começa um *array*;

- ] (colchete de fechamento) - finaliza um *array*;
- : (dois pontos) - separa um nome e um valor em um par nome/valor;
- , (vírgula) - separa os pares nome/valor em um objeto ou separa os valores em um *array*.

Como um formato de intercâmbio de dados, JSON pode ser um documento armazenado em um sistema de arquivos, cuja extensão é de fácil lembrança: “.json”. Para cumprir a restrição de representação numa interface uniforme, o tipo MIME para JSON é “application/json” (BASSETT, 2015).

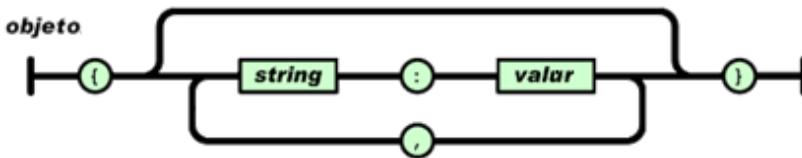
Para Silva (2009) e Smith (2015) o formato JSON foi criado para suportar duas estruturas, a saber:

- um conjunto de pares nome/valor (ou ainda chave/valor) não ordenados, que, em linguagem de programação, caracterizam um objeto ou *array* associativo;
- uma lista ordenada de valores também conhecida como *array*.

Esclarece Smith (2015) que Crockford descreveu as duas representações estruturais de JSON através de uma série de diagramas de sintaxe, dos quais destacam-se os dois abaixo descritos.

A Figura 7 ilustra a representação gramatical para uma coleção de pares de string/valor.

Figura 7 – Diagrama da sintaxe de uma coleção de pares de string/valor.



Fonte: Smith (2015, p. 56).

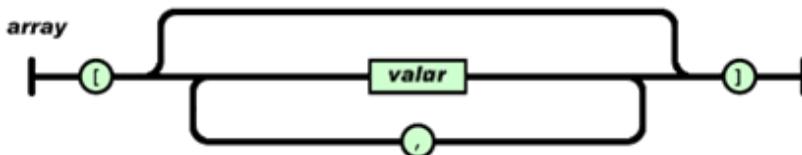
O diagrama descreve, uma coleção que começa com a utilização da chave de abertura ({), e termina com a utilização da chave de fechamento (}). O conteúdo do conjunto pode ser composto por qualquer um dos três possíveis caminhos designados:

- o caminho superior ilustra que a coleção pode ficar desprovida de qualquer par de string/valor;
- o caminho central ilustra que a coleção pode ser a de um único par de string/valor;
- o caminho inferior ilustra que, após a inclusão de um único par de string/valor, a coleção permite a adição que qualquer número de

pares de string/valor, antes de chegar ao fim. Cada par de string/valor é separado um do outro por meio de uma vírgula (,).

A Figura 8 ilustra a representação gramatical para uma lista ordenada de valores. Verifica-se que uma lista ordenada começa com o uso do colchete de abertura ([) e termina com o uso do colchete de fechamento (]).

Figura 8 – Diagrama da sintaxe de uma lista ordenada.



Fonte: Smith (2015, p. 57).

O conteúdo da lista pode ser determinado por qualquer um dos três possíveis caminhos designados:

- o caminho superior mostra que a lista pode permanecer vazia;
- o caminho central ilustra que lista ordenada pode possuir apenas um valor singular;
- através do caminho inferior verifica-se que a lista pode possuir qualquer número de valores, e que estes devem ser separados com o uso de uma vírgula (,).

O formato de dados JSON expressa ambos, os objetos e as matrizes (*arrays*), na forma de literal. Na verdade, JSON usa literais para capturar todos os valores *JavaScript*, exceto para o objeto *Date*, uma vez que carece de uma forma literal (SMITH, 2015).

Do exposto nesta Seção, optou-se pela utilização do formato de intercâmbio de dados JSON na construção do protótipo pelos motivos acima descritos e porque todas as Autoridades de Registro DOI disponibilizam, através de uma interface uniforme REST, a hipermídia do tipo “application/json”.

### 4.3 CLIENT URL REQUEST LIBRARY (CURL)

Nesta Seção procura-se esclarecer alguns tópicos que gravitam em torno de uma tecnologia computacional que permite a transferência de dados de forma simples e que foi utilizada no desenvolvimento do protótipo que viabiliza a obtenção dos metadados vinculados a um

determinado DOI. Assim sendo, fecha-se o ciclo de tecnologias computacionais utilizadas neste empreendimento.

Curl é o nome de um projeto que produz dois produtos: libcurl e curl. O nome nasceu de uma alusão a “*Client for URLs*”, originalmente com o URL escrito em letras maiúsculas para torná-lo óbvio que lida com URLs. O fato de ele também pode ser pronunciado “ver URL” (a pronúncia da letra c em inglês é muito semelhante a do verbo “see” – que significa ver em português) também ajudou, ele funciona como uma abreviação para “*Client URL Request Library*” ou por que não a versão recursiva: “*Curl URL Request Library*” (CURL, 2016; STENBERG, 2015 e 2016).

O projeto foi inicialmente desenvolvido por Daniel Stenberg, em 1996, o qual se mantém líder do mesmo até os dias atuais e já conta com mais de 1300 colaboradores. Na época Daniel ocupava seu tempo livre escrevendo um *bot* IRC para a aquisição de taxas de câmbio, tão precisas quanto possível, de uma página *Web* que as hospedava. Ele só precisava automatizar a sua recuperação. Aproveitou a existência de uma ferramenta de linha de comando, de código aberto, o *HttpGet*, que o brasileiro Rafael Sagula tinha escrito e, recentemente, lançado a versão 0.1. Fez alguns pequenos ajustes e obteve exatamente o que ele precisava. O dia 20 de março de 1998 é tido como o dia do nascimento de Curl (GIBBS, 2008; STENBERG, 2016).

### 4.3.1 Biblioteca libcurl

A libcurl é uma biblioteca de transferência de URL do lado do cliente, gratuita e de fácil utilização, dando apoio a uma extensa gama de protocolos e suportando vários certificados, tais como: HTTPS, HTTP POST, HTTP PUT, upload FTP, Kerberos SPNEGO, formulário baseado em HTTP de *upload*, *proxies*, *cookies*, autenticação de usuário e senha, tunelamento *proxy* HTTP e vários outros. Sua portabilidade é de fácil verificação, pelo fato de sua construção e funcionamento ocorrerem de forma idêntica em várias plataformas, incluindo Solaris, NetBSD, FreeBSD, OpenBSD, Darwin, HP-UX, IRIX, AIX, Tru64, Linux, UnixWare, Hurd, Windows, Amiga, OS/2, BeOS, Mac OS X, Ultrix, QNX, OpenVMS, RISC OS, Novell NetWare, DOS, OSF, Android, Minix, IBM TPF e muito mais (CURL, 2016; GIBBS, 2008).

A libcurl é totalmente livre, *thread* segura, compatível com IPv6, rica em recursos, com suporte estruturado em listas de discussão, rápida e que dispõe de uma interface fácil para uma série de protocolos comuns da Internet. É uma biblioteca de transferência de arquivos baseada em C,

multiplataforma e, provavelmente, a mais utilizada pelos variados *softwares*, sejam eles *open source* ou comerciais (STENBERG, 2016).

Várias linguagens de programação tem feito a tradução de libcurl facilitando sua utilização pelos usuários destas, as vezes até de forma transparente. Uma das linguagens pioneiras na utilização deste artifício foi PHP, possibilitando o acesso às funções curl através dos aplicativos criados com esta linguagem de programação.

O módulo de PHP para implementar curl (ou PHP/CURL, ou ainda ext/curl) foi inicialmente escrito por Sterling Hughes e é capaz de executar várias solicitações usando o mesmo identificador, pelo menos na versão PHP 4.3.8 e superiores. Após uma transferência, define-se novas opções para fazer outra transferência e isso fará com que libcurl reutilize a mesma conexão, se houver a possibilidade (CURL, 2016).

Uma vez que PHP/CURL é um módulo que vem com o pacote PHP regular, sua utilização depende da instalação correta de libcurl. A adoção de libcurl pelo projeto PHP ocorreu, e se mantém atualmente, com o lançamento do PHP 4.0.2. (STENBERG, 2016).

Salienta Stenberg (2016) a importância de considerar que os produtos do projeto curl (curl e libcurl) tentam evitar o manuseio dos dados reais que são transferidos. Eles não tem, por exemplo, nenhum conhecimento sobre HTML ou qualquer outra coisa do conteúdo que é popular para transferir através de HTTP, mas sabe tudo sobre como transferir esses dados através de HTTP. Vale esclarecer que a ferramenta de linha de comando foi criada primeiro e dela foi extraído o motor de rede e convertido em uma biblioteca, no ano de 2000. A partir deste momento, a ferramenta de linha de comando tornou uma fina camada de lógica para fazer uma ferramenta que encapsula a biblioteca responsável por todo o trabalho pesado.

### **4.3.2 Curl, a ferramenta de linha de comando**

Curl é uma ferramenta de linha de comando para obter ou enviar arquivos usando a sintaxe URL. Usa libcurl e suporta toda a sua ampla gama de protocolos comuns da Internet. Curl foi concebido como uma ferramenta de linha de comando e vai manter-se sem uma GUI. Facilita sua saída para outro programa ou redireciona para outro arquivo para o próximo programa interpretar (GIBBS, 2008; CURL, 2016).

Segundo Stenberg (2016), curl possui implementações que compilam e são executadas em uma ampla gama de sistemas operacionais, incluindo todos os sistemas Unix modernos (além de vários mais antigos), Windows, Amiga, BeOS, OS/2, OS X, QNX etc.

Fato este que viabiliza sua utilização através de linhas de comando ou *scripts* para transferir dados. Ele também é usado em *car infotainment*, televisores, roteadores, impressoras, equipamentos de áudio, telefones celulares, tablets, players de mídia e é a espinha dorsal de transferência de dados na internet para milhares de aplicações de software que afetam totalmente mais de um bilhão de usuários.

Ensina Gibbs (2008) que ao usar curl em um comando ou janela de terminal, entra-se em uma linha de comando da seguinte maneira (este comando deve estar em uma única linha):

```
curl -u email:passw -d status="text" http://twitter.com/statuses/update.xml
```

Ao interpretar o comando recebido, curl descobre qual protocolo usar a partir do URL fornecido, e quando “http:” é especificado será, salvo indicação em contrário (usando 0), padrão para usar HTTP 1.1. O primeiro argumento, -u, instrui curl para usar os próximos dois parâmetros (separados por dois pontos) como o nome da conta e a senha para autenticação básica. O próximo argumento, -d, especifica os dados para ser enviado através de uma solicitação POST. O que curl espera é uma resposta ao pedido POST que deverá ser a página solicitada (update.xml) enviada para stdout. Esta página será codificada de acordo com o tipo de arquivo usado no pedido, assim como foi solicitado update.xml será recebido no formato XML.

Salienta Stenberg (2016), que nada obriga os usuários a referenciar o uso de curl ou libcurl em seus serviços ou produtos. Algumas empresas informam sobre sua utilização. No entanto, normalmente a descoberta é acidental, através de leitura sobre os diálogos, documentação e contratos de licença. A lista de empresas que estão usando curl e/ou libcurl, em ambientes comerciais, vai bem além de 200 nomes, dentre as quais destaca-se uma pequena seleção: Adobe, AOL, Apple, AT & T, BBC, Blackberry, BMW, Bosch, Chevrolet, Cisco, Facebook, Google, HP, Huawei, HTC, IBM, Intel, LG, Mercedes Benz, Motorola, Netflix, Nintendo, Oracle, Panasonic, Philips, Pioneer, RBS, Samsung, SanDisk, SAP, SAS Institute, SEB, Sharp, Siemens, Sony, Spotify, Sol, Swisscom, Tomtom, Toshiba, Vmware, Yahoo, Yamaha.

Neste capítulo fez-se uma rápida abordagem das tecnologias computacionais empregadas na construção do protótipo utilizado na obtenção dos metadados vinculados ao DOI informado pelo usuário do MORE e que serão utilizados para preenchimento dos campos do formulário específico utilizado na elaboração e geração da referência bibliográfica.

## 5 PROCEDIMENTOS METODOLÓGICOS

Este capítulo tem por objetivo fazer uma descrição das tarefas que foram planejadas e realizadas durante a consecução desta pesquisa. Salienta-se que algumas etapas antecederam o desenvolvimento propriamente dito e dentre elas destaca-se a elaboração do projeto de pesquisa.

Para a consolidação da pesquisa foi utilizado o sistema computacional MORE para validar as questões de pesquisa. O MORE conta com mais de 160.000 usuários cadastrados e cerca de 6.000.000 de referências bibliográficas armazenadas. Pressupõe-se que o subsistema desenvolvido e outras TICs podem impactar na usabilidade do MORE e influenciar a utilidade percebida e a satisfação do usuário.

Cabe esclarecer, que no contexto deste trabalho, os termos acima possuem os seguintes significados:

- Usabilidade é um aspecto da qualidade do sistema e relaciona-se a facilidade encontrada pelo usuário na utilização deste;

- Utilidade Percebida é o grau de acreditar que o desempenho da tarefa é reforçado pela utilização de um sistema específico (RAI et al., 2002). Assim, os usuários que tenham um determinado nível de utilidade percebida acreditam que existe uma relação positiva entre o uso e o desempenho;

- Satisfação do Usuário é o nível de satisfação que um usuário tem com um sistema em relação ao que o usuário espera após a primeira utilização do sistema (PARK et al., 2011).

Somado a estas considerações, verificou-se que o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), através de sua Plataforma Lattes, vem utilizando identificadores digitais de objetos para auxiliar no preenchimento e na validação de produção científica ali cadastrada.

O panorama aqui apresentado contribuiu para o surgimento de uma série de questionamentos sobre a utilização das Tecnologias de Informação e Comunicação e seu impacto sobre a percepção do usuário do MORE a respeito desta utilização. Após uma reflexão feita a luz de trabalhos recentes sobre a utilização de identificadores digitais de objetos, optou-se por desenvolver um sistema, que integrado ao MORE, favoreça a gestão e a utilização dos metadados vinculados a um nome DOI<sup>®</sup>, cuja estrutura de suporte é gerida e mantida pela IDF, suas Autoridades de Registro e produtores associados. Lembra-se que, atualmente, a maioria dos associados a IDF são editoras de periódicos

científicos, portanto, geradores de recursos informacionais cujos metadados são de grande interesse da Biblioteconomia.

Após verificadas estas questões decidiu-se pelo desenvolvimento de uma pesquisa para verificar junto aos usuários do MORE a percepção destes em relação a utilização das TICs na tarefa de gerenciamento e utilização de metadados vinculados a um nome DOI<sup>®</sup>, que identifique um recurso informacional de interesse da Biblioteconomia e, por consequência, da Ciência da Informação, da Ciência da Computação e das TICs. A partir deste momento foi elaborado o projeto de pesquisa para orientar na condução das tarefas atinentes a este trabalho.

Para a execução deste trabalho foi, inicialmente, buscado na bibliografia existente o estado da arte sobre Recuperação da Informação, Identificação de Recursos de Informação e Tecnologias Computacionais disponíveis para dar suporte aos Sistemas de Recuperação da Informação.

Verifica-se na bibliografia consultada que os autores concordam que a pesquisa bibliográfica é aquela que se utiliza de publicações existentes, através da identificação, localização e compilação dos dados disponíveis em artigos científicos (impressos e/ou *online*), livros, publicações dos órgãos oficiais, teses, periódicos especializados, etc. (PÁDUA, 2012; FARIAS FILHO; ARRUDA FILHO, 2013; SEVERINO, 2007; CERVO; BERVIAN; SILVA, 2007).

Em consonância com os autores supramencionados buscou-se o entendimento da variação semântica dos termos abarcados pela recuperação da informação quando utilizados nas mais diversas áreas do conhecimento, tendo em vista a interdisciplinaridade do tema. Inicialmente busca-se uma definição para o termo informação e suas variações contextuais. Busca-se na Ciência da Informação e na Biblioteconomia o suporte teórico e doutrinário para a representação documentária, através de suas linguagens próprias. Faz-se um levantamento das possibilidades oferecidas pela recuperação da informação propriamente dita desde os primeiros ensaios dos catálogos até os dias atuais e suas folksonomias, amplamente utilizadas no ambiente *Web 2.0*.

Posteriormente, procura-se saber como os recursos de informação estão sendo identificados e verifica-se o uso intensivo da metainformação (metadados), cuja categorização e criação de padrões são evidenciados na bibliografia disponível. Destaca-se que os metadados oferecem o suporte necessário a utilização de identificadores de objetos. Sua utilização em sistemas de informação interligados por

rede possibilita o uso de identificadores digitais de objetos, dos quais destaca-se o DOI, o UCI e o cID.

As tecnologias computacionais utilizadas na implementação do protótipo também foram objeto de estudo. O motivo pelo qual o estilo arquitetônico REST foi estudado é devido a sua utilização pelas Autoridades de Registro DOI na implementação de uma *Application Programming Interface* (API) REST que possibilita o fornecimento dos metadados referentes a um determinado nome DOI, mediante uma requisição que especifica um tipo MIME. Para esta requisição utilizou-se as funções do módulo PHP/CURL que oferece todas as possibilidades levantadas na Seção 4.3. O tipo MIME requerido é o “*application/json*” estudado na Seção 4.2.

Na sequência, foi desenvolvido um protótipo de sistema que viabiliza a aquisição dos metadados referentes ao recurso informacional identificado por um nome DOI. Este sistema, após a obtenção de toda a estrutura de metadados intercambiada com a AR, seleciona e utiliza aqueles metadados interessantes à elaboração da referência bibliográfica relativa ao recurso identificado.

O protótipo foi integrado ao MORE e submetido à utilização do público em geral, para ser avaliado, em situação normal de produção, através de um questionário *online*, de 10 (dez) questões fechadas e 01 (uma) questão aberta, disponibilizado ao usuário que fizer uso do recurso. A abordagem empregada teve a finalidade de verificar o grau de satisfação do usuário no momento de sua utilização.

No entendimento de Rai et al. (2002) a satisfação do usuário pode ser medida direta ou indiretamente. No tocante ao questionário utilizado nesta pesquisa as 08 (oito) primeiras questões foram dedicadas à medida indireta e as 02 (duas) últimas à medida direta.

Segundo Vieira (2009) questionário “é um instrumento de pesquisa constituído por uma série de questões sobre determinado tema”. Cabe aos participantes da pesquisa responder às questões apresentadas e restituir o questionário devidamente preenchido ao entrevistador.

As respostas às questões fechadas do questionário utiliza uma escala de *Likert* de 6 (seis) níveis, evitando a tendência a posicionamentos centrais. Escala de Likert foi a forma proposta por Rensis Likert para mensurar atitudes, opiniões e avaliações. Muito utilizada nas Ciências Sociais, geralmente como resposta a questões fechadas e escalonadas de acordo com a intensidade da percepção do entrevistado (GÜNTHER, 2003).

A opção pela apresentação *online* do questionário deu-se em virtude de todos os usuários do MORE dispor de acesso à *Web* e das facilidades oferecidas pelo *Google Drive*<sup>®</sup> através do *Google Form*<sup>®</sup> cujas respostas são registradas automaticamente em uma Planilha *Google*<sup>®</sup> que pode ser baixada e seus dados tratados *offline*. Considera Vieira (2009) que o levantamento de dados pela Internet apresenta as seguintes vantagens: “os questionários são facilmente distribuídos, a coleta e o processamento dos dados são rápidos”. No âmbito desta pesquisa a desvantagem de que a “pesquisa fica restrita aos usuários da Internet, que não representam a população em geral” não se aplica, uma vez que todos os usuários do MORE necessitam acesso a rede para fazer uso do sistema.

O Apêndice B replica o questionário apresentado aos usuários do MORE cujas respostas contribuíram para o êxito deste trabalho. Os enunciados das questões fechadas são frases afirmativas às quais os usuários apontaram o grau de concordância para cada uma delas, variando de “Concordo Totalmente” (5) a “Discordo Totalmente” (0), inclusive. Os enunciados das questões fechadas são listados a seguir:

1. O tempo utilizado para preencher o formulário diminuiu;
2. A utilização do DOI facilitou o preenchimento do formulário;
3. A utilização do DOI tornou mais prática a geração da referência bibliográfica;
4. Havendo oportunidade, voltarei a utilizar um DOI para o autopreenchimento do formulário;
5. Sou favorável à implementação de funcionalidades deste tipo;
6. A utilização da funcionalidade apresentada melhora minha produtividade;
7. Gostaria de contar com esta funcionalidade para outros tipos de fontes bibliográficas;
8. O autopreenchimento de alguns campos do formulário ajuda a diminuir a ocorrência de erros de ortografia;
9. Fiquei satisfeito com a utilização da funcionalidade de autopreenchimento;
10. Geralmente fico satisfeito com as novas funcionalidades implementadas pela equipe do MORE.

A questão aberta proporciona um espaço para críticas e sugestões que possam vir a melhorar aspectos relativos à ferramenta utilizada.

Devido a dificuldade em contatar toda a população alvo desta pesquisa, uma vez que conta com mais de 160.000 usuários, optou-se pelo expediente de utilizar uma amostra desta população. Segundo

Paranhos et al. (2014) “existem duas principais razões para utilizar amostras na pesquisa científica: economia de tempo e economia de recursos”.

Esclarece, ainda, Paranhos et al. (2014) que o cálculo do tamanho da amostra para uma população superior a 100.000 deve-se utilizar a fórmula para populações infinitas, conforme a equação (1):

$$n = \frac{\sigma^2 * p * q}{E^2} \quad (1)$$

Onde  $n$  é o tamanho da amostra,  $\sigma$  representa o nível de confiança escolhido pelo pesquisador (em geral 90%, 95% ou 99%),  $p$  representa a proporção das características pesquisadas no universo (%),  $q$  representa a proporção do universo que não possui a característica de interesse ( $q = 1 - p$ ) e  $E$  representa o erro de estimação permitido (em geral, 2%, 3% ou 4%).

Pela equação apresentada por Paranhos et al. (2014), para uma população superior a 100.000, desejando-se um nível de confiança de 99% e um erro de estimação permitido de 2%, para o pior caso de relação entre  $p$  e  $q$ , o tamanho da amostra deve ser superior a 612.

Pelo exposto, entende-se que a amostra deste trabalho pode ser considerada representativa uma vez que conta com 1.033 questionários correta e espontaneamente respondidos no período de 18/05/2015 a 04/11/2015.

Encerrado o período destinado a coleta dos dados, através da aplicação do questionário supracitado, os dados adquiridos foram tratados e analisados conforme estudo desenvolvido no capítulo 7, desta dissertação.

A análise desenvolvida no capítulo 7 forneceu subsídios que foram determinantes para o cumprimento do objetivo geral. Salienta-se, ainda, que as conclusões a que se chegou derivam da realização desta análise.

A medida que as tarefas eram realizadas, anotações eram feitas e o relatório era escrito para posteriores revisões, pelo autor e por seu orientador, até entender-se estar em condições de apreciação por uma banca examinadora.



## 6 DESENVOLVIMENTO DO PROTÓTIPO

A construção do protótipo destinado ao intercâmbio de metadados com as Autoridades de Registro da *International DOI Foundation, Inc.* (IDF) ocorreu em ambiente de laboratório. A finalidade deste protótipo é suprir as necessidades de dados no autopreenchimento dos campos do formulário de geração de referências bibliográficas a artigo de revistas (*Journals*).

Utilizou-se o *Digital Object Identifier (DOI)*<sup>®</sup> na construção do protótipo porque é o identificador que abrange o maior número de objetos de Biblioteconomia, atualmente conta com mais de 77 milhões de enlaces. A composição deste identificador é de fácil entendimento e segue regras bem definidas pela *International DOI Foundation (IDF)* (WEBER, 2013).

Para dar suporte ao subsistema proposto e viabilizar sua execução em ambiente de produção utilizou-se o MORE. A escolha deste sistema deve-se ao fato de ter sido trabalhado em Felisberto (2013), estar em produção há mais de 12 anos, cobrir um grande número de usuários ligados a educação e a pesquisa, além de outras características citadas neste trabalho.

Segundo Alves, Mendes e Alves (2006) as dificuldades encontradas pelos usuários de biblioteca na utilização da NBR 6023 (a norma para elaboração de referências) motivou-os a construir “uma ferramenta gratuita e on-line, que pudesse auxiliar alunos, professores e demais usuários, na tarefa de elaborar e organizar referências”.

O desenvolvimento do MORE deu-se com a utilização de ferramentas licenciadas como softwares livres, cuja comunidade de desenvolvimento de softwares tem utilizado e dado contínuo suporte, além de oferecer frequentes atualizações.

Todo o projeto MORE segue o princípio da inclusão social em todos os seus aspectos e em virtude disso destina-se, fundamentalmente, a uma parcela da sociedade que necessita uma ferramenta de gerenciamento de referências bibliográficas gratuita, na língua portuguesa, cuja formatação siga a NBR 6023 (2002) e que mantenha suas coleções de referências armazenadas e disponíveis na *Web*.

O crescimento do número de usuários cadastrados e o conseqüente aumento no número de acessos diários foram decisivos na tomada de decisão no sentido de reestruturá-lo em 2013. Na oportunidade passou por uma modernização que resultou em sua modularização em três seguimentos, a saber: módulo de suporte, módulo de controle de referências e módulo de controle de usuários. No ano de

2014 atualizou-se o *framework* CSS3 utilizado em seu desenvolvimento e foi adicionada responsividade a todas as páginas, permitindo acesso ao MORE através de dispositivos móveis.

O módulo de suporte preocupa-se em manter o usuário informado e em dirimir dúvidas. Neste módulo foram incluídas páginas referentes a “Ajuda”, “Links Úteis”, “Informações sobre o MORE”, “Contato”, “FAQ” e um “Tutorial”. O formulário de “Contato”, mostrado na Figura 9, está disponível para esclarecimento de dúvidas e/ou sugestões individuais.

Figura 9 – Tela do MORE referente ao formulário de contato.

The image shows a screenshot of a web browser displaying the contact form of the MORE system. The browser's address bar shows the URL '127.0.0.1/more/suporte/contato'. The page header includes the MORE logo and the title 'Mecanismo Online para Referências'. A navigation menu contains links for 'Início', 'Criar Referências', 'Pesquisar', 'Minha Conta', 'Ajuda', 'Fórum', 'Blog', 'Links', 'Sobre o MORE', and 'Contato'. The main content area is titled 'FALE CONOSCO' and features a form with three input fields: 'Nome', 'E-mail', and 'Mensagem'. There are two buttons at the bottom of the form: 'Enviar Mensagem' and 'Cancelar Mensagem'. The browser's taskbar at the bottom shows various application icons and the system clock indicating the date 17/10/2013.

Fonte: *print screen* da aplicação em execução no navegador *Mozilla Firefox*.

O módulo de controle de referências destina-se ao processo de formatação e geração das referências bibliográficas. Durante este trabalho o módulo em tela sofreu as atualizações necessárias a integração do subsistema responsável por promover o intercâmbio dos metadados com as AR da IDF.

Cada tipo de fonte bibliográfica possui formulário próprio. Neste trabalho a fonte em evidência é a referente a artigos de periódicos, cuja respectiva referência é elaborada através de formulário próprio apresentado ao usuário e mostrado na Figura 10. Salienta-se que o MORE verifica se todos os campos obrigatórios foram preenchidos e faz algumas correções da escrita quanto a forma, tais como: coloca o sobrenome dos autores em letras maiúsculas e no início, separado do

restante do nome por vírgula; faz abreviaturas dos meses; capitaliza (somente a primeira letra maiúscula) os títulos; etc.

Figura 10 – Tela do MORE referente ao formulário de referência a artigo de revistas.

The screenshot shows a web application interface for creating a reference. At the top, there is a navigation bar with links: Início, Criar Referências, Pesquisar, Minha Conta, Ajuda, Links, Sobre, Contato, Tutorial, and FAQ. Below the navigation bar, the current location is indicated as 'Sua localização: Artigo\_revista » Inserir Artigo Revista'.

The main form is titled 'ELABORAR REFERÊNCIA A ARTIGO DE REVISTAS'. It contains the following fields and sections:

- DOI Field:** A text input field with the example 'Ex: 10.1500/1981.8344/1915' and a 'Buscar' button.
- Escolha uma Coleção de Referências:** A dropdown menu currently showing 'Dissertação'.
- Título do Artigo \*:** A text input field with a sub-label 'A geração e o gerenciamento de referência:'.
- Nº do Fascículo \*:** A text input field with the value '42'.
- Nº do Volume \*:** A text input field with the value '20'.
- Tipo do Autor:** A dropdown menu currently showing 'Pessoa(s) Física(s)'.
- Subtítulo do Artigo:** A text input field with a sub-label 'Se for nome próprio a 1ª letra será maiúscula:'.
- Periodicidade:** A text input field with the example 'Ex: semanal'.
- Responsabilidade Intelectual:** A dropdown menu currently showing 'Autor da Obra'.
- Título da Revista:** A text input field with a sub-label 'Enc. Bibl:'.
- Diã, Mês \*, Ano \*:** Three text input fields with values '14', '4', and '2015' respectively.
- Mais de 3 Autores:** A dropdown menu currently showing 'Sim'.
- Subtítulo da Revista:** A text input field with the example 'R. Eletr. Bibliotecon. Ci. Inf.'.
- Página Inicial \*:** A text input field with the value '79'.
- Página Final \*:** A text input field with the example 'Ex: 345'.
- Autor \*:** A text input field with the value 'Proxério Manoel Felsberto'.
- Local (Cidade) \*:** A text input field with the value '[S.l.]'.
- Notas:** A text input field with the value 'Universidade Federal de Santa Catarina (UF'.

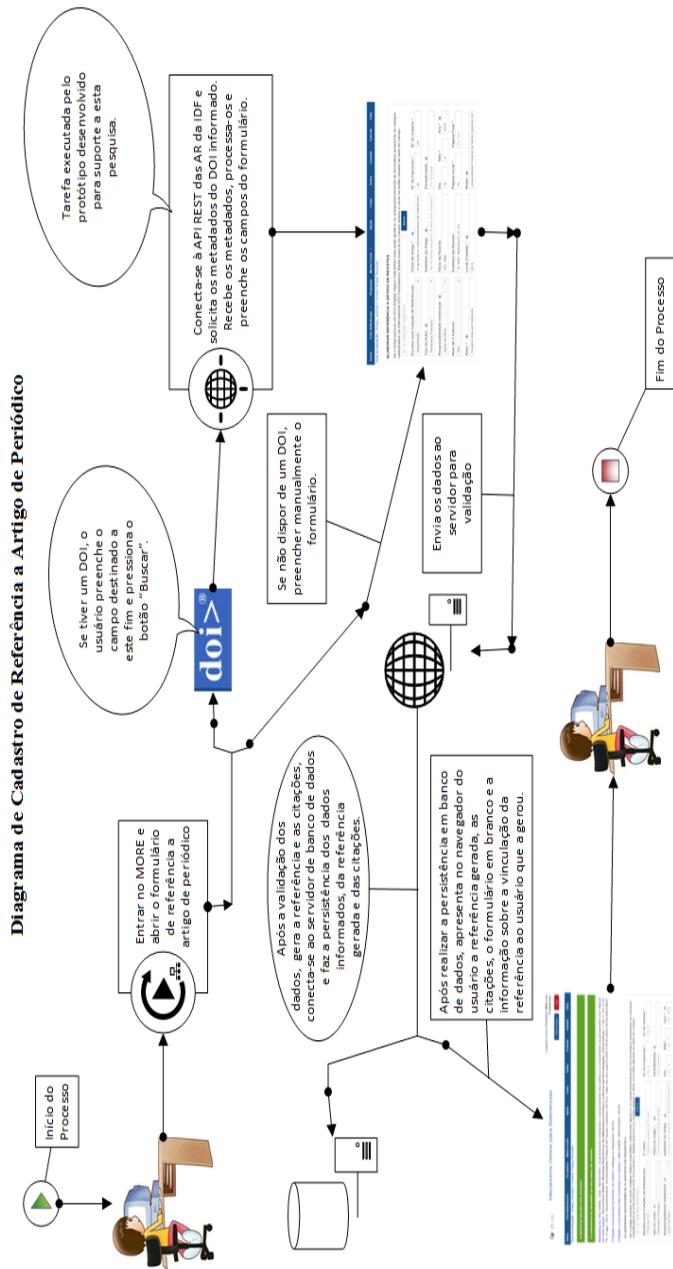
Fonte: *print screen* da aplicação em execução no navegador *Google Chrome*.

O Diagrama de Cadastro de Referência a Artigo de Periódico apresentado na Figura 11 possibilita, de maneira didática, a identificação da tarefa realizada pelo protótipo e em que parte do processo ele está inserido.

Para um entendimento mais técnico um diagrama de máquina de estado referente ao processo de geração de referência a artigo de periódico foi elaborado para que se tenha uma visão geral de todo o processo e do ponto de referência onde ocorre a interação do protótipo com os repositórios externos, além de todas as interações referente ao processo representado. Estado pode ser entendido como uma situação durante o ciclo de um processo, em que se espera a ocorrência de um evento, ou então, executa uma tarefa em resposta a um evento, ou ainda, satisfaz a uma determinada condição.

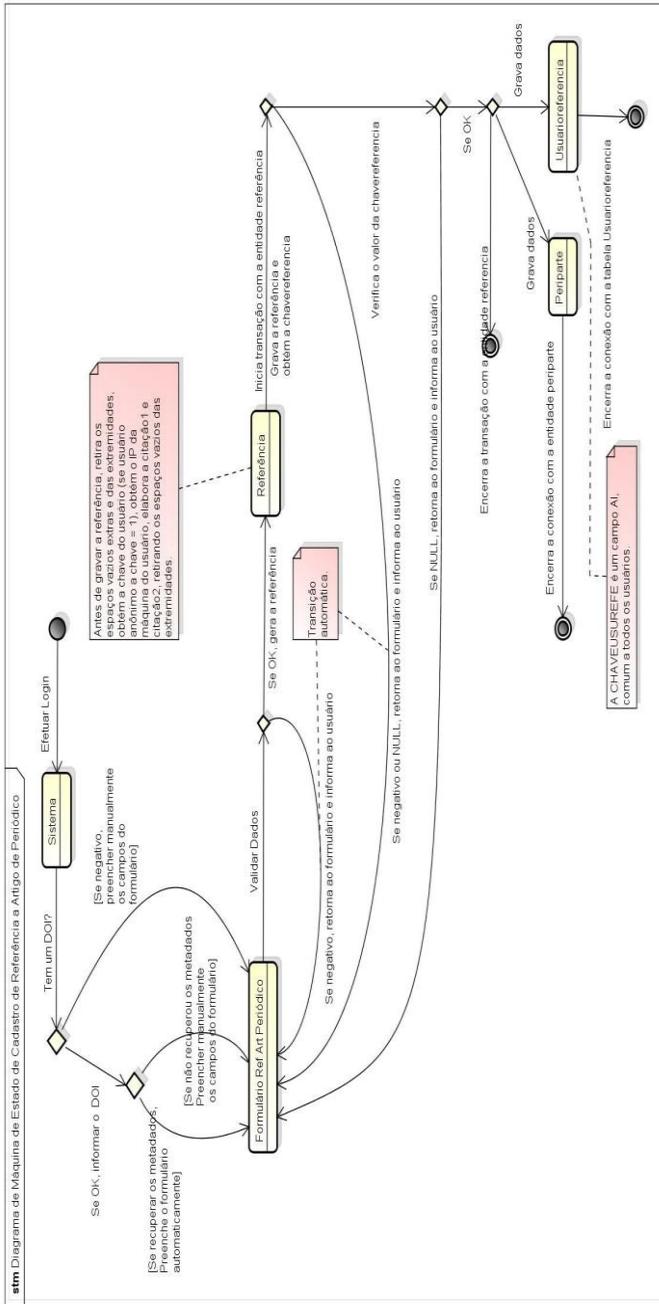
A Figura 12 mostra o Diagrama de Máquina de Estado de Cadastro de Referência a Artigo de Periódico onde pode ser verificada a posição do subsistema responsável por intercambiar metadados de um nome DOI com as AR da IDF em relação ao processo de geração de referência.

Figura 11 – Diagrama de Cadastro de Referência a Artigo de Periódico.



Fonte: elaborado pelo autor

Figura 12 – Diagrama de Máquina de Estado de Cadastro de Referência a Artigo de Periódico.

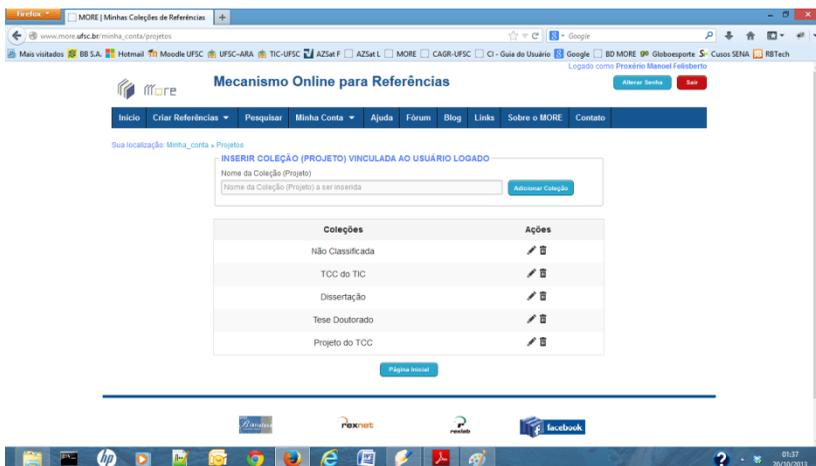


Fonte: elaborado pelo autor

O módulo de controle de usuários está voltado para as funcionalidades que envolvem a autenticação para que seja possível acessá-las em sua plenitude. Encontram-se no canto superior direito de cada página o nome do usuário autenticado, um enlace para alterar senha e um enlace para fechar a sessão, se estiver autenticado. Caso o usuário não tenha realizado sua autenticação o nome mostrado será “Anônimo” e apenas o enlace de *login* será apresentado.

Destaca-se a funcionalidade “Pesquisar” que será objeto de algumas considerações, ainda neste capítulo. No menu “Minha Conta” encontram-se as funcionalidades “Gerenciar Referências”, “Criar e Editar Coleções”, “Cadastrar Usuário” e “Atualizar Dados Usuário”. Na página acessada através do enlace “Criar e Editar Coleções” é permitido, ao usuário, inserir uma nova coleção (projeto), editar ou excluir uma coleção existente, conforme pode ser visto na Figura 13.

Figura 13 – Tela do MORE de inclusão e gerenciamento de coleções.



Fonte: *print screen* da aplicação em execução no navegador *Mozilla Firefox*.

Através do enlace “Gerenciar Referências”, conforme apresentado na Figura 14, o usuário tem acesso a suas referências de acordo com a coleção a qual estão vinculadas. As referências referentes a uma determinada coleção são carregadas em uma tabela na margem direita onde cada linha contém uma referência e os atalhos para edição e exclusão da mesma.

Figura 14 – Tela de gerenciamento de coleções pessoais.

Logado como **Proxério Manoel Feisberto**

 Mecanismo Online para Referências [Alterar Senha](#) [Sair](#)

**Início**   **Criar Referências** ▾   **Pesquisar**   **Minha Conta** ▾   **Ajuda**   **Links**   **Sobre**   **Contato**   **Tutorial**   **FAQ**

Sua localização: Minha\_conta > Gerenciar

### Minhas Coleções

- [TCC do TIC - 7 registros](#)
- [Dissertação - 101 registros](#)
- [Trabalho EDS - 4 registros](#)
- [Artigo Gestão da Tecnologia - 10 registros](#)
- [Artigo ERC - 8 registros](#)
- [Seminário Integrador - 4 registros](#)
- [Não Classificada - 10 registros](#)

### Dissertação

Referências	Ações
ALONSO, Jorge Alberto González; SANTACRUZ, Mauro Pazmiño. Cálculo e interpretación del Alfa de Cronbach para el caso de validación de la consistencia interna de un cuestionario, con dos posibles escalas tipo Likert. <b>Revista Publicando</b> , Quito, v. 2, n. 1, p. 62-77, 2015. Trimestral.	 
ALVES, Maria Bernadete Martins; MENDES, Leandro Luiz; ALVES, João Bosco da Mota. MORE: Mecanismo on-line para referências. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 14., 2006, Salvador. <b>Anais...</b> [s.l.], 2006. p. 1 - 12. Disponível em: <http://xa.yimg.com/kq/groups/25169972/1250144979/name/4.ALVES>. Acesso em: 12 mar. 2016.	 
ALVES, Rachel Cristina Vesú. <b>Metadados como elementos do processo de catalogação</b> . 2010. 134 f. Tese (Doutorado) - Curso de Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Cap. 7.	 
ALVES, Rachel Cristina Vesu; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. <b>Metadados no Domínio Bibliográfico</b> . Rio de Janeiro: Intertexto, 2013. 196 p.	 
ALVES, Rachel Cristina Vesu; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. METADADOS: ORGANIZAÇÃO E ACESSO À INFORMAÇÃO NO DOMÍNIO BIBLIOGRÁFICO. In: ENCONTRO INTERNACIONAL DE CATALOGADORES, 9., 2013, Rio de Janeiro. <b>Anais...</b> Rio de Janeiro: Fundação Biblioteca Nacional do Brasil, 2013. p. 1 - 15. Disponível em: <http://www.enacat.ufscar.br/index.php/eic-enacateic-enacat/paper/viewFile/52/22>. Acesso em: 17 abr. 2015.	 
ARAÚJO JÚNIOR, Rogério Henrique de; TARAPANOFF, Kira. Precisão no processo de busca e recuperação da informação: uso da mineração de textos. <b>Ciência da Informação</b> , Brasília, v. 35, n. 3, p. 326-247, dez. 2006. Quadrimestral.	 
ARAUJO, Vera Maria Araujo Pigozzi de. Sistemas de recuperação da informação: uma discussão a partir de parâmetros enunciativos. <b>Transinformação</b> , Campinas, v. 2, n. 24, p. 137-143, 2012. Bimestral.	 
ARELLANO, Miguel Angel. Preservação de documentos digitais. <b>Ciência da Informação</b> , Brasília, v. 33, n. 2, p. 15-27, ago. 2004. Quadrimestral.	 
ASSUMPÇÃO, Fabrício Silva; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Representação no domínio bibliográfico: um olhar sobre os Formatos MARC 21. <b>Perspectivas em Ciências da Informação</b> , Belo Horizonte, v. 20, n. 1, p. 54-74, maio 2015. Trimestral. FapUNIFESP (SciELO). <a href="http://dx.doi.org/10.1590/1981-5344/2054">http://dx.doi.org/10.1590/1981-5344/2054</a> .	 
BARITÉ, Mario et al. Garantia literária: elementos para uma revisão crítica após um século. <b>Transinformação</b> , Campinas, v. 22, n. 2, p. 123-138, ago. 2010. Quadrimestral.	 

1 2 3 4 5 6 Próximo Última

[Página Inicial](#)   [Enviar Referências por E-Mail](#)

---







Copyright © 2005-2013 todos os direitos reservados.  
Versão 2.0

Fonte: print screen do MORE em execução no navegador *Google Chrome*.

O protótipo, propriamente dito, e utilizado da recuperação da metainformação relativa ao recurso informacional identificado pelo DOI informado pelo usuário, trata-se de um *script*, escrito na linguagem PHP

que se utiliza das funções do módulo CURL implementado por esta. No momento que o usuário informa um nome DOI e o submete para efetuar a busca o protótipo dispara uma requisição contra os principais repositórios das Autoridades de Registro DOI solicitando os metadados da obra vinculada ao identificador informado e cujos dados devem estar contidos em um objeto JSON. A Figura 15 mostra um exemplo de objeto JSON recebido pelo protótipo ao disparar uma requisição informando o nome DOI® “10.5007/1518-2924.2015v20n42p79”. Ressalta-se que, apesar da figura mostrar o objeto em dez linhas (apenas para possibilitar a visualização do objeto na sua totalidade), não há quebra de linha no objeto, uma vez que é representado em uma única linha (sem quebra e sem nova linha).

Figura 15 – Exemplo de objeto JSON.

```
{
  "indexed": true,
  "date-parts": [
    [
      2015,
      12,
      20
    ]
  ],
  "date-time": "2015-12-20T10:12:55Z",
  "timestamp": 1450606375272,
  "reference-count": 0,
  "publisher": "Universidade Federal de Santa Catarina (UFSC)",
  "issue": "42",
  "DOI": "10.5007/1518-2924.2015v20n42p79",
  "type": "journal-article",
  "created": [
    [
      2015,
      4,
      23
    ]
  ],
  "date-time": "2015-04-23T11:43:46Z",
  "timestamp": 1429789426000,
  "page": "79",
  "source": "CrossRef",
  "title": "A geral\u00e7\u00e3o e o gerenciamento de refer\u00eancias bibliogr\u00e1ficas com o Mecanismo Online para Refer\u00eancias - MORE",
  "prefix": "http://dx.doi.org/10.5007",
  "volume": "20",
  "author": [
    [
      {
        "affiliation": [
          [
            "family": "Felisberto",
            "given": "Prox\u00edmo Manoel"
          ],
          {
            "affiliation": [
              [
                "family": "Marcelino",
                "given": "Roderval"
              ],
                {
                  "affiliation": [
                    [
                      "family": "Alves",
                      "given": "Maria Bernardete Martins"
                    ],
                    {
                      "affiliation": [
                        [
                          "family": "Alves",
                          "given": "Jo\u00e3o Bosco da Mota"
                        ],
                        {
                          "affiliation": [
                            [
                              "family": "Silva",
                              "given": "Juares Bento da"
                            ],
                            {
                              "affiliation": [
                                [
                                  "family": "Gruber",
                                  "given": "Wilson"
                                ],
                                {
                                  "member": "http://dx.doi.org/10.5007/1518-2924.2015v20n42p79",
                                  "published-online": true,
                                  "date-parts": [
                                    [
                                      2015,
                                      4,
                                      14
                                    ]
                                  ],
                                  "container-title": "Enc. Bibli. R. Eletr. Bibliotecon. Ci. Inf.",
                                  "link": [
                                    [
                                      {
                                        "intended-application": "text-mining",
                                        "content-version": "vor",
                                        "content-type": "application/pdf",
                                        "URL": "https://periodicos.ufsc.br/index.php/leib/article/viewFile/36197/29137"
                                      }
                                    ]
                                  ],
                                  "deposited": true,
                                  "date-parts": [
                                    [
                                      2015,
                                      4,
                                      23
                                    ]
                                  ],
                                  "date-time": "2015-04-23T11:43:56Z",
                                  "timestamp": 1429789436000,
                                  "score": 1.0,
                                  "sub-title": [
                                    [
                                      {
                                        "issued": [
                                          [
                                            2015,
                                            4,
                                            14
                                          ]
                                        ],
                                        "URL": "http://dx.doi.org/10.5007/1518-2924.2015v20n42p79",
                                        "ISSN": [
                                          [
                                            "1518-2924"
                                          ]
                                        ]
                                      }
                                    ]
                                  ]
                                }
                              ]
                            ]
                          ]
                        ]
                      ]
                    ]
                  ]
                ]
              ]
            ]
          ]
        ]
      ]
    ]
  ]
}
```

Fonte: Autoridade de Registro DOI – CrossRef (2016).

O Apêndice A traz um excerto do código fonte do protótipo, cuja rotina implementada cumpre todas as tarefas descritas acima.

Ao receber o objeto JSON este é transformado em um objeto PHP e utilizado segundo as necessidades do sistema. Cada par termo/valor é percorrido e, se for de alguma utilidade na composição da referência bibliográfica, será utilizado no preenchimento do respectivo campo do formulário, conforme pode ser verificado na Figura 16.

O tempo de resposta às buscas realizadas vinham aumentando significativamente devido ao grande volume de referências armazenadas, bem como a quantidade de usuários que diariamente o utilizam. Esta quantidade é sazonal e há épocas que ocorrem picos superiores a 7.000 (sete mil) sessões de usuários abertas durante o dia, conforme pode ser verificado na Figura 17, cujo relatório é fornecido pelo *Google Analytics* e refere-se ao mês de novembro de 2015. No dia 16, segundo o relatório, ocorreram 7.593 acessos ao sistema MORE.

Figura 16 – Formulário parcialmente preenchido com o uso de um nome DOI.

[Início](#)
[Criar Referências](#)
[Pesquisar](#)
[Minha Conta](#)
[Ajuda](#)
[Links](#)
[Sobre](#)
[Contato](#)
[Tutorial](#)
[FAQ](#)

Sua localização: Artigo\_revista » Inserir Artigo Revista

### ELABORAR REFERÊNCIA A ARTIGO DE REVISTAS

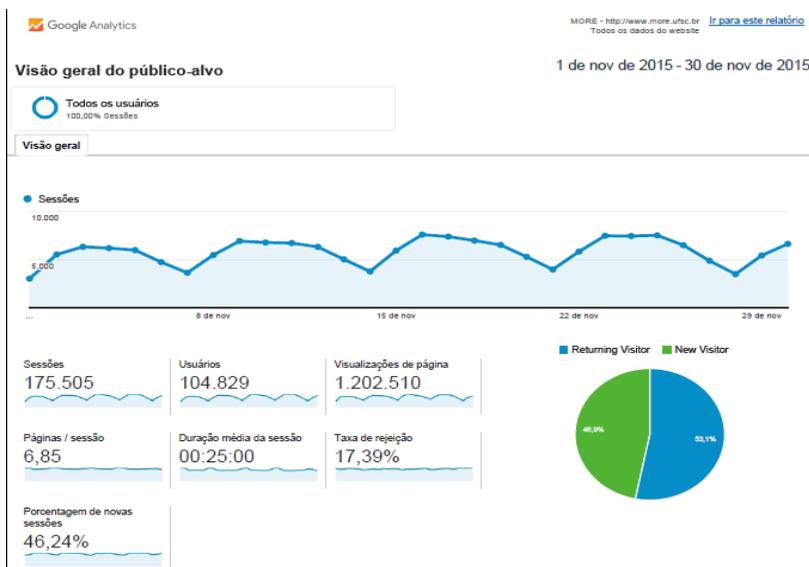
Se o artigo possuir um DOI (Digital Object Identifier) este pode ajudá-lo no autopreenchimento do formulário (preenche os campos cadastrados na *International DOI Foundation*). Basta inseri-lo no campo abaixo e clicar no botão situado ao lado do campo.

Ex: 10.1590/1981-5344/1615

Escolha uma Coleção de Referências Dissertação	Título do Artigo * A geração e o gerenciamento de referência:	N° do Fascículo * 42	N° do Volume * 20
Tipo do Autor Pessoa(s) Física(s)	Subtítulo do Artigo Se for nome próprio a 1ª letra será maiúscul	Periodicidade Ex: semanal	
Responsabilidade Intelectual Autor da Obra	Título da Revista Enc. Bibli	Dia 14	Mês * 4
Mais de 3 Autores Sim	Subtítulo da Revista R. Eletr. Bibliotecon. Ci. Inf.	Ano * 2015	
Autor * Proxério Manoel Felisberto	Local (Cidade) * [S.I.]	Página Inicial * 79	Página Final * Ex: 345
		Notas Universidade Federal de Santa Catarina (UF	

Fonte: print screen da aplicação no sistema operacional Windows 10.

Figura 17 – Extrato do relatório do *Google Analytics*, referente a utilização do sistema MORE durante o mês de novembro de 2015.



Fonte: *Google Analytics* (2015).

Pode ser verificado, ainda, neste extrato alguns outros dados estatísticos para subsidiar nível de importância alcançado pelo MORE no momento atual, bem como mostrar o grau de utilização diária da ferramenta.

Neste contexto, optou-se por modernizar o sistema de busca do MORE baseando-o no modelo espaço vetorial (para um melhor entendimento deste modelo, ver Seção 2.4).

Na realização desta tarefa utilizou-se como suporte o Servidor de Recuperação de Informação Apache Solr<sup>®</sup>, cuja instalação e configuração customizada ocorreu em um servidor de aplicativos Java Apache Tomcat<sup>®</sup> 8. Posteriormente, os dados armazenados na relação “referencia” do banco de dados relacional MySQL do MORE foram recuperados através de um script em PHP que permitiu percorrer os mais de cinco milhões de registros e gerar arquivos tipo *JSON* (*JavaScript Object Notation*) com uma centena de documentos em cada arquivo para posterior indexação. Assim, cada registro da relação é transformado em um documento a ser indexado. Este documento consta de três campos: “CHAVEREFERENCIA” (*unique key*), “REFE” (campo de texto que contém uma referência bibliográfica gerada) e “TIPOTABELA” (identifica a relação que armazena os dados informados pelo usuário para a geração de uma referência bibliográfica). A utilização do formato *JSON* deve-se às características por ele apresentadas e abordadas na Seção 4.2.

Na sequência foram inseridos os mais de cinco milhões de documentos na base de dados do Apache Solr<sup>®</sup> a fim de que o índice invertido fosse construído pelo Apache Lucene<sup>®</sup>, adicionando-os e indexando-os para posterior utilização na recuperação da informação. Esta tarefa foi realizada com o apoio do “curl”.

Para esclarecer de que forma o Solr<sup>®</sup> implementa suas principais características, Serafini (2013) entende que é importante ter uma visão geral sobre o que é um índice Lucene (índice invertido), e como ele é feito. Os conceitos fundamentais do Lucene são os seguintes:

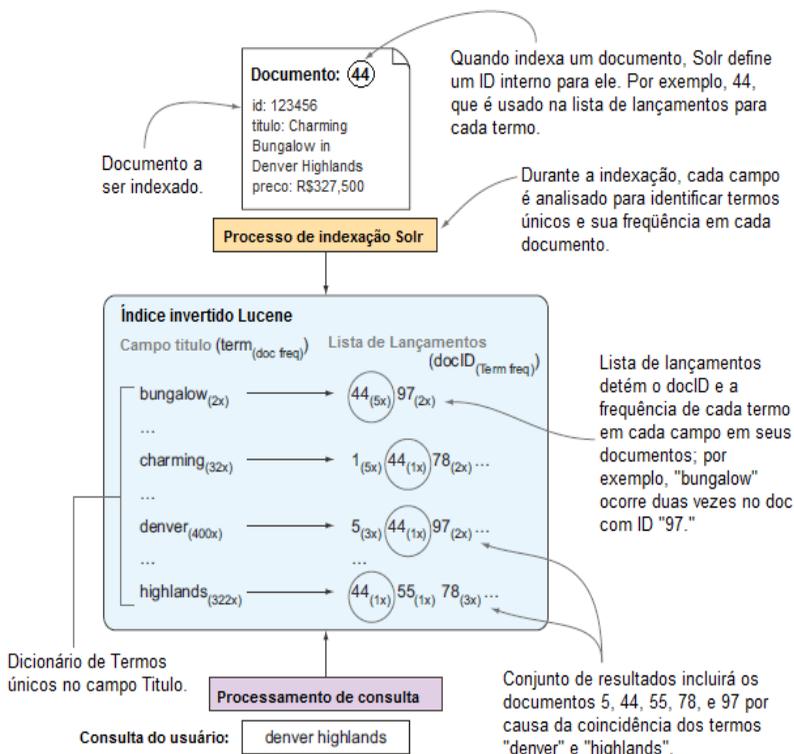
- Documento: é a principal estrutura utilizada, tanto para as buscas como para os índices. Um documento é uma representação na memória dos valores dos dados que precisam ser utilizadas nas pesquisas;

- Campo: este tem seu próprio nome e valor, e é composto por pelo menos um termo. O campo é análogo ao atributo de uma relação em um Banco de Dados Relacional. Assim, cada documento pode ser visto como uma simples lista de pares campo-valor. Um campo pode ser monovalorado ou multivalorado;

- Termo: é a unidade básica para a indexação;
- Índice: é a estrutura em memória onde Lucene (e Solr®) executam as buscas. Pode-se, então, pensar em um documento como um único registro no índice.

A Figura 18 apresenta uma visão geral do processo de indexação.

Figura 18 – Estrutura de chave de dados para recuperação da informação.



Fonte: Adaptado de Grainger e Potter (2014).

A tarefa foi finalizada com o desenvolvimento de um subsistema, o qual foi integrado ao MORE, com a finalidade de agilizar o processo de busca e disponibilização do resultado ao interessado. Na construção deste subsistema foi utilizada a biblioteca *Solarium*, facilitando a comunicação e a integração do subsistema com o Apache Solr®. Vale salientar que a utilização deste mecanismo de recuperação de

informação provê as tarefas de indexação de novos documentos, ou mesmo, atualização ou exclusão, no índice invertido.

Testes realizados e o acompanhamento do subsistema em ambiente de produção comprovam que a utilização do servidor de recuperação de informação Apache Solr<sup>®</sup> contribuiu para uma sensível redução no tempo de resposta das consultas realizadas e, tem demonstrado um desempenho ainda melhor, na repetição destas consultas. Verifica-se que o modelo espaço vetorial, utilizado pelo Apache Solr<sup>®</sup> para atribuir um grau de relevância para cada documento recuperado, permite a recuperação por coincidência parcial e a ordenação do conjunto ocorre segundo o grau de relevância. Assim sendo, mesmo que o usuário cometa algum equívoco na digitação dos termos que compõem a consulta o sistema calcula o grau de relevância em função dos termos corretos, evitando frustrações indesejadas.

## 7 RESULTADOS

Este capítulo é dedicado ao tratamento e a análise dos dados coletados através do instrumento abordado na Seção 5.3.

Os dados aqui referidos foram coletados através de um questionário disponibilizado *online* no *Google Forms* e solicitado aos usuários do MORE que, ao usar um nome DOI<sup>®</sup> no preenchimento do formulário de geração de referência bibliográfica a artigo de revista (*journals/artigo científico*), contribuíssem respondendo a um questionário, com 10 perguntas fechadas e uma aberta (esta para críticas e sugestões), sobre a experiência recém vivenciada.

Para as questões fechadas foi utilizada a escala de Likert de 6 níveis, graduadas de 0 a 5 (para evitar centralidade de opiniões). Valorando os níveis de 0 a 5 o somatório das respostas conduz a um escore situado na faixa entre 0 e 50, inclusive. Para este trabalho foram contabilizados 1033 questionários respondidos. Ressalta-se, ainda, que as perguntas fechadas foram elaboradas no sentido de se obter a posição do usuário sobre sua percepção da utilidade e satisfação em relação ao subsistema que possibilita o preenchimento automático de campos do formulário de geração de referência a artigos de periódicos.

Inicialmente os dados colhidos foram trabalhados e tratados individualmente por questão (pergunta) apresentada no questionário. Na Tabela 2 pode-se verificar alguns dados estatísticos para cada uma destas questões. São eles: média, desvio padrão, curtose e assimetria.

Uma definição para cada termo acima mencionado pode ser encontrada em Levin, Fox e Forde (2012), onde:

- Média é “a soma de um conjunto de escores dividida pelo número total de escores no conjunto”;

- Desvio Padrão “representa a variabilidade do ‘termo médio’ em uma distribuição, pois ele mede a média dos desvios a contar da média”;

- Curtose de uma distribuição simétrica está relacionada a sua característica de pico, ou seja, quanto menor a curtose mais achatada é a curva que representa aquela distribuição (platicúrtica). Ao contrário, valores de curtose mais elevados significa que a curva apresenta altos picos (leptocúrticas), ou seja, valores mais elevados para a variável representada;

- Assimetria está relacionada a inclinação da curva, para direita ou para a esquerda, em virtude da concentração de escores em uma direção, gerando uma “cauda” no extremo oposto onde poucos escores estão localizados. A direção da assimetria é dada pela direção da “cauda”. Assim, uma distribuição negativamente assimétrica (inclinada

para a esquerda) representa uma alta frequência de escores altos e uma baixa frequência de escores de menor valor. Em contrapartida, uma distribuição positivamente assimétrica (inclinada para a direita) representa uma alta frequência de escores baixos e uma baixa frequência de escores de maior valor.

Tabela 2 – Dados estatísticos individualizados por questão.

<i>Questões</i>	<i>Média</i>	<i>Desvio Padrão</i>	<i>Curtose</i>	<i>Assimetria</i>
1. O tempo utilizado para preencher o formulário diminuiu.	4,448	1,213	5,837	-2,549
2. A utilização do DOI facilitou o preenchimento do formulário.	4,378	1,362	4,301	-2,332
3. A utilização do DOI tornou mais prática a geração da referência bibliográfica.	4,424	1,313	5,162	-2,495
4. Havendo oportunidade, voltarei a utilizar um DOI para o autopreenchimento do formulário.	4,524	1,188	7,071	-2,802
5. Sou favorável à implementação de funcionalidades deste tipo.	4,586	1,088	8,760	-3,035
6. A utilização da funcionalidade apresentada melhora minha produtividade.	4,516	1,143	7,461	-2,816
7. Gostaria de contar com esta funcionalidade para outros tipos de fontes bibliográficas.	4,600	1,080	9,207	-3,112
8. O autopreenchimento de alguns campos do formulário ajuda a diminuir a ocorrência de erros de ortografia.	4,513	1,091	7,758	-2,787
9. Fiquei satisfeito com a utilização da funcionalidade de autopreenchimento.	4,479	1,165	6,735	-2,680
10. Geralmente fico satisfeito com as novas funcionalidades implementadas pela equipe do MORE.	4,520	1,062	8,276	-2,846
Totalização dos itens de Likert	44,987	10,066	7,594	-2,752

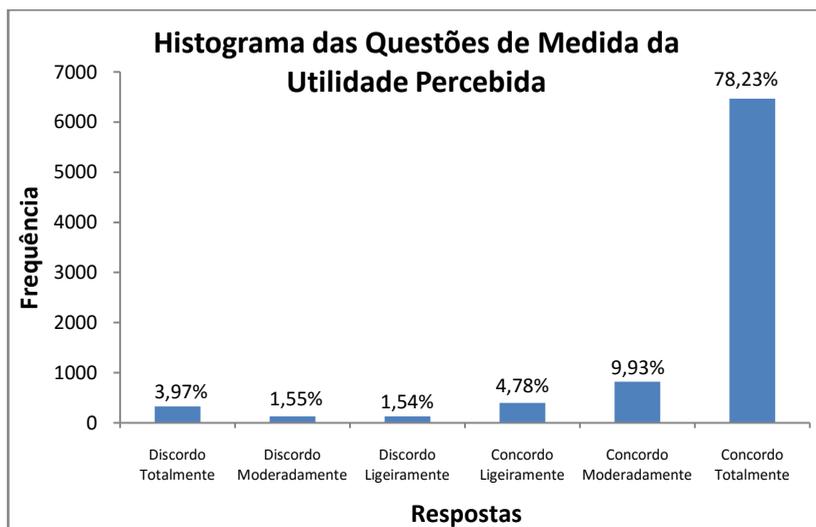
Fonte: elaborada pelo autor.

Dos dados apresentados na Tabela 2 verifica-se que as médias relativas a cada questão oscila entre 4,378 (Questão nº 2) e 4,600 (Questão nº 7). Estas médias elevadas são ratificadas pelo alto valor da curtose e pela assimetria negativa da curva confirmando uma concentração de avaliações correspondentes ao nível máximo da escala. O fato de não existir grandes discrepâncias entre as médias, permite empiricamente deduzir que há afinidade entre as questões. No decorrer

deste capítulo será demonstrado, através de cálculos estatísticos, o grau de afinidade existente entre as questões, bem como a confiabilidade e a validade do instrumento de coleta de dados empregado.

As questões fechadas do questionário foram agrupadas em dois conjuntos. O primeiro conjunto, cuja consolidação encontra-se representada na Figura 19, apresenta os resultados das questões de 1 a 8. A finalidade deste conjunto é avaliar indiretamente o grau de satisfação dos usuários através do nível de utilidade percebida e verificar, no escopo deste trabalho, o entendimento de Park et al. (2011) em que a “utilidade percebida influencia a satisfação dos usuários”. O Histograma da consolidação deste bloco de questões sugere uma variabilidade mínima no percentual das frequências representativas de cada nível da escala de Likert.

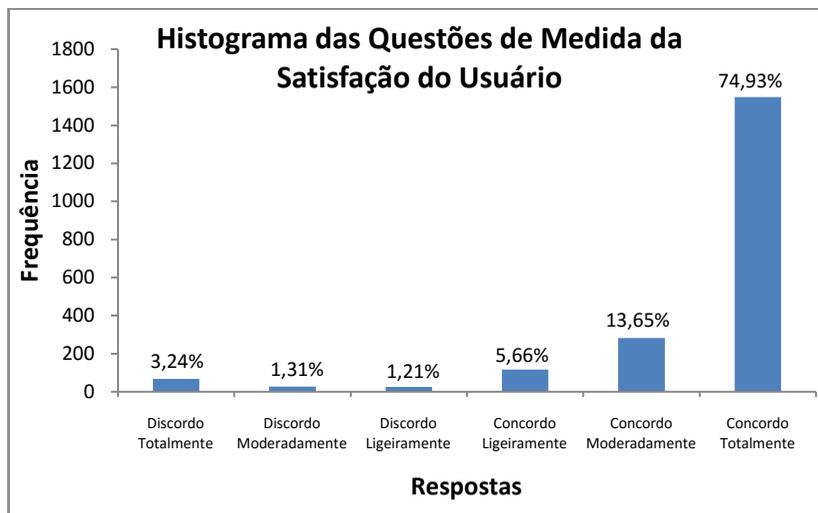
Figura 19 – Histograma das Questões de Medida da Utilidade Percebida.



Fonte: elaborada pelo autor.

O segundo conjunto, cuja consolidação encontra-se representada na Figura 20, apresenta os resultados das questões 9 e 10. A finalidade deste conjunto é avaliar de forma direta o grau de satisfação dos usuários. O Histograma da consolidação deste bloco de questões mantém a tendência da curva verificada na consolidação do bloco anterior.

Figura 20 – Histograma das Questões de Medida da Satisfação do Usuário.



Fonte: elaborada pelo autor.

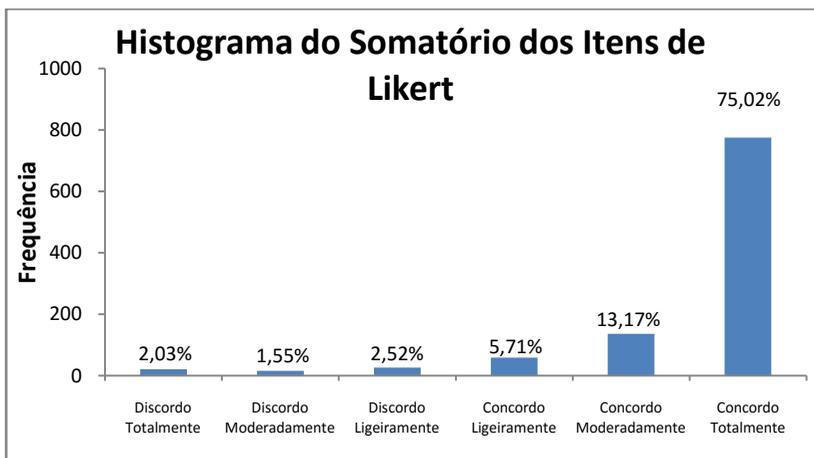
A Figura 21 apresenta o somatório dos itens de Likert dos resultados obtidos, onde pode ser verificada a percentagem de avaliações dos usuários segundo o somatório das opções escolhidas dentre as opções a eles oferecidas: 75,02% concordam totalmente (valores 46 a 50); 13,17% concordam moderadamente (valores de 36 a 45); 5,71% concordam ligeiramente (valores de 26 a 35); 2,52% discordam ligeiramente (valores de 16 a 25); 1,55% discorda moderadamente (valores de 6 a 15); e, 2,03% discordam totalmente (valores de 0 a 5).

O processamento dos resultados acima, com o *software* estatístico “PSPP”, disponível livremente na *web* para baixar e instalar em máquinas isoladas, através da *GNU GENERAL PUBLIC LICENSE*, versão 3, de 29 de junho de 2007, conduziram à obtenção dos seguintes dados estatísticos descritivos para totalização das respostas que compõem a escala de Likert: média de 44,987; desvio padrão de 10,066; curtose de 7,594; e, coeficiente de assimetria da curva de -2,752, o que sugere uma forte tendência de valorações próximas ao limite superior do intervalo.

O gráfico comparativo dos dados consolidados para utilidade percebida, satisfação do usuário e somatório dos itens de Likert é

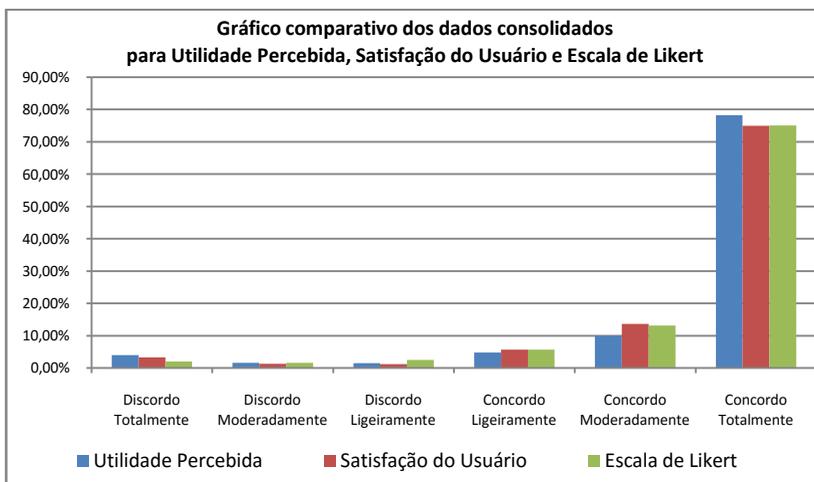
mostrado na Figura 22, o qual ratifica a tendência da frequência para cada escore da escala.

Figura 21 – Histograma do Somatório dos Itens de Likert.



Fonte: elaborada pelo autor.

Figura 22 – Gráfico comparativo dos dados consolidados para Utilidade Percebida, Satisfação do Usuário e Somatório dos Itens de Likert.



Fonte: elaborada pelo autor.

Ressalta-se, ainda, que o menor valor para o coeficiente de correlação de Pearson entre as questões constantes do questionário, conforme pode ser verificado na Tabela 3, é 0,488 e que 88,89% destes valores é igual ou superior a 0,600, indicando que a correlação entre as questões é sempre no sentido positivo e é muito forte, forte ou moderada.

Tabela 3 – Coeficiente de Correlação de Pearson entre as questões do questionário.

Questões	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Q1	1,000									
Q2	0,725	1,000								
Q3	0,709	0,888	1,000							
Q4	0,629	0,814	0,843	1,000						
Q5	0,570	0,729	0,757	0,830	1,000					
Q6	0,600	0,727	0,760	0,793	0,809	1,000				
Q7	0,555	0,674	0,710	0,779	0,859	0,860	1,000			
Q8	0,508	0,599	0,624	0,668	0,722	0,722	0,730	1,000		
Q9	0,675	0,759	0,784	0,756	0,701	0,769	0,722	0,681	1,000	
Q10	0,488	0,607	0,648	0,667	0,695	0,668	0,708	0,673	0,695	1,000

Fonte: elaborada pelo autor.

Segundo Levin, Fox e Forde (2012) o coeficiente de correlação de Pearson é utilizado para “determinar a intensidade e a direção da relação entre duas variáveis, ambas medidas ao nível intervalar”. A Tabela 4 traz a faixa de valores do coeficiente de correlação para os diversos níveis de intensidade.

A Tabela 5 mostra o coeficiente de correlação de Pearson entre os dados consolidados das questões destinadas a medida de utilidade percebida e as questões destinadas a medida direta de satisfação do usuário no valor de 0,998 o que revela uma correlação, no sentido positivo, muito forte (quase perfeita) entre esses dois grupos de dados consolidados. O coeficiente verificado acima corrobora o afirmado por Park et al. (2011) em que a “utilidade percebida influencia a satisfação dos usuários e representa benefício líquido para os indivíduos,

organizações e sociedade”, ou seja, “é a hipótese de haver uma relação causal entre a utilidade percebida e satisfação do usuário”.

Tabela 4 – Intensidade e direção dos valores do coeficiente de correlação.

<i>Correlação de Pearson</i>				
<i>Correlação Negativa</i>			<i>Correlação Positiva</i>	
<i>Coefficiente</i>	<i>Intensidade</i>		<i>Coefficiente</i>	<i>Intensidade</i>
- 1,00	Perfeita		+ 1,00	Perfeita
- 0,80	Muito Forte		+ 0,80	Muito Forte
- 0,60	Forte		+ 0,60	Forte
- 0,30	Moderada		+ 0,30	Moderada
- 0,10	Fraca		+ 0,10	Fraca

Fonte: adaptada de Levin, Fox e Forde (2012, p. 303-304).

Tabela 5 – Coeficiente de Correlação de Pearson entre as medidas de Utilidade Percebida e Satisfação do Usuário.

	<i>Utilidade Percebida</i>	<i>Satisfação do Usuário</i>
<b>Utilidade Percebida</b>	<b>1,000</b>	
<b>Satisfação do Usuário</b>	<b>0,998</b>	<b>1,000</b>

Fonte: elaborada pelo autor.

O grau de confiabilidade do questionário foi avaliado através do cálculo do Alfa de Cronbach. Esclarecem Alonso e Santacruz (2015) que o alfa de Cronbach “aparece, frequentemente na literatura, como uma forma simples e confiável para a validação do constructo de uma escala e como uma medida que quantifica a correlação existente entre os itens que compõem esta escala”. Este coeficiente varia entre 0 e 1, inclusive. Quanto maior o coeficiente maior a confiabilidade. Um valor do alfa de Cronbach, entre 0,70 e 0,90, indica uma boa consistência interna para uma escala unidimensional.

Para Vieira (2009) um questionário “tem confiabilidade se os escores conferidos às declarações que medem a mesma dimensão têm consistência”, ou seja, confiabilidade é “o grau em que um instrumento mede seja lá o que deva medir, de forma consistente”. Esclarece, ainda,

a autora que “um instrumento é válido para obter determinada medida quando faz a medição dessa medida com confiabilidade”.

No âmbito desta pesquisa o alfa de Cronbach foi calculado em função de todas as 1033 respostas a todas as questões fechadas do questionário. Cada questão é considerada individualmente, de modo a verificar sua afinidade com as demais e, finalmente, calcula-se o valor do coeficiente para o conjunto de questões como um todo. O valor encontrado foi de 0,96, indicando que o questionário aplicado apresenta um alto grau de confiabilidade.

Quanto a proposição da questão aberta (Questão nº 11) no sentido de deixar o usuário a vontade para fazer suas críticas e/ou sugestões, para melhorar as funcionalidades e a usabilidade do MORE, verifica-se um grande interesse do público em geral em colaborar com as mais variadas ideias, algumas coerentes e outras um pouco menos. A frequência das opiniões favoráveis e desfavoráveis seguem, conforme a especificidade desta questão, a curva verificada nas questões anteriores.

Observa-se que alguns usuários não se dão conta que determinadas funcionalidades deixam de ser implementadas por circunstâncias alheias a vontade da equipe de desenvolvimento, como por exemplo, implementar rotina semelhante a que usa o DOI para outros identificadores (ISBN, ISSN, etc) que não se enquadram como identificadores digitais de objeto e, portanto, não são suportados por uma estrutura adjacente para possibilitar o intercâmbio dos metadados referentes ao recurso informacional vinculado ao identificador.

Salienta-se a existência de várias respostas indicando uma boa aceitação do emprego das TICs no auxílio ao gerenciamento e geração de referências bibliográficas, bem como a utilização dos metadados relativos a um nome DOI no preenchimento automático dos campos do formulário empregado na geração da referência bibliográfica a artigo de periódico.

Segue uma amostra destas respostas:

- “Parabéns, essa nova funcionalidade é perfeita. Obrigado.”
- “Achei muito proveitoso já que as normas da ABNT são de difícil entendimento para algumas pessoas.”
- “Foi ótimo, pois me poupam um tempo considerável de digitação.”
- “Estou agradecida por esse melhoramento.”
- “Ferramenta muito boa. Ajudou bastante no desenvolvimento da minha monografia! Obrigada!”
- “Seria melhor ter a opção de incluir todos os autores, no caso da obra ter mais de três.”

- “Excelente ideia, muito obrigada!”
- “Infelizmente não houve preenchimento do n. do fascículo nem do volume.”
- “A tecnologia está sempre andando a nosso favor, e essa ferramenta facilitou muito o andamento.”
- “Adorei!!!!”
- “Muito grata a todos pela praticidade que é novidade e seria muito produtivo que todos os artigos obtivesse este código DOI. Desde já obrigada.”
- “Facilitou demais o preenchimento, Parabéns!!!”
- “Não ajudou em nada. Tive que preencher todos os itens.”
- “Nem tudo ficou certo. Faltaram alguns dados.”
- “Amei esta nova ferramenta. Incrível.”
- “Ideia GENIAL! Muito obrigada!”

Destaca-se, ainda, a sugestão de uma pesquisadora referência nacional em pesquisa sobre a utilização do DOI<sup>®</sup>, nos seguintes termos: “Sugiro ainda que o link de acesso seja o que tenha o DOI<sup>®</sup>, por exemplo <http://dx.doi.org/10.5007/1518-2924.2015v20n42p79>, isso irá auxiliar nas métricas. No mais, parabêniso a equipe. E estou divulgando essa nova funcionalidade. Att.” (omite-se aqui o nome da pesquisadora em atenção ao compromisso de confidencialidade das respostas assumido no cabeçalho do instrumento de pesquisa).

Do acima exposto e em virtude do contido em Brito et al. (2015, p. 40) e na Seção 3.3.3 deste trabalho foi alterada a rotina de geração de referência a fim de constar ao final desta um enlace ativo com o DOI<sup>®</sup> informado pelo usuário o que dispensa a informação sobre o URL e a data de acesso e auxilia nas métricas de acesso e citação.

A respeito de respostas do tipo “Não ajudou em nada” ou “Nem tudo ficou certo. Faltaram alguns dados” pode-se destacar possíveis causas, tais como:

- as especificidades da norma técnica da ABNT que difere de outras normas internacionais (APA, Vancouver e outras);
- alguns dos responsáveis pela geração do arquivo XML utilizado no cadastro dos metadados do recurso informacional identificado não seguem o padrão preconizado pela ABNT (até porque vários periódicos especializados seguem modelo de referências próprio (IEEE, por exemplo);
- a IDF exige um conjunto mínimo de metadados que não são, na maioria das vezes, suficientes para o preenchimento de todos os campos do formulário de geração da referência bibliográfica;

- as normas da IDF não padronizam a escrita dos metadados (por exemplo, não abreviar o nome do periódico).

Do abordado neste capítulo, a análise dos dados coletados indica que a utilização das TICs na gestão dos metadados contidos nos identificadores digitais de objetos, utilizados na Biblioteconomia, podem impactar positivamente sobre a usabilidade do Mecanismo Online para Referências (MORE) e esta, por sua vez, sobre a utilidade percebida e sobre a satisfação do usuário.

## 8 CONSIDERAÇÕES FINAIS

Esta pesquisa foi desenvolvida com base no pressuposto de que os sistemas que fazem uso de informações recuperadas dos repositórios das autoridades de registro da *International DOI Foundation*, aliado ao suporte oferecido pelas Tecnologias da Informação e Comunicação na recuperação de metadados de fontes bibliográficas, podem contribuir para melhorar a usabilidade do sistema e elevar o grau de utilidade percebida e de satisfação de seus usuários.

Para concretizar o desiderato proposto na Seção 1.2.1 foi, inicialmente, desenvolvido um protótipo capaz de extrair os metadados associados a um nome DOI conforme especificado no capítulo 6. Dos dados coletados junto aos usuários percebe-se que o referido protótipo realiza a tarefa proposta de maneira satisfatória contribuindo, de acordo com o esperado, para o êxito deste trabalho.

A adição de rotinas e/ou subsistemas ao sistema em uso pressupõe um incremento no consumo de recursos computacionais, demandando um maior lapso de tempo de processamento e, por consequência, aumento no tempo de resposta. No entanto, Fielding (2000) considera que o desempenho percebido pelo usuário difere do desempenho, propriamente dito, do sistema e da rede, uma vez “que o desempenho de uma ação é medido em termos de impacto sobre o usuário frente a uma aplicação, em vez da taxa a qual a rede move informações”.

Neste sentido, ao avaliar se a integração do protótipo ao sistema em uso altera o comportamento de algum requisito não funcional com base nos dados coletados, em especial aqueles colhidos através da questão nº 11, conclui-se que, em consonância com as considerações de Fielding (2000) e o universo de usuários consultados, a referida integração não trouxe qualquer impacto negativo aos requisitos não funcionais do MORE. Ao contrário, avalia-se que houve um impacto positivo na usabilidade do sistema.

Durante o período destinado a coleta dos dados o acompanhamento diário dos questionários respondidos permitiu levantar alguns dos benefícios propiciados pela gestão dos metadados contidos nos identificadores digitais de objetos, utilizados em Biblioteconomia, aos usuários do MORE e à comunidade em geral. Entre eles pode-se citar:

- Economia de tempo e agilização no preenchimento do formulário;

- Evidências de incremento positivo na facilidade, na praticidade e na produtividade na condução do processo de geração de referência bibliográfica;

- Contribuição para a redução de erros de digitação uma vez que os dados recuperados são previamente auditados pelas Autoridades de Registro antes de sua inclusão nos respectivos repositórios;

- A utilização de um enlace ativo na referência bibliográfica permite a recuperação rápida do recurso informacional referenciado, se este for um recurso aberto (caso contrário necessitará permissão de acesso ao recurso o que poderá demandar algum tempo adicional);

- O interesse demonstrado pelos usuários do MORE através da quantidade de *feedbacks* recebidos e das sugestões e críticas neles contidas permite concluir que há uma contribuição positiva na autoconfiança desses usuários no momento da formatação da lista das referências bibliográficas citadas nos respectivos trabalhos acadêmicos e/ou científicos.

Em sintonia com Park et al. (2011) no sentido de que a utilização da *Web 2.0* contribuiu e ainda contribui para o fenômeno produtivo de recursos informacionais digitais e que para agilizar e equacionar o gerenciamento deste grande volume de informações torna-se cada vez mais urgente e essencial a adoção de sistemas de identificação e gestão destes objetos.

Vale salientar o fato de que, até o presente momento, poucas investigações foram feitas para verificar os benefícios, as desvantagens e os fatores de sucesso na adoção de sistemas de identificadores digitais de objetos. A utilização destes sistemas para além dos publicadores de periódicos especializados e provedores de conteúdo digital proprietário ainda é incipiente, mesmo a nível mundial.

No Brasil destaca-se o pioneirismo desta Universidade Federal na sua filiação a IDF e a adoção do DOI<sup>®</sup> na identificação dos recursos informacionais contidos em seus periódicos. Ao final de 2013 verifica-se que é a instituição nacional com o maior número de nomes DOI<sup>®</sup> atribuídos seguida pela Universidade Estadual de Maringá (UEM).

Da análise dos dados obtidos e processados verifica-se um elevado grau de correlação entre o nível de utilidade percebida e a satisfação do usuário. Constata-se, ainda, que 93,9% dos usuários amostrados neste trabalho entendem que a funcionalidade que faz uso dos metadados vinculados ao nome DOI por eles informado contribui positivamente para melhorar a usabilidade do sistema e, por consequência, para elevar o grau da utilidade percebida e da satisfação do usuário.

Do exposto e no intuito de responder a pergunta condutora dos trabalhos realizados nesta pesquisa pode-se concluir que a utilização dos metadados contidos nos identificadores digitais de objetos, utilizados na Biblioteconomia, apoiada pelas Tecnologias da Informação e Comunicação, contribui positivamente para melhorar a usabilidade dos sistemas que utilizam informação recuperada dos repositórios das Autoridades de Registro da IDF (no caso deste trabalho o sistema MORE) e esta, por sua vez, incrementa o grau da utilidade percebida e da satisfação do usuário.

Como perspectiva de trabalhos futuros vislumbra-se a possibilidade de utilização dos metadados contidos nos identificadores digitais de objetos de Biblioteconomia para a geração de referências bibliográficas a outras fontes bibliográficas que dispuserem desta estrutura de dados junto a uma nova coleta de dados a fim de confirmar, ou refutar, a tendência ora apresentada.

Outra possibilidade a ser trabalhada seria uma pesquisa junto às instituições nacionais que adotaram sistemas de identificadores digitais de objetos a fim de levantar os fatores que contribuíram positivamente ou negativamente para o êxito desta adoção e quais os benefícios advindos de tais sistemas.

Uma oportunidade que se apresenta com relação ao MORE é a possibilidade de armazenamento do resumo e das palavras-chave da obra referenciada com a finalidade de formar agrupamentos de modo a promover sugestões sobre obras que tenham alguma afinidade com aquela cuja referência está sendo gerada.

Finalmente, pode-se trabalhar no intuito de promover a integração entre os repositórios dos sistemas de identificadores digitais de objetos que tem se destacado no cenário mundial, tais como o DOI, o cID e o UCI.



## REFERÊNCIAS

ALONSO, Jorge Alberto González; SANTACRUZ, Mauro Pazmiño. Cálculo e interpretación del Alfa de Cronbach para el caso de validación de la consistencia interna de un cuestionario, con dos posibles escalas tipo Likert. **Revista Publicando**, Quito, v. 2, n. 1, p.62-77, 2015. Trimestral.

ALVES, Maria Bernadete Martins; MENDES, Leandro Luiz; ALVES, João Bosco da Mota. MORE: Mecanismo on-line para referências. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 14., 2006, Salvador. **Anais...** . p. 1 - 12. Disponível em: <<http://xa.yimg.com/kq/groups/25169972/1250144979/name/4.ALVES>>. Acesso em: 12 mar. 2016.

ALVES, Rachel Cristina Vesú. **Metadados como elementos do processo de catalogação**. 2010. 134 f. Tese (Doutorado) - Curso de Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Cap. 7.

ALVES, Rachel Cristina Vesú; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. **Metadados no Domínio Bibliográfico**. Rio de Janeiro: Intertexto, 2013. 196 p.

ALVES, Rachel Cristina Vesú; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. METADADOS: organização e acesso à informação no domínio bibliográfico. In: ENCONTRO INTERNACIONAL DE CATALOGADORES, 9., 2013, Rio de Janeiro. **Anais...** . Rio de Janeiro: Fundação Biblioteca Nacional do Brasil, 2013. p. 1-15. Disponível em: <<http://www.enacat.ufscar.br/index.php/eic-enacat/eic-enacat/paper/viewFile/52/22>>. Acesso em: 17 abr. 2015.

ARAÚJO JÚNIOR, Rogério Henrique de; TARAPANOFF, Kira. Precisão no processo de busca e recuperação da informação: uso da mineração de textos. **Ciência da Informação**, Brasília, v. 35, n. 3, p.326-247, dez. 2006. Quadrimestral.

ARAUJO, Vera Maria Araujo Pigozzi de. Sistemas de recuperação da informação: uma discussão a partir de parâmetros enunciativos. **Transinformação**, Campinas, v. 2, n. 24, p.137-143, 2012. Bimestral.

ARELLANO, Miguel Angel. Preservação de documentos digitais. **Ciência da Informação**, Brasília, v. 33, n. 2, p.15-27, ago. 2004. Quadrimestral.

ASSUMPCÃO, Fabrício Silva; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Representação no domínio bibliográfico: um olhar sobre os Formatos MARC 21. **Perspectivas em Ciências da Informação**, Belo Horizonte, v. 20, n. 1, p.54-74, maio 2015. Trimestral. <http://dx.doi.org/10.1590/1981-5344/2054>.

BARITÉ, Mario et al. Garantia literária: elementos para uma revisão crítica após um século. **Transinformação**, Campinas, v. 22, n. 2, p.123-138, ago. 2010. Quadrimestral.

BASSETT, Lindsay. **Introduction to JavaScript Object Notation**. Sebastopol: O'Reilly Media, Inc., 2015. 108 p.

BATTLE, Robert; BENSON, Edward. Bridging the semantic Web and Web 2.0 with Representational State Transfer (REST). **Web Semantics: Science, Services and Agents on the World Wide Web**, [S.l.], v. 6, n. 1, p.61-69, fev. 2008. Elsevier BV. <http://dx.doi.org/10.1016/j.websem.2007.11.002>.

BELKIN, N. J.. INFORMATION CONCEPTS FOR INFORMATION SCIENCE. **Journal of Documentation**, London, v. 34, n. 1, p.55-85, jan. 1978. <http://dx.doi.org/10.1108/eb026653>.

BRITO, Ronnie Fagundes de et al. **Manual do usuário do Digital Object Identifier (DOI®)**. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2015. 74 p.

BRITO, Ronnie Fagundes de; GUEDES, Mariana Giuberti; SHINTAKU, Milton. **Atribuição de identificadores digitais para publicações científicas: DOI para o SEER/OJS**. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2013. 40 p. Disponível em: <[http://livroaberto.ibict.br/bitstream/1/1016/7/Atribuição de identificadores digitais para publicações](http://livroaberto.ibict.br/bitstream/1/1016/7/Atribuição_de_identificadores_digitais_para_publicações)>. Acesso em: 06 abr. 2013.

CA TECHNOLOGIES. **Um guia para a REST e o design de API**. São Paulo: [S.e], 2015. 12 p. Disponível em: <<http://transform.ca.com/rs/>>

catech/images/A Guide to REST and API Design.pdf>. Acesso em: 18 jun. 2015.

CAPURRO, Rafael. Information Technology and Technologies of the Self. **Journal of Information Ethics**, [S.l.], v. 5, n.2, p.19-28, 1996.

CAPURRO, Rafael; HJØRLAND, Birger. The concept of information as we use in everyday. In: ANNUAL Review of Information Science and Technology. Medford: Blaise Cronin, 2003. Cap. 8. p. 343-411.

CARVALHO, Lidiane dos Santos; LUCAS, Elaine R. de Oliveira; GONÇALVES, Lucas Henrique. Organização da Informação para Recuperação em Redes de Produção e Colaboração na Web. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 15, n. 1, p.71-86, jun. 2010. Semestral.

CASTILLO, P.A. et al. Using SOAP and REST web services as communication protocol for distributed evolutionary computation. **International Journal of Computers & Technology**, [S.l.], v. 10, n. 5, p.1752-1770, 25 ago. 2013. Semanal.

CERVO, Amado Luiz; BERVIAN, Pedro Alcino; SILVA, Roberto da. **Metodologia Científica**. 6. ed. São Paulo: Pearson Prentice Hall, 2007. 162 p.

CHANDRAKAR, Rajesh. Digital object identifier system: an overview. **The Electronic Library**, Ahmedabad, v. 24, n. 4, p.445-452, jul. 2006. <http://dx.doi.org/10.1108/02640470610689151>.

CONTENT ID FORUM (CIDF). **CIDF SPECIFICATION 2**: Rev. 1.1. Tokio, Japão, 2007. 148 p. Disponível em: <<http://www.npo-ba.org/cid/cIDfSpecV2R11E.pdf>>. Acesso em: 09 nov. 2015.

CURL: groks those URLs. 2016. Disponível em: <<https://curl.haxx.se/>>. Acesso em: 05 jan. 2016.

DAIGLE, L. et al. **RFC 3406**: Uniform Resource Names (URN) Namespace Definition Mechanisms. [S.l.]: RFC Editor, 2002. 22 p. Disponível em: <<https://www.ietf.org/rfc/rfc3406.txt>>. Acesso em: 23 nov. 2015.

DAMASIO, Edilson. CrossRef, DOI (Digital Object Identifier) e serviços: estudo comparativo Luso-Brasileiro. **Incid: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 4, n. 2, p.126-142, 20 dez. 2013. Semestral. Universidade de São Paulo. SIBiUSP. <http://dx.doi.org/10.11606/issn.2178-2075>.

DODEBEL, Vera Lúcia Doyle. **Tesauro**: linguagem de representação da memória documentária. Rio de Janeiro: Interciência, 2014. 127 p. 1ª Reimpressão.

DZIEKANIAK, Gisele. A Organização da informação e a comunicação científica: implicações para os profissionais e usuários da informação. **Em Questão**, Porto Alegre, v. 16, n. 1, p.45-59, 2010. Bimestral.

FACHIN, Gleisy Regina Bories. Recuperação Inteligente da Informação e Ontologias: um levantamento na área da Ciência da Informação. **Biblios**, Rio Grande, v. 1, n. 23, p.259-283, 2009.

FARIAS FILHO, Milton Cordeiro; ARRUDA FILHO, Emílio J. M.. **Planejamento da pesquisa científica**. São Paulo: Atlas, 2013. 157 p.

FELISBERTO, Proxério Manoel. **Mecanismo Online para Referências - MORE**: reengenharia e ampliação do sistema. 2013. 104 f. TCC (Graduação) - Curso de Tecnologias da Informação e Comunicação, Campus Araranguá, Universidade Federal de Santa Catarina, Araranguá, 2013. Cap. 6. Disponível em: <[https://repositorio.ufsc.br/bitstream/handle/123456789/131062/Proxerio\\_Felisberto.pdf](https://repositorio.ufsc.br/bitstream/handle/123456789/131062/Proxerio_Felisberto.pdf)>. Acesso em: 27 abr. 2014.

FERREIRA, Miguel. **Introdução à preservação digital**: Conceitos, estratégias e actuais consensos. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006. 88 p.

FIELDING, Roy Thomas. **Architectural Styles and the Design of Network-based Software Architectures**. 2000. 180 f. Tese (Doutorado) - Curso de Information and Computer Science, University of California, Irvine, 2000. Cap. 6.

GHORAB, M. Rami et al. Personalised Information Retrieval: survey and classification. **Springer Science**, [S.l.], p.381-443, 13 maio 2012.

GIBBS, Mark. Updating Twitter with cURL and Wget. **Network World**, [S.l.], p.30-30, 21 maio 2008. Disponível em: <<http://go.galegroup.com/ps/i.do?id=GALE|A214550911&v=2.1&u=capes&it=r&p=AONE&sw=w&asid=43566946bff9>>. Acesso em: 12 jan. 2016.

GRAINGER, Trey; POTTER, Timothy. **Solr in Action**. Shelter Island: Manning Publications Co., 2014. 639 p.

GRIFFIN, Keith; FLANAGAN, Colin. Defining a call control interface for browser-based integrations using representational state transfer. **Computer Communications**, [S.l.], v. 34, n. 2, p.140-149, fev. 2011. Elsevier BV. <http://dx.doi.org/10.1016/j.comcom.2010.03.029>.

GÜNTHER, Hartmut. Como Elaborar um Questionário. In: GÜNTHER, Hartmut (Org.). **Planejamento de Pesquisa para as Ciências Sociais**. Brasília, DF: UNB, Laboratório de Psicologia Ambiental, 2003. p. 1-35.

IKEMATU, Ricardo Shoiti. Gestão de Metadados: Sua Evolução na Tecnologia da Informação. **Revista de Ciência da Informação: DataGramZero**, Rio de Janeiro, v. 2, n. 6, p.1-7, dez. 2001. Bimestral. Disponível em: <[http://www.dgz.org.br/dez01/Art\\_02.htm](http://www.dgz.org.br/dez01/Art_02.htm)>. Acesso em: 29 set. 2015.

INTERNATIONAL DOI FOUNDATION (IDF). **Digital Object Identifier System Handbook**. 2015. Disponível em: <<https://www.doi.org/hb.html>>. Acesso em: 29 out. 2015.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **ISO 26324**: Information and documentation: digital object identifier system. Geneva, 2012.

JACSO, Peter. **The Digital Object Identifier**: this method of uniquely identifying a document could be the strongest link. (Internet Insights). 2002. Information Today, Inc. Academic OneFile.. Disponível em: <<http://go.galegroup.com/ps/i.do?id=GALE|A94961074&v=2.1&u=capes&it=r&p=AONE&sw=w&asid=232cf7d29201fc>>. Acesso em: 15 abr. 2015.

KANG, S.. National Computerization Agency. **RFC 4179**: Using Universal Content Identifier (UCI) as Uniform Resource Names (URN).

Seoul: The Internet Society, 2005. 7 p. Disponível em: <<http://www.rfcbase.org/txt/rfc4179.txt>>. Acesso em: 15 nov. 2015.

KEEFER, Alice. Gestão de enlaces entre artigos eletrônicos: el sistema CrossRef. **El Profesional de la Información**, [S.l.], v. 10, n. 4, p.32-33, abr. 2001. Disponível em: <<http://www.elprofesionalde la informacion.com/contenidos/2001/abril/4.pdf>>. Acesso em: 10 abr. 2014.

LANCASTER, F. W.. **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Briquet de Lemos / Livros, 2004. Tradução de: Antônio Agenor Briquet de Lemos.

LEVIN, Jack; FOX, James Alan; FORDE, David R.. **Estatística para ciências humanas**. 11. ed. São Paulo: Pearson Education do Brasil, 2012. 460 p. Tradução de Jorge Ritter.

LOPES, Ilza Leite. Estratégia de busca na recuperação da informação: revisão da literatura. **Ciência da Informação**, Brasília, v. 31, n. 2, p.60-71, ago. 2002. Quadrimestral.

MA, Kun; YANG, Bo. A simple scheme for bibliography acquisition using DOI content negotiation proxy. **The Electronic Library**, [S.l.], v. 32, n. 6, p.806-824, 3 nov. 2014. Emerald. <http://dx.doi.org/10.1108/el-09-2012-0121>.

MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinrich. **An Introduction to Information Retrieval**. Cambridge: Cambridge University Press, 2009. 544 p.

MARTINS, Elaine Cristina Domingues; CARVALHO, Tatiana. Recuperação da informação em psicologia: LILACS e Index Psi Revistas Técnico-Científicas. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 2, p.118-130, jun. 2014. Trimestral.

MARTINS, Marcio Souza; LIMA, Vânia Mara Alves. The social approach in information retrieval: front and trends research. **Biblios**, Lima, n. 52, p.1-15, 29 out. 2013. Trimestral. University of Pittsburgh. <http://dx.doi.org/10.5195/biblios.2013.104>.

MORI, Alexandre; CARVALHO, Cedric Luiz de. **Metadados no Contexto da Web Semântica**. Goiânia: Universidade Federal de Goiás, 2004. 19 p.

NATIONAL INFORMATION STANDARDS ORGANIZATION (NISO). **Common Identifiers and Their Acronyms**: NISO Roundtable on Identifiers. US, 2006. 3 p. Disponível em: <<http://www.niso.org/news/events/niso/past/ID-06-wkshp/identifiers.pdf>>. Acesso em: 25 out. 2015.

NATIONAL INFORMATION STANDARDS ORGANIZATION. NISO. **Understanding Metadata**. Bethesda (USA): Niso Press, 2004. 20 p. Disponível em: <<http://www.niso.org>>. Acesso em: 07 abr. 2014.

NEUMANN, Janna; BRASE, Jan. DataCite and DOI names for research data. **Journal of Computer-aided Molecular Design**, Hannover, v. 28, n. 10, p.1035-1041, 20 jul. 2014. Springer Science, Business Media. <http://dx.doi.org/10.1007/s10822-014-9776-5>.

OKADA, Susana Yuri; ORTEGA, Cristina Dotta. Análise da Recuperação da Informação em Catálogo online de Biblioteca Universitária. **Informação & Informação**, Londrina, v. 14, n. 1, p.18-35, jul. 2009. Bimestral.

OLIVEIRA, Dalgiza Andrade; ARAUJO, Ronaldo Ferreira de. Construção de Linguagens Documentárias em Sistemas de Recuperação da Informação: a importância da garantia do usuário. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 17, n. 34, p.17-30, ago. 2012. Quadrimestral. <http://dx.doi.org/10.5007/1518-2924.2012v17n34p17>.

OLIVEIRA, Luis Henrique Gonçalves de. **Extração de Metadados utilizando uma ontologia de domínio**. 2009. 67 f. Dissertação (Mestrado) - Curso de Ciências da Computação, Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009. Cap. 2. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/22814/000740674.pdf>>. Acesso em: 09 abr. 2014.

ONG, Shyue Ping et al. The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles.

**Computational Materials Science**, [S.l.], v. 97, p.209-215, fev. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.commatsci.2014.10.037>.

PÁDUA, Elisabete Matallo Marchesini de. O trabalho monográfico como iniciação à pesquisa científica. In: CARVALHO, Maria Cecília Maringoni de (Org.). **Construindo o saber: Metodologia Científica - Fundamentos e Técnicas**. 24. ed. Campinas, SP: Papirus, 2012. Cap. 5. p. 185-213.

PARANHOS, Ranulfo et al. Corra que o survey vem aí.: Noções básicas para cientistas sociais. **Revista Latinoamericana de Metodología de la Investigación Social**, Argentina, n. 6, p.07-24, mar. 2014. Semestral. Año 3.

PARK, Sungbum et al. Examining success factors in the adoption of digital object identifier systems. **Computer Networks and Communications: Electronic Commerce Research and Applications**, Amsterdam, v. 10, n. 6, p.626-636, 27 maio 2011. Bimestral. <http://dx.doi.org/10.1016/j.eierap.2011.05.004>.

PARK, Sungbum et al. ID-based interoperation between digital and physical resources in ubiquitous environment. In: INTERNATIONAL CONFERENCE ON CONVERGENCE AND HYBRID INFORMATION TECHNOLOGY, 3., 2008, Busan (Korea). **Proceedings...** . Busan (korea): Iccit, 2008. p. 781 - 786.

PASKIN, Norman. Digital Object Identifier (DOI®) System. In: **Understanding Information Retrieval Systems**. Oxford: Informa UK Limited, 2011. p. 605-613. <http://dx.doi.org/10.1201/b11499-49>.

PASKIN, Norman. Digital Object Identifiers for scientific data. **Data Science Journal**, [S.l.], v. 4, p.12-20, 2005. Disponível em: <<http://www.doi.org/topics/041110CODATAarticleDOI.pdf>>. Acesso em: 10 abr. 2014.

PASKIN, Norman. The digital object identifier system: digital technology meets content management. **Interlending & Document Supply**, Kidlington, v. 27, n. 1, p.13-16, mar. 1999. Emerald. <http://dx.doi.org/10.1108/02641619910255829>.

PELTONEN, Jaakko; LIN, Ziyuan. Information retrieval approach to meta-visualization. **Springer**, Aalto, p.1-41, 03 out. 2014.

PONTES JUNIOR, João de; CARVALHO, Rodrigo Aquino de; AZEVEDO, Alexander William. Da recuperação da informação à recuperação do conhecimento: reflexões e propostas. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 18, n. 4, p.2-17, dez. 2013. Trimestral.

POULOS, Marios; KORFIATIS, Nikolaos; BOKOS, George. Towards text copyright detection using metadata in web applications. **Program: electronic library and information systems**, [S.l.], v. 45, n. 4, p.439-451, ago. 2011. <http://dx.doi.org/10.1108/00330331111182111>.

PPGTIC. **Linhas de Pesquisa do Programa de Pós-Graduação em Tecnologias da Informação e Comunicação**. 2016. Disponível em: <<http://ppgtic.ufsc.br/areas/>>. Acesso em: 26 mar. 2016.

RAI, Arun et al. Assessing the Validity of IS Success Models: An Empirical Test and Theoretical Analysis. **Information Systems Research**, Catonsville (USA), v. 13, n. 1, p.50-69, mar. 2002. Institute for Operations Research and the Management Sciences (INFORMS). <http://dx.doi.org/10.1287/isre.13.1.50.96>.

RAMOS, Clérison; MUNHOZ, Deise Parula. A Subjetividade da Relevância na Recuperação da Informação: análise a partir de imagens representativas. **Biblios: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 25, n. 1, p.69-79, jun. 2011. Semestral.

RIBEIRO, Fernanda. O uso da classificação nos arquivos como instrumento de organização, representação e recuperação da informação. In: CONGRESSO ISKO ESPANHA E PORTUGAL, 1., 2013, Porto. **Informação e/ou Conhecimento: as duas faces de Jano**. Porto: Cetac Media, 2013. p.5-16.

RODRIGUES, Bruno César; CRIPPA, Giulia. A recuperação da informação e o conceito de informação: o que é relevante em mediação cultural?. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 16, n. 1, p.45-64, 2011. Trimestral.

ROSENBLATT, Bill. Enterprise Content Integration with the Digital Object Identifier. **Giantsteps Media Technology Strategies**, New York, p.1-26, 20 jun. 2002. Content Directions, Inc.

RYKOWSKI, Jarogniew. Managing, Searching and Accessing IoT Devices Via REST Resources. **International Journal Of Computer Networks & Communications (IJCNC)**, Tamil Nadu, India, v. 7, n. 3, p.171-191, 31 maio 2015. Academy and Industry Research Collaboration Center. <http://dx.doi.org/10.5121/ijcnc.2015.7313>.

SARACEVIC, Tefko. Information Science. **Journal of the American Society for Information Science**, New Brunswick, v. 50, n. 12, p.1051-1063, 1999.

SAYÃO, Luís Fernando. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 15, n. 30, p.1-31, 2010. Bimestral. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/12528/19527>>. Acesso em: 06 abr. 2014.

SERAFINI, Alfredo. **Apache Solr Beginner's Guide**. Birmingham: Packt Publishing, 2013. 304 p.

SEVERANCE, C.. Discovering JavaScript Object Notation. **Computer**, [S.l.], v. 45, n. 4, p.6-8, abr. 2012. Institute of Electrical & Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/mc.2012.132>.

SEVERANCE, Charles. Roy T. Fielding: Understanding the REST Style. **Computer**, [S.l.], v. 48, n. 6, p.7-9, jun. 2015. Institute of Electrical & Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/mc.2015.170>.

SEVERINO, Antônio Joaquim. **Metodologia do Trabalho Científico**. 23. ed. São Paulo: Cortez, 2007. 304 p.

SIDMAN, David; DAVIDSON, Tom. A Practical Guide to Automating the Digital Supply Chain with the Digital Object Identifier (DOI). **Publishing Research Quarterly: Summer**, Brooklyn, p.9-23, 2001.

SILVA, Maurício Samy. **Ajax com jQuery**: requisições Ajax com a simplicidade de jQuery. São Paulo: Novatec Editora, 2009. 327 p.

SMITH, Ben. **Beginning JSON**: learn the preferred data format of the web. New York: Apress, 2015. 353 p.

SOUSA, Renato Tarciso Barbosa de; ARAÚJO JÚNIOR, Rogério Henrique de. A classificação e a taxonomia como instrumentos efetivos para a recuperação da informação arquivística. **Ciência da Informação**, Brasília, v. 41, n. 1, p.148-160, 2013. Quadrimestral.

STENBERG, Daniel. **CURL: 17 YEARS OLD TODAY**. 2015. Disponível em: <<http://daniel.haxx.se/blog/2015/03/20/>>. Acesso em: 25 abr. 2015.

STENBERG, Daniel. **Everything cURL**. Estocolmo: Ebook, 2016. 179 p. Disponível em: <<http://bookcurl.haxx.se>>. Acesso em: 04 fev. 2016.

STREHL, Leticia. As folksonomias entre os conceitos e os pontos de acesso: as funções de descritores, citações e marcadores nos sistemas de recuperação da informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 16, n. 2, p.101-114, jun. 2011. Trimestral.

TEIXEIRA, Cenidalva Miranda de Sousa; SCHIEL, Ulrich. A Internet e seu impacto nos processos de recuperação da informação. **Ciência da Informação**, Brasília, v. 26, n. 1, p.09-20, 1997. Quadrimestral.

TEIXEIRA, Fábio Augusto Guimarães. **A Recuperação da Informação e a colaboração de usuários na Web**. 2010. 160 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2010. Cap. 7.

TEIXEIRA, Robson da Silva. Serviço de Recuperação da Informação na Biblioteca de um Laboratório Farmacêutico: um Estudo Prático. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 2, n. 2, p.80-89, jun. 2005. Semestral.

TESTAL, Cristina García. Digital object identifier. **El Profesional de la Información**, Barcelona, v. 10, n. 7, p.26-31, ago. 2001. Bimestral. Sistemas de Información.

TONKIN, Emma. Persistent Identifiers: Considering the Options. **Ariadne**, Bath, Uk, n. 56, p.1-10, 30 jul. 2008. Trimestral. Disponível em: <<http://www.ariadne.ac.uk/issue56/tonkin/>>. Acesso em: 10 abr. 2014.

VAN RIJSBERGEN, C. J.. **Information Retrieval**. 2. ed. Glasgow: University Of Glasgow, 1979. 147 p.

VIEIRA, Sonia. **Como Elaborar Questionários**. São Paulo: Atlas, 2009. 159 p.

WANG, Jue. Digital Object Identifiers and Their Use in Libraries. **Serials Review**, [S.l.], v. 33, n. 3, p.161-164, set. 2007. Informa UK Limited. <http://dx.doi.org/10.1016/j.serrev.2007.05.006>.

WEBBER, Jim; PARASTATIDIS, Savas; ROBINSON, Ian. **REST in Practice: Hypermedia and Systems Architecture**. Sebastopol: O'Reilly Media, Inc., 2010. 430 p.

WEBER, Claudiane. Biblioteca Universitária e aplicação do Digital Object Identifier - DOI. In: AMBONI, Narcisa de Fátima (Org.). **Gestão de Bibliotecas Universitárias: experiências e projetos da UFSC**. Florianópolis: Biblioteca Universitária - UFSC, 2013. Cap. 6. p. 79-86.

WEIKUM, Gerhard et al.. Database and Information Retrieval Methods for Knowledge Discovery. **Communications of the ACM**, [S.l.], v. 52, n. 4, p.56-64, abr. 2009. Mensal.

WU, Mingfang et al. Cost and Benefit Estimation of Experts' Mediation in an Enterprise Search. **Journal of the Association for Information Science and Technology**, Wiley Online Library, v. 65, p.146-163, 2013.

XAVIER, Raphael Figueiredo. **Análise de Métodos de Produção de Interfaces Visuais para Recuperação da Informação**. 2009. 78 f. Dissertação (Mestrado) - Curso de Mestrado em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2009. Cap. 5.

## APÊNDICE A – Excerto do Código Fonte do Protótipo

```

.....
if(!empty($doi)){
    $output = 0;
    $ch = curl_init(); // inicia módulo curl
    $url = "http://data.crossref.org/".$doi; //local do recurso

    $header[0] = "Accept: application/vnd.citationstyles.csl+json";

    curl_setopt($ch, CURLOPT_URL,$url);
    curl_setopt($ch, CURLOPT_HTTPHEADER, $header);
    curl_setopt($ch, CURLOPT_RETURNTRANSFER, true);
    // retorna no formato de texto
    $output = curl_exec ($ch); // submete a requisição

    $output2 = json_decode($output, TRUE); // transforma JSON em PHP

    if(!is_array($output2)){ // se não for um array
        $url =
"http://data.datacite.org/application/vnd.citationstyles.csl+json/".$doi;

        $header[0] = "Accept: application/vnd.citationstyles.csl+json";
        curl_setopt($ch, CURLOPT_URL,$url);
        curl_setopt($ch, CURLOPT_HTTPHEADER, $header);
        curl_setopt($ch, CURLOPT_RETURNTRANSFER, true);
        $output = curl_exec ($ch); // execute

        $output2 = json_decode($output, TRUE);
    }

    if(!is_array($output2)){
        $url = "http://data.medra.org/".$doi;
        $header[0] = "Accept: application/vnd.citationstyles.csl+json";
        curl_setopt($ch, CURLOPT_URL,$url);
        curl_setopt($ch, CURLOPT_HTTPHEADER, $header);
        curl_setopt($ch, CURLOPT_RETURNTRANSFER, true);
        $output = curl_exec ($ch);

        $output2 = json_decode($output, TRUE);
    }

    curl_close ($ch); // encerra o manipulador curl

    $dados['msg_doi'] = ''; // cria o objeto que será enviado ao usuário

    if(!is_array($output2)){ // verifica o recebimento da info solicitada
        // caso negativo, informa o usuário
        $dados['msg_doi'] = '<div class="alert-box alert"><p>
O DOI informado ainda não está cadastrado na
Autoridade de Registro da IDF.</p></div>';
    }
}
}
.....

```



## APÊNDICE B – Questionário

### **Pesquisa sobre funcionalidade implementada no MORE**

A presente pesquisa foi elaborada pela equipe do MORE a fim de colher as opiniões dos usuários sobre a funcionalidade de autopreenchimento de alguns campos do formulário de geração de referência a artigos de revistas ("journals").

Esclarecemos que as respostas são confidenciais e os dados apurados servirão, única e exclusivamente, para fins científicos (avaliar a percepção de utilidade da funcionalidade implementada e a satisfação do usuário) e para orientar a equipe do MORE na concepção de novas funcionalidades para o sistema.

Para validação dos dados, solicita-se o preenchimento de todas as questões objetivas (assinalando uma das opções apresentadas).

Questão nº 1 - O tempo utilizado para preencher o formulário diminuiu.

*Marcar apenas uma opção.*

- Concordo totalmente
- Concordo moderadamente
- Concordo ligeiramente
- Discordo ligeiramente
- Discordo moderadamente
- Discordo totalmente

Questão nº 2 - A utilização do DOI facilitou o preenchimento do formulário.

*Marcar apenas uma opção.*

- Concordo totalmente
- Concordo moderadamente
- Concordo ligeiramente
- Discordo ligeiramente
- Discordo moderadamente
- Discordo totalmente

Questão nº 3 - A utilização do DOI tornou mais prática a geração da referência bibliográfica.

*Marcar apenas uma opção.*

- Concordo totalmente
- Concordo moderadamente

- Concordo ligeiramente
- Discordo ligeiramente
- Discordo moderadamente
- Discordo totalmente

Questão nº 4 - Havendo oportunidade, voltarei a utilizar um DOI para o preenchimento do formulário.

*Marcar apenas uma opção.*

- Concordo totalmente
- Concordo moderadamente
- Concordo ligeiramente
- Discordo ligeiramente
- Discordo moderadamente
- Discordo totalmente

Questão nº 5 - Sou favorável à implementação de funcionalidades deste tipo.

*Marcar apenas uma opção.*

- Concordo totalmente
- Concordo moderadamente
- Concordo ligeiramente
- Discordo ligeiramente
- Discordo moderadamente
- Discordo totalmente

Questão nº 6 - A utilização da funcionalidade apresentada melhora minha produtividade.

*Marcar apenas uma opção.*

- Concordo totalmente
- Concordo moderadamente
- Concordo ligeiramente
- Discordo ligeiramente
- Discordo moderadamente
- Discordo totalmente

Questão nº 7 - Gostaria de contar com esta funcionalidade para outros tipos de fontes bibliográficas..

*Marcar apenas uma opção.*

- Concordo totalmente
- Concordo moderadamente
- Concordo ligeiramente

- ( ) Discordo ligeiramente
- ( ) Discordo moderadamente
- ( ) Discordo totalmente

Questão nº 8 - O autopreenchimento de alguns campos do formulário ajuda a diminuir a ocorrência de erros de ortografia.

*Marcar apenas uma opção.*

- ( ) Concordo totalmente
- ( ) Concordo moderadamente
- ( ) Concordo ligeiramente
- ( ) Discordo ligeiramente
- ( ) Discordo moderadamente
- ( ) Discordo totalmente

Questão nº 9 - Fiquei satisfeito com a utilização da funcionalidade de autopreenchimento.

*( ) Marcar apenas uma opção.*

- ( ) Concordo totalmente
- ( ) Concordo moderadamente
- ( ) Concordo ligeiramente
- ( ) Discordo ligeiramente
- ( ) Discordo moderadamente
- ( ) Discordo totalmente

Questão nº 10 - Geralmente fico satisfeito com as novas funcionalidades implementadas pela equipe do MORE.

*Marcar apenas uma opção.*

- ( ) Concordo totalmente
- ( ) Concordo moderadamente
- ( ) Concordo ligeiramente
- ( ) Discordo ligeiramente
- ( ) Discordo moderadamente
- ( ) Discordo totalmente

Questão nº 11 - Fique a vontade para fazer suas críticas e/ou sugestões, no sentido de melhorar as funcionalidades e a usabilidade do MORE.

.....  
.....  
.....



## ANEXO A - Exemplo de Registro MARCXML

```

<?xml version="1.0" encoding="UTF-8"?>
<record xmlns="http://www.loc.gov/MARC21/slim"
  xmlns:cinclude="http://apache.org/cocoon/include/1.0"
  xmlns:zs="http://www.loc.gov/zing/srw/">
  <leader>01189cam a2200325 a 4500</leader>
  <controlfield tag="001">15922449</controlfield>
  <controlfield tag="005">20141221150918.0</controlfield>
  <controlfield tag="008">090925s2009 nyua 000 1 eng d</controlfield>
  <datafield tag="906" ind1=" " ind2=" ">
    <subfield code="a">7</subfield>
    <subfield code="b">cbc</subfield>
    <subfield code="c">copycat</subfield>
    <subfield code="d">2</subfield>
    <subfield code="e">ncip</subfield>
    <subfield code="f">20</subfield>
    <subfield code="g">y-gencatlg</subfield>
  </datafield>
  <datafield tag="925" ind1="0" ind2=" ">
    <subfield code="a">acquire</subfield>
    <subfield code="b">2 shelf copies</subfield>
    <subfield code="x">AY, 2014-12-21</subfield>
  </datafield>
  <datafield tag="955" ind1=" " ind2=" ">
    <subfield code="b">rg04 2009-09-25 z-processor</subfield>
    <subfield code="i">rg04 2009-09-25 to CALM</subfield>
    <subfield code="a">Barcode 00240193838 returned from bindery 2009-09-
29;</subfield>
    <subfield code="t">Copy 2 to CALM</subfield>
  </datafield>
  <datafield tag="010" ind1=" " ind2=" ">
    <subfield code="a"> 2009464840</subfield>
  </datafield>
  <datafield tag="020" ind1=" " ind2=" ">
    <subfield code="a">9780385504225</subfield>
  </datafield>
  <datafield tag="035" ind1=" " ind2=" ">
    <subfield code="a">(OCoLC)ocn399843314</subfield>
  </datafield>
  <datafield tag="040" ind1=" " ind2=" ">
    <subfield code="a">UPZ</subfield>
    <subfield code="c">UPZ</subfield>
    <subfield code="d">DOV</subfield>
    <subfield code="d">JZ6</subfield>
    <subfield code="d">DCB</subfield>
    <subfield code="d">KEU</subfield>
    <subfield code="d">UUC</subfield>
    <subfield code="d">BTCTA</subfield>
    <subfield code="d">YDXCP</subfield>
    <subfield code="d">ABG</subfield>
    <subfield code="d">PN4</subfield>
    <subfield code="d">DLC</subfield>
  </datafield>
  <datafield tag="042" ind1=" " ind2=" ">
    <subfield code="a">lccopycat</subfield>
  </datafield>
  <datafield tag="043" ind1=" " ind2=" ">
    <subfield code="a">n-us-dc</subfield>
  </datafield>
  <datafield tag="050" ind1="0" ind2="0">
    <subfield code="a">PS3552.R685434</subfield>
    <subfield code="b">L67 2009</subfield>
  </datafield>
  <datafield tag="082" ind1="0" ind2="0">
    <subfield code="a">813/.54</subfield>
    <subfield code="2">22</subfield>
  </datafield>
  <datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">Brown, Dan,</subfield>
    <subfield code="d">1964-</subfield>

```

```

</datafield>
<datafield tag="245" ind1="1" ind2="4">
  <subfield code="a">The lost symbol :</subfield>
  <subfield code="b">a novel /</subfield>
  <subfield code="c">Dan Brown.</subfield>
</datafield>
<datafield tag="250" ind1=" " ind2=" " >
  <subfield code="a">1st ed.</subfield>
</datafield>
<datafield tag="260" ind1=" " ind2=" " >
  <subfield code="a">New York :</subfield>
  <subfield code="b">Doubleday,</subfield>
  <subfield code="c">c2009.</subfield>
</datafield>
<datafield tag="300" ind1=" " ind2=" " >
  <subfield code="a">509 p. :</subfield>
  <subfield code="b">ill. ;</subfield>
  <subfield code="c">25 cm.</subfield>
</datafield>
<datafield tag="520" ind1=" " ind2=" " >
  <subfield code="a">Symbolist Robert Langdon returns in this new thriller
    follow-up to The Da Vinci Code.</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="0">
  <subfield code="a">Cryptographers</subfield>
  <subfield code="v">Fiction.</subfield>
</datafield>
<datafield tag="650" ind1=" " ind2="0">
  <subfield code="a">Freemasonry</subfield>
  <subfield code="v">Fiction.</subfield>
</datafield>
<datafield tag="651" ind1=" " ind2="0">
  <subfield code="a">Washington (D.C.)</subfield>
  <subfield code="v">Fiction.</subfield>
</datafield>
<datafield tag="655" ind1=" " ind2="7">
  <subfield code="a">Suspense fiction.</subfield>
  <subfield code="2">gsafd</subfield>
</datafield>
<datafield tag="655" ind1=" " ind2="7">
  <subfield code="a">Adventure fiction.</subfield>
  <subfield code="2">gsafd</subfield>
</datafield>
</record>

```

Fonte: Adaptado de <<https://lcn.loc.gov/2009464840/marcxml>>. Acesso em 31 jan. 2016.