

UNIVERSIDADE ESTADUAL PAULISTA  
CAMPUS DE MARÍLIA  
FACULDADE DE FILOSOFIA E CIÊNCIAS

ZAIRA REGINA ZAFALON

SCAN FOR MARC: PRINCÍPIOS SINTÁTICOS E SEMÂNTICOS DE REGISTROS  
BIBLIOGRÁFICOS APLICADOS À CONVERSÃO DE DADOS ANALÓGICOS PARA  
O FORMATO MARC21 BIBLIOGRÁFICO

MARÍLIA  
2012

ZAIRA REGINA ZAFALON

SCAN FOR MARC: PRINCÍPIOS SINTÁTICOS E SEMÂNTICOS DE REGISTROS  
BIBLIOGRÁFICOS APLICADOS À CONVERSÃO DE DADOS ANALÓGICOS PARA  
O FORMATO MARC21 BIBLIOGRÁFICO

Tese apresentada ao Programa de Pós-graduação em Ciência da Informação da Universidade Estadual Paulista como requisito parcial para a obtenção do título de Doutor em Ciência da Informação

Área de concentração: Informação, Tecnologia e Conhecimento

Orientadora: Profa. Dra. Plácida Leopoldina Ventura Amorim da Costa Santos

MARÍLIA

2012

Zafalon, Zaira Regina  
Z22s      Scan for MARC : princípios sintáticos e semânticos de registros bibliográficos aplicados à conversão de dados analógicos para o Formato MARC21 Bibliográfico / Zaira Regina Zafalon. -- 2012.  
169 f. : il.

Tese (Doutorado em Ciência da Informação)–Faculdade de Filosofia e Ciências da Universidade Estadual Paulista, Marília, 2012.

Orientadora: Profa. Dra. Plácida Leopoldina Ventura Amorim da Costa Santos.

1. Registros bibliográficos - Sintaxe 2. Registros bibliográficos - Semântica 3. Registros bibliográficos - Conversão 4. AACR2 5. ISBD 6. MARC21 Bibliográfico I. Título.

CDD 025.3

ZAIRA REGINA ZAFALON

SCAN FOR MARC: PRINCÍPIOS SINTÁTICOS E SEMÂNTICOS DE REGISTROS  
BIBLIOGRÁFICOS APLICADOS À CONVERSÃO DE DADOS ANALÓGICOS PARA  
O FORMATO MARC21 BIBLIOGRÁFICO

Tese apresentada ao Programa de Pós-  
graduação em Ciência da Informação da  
Universidade Estadual Paulista como  
requisito parcial para a obtenção do título de  
Doutor em Ciência da Informação

Área de concentração: Informação,  
Tecnologia e Conhecimento

---

Profa. Dra. Plácida Leopoldina Ventura Amorim da Costa Santos (orientadora)  
Faculdade de Filosofia e Ciências da Universidade Estadual Paulista

---

Profa. Dra. Dulce Maria Baptista  
Faculdade de Ciência da Informação da Universidade de Brasília

---

Prof. Dr. Edberto Ferneda  
Faculdade de Filosofia e Ciências da Universidade Estadual Paulista

---

Profa. Dra. Elisa Campos Machado  
Universidade Federal do Estado do Rio de Janeiro

---

Prof. Dr. Ricardo César Gonçalves Santana  
Campus Experimental de Tupã da Universidade Estadual Paulista

Marília, 29 de junho de 2012



Aos que me ensinaram  
a importância de se  
pensar o bem,  
falar o bem e  
fazer o bem.

## AGRADECIMENTOS

Toda e qualquer forma que venha a ser adotada para registrar o meu reconhecimento pelo quanto foi feito por mim durante o período de desenvolvimento deste estudo se esvai no transcorrer do tempo e se dilui no espaço físico do papel. O mesmo não acontece com as marcas deixadas em mim.

Obrigada a todos!

Este caminhar, como em outros de minha vida, foi marcado por diferentes situações, muitas vezes antagônicas: de alegria e de tristeza, de ganhos e perdas; mas todas, sem dúvida alguma, foram de aprendizado e amadurecimento.

Pela oportunidade de um novo recomeço a cada minuto, agradeço a *Deus*.

Pelos ensinamentos deixados, pelo exemplo de vida, retidão e resignação diante da vida, agradeço a *Cristo*.

Pela presença marcante, pelas orientações, sinalizações e ponderações sobre o melhor caminho, por saber e por conseguir me conduzir, pelo voto de confiança, concedido quando até eu mesma desacreditava que pudesse dar certo, agradeço à Profa. Dra. *Plácida Leopoldina Ventura Amorim da Costa Santos*, sem dúvida, exemplo a ser seguido.

Pela oportunidade que me foi dada para aprender e discutir, pela rica convivência e por tornarem-se exemplos profissionais, agradeço aos docentes do Programa de Pós-Graduação em Ciência da Informação da UNESP: *Edberto Ferneda*, *Ricardo César Gonçalves Sant'Ana* e *Silvana Aparecida Borsetti Gregório Vidotti*, e aos docentes visitantes: *Ana Alice Rodrigues Pereira Baptista* (Universidade do Minho, Portugal), *Alan Story* (Universidade de Kent, Reino Unido) e *Nanci Oddone* (Universidade Federal da Bahia).

Pelas contribuições na qualificação, todas pertinentes e que espero ter saldado, pela leitura atenta e crítica, agradeço à Profa. Dra. Dulce Maria Baptista e ao Prof. Dr. Ricardo César Gonçalves Santana.

Por comporem a Banca Examinadora, agradeço à Profa. Dra. Dulce Maria Baptista, ao Prof. Dr. Edberto Ferneda, à Profa. Dra. Elisa Campos Machado, e ao Prof. Dr. Ricardo César Gonçalves Santana.

Pelo incentivo institucional à formação contínua e por oferecer o Programa de Apoio à Capacitação de Servidores, agradeço à UFSCar.

Pela colaboração, apoio e compreensão decorrentes do afastamento parcial das atividades, agradeço ao Departamento de Ciência da Informação da UFSCar, ao qual estou vinculada.

Pelas conversas, pelo incentivo, pela compreensão tanto de minha ausência quanto de minha presença fugaz, pelos depoimentos de já terem passado por isso, agradeço aos colegas de departamento: *Ariadne Chloe Mary Furnival, Camila Carneiro Dias Rigolin, Carlos Roberto Massao Hayashi, Leandro Innocentini Lopes de Faria, Luciana de Souza Gracioso, Luzia Sigoli Fernandes Costa, Maria Cristina Comunian Ferraz, Maria Cristina Piumbato Innocentini Hayashi, Nádea Regina Gaspar, Rogério Aparecido Sá Ramalho, Roniberto Morato do Amaral, Sérgio Luís da Silva, Vera Regina Casari Boccato, Wanda Aparecida Machado Hoffmann.*

Por me acolherem como aluna, com dúvidas, questões, anseios e vontade de aprender, agradeço aos colegas da UFSCar: *Flávia Bezerra de Menezes Hirata Vale, Oto Araújo Vale e Lúcia Helena Machado Rino.*

Por dispor-se a discutir aspectos iniciais sobre a teoria de Saussure, agradeço a *Nádea Regina Gaspar.*

Por partilharem comigo a vivência do ensino, da pesquisa e da extensão na Biblioteconomia e Ciência da Informação, na UFSCar e no UNIFAI, mas, também, por compreenderem as minhas fugas necessárias para terminar essa pesquisa, agradeço aos que estão e aos que já foram alunos e colegas.

Pela parceria, pelo quanto me ensina a viver, a conviver e a bem viver, por ser quem é, por me fazer rir quando menos me é possível, por fazer parte do meu dia-a-dia, por torná-lo mais feliz, divertido e leve e por completar a minha vida, agradeço ao meu filho, *Pedro Zafalon Gomes.*

Pelo apoio durante a jornada São Carlos-Marília, pelo suporte ao Pedro em minhas ausências, por acreditarem que eu seria capaz, por me incentivarem a continuar, por ser para quem recorro e de quem, às vezes, fujo, agradeço à minha família: meu pai, *Didier*; minha mãe, *Neide*; meus irmãos, *Luiz Antonio, Marcia e Junior*; meus sobrinhos, *Gabriel, Camila, Thiago e Júlia*; minha avó, *Victória*, minha cunhada *Patrícia.*

Por me incentivar a sonhar, e por ter sonhado comigo enquanto houve tempo, pela convivência, e também, pelo que aprendi com a sua ausência, agradeço ao *Marcio Rillo (in memoriam).*

Por cuidar de mim e do Pedro, de nossos amigos animais e de onde moramos, agradeço ao empenho da *Fátima Aparecida Dias de Souza*.

Por topar fazer parte das minhas tentativas e erros, por não se esquivar dos meus pedidos, mesmo que inicialmente mal formulados, por ser incansável e sempre solícito, agradeço ao amigo *Jairo da Silva*, parceiro de anos e também na empreitada do Scan for MARC.

Pela amizade, parceria e presença constantes, pelo companheirismo e pela concessão de ombro amigo, pela disposição em discutir, pensar e repensar a doutrina, a ciência e a filosofia que partilhamos, por me ensinar a ser uma pessoa melhor e por me compreender em meus erros e acertos, agradeço à amiga *Rosângela Castilho Alcaraz Moraes*.

Pelo apoio, colaboração e incentivo nos momentos de ansiedade agradeço ao *Edberto Ferneda*, que também se tornou amigo.

Pela iniciação na área de catalogação e pelo reencontro amigo, agradeço a *Sônia Maria Trombelli*.

Pela amizade, pela oportunidade de compartilhar a área de catalogação e por acreditar que a loucura de uma idéia poderia virar um tema de tese, agradeço à *Eliane Serrão Alves Mey*.

Por partilhar a vivência acadêmica em Marília de modo tão próximo, por percorrer, na confiança de que chegaríamos logo, a longa estrada até São Carlos, por compartilhar das minhas, e por nos sentirmos sempre em família, agradeço à *Carolina Segatto Vianna*.

Por nos conhecermos em sala de aula, por vivenciarmos discussões, tolices e bobagens, risadas, caronas e amizade, agradeço aos colegas: *Ana Maria Ferreira, André Coneglian, André Sávio Bueno, Camila Araújo dos Santos, Camila Narukawa, Fernando Alves da Gama, Fernando Luiz Vechiato, Flávia Bastos, Januário Albino Nhacuongue, Letícia Molina, Luana Woida, Miriam Mattos, Mariana Lousada, Régis Garcia, Tamara Guaraldo e Walter Oliveira*.

Pela amizade, mas, neste caso, pelos constantes auxílios com a bibliografia, concedidos de modo incondicional, agradeço ao *Nivaldo Aparecido Coelho*, da Biblioteca da EESC/USP, e à *Eliane Colepícolo*, à *Jandira Ferreira de Jesus Rossi*, à *Sheila Regina Rizzo César*, à *Silvana Alice Maragno*, à *Gislene Taís de Souza Sperandio*, da Biblioteca Comunitária da UFSCar.

"[...] é o ponto de vista que cria o objeto".  
(SAUSSURE, 2010, p. 15).

"Não há nem conhecimento nem descrição científica possível de um objeto qualquer sem recurso a um princípio estrutural.

(HJELMSLEV, 1991, p. 116).

## RESUMO

A pesquisa apresenta como tema nuclear o estudo do processo de conversão de registros bibliográficos. Delimita-se o objeto de estudo pelo entendimento da conversão de registros bibliográficos analógicos para o formato MARC21 Bibliográfico, a partir da análise sintática e semântica de registros descritos segundo padrões de estrutura de metadados descritivos e padrões de conteúdo. A tese nesta pesquisa é a de que os princípios sintáticos e semânticos de registros bibliográficos, definidos pelos esquemas de descrição e de visualização na catalogação, presentes nos padrões de estrutura de metadados descritivos e nos padrões de conteúdo, determinam o processo de conversão de registros bibliográficos para o Formato MARC21 Bibliográfico. Em vista desse panorama, a proposição desta pesquisa é desenvolver um estudo teórico sobre a sintaxe e a semântica de registros bibliográficos, pelo viés da Linguística, com Saussure e Hjelmslev, que subsidiem a conversão de registros bibliográficos analógicos para o Formato MARC21 Bibliográfico em um interpretador computacional. Com esta proposta, estabelece-se, como objetivo geral, desenvolver um modelo teórico-conceitual de sintaxe e semântica em registros bibliográficos, a partir de estudos lingüísticos saussureanos e hjelmslevianos das manifestações da linguagem humana, que seja aplicável a um interpretador computacional voltado à conversão de registros bibliográficos ao formato MARC21 Bibliográfico. Para o alcance de tal objetivo recorre-se aos seguintes objetivos específicos, reunidos em dois grupos e voltados, respectivamente ao modelo teórico-conceitual da estrutura sintática e semântica de registros bibliográficos, e ao processo de conversão de seus registros: explicitar a relação entre a sintaxe e a semântica de registros bibliográficos à luz dos aspectos lingüísticos estudados por Saussure e por Hjelmslev, aplicáveis aos meios analógicos e digitais; apresentar princípios característicos da sintaxe e da semântica em registros bibliográficos; examinar, por meio da literatura disponível, os métodos e os processos de conversão de registros bibliográficos ao formato MARC21 Bibliográfico; definir método de processamento de imagens de registros bibliográficos analógicos em *software* de reconhecimento ótico de caracteres (OCR); apresentar técnica de interpretação computacional que seja voltada à conversão de registros bibliográficos analógicos para o Formato MARC21 Bibliográfico que

considere os esquemas sintáticos e semânticos. Em vista dos objetivos delineados, o percurso metodológico da pesquisa baseia-se na abordagem qualitativa, de caráter exploratório, descritivo e experimental e recorre à pesquisa bibliográfica. Como resultados, espera-se contribuir de modo significativo à Ciência da Informação, considerando aspectos sociais, inerentes ao conhecimento humano e à solução do problema apresentado. O maior benefício social ensejado por essa pesquisa relaciona-se, intimamente, aos aspectos do compartilhamento de registros bibliográficos, bem como à possibilidade de ampliar o acesso aos metadados dos recursos informacionais representados pelas diversas unidades de informação. Contribuições no plano teórico podem ser vislumbradas diante do desenvolvimento de questões inerentes aos aspectos sintáticos e semânticos de registros bibliográficos, e por envolver, a um só tempo, a interdisciplinaridade entre a Ciência da Informação, a Ciência da Computação e a Linguística. Contribuições no campo prático são identificadas pelo fato de o estudo abranger o desenvolvimento do Scan for MARC, um interpretador computacional que possa ser adotado por quaisquer instituições que queiram utilizar-se de procedimento de conversão de bases de dados de registros bibliográficos, para o formato MARC21 Bibliográfico a partir dos esquemas de descrição e de visualização de registros bibliográficos (AACR2r e ISBD), caráter que se considera inovador na pesquisa.

**Palavras-chave:** Conversão de registros bibliográficos. Formato MARC21 Bibliográfico. Análise sintática de registros bibliográficos. Análise semântica de registros bibliográficos. ISBD. AACR2r. Padrão de estrutura de metadados descritivos. Padrão de conteúdo.

## **ABSTRACT**

The research presents as its central theme the study of the bibliographic record conversion process. The object of study is framed by an understanding of analogic bibliographic record conversion to the Bibliographic MARC21 format, based on a syntactic and semantic analysis of records described according to descriptive metadata structure standards and content standards. The thesis in this research is that the syntactic and semantic principles of bibliographic records, defined by description and visualization cataloguing schemes, present in the descriptive metadata structure standards and content standards, determine the bibliographic record conversion process to the MARC21 Bibliographic Format. In the light of this, the purpose of this research is to develop a theoretical study of the syntax and semantics of bibliographic records, grounded in Linguistic theories of Saussure and Hjelmslev, which can underlie analogic bibliographic record conversion to the MARC21 Bibliographic Format using a computational interpreter. To this end, the general aim was to develop a theoretical-conceptual model of the syntax and semantics of bibliographic records, based on saussurean and hjelmslevian linguistic studies of human language manifestations, which can be applicable to a computational interpreter designed for the conversion of bibliographic records to the MARC21 Bibliographic Format. To attain this goal, the following specific objectives were identified, in two groups and related to the theoretical-conceptual model of bibliographic record syntax and semantics and to the conversion process of the records, respectively: to make explicit the relationship between the syntax and semantics of bibliographic records in the light of linguistic aspects studied by Saussure and Hjelmslev, applicable to analogical and digital environments; to present principles characteristic of the syntax and semantics of bibliographic records; to examine, from the available literature, the methods and processes of conversion of bibliographic records to the MARC21 Bibliographic Format; to define an image processing method for analogic bibliographic records in Optical Character Recognition (OCR) software; to present a technique for computational interpretation aimed at the conversion of analogic bibliographic records to the MARC21 Bibliographic Format which considers syntactic and semantic schemes. Given the aforementioned objectives, the methodological trajectory of the research is based on



the qualitative approach, of an exploratory, descriptive and experimental nature, and with recourse to the literature. With the results, it is hoped that a significant contribution will be made to Information Science, considering social aspects inherent to human knowledge and to the solution of the problem presented. The greatest social benefit sought with this research is inextricably related to aspects of bibliographic record sharing, as well as to the possibility of broadening access to the metadata of information resources found in a variety of information units. Contributions on the theoretical plane can be envisaged regarding the development of questions inherent to the syntactic and semantic aspects of bibliographic records, and by involving, at the same time, interdisciplinarity between Information Science, Computer Science and Linguistics. Contributions to the practical field are identified by the fact the study covers the development of the Scan for MARC, a computational interpreter that can be adopted by any institution that wishes to use the conversion procedure for bibliographic record databases to the MARC21 Bibliographic Format from description and visualization schemes of bibliographic records (AACR2r and ISBD), an aspect of the research which is considered innovative.

**Keywords:** Bibliographic record conversion. MARC21 Bibliographic Format. Syntactic analysis of bibliographic records. Semantic analysis of bibliographic records. ISBD. AACR2r. Descriptive metadata structure standard. Content standard.

## LISTA DE FIGURAS

Figura 1 – Nomes das etiquetas de campo do sistema ‘X’ .....	27
Figura 2 – Nomes das etiquetas de campo do sistema ‘Y’ .....	28
Figura 3 – Representação de um documento segundo o <i>padrão de conteúdo</i> definido pelas AACR2r e o <i>padrão de estrutura de metadados descritivos</i> indicados pela ISBD.....	44
Figura 4 – Esquema clássico da comunicação .....	47
Figura 5 – Esquema do processo comunicativo entre recursos informacionais e usuários nas instituições de patrimônio cultural .....	47
Figura 6 – Esquema do processo comunicativo para a catalogação cooperativa.....	58
Figura 7 – Os três conceitos do modelo sógnico diádico de Saussure .....	69
Figura 8 – Os três termos do modelo para a representação da informação .....	70
Figura 9 – Representação gráfica das relações entre <i>obra</i> , <i>expressão</i> e <i>manifestação</i> ( <i>Romeo and Juliet</i> ).....	71
Figura 10 – Representação gráfica das relações entre <i>obra</i> , <i>expressão</i> e <i>manifestação</i> ( <i>Ninth symphony</i> ).....	72
Figura 11 – Imagens de registros bibliográficos obtidos em impressora multifuncional .....	86
Figura 12 – Imagens de registros bibliográficos capturadas a partir de câmera de dispositivo móvel com distância aproximada de 30 cm.....	88
Figura 13 – Imagens de registros bibliográficos capturadas a partir de câmera de dispositivo móvel com distância aproximada de 20 cm.....	89
Figura 14 – Imagens de registros bibliográficos capturadas a partir de câmera de dispositivo móvel com distância aproximada de 10 cm.....	91
Figura 15 – Imagens de registros bibliográficos capturadas a partir de câmera fotográfica digital com distância aproximada de 30 cm.....	93
Figura 16 – Imagens de registros bibliográficos capturadas a partir de câmera fotográfica digital com distância aproximada de 20 cm.....	94
Figura 17 – Imagens de registros bibliográficos capturadas a partir de câmera fotográfica digital com distância aproximada de 10 cm.....	96
Figura 18 – Imagem de registro bibliográfico analógico utilizado para os testes do processamento de imagem com reconhecimento de caracteres .....	99

Figura 19 – Processamento da imagem do Registro 1, capturada em scanner de impressora multifuncional, visualizável em arquivo (download) .....	101
Figura 20 – Processamento da imagem do Registro 2, capturada em scanner de impressora multifuncional, visualizável em arquivo (download) .....	101
Figura 21 – Processamento da imagem do Registro 3, capturada em scanner de impressora multifuncional, visualizável em arquivo (download) .....	102
Figura 22 – Processamento da imagem do Registro 1, capturada em câmera de dispositivo móvel em distância aproximada de 30 cm, visualizável em arquivo (download).....	103
Figura 23 – Processamento da imagem do Registro 2, capturada em câmera de dispositivo móvel em distância aproximada de 30 cm, visualizável em arquivo (download).....	103
Figura 24 – Processamento da imagem do Registro 3, capturada em câmera de dispositivo móvel em distância aproximada de 30 cm, visualizável em arquivo (download).....	103
Figura 25 – Processamento da imagem do Registro 1, capturada em câmera de dispositivo móvel em distância aproximada de 20 cm, visualizável em arquivo (download).....	105
Figura 26 – Processamento da imagem do Registro 2, capturada em câmera de dispositivo móvel em distância aproximada de 20 cm, visualizável em arquivo (download).....	105
Figura 27 – Processamento da imagem do Registro 3, capturada em câmera de dispositivo móvel em distância aproximada de 20 cm, visualizável em arquivo (download).....	105
Figura 28 – Processamento da imagem do Registro 1, capturada em câmera de dispositivo móvel em distância aproximada de 10 cm, visualizável em arquivo (download).....	106
Figura 29 – Processamento da imagem do Registro 2, capturada em câmera de dispositivo móvel em distância aproximada de 10 cm, visualizável em arquivo (download).....	106
Figura 30 – Processamento da imagem do Registro 3, capturada em câmera de dispositivo móvel em distância aproximada de 10 cm, visualizável em arquivo (download).....	107

Figura 31 – Processamento da imagem do Registro 1, capturada em câmera fotográfica digital em distância aproximada de 30 cm, visualizável em arquivo (download).....	108
Figura 32 – Processamento da imagem do Registro 2, capturada em câmera fotográfica digital em distância aproximada de 30 cm, visualizável em arquivo (download).....	109
Figura 33 – Processamento da imagem do Registro 3, capturada em câmera fotográfica digital em distância aproximada de 30 cm, visualizável em arquivo (download).....	109
Figura 34 – Processamento da imagem do Registro 1, capturada em câmera fotográfica digital em distância aproximada de 20 cm, visualizável em arquivo (download).....	110
Figura 35 – Processamento da imagem do Registro 2, capturada em câmera fotográfica digital em distância aproximada de 20 cm, visualizável em arquivo (download).....	110
Figura 36 – Processamento da imagem do Registro 3, capturada em câmera fotográfica digital em distância aproximada de 20 cm, visualizável em arquivo (download).....	111
Figura 37 – Processamento da imagem do Registro 1, capturada em câmera fotográfica digital em distância aproximada de 10 cm, visualizável em arquivo (download).....	112
Figura 38 – Processamento da imagem do Registro 2, capturada em câmera fotográfica digital em distância aproximada de 10 cm, visualizável em arquivo (download).....	112
Figura 39 – Processamento da imagem do Registro 3, capturada em câmera fotográfica digital em distância aproximada de 10 cm, visualizável em arquivo (download).....	112
Figura 40 – Primeiro roteiro para o <i>script</i> de conversão de registros bibliográficos	117
Figura 41 – Resultado de conversão de registro bibliográfico conforme versão 1 do <i>script</i> .....	117
Figura 42 – Segundo roteiro para o <i>script</i> de conversão de registros bibliográficos .....	118
Figura 43 – Resultado de conversão de registro bibliográfico conforme versão 2 do <i>script</i> .....	119

Figura 44 – Terceiro roteiro para o <i>script</i> de conversão de registros bibliográficos	121
Figura 45 – Resultado de conversão de registro bibliográfico conforme <i>versão 2 do script</i> .....	122
Figura 46 – Resultado de conversão de registro bibliográfico conforme <i>versão 3 do script</i> .....	123
Figura 47 – Quarto roteiro para o <i>script</i> de conversão de registros bibliográficos...	124
Figura 48 – Resultado de conversão de registro bibliográfico conforme <i>versão 3 do script</i> .....	124
Figura 49 – Indicação de dados de localização de um registro bibliográfico.....	125
Figura 50 – Resultado de conversão de registro bibliográfico conforme <i>versão 4 do script</i> .....	127

## LISTA DE QUADROS

Quadro 1 – Representação de um documento segundo o <i>padrão de estrutura de metadados descritivos</i> MARC21 Bibliográfico e o <i>padrão de conteúdo</i> AACR2r.....	45
Quadro 2 – Representação dos recursos informacionais em meio automatizado.....	82
Quadro 3 – <i>Softwares</i> para reconhecimento ótico de caracteres em arquivos de imagem .....	98
Quadro 4 – Resumo do resultado dos testes realizados em <i>softwares</i> para reconhecimento ótico de caracteres.....	99
Quadro 5 – Resultado da análise sintática e sua relação semântica nos dados de descrição de um registro bibliográfico .....	116

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	21
<b>2 REPRESENTAÇÃO E CATÁLOGOS NO PROCESSO COMUNICATIVO DE BIBLIOTECAS</b> .....	33
2.1 REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS .....	33
2.2 PADRÕES PARA A REPRESENTAÇÃO BIBLIOGRÁFICA .....	37
2.3 CATÁLOGOS E OBJETIVOS BIBLIOGRÁFICOS .....	46
2.4 COOPERAÇÃO, COMPARTILHAMENTO E INTERCÂMBIO DE REGISTROS BIBLIOGRÁFICOS .....	53
2.5 MÉTODOS E PROCESSOS DE CONVERSÃO DE REGISTROS BIBLIOGRÁFICOS .....	60
<b>3 UM OLHAR DA LINGUÍSTICA SOBRE OS REGISTROS BIBLIOGRÁFICOS</b> ....	66
3.1 CONTRIBUIÇÃO SAUSSUREANA À REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS .....	66
3.2 SEMÂNTICA HJELMSLEVIANA E OS REGISTROS BIBLIOGRÁFICOS .....	73
3.3 PRINCÍPIOS SINTÁTICOS E SEMÂNTICOS DE REGISTROS BIBLIOGRÁFICOS: A REPRESENTAÇÃO COMO ESTRUTURA SEMÂNTICA .....	78
<b>4 SCAN FOR MARC: INTERPRETADOR DE PRINCÍPIOS SINTÁTICOS E SEMÂNTICOS DE REGISTROS BIBLIOGRÁFICOS</b> .....	84
4.1 CAPTURA DE IMAGENS DE REGISTROS BIBLIOGRÁFICOS.....	85
4.1.1 Captura de imagens de registros bibliográficos em scanner de impressora multifuncional .....	86
4.1.2 Captura de imagens de registros bibliográficos em câmera de dispositivo móvel .....	87
4.1.3 Captura de imagens de registros bibliográficos em câmera fotográfica digital.....	92
4.2 ANÁLISE DE <i>SOFTWARES</i> DE RECONHECIMENTO DE CARACTERES.....	97
4.3 PROCESSAMENTO DAS IMAGENS DE REGISTROS BIBLIOGRÁFICOS.....	100
4.3.1 Processamento de imagens de registros bibliográficos capturados em scanner de impressora multifuncional.....	101
4.3.2 Processamento de imagens de registros bibliográficos capturados em câmera de dispositivo móvel .....	102

4.3.3 Processamento de imagens de registros bibliográficos capturados em câmera fotográfica digital .....	108
4.3.4 Resultado geral do processamento de imagens de registros bibliográficos...	113
4.4 TRATAMENTO SINTÁTICO E SEMÂNTICO DE REGISTROS BIBLIOGRÁFICOS PARA O FORMATO MARC21 BIBLIOGRÁFICO .....	114
<b>5 CONSIDERAÇÕES FINAIS</b> .....	129
<b>REFERÊNCIAS</b> .....	133
<b>APÊNDICES</b> .....	145
APÊNDICE A – RESULTADO DO TESTE COM O ABBY FineReader 11.....	146
APÊNDICE B – RESULTADO DO TESTE COM O ABBYY FineReader Online.....	147
APÊNDICE C – RESULTADO DO TESTE COM O Cognitive Open OCR (Cuneiform) 0.1 .....	150
APÊNDICE D – RESULTADO DO TESTE COM O FreeOCR .....	152
APÊNDICE E – RESULTADO DO TESTE COM O FreeOCR 3.1.....	153
APÊNDICE F – RESULTADO DO TESTE COM O Leadtools.....	154
APÊNDICE G – RESULTADO DO TESTE COM O OnlineOCR.net .....	158
APÊNDICE H – RESULTADO DO TESTE COM O ScreenOCR 9.1 .....	159
APÊNDICE I – RESULTADO DO TESTE COM O Sci2ools (i2OCR).....	160
APÊNDICE J – RESULTADO DO TESTE COM O SimpleOCR 3.5.....	162
APÊNDICE K – RESULTADO DO TESTE COM O TopOCR 3.1.....	163
APÊNDICE L – RESULTADO DO TESTE COM O WeOCR Server .....	164
APÊNDICE M – PEDIDO DE REGISTRO DE PROGRAMA DE COMPUTADOR ..	165
APÊNDICE N – PEDIDO DE REGISTRO DE MARCA .....	168



## 1 INTRODUÇÃO

Esta pesquisa apresenta como tema nuclear o estudo do processo de conversão de registros bibliográficos e delimita-se pelo entendimento da conversão de registros bibliográficos para o Formato MARC21 Bibliográfico<sup>1</sup>, a partir da análise sintática, definida pela *International Standard of Bibliographical Description* (ISBD) [Padrão Internacional de Descrição Bibliográfica], e pela análise semântica, definida pelas *Anglo American Cataloging Rules, 2nd Edition* (AACR2) [Regras de Catalogação Anglo-Americana, 2ª edição revista -- traduzido no Brasil em 2004]. Insere-se, portanto, na linha de pesquisa Informação e Tecnologia da área de concentração Informação, Tecnologia e Conhecimento. A relação que se estabelece entre o trabalho desenvolvido e a linha de pesquisa oferecida diz respeito, em um primeiro momento, ao estudo da representação de recursos informacionais e do compartilhamento e conversão de registros bibliográficos em ambiente digital e, em um segundo, não menos importante, como sendo voltado ao desenvolvimento tecnológico. A análise e o desenvolvimento de aspectos teóricos e metodológicos que envolvem tais processos, utilizando-se de ferramentas e métodos inerentes às tecnologias da informação e comunicação, também se enquadram na linha de pesquisa assinalada.

O estudo das formas de representação dos recursos informacionais, bem como a geração, o controle, a transmissão, os processos de gerenciamento de registros bibliográficos em meio automatizado e seu uso, apresenta relação intrínseca à abordagem apresentada por Borko (1968, p. 3, tradução e grifos nossos) sobre a Ciência da Informação:

[...] disciplina que investiga as propriedades e o comportamento informacional, as forças que governam os fluxos de informação, e os *significados do processamento da informação*, visando à acessibilidade e a usabilidade ótima. A Ciência da Informação está preocupada com o corpo de conhecimentos relacionados à origem, à coleção, à *organização*, ao *armazenamento*, à *recuperação*, à *interpretação*, à *transmissão*, à *transformação*, e à utilização da

---

<sup>1</sup> MARC é acrônimo de MACHine READable Cataloging Record. O Formato MARC 21 agrega formatos destinados a dados bibliográficos, dados de autoridade, para holdings, para dados de classificação e para informações de comunidade. Nesta pesquisa serão estudados dados bibliográficos.

informação. Isto inclui a pesquisa sobre a *representação da informação em ambos os sistemas*, tanto *naturais* quanto *artificiais*, o uso de *códigos* para a transmissão eficiente da mensagem, bem como o estudo do processamento e de técnicas aplicadas aos computadores e seus sistemas de programação.

Desse modo, esta pesquisa insere-se no contexto social contemporâneo, no qual se faz uso e aplicação de tecnologias no seu cotidiano. Os recursos tecnológicos e midiáticos, por meio de estruturas computacionais, permeiam a produção, a organização, a distribuição, o acesso, o armazenamento, a preservação, o uso e o reuso dos recursos informacionais por meio de métodos de representação e de recuperação, construindo, conseqüentemente, novo contexto sócio-cultural, educacional, econômico e tecnológico.

Aliado a esses fatores há o crescimento exponencial de acervos documentais em unidades de informação, o que requer a adequação de processos biblioteconômicos que satisfaçam, de modo eficiente e eficaz, a demanda informacional por parte de diferentes públicos. Com as novas tecnologias de informação e comunicação surgem, também, as facilidades de compartilhamento de registros bibliográficos entre as mais variadas unidades de informação.

Esforços no sentido de se estudar temáticas relacionadas com o compartilhamento e com a conversão de registros bibliográficos, a partir de ações cooperativas ou não, foram, e continuam sendo, envidados (nesta pesquisa serão abordados nos tópicos 2.4 e 2.5).

Na literatura nacional, diagnósticos e situações de conversão de bases de dados são relatados no sentido de apresentar as experiências e os procedimentos metodológicos utilizados em diversas instituições. Estas publicações, de certa forma, tendem a deixar de explicitar as concepções de ordem teórico-metodológica envolvidas nesse processo. Identifica-se nesses textos, entretanto, que cada instituição que necessite estabelecer o processo de conversão de seus dados bibliográficos para o Formato MARC21 Bibliográfico parte do estudo, conhecimento e delineamento da estrutura do seu banco de dados da base a ser convertida. Essa situação pode ser vislumbrada a partir de Silva e Favaretto (2005) quando afirmam que, diante da diversidade da estrutura das bases de dados, as ações de conversão são dificultadas.

A análise do sistema atual – que são as tabelas e campos, os relacionamentos entre tabelas (somente aqueles envolvidos no processo de migração) e sua política de funcionamento – deve ser

documentada e mapeada para que o fornecedor do novo *software* possa fazer a ligação com as tabelas e campos equivalentes no novo sistema. Ou seja, as necessidades da biblioteca já são especificadas em relação ao futuro sistema. Também é importante especificar informações adicionais, como códigos de subbibliotecas, códigos de tipos de obra, entre outros. (SILVA; FAVARETTO, 2005, p. 108-109).

Note-se que normas, padrões, metodologias e critérios são requeridos em unidades de informação e tornam-se essenciais em processos de gerenciamento, tratamento, recuperação, provisão, uso e compartilhamento de registros informacionais em diversos suportes. Buckland (1997) afirma que o processo de automação é facilitado diante da adoção de códigos de catalogação padronizados, o que resulta em procedimentos eficazes que intrinsecamente apresentam maior relação com aspectos estratégicos da biblioteca que com os computadores em si.

Formatos para o reconhecimento, o processamento automatizado e o intercâmbio de registros bibliográficos têm sido desenvolvidos desde a década de 1960, capitaneados pela *Library of Congress* (EUA). O padrão de estrutura de metadados descritivos que, se não internacional, tornou-se internacionalizado no processo de compartilhamento de registros bibliográficos é o MARC, acrônimo de *Machine Readable Cataloging*. Tal padrão é um dos instrumentos indispensáveis ao catalogador que intenciona prover de meios automatizados de acesso aos, e compartilhamento de, registros bibliográficos na unidade de informação em que atua. Embora este padrão seja adotado, também, por instituições brasileiras, a grande parte delas parece ainda não contar com tal estrutura.

Essa situação parece ser confirmada por Côte et al. (2002), que alertam para o fato de que bibliotecas partilham da insuficiência financeira e de pessoal, bem como de conflitos organizacionais para o desempenho satisfatório de suas funções, desafios muitas vezes decorrentes de aspectos socioculturais.

Diante do exposto, e do cenário de atuação profissional do bibliotecário diante de aspectos das novas tecnologias da informação e da comunicação, questiona-se:

- é possível proceder a conversão de registros bibliográficos para o formato MARC21 Bibliográfico por meio de uma mesma metodologia que seja aplicável a diferentes bases?
- é possível estabelecer-se um conteúdo sintático e semântico do registro bibliográfico que possa nortear o processo de conversão ao Formato MARC21 Bibliográfico?

- é possível aplicar-se um modelo teórico-conceitual de sintaxe e semântica de registros bibliográficos a uma ferramenta computacional que permita estabelecer a conversão para o Formato MARC21 Bibliográfico?

Nesse sentido, estabelecem-se como premissas os seguintes fatos:

- a) não há padronização única adotada para a definição da estrutura de bases de dados em sistemas gerenciadores de bibliotecas e outras unidades de informação;
- b) no domínio da catalogação os registros bibliográficos são elaborados a partir de convenções advindas de comunidades de prática;
- c) há esquemas tradicionais e internacionais de descrição e de visualização de registros bibliográficos, identificados nos padrões de estrutura de metadados descritivos e nos padrões de conteúdo.

Assim, a tese desta pesquisa é a de que os princípios sintáticos e semânticos de registros bibliográficos, definidos pelos esquemas de descrição e de visualização na catalogação, presentes nos padrões de estrutura de metadados descritivos e nos padrões de conteúdo, determinam o processo de conversão de registros bibliográficos para o Formato MARC21 Bibliográfico.

Apresenta-se como hipótese o fato de que a partir de princípios e esquemas sintáticos e semânticos, inerentes aos registros bibliográficos, é possível traçar-se sua interferência no processo de conversão para o Formato MARC21 Bibliográfico.

Em vista desse panorama, a proposição desta pesquisa é analisar e desenvolver um estudo teórico-conceitual sobre a sintaxe e a semântica de registros bibliográficos que venham a ser confirmados em um interpretador computacional de conversão de registros bibliográficos impressos para o Formato MARC21 Bibliográfico. O conceito de interpretador computacional nesta pesquisa é adotado para refletir o processo, mediado por meios computacionais, que interpreta um registro bibliográfico analógico por meio de uma estrutura definida pelas marcações sintáticas, identificadas pelos sinais de pontuação presentes nos registros bibliográficos, de tal forma que permite inferir o valor semântico do recurso

informacional representado de modo que seja garantida a fidedignidade da representação.<sup>2</sup>

Com esta proposta, estabelece-se, como objetivo geral, desenvolver um modelo teórico-conceitual de sintaxe e semântica em registros bibliográficos, a partir de estudos lingüísticos saussureanos e hjelmslevianos das manifestações da linguagem humana, que seja aplicável a um interpretador computacional voltado à conversão de registros bibliográficos ao formato MARC21 Bibliográfico.

Para o alcance de tal objetivo recorre-se aos seguintes objetivos específicos, reunidos em dois grupos e voltados, respectivamente, ao modelo teórico-conceitual da estrutura sintática e semântica de registros bibliográficos, e ao processo de conversão de registros bibliográficos:

- explicitar a relação entre a sintaxe e a semântica de registros bibliográficos à luz dos aspectos lingüísticos, estudados por Saussure e por Hjelmslev, aplicáveis aos meios analógicos e digitais;
- apresentar princípios característicos da sintaxe e da semântica em registros bibliográficos;
- examinar, por meio da literatura disponível, os métodos e os processos de conversão de registros bibliográficos ao formato MARC21 Bibliográfico;
- definir método de processamento de imagens de registros bibliográficos analógicos em *software* de reconhecimento ótico de caracteres (OCR);
- apresentar técnica de interpretação computacional que seja voltada à conversão de registros bibliográficos analógicos para o Formato

---

<sup>2</sup> A Ciência da Computação estuda algoritmos para lingüística computacional, no qual há destaque para o processo de análise sintática de textos, chamado de *parsing*, e para o *parser*, o analisador sintático. Parsers são adotados para o processamento de linguagem natural, mais especificamente para a análise de textos e definição de sua estrutura. (METSKEER, 2001). Nas palavras de Kakkonen (2007, p. 1), *parser* é um software que analisa automaticamente a descrição estrutural de um dado conjunto de caracteres no contexto de uma gramática específica. O resultado gerado por um *parser* é chamado de *parse* e descreve a estrutura de um fragmento de idioma analisado em particular. Não será adotado o termo *parsing* nesta pesquisa tendo como base a afirmação de Othero e Menuzzi (2005, p. 39) de que “A palavra *parsing* em si não remete ao processamento sintático mediado por computador (ou processamento sintático computacional). O termo vem da expressão latina *pars orationes* (partes do discurso) e tem suas raízes na tradição clássica.”

MARC21 Bibliográfico que considere os esquemas sintáticos e semânticos.

Tendo em vista a afirmação de Marconi e Lakatos (2001, p. 83) de que “não há ciência sem o emprego de métodos científicos” e que o “método é o conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo”, o percurso metodológico definido nesta pesquisa baseia-se nas considerações de Gonsalves (2011) e de Silva e Menezes (2005).

A pesquisa apresenta abordagem qualitativa por assumir uma relação dinâmica com o mundo real, por meio da interpretação e atribuição de significados aos fenômenos estudados. Segundo seus objetivos, a pesquisa possui caráter exploratório, uma vez que “se caracteriza pelo desenvolvimento e esclarecimento de ideias, com objetivo de oferecer uma visão panorâmica, uma primeira aproximação a um determinado fenômeno que é pouco explorado” e por oferecer, por meio da pesquisa bibliográfica, “dados elementares que dão suporte para a realização de estudos mais aprofundados sobre o tema,” (GONSALVES, 2011, p. 67).

Por considerar, ainda, os objetivos, a pesquisa encarrega-se do caráter descritivo, por “descrever as características de um objeto de estudo.” (op. cit., p. 68). O caráter experimental está presente por referir-se, também, a “um fenômeno que é reproduzido de forma controlada, submetendo os fatos à experimentação (verificação), buscando, a partir daí, evidenciar as relações entre os fatos e as teorias.” (op. cit., p. 69).

A opção por tal pesquisa decorre da experiência profissional dos últimos anos na qual foi possível observar que grande parcela das unidades de informação parece não usufruir das oportunidades advindas do compartilhamento de registros bibliográficos e, conseqüentemente, do estabelecimento de redes voltadas a este fim, por não fazerem uso do formato MARC21 Bibliográfico. Parece que isto seja decorrência dos mais variados motivos, dentre eles, destacam-se:

- o desconhecimento dos padrões ou das vantagens em adotá-los;
- as limitações oriundas de infra-estrutura tecnológica;
- as restrições orçamentárias para o tratamento dos dados, por requerer pessoal capacitado;

- a falta de habilidades de gestão para o gerenciamento de diferenciados recursos necessários ao processo de mudança ou implantação de sistemas automatizados;
- a inexperiência no processo de seleção de *software* de gerenciamento que atenda aos requisitos de compartilhamento de dados e uso do formato MARC21 no delineamento das bases de dados; e
- o desconhecimento dos processos necessários para o compartilhamento, a conversão e a migração de dados bibliográficos, apesar de se ter conhecimento da existência de padrões e das vantagens na adoção.

Nesse sentido, e diante da recorrente necessidade de compartilhamento de registros bibliográficos, esta pesquisa justifica-se por considerar a possibilidade de se efetuar a conversão desses dados para o formato MARC21 Bibliográfico. Processos de conversão de bases de dados bibliográficos, relatados na literatura, apresentam a experiência da conversão a partir da estrutura da base de dados, que se constitui, de maneira geral, no estudo da estrutura da base de dados para que, então, sejam realizados: [1] o reconhecimento das etiquetas de campo (TAGs) utilizadas, [2] a análise de equivalências entre elas e o formato MARC21 Bibliográfico, e [3] o rearranjo dos dados no padrão MARC21 Bibliográfico. Entretanto, note-se a ausência de padrões para nomear as etiquetas dos campos na referida estrutura o que implica a elaboração de um novo *script* a cada conversão de dados bibliográficos.

Para exemplificar como se dá esse processo observe-se a Figura 1.

Figura 1 – Nomes das etiquetas de campo do sistema 'X'

CLASS	AUTOR	TITUL	ANOAQ	AQUIS	LOCAL	EDITO
270.02	Pietrini,	A idade	2004	Compra	São Paulo	Paulus
221.93	Lenhardt,	A Torah	2004	Compra	São Paulo	Paulus
262.01	Kehl,	A Igreja :	2004	Compra	São Paulo	Paulus
225.83	Arens,	Ásia	2004	Compra	São Paulo	Loyola

ANOPB	PAGIN	CDDED	IDIOM	TIPO	BIBLI	ASSUN
1998	249 p.	21	Port.	Livro	Bibliograf	Curso de
1997	155 p.	21	Port.	Livro		Documen
1997	414 p.	21	Port.	Livro	Bibliograf	Eclesiolog
1998	209 p.	21	Port.	Livro		Sociologia

Fonte: Elaborado pelo próprio autor.

Na Figura 1 pode ser observado que os nomes das etiquetas de campo constantes da estrutura da base de dados 'X' são: CLASS (para o número de



Classificação), AUTOR (para o nome do(s) autor(es)), TITUL (para o título), ANO AQ (para o ano de aquisição), AQUIS (para a forma de aquisição), LOCAL (para o local de publicação), EDITO (para a editora), ANOPB (para o ano de publicação), PAGIN (para a descrição das páginas), CDDED (para a identificação da edição da Tabela de Classificação Decimal de Dewey), IDIOM (para o idioma do item), TIPO (para o tipo de material), BIBLI (para o tipo de bibliografia do curso) e ASSUN (para identificação temática da obra).

Veja-se a Figura 2 e analise-se os nomes de etiquetas de campo, comparativamente, em relação à Figura 1.

Figura 2 – Nomes das etiquetas de campo do sistema ‘Y’

Assunto	Autor(a)	Classificação	Not. Autor	Série	Editora	Título
Antropolo	Mancini, J	233	M237e	Coleção	Paulinas	Existência
Bíblia	Rolla, Arn	22.08	R655a	Crer e co	Paulinas	A Bíblia e a
Deus	Catão, Fra	231	C354f	Iniciação	Paulinas	Falar de D
Ética famil	Salomão,	241.91:173	S17r	Questõe	Vozes	Família
Êxodo	Pixley, Ge	222.12	P764e	Grande q	Santuári	Êxodo

Fonte: Elaborado pelo próprio autor.

A Figura 2 demonstra os seguintes nomes das etiquetas de campo constantes da estrutura da base de dados ‘Y’: Assunto, Autor(a), Classificação, Not. Autor, Série, Editora, Curso e Título. Pode-se observar que há variação significativa e nenhuma padronização para a nomeação das etiquetas de campos nas duas estruturas de base de dados indicadas, o que dificultaria o compartilhamento de registros bibliográficos. Na primeira, pode-se observar que as etiquetas de campo são identificadas por, no máximo, cinco letras maiúsculas, padrão de estrutura que não ocorre no segundo exemplo.

Explicite-se o processo que se apresenta, sob nosso ponto de vista, como um problema: em um processo de conversão de dados bibliográficos em uma determinada instituição é requerido o desenvolvimento de um *script* que faça esse processamento em lote. Tal *script* não servirá para outra instituição, caso a estrutura das bases em questão seja diferente em algum ponto. O *script*, então, torna-se descartável; seu modelo poderá ser (re)aproveitado, porém, o processo acima descrito se repetirá a cada vez que se requeira conversão dos dados de uma nova base. Sob esse ponto de vista, essa técnica é onerosa, pois a geração de *script* para conversão torna-se única, o que, portanto, não contribui para os processos de



conversão que tenham como ponto de partida o conteúdo do campo a ser convertido, conforme se propõe estudar aqui.

Recorrer a mecanismos computacionais nas atividades do dia-a-dia das bibliotecas, como substitutos das atividades manuais e mecânicas, já foram indicadas por Ranganathan e Gopinath (1967, p. 543), para quem esse processo exige viabilidade econômica e, em última instância, depende do estágio de desenvolvimento de um país.

No âmbito técnico-funcional-administrativo das bibliotecas, o emprego de tecnologias de informação e comunicação tornou-se constante, o que favorece tanto o surgimento quanto a inovação de várias atividades cooperativas. Segundo García Martínez (2005), o entendimento da cooperação bibliotecária deve ser antecedido pela preocupação com o cumprimento e com a oferta de serviços de qualidade aos seus usuários e, nesse sentido, é necessário associar-se a outras unidades de informação para que tenham melhores condições de atender às necessidades informacionais de seus usuários.

Segundo Cunha (1979) e Ferreira et al. (1979), a catalogação cooperativa no Brasil iniciou-se em 1942, com Serviço de Intercâmbio de Catalogação (SIC), inicialmente coordenado pelo Departamento Administrativo do Serviço Público (DASP), em convênio com o Departamento de Imprensa Nacional (DIN), e, com destaque para um período grandioso, entre os anos de 1947 e 1952, com esforços agregados da Fundação Getúlio Vargas. A partir de 1954, o SIC, coordenado agora pelo Instituto Brasileiro de Bibliografia e Documentação (IBBD)<sup>3</sup>, foi fundamental para o estabelecimento de projetos e serviços cooperativos nacionais, apesar de serem observados os princípios de catalogação cooperativa desde a década de 40.

Nos últimos anos, diversas unidades de informação têm procurado adequar-se aos padrões internacionais tanto de descrição quanto de compartilhamento de dados bibliográficos. No Brasil, parece-nos que as bibliotecas universitárias são as que mais se utilizam do Formato MARC21 Bibliográfico, provavelmente como reflexo das ações do Plano Nacional de Bibliotecas Universitárias (PNBU); mesmo assim, nem todas dele se servem. Considera-se que diversas unidades de informação, tanto bibliográficas quanto documentais, poderiam beneficiar-se amplamente do

---

<sup>3</sup> Atualmente denominado Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

compartilhamento de registros de registros bibliográficos, pelo menos ao nível nacional. Porém, sem participar de redes de unidades de informação, o compartilhamento de dados bibliográficos em meios automatizados torna-se prejudicado.

Assim sendo, julga-se que a proposta aqui apresentada possa ter contribuição significativa, considerando aspectos sociais, inerentes ao conhecimento humano e à solução do problema apresentado. O maior benefício social ensejado por essa pesquisa relaciona-se, intimamente, aos aspectos do compartilhamento de registros bibliográficos, bem como ao amplo acesso aos recursos informacionais representados pelas diversas unidades de informação, ocasionando, assim, maiores oportunidades de oferta de serviços aos usuários.

Tal estudo, a nosso ver, deflagra contribuições tanto no plano teórico quanto no prático.

No campo teórico, por vislumbrar o desenvolvimento de questões sobre aspectos sintáticos e semânticos de registros bibliográficos, e por envolver, a um só tempo, a interdisciplinaridade entre a Ciência da Informação, a Ciência da Computação e a Lingüística, de modo a reafirmar a proposta de Borko (1968) e de Saracevic (1996).

Para Borko (1968, p. 3) a Ciência da Informação é uma ciência interdisciplinar derivada e relacionada com áreas como a matemática, a lógica, a lingüística, a psicologia, a informática, a pesquisa de operações, as artes gráficas, a comunicação, a biblioteconomia, a administração e outras áreas afins. A Ciência da Informação apresenta, nesse sentido, tanto um componente de ciência pura, por investigar o assunto sem considerar a sua aplicação, quanto um componente de ciência aplicada, que desenvolve serviços e produtos.

Para Saracevic (1996) a Ciência da Informação apresenta-se como interdisciplinar pelo fato de [1] sua natureza estar “inexoravelmente ligada à tecnologia da informação” (op. cit., p. 42), [2] ser “participante ativa e deliberada na evolução da sociedade da informação” (loc. cit.), e [3] dada a sua “dimensão social e humana” (loc. cit.). O autor apresenta, ainda, características de inter-relação da Ciência da Informação com a Biblioteconomia, com a Ciência da Computação e com a Comunicação. Do ponto de vista deste trabalho, dois aspectos apresentados pelo autor tornam-se relevantes: o fato de que “[...] a ciência da computação trata de

algoritmos que transformam informações enquanto a CI trata da natureza mesma da informação e sua comunicação para uso pelos humanos” (op. cit., p. 50), o que se entende seja condição intrínseca à proposta aqui apresentada; e que a Comunicação trata da comunicabilidade humana, uma vez que “[...] as relações entre o fenômeno e o processo – informação e comunicação – definem as relações entre CI e comunicação” (op. cit., p. 53).

A contribuição no campo prático é observada pelo fato de a pesquisa abranger o desenvolvimento de interpretador computacional que possa ser adotado por quaisquer instituições que queiram utilizar-se de procedimento de conversão de bases de dados de registros bibliográficos, para o formato MARC21 Bibliográfico a partir dos esquemas de descrição (AACR2) e de visualização de registros bibliográficos (ISBD), caráter que se considera inovador na pesquisa.

De acordo com o que se propõe nesta pesquisa, após este primeiro capítulo, de cunho introdutório e de contextualização na Ciência da Informação, indica-se o referencial teórico, desenvolvido no capítulo segundo, intitulado *Representação e catálogos no processo comunicativo de bibliotecas*, no qual se discute a representação de recursos informacionais; os padrões para a representação bibliográfica; os catálogos e os objetivos bibliográficos; a cooperação, o compartilhamento e o intercâmbio de registros bibliográficos; e os métodos e processos de conversão de registros bibliográficos.

A contribuição teórica da pesquisa está delineada no terceiro capítulo, intitulado *Um olhar da linguística sobre os registros bibliográficos*, no qual se recorre aos estudos linguísticos da comunicação humana, com Saussure, e da concepção estruturalista, com Hjelmslev, para a construção do arcabouço teórico da sintaxe e da semântica dos registros bibliográficos. Neste capítulo propõe-se a análise da contribuição saussureana à representação dos recursos informacionais; a compreensão da semântica hjelmsleviana e os registros bibliográficos; e a apresentação dos princípios sintáticos e semânticos de registros bibliográficos, a partir da representação como estrutura semântica.

No capítulo quatro, intitulado *Scan for MARC: interpretador de princípios sintáticos e semânticos de registros bibliográficos*, recebe destaque a abordagem metodológica e de aplicação do constructo teórico apresentado, no qual se explicita a captura de imagens de registros bibliográficos; a análise de *softwares* de

reconhecimento de caracteres; o processamento das imagens de registros bibliográficos; e o tratamento sintático e semântico de registros bibliográficos para o Formato MARC21 Bibliográfico.

O último capítulo dedica-se às *considerações finais* da pesquisa.

## 2 REPRESENTAÇÃO E CATÁLOGOS NO PROCESSO COMUNICATIVO DE BIBLIOTECAS

Instituições de patrimônio cultural, dentre estas as bibliotecas, representam o papel de agentes sociais de comunicação e, por conseguinte, intermediam recursos informacionais para uso pela humanidade. Para tanto, é necessário conhecer os usuários, os documentos a serem representados, e os processos de representação. Este capítulo, destinado a dar a conhecer esta inter-relação, apresenta-se dividido em cinco momentos: o primeiro, volta-se à representação dos recursos informacionais; o segundo, dedica-se ao estudo dos padrões, regras e esquemas para a representação bibliográfica; o terceiro, traz os catálogos e os objetivos bibliográficos; o quarto, destina-se à cooperação e ao compartilhamento de registros bibliográficos com a conseqüente formação de redes para intercâmbio de dados bibliográficos; por fim, no quinto, apresentam-se os processos e as metodologias de conversão de registros bibliográficos.

### 2.1 REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS

Representar recursos informacionais é atividade fundamental para o estabelecimento do processo comunicativo entre unidades de informação e usuários e freqüentadores de instituições de patrimônio cultural e, para que estas atuem como agentes sociais de comunicação, deve pautar suas atividades na conveniência do usuário. Para tanto, requer conhecimento do público a que se destina, das especificidades do gênero e do formato dos documentos de que dispõe, e dos padrões a serem utilizados na representação. Segundo Castro e Santos (2005, p. 68), pela representação de itens documentários é possível “simplificar a busca e a recuperação dos mesmos.”

Nesse sentido, o constructo da representação dos recursos informacionais em todos os seus aspectos, tanto descritivos quanto de conteúdo, em outras palavras, a representação documental, visa a duas ações básicas em bibliotecas: comunicar aos usuários os documentos de que dispõem, de modo individualizado, por meio da identificação das diferenças inerentes a cada documento, bem como por suas

manifestações e itens; e reunir os documentos, por meio de suas características de conteúdo, ou de suas expressões e obras. Dessa maneira, é possível que a representação bibliográfica dos recursos informacionais promova o estabelecimento do processo de comunicação. Em suma: a representação não deve se propor a tomar o lugar do representado, mas buscar a forma de tornar o representado reconhecível diante de dada situação, contexto e público (MEY; ZAFALON, 2009). O ato representacional dos recursos informacionais recorre a aspectos descritivos e de conteúdo do objeto informacional e se efetiva por relações entre o registro documental e o seu representado, bem como entre os próprios recursos bibliográficos. Isso decorre do fato de a representação subsidiar a colocação de 'algo' no 'lugar de', para que seja possível buscar formas de tornar o representado reconhecível diante de dada situação, contexto e público. Então, o que seria esse processo de representação?

Segundo Mey (1999), representação bibliográfica pode ser definida como “o conjunto de informações padronizadas que representam um registro do conhecimento, permitindo sua seleção, identificação, localização e recuperação.” Outra conceituação apresentada pela autora conceitua representação bibliográfica como:

estudo, preparação e organização de mensagens codificadas, com base em itens existentes ou passíveis de inclusão em um ou vários acervos, de forma a permitir intersecção entre as mensagens contidas nos itens e as mensagens dos usuários. (MEY, 1995, p. 5).

Tais conceituações parecem-nos recorrer àquela apresentada no Glosario ALA de Bibliotecología y Ciencias de la Información, editado por Young (1988, p. 56), no qual catalogação, como também é identificada a representação bibliográfica, apresenta-se como sendo as atividades desenvolvidas na preparação dos registros bibliográficos para sua incorporação a um catálogo. Garrido Arilla (1999), contudo, considera catalogação como o processo de descrição de elementos que permitem identificar e estabelecer formas de recuperação de um recurso informacional. No entanto, a autora (op. cit.) apresenta outra conceituação, mais elucidativa, ao indicar que a descrição bibliográfica é a ação na qual se expõem elementos de unidades informativas que permitem reconhecer e particularizar formalmente um documento em referências específicas que aludem os dados externos do documento e se distinguem dos restantes. Porém, outro enfoque sobre a descrição bibliográfica é

apontado por International Federation of Library Associations and Institutions (2005) ao afirmar que se trata de um conjunto de dados bibliográficos compostos por entidades que descrevem ou identificam uma ou mais obras, expressões, manifestações ou itens.

Entende-se, ainda com Garrido Arilla (1999), que alguns estágios são necessários à catalogação: análise do documento; identificação do tipo de documento e eleição dos padrões adequados à sua descrição; consolidação do nível de descrição; identificação dos elementos essenciais ao nível de descrição; registro dos resultados da análise no suporte escolhido, manual ou automatizado; certificação da exatidão dos dados extraídos.

O processo de representação documental, também identificado como catalogação, resulta em um registro que compreende três conjuntos de elementos: a descrição bibliográfica, os pontos de acesso e os dados de localização para uma ou mais obras, expressões, manifestações ou itens. (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2005; MEY, 1995). Tais conjuntos são responsáveis, respectivamente, pela individualização de diferentes manifestações de uma obra, pela reunião de suas semelhanças, e pela obtenção do documento.

No mesmo sentido, Santos e Correia (2009), apresentam como funcionalidades de tais conjuntos em ambientes informacionais: a unicidade, por meio da descrição bibliográfica das características do recurso bibliográfico; o instanciamento, pelos pontos de acesso definidos pelos termos que representam conteúdo e forma do documento; e o armazenamento de um item, por meio dos dados de localização. Garrido Arilla (1999), por sua vez, afirma que o registro bibliográfico é uma representação codificada e econômica de um documento: representação pois reproduz certas características do documento; codificada porque identifica o documento a que se propôs identificar; econômica porque busca proporcionar informação de em um mínimo espaço e com o mínimo de símbolos possíveis.

Nesse processo de representação em face da comunicação, não se pode perder de vista que a adoção de diretrizes e códigos fornece o suporte necessário para a elaboração do registro representativo do documento. Essa representação é muito mais complexa do que pode aparentar. Embora o ato representacional se

realize por uma pessoa, que pretende certa imparcialidade, sabe-se de antemão tratar-se de utopia, ou mesmo falácia, uma vez que permanece envolto pela própria história, cultura e pelo mundo em que se insere, assim como os próprios códigos. Portanto, o olhar do catalogador não se deve sobrepor aos daqueles para quem se fazem as representações de fato: os usuários<sup>4</sup> (ZAFALON; MEY, 2009). Essa visão é partilhada com Coleman (2005), para quem a criação de metadados é, muitas vezes, subjetiva e não objetiva, afinal, sem regras e normas claras, dois criadores de metadados podem descrever o mesmo recurso de forma bastante diferente. O termo *subjetivo* é usado para indicar que um determinado elemento pode ser descrito com base no julgamento pessoal do catalogador do recurso, e *objetivo* é utilizado para designar a objetividade, quando os dados ou o conteúdo encontrado no recurso podem ser identificados de forma semelhante pelos criadores de metadados e catalogadores. Adotar padrões, portanto, torna-se fundamental em instituições de patrimônio cultural uma vez que isso minimiza eventuais diferenças na representação de documentos. Frège (1978, p. 65) apresenta a relação entre representação e sentido ao afirmar que

Quando dois homens representam a mesma coisa, ainda assim cada um tem sua própria representação. De fato, às vezes é possível estabelecer diferenças entre as representações, ou mesmo as sensações de homens diferentes; mas uma comparação precisa não é possível, porque não podemos ter ambas as representações juntas na mesma consciência.

Para a operacionalização dessa representação utilizam-se normas, padrões e recursos, que se tornaram internacionalizados, e são fundamentais para a qualidade dos registros bibliográficos e a possibilidade de compartilhamento. Esse processo de representação, sistematizado desde meados do século XVI, apresenta formas ditadas pelos códigos de catalogação, agregadores de regras, interpretações, exemplos indicativos, interpretativos e explicativos sobre as possibilidades de representação da informação e visa, por um lado, a individualização de recursos informacionais, pelos seus diferenciados aspectos físicos, ou de suas

---

<sup>4</sup> Neste caso, a palavra *usuários* assume múltiplos olhares: o olhar do leitor, ou do destinatário da obra do autor; o olhar do autor, ou da identidade bibliográfica, ou da obra em si, que espera ser encontrada; e o olhar de todas as outras instituições documentais no mundo, porque nenhum centro de documentação exerce suas funções de maneira isolada.



manifestações, ou de seus itens; e, por outro, a reunião por meio de suas características similares, de conteúdo, ou de suas expressões ou de suas obras.

No processo de elaboração da representação bibliográfica aspectos como a integridade, a clareza, a precisão, a lógica e a consistência dos conteúdos dos documentos a serem representados, devem ser resguardados, tendo em vista, constantemente, o usuário e as necessidades por ele demonstradas, requeridas ou determinadas (MEY, 1995).

## 2.2 PADRÕES PARA A REPRESENTAÇÃO BIBLIOGRÁFICA

A produção e a proliferação de informação em ambientes digitais exigem a adoção de padrões para representação da informação de modo que, assim, tanto recuperação quanto acesso estejam garantidos pela identificação unívoca do recurso. A aceitação e a adoção de padrões apresentam importância vital para os requisitos de sistemas cooperativos e centralizados e, por tal motivo, Hunter e Bakewell (1983, p. 22) afirmam que os padrões devem se relacionar aos esquemas de conteúdo e de arranjo da entrada de catálogo.

A relação intrínseca entre representação e recuperação de documentos exige que se leve em conta as ferramentas de descrição e, também, de estrutura de registros bibliográficos, o que promove consistência, precisão e relevância aos resultados obtidos em resposta a uma consulta.

Padrões são criados para serem usados. Entretanto, Foulonneau e Riley (2008) ressaltam que é desnecessário que instituições criem os seus próprios padrões de metadados. Ocasionalmente, o desenvolvimento local é uma opção razoável, mas isso requer investimento de tempo, além do que, aumenta a possibilidade de ocorrerem riscos que não são possíveis de serem encontrados quando se usa padrões que já estão em vigor.

Compartilhamento de registros de metadados com outras instituições é um fator a se considerar na decisão sobre qual padrão usar, e se vai ou não desenvolver um novo esquema. Foulonneau e Riley (2008) ressaltam que a prática de sua criação é um tanto difícil, com alto custo e, por vezes, as instituições de patrimônio cultural desconhecem os benefícios de sua descoberta e da administração de dados por meio da criação de metadados, o que requer conhecer

sistemas de implementação e o ambiente para saber como os metadados (as formas de representação bibliográfica) podem ajudar a promover serviços.

Alguns fatores são decisivos para a escolha dos padrões para a representação bibliográfica, segundo Foulonneau e Riley (2008):

- a) *a instituição*: diferentes tipos de instituições de patrimônio cultural, apesar de terem objetivos similares, apresentam práticas diversificadas para a descrição de recursos: as práticas descritivas de bibliotecas são mais padronizadas, apesar de, ainda, terem diferenciações entre si, por compartilharem códigos de catalogação e formato do registro, além de terem longa tradição de catalogação cooperativa para reduzir duplicação de esforços; os arquivos tem prática de descrição menos formal e pouca tradição de compartilhamento de informações, prevalecendo a prática da descrição multi-nível; os museus apresentam um grau ainda maior de falta de orientações normativas para a criação de dados representativos de seu acervo, tem pouca experiência quanto ao compartilhamento de registros com seus pares de instituições, o processo de catalogação tende a ser distribuído e envolve, tipicamente, contribuição de curadores e profissionais da informação com várias capacitações;
- b) *o padrão*: aspectos quanto ao propósito, à estrutura, ao contexto e ao histórico do padrão devem ser considerados na decisão e adequação do padrão a um projeto, pois são questões decisivas para a interoperabilidade, dada a tecnologia usada. Em comunidades de patrimônio cultural, muitos formatos de metadados são definidos com esquema XML DTD ou W3C XML e de acordo com o que se espera do padrão de metadados. Outros tipos de padrões, como os vocabulários controlados, tendem a ser expressos com protocolos muito especializados. Desenvolver ou manter padrões por um corpo respeitável na comunidade de patrimônio cultural (como a Library of Congress) sugere certo grau de autoridade e sustentabilidade. Aqueles desenvolvidos por membros de uma associação profissional representativa de uma comunidade especializada (como a Visual Resources Association), ou aqueles que surgem de uma comunidade

de prática (como a Digital Library Federation), frequentemente transmitem o conhecimento especializado dos recursos ou do ambiente técnico a que o padrão se destina. Fatores como a atualização do padrão (raramente podem até significar que o padrão sofre o risco de ser abandonado e muito frequentemente pode-se ter problemas de manutenção pelos criadores dos metadados) e a quantidade de implementadores que um determinado padrão tem também devem ser considerados;

- c) *os documentos a serem representados*: muitos documentos apresentam especialidades de gênero e de formato, o que requer atenção na escolha do padrão. É extremamente difícil criar metadados que sirvam simultaneamente a usuários especializados em um domínio específico de conhecimento e a usuários genéricos e casuais. É necessário pensar sobre a audiência primária do material e, conseqüentemente, sobre a escolha de padrões de metadados que suportem a terminologia e a recuperação necessária ao atendimento dos usuários. Para usuários genéricos, existem diferenças entre os formatos: material de arquivo requer aproximações diferentes de registros de dados científicos; um texto tem características importantes para a pesquisa como, por exemplo, a linguagem; a fotografia histórica tem na localização ou nas pessoas informações de interesse primário; materiais publicados são diferentes de materiais não publicados; em registros sonoros, detalhes técnicos, como velocidade de execução, são essenciais. Em suma: é necessário que se conheçam os elementos essenciais para que depois seja escolhido o padrão;
- d) *os objetivos do projeto*: fatores como a repetibilidade dos elementos e a necessidade de delinear múltiplas versões de um recurso; o depósito de metadados em ambientes agregados, incluindo motores de busca na Web em geral; os protocolos de pesquisa e de coleta de metadados podem ser considerados no momento da escolha de um padrão de metadados. Escolher padrão de metadados exige compromisso. Quando se selecionam padrões para um projeto, é importante balancear a necessidade da generalidade com a especificidade.

Selecionar um padrão permite que este seja granular, completo e consistente para que possa justificar os serviços aos seus usuários.

Os padrões de metadados podem ser categorizados de diferentes formas, segundo Foulonneau e Riley (2008). Entretanto, nesta pesquisa, atentar-se-á aos *padrões de estrutura de metadados descritivos* e aos *padrões de conteúdo*.

*Padrões de estrutura de metadados descritivos* são padrões que listam elementos considerados importantes para a descrição do recurso, incluindo características físicas e de conteúdo. Os padrões tipicamente especificados com elementos recorrem à repetibilidade e à ordem em que devem aparecer. Os padrões de estrutura de metadados descritivos mais gerais, indicados por Foulonneau e Riley (2008), são o MARC Bibliográfico, o MARCXML, o MODS e o Dublin Core, e os mais especializados são o VRA Core (para recursos visuais), o CDWA lite (para arte e arquitetura), o GEM (para objetos de aprendizagem), o IMS Learning Resource Metadata (para objetos de aprendizagem), o ETD-MS (para dissertações e teses eletrônicas) e o DDI (para conjuntos de dados de ciências sociais e comportamentais).

Os *padrões de conteúdo*, por sua vez, fornecem regras para a sintaxe de uma entrada em um campo de metadado. Seu propósito é promover consistência em registros metadados para permitir uma melhor pesquisa e recuperação pelos usuários. Esquemas de metadados sem regras de conteúdo não são muito utilizados. Padrões de conteúdo são comumente usados em padrões de estrutura de metadados descritivos, mas também são aplicáveis a outros tipos, tais como padrões usados em datas, que podem ser usados como metadados de registros técnicos. Em bibliotecas é familiar o padrão de conteúdo e sua relação com o padrão de estrutura de metadados, como as AACR2 e sua relação com o MARC. Os padrões de conteúdo usados em larga escala são o AACR2, o CCO (Cataloging cultural objects) e o DA:CS (Describing Archives: content standards), e os utilizados em pequena escala são o Archival Moving Image Materials: a cataloging manual (suplemento do AACR2 para catalogação de materiais de imagem em movimento), o W3CDTF (para codificação de sintaxe para datas), o DOI (para codificação de sintaxe para identificadores), o DCMI Type (para esquema de codificação de vocabulário para tipos de recursos) e o AAT (para esquema de codificação para assuntos de arte e arquitetura). Danskin (2009) apresenta uma divisão dos padrões

de conteúdo em padrões descritivos e padrões de assunto e afirma que tais são adotados para indicar qual informação deve ser registrada e como deve ser apresentada em um registro bibliográfico. Danskin (2009) afirma que por meio dos padrões de conteúdo é possível garantir a consistência nas bibliografias publicadas. Dentre os padrões descritivos, Danskin (2009) identifica a ISBD, reconhecida pelo fato de oferecer consistência para compartilhamento de informações bibliográficas. A ISBD determina os elementos de dados que devem ser transcritos em uma sequência específica como a base da descrição de um recurso que está sendo catalogado e ressalta que é o padrão no qual códigos de catalogação estão baseados.

Santos e Corrêa (2009) ressaltam “a necessidade do estabelecimento de esquemas e regras de descrição no que se refere à forma e ao conteúdo das manifestações informacionais” por conta da (re)transmissão de dados bibliográficos.

Nesse sentido, e com tal finalidade, encaixam-se os códigos de catalogação, que podem ser definidos como um conjunto de regras, interpretações e, por vezes, exemplos indicativos e explicativos, destinados à elaboração de registros bibliográficos (ZAFALON; MEY, 2009). A partir da revisão dos Princípios de Paris (1961), publicada em 2009 como *Declaração dos Princípios Internacionais de Catalogação*, os códigos de catalogação devem ter como diretrizes: [1] a conveniência dos usuários, [2] o uso comum, [3] a representação, [4] a precisão, [5] a suficiência e a necessidade, [6] a significância, [7] a economia, [8] a consistência e padronização, e [9] a integração. Vale mencionar que o usuário deve ser o foco primordial da atividade de catalogação (ZAFALON; MEY, 2009; INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2009).

Para o constructo desta pesquisa, foram estudadas as AACR2r, como *padrão de conteúdo*, e a ISBD e o Formato MARC21 Bibliográfico, como *padrões de estrutura de metadados descritivos*.

As AACR2r, revisão de 2002 das *Anglo-American Cataloging Rules, 2nd edition*, publicadas no Brasil em 2004 com o título Código de Catalogação Anglo-Americano (CCAA), apresentam, por meio de orientações, regras e exemplos, a descrição do conteúdo (apresentada nas seções de parágrafos de um registro em cartão, inclui o título, indicação de responsabilidade, edição, detalhes específicos do material, informação de publicação, descrição física, série, notas e números

normalizados) e, também, a escolha, a elaboração e a atribuição de pontos de acesso a um documento (representando, por sua vez, nomes de pessoas, de lugares geográficos, instituições e eventos), o que permite que sejam criados os direcionamentos para a construção de catálogos bibliográficos. Pelo fato de as AACR2r apresentarem, inclusive historicamente, relação direta com as ISBDs, pode-se dizer, de certo modo, que, para os catálogos manuais, elas assumem tanto aspectos do *padrão de estrutura de metadados descritivos* quanto de *padrões de conteúdo*. Observe-se, somente para reflexão, que a própria forma de compreensão das ISBDs é voltada aos aspectos estruturais de um registro bibliográfico:

[...] as áreas e os elementos são indicados pela pontuação e pela posição. Esta determinação se deve à crença de que os sistemas automatizados reconheceriam as áreas e os elementos graças à pontuação que os precede e à posição que ocupam no registro bibliográfico. A crença revelou-se demasiado otimista, mas a pontuação clarifica o entendimento do registro e a posição é a mesma das práticas usuais. (MEY, 1995, p. 44).

Compreende-se a ISBD como *padrão de estrutura de metadados descritivos* a partir de Swanson (1973) e Langker (1974) para quem a ISBD especifica os elementos de uma descrição bibliográfica, prescreve a ordem em que devem ser apresentados, mas, principalmente, porque indica os sinais de pontuação<sup>5</sup> pelo qual os elementos devem ser demarcados. Desse modo, a ISBD visa três objetivos: fazer registros de diferentes fontes intercambiáveis, facilitar a sua interpretação acima das barreiras lingüísticas e facilitar a conversão de tais registros à forma legível por máquina. Langker (1974) ressalta que a pontuação é usada para fins estruturais para delimitar os campos e subcampos (de modo a auxiliar um operador de máquina a registrar em formato legível por máquina). Spaulding (1973), por sua vez, explicita o esforço de Henriette Avram para que, recorrendo-se à codificação ou marcação de registros bibliográficos, houvesse redução de esforço intelectual com a conversão em formato legível por máquina. Para Avram, ao recorrer a um padrão de pontuação entre os campos e subcampos para a descrição de um registro bibliográfico, é possível o processo de codificação destes elementos de forma mecanizada e liberto de quaisquer considerações de idioma. Essa abordagem é recusada por Swanson

---

<sup>5</sup> Considerar a ISBD como um padrão de estrutura de metadados descritivos por conta da pontuação assume caráter fundamental para o desenvolvimento desta pesquisa, uma vez que, segundo Trask (2008, p. 232), a pontuação é “um sistema convencional de marcas que representam informações sobre a estrutura de um texto escrito.”

(1973) para quem a crença de que a ISBD permite simplificar a conversão dos dados analógicos para o formato legível por máquina é tecnicamente irrealista, afinal seria necessário recorrer a um dispositivo OCR (reconhecimento ótico de caracteres), desconsiderado pela equipe de Avram ao indicar que nenhum dispositivo de escaneamento ótico seria adequado para escanear catálogos em fichas. Entretanto, entende-se que a pontuação prescrita nas ISBDs cumpre a dupla finalidade de proporcionar meios para precisar elementos bibliográficos, independentemente da linguagem, tanto para seres humanos quanto para máquinas.

Para a descrição e a recuperação de registros bibliográficos em meio automatizado, é requerida, entretanto, aliada às AACR2r, a adoção de um *padrão de estrutura de metadados descritivos*, e, para esta pesquisa, interessa estudar, também, o Formato MARC21 Bibliográfico.

O Formato MARC21 Bibliográfico abrange aspectos de leitura e de interpretação de dados disponíveis em registros bibliográficos, por meio computacional.

A informação de um catálogo bibliográfico não pode ser simplesmente digitada em um computador para produção de um catálogo automatizado. O computador necessita de um meio para interpretar a informação contida em um registro bibliográfico. O registro MARC contém um guia para a interpretação destes dados antes de cada peça da informação bibliográfica. (ZAFALON, 2008, p. 14).

Aspectos inerentes à estrutura de um registro MARC podem ser notados pela flexibilidade da estrutura do arquivo e pelo número e tamanho ilimitados dos campos.

O computador não pode contar com certo tipo de informação para começar e terminar cada uma das posições em todo registro bibliográfico. A indicação de responsabilidade não poderá começar sempre no 145º caractere de um registro e terminar na 207ª posição, por exemplo. (ZAFALON, 2008, p. 15).

O processo de leitura e interpretação computacional de um registro bibliográfico em Formato MARC é facilitado pelas marcações inerentes a ele.

Se um registro bibliográfico foi marcado corretamente e salvo em um arquivo de computador, programas de computador podem ser utilizados para pontuar e formatar a informação corretamente para impressão em uma ficha catalográfica, ou para mostrar a informação na tela do computador. Programas podem ser usados para pesquisar e para recuperar certos tipos de informação em campos específicos,



e também para mostrar listas de itens encontrados em um critério de busca. (ZAFALON, 2008, p. 15).

Observe-se, que as convenções adotadas nas marcações de um registro bibliográfico, como *padrão de estrutura de metadados descritivos*, quer seja pela ISBD, quer seja pelo Formato MARC21 Bibliográfico, tornam o registro, o gerenciamento e a recuperação dos dados bibliográficos mais eficientes. Note-se como se dá esse processo, conforme ilustra a Figura 3 e o Quadro 1, a seguir.

A Figura 3 apresenta a representação bibliográfica de um documento a partir das AACR2r, para a descrição de conteúdo, e da ISBD, como padrão de estrutura de metadados descritivos.

Figura 3 – Representação de um documento segundo o *padrão de conteúdo* definido pelas AACR2r e o *padrão de estrutura de metadados descritivos* indicados pela ISBD

925 T49c	<p>Tiner, John Hudson</p> <p>100 cientistas que mudaram a história do mundo / John Hudson Tiner ; tradução de Marise Chinetti de Barros. - Rio de Janeiro : Ediouro, c2002.</p> <p>259 p. - (Coleção 100 mais)</p> <p>Tradução de: 100 scientists who shaped world history.</p> <p>ISBN 85-00-01493-9</p> <p>1. CIENTISTAS - BIOGRAFIA 2. CIÊNCIA - HISTÓRIA I. Título. II. Título: Cem cientistas que mudaram a história do mundo. III. Série.</p>
-------------	---

Fonte: Elaborado pelo próprio autor.

Nesta forma de representação é possível identificar, no nível do *registro*, um grupo de informações voltadas aos dados de localização, de descrição do item, propriamente dito, e das formas pelo qual o item poderá ser recuperado pelos pontos de acesso definidos. O *gerenciamento* pode ser encontrado, sumariamente, no controle dos dados de localização, para fins de identificação única do item, na equivalência entre o número de classificação decimal e da notação de autor; e, na forma autorizada dos pontos de acesso ao registro. A *recuperação*, por sua vez, pode ser visualizada nos pontos de acesso estabelecidos como principal (de autoridade) e secundários (assunto, título e série).

O quadro 1 apresenta, de modo esquematizado, a representação de um documento segundo o MARC21 Bibliográfico, como *padrão de estrutura de metadados descritivos*, e as AACR2r, como *padrão de conteúdo*.



Quadro 1 – Representação de um documento segundo o *padrão de estrutura de metadados descritivos* MARC21 Bibliográfico e o *padrão de conteúdo* AACR2r<sup>6</sup>

<b>Padrão de estrutura de metadados descritivos</b>				<b>Padrão de conteúdo</b>
<b>Formato MARC 21 Bibliográfico</b>				<b>AACR2r</b>
<b>TAG</b>	<b>Ind. 1</b>	<b>Ind. 2</b>	<b>Subcampo</b>	<b>Conteúdo do registro bibliográfico</b>
020	∅	∅	\$a	85-00-01493-9
090	∅	∅	\$a	925
			\$b	T49c
100	1	∅	\$a	Tiner, John Hudson
245	1	0	\$a	100 cientistas que mudaram a história do mundo
			\$c	John Hudson Tiner ; tradução de Marise Chinetti de Barros
260	∅	∅	\$a	Rio de Janeiro
			\$b	Ediouro
			\$c	c2002
300	∅	∅	\$a	259 p.
490	0	∅	\$a	Coleção 100 mais
500	∅	∅	\$a	Tradução de: 100 scientists who shaped world history
650	1	4	\$a	CIENTISTAS
			\$b	BIOGRAFIA
650	2	4	\$a	CIÊNCIA
			\$x	HISTÓRIA
740	0	∅	\$a	Cem cientistas que mudaram a história do mundo

Fonte: Elaborado pelo próprio autor.

Um registro descrito segundo o Formato MARC21 Bibliográfico recebe as marcações (presentes nas quatro colunas à esquerda) para que seja possível ao computador interpretá-lo. Nesta outra forma de representação, interpretável por meios computacionais, também é possível identificar, no nível do *registro*, informações voltadas aos dados de localização, de descrição do item, propriamente dito, e das formas pelo qual o item poderá ser recuperado pelos pontos de acesso definidos. O *gerenciamento* pode ser encontrado, sumariamente, no controle dos dados de localização, para fins de identificação única do item, na equivalência entre o número de classificação decimal e da notação de autor; e, na forma autorizada dos pontos de acesso ao registro; e na correlação entre o conteúdo e as marcações do Formato MARC21 Bibliográfico. A *recuperação* apresenta uma inovação, em se tratando de interpretação pela máquina, por permitir que sirvam como forma de recuperação não somente os pontos de acesso estabelecidos como principal (de autoridade) e secundários (assunto, título e série), mas também, outros campos que

<sup>6</sup> O símbolo ∅ foi adotado para indicar que, no Formato MARC, é necessário deixar um caractere em branco.

podem ser definidos como indexáveis, tais como local de publicação, editora, data, etc. Dessa maneira, as formas de recuperação podem ser ampliadas, sendo possível atender uma pesquisa com alto grau de especificidade. Exemplifique-se: o usuário de uma instituição bibliográfica, que dispõe de um registro conforme a descrição acima, deseja satisfazer o seguinte requisito de pesquisa: *qual o item documental que trata de 'História da ciência' como assunto secundário? Ou, ainda: quais obras traduzidas essa unidade dispõe?* Estas interpretações são possíveis de serem obtidas em um registro MARC.

Entende-se, por fim, que as convenções adotadas nas marcações de um registro bibliográfico, quer seja pela ISBD, quer seja pelo Formato MARC21 Bibliográfico, promovem, cada uma a seu modo, o desenvolvimento de catálogos e o atendimento aos objetivos bibliográficos, a serem discutidos no próximo tópico.

### 2.3 CATÁLOGOS E OBJETIVOS BIBLIOGRÁFICOS

As bibliotecas utilizam-se de catálogos, instrumentos disponíveis em diversas formas ou suportes (livros, folhas soltas, fichas ou eletrônicos) com o objetivo de garantir a recuperação dos registros bibliográficos pelo público a que se destina.

Garrido Arilla (1999) afirma que o principal objetivo do catálogo é facilitar a identificação e a localização de um documento do qual se conhece seu autor, título ou assunto. Nesse sentido, pode ser definido como um conjunto de registros bibliográficos de documentos de um acervo, organizados a partir de determinadas normas ou critérios e reconhecido como instrumento relevante para a recuperação de informações. Segundo Mey (1986, p. 3), “[...], os catálogos são listas organizadas de acordo com uma ordem qualquer [previamente estabelecida], permitindo a quem os lê ter idéia do objeto a que se referem, sem acesso direto ao objeto em si.” Entretanto, essa abordagem pode ser mais bem elucidada:

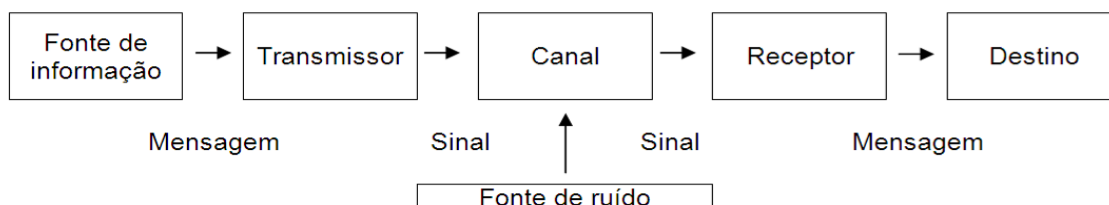
[...] um canal de comunicação estruturado, que veicula mensagens contidas nos itens, e sobre os itens, de um ou vários acervos, apresentando-as sob forma codificada e organizada, agrupada por semelhanças, aos usuários desse(s) acervo(s). (MEY, 1995, p. 9).

O processo comunicativo entre os usuários e os objetos documentais é possível de ser estabelecido pelo fato de catálogos serem formados por registros representativos dos documentos de que uma unidade informacional dispõe; em outras palavras, considera-se que a representação da informação trace mecanismos

de comunicação e de mediação entre demanda informacional de usuários e os documentos, por meio de catálogos, resultado dos constructos da representação bibliográfica.

A Figura 4 mostra o esquema clássico de comunicação:

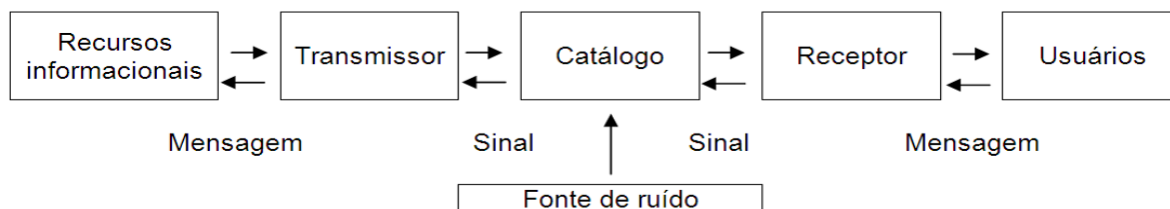
Figura 4 – Esquema clássico da comunicação



Fonte: Adaptado pelo autor a partir de Shannon (1948; 1969, p. 34).

Ao apresentar-se o ponto de vista do processo comunicativo que o catálogo exerce em instituições de patrimônio cultural, por conta da relação entre itens documentais e necessidades informacionais dos usuários, o esquema anterior pode assumir a configuração presente na Figura 5:

Figura 5 – Esquema do processo comunicativo entre recursos informacionais e usuários nas instituições de patrimônio cultural



Fonte: Elaborado pelo próprio autor.

A representação bibliográfica torna-se, assim, atividade fundamental em unidades de informação por veicular mensagem codificada sobre os recursos informacionais disponíveis e as necessidades de seus usuários. Cabe ao catalogador definir os elementos essenciais que satisfaçam as necessidades de seu público específico para o cumprimento de tal ação comunicativa. Nesse sentido, as informações de um registro documental voltadas a um público leigo distinguem-se das informações de um registro documental voltadas a um especialista. Atentar-se a estes aspectos na construção de um registro torna-se fundamental à estruturação de catálogos.

Por meio dos catálogos é dada a oportunidade aos usuários de encontrarem, identificarem, selecionarem e obterem os recursos bibliográficos necessários, o que, hoje em dia, pode incluir a navegação em um catálogo. Tais objetivos, consagrados

dessa forma atualmente, contam com um percurso histórico e de contribuição de vários agentes.<sup>7</sup>

Hunter e Bakewell (1983, p. 8-9) apresentam questionamentos sobre a indispensabilidade dos catálogos, dentre os quais se destaca o fato de que os usuários podem descobrir muito rapidamente, através de um terminal de computador, o que está disponível sobre um determinado autor ou assunto. Porém, o que usuários muitas vezes querem saber – e o que o computador com frequência não lhes diz – é o que está disponível em sua biblioteca local e onde é arquivado. É desejável ter um catálogo para mostrar o que está disponível e onde está arquivado, fornecendo diversas formas de acesso prováveis ao documento. Entretanto, existe outra razão importante para a manutenção de um catálogo: pode fornecer entradas analíticas para as partes de um item que são particularmente relevantes para a biblioteca.

Freedman (1984) indica que o delineamento dos catálogos, acerca de sua função e organização foi dado por quatro estudiosos: Antonio Panizzi, um anarquista italiano que imigrou para a Grã-Bretanha, onde foi o maior bibliotecário; Charles Coffin Jewett, definido por ele como um funileiro com tecnologia falha, e a quem Fred Kilgour admitiu que seguiu como modelo para estabelecer a rede OCLC (*Online Computer Library Center*); Charles Ammi Cutter, o único teórico da catalogação que teve alguma percepção sobre as bibliotecas públicas; e Seymour Lubetzky, um imigrante judeu polonês e, certamente, o melhor pensador do século XX sobre as funções de um catálogo bem como o que é necessário para defini-lo.

Destacam-se informações sumarizadas, a partir de Hunter e Bakewell (1983) para uma compreensão primeira sobre catálogos e objetivos bibliográficos.

Dentre as diversas contribuições para a elaboração de catálogos e seus objetivos, destaca-se a importante colaboração de Panizzi, ao traçar, em 1841, os primeiros objetivos bibliográficos, voltados à necessidade de um catálogo reunir e diferenciar itens similares. Posteriormente, em 1876, Cutter formula objetivos baseados nas necessidades do usuário e do que é preciso atender ao consultar um catálogo. Cutter explicita como sendo objetivos de um sistema bibliográfico:

---

<sup>7</sup> Dados históricos sobre os catálogos e seus objetivos bibliográficos podem ser observados de forma mais consistente em Garberson (2006), Svenonius (2000), Freedman (1984) e Hunter e Bakewell (1983).

- [1] permitir a uma pessoa encontrar um livro a partir
  - [a] do autor,
  - [b] do título ou do
  - [c] tema conhecidos;
- [2] mostrar o que a biblioteca tem
  - [d] de um determinado autor,
  - [e] sobre um determinado assunto,
  - [f] de determinado tipo de literatura;
- [3] auxiliar a escolha de um livro
  - [g] tanto pela sua edição (bibliograficamente)
  - [h] quanto pelo seu caráter (literário ou tópico).

Porém, em 1960, Lubetzky traz à tona a discussão acerca da diferença entre obra e livro (implícita em Cutter) e afirma a prioridade de informações de conteúdo, como nos remete o número de classificação. Na revisão proposta por Lubetzky os seguintes objetivos bibliográficos são apresentados:

- [1] facilitar a localização de uma publicação particular, i.e., uma edição particular de uma obra, que está na biblioteca, e
- [2] relacionar e trazer junto edições que a biblioteca tem de um dado trabalho e os trabalhos que ela tem de certo autor.

Tais princípios foram ligeiramente modificados na Conferência sobre Princípios de Catalogação, ocorrida em 1961, em Paris, os quais definem que o catálogo deve ser um instrumento eficiente para verificar:

- [1] se a biblioteca contém um livro particular especificado por autor e título, *ou*, se o autor não está nomeado no livro, seu título está sozinho, *ou*, se autor e título forem inapropriados ou insuficientes para a identificação, um substituto adequado para o título, e
- [2] quais obras de um autor particular e quais edições de uma obra particular estão na biblioteca.

Estes objetivos, porém, declaram, segundo Svenonius (2000), parcialmente a função que um catálogo deve apresentar, por serem restritos à organização de documentos por autor e título. Essa autora, afirma, ainda, que em 1997, com o intuito de revisar e racionalizar a prática biblioteconômica houve nova revisão dos

princípios pela IFLA, no qual se estabelecem os seguintes (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATION, 1998):

- [1] encontrar entidades que correspondam ao critério de busca estabelecido pelo usuário,
- [2] identificar uma entidade,
- [3] selecionar uma entidade adequada às necessidades do usuário, e
- [4] adquirir ou obter acesso à entidade descrita.

Svenonius (2000), porém, procede a análise crítica da correlação entre estes princípios e os constantes nos Princípios de Paris em vários momentos:

- [a] ao restabelecer os objetivos de escolha ao substituir selecionar por assistir na escolha de;
- [b] ao generalizar o vocabulário para que seja também pertinente ao global, ao não livro, ao ambiente digital;
- [c] ao resolver a ambigüidade que atribui ao objetivo tradicional de *encontro*, que pode ser entendido por *encontrar* a localização do documento na base de dados, *identificar* um documento, ou, *verificar* se o documento está disponível.

Na análise da autora, apresenta-se uma problemática quanto ao tradicional objetivo de *encontrar* o que está sendo procurado, seja um documento particular, definido por critérios como autor, obra e assunto. O primeiro objetivo da IFLA integra isto em um único objetivo. Entretanto, diminui a importância de conceito de arranjo (arrumação, colocação). Este conceito, para Svenonius (2000) está imbricado ao discurso bibliográfico. É particularmente útil pela ênfase que dá ao que, em primeira instância, seja o ato primário de organização da informação – trazer coisas juntas. Esta conotação é importante por tradição e valiosa demais para se perder.

Além de quebrar com a tradição, o primeiro objetivo da IFLA não especifica os registros de entidades para ser encontrado, mas relega esta tarefa ao modelo de relacionamento entre entidade e atributo. A problemática está no ponto de vista do design da base de dados. O *design* do objetivo de uma base de dados é determinado pela ontologia e não o contrário; desde que sejam dados os objetivos, modelos alternativos podem ser desenvolvidos para propósitos diferenciados. Além disso, a declaração dos princípios destes objetivos deve incorporar as necessidades dos usuários. Deve indicar apenas o que os usuários precisam encontrar. (SVENONIUS, 2000, p. 17-18, tradução livre nossa).

Svenonius (2000) reescreve, portanto, o objetivo proposto como *encontrar*, definido pela IFLA em 1997 (publicado em 1998, cf. INTERNATIONAL FEDERAL OF LIBRARY ASSOCIATION, 1998).

Outra contribuição da autora está na identificação de lacunas ao longo dos anos quanto às características inerentes a um sistema bibliográfico e, nesse sentido, propõe o princípio da *navegação*, haja vista que os quatro existentes (encontrar, identificar, selecionar e obter) mostram-se, a partir de seu ponto de vista, como insuficientes (SVENONIUS, 2000). Segundo a autora, esta proposta baseia-se em discussão registrada por Pierce Butler<sup>8</sup>, que apresenta a necessidade de um objetivo que caracterize o meio pelo qual se navega em um universo bibliográfico. A metáfora é adequada para descrever como ir do ponto A ao ponto B e para alcançar o destino – o documento desejado. O argumento para reconhecer explicitamente o objetivo da navegação tem duas partes: a primeira, desenhada para o comportamento de busca de usuários, e a segunda, para análise dos códigos tradicionais para a descrição bibliográfica.

A autora enfatiza que alguns usuários buscam por informação conhecendo exatamente o que querem, mas outros não sabem bem ou não são capazes de reconhecer o objeto de sua pesquisa, ou, ainda, não são capazes de reconhecer imediatamente quando encontram. Usuários esperam orientação e sistemas bibliográficos deram tradicionalmente essa orientação. Outra razão para postular o objetivo da navegação é que códigos de regras bibliográficas utilizadas para organizar documentos assumem essa existência. (SVENONIUS, 2000).

Idealmente, em cada regra de um código deve ser possível apontar um objetivo que o justifique. Muitas das regras não o justificam, mas algumas, atualmente, tem um propósito legítimo, no caso de os próprios objetivos serem questionados. Entre regras com um propósito legítimo estão aquelas que estabelecem relacionamentos bibliográficos. Tais regras podem ser encontradas em códigos tanto para a descrição autor-título quanto na descrição de assunto. Eles incluem regras que especificam a relação entre obras bem como as relações entre nomes dos atributos da obra, como autores e assuntos.

---

<sup>8</sup> Butler, P. The bibliographical function of the library. **Journal of Cataloging and Classification**, v. 9, p. 3-11, 1953.



Svenonius (2000) ressalta que relações entre obras incluem relações generalizadas (*é uma subclasse de*), relações de agregação (*é parte de*), e relações associativas (*é uma seqüência de, é uma adaptação de, é uma abreviação de, é descrito por*). Relações entre nomes de atributos de obra incluem equivalência, hierarquia e relações associativas.

O objetivo das regras que constituem essas relações é mapear o universo bibliográfico, isto é, para facilitar a navegação. Desse modo, o objetivo da navegação fornece garantia ao usuário e ao código.

A contribuição de Svenonius (2000) teve repercussão junto à IFLA, na publicação, em 2009, da *Declaração dos Princípios Internacionais de Catalogação*, que prima pelo destaque que se deve dar à conveniência dos usuários e que, diante dos *Princípios de Paris*, os substitui e amplia.

**4.1 Encontrar** recursos bibliográficos numa coleção como resultado de uma pesquisa, utilizando atributos e relações entre recursos:

**4.1.1** Para **encontrar** um determinado recurso.

**4.1.2** Para **encontrar** conjuntos de recursos representando todos os recursos que pertencem à mesma obra;  
todos os recursos que representam a mesma expressão;  
todos os recursos que exemplificam a mesma manifestação;  
todos os recursos associados a determinada pessoa, família ou colectividade (entidade);  
todos os recursos sobre um determinado assunto;  
todos os recursos definidos por outros critérios (língua, lugar de publicação, data de publicação, tipo de conteúdo, tipo de suporte, etc.), normalmente como uma delimitação secundária de um resultado de pesquisa.

**4.2 Identificar** um recurso bibliográfico ou agente (ou seja, confirmar que a entidade descrita corresponde à entidade procurada ou distinguir entre duas ou mais entidades com características similares);

**4.3 Selecionar** um recurso bibliográfico que seja apropriado às necessidades do utilizador (usuário), (ou seja, escolher um recurso que esteja de acordo com as necessidades do utilizador (usuário), no que diz respeito ao conteúdo, suporte, etc. ou rejeitar um recurso que seja inadequado às necessidades do utilizador (usuário));

**4.4 Adquirir** ou **obter** acesso a um item descrito (ou seja, fornecer informação que permitirá ao utilizador (usuário) adquirir um item por meio de compra, empréstimo, etc. ou aceder (acessar) eletronicamente a um item por meio de uma ligação em linha a uma fonte remota); ou acessar (aceder), adquirir ou obter dados bibliográficos ou de autoridade;

**4.5 Navegar** num catálogo ou para além dele (quer dizer, através da organização lógica dos dados bibliográficos e de autoridade e da apresentação de formas claras de se navegar, incluindo a apresentação de relações entre obras, expressões, manifestações, itens, pessoas, famílias, entidades (colectividades), conceitos,



objetos, eventos e lugares). (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATION, 2009, p. 3-4).

Observe-se a correlação que pode ser estabelecida entre os objetivos bibliográficos definidos por Cutter, em 1876, com destaque aos usuários, e aqueles constantes na *Declaração dos Princípios [...]*, de 2009. É possível notar que alguns movimentos para o estabelecimento da catalogação e dos catálogos estão imbricados. Note-se que, sem as *91 regras* de Panizzi não teria sido possível publicar o *Anglo-American code*, de 1908, ou, ainda, as *Regras da ALA*, de 1942 e de 1949. Estes, por sua vez, sedimentaram os trabalhos de Osborn (1941), Lubetzky (1953) e dos Princípios de Paris (1961), o que levou às *Anglo-American Cataloging Rules*, publicadas em 1967 e 1978. Desse modo, pode-se observar que a riqueza da catalogação está no século XIX, período em que, segundo Hunter e Bakewell (1983), é marcado pela contribuição de muitos bibliotecários para a melhoria de métodos; pela influência exercida por instituições como a Library of Congress e a British Library (àquela época denominada British Museum Library); por organizações nas quais bibliotecários uniram-se em conferências para que fosse possível conhecer, discutir e trocar pontos de vista; pela tentativa de normalizar as práticas de catalogação em códigos, entre tantos outros exemplos.

Sem dúvida alguma, por conta destas contribuições e da adoção de princípios bibliográficos que norteiam a elaboração de catálogos bibliográficos, as possibilidades de cooperação e o compartilhamento de registros bibliográficos, tornam-se factíveis. Essas facilidades são apresentadas no próximo tópico.

#### 2.4 COOPERAÇÃO, COMPARTILHAMENTO E INTERCÂMBIO DE REGISTROS BIBLIOGRÁFICOS

Na contemporaneidade, com o advento e a difusão das tecnologias de informação e comunicação, é possível a otimização de muitas atividades nas bibliotecas, dentre elas o processo de catalogação, que, em um sistema automatizado de gerenciamento de bibliotecas, contemplam não só a entrada de dados, mas a catalogação cooperativa e o recebimento de dados do sistema legado (MARCONDES, 1997). Entretanto, essa situação colaborativa não é nova.

Karisiddappa (2008, p. 283-285) relata, de forma geral, que a cooperação entre bibliotecas remonta a 200 a.C., com a *Alexandria Library* e a *Pergamum*

*Library*. Retrata que, durante o século XIII, a cooperação também esteve presente nos monastérios e que a união de catálogos de bibliotecas é do século XVII. Um marco na área cooperativa foi a compilação feita por Henry Bolton, em 1885, da obra intitulada *A Catalogue of Scientific and Technical Periodicals*. Em seus breves apontamentos, o autor relembra-nos dos inícios dos projetos de catalogação feitos pela *Library of Congress* no início do século XX; da publicação de Melvil Dewey, em 1886, sobre a cooperação entre bibliotecas; da contribuição de E. A. Mac sobre sua visão acerca da “cooperação versus competição” (ambos publicados na *Library Journal*); da afirmação de Ranganathan, na primeira metade do século XX, sobre o fato de os serviços de biblioteca, a organização bibliográfica e a cooperação bibliográfica não reconhecerem fronteiras nacionais ou políticas, por serem internacionais.

Segundo Cunha (1979, p. 8), foi em 1934, em decorrência da união entre a ALA e a Library of Congress que “concretizou-se o sonho de Charles C. Jewett, pioneiro da Catalogação Cooperativa nos Estados Unidos”, para quem “um livro deveria ser catalogado uma única vez, a fim de que o [tempo] despendido por uma biblioteca na catalogação de uma obra não tornasse a onerar não só essa mesma biblioteca, como também a qualquer outra” (loc. cit.).

A catalogação cooperativa, apresentada como sinônimo de catalogação coletiva ou catalogação compartilhada no Glosario ALA de Bibliotecología y Ciencias de la Información (YOUNG, 1988, p. 89), encontra-se definida como a catalogação original de itens bibliográficos, resultado da ação conjunta de um grupo de bibliotecas independentes que disponibilizam seus registros bibliográficos a outros membros do grupo, e, também, às vezes, às bibliotecas que não fazem parte do mesmo. Pode apresentar-se como catalogação compartilhada por compartilharem, também, responsabilidade e produtos da catalogação. No mesmo documento, ressalta-se que a catalogação cooperativa, nos Estados Unidos, é um tipo específico de catalogação desenvolvida pela Library of Congress com outras bibliotecas nacionais como parte do Programa Nacional de Aquisição e Catalogação para o intercâmbio de registros bibliográficos (op. cit., p. 313).

A partir do caráter de cooperação interbibliotecária, padrões tornam-se “essenciais para garantir maior eficiência e conectividade” (CAFÉ; SANTOS; MACEDO, 2001, p. 74). Entretanto, Castro e Santos (2005, p. 70), ao afirmarem

que, apesar de dados estruturados em meio magnético possibilitarem o intercâmbio de registros bibliográficos, caso a estrutura seja tão-somente semelhante, a compatibilidade não fica garantida. Este ponto de vista é reiterado na afirmação de Zumer e Lupovici (2009), para quem, embora o formato seja a estrutura que permite o processamento de dados bibliográficos processados em um computador, bem como muitos formatos forneçam a estrutura de dados para ser compreendida pelo computador, não há como afirmar que a estrutura seja suficiente para garantir a interoperabilidade de dados entre bibliotecas. É a opção de adesão a padrões, segundo Paranhos (2004, p. 27), que proporciona “benefícios e diminuição de riscos” e isso decorre, tanto por evitar o retrabalho, quanto por facilitar ações futuras de migração de *softwares* de gerenciamento integrado de unidades de informação.

O formato para descrição de registros bibliográficos em meio automatizado mais usado, internacionalmente, é o MARC21 e, por meio dele é possível realizar intercâmbio desses registros. Café, Santos e Macedo (2001, p. 74) afirmam que adotar o formato MARC garante segurança ao processo de migração de dados do sistema legado para o novo. Dziekaniak et al. (2008) relembram que, pelo uso do formato MARC e de protocolos de comunicação de dados, há “salto qualitativo no processo de disseminação da informação eletrônica”, o que, a nosso ver favorece a interoperabilidade.

Interoperabilidade, para Miller (2000, tradução livre nossa), é o processo contínuo que assegura que sistemas, procedimentos e cultura de uma organização sejam gerenciados de forma a maximizar oportunidades de intercâmbio e de (re)uso de informações, seja interna ou externamente. Essas características favorecem a “economia no tratamento de informação otimizando a troca desta informação.” (DZIEKANIAK et al., 2008). Conclui-se que a interoperabilidade deva fazer parte da cultura organizacional de cada uma das instituições parceiras por adotar padrões, que, segundo Noerr (2003, p. 93, tradução nossa), apresentam a função de

unificar a representação, a manipulação ou a transmissão de algum item de informação de forma que dois ou mais diferentes sistemas possam ‘entendê-lo’ da mesma maneira. Eles [os padrões] são a base da interoperabilidade, da portabilidade, da modularidade, dos blocos de construção, dos objetos e de todos os outros nomes inventados para descrever como dois itens de um software são capazes de, simplesmente, trabalharem juntos.

Ao processo interoperável, que pode ser estabelecido entre várias unidades de informação, é requerido o uso de formato bibliográfico. Marcondes (1997) define

formato bibliográfico como “um padrão para intercâmbio de dados entre sistemas computadorizados através de um arquivo em meio legível por computador.” Segundo os apontamentos feitos por Marcondes (1997), Côrte et al. (1999), Café, Santos e Macedo (2001) e Dziekaniak et al. (2008), sumariza-se os seguintes componentes do formato bibliográfico, necessários ao compartilhamento por meio da interoperabilidade:

- [1] o padrão de representação do conteúdo do documento, pelo AACR2, por exemplo;
- [2] o formato de catalogação legível por máquina, pelo MARC21, como padrão de estrutura de metadados descritivos;
- [3] o formato de intercâmbio do registro, pela ISO 2709, identificado como protocolo síncrono com arquitetura cliente-servidor; e
- [4] o protocolo de comunicação e recuperação de informação bibliográfica, pelo Z39.50, que garante que máquinas procedam a transferência de dados.

Assim, a interoperabilidade, a cooperação e o compartilhamento requerem a adoção de padrões e de formatos bibliográficos. Em uma atividade interoperável, a catalogação cooperativa é privilegiada por caracterizar, em muitas ocasiões, um pré-requisito para outras formas de cooperação, pois, sem a informação adequada poderiam ser inúteis os convênios de intercâmbio ou os programas de desenvolvimento cooperativo das coleções (MORALEJO ÁLVAREZ; MARQUINA GARCÍA; ABAD HIRALDO, 1989).

Watson (2001, tradução livre nossa) ressalta a importância de administradores de bibliotecas apoiarem ações cooperativas. Para tanto, enumera cinco razões a serem consideradas diante da opção de participarem ou não de programas cooperativos de catalogação.

- Razão 1: Porque é caro!: participar de programas cooperativos de catalogação permite que bibliotecas criem e compartilhem experiências que não seriam possíveis para cada uma isoladamente. Embora haja alguma despesa inicial, o resultado mostra uma redução geral nos custos de catalogação, tanto nos registros bibliográficos quanto nos de autoridade;
- Razão 2: Porque é trabalhoso!: participar de programas cooperativos de catalogação fornece um mecanismo para captura e compartilhamento amplo

de resultados de alta qualidade desenvolvidos pela equipe de catalogação. A catalogação, em geral, é uma atividade que requer trabalho inicial intenso que quando compartilhada, inclusive quanto aos padrões, torna-se mais amena para todos;

- Razão 3: Porque envolve custo excessivo!: ter as atividades de catalogação em uma instituição exigem um catalogador muito experiente e, talvez, um dos mais bem pagos da instituição por dedicar grande parte do seu tempo ao programa, o que aumenta os custos indiretos. Enfim, investir é necessário para que a participação em um programa de catalogação cooperativa pague seus próprios dividendos em um curto período de tempo;
- Razão 4: Porque leva tempo a produção da catalogação!: participar de programas de catalogação cooperativa permite que o tempo dedicado à catalogação seja otimizado, uma vez que, normalmente, os bibliotecários envolvem-se em muitas outras atividades na biblioteca. Além disso, há melhora no nível dos registros, cuja qualidade dependia exclusivamente de sua própria unidade;
- Razão 5: Porque catalogadores perdem a auto-estima!: participar de programas de catalogação cooperativa oferece a oportunidade aos catalogadores de se tornarem membros de um grupo de forma a abrandar e reservar algum tempo para considerar a sua auto-imagem, o valor agregado das atividades de catalogação e sobre o que é ou não essencial em um ambiente cooperativo. Afinal, diante da perspectiva de participação em programas de catalogação cooperativa, catalogadores ficaram mais animados e entusiasmados com o desenvolvimento de suas atividades.

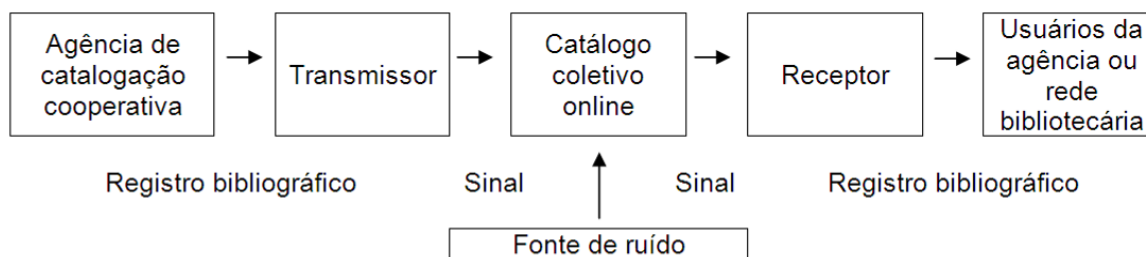
A catalogação compartilhada tem seu espectro ampliado por decorrência da automação, e, por constituir-se pré-requisito para outras formas de cooperação, supõe, sempre, um esforço para a normalização dos dados. Vasconcellos (1996) ressalta que, diante da necessidade de as bibliotecas compartilharem registros bibliográficos, houve a organização em consórcios ou redes, o que proporciona “reaproveitamento dos registros bibliográficos.” Esse tipo de catalogação pode ser o resultado de um acordo entre um número de centros para compartilharem esse trabalho, de forma que a catalogação do documento efetuada por um dos elos dessa rede sirva para os restantes. O produto dessa cooperação é a criação de um

catálogo coletivo online que reúne fundos de todos os centros participantes e leva ao entendimento do conceito de base de dados cooperativa, proposto por Souza e Mostafa (1999, p. 128):

Entende-se por base de dados cooperativa a reunião dos registros bibliográficos de diferentes unidades de informação em meio magnético, constituindo-se em um catálogo coletivo onde os mesmos podem ser consultados e aproveitados por qualquer uma das unidades de informação integrantes da rede.

A catalogação cooperativa permite que não seja catalogado o que já está catalogado em algum ponto da rede e serve de canal de transmissão de documentos entre certo número de bibliotecas. Garrido Arilla (2001) detectou um paralelo entre o processo comunicativo e a catalogação cooperativa, que pode ser visualizado na Figura 6:

Figura 6 – Esquema do processo comunicativo para a catalogação cooperativa



Fonte: Adaptado pelo autor a partir de Garrido Arilla (2001).

Borgman (1997) afirma que, ao compartilhar dados catalográficos, substancial eficiência e controle de custos internos são obtidos, pois ao se evitar representar um item que está disponível em outra unidade de informação, membro de uma rede de cooperação, poupam-se custos com catalogação, uma das atividades mais caras da biblioteca pelo fato de recorrer à identificação unívoca do item. Partindo-se do pressuposto de que a maioria dos itens de uma unidade é publicada em várias cópias idênticas, torna-se altamente valioso contar com o fato de que se um registro é criado com alta qualidade, outra unidade poderá usá-lo sem ter que recriá-lo, o que permite, também, uma descrição consistente do mesmo.

Algumas vantagens da catalogação cooperativa foram observadas em Souza, Catarino e Santos (1997, p. 97 *et seq.*). Dentre elas, destaca-se o fato de que as unidades de informações envolvidas cooperativamente apresentam ações voltadas para saldar deficiências de recursos, por meio do compartilhamento, bem como as oportunidades para sua otimização. Observa-se, também, a redução de esforços e

de custos com o processamento técnico, por meio do aperfeiçoamento dos serviços cooperados entre as instituições participantes.

Segundo Krzyzanovski (2007, p. 22), Rubens Borba de Moraes, em 1943, já apresentou a necessidade de que um sistema de bibliotecas trabalhasse em conjunto para suprir as necessidades de outras bibliotecas, formando uma rede bibliotecária. Em 1997, a mesma autora, com outros colegas, afirma que é fundamental a cooperação interbibliotecária no que se refere a acordos interinstitucionais com ações voltadas às coleções, às bases de dados, às redes e aos sistemas de informação e aos profissionais (KRZYZANOVSKI et al., 1997, p. 175). Entretanto, em pesquisa feita junto às bibliotecas universitárias da região sul do Brasil, Dziekaniak et al. (2008), constatam que a interoperabilidade, em bibliotecas que se utilizam de *softwares* que possibilitam o uso de padrões compartilháveis, ainda não é privilegiada por grande parcela delas. Também há outra dificuldade:

[...] sistemas serem desenvolvidos para uma única biblioteca, de maneira particular, sem projetar e prever a possibilidade/necessidade de relacionamento entre arquivos de outro sistema com outras bibliotecas, como o uso do padrão MARC juntamente com o uso de protocolos de comunicação possibilitaria. (DZIEKANIAK et al., 2008).

A constatação de Dziekaniak et al. (2008) reforça o ponto de vista apresentado de que muitas bibliotecas ainda não se utilizam do Formato. Em sua pesquisa, na região sul brasileira, se utilizam do formato MARC21 somente 68% das bibliotecas do Paraná, 62% em Santa Catarina e 60% no Rio Grande do Sul. Outra questão que se julga agravante no processo de compartilhamento é o fato de a pesquisa detectar que muitas bibliotecas, em processo inicial de informatização, optaram pelo uso do Winlisis e, nesse caso, concorda-se com Dziekaniak et al. (2008):

Quanto ao uso deste tipo de software, - apenas de cadastro e recuperação na base de dados -, seu uso corrobora para a ausência de interoperabilidade entre as bibliotecas, uma vez que não permite a padronização dos dados através da utilização de um modelo de metadados, como é o caso do MARC, além de limitar a gestão de uma biblioteca.

Côrte et al. (1999, p. 246) ressaltam a necessidade de que, em um processo de automação de uma unidade de informação, “Qualquer que seja a opção pelo software, a biblioteca precisará incluir estes registros no novo sistema”, processo ao



qual se denomina conversão retrospectiva. Os autores abordam algumas opções para esta atividade:

Para as fichas, a opção mais tradicional é a digitação de todos os registros no novo sistema. Esta atividade pode ser mais demorada, mas, por vezes, é recomendável, pela garantia da qualidade do resultado final. (CÔRTE et al., 1999, p. 247).

Outra solução é a digitalização destas fichas. O arquivo resultante deste processo será convertido eletronicamente e incorporado à base de dados. Pode-se também, buscar em outras bases, pela consulta direta ou digitalização das fichas, o registro catalográfico e incorporá-lo à nova. Para o acervo que está em meio magnético, um programa de conversão de dados pode ser utilizado. A decisão de qual caminho será tomado para fazer a CR é muito importante. (op. cit., p. 248).

Em face de solução para o que se mostra, sob o ponto de vista desta pesquisa, como problema diante de dois apontamentos feitos pelas autoras é que se busca novo processo de conversão de dados bibliográficos, a partir de análises sintáticas e semânticas. O próximo tópico volta-se aos estudos da conversão de registros bibliográficos.

## 2.5 MÉTODOS E PROCESSOS DE CONVERSÃO DE REGISTROS BIBLIOGRÁFICOS

Oliveira et al. (1998, p. 41) indicam que bibliotecários perceberam a necessidade de conversão retrospectiva de seus acervos na década de 1980 ao se depararem com “dificuldades em registrar, no formato eletrônico, grande parte da produção intelectual da humanidade depositada nas bibliotecas e livrarias.”

Considerando-se que o tratamento técnico e o registro de informações bibliográficas, sem dúvida, são as atividades nas quais mais incidem os custos de um processo de automação, e do pressuposto de que o “princípio fundamental da moderna catalogação, segundo o qual não é economicamente aceitável elaborar um novo registro bibliográfico para um documento quando estiver disponível um registro satisfatório codificado por outra fonte” (VASCONCELLOS, 1996, p. 450), torna-se fundamental garantir que os dados em meio digital do acervo sejam (re)utilizados.

Nesse sentido, um requisito importante diz respeito à estrutura e ao formato de descrição do conteúdo na representação dos recursos informacionais, uma vez que isso evita a redigitação dos dados. Para tanto, é requerida a garantia da base



tecnológica e metodológica fornecida pela adoção de padrões, que, por sua natureza, promovem compatibilidade.

Segundo Dziekaniak et al. (2008), Paranhos (2004), Côte et al. (1999) e Marcondes (1997), para fins de intercâmbio de registros bibliográficos, os padrões a serem adotados, indicados por privilegiarem o compartilhamento de dados, envolvem o Formato MARC 21 – padrão de estrutura de um registro bibliográfico legível por computador; a ISO 2709 – padrão que apresenta uma estrutura interpretável por computadores para a comunicação entre diferentes sistemas; e o protocolo Z39.50 – protocolo que autentica a comunicação em uma estrutura cliente-servidor, com fins de identificação e recuperação de registros bibliográficos. Entretanto, Prado e Abreu (2002), ao apresentarem os formatos padronizados em bibliotecas universitárias do Estado de Santa Catarina, destacam a adoção do Formato MARC por mais de 70% das instituições pesquisadas, sendo, porém, que somente 11% fazem uso do protocolo Z39.50.

Em se tratando de intercâmbio de dados bibliográficos, uma das principais atividades envolve a conversão de dados bibliográficos. Porém, faz-se necessário esclarecer alguns termos, usados, algumas vezes, como sinônimos: migração e conversão de dados bibliográficos.

Migração, em Cunha e Cavalcanti (2008, p. 250), refere-se à

transferência de uma aplicação de um computador ou de um sistema operacional para outro. O objetivo é preservar a integridade dos objetos digitais e fazer com que os usuários continuem com a possibilidade de recuperar e utilizar a informação digital armazenada, apesar das contínuas mudanças tecnológicas.

Neste caso, essa definição apresenta-se com grande preocupação voltada à preservação digital, a qual exige uma “política formal de migração ou atualização regular dos equipamentos, programas e suportes” (CUNHA; CAVALCANTI, 2008, p. 290). Entretanto, em outro momento, os autores indicam que migração refere-se à “conversão dos dados de um formato para outro, quando se adquire um novo sistema de automação ou se muda para nova versão.” (op. cit., p. 250). Essa acepção mostra-se confusa, para fins desta pesquisa, por envolver os termos *conversão* e *formato*, que parecem voltar-se, com maior propriedade, à conversão de dados bibliográficos.

Algumas das acepções presentes em Cunha e Cavalcanti (2008, p. 107) para conversão são:

- a) “processo de mudança de um método ou forma de representação, ou seja, de processamento, para outro método”;
- b) “modificação das formas de registro da informação, no que se refere à tradução, codificação, estrutura dos dados, etc., objetivando a possibilidade de intercâmbio entre dois ou mais serviços ou sistemas, que utilizam convenções e métodos diversos”;
- c) “processo pelo qual um registro bibliográfico impresso é transformado em registro bibliográfico legível por máquina.”

No enfoque desta pesquisa, a primeira definição apresentada não seria a mais indicada, uma vez que indica a possibilidade de se proceder à representação documental. Entretanto, a segunda e a terceira indicações parecem-nos mais cabíveis, por abordarem a codificação e a estruturação dos dados de forma a serem passíveis de troca do modo manual para o meio automatizado. Essas acepções são compatíveis com as indicadas por Oliveira et al. (1998) e Côrte et al. (2002) que apresentam a expressão conversão retrospectiva para designar o processo de mudança do suporte no qual um registro bibliográfico está inscrito ou, ainda, o processo de conversão de registros que ainda estão disponíveis para consulta manual, por meio de fichas, para o processo automatizado, em meio magnético, legível por máquina. Esse ponto de vista também é compartilhado por Garrido Arilla (2001), que estabelece que os objetivos a serem alcançados com a conversão retrospectiva são a criação de um único arquivo legível por máquina do catálogo manual existente, a não duplicação de tarefas, e a possibilidade de se evitar custos de catalogação, uma vez que os registros serão copiados e não digitados.

Entende-se que o processo de conversão possa fazer parte do processo de migração. Entretanto, nesta pesquisa, migração referir-se-á, tão somente, ao tratamento dos dados tendo em vista a mudança de *software*. Conversão, por sua vez, será entendida como o processo em que a alteração se dá no nível da estrutura do registro, o que não envolve a alteração da descrição de seu conteúdo. A conversão de dados bibliográficos assume-se, portanto, como meio de se proceder a mudança de padrão de estrutura de metadados descritivos dos recursos informacionais.

Para entender, de modo prático, como funciona a conversão retrospectiva recorre-se a Marcondes (1997, p. 4, grifo do autor):

Num processo de conversão retrospectiva geralmente dados são migrados do sistema antigo para o novo sistema através de arquivos estruturados segundo um Formato de Intercâmbio. Neste processo, deve-se levar em conta então que há necessidade de haver compatibilidade, a nível a- do 'lay out' do registro em meio legível por computador; b- do elenco de campos que formam os dois sistemas, isto é, dos seus conteúdos; e finalmente, c- das regras de transcrição destes conteúdos. Caso somente parte destes requisitos de compatibilidade sejam atendidos, há necessidade de um trabalho de conversão e compatibilização, que pode incluir trabalho de análise programação e eventualmente, digitação e edição de dados.

Outros trabalhos apresentam aspectos do processo de conversão e, mais especificamente, a metodologia adotada nessa atividade.

Castro e Santos (2005) destacam que automatizar uma unidade de informação e não recorrer à conversão retrospectiva é impensável, haja vista que, por meio dela, é possível intercambiar dados em meio digital.

Verificamos que a conversão retrospectiva ainda é uma necessidade nacional e que muitas bibliotecas e muitos profissionais desconhecem esta prática, considerando que na automação de uma biblioteca, o serviço de conversão deve caminhar em paralelo com a automação. (CASTRO; SANTOS, 2005, p. 82).

Nesse sentido, Paranhos (2004) reconfirma que a aderência aos padrões é necessária à compatibilidade de registros e, conseqüentemente, ao intercâmbio, uma vez que, dependendo da forma com que uma base bibliográfica é construída, pode haver retrabalho.

Silva, Alvez e Viapiana (2008) descrevem que houve insucesso em seu primeiro processo de migração, especificamente quanto à carga de dados do *software* legado, pelo fato de este ser incompatível com o Formato MARC. Outra experiência, de Silva e Favaretto (2005), descreve metodologicamente, a migração dos dados bibliográficos, ambos com registros incrementados em Formato MARC, no qual se destacam a decisão institucional pela migração; visitas a outras instituições e avaliação dos *softwares* adotados; elaboração de um plano de migração, no qual constam mapeamento de pontos de alteração, cronograma, análise de requisitos de *hardware* e de *software*; estabelecimento de novos fluxos de trabalho; análise da estrutura do sistema legado; análise dos dados bibliográficos; extração e exportação da base legada para testes e simulações (fase crítica de

sucesso); conversão<sup>9</sup> e treinamento. Castro e Santos (2005, p. 79) destacam como fases para a conversão retrospectiva ações de “pesquisa do registro em bases servidoras, cópia do registro, edição do registro e importação do registro.” Como procedimentos, destacam:

Quando é encontrado um registro IT [identidade total], o mesmo é captado e adequado ao padrão de registros da UNESP para posteriormente ser inserido no Catálogo Coletivo UEP01 [base de dados bibliográficos]. Se o registro for um RA [registro aproveitável], será captado e adequado aos moldes do documento da unidade. Se for um NL [não localizados], é pesquisado na próxima base. Esgotadas as opções de localização, é realizada a catalogação original, para ser inserido na base como um novo registro, atribuindo as características necessárias para a cooperação com outras bases. (CASTRO; SANTOS, 2005, p. 81).

Hübner (2002) destaca que o sucesso de um projeto de conversão retrospectiva depende da identificação de informações do acervo, tais como quantidade, tipo e suporte; da tabulação dos detalhes a serem sanados na conversão; da aplicação de teste piloto em amostra de registros; da determinação de rotinas e fluxos de trabalho; da estimativa de taxa de conversão e de catalogação; e, por fim, do cálculo de tempo e custos envolvidos. A análise de custos e prazos também é reiterada por Oliveira et al. (1998, p. 42), que ressaltam que o método de conversão retrospectiva a ser adotado deve adequar-se “às características e necessidades das bibliotecas”, de modo a garantir “prazos de execução razoáveis, custo baixo e [...] a qualidade das informações.” Castro e Santos (2005) destacam que utilizar processos de conversão retrospectiva promove economia tanto de recursos financeiros quanto humanos.

Perez e Lima (2002), por sua vez, destacam que a tarefa de conversão retrospectiva é onerosa e intensa e requer a observância de dois aspectos fundamentais na definição do método a ser adotado: um deles recai no custo que se torna alto em se tratando de tratamento externo dos dados; o outro aborda a necessidade de, dependendo da instituição escolhida como base para a cópia dos registros, fazer a adequação de pontos de acesso de assunto, de notas e outras informações locais.

---

<sup>9</sup> Note-se que os termos conversão e migração foram tomados como sinônimos pelos autores.

Garrido Arilla (2001), ao apresentar o processo de conversão retrospectiva afirma que, aparentemente, o trabalho de catalogação é quase nulo, uma vez que o registro já está feito. Entretanto, vários trabalhos técnicos tornam-se necessários. Um deles é o confronto dos documentos com a base de dados fonte e a captura dos registros correspondentes à chave de busca que seja mais apropriada à revocação com maior índice de precisão. Outra atividade é a análise do grau de adequação dos registros com a base de dados destino e, conseqüentemente, a sua correção, momento em que é necessário suprimir alguns campos, modificar o conteúdo de outros, além de acrescentar informação local, etc. Esse aspecto também aparece em Souza e Mostafa (1999) que ressaltam que, embora a participação em redes de bibliotecas favoreça atividades de catalogação cooperativa, principalmente quanto à descrição do recurso, problemas quanto à definição dos pontos de acesso podem se apresentar. Entretanto, Castro e Santos (2005, p. 84) afirmam que “o processo de conversão retrospectiva é considerado um método vantajoso para a automação de uma unidade de informação e que o profissional bibliotecário está cada vez mais utilizando dessa ferramenta.”

Ao finalizarem-se, brevemente, aspectos inerentes à catalogação, seus padrões e produtos, bem como a cooperação, compartilhamento e intercâmbio de registros representativos dos documentos e os processos e metodologias de conversão de registros bibliográficos, parte-se para o segundo ponto, concernente à adoção de aspectos sintáticos e semânticos do registro bibliográfico.

### 3 UM OLHAR DA LINGUÍSTICA SOBRE OS REGISTROS BIBLIOGRÁFICOS

Este capítulo destina-se aos estudos da contribuição lingüística da comunicação humana, feitos por Saussure, e da concepção semântica estruturalista, com Hjelmslev, para a elaboração do arcabouço teórico da sintaxe e da semântica dos registros bibliográficos.

#### 3.1 CONTRIBUIÇÃO SAUSSUREANA À REPRESENTAÇÃO DE RECURSOS INFORMACIONAIS

Entende-se que, da mesma forma que a lingüística se forma pelas manifestações da linguagem humana (SAUSSURE, 2010, p. 13), o papel social das instituições de patrimônio cultural forma-se a partir da representação de tais manifestações, levando-se em conta aquelas registradas, independentemente do meio e do suporte em que isso seja feito. As manifestações da linguagem humana, por meio de seus registros, permitem a descrição, a identificação, o uso, o reuso, a disseminação e o compartilhamento entre as mais diversificadas instituições de patrimônio cultural.

O fenômeno da representação dos recursos informacionais integra duas faces que se correspondem e se complementam: a *obra* e a *manifestação*, ambas reconhecidas pela International Federation of Library Associations and Institutions (1998, 2005, 2009) como produtos do esforço intelectual ou artístico.<sup>10</sup>

*Obra* é a criação intelectual ou artística que reflete o conteúdo e é identificada como uma entidade abstrata. Para Smiraglia (2002), *obra* é o conhecimento deliberadamente criado para representar um conjunto coordenado de idéias (ou seja, o conteúdo ideacional), que veiculado por meio do texto<sup>11</sup> tem a finalidade de

---

<sup>10</sup> Pelo fato de o estudo entre as relações entre as entidades *obra*, *expressão*, *manifestação* e *item* não ser objeto desta pesquisa recomenda-se a leitura de International Federation of Library Associations and Institutions (1998), Mey (1998), Smiraglia (2002), Rios Hilário (2007), Tillet ([2007?]), Maxwell (2008), Denton (2009), Santos e Corrêa (2009) e Mey e Silveira (2009).

<sup>11</sup> Com o intuito de aclarar o significado de texto neste trabalho recorreremos a Trask (2008, p. 291), para quem o texto é “uma porção contínua da língua falada ou escrita, especialmente quando tem um começo e um fim reconhecíveis.” O texto é “uma abstração”.

ser comunicado ao consumidor. Um documento pode conter uma ou mais *obras*, e uma obra pode existir em um ou mais documentos, o que significa que é possível que exista em várias instâncias.

*Manifestação* é a corporificação de uma *Obra*, que só é possível de ser conhecida se manifestada, ou, de outro modo, a *manifestação* só existe a partir da concepção de uma *obra*; a *obra* só pode ser reconhecida por meio da *manifestação*. A manifestação assume a forma física.

Com base na abordagem dada por Smiraglia (2002), entende-se que eventuais equívocos sobre a relação entre *obra*, *expressão*, *manifestação* e *item* possam ser saldados do seguinte modo: uma *obra* começa como um conjunto de impressões (conceitos ideacionais) na mente de seu criador (intenção autoral); uma vez que o criador refletiu suficientemente sobre essas impressões, de modo a assumir a forma mais adequada de apresentação ordenada de seu conceito, então esta *obra* tem condições de assumir as características de *expressão* (conteúdo ideacional). Essa forma poderá vir a ser realizada em uma expressão alfa-numérica, musical, sonora, imagética, etc. Tendo a *obra* sido expressa, ocorre a corporificação da obra, ou seja, sua *manifestação* concreta, a qual recorre a um conjunto específico de cadeias semânticas e ideacionais em um conjunto ordenado fisicamente realizado ou, melhor dizendo, assume uma instância física, inclusive em meio eletrônico e digital. A *manifestação*, por sua vez, pode ser corporificada em um ou mais *itens*.<sup>12</sup>

Assim, a representação da informação só é possível por meio da compreensão da correspondência entre *obra* e *manifestação*. Não se pode, portanto, reduzir a representação a uma ou a outra face: *obra* é resultado do pensamento, sem que, entretanto, tenha sido manifestada, vindo a público. Surge, por conseguinte, daí, a correspondência entre a *obra*, unidade complexa mental, e a *manifestação*, unidade complexa física. Entenda-se, portanto, que a *manifestação* é o “traje” que uma *obra* ocupa.

Partindo-se desta perspectiva, pode-se avistar um dilema: o que de fato se representa? Ou envolve-se com a explicação de que é a *obra*, por ser concepção mental primeira; ou arrisca-se a perceber que é a *manifestação*, registro em suporte

---

<sup>12</sup> A *expressão* e o *item*, embora inerentes à compreensão do recurso informacional, não são objeto de estudo nesta pesquisa.

físico da concepção mental. Parte-se do ponto de vista nesta pesquisa, de se observar a representação a partir da *manifestação*.

Representação da informação é, portanto, o ato de articular formas de descrição a partir de instrumentos que permitam tornar cognoscível um recurso informacional sem que seja necessário recorrer ao documento original para identificá-lo.

Compreende-se, portanto, que os recursos informacionais constituem-se de *manifestações* socializáveis de obras, estas, por sua vez, individuais ou coletivas. Quando assim se procede, opta-se pela compreensão de que o objeto da Ciência da Informação volta-se aos documentos e, ao considerar-se como seu objeto a informação registrada, os elementos que serão representados subordinam-se a esta questão. Do exposto, podem-se inferir mais algumas compreensões importantes acerca de *obra* e de *manifestação*.

Considere-se, por exemplo, a produção de elementos para a *manifestação*: a concepção mental é tão exterior à *manifestação* quanto os suportes que servem para registrá-la. Neste sentido, a *obra* independe da forma material em que é expressa e em que é manifestada.

Outra separação entre *obra* e *manifestação* é possível de ser indicada a partir do conhecimento e da aproximação que se tem de uma *obra* a partir das variadas *manifestações* possíveis. Se não, observe-se: é possível se conhecer a *Sinfonia nº 9 em Ré Menor, op. 125*, de Ludvig van Beethoven, a partir da apresentação da *Orquestra Sinfônica do Estado de São Paulo* ou, ainda, pela *Berliner Philharmoniker*. Entenda-se, portanto, que esta *obra* de Beethoven só foi possível de ser (re)conhecida por meio do recurso informacional feito por aquele criador expresso em partituras. Outra forma de essa *obra* chegar ao (re)conhecimento público seria por outra expressão: o registro sonoro daquelas apresentações.

Retoma-se, pelo exposto, que o papel das instituições de patrimônio cultural, entenda-se como o de comunicar fatos e registros às futuras gerações da humanidade, só é possível de se realizar a partir de registros bibliográficos que, por sua vez, são responsáveis por intermediar recursos informacionais a usuários com demandas específicas.

O enfoque entre *obra* e *manifestação* e sua relação com o processo comunicativo norteiam o estudo de uma possibilidade teórica advinda do campo da



linguagem para se pensar os registros da informação, como o sugerido por Ferdinand de Saussure, em 1916, relacionando-o com a questão do *significado* e do *significante*. Para tanto, estuda-se a correspondência entre *obra* e *manifestação*, *significado*, *significante*.

Nöth (1996) representa os três conceitos expostos no modelo sígnico diádico definidos por Saussure, como indicado na Figura 7:

Figura 7 – Os três conceitos do modelo sígnico diádico de Saussure

SIGNO	SIGNIFICADO
	SIGNIFICANTE

Fonte: Nöth (1996, p. 32).

Do mesmo modo que Saussure (2010, p. 81 *et seq.*) apresenta princípios como a *arbitrariedade do signo* e o caráter *linear do significante*, buscar-se-á explicitar a correlação entre *obra* e *manifestação* presentes nesses princípios de Saussure.

- a) o primeiro princípio: “O laço que une o significante ao significado é arbitrário” (op. cit., p. 81). Observa-se, como exemplo, a arbitrariedade da *manifestação* em relação à *obra*. Neste sentido, a idéia de uma *obra*, como *100 cientistas que mudaram a história do mundo*, de John Hudson Tiner, não apresenta relação direta com, tão somente, uma forma de *manifestação*; essa ligação é arbitrária e pode assumir outras tantas formas: um roteiro de uma peça teatral ou de um cinema; um musical; um livro; entre outras. Na prática, a forma de *manifestação* selecionada pode ser qualquer uma, desde que evoque a *obra* por meio da corporificação;
- b) o segundo princípio: “O significante [...] desenvolve-se no tempo [...]” (op. cit., p. 84). Esse princípio, de tão notável relevância quanto o primeiro, é passível de ser notado pela linearidade que um registro assume, há uma sequência para o registro da *obra* na *manifestação* do produto mental (conteúdo ideacional), independentemente de sua forma.

Ao buscar-se adaptar o modelo de Saussure à representação dos recursos informacionais obtém-se o modelo presente na Figura 8:

Figura 8 – Os três termos do modelo para a representação da informação

REPRESENTAÇÃO	OBRA
	MANIFESTAÇÃO

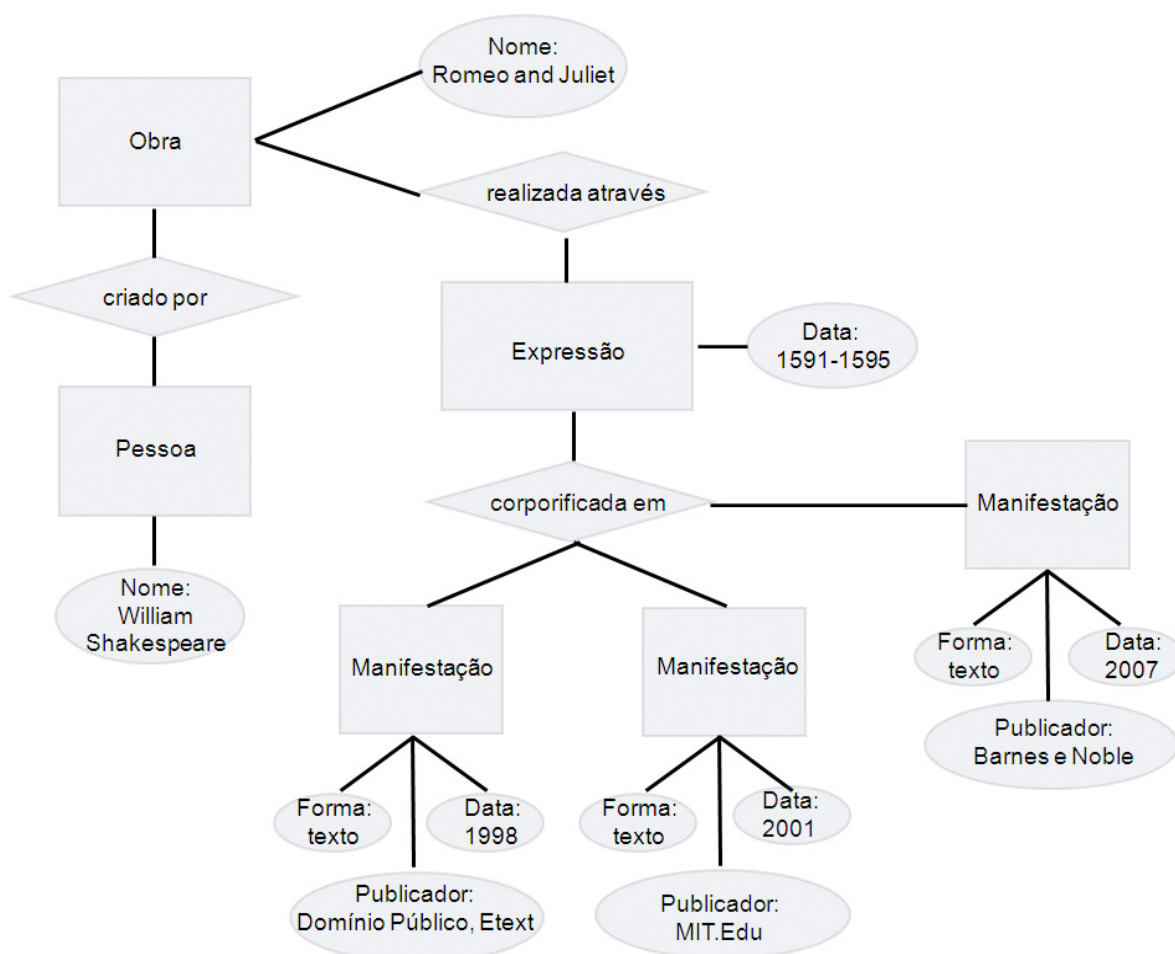
Fonte: Elaborado pelo próprio autor.

A *obra* faz menção ao conceito mental, ou, para remeter a Saussure, ao *significado*, ao conceito; a *manifestação*, por sua vez, remete ao *significante*, à imagem acústica registrada. A *obra*, reduzida a um princípio essencial para a *manifestação*, apresenta correspondência entre tantas formas de *expressão* quantas forem possíveis. Como exemplo, tome-se a obra *Romeo and Juliet*, de William Shakespeare, criada entre 1591 e 1595. Essa obra, originalmente em inglês (*expressão*), apresenta expressões correspondentes em diversificados idiomas, cada qual podendo assumir diversas manifestações, e, inclusive, adaptações, o que se desdobra como nova obra, para teatro, cinema, música, literatura, história em quadrinhos. Entende-se, deste modo, o processo de representação da *obra* pela *manifestação*.

Com o intuito de ilustrar as relações de equivalência atinentes à *manifestação*, à *expressão* e à *obra*, baseados no diagrama de entidade-relacionamento proposto por Maxwell (2008, p. 75), propõe-se a observação das análises a seguir: da *obra Romeo and Juliet*, de William Shakespeare, e da *obra Ninth symphony*, de Ludvig van Beethoven.

Na Figura 9 analise-se a *obra Romeo and Juliet*, a qual apresenta a relação de três *manifestações* para uma *expressão*.

Figura 9 – Representação gráfica das relações entre *obra*, *expressão* e *manifestação* (*Romeo and Juliet*).

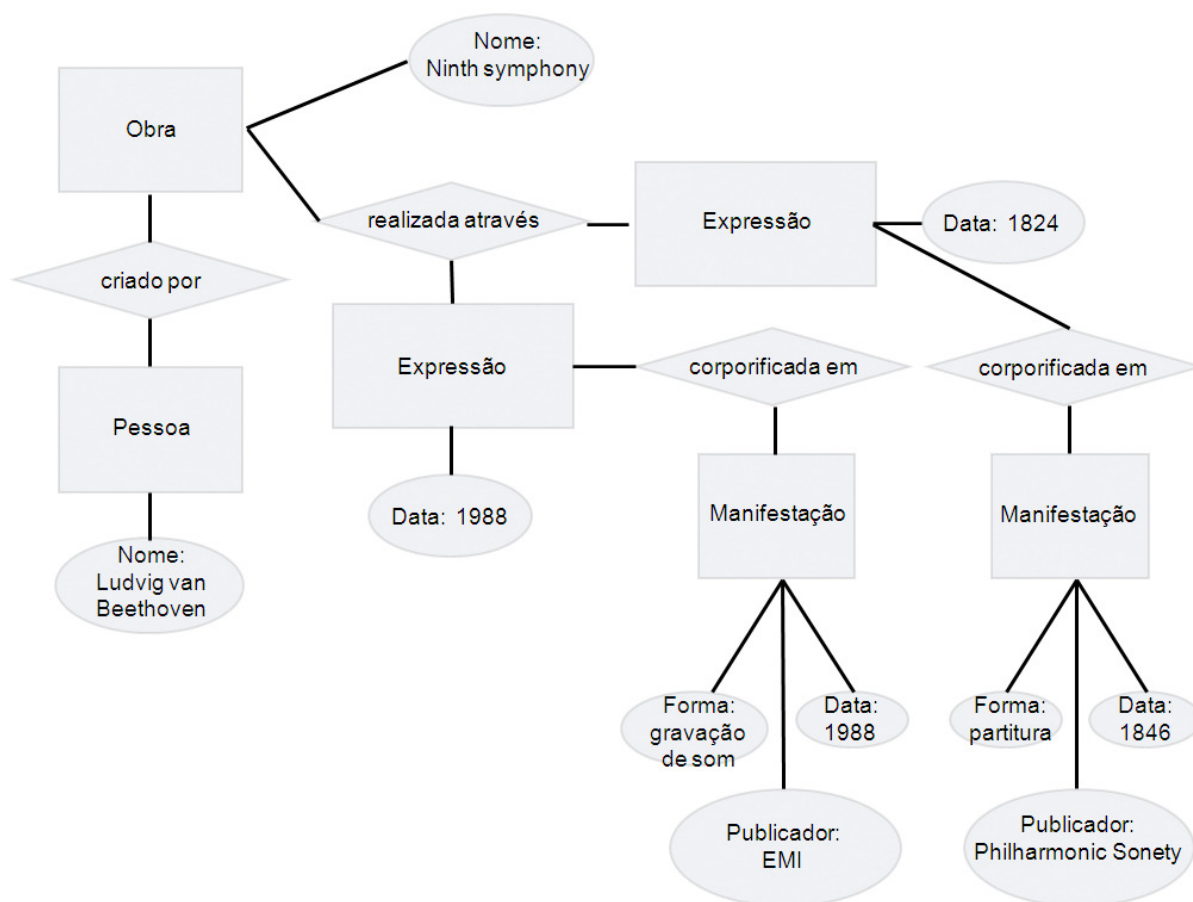


Fonte: Adaptação de Maxwell (2008, p. 75).

A obra conhecida por *Romeo and Juliet*, criada por *William Shakespeare*, foi expressa entre os anos 1591 e 1595. Em 1998 assume, por meio de *manifestação*, a forma de texto digital, quando é publicada no site Domínio Público, pela Etext. O acesso a tal obra é dado pelo endereço [www.dominiopublico.gov.br/download/texto/gu001513.pdf](http://www.dominiopublico.gov.br/download/texto/gu001513.pdf). Esta obra também assume outra *manifestação*, também na forma de texto digital, em 2001, quando foi publicada pelo MIT.edu. Está disponível no endereço [shakespeare.mit.edu/romeo\\_juliet/index.html](http://shakespeare.mit.edu/romeo_juliet/index.html). Por fim, outra *manifestação* desta obra é possível de ser encontrada em um texto impresso, datado de 2007 e publicado pela Barnes & Noble, cuja representação está disponível no site da Library of Congress no endereço [lccn.loc.gov/2006009007](http://lccn.loc.gov/2006009007).

A figura 10 apresenta-se a relação de duas *expressões* com a obra *Ninth symphony*, de Ludvig van Beethoven.

Figura 10 – Representação gráfica das relações entre *obra*, *expressão* e *manifestação* (Ninth symphony).



Fonte: Adaptação de Maxwell (2008, p. 75).

Este diagrama nos apresenta as relações para a *obra Ninth symphony*, criada por Ludvig van Beethoven. Essa *obra* assume duas *expressões*: uma como partitura, datada de 1824; outra como gravação de som, de 1988. A *expressão* em partitura assume *manifestação* pela Philharmonic Sonety, em 1846, cujo *item* está disponível no endereço <http://tinyurl.com/7qj5sxc>. A *expressão* com *manifestação* em gravação de som é publicada pela EMI, em 1988 e sua representação está disponível no endereço [lcn.loc.gov/2008640988](http://lcn.loc.gov/2008640988).

Observe-se as relações entre as representações das informações presentes na Figura 9 e na Figura 10, ambos com destaque para as relações de *arbitrariedade* entre a *obra* e as *manifestações* e do fato de a *manifestação* desenvolver-se no tempo (dada a *linearidade* que um recurso informacional assume).

Tomando-se, ainda, Saussure (2010, p. 27), como solo desta compreensão, reafirma-se, acerca da afirmação “a língua existe na coletividade sob a forma duma soma de sinais depositados em cada cérebro”, que a representação dos recursos

informativos, origina-se, também, no cérebro de cada um. Esse produto cerebral pode ser socializado de várias formas; entretanto, para fins didáticos, serão divididas em duas vertentes: o produto cerebral registrado e o não registrado. Tal produto difere de acordo com as possibilidades de expressões e suas correlatas *manifestações*<sup>13</sup>.

O catalogador, assim sendo, precisa conhecer as *manifestações* para, delas, recorrer aos mecanismos de representação, o que possibilitará dar a conhecer aos outros os recursos informativos. Dessa maneira, embora a *manifestação* seja, por si só, estranha à *obra*, é impossível abstraí-la das várias *manifestações*, processo pela qual a *obra* é constantemente passível de ser representada.

*Obra e manifestação*, dois sistemas distintos, são, porém, complementares para a formação do objeto documental. A *obra* pode ter uma tradição oral, fixada diversamente da tradição escrita em um suporte, e, mesmo assim, ser transferida para outras gerações. Embora essas de tradição oral possam ser objeto das instituições de patrimônio cultural, conquanto não estiverem registradas, não são passíveis de representação. No contexto exposto, observa-se a representação via estrutura semântica na Ciência da Informação.

### 3.2 SEMÂNTICA HJELMSLEVIANA E OS REGISTROS BIBLIOGRÁFICOS

Na área de lingüística, há estudos voltados à semântica estruturalista que, por sua vez, abordam a semântica de forma concreta e analisam a semântica lexical a partir da idéia central de que a linguagem deve ser vista como um sistema (GEERAERTS, 2010; TAMBA-MECZ, 2006). Assim, a linguagem natural pode ser entendida como um sistema simbólico, com propriedades e princípios próprios que determinam como um signo lingüístico funciona. A língua, nesta visão, é estabelecida a partir de convenções, tais quais as práticas sociais transmitidas de geração em geração, e não por deliberações pessoais. Surge a partir da semântica

---

<sup>13</sup> As variadas formas de *manifestação* de uma *obra* podem ser singularizadas em duas: analógica e digital. Essas formas de *manifestação* direcionam todas as outras, a saber: registros em formato analógico: livros, folhetos, folhas impressas, manuscritos, materiais cartográficos, música, materiais gráficos, microformas, recursos contínuos, gravações de som e de imagem. Registros em formato digital: todas essas *manifestações*, além daquelas disponíveis somente em meio eletrônico.

histórico-filológica e tem sua base associada à obra de Ferdinand de Saussure (1916), a qual apresentava, dentre outras, duas abordagens possíveis para o estudo do léxico, a sincrônica e a diacrônica.

A visão de estrutura na lingüística estruturalista evidencia, segundo Hjelmslev (1991, p. 115), sua posição de “*entidade autônoma de dependências internas*”. Pela redução de classes abertas em classes fechadas torna-se possível a descrição estrutural (LOPES, 2008).

Na descrição estrutural do plano de expressão conseguiu-se operar essa redução, concebendo-se os signos como compostos de elementos dos quais um efetivo relativamente baixo bastava para efetuar a descrição. (LOPES, 2008, p. 239).

Recorre-se, também, a Hjelmslev (1991, p. 116) para compreender melhor a relação entre objeto, estrutura e descrição científica, quando o autor afirma, diante da lingüística, que

Não há nem conhecimento nem descrição científica possível de um objeto qualquer sem recurso a um princípio estrutural. [...] Toda descrição científica pressupõe que o objeto da descrição seja concebido como uma estrutura [...] ou como parte de uma estrutura [...].

Diante de tal afirmativa observa-se a adoção de princípios formais em uma relação parte-todo, o que supõe uma afinidade intrínseca à representação documental na Ciência da Informação.

Denota-se um paralelo entre a semântica estruturalista e a Ciência da Informação, tendo em vista que ambas fundamentam-se em constructos simbólicos e com visão sistêmica, senão, veja-se: tal qual a linguagem, a representação também recorre a um sistema, no qual é possível estudar, analisar e representar unidades informacionais, simbólicas, presentes nas mais diferentes *manifestações de obras*. Nesse sentido, a representação documental, baseada em convenções, normas e padrões, iguala-se à língua. A representação permite a análise sincrônica do documento, com traços inerentes ao recurso informacional e seu suporte, no qual estão presentes as relações entre *obra* e *manifestação*. A sincronia, na representação, significa a impossibilidade de dissociação entre representante e representado; entre obra e manifestação.

A partir da concepção estruturalista é possível entender, com Saussure, que o significado de um signo lingüístico deve ser analisado de forma sincrônica com o mundo real e é determinado a partir da sua posição quanto às estruturas lingüísticas

das quais faz parte.<sup>14</sup> Ao priorizar o estudo sincrônico, Saussure marca espaço para o estudo do funcionamento da língua e não como a língua se modifica. Nesse sentido, o estudo sincrônico da representação propõe o estudo dos registros bibliográficos a partir de suas relações, tanto no próprio registro quanto em relação ao objeto descrito. A semântica estruturalista de registros bibliográficos volta-se, assim, ao estudo descritivo do funcionamento dos catálogos.

No estudo estruturalista dos registros bibliográficos percebe-se a intenção de identificar a estrutura do registro, sua relação com outros registros e a relação com o documento. Desse modo, a sintaxe do registro bibliográfico não dá conta do catálogo; é a semântica que permite o contexto e as sinapses entre os vários registros bibliográficos; é a semântica que dá conta dos processos mentais segundo os quais se produz, constitui, compreende e descreve a representação de um recurso informacional. Nesta pesquisa, denomina-se papel semântico o estudo das diferentes relações que podem ser estabelecidas entre os registros bibliográficos, entre o registro bibliográfico e o recurso informacional, e entre os elementos do próprio registro bibliográfico.

Quanto à abordagem da lingüística estrutural, Hjelmslev (1991, p. 29, grifos do autor) define-a como

um conjunto de pesquisas que repousam em uma *hipótese* segundo a qual é cientificamente legítimo descrever a linguagem como sendo *essencialmente* uma *entidade autônoma de dependências internas* ou, numa palavra, uma *estrutura*.

Nesse sentido, considera-se a possibilidade de se estudar a representação documental a partir de uma *teoria dos níveis*, do nível subsemântico (entre os elementos de um registro bibliográfico) ao nível supersemântico (a relação entre os

---

<sup>14</sup> Quanto a este ponto Geeraerts (2010, p. 48) apresenta a semelhança entre o jogo de xadrez e a lingüística, apresentado, anteriormente, por Saussure (2010, p. 31-32). O autor afirma que descrever as regras do jogo de xadrez é uma forma adequada e suficiente para descrevê-lo, sem que fatores extrínsecos ao sistema de regras em si sejam necessários para explicar o seu funcionamento. Da mesma forma, continua o autor, a lingüística deve principalmente descrever a linguagem natural como um sistema simbólico, sem que seja necessário recorrer a fatores que estão fora deste sistema simbólico. Como tal, a lingüística em si pode ser considerada uma disciplina autônoma: não requer metodologia de outras disciplinas. Além disso, a imagem do jogo de xadrez ilustra como os signos da língua podem ser estudados: o valor de uma peça separada em um jogo de xadrez só pode ser definida com respeito ao corpo de regras como um todo. (Tradução livre em texto sinóptico).



vários registros bibliográficos, a partir de suas semelhanças e diferenças), passando pelo nível semântico (objeto que está sendo descrito e a descrição em si).<sup>15</sup>

O objetivo da ênfase estruturalista para a análise semântica dos registros bibliográficos pode ser definido como: estudar a descrição de registros bibliográficos efetivamente realizados, no qual se considera a influência do catálogo, como meio para o estabelecimento de mensagens presentes nos recursos informacionais e nas necessidades informacionais dos usuários.

A semântica estruturalista de registros bibliográficos envolve três posições teóricas que podem ser distinguidas da seguinte forma:

1. a teoria das unidades semânticas, na qual se estuda a relação entre as várias unidades que compõem um registro (subsemântica);
2. a análise componencial da representação bibliográfica, na qual se estuda a relação entre um recurso informacional e o seu registro bibliográfico (semântica);
3. a semântica relacional, na qual se estuda a relação entre um registro bibliográfico e outros em um catálogo (supersemântica).

Nesse sentido, é possível admitir três planos de diferença semântica em registros bibliográficos: entre o referente e a representação, entre o todo e as partes da representação e, entre as representações presentes no catálogo. São estes aspectos semânticos que reduzem a alteridade de um registro bibliográfico<sup>16</sup> que, em um catálogo, faz com que o disperso e o aparente sejam marcados por identidades que lhe são próprias.

A teoria das unidades semânticas volta-se ao estudo de um conjunto de itens lexicais de um registro bibliográfico como sendo semanticamente relacionados, cujos significados são mutuamente interdependentes, e que juntos fornecem a estrutura conceitual de um determinado registro bibliográfico.

A análise componencial da representação bibliográfica, por sua vez, referencia a afinidade entre a representação e o documento por meio da descrição detalhada, uma vez que seus componentes são representativos de um item

---

<sup>15</sup> Faz-se alusão aos termos sistema, subsistema e supersistema adotados em Teoria dos Sistemas para a definição dos níveis semânticos em registros bibliográficos e adota-se subsemântica, semântica e supersemântica em registros bibliográficos.

<sup>16</sup> Entende-se que a alteridade de registros bibliográficos refira-se à possibilidade de um registro bibliográfico ser de outro referente que não o dele mesmo.



documental. Isso significa que, conhecido um documento, as relações entre os seus elementos e a representação deverão ser fornecidos em detalhes, não bastando elencar os itens e dizer que estão em mútua oposição. A descrição deve indicar elementos que identifiquem e definam o documento, por meio das unidades semânticas, ou seja, dos componentes representativos de um item documental.

A semântica relacional pode ser entendida como um desenvolvimento lógico entre as relações que podem ser estabelecidas entre os vários registros em um catálogo, no qual se explicitam as relações internas entre os elementos de um registro, sendo possível identificar oposições e paridades. Recorre, portanto, ao aparato de descrição nas relações entre as unidades semânticas, por meio da identidade de significados (semelhanças) e da oposição de significados (diferenças). Desenvolve a idéia de descrição das relações estruturais entre os vários registros bibliográficos, restritos aos elementos usados tanto na descrição quanto no estabelecimento dos pontos de acesso.

Essa concepção de teoria semântica volta-se ao tratamento do sentido da representação a partir das unidades de significação (cada um dos elementos descritos) que, por sua vez, ligam-se ao objeto documental e permitem correlacionar elementos em um catálogo, ou seja, preocupa-se com as relações entre as *obras* e as *manifestações* no qual se analisa o conteúdo de um item documental. Por apresentar-se como um estudo sistêmico, com embasamento estrutural e sincrônico, como *ciência das significações representacionais*, busca, nas regras que determinam a função sintática, as relações semânticas que, por sua vez, contam com uma organização própria.

Tais conteúdos semânticos podem ser analisados a partir de semelhanças e diferenças específicas, conceitos gerais, integridade na descrição dos elementos e seu significado, onde, cada um deles assume valores, tais como as características distintivas elencadas na *manifestação*: indicação de responsabilidade, título, outras informações sobre o título, indicação de edição, local de publicação, nome do editor, data, entre outros.

Esse processo de decomposição de um registro bibliográfico em componentes (unidades semânticas) faz uso de traços semânticos, demarcados por aspectos sintáticos. Assim, na representação, fatores externos à *manifestação*, inerentes, por conseguinte, à *obra*, interferem em sua estrutura e em seu estudo.

Deste modo, o valor de cada elemento semântico diz respeito a uma estrutura maior, identificada como sistema, e assume determinada função, o que estabelece seu significado em dado contexto. A representação para ser parte de um sistema deve assumir características da *obra* e da *manifestação*, ou, adotando-se expressões da Linguística, do *significado* e do *significante*.

### 3.3 PRINCÍPIOS SINTÁTICOS E SEMÂNTICOS DE REGISTROS BIBLIOGRÁFICOS: A REPRESENTAÇÃO COMO ESTRUTURA SEMÂNTICA

Tal qual *significado* e *significante* são termos que levam à dimensão semântica da semiótica e da semiologia, na área da Linguística, *obra* e *manifestação* levam à dimensão da semântica da representação na Ciência da Informação. A semântica na Ciência da Informação é dada pela forma da e na representação da informação.

A semântica, pensada por este prisma, refere-se à estrutura de um sistema que relaciona *significado* e *significante*, *obra* e *manifestação*. Entenda-se que na semântica da representação dos recursos informacionais, ou na semântica dos registros bibliográficos, o *significado* é dado pelo valor do *significante*, ou, a *manifestação* é o valor da *obra* no processo de representação. Esses valores semânticos em um registro bibliográfico formam uma rede de relações estruturais com outros registros bibliográficos, a qual se denomina supersemântica do registro bibliográfico.

A semântica requer a adoção de sintaxe para a definição dos valores semânticos; em outras palavras, a *sintaxe* está presente no *padrão de estrutura de metadados descritivos* e a *semântica* nos *padrões de conteúdo*.

Sintaxe, em seu sentido etimológico, volta-se ao arranjo, à disposição, à organização, à ordenação (HOUAISS, 2008). No âmbito desta pesquisa, a sintaxe do registro bibliográfico diz respeito à ordem dos elementos dispostos para a representação dos recursos informacionais. Entende-se, portanto, que a sintaxe do registro bibliográfico seja parte da Ciência da Informação voltada ao estudo da forma, arranjo e disposição em que cada elemento deva ser descrito quando da representação do recurso informacional. Neste sentido, faz parte do sistema biblioteconômico que determina relações formais entre a representação de cada

uma das partes do documento representado. Esses elementos são organizados segundo padrões de estrutura de metadados estabelecidos, no caso a ISBD. Vejam-se alguns exemplos:

- **Shakespeare, William** – é uma forma de representação dos elementos de responsabilidade intelectual de uma *obra*, representado como ponto de acesso ao documento (*manifestação*); tal sintaxe é dada pela Regras 21.1, 22.1, 22.4 e 22.5 das AACR2r;

- **Romeu e Julieta / William Shakespeare** – é a forma de representação do elemento título e da responsabilidade de uma *obra*, em determinada *manifestação*, dada pelas Regras 1.1B, 1.1E, 2.1A1, 2.1B, 2.1E das AACR2r;

- . – **3. ed.** – é a forma de representação do elemento indicativo de edição da *manifestação* de uma *obra*, dada pelas Regras 1.2A1, 1.2B1, 2.2A1, 2.2.B1, nota da Regra C.8A das AACR2r;

- . – **São Paulo : Martin Claret, 2002** – são formas de representação dos elementos local de publicação, nome do editor e data de publicação da *manifestação* de uma *obra*, dadas pelas Regras 1.4A1, 1.4B, 1.4C, 1.4D, 1.4F, 2.4A1, 2.4B, 2.4C, 2.4D e 2.4F das AACR2r;

- . – **160 p.** – é a forma de representação do elemento extensão de um *item* de uma *manifestação* de uma *obra*, dada pelas Regras 1.5A1, 1.5B1, 1.5B2, 2.5A1, 2.5B1, 2.5B2 das AACR2r;

- . – **ISBN 8572325271. – ISBN 9798572325270 (ISBN-13)** – é a forma de representação do elemento número normalizado da *manifestação* de uma *obra*, dada pelas Regras 1.8A1, 1.8B1, 1.8B2, 2.8A1, 2.8B1, das AACR2r.

Os aspectos sintáticos de um registro bibliográfico, conforme podem ser observados acima, remetem à estrutura semântica. A frase . – **São Paulo : Martin Claret, 2002**, marcada pela pontuação definida pela ISBD, também presente nas AACR2r, especificamente quanto ao uso do ‘. – ’ (ponto espaço travessão espaço), dos ‘ : ’ (dois pontos), da ‘ , ’ (vírgula), permite que se faça a análise do conteúdo semântico, indicado pelas AACR2r, especificamente: o ‘ponto espaço travessão espaço’ deve anteceder o início da área de publicação, distribuição etc., a qual indica que o seu primeiro elemento seja o local de publicação; os ‘dois pontos’ marcam que a informação seguinte é o nome do editor; e, por sua vez, a ‘vírgula’ antecede a indicação do ano de publicação.

Semântica, por sua acepção etimológica, apresenta aspectos da significação. Na lingüística, pode ser entendida como “estudo da significação como parte dos sistemas das línguas naturais”; “componente do sentido das palavras e da interpretação das sentenças e dos enunciados”; “significado das palavras, por oposição à sua forma” (HOUAISS, 2008). Na filosofia, trata da “teoria abstrata da significação ou da relação entre os signos e seus referentes (em oposição à sintaxe e à pragmática)” (HOUAISS, 2008).

No contexto desta investigação a semântica de registros bibliográficos relaciona-se tanto ao [1] significado estabelecido entre cada um dos elementos representados e o documento representado, propondo-se um conteúdo semântico do registro bibliográfico, identificado como a semântica do registro bibliográfico em si, quanto ao [2] significado entre o elemento definidor e o valor dado para expressar este elemento, apresentado como subsemântica, no qual se estudam as unidades semânticas, e quanto ao [3] significado que pode ser estabelecido entre os vários registros de um catálogo, identificado como supersemântica. Em outras palavras: o registro bibliográfico de um documento só pode assumir o significado de representação daquele documento dado o sentido que pode ser estabelecido entre *representado* e *representante*, ou, entre *obra* e *manifestação*; pelo significado entre o elemento a ser representado e o seu valor (dado, por sua vez, pela sintaxe), e uma metasemântica, no sentido das relações de metalinguagem que podem ser estabelecidas entre diferentes registros.

Se não, vejam-se:

- **Romeu e Julieta / William Shakespeare** – estes elementos assumem, no contexto de um registro bibliográfico, a representação que dá o significado de título e de responsabilidade da *obra* presente na *manifestação*;

- . – **São Paulo : Martin Claret, 2002** – estes elementos, por sua vez, assumem, no contexto de um registro bibliográfico, a representação que dá o significado de elementos como local de publicação, nome do editor e data de publicação, dados da *manifestação* de uma *obra*.

Nestes dois casos, explicita-se: [1] a semântica entre os elementos de uma *obra* (título, responsabilidade, local de publicação, nome do editor e data de publicação) e o seu valor (o seu conteúdo designativo e representativo da *obra*); e [2] a semântica em um registro bibliográfico, na qual, a representação bibliográfica,

para que assuma a sua vertente representacional e significativa de um representado, apresente relações entre os valores dos elementos. Ou seja, no exemplo acima, a *obra* intitulada **Romeu e Julieta** apresenta relação intrínseca com a sua *manifestação*, dada pelos elementos da publicação, tais como, **São Paulo: Martin Claret, 2002**.

Em suma, as relações semânticas de um registro bibliográfico existem entre cada um dos elementos e o seu valor no documento que está sendo representado (semântica do registro bibliográfico); é estabelecida, também, uma relação semântica entre as várias unidades de um registro bibliográfico (subsemântica do registro bibliográfico); e uma outra relação semântica entre os vários registros bibliográficos que referem-se às várias manifestações de uma obra (supersemântica do registro bibliográfico).

Nesse sentido, a linguagem bibliográfica ultrapassa o nível sintático e faz com que se entenda que um registro apresente níveis semânticos, necessário ao entendimento do documento representado sintática e semanticamente. Desse modo, cada elemento sintático assume um conteúdo semântico diante de cada elemento definidor da representação, e este elemento, por sua vez, quando contextualizado e contraposto ao documento representado, assume significado entre o registro e o objeto.

O registro bibliográfico cinge, então, tanto questões sintáticas, por referir-se aos padrões de estruturas dos metadados de cada elemento do documento ou objeto a ser descrito, quanto questões semânticas, por permitir analisar a coesão e o significado indicado entre elementos do representante e do representado e entre o próprio representante e representado. Cada elemento sintático, quando contextualizado e contraposto ao documento representado, assume um significado concreto entre o registro e o objeto.

Analisa-se, de modo sintético, o processo sintático e semântico em registros bibliográficos na representação em meios automatizados, segundo o Formato MARC21 Bibliográfico:

Quadro 2 – Representação dos recursos informacionais em meio automatizado

TAG	Indicador 1	Indicador 2	Subcampo	Valor
020	Ø	Ø	a	8572325271
020	Ø	Ø	a	9798572325270
100	1	Ø	a	Shakespeare, William
245	1	0	a	Romeu e Julieta
			c	William Shakespeare
250	Ø	Ø	a	3. ed.
260	Ø	Ø	a	São Paulo
			b	Martin Claret
			c	2002
300	Ø	Ø	a	160 p.

Fonte: Elaborado pelo próprio autor.

A sintaxe em cada valor da TAG<sup>17</sup> (rótulo) do Formato MARC21 Bibliográfico é dada pela ISBD, presente, também, nas AACR2r. A relação semântica é dada em três momentos<sup>18</sup>: o primeiro (subsemântico), entre a TAG, os indicadores 1 e 2, os delimitadores e os códigos de subcampo (dados pelo Formato MARC21 Bibliográfico) e o valor do campo; o segundo (semântico), entre os campos do registro bibliográfico (representante) em Formato MARC21 Bibliográfico e o representado; o terceiro (supersemântico) entre essa *manifestação* e outras possíveis de existirem no acervo de uma instituição de patrimônio cultural, como, por exemplo, a *manifestação* em meio eletrônico.

Confirma-se a premissa de que o registro bibliográfico atinge tanto as questões sintáticas, por referir-se às estruturas ou padrões para expressar cada elemento do documento ou objeto a ser descrito, quanto às questões semânticas, por debruçar-se sobre a coesão e o significado entre o representante e o representado.

Além de se representar a *obra*, via *manifestações*, deve-se também disseminá-la, para que a *obra* faça parte do processo comunicativo.

Como se pode entender a conversão, a partir da análise sintática e semântica dos registros bibliográficos, a ponto de ser possível conduzir processos de conversão de registros bibliográficos ao Formato MARC21 Bibliográfico? Entende-se que ocorra pela utilização das marcações dadas pelo padrão de estrutura de metadados descritivos, presentes nas AACR2r e nas ISBDs, e, conseqüentemente,

<sup>17</sup> TAG é um código de três dígitos que identifica o campo – o tipo de dado – que segue no registro bibliográfico.

<sup>18</sup> Recorre-se aos FRBR, aos FRAD e às AACR2r para que seja possível o estabelecimento dessas relações.

pela inferência semântica, prevista em uma aplicação computacional.<sup>19</sup> Fez-se tal opção por julgar que a metodologia desenvolvida nessa pesquisa vá ao encontro do pressuposto de que, para o armazenamento e acesso aos recursos informacionais, se faz necessária à sua representação. Nesse ponto, concorda-se com o exposto por Fernandes (2005, p. 8) ao afirmar que, para a representação do conhecimento, é necessária a adoção “de convenções sintáticas e semânticas que tornam possível descrever coisas”.

Tendo sido discutidos os aspectos referenciais da pesquisa (no tópico 2), e a concepção teórica da tese (no tópico 3), parte-se para a apresentação do Scan for MARC, o interpretador de registros bibliográficos analógicos para o Formato MARC21 Bibliográfico, proposto para o estudo da aplicação de aspectos teórico-conceituais sintáticos e semânticos na conversão de registros bibliográficos.

---

<sup>19</sup> Do mesmo modo que se faz este questionamento referente à conversão de registros bibliográficos, ocorre-nos remeter ao paradigma identificado por Saussure: “como podem os falantes continuar usando uma língua de maneira eficaz, quando essa língua está constantemente mudando?” A partir deste modelo Saussure identifica que a língua é um “sistema altamente estruturado, onde cada elemento se define em grande medida pela maneira como está relacionado com outros elementos.” (TRASK, 2008, p. 222).

#### 4 SCAN FOR MARC: INTERPRETADOR DE PRINCÍPIOS SINTÁTICOS E SEMÂNTICOS DE REGISTROS BIBLIOGRÁFICOS

Neste momento, vislumbra-se aplicar aspectos teórico-conceituais dos princípios sintáticos e semânticos de registros bibliográficos à conversão de registros bibliográficos ao Formato MARC21 Bibliográfico em um interpretador computacional.

Para tanto, parte-se do pressuposto teórico estabelecido, de se considerar os esquemas sintáticos e semânticos dos registros bibliográficos e não a estrutura da base de dados legada para que seja possível realizar a conversão de registros bibliográficos ao Formato MARC21 Bibliográfico.

Retome-se alguns pontos teóricos a fim de ser possível delinear o comportamento do interpretador computacional sintático e semântico que se busca desenvolver:

- a) a sintaxe em registros bibliográficos é identificada pela pontuação, indicada em padrões de estrutura de metadados descritivos, neste caso, a ISBD e as regras de pontuação das AACR2; estuda a forma, o arranjo e a disposição de cada elemento descrito na representação do recurso informacional e, com base em Saussure, compreende-se que a sintaxe estabeleça as relações formais e síncronas entre significado, significante, obra e manifestação;
- b) a semântica em registros bibliográficos, estabelecida a partir da teoria dos níveis do estudo estruturalista dos registros bibliográficos, com base em Hjelmslev, ocorre em três momentos:
  - a. subsemântica – estudo da teoria das unidades semânticas, ou dos elementos semanticamente relacionados presentes em um registro bibliográfico; significado estabelecido entre cada um dos elementos de um registro e os valores de representação que assume;
  - b. semântica – descreve as relações estruturais estabelecidas entre o registro bibliográfico e o recurso informacional;
  - c. supersemântica – relação semântica entre os vários registros bibliográficos representados em um catálogo, por meio de



identidade de significados (semelhanças) e de oposição de significados (diferenças).

O interpretador computacional sintático e semântico de registros bibliográficos será identificado pelo nome de *Scan for MARC*, por aludir ao método de escaneamento dos registros bibliográficos e sua consequente conversão ao Formato MARC21 Bibliográfico.

O interpretador, em sua versão *beta*<sup>20</sup>, envolve, sumariamente, as seguintes ações:

- a) captura de imagens de registros bibliográficos analógicos, apresentados em fichas catalográficas;
- b) análise de resultados de processamentos de imagem em *software* de reconhecimento de caracteres, o que permite a conversão de imagem para texto editável;<sup>21</sup>
- c) processamento das imagens de registros bibliográficos selecionados para teste (processo inicial da construção da representação);
- d) tratamento sintático e semântico dos registros bibliográficos digitais, checagem dos resultados e avaliação de ajustes.

Tais ações consubstanciam procedimentos de reprodução (itens *a* e *b*) e de representação (itens *c* e *d*). Apresentam-se, a seguir os resultados dos testes e ensaios de cada uma destas ações.

#### 4.1 CAPTURA DE IMAGENS DE REGISTROS BIBLIOGRÁFICOS

A fase de testes para a captura de imagens dos registros bibliográficos dividiu-se em três etapas: em *scanner* de impressora funcional, em câmera de dispositivo móvel e em câmera fotográfica digital. Comentários acerca da melhor resolução quanto ao escaneamento não serão apresentadas neste tópico, pois não

---

<sup>20</sup> A *versão beta* de um produto pressupõe o fato de ainda estar em fase de desenvolvimento e de testes.

<sup>21</sup> O processo de reconhecimento ótico de caracteres também pode ser compreendido como uma representação de baixo nível. Entretanto, nesta pesquisa, os procedimentos de captura de imagens e de reconhecimento ótico de caracteres serão compreendidos como reprodução de registros bibliográficos, com destaque à mudança do suporte no qual estão registrados, em meio analógico e digital, haja vista que não há alteração no padrão de estrutura de metadados descritivos.

é seu objetivo. Estes serão apresentados em tópico posterior, por conta da verificação do fato de a definição das imagens apresentar, ou não, interferência no reconhecimento de caracteres.

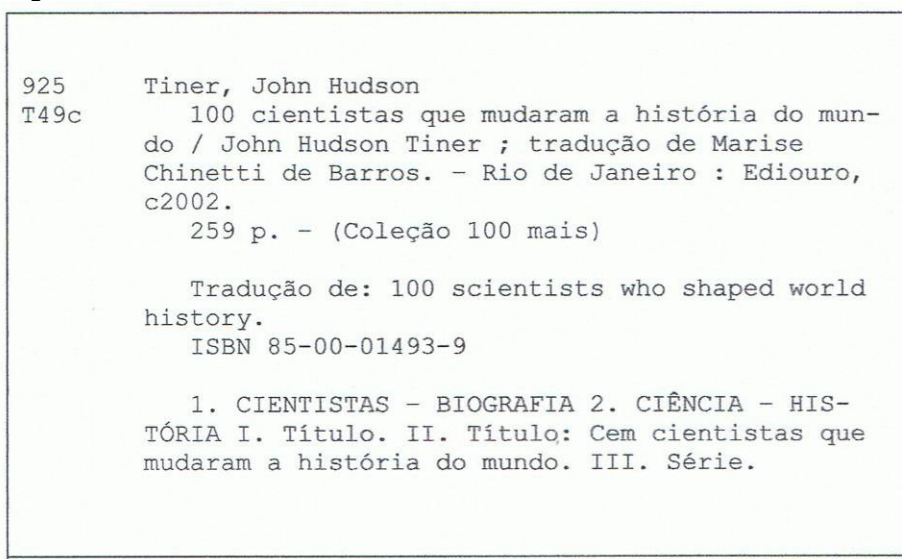
#### 4.1.1 Captura de imagens de registros bibliográficos em scanner de impressora multifuncional

Os testes de escaneamento de registros bibliográficos em scanner de impressora funcional foram realizados em equipamento com recurso de *scanner* de mesa, que apresentava a seguinte configuração (BROTHER, 2004):

- intensidade de cor: processamento de 36 bits para entrada e de 24 parabits para saída;
- resolução: até 19200 x 19200 ppp (interpolado) e até 600 x 2400 ppp (óptico);
- velocidade de digitalização: em cores, até 5.93 segundos; em preto, até 3.76 segundos (tamanho A4 em 100 x 100 ppp);
- escala de cinza: 256 níveis.

Nos testes de escaneamento do registro bibliográfico em scanner de impressora funcional foram obtidos os resultados constantes da Figura 11:

Figura 11 – Imagens de registros bibliográficos obtidos em impressora multifuncional  
Registro 1



## Registro 2

302.222 Weil, Pierre  
W439c O corpo fala : a linguagem silenciosa da comunicação não-verbal / Pierre Weil, Roland Tompakow. - 55. ed. - Petrópolis : Vozes, 2002. 288 p.

ISBN 85-326-0208-8

1. COMUNICAÇÃO NÃO-VERBAL I. Tompakow, Roland. II. Título.

## Registro 3

401.41 Discurso e texto : multiplicidade de sentidos na  
D639 ciência da informação / Nádea Regina Gaspar, Lucília Maria Sousa Romão (organizadoras). - São Carlos : EdUFSCar, 2008. 259 p.

ISBN 978-85-7600-137-9

1. ANÁLISE DO DISCURSO 2. CIÊNCIA DA INFORMAÇÃO I. Gaspar, Nádea Regina, org. II. Romão, Lucília Maria Sousa, org.

Fonte: Elaborado pelo próprio autor.

Entende-se que os resultados neste equipamento tenham sido positivos, consideradas questões de luminosidade e qualidade da imagem.

#### 4.1.2 Captura de imagens de registros bibliográficos em câmera de dispositivo móvel

Os testes de escaneamento de registros bibliográficos em câmera de dispositivo móvel foram realizados em um iPhone 4, com as seguintes configurações (APPLE, 2011):

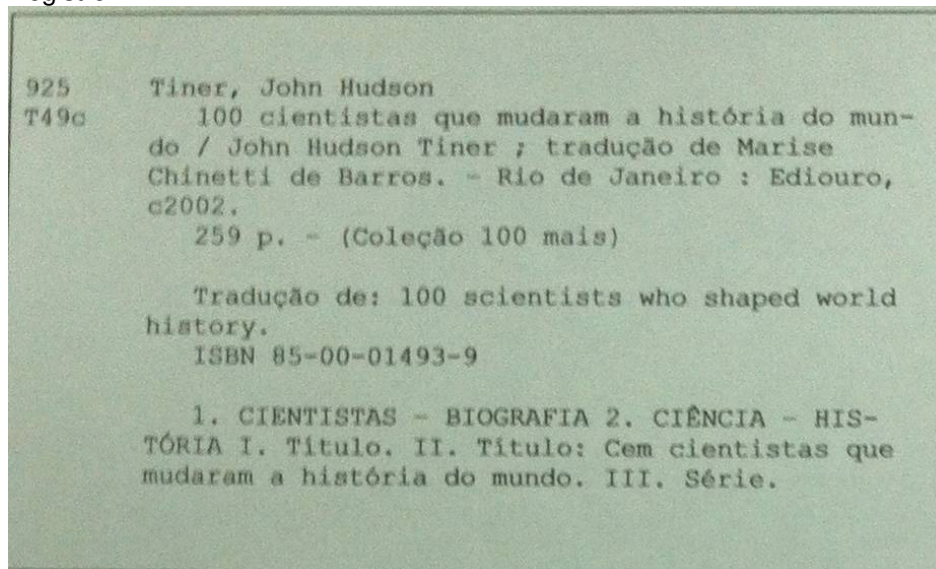
- câmera com 5 megapixels;
- flash de LED integrado;
- zoom 5x.

Foram realizados três testes: com imagens capturadas em distância aproximada de 30, 20 e 10 cm.

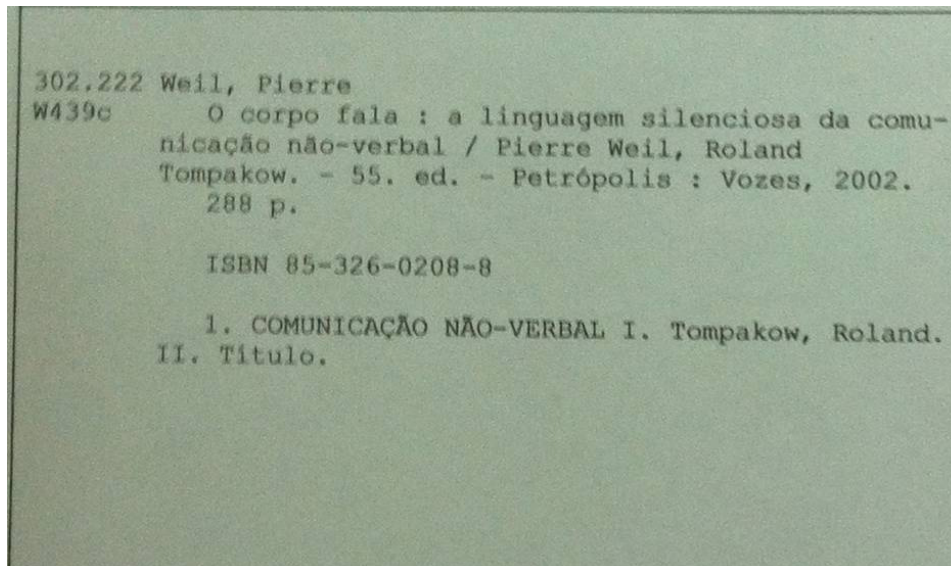
O teste, com imagem capturada a uma distância aproximada de 30 cm, resultou nas imagens constantes da Figura 12.

Figura 12 – Imagens de registros bibliográficos capturadas a partir de câmera de dispositivo móvel com distância aproximada de 30 cm

Registro 1

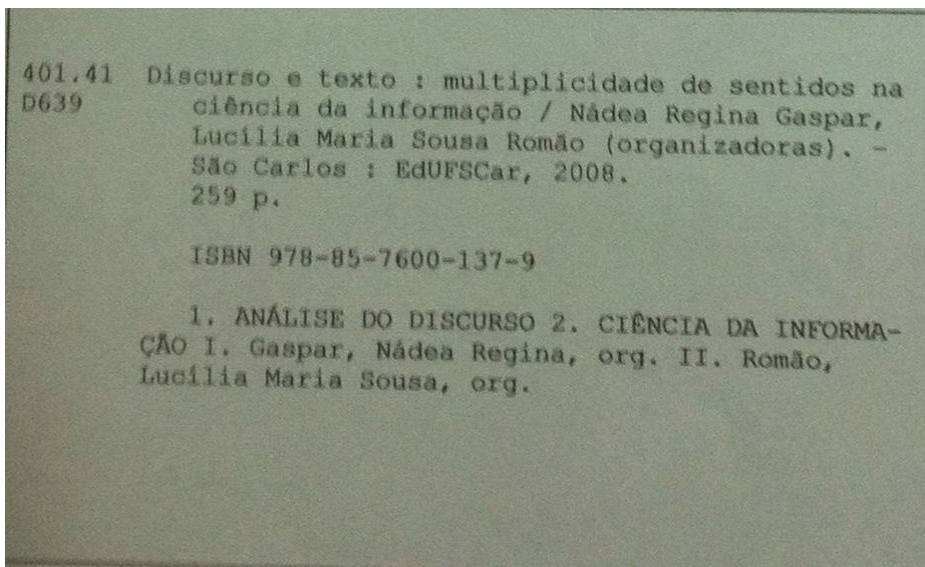


Registro 2





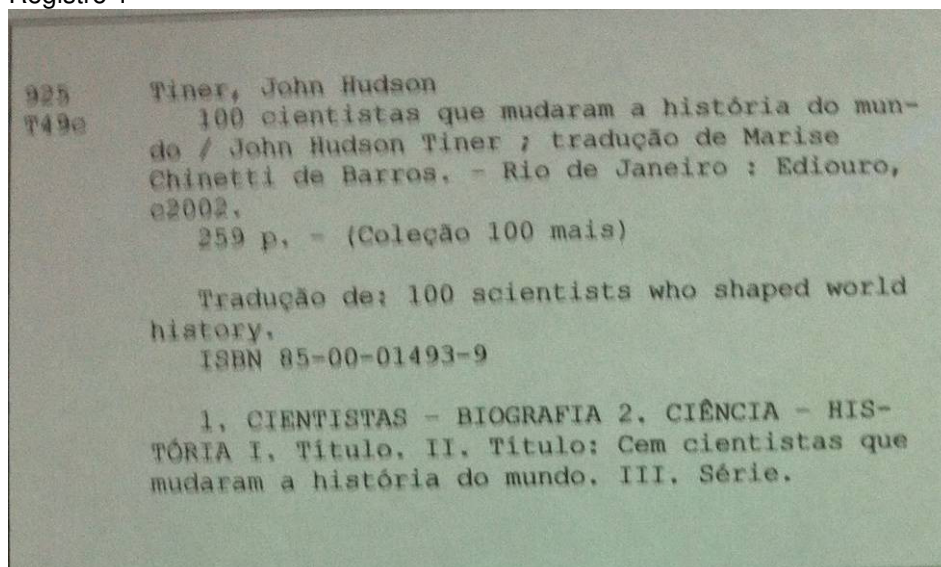
## Registro 3



Fonte: Elaborado pelo próprio autor.

O teste, com imagem capturada a uma distância aproximada de 20 cm, resultou nas imagens constantes da Figura 13.

Figura 13 – Imagens de registros bibliográficos capturadas a partir de câmera de dispositivo móvel com distância aproximada de 20 cm



## Registro 2

302.222 Weil, Pierre  
W439c O corpo fala : a linguagem silenciosa da comunicação não-verbal / Pierre Weil, Roland Tompakow. - 55. ed. - Petrópolis : Vozes, 2002.  
288 p.  
  
ISBN 85-326-0208-8  
  
1. COMUNICAÇÃO NÃO-VERBAL I. Tompakow, Roland.  
II. Título.

## Registro 3

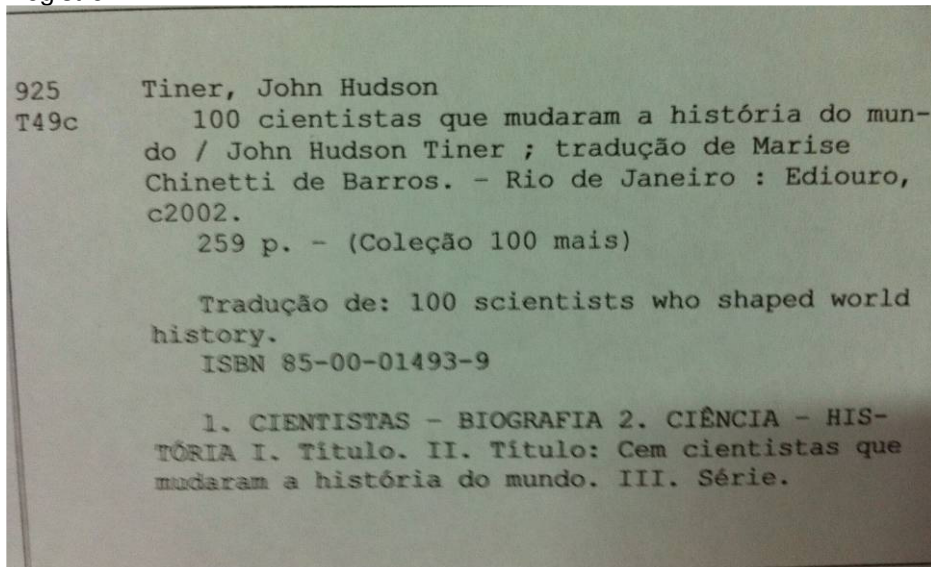
401.41 Discurso e texto : multiplicidade de sentidos na  
D639 ciência da informação / Nádea Regina Gaspar,  
Lucília Maria Sousa Romão (organizadoras). -  
São Carlos : EdUFSCar, 2008.  
259 p.  
  
ISBN 978-85-7600-137-9  
  
1. ANÁLISE DO DISCURSO 2. CIÊNCIA DA INFORMAÇÃO I. Gaspar, Nádea Regina, org. II. Romão, Lucília Maria Sousa, org.

Fonte: Elaborado pelo próprio autor.

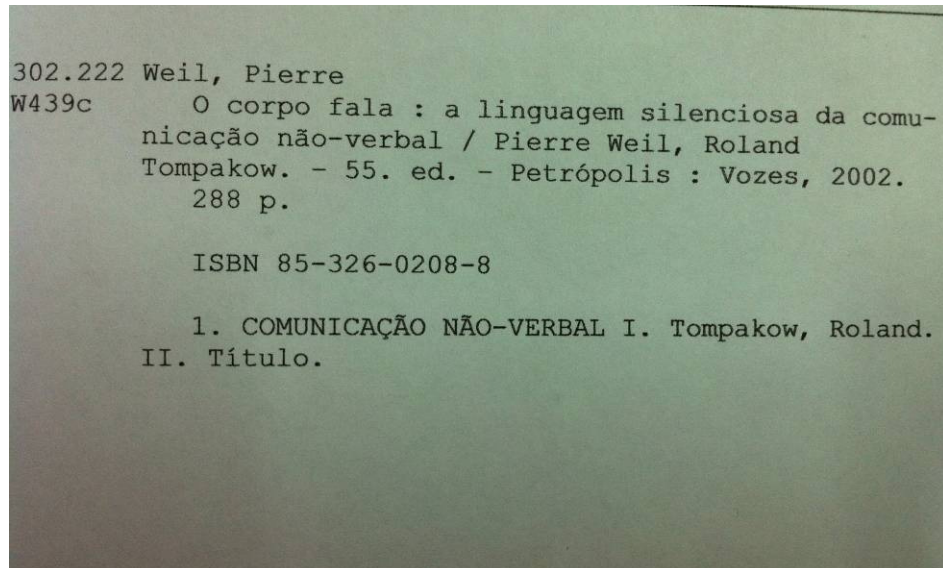
O teste, com imagem capturada a uma distância aproximada de 10 cm, resultou nas imagens constantes da Figura 14.

Figura 14 – Imagens de registros bibliográficos capturadas a partir de câmera de dispositivo móvel com distância aproximada de 10 cm

Registro 1

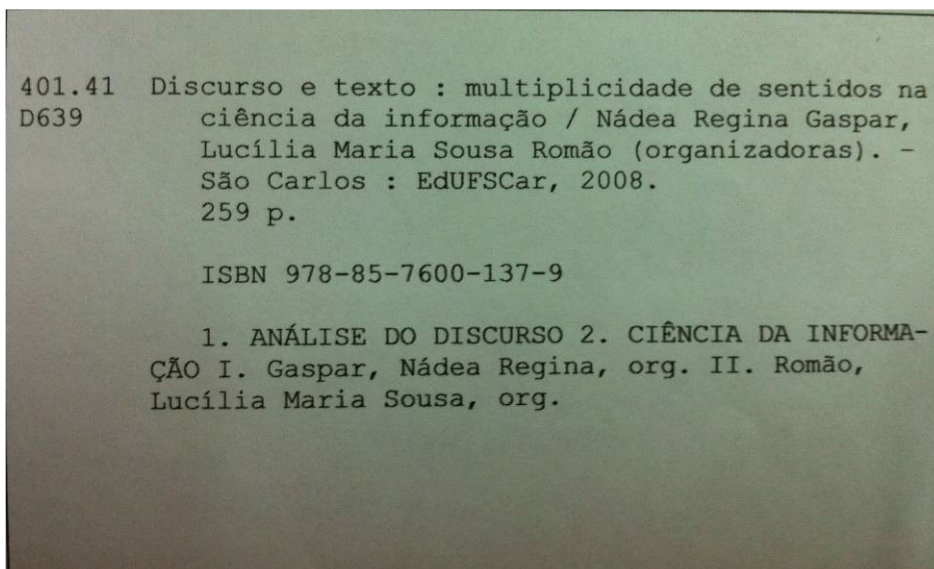


Registro 2





## Registro 3



Fonte: Elaborado pelo próprio autor.

Entende-se que os resultados neste equipamento tenham sido razoáveis, uma vez que, consideradas questões de luminosidade, qualidade da imagem e distância, estes se apresentaram escuros e com baixa definição dos caracteres.

#### 4.1.3 Captura de imagens de registros bibliográficos em câmera fotográfica digital

Os testes de escaneamento de registros bibliográficos em câmera fotográfica digital foram realizados em um equipamento Sony Cyber-Shot, Modelo DSC-W210, a qual apresentava a seguinte configuração (SONY, c2011):

- pixels efetivos: aproximadamente 12,1 megapixels;
- processador Bionz;
- conversão A/D (DXP): 14 bits;
- nitidez RAW NR;
- focagem pré-programada (em m): 0,5 / 1 / 3 / 7 / distância ilimitada;
- balanço de brancos: automático, luz do dia, nublado, fluorescente, incandescente, flash;
- sensibilidade ISO: automático, 100, 200, 400, 800, 1600, 3200);
- sistema de focagem automática;
- zoom óptico: 4x;
- zoom de precisão digital: aproximadamente 8x (total);
- zoom inteligente: até 25x (com VGA);

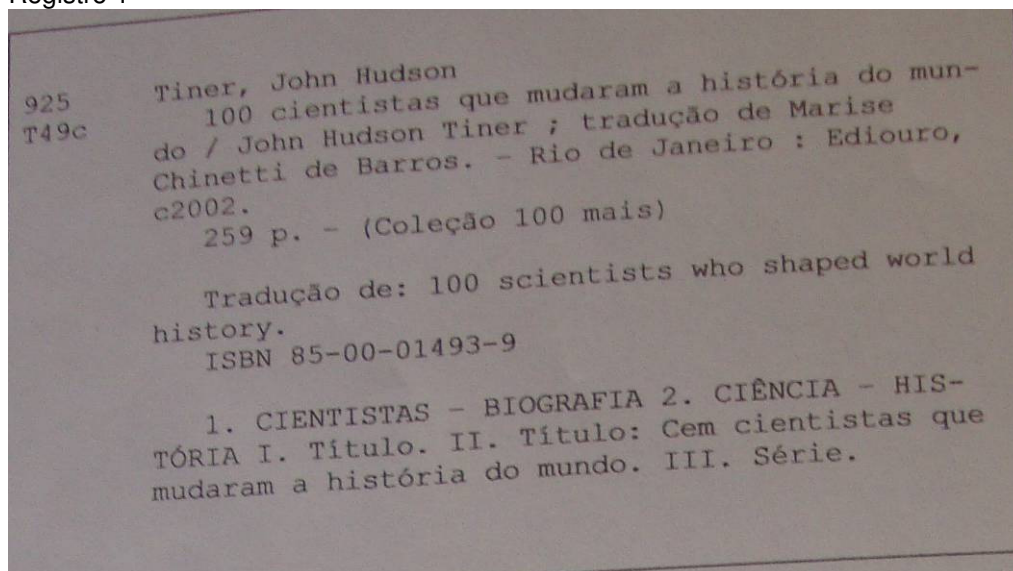


- sensor de imagem por CCD super HAD.

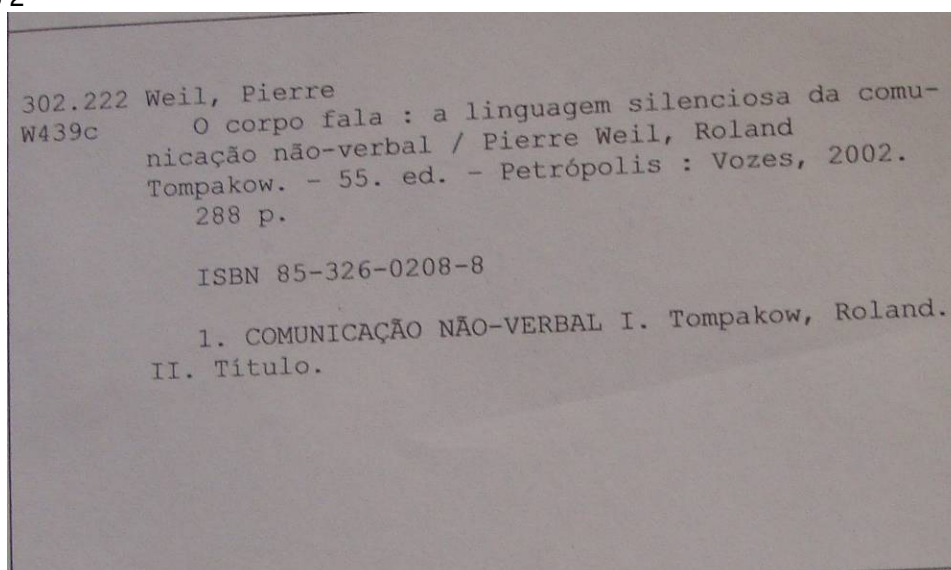
O teste, com imagem capturada a uma distância de aproximadamente 30 cm, resultou nas imagens constantes da Figura 15.

Figura 15 – Imagens de registros bibliográficos capturadas a partir de câmera fotográfica digital com distância aproximada de 30 cm

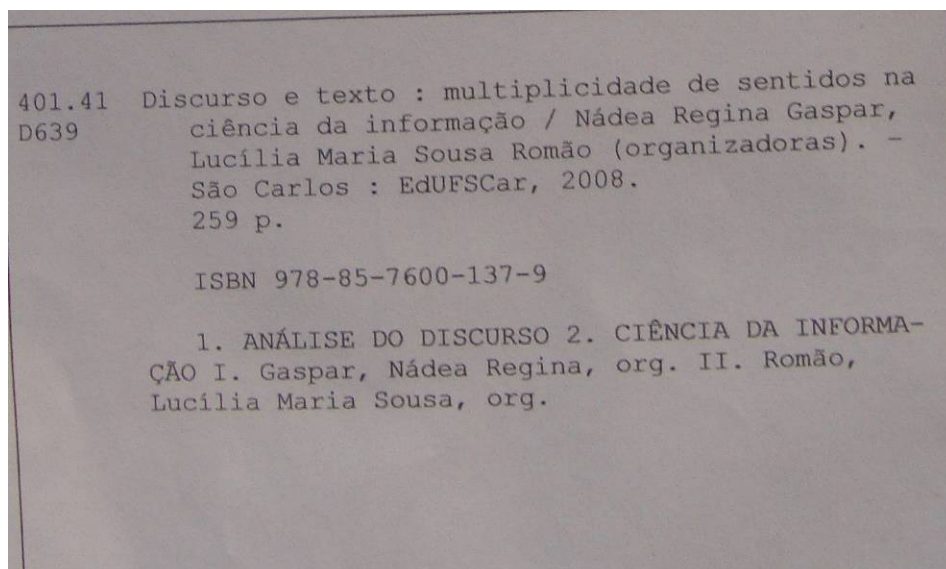
Registro 1



Registro 2



## Registro 3

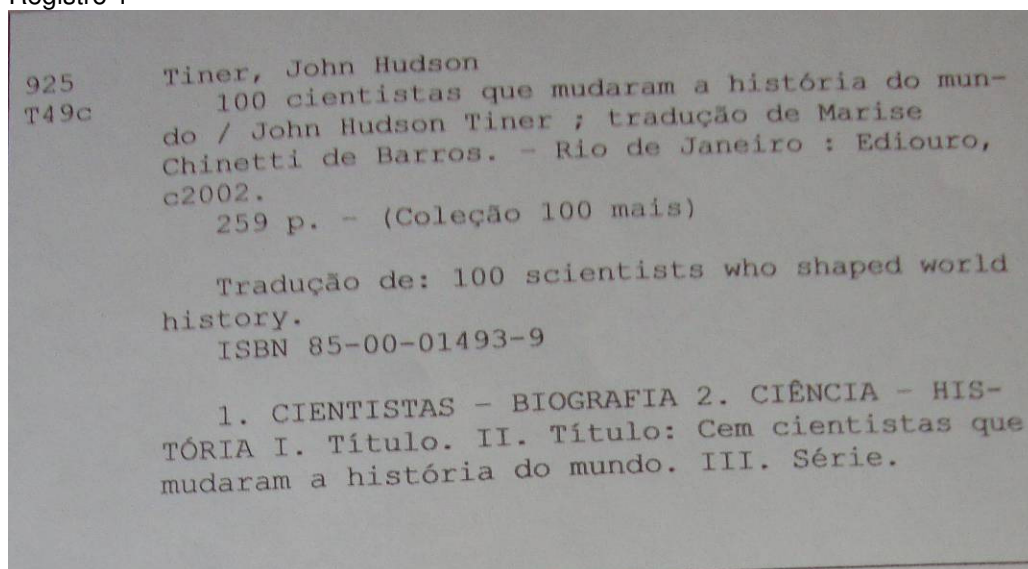


Fonte: Elaborado pelo próprio autor.

O teste, com imagem obtida a uma distância de aproximadamente 20 cm, resultou nas imagens constantes da Figura 16.

Figura 16 – Imagens de registros bibliográficos capturadas a partir de câmera fotográfica digital com distância aproximada de 20 cm

## Registro 1



## Registro 2

302.222 Weil, Pierre  
W439c O corpo fala : a linguagem silenciosa da comunicação não-verbal / Pierre Weil, Roland Tompakow. - 55. ed. - Petrópolis : Vozes, 2002.  
288 p.

ISBN 85-326-0208-8

1. COMUNICAÇÃO NÃO-VERBAL I. Tompakow, Roland.  
II. Título.

## Registro 3

401.41 Discurso e texto : multiplicidade de sentidos na  
D639 ciência da informação / Nádea Regina Gaspar,  
Lucília Maria Sousa Romão (organizadoras). -  
São Carlos : EdUFSCar, 2008.  
259 p.

ISBN 978-85-7600-137-9

1. ANÁLISE DO DISCURSO 2. CIÊNCIA DA INFORMAÇÃO I. Gaspar, Nádea Regina, org. II. Romão, Lucília Maria Sousa, org.

Fonte: Elaborado pelo próprio autor.

O teste, com imagem tirada a uma distância de aproximadamente 10 cm, resultou nas imagens constantes da Figura 17.

Figura 17 – Imagens de registros bibliográficos capturadas a partir de câmera fotográfica digital com distância aproximada de 10 cm

Registro 1

925 Tiner, John Hudson  
T49c 100 cientistas que mudaram a história do mundo / John Hudson Tiner ; tradução de Marise Chinetti de Barros. - Rio de Janeiro : Ediouro, c2002.  
259 p. - (Coleção 100 mais)

Tradução de: 100 scientists who shaped world history.  
ISBN 85-00-01493-9

1. CIENTISTAS - BIOGRAFIA 2. CIÊNCIA - HISTÓRIA I. Título. II. Título: Cem cientistas que mudaram a história do mundo. III. Série.

Registro 2

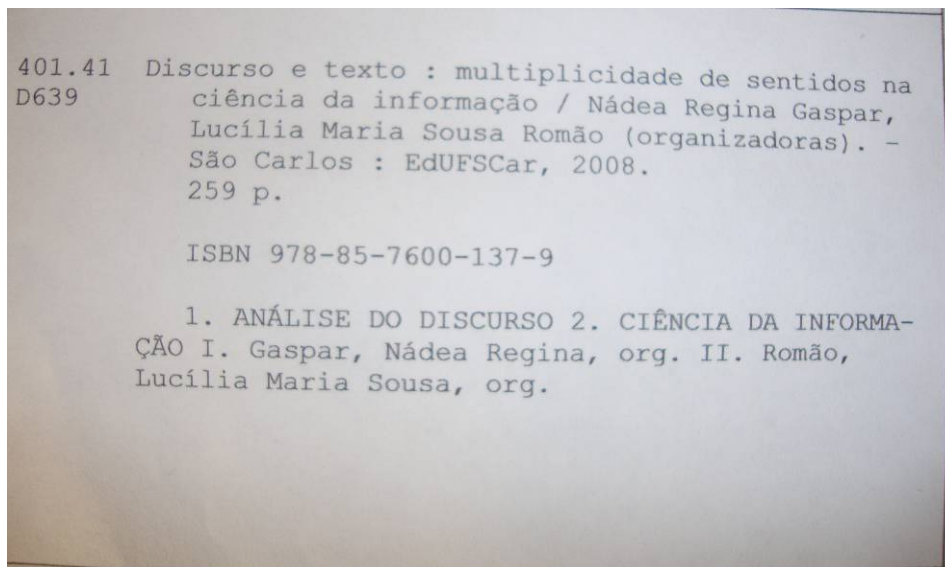
302.222 Weil, Pierre  
W439c O corpo fala : a linguagem silenciosa da comunicação não-verbal / Pierre Weil, Roland Tompakow. - 55. ed. - Petrópolis : Vozes, 2002.  
288 p.

ISBN 85-326-0208-8

1. COMUNICAÇÃO NÃO-VERBAL I. Tompakow, Roland. II. Título.



## Registro 3



Fonte: Elaborado pelo próprio autor.

Entende-se que os resultados neste equipamento tenham sido medianos, uma vez que, consideradas questões de luminosidade, qualidade da imagem e distância, estes se apresentaram com brilho e baixa definição dos caracteres.

Após a captura de imagens a partir de diversos recursos (*scanner* de impressora multifuncional, câmera de dispositivo móvel e máquina fotográfica digital), procederam-se testes para o processamento de imagens com reconhecimento de caracteres, expressos no próximo tópico.

#### 4.2 ANÁLISE DE *SOFTWARES* DE RECONHECIMENTO DE CARACTERES

Para que seja possível o processamento do registro bibliográfico em formato analógico é requerida a adoção de um *software* de reconhecimento de caracteres, processo que usa tecnologia reconhecida como *Optical Character Recognition* (OCR). A fase de testes para análise de *softwares* de reconhecimento de caracteres a partir do processamento das imagens de registros bibliográficos que foram capturadas foi dividida em dois momentos: o de análise de *softwares* gratuitos (*freeware*) ou de licença livre (*free software*), e o de *softwares* proprietários, com análise em versões *trial*. A identificação da lista de *softwares* voltados ao reconhecimento de caracteres (OCR) foi feita a partir de buscas aleatórias em sites de busca. Os *softwares* identificados estão no Quadro 3.

Quadro 3 – *Softwares* para reconhecimento ótico de caracteres em arquivos de imagem

<b>Nome</b>	<b>Licença</b>	<b>Online</b>	<b>Acesso</b>
ABBYY FineReader 11	Proprietário	Não	Disponível em: <a href="http://www.abbyy.com.br/finereader/">http://www.abbyy.com.br/finereader/</a>
ABBYY FineReader Online	Proprietário	Sim	Disponível em: <a href="http://finereader.abbyyonline.com/en/Task/Queue">http://finereader.abbyyonline.com/en/Task/Queue</a>
Cognitive Open OCR (Cuneiform) 0.1	Gratuito	Não	Disponível em: <a href="http://cognitive-openocr-cuneiform.en.softonic.com/download">http://cognitive-openocr-cuneiform.en.softonic.com/download</a>
Free WebOCR Online	Proprietário	Sim	Disponível em: <a href="http://www.expervision.com/ocr-software/webocr-onlineocr">http://www.expervision.com/ocr-software/webocr-onlineocr</a>
FreeOCR	Gratuito	Sim	Disponível em: <a href="http://www.free-ocr.com/">http://www.free-ocr.com/</a>
FreeOCR 3.1	Gratuito	Não	Disponível em: <a href="http://www.paperfile.net/freeocr.exe">http://www.paperfile.net/freeocr.exe</a>
GOOCR	Gratuito	Não	Disponível em: <a href="http://jocr.sourceforge.net/">http://jocr.sourceforge.net/</a>
jOCR1	Gratuito	Não	Disponível em: <a href="http://download.cnet.com/3001-2192_4-10768898.html?spi=e3428713a90b77d49b8ee8e74827db45">http://download.cnet.com/3001-2192_4-10768898.html?spi=e3428713a90b77d49b8ee8e74827db45</a>
Leadtools	Proprietário	Não	Disponível em: <a href="http://www.leadtools.com/sdk/ocr/default.htm">http://www.leadtools.com/sdk/ocr/default.htm</a>
OCR Terminal	Proprietário	Não	Disponível em: <a href="http://www.abbyyonline.com/en/home/ocrterminal">http://www.abbyyonline.com/en/home/ocrterminal</a>
OnlineOCR.net	Gratuito	Sim	Disponível em: <a href="http://www.onlineocr.net/default.aspx">http://www.onlineocr.net/default.aspx</a>
ScreenOCR 9.1	Proprietário	Não	Disponível em: <a href="http://www.screenocr.com">http://www.screenocr.com</a>
Sci2ools (i2OCR)	Gratuito	Sim	Disponível em: <a href="http://www.sciweavers.org/free-online-ocr">http://www.sciweavers.org/free-online-ocr</a>
Sci2ools 3.0	Gratuito	Não	Disponível em: <a href="http://sci2ools.softonic.com.br/download">http://sci2ools.softonic.com.br/download</a> <a href="https://chrome.google.com/webstore/detail/jpkakegfnfcmcdgicgdbnnpjiaeldi?hl=en#">https://chrome.google.com/webstore/detail/jpkakegfnfcmcdgicgdbnnpjiaeldi?hl=en#</a>
SimpleOCR 3.5	Gratuito	Não	Disponível em: <a href="http://www.charactell.com/scanstore/">http://www.charactell.com/scanstore/</a>
TopOCR 3.1	Gratuito	Não	Disponível em: <a href="http://www.brothersoft.com/topocr-download-47055-s1.html">http://www.brothersoft.com/topocr-download-47055-s1.html</a>
WeOCR Server	Gratuito	Sim	Disponível em: <a href="http://ocr1.sc.isc.tohoku.ac.jp/e1/">http://ocr1.sc.isc.tohoku.ac.jp/e1/</a>

Fonte: Elaborado pelo próprio autor.

Utilizou-se, para fins de uma primeira percepção da qualidade do *software* de reconhecimento de caracteres, principalmente quanto ao resultado de

processamento de sinais diacríticos<sup>22</sup>, o registro constante da Figura 18, selecionado por conta da presença de tais sinais.

Figura 18 – Imagem de registro bibliográfico analógico utilizado para os testes do processamento de imagem com reconhecimento de caracteres

```

330      Galache, G
G146b    Brasil processo e integração : estudos de
          problemas brasileiros / G Galache, M. André. --
          São Paulo : Loyola, 1972.
          414 p. : il. ; 21 cm

          Inclui bibliografia ao final dos capítulos.

          1. PROBLEMAS BRASILEIROS I. André, M. II.
          Título.
  
```

Fonte: Elaborado pelo próprio autor.

Os resultados dos testes, desenvolvidos a partir de análise quanto à pontuação, aos sinais diacríticos, à troca de letras, ao espaçamento entre informações e margens, estão indicados de forma sinóptica no Quadro 4, o qual indica, também, o apêndice com o registro dos resultados obtidos.

Quadro 4 – Resumo do resultado dos testes realizados em *softwares* para reconhecimento ótico de caracteres

Nome	Considerações	Apêndice
ABBYY FineReader 11	Resultado insatisfatório: pontuação, sinais diacríticos e troca de letras	A
ABBYY FineReader Online	Resultado insatisfatório: espaçamento entre informações (número de chamada) e margens	B
Cognitive Open OCR (Cuneiform) 0.1	Resultado insatisfatório: sinais diacríticos, troca de letras, espaçamento entre informações (número de chamada) e margens	C
FreeOCR	Resultado insatisfatório: pontuação, troca de letras	D
FreeOCR 3.1	Resultado insatisfatório: pontuação, troca de letras, espaçamento entre informações	E
Leadtools	Resultado insatisfatório: pontuação, troca de letras, espaçamento entre informações	F

<sup>22</sup> Os sinais diacríticos são, segundo iDicionário Aulete (2008), sinais gráficos adotados para conferir valor fonético ou fonológico a uma letra. Podem ser identificados por acento agudo (´), grave (`), circunflexo (^), cedilha (,), trema (¨) e til (~).

OnlineOCR.net	Resultado satisfatório: pontuação, sinais diacríticos, troca de letras, espaçamento entre informações; problema com margens (ausentes diante da opção de download do resultado)	G
ScreenOCR 9.1	Resultado insatisfatório: pontuação, sinais diacríticos, troca de letras, espaçamento entre informações e margens	H
Sci2ools (i2OCR)	Resultado insatisfatório: pontuação, troca de letras, espaçamento entre margens	I
SimpleOCR 3.5	Resultado insatisfatório: pontuação, sinais diacríticos, troca de letras, espaçamento entre informações e margens	J
TopOCR 3.1	Resultado insatisfatório: pontuação, troca de letras, espaçamento entre informações e margens	K
WeOCR Server	Resultado insatisfatório: sinais diacríticos, troca de letras, espaçamento entre margens	L

Fonte: Elaborado pelo próprio autor.

Dentre os aplicativos identificados, constantes do Quadro 3, não foi possível proceder a avaliação dos seguintes *softwares*:

- Free WebOCR Online: ao final do teste, o email recebido indica o arquivo submetido (imagem) e não o arquivo interpretado e reconhecido;
- GOCR: não foi possível instalar o aplicativo e nenhum problema foi aparentemente identificado;
- jOCR1: indica que não há linguagem OCR no MSOffice capaz de reconhecer o arquivo;
- OCR Terminal: faz parte agora do ABBY FineReader Online;
- Sci2ools 3.0: trabalha somente com arquivos disponíveis na *web*.

Tendo sido identificado que o *software* OnlineOCR.net promoveu melhor resultado (cf. Apêndice G) diante do teste inicial, procederam-se testes a partir dos três registros escaneados, constantes do tópico 4.1.

#### 4.3 PROCESSAMENTO DAS IMAGENS DE REGISTROS BIBLIOGRÁFICOS

Este tópico destina-se ao registro dos testes de processamento de imagens dos registros bibliográficos selecionados, diante do OnlineOCR.net, *software* de reconhecimento de caracteres em arquivos de imagens que satisfaz aos requisitos de testes e análises constantes do tópico anterior (cf. 4.2). Apresenta-se dividido a partir do recurso no qual as imagens foram capturadas: scanner de impressora

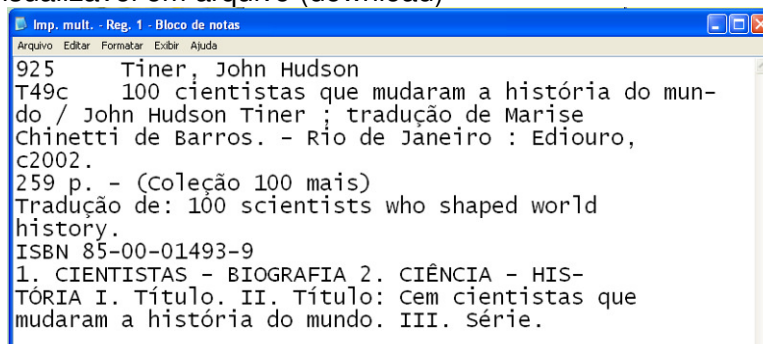


multifuncional, câmera fotográfica digital e câmera de dispositivo móvel. Os testes concentram-se nos resultados obtidos em arquivos de *download*, uma vez que, para o processamento sintático e semântico é requerida a adoção de arquivos, e não de visualização na interface web.

#### 4.3.1 Processamento de imagens de registros bibliográficos capturados em scanner de impressora multifuncional

O teste realizado no aplicativo OnlineOCR.net, da imagem do Registro 1, capturada em scanner de impressora multifuncional, apresenta o resultado presente na Figura 19:

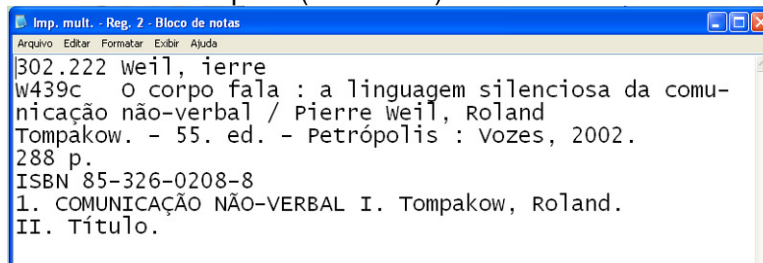
Figura 19 – Processamento da imagem do Registro 1, capturada em scanner de impressora multifuncional, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O segundo teste realizado, da imagem do Registro 2, capturada em scanner de impressora multifuncional, apresenta o resultado presente na Figura 20:

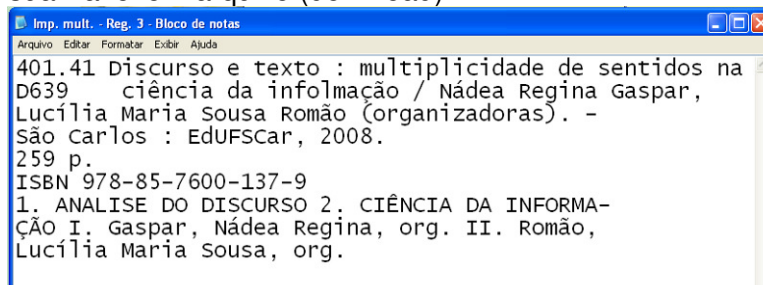
Figura 20 – Processamento da imagem do Registro 2, capturada em scanner de impressora multifuncional, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O teste realizado com a imagem do Registro 3, capturada em scanner de impressora multifuncional, apresenta o resultado indicado na Figura 21:

Figura 21 – Processamento da imagem do Registro 3, capturada em scanner de impressora multifuncional, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

De modo geral, a análise que se pode fazer diante da multiplicidade de características específicas a cada um dos registros bibliográficos selecionados (Registro 1, Registro 2 e Registro 3), permite os seguintes comentários:

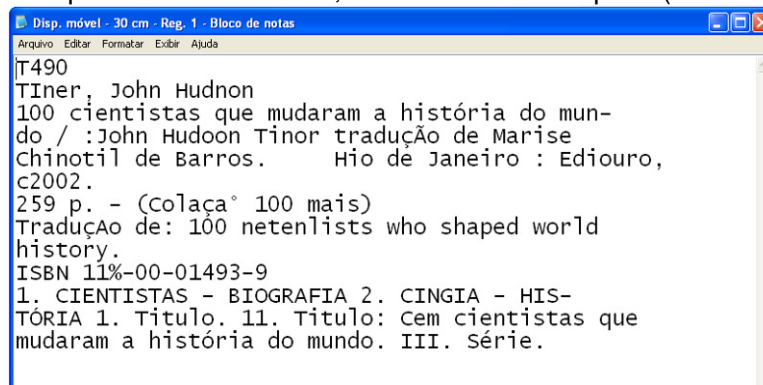
- a) os resultados do processamento das imagens do Registro 1 mostraram-se satisfatória;
- b) quanto ao resultado do Registro 2 há a indicação de omissão de letra inicial do prenome do autor, 'ierre' ao invés de 'Pierre';
- c) o resultado do processamento da imagem do Registro 3 apresenta problemas: troca de letra ('r' por 'l', na palavra 'informação'), omissão de sinal diacrítico ('', na palavra 'ANÁLISE').

Buscar-se-á a verificação do processamento de imagens capturadas por outros métodos para que seja possível contextualizar o índice de satisfação do resultado obtido acima.

#### 4.3.2 Processamento de imagens de registros bibliográficos capturados em câmera de dispositivo móvel

O teste realizado no aplicativo OnlineOCR.net, da imagem do Registro 1, capturada em câmera de dispositivo móvel em distância aproximada de 30 cm, apresenta o resultado presente na Figura 22:

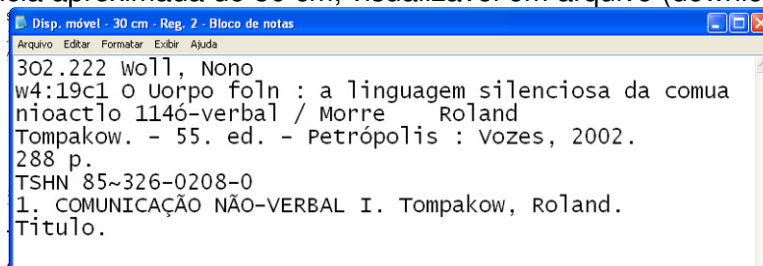
Figura 22 – Processamento da imagem do Registro 1, capturada em câmera de dispositivo móvel em distância aproximada de 30 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O mesmo teste realizado, a partir da imagem do Registro 2, apresenta o resultado presente na Figura 23:

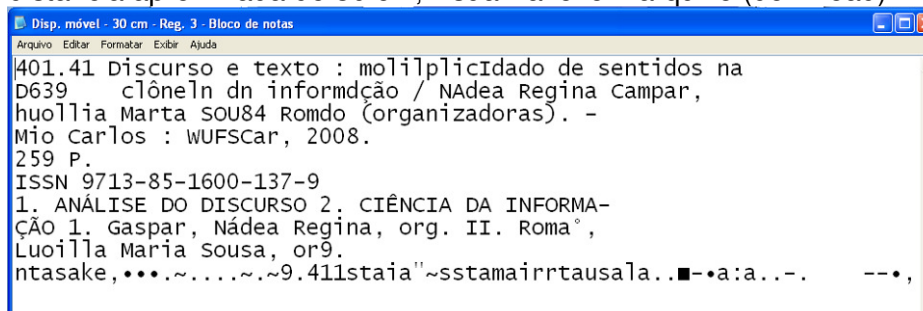
Figura 23 – Processamento da imagem do Registro 2, capturada em câmera de dispositivo móvel em distância aproximada de 30 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O teste, realizado com a imagem do Registro 3, apresenta o resultado indicado na Figura 24:

Figura 24 – Processamento da imagem do Registro 3, capturada em câmera de dispositivo móvel em distância aproximada de 30 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

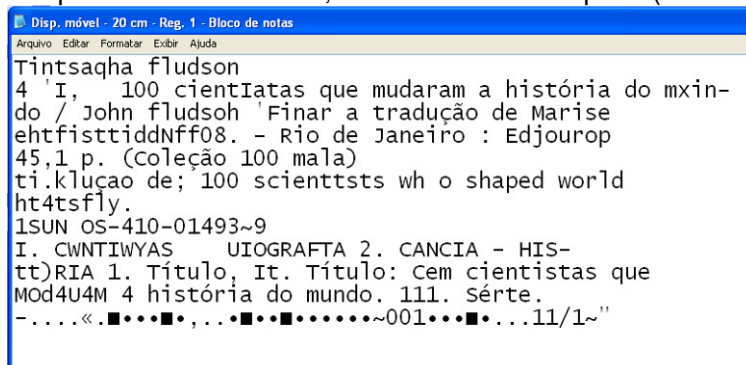
De modo geral, a análise que se pode fazer diante da multiplicidade de características específicas a cada um dos registros bibliográficos selecionados (Registro 1, Registro 2 e Registro 3), e segundo a forma de captura das imagens, capturadas em câmera de dispositivo móvel com distância aproximada de 30 cm, é possível estabelecer os seguintes comentários:

- a) o resultado do processamento da imagem do Registro 1 mostra que houve omissão, inserção e troca de algumas informações:
- omissão de caracteres (número de classificação de assunto '925'; pontuação ';' e '-'; letra 'Ê');
  - inserção de caractere (':' antes de 'John' na indicação de responsabilidade);
  - troca de letras ('c' por '0'; 'i' por 'l'; 's' por 'n'; 's' por 'o'; 'e' por 'o'; 'ã' por 'Ã'; 't' por 'i'; 'i' por 'l'; 'R' por 'H'; 'e' por 'a'; 'ã' por 'a'; 'o' por 'º'; 'C' por 'G'; 'l' por '1'; e 'í' por 'i');
  - troca de palavras ('scientists' por 'netenlists'; '85-00-01493-9' por '11%-00-01493-9');
- b) o resultado do processamento da imagem do Registro 2 mostra que houve omissão e troca de algumas informações:
- troca de caracteres ('0' por 'O'; 'e' por 'o'; 'i' por 'l'; 'c' por 'U'; 'a' por 'o'; 'a' por 'n'; '-' por 'a'; 'c' por 'o'; 'ç' por 'c'; 'ã' por 'tl'; 'n' por '11'; 'ã' por '4'; 'o' por 'ó'; '-' por '~', 'í' por 'i');
  - troca de palavras ('Pierre' por 'Nono' e, depois, por 'Morre'; 'W439c' por 'w4:19c1');
  - omissão de caracteres (',');
  - omissão de palavras ('Weil'; );
- c) o resultado do processamento da imagem do Registro 3 mostra que houve troca e acréscimo de algumas informações:
- troca de letras ('u' por 'o'; 't' por 'i'; 'i' por 'l'; 'e' por 'o'; 'ê' por 'ô'; 'e' por 'c'; 'n' por 'a'; 'd' por 'a'; 'á' por 'A'; 'G' por 'C'; 's' por 'm'; 'L' por 'h'; 'c' por 'o'; 'í' por 'l'; 'i' por 't'; 'o' por 'O'; 'u' por 'U'; '8' por 's'; '4' por 'a'; 'd' por 'ã'; 'Sã' por 'Mi'; 'Ed' por 'W'; 'B' por 'S'; '8' por '13'; '1' por '7'; 'l' por '1'; 'ã' por 'a'; 'o' por 'º'; 'í' por 'i'; 'g' por '9'; 'º' por 'º'; 'º' por 'º'; 'º' por 'º'; 'º' por 'º'; 'º' por 'º'; 'º' por 'º'; 'º' por 'º';
  - acrécimo de informações ('ntasake,••.~....~.~9.411staia"~sstamairrtausala..■-•a:a..- --•,').

Prosseguiu-se a análise dos testes realizados com imagens capturadas em câmera de dispositivo móvel, desta vez com distância aproximada de 20 cm.

O teste realizado no aplicativo OnlineOCR.net, da imagem do Registro 1, capturada em câmera de dispositivo móvel em distância aproximada de 20 cm, apresenta o resultado presente na Figura 25:

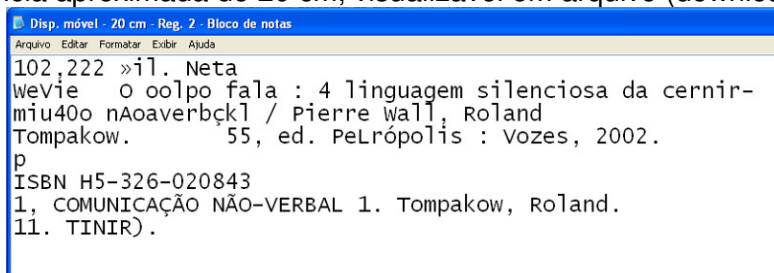
Figura 25 – Processamento da imagem do Registro 1, capturada em câmera de dispositivo móvel em distância aproximada de 20 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O mesmo teste realizado, da imagem do Registro 2, apresenta o resultado presente na Figura 26:

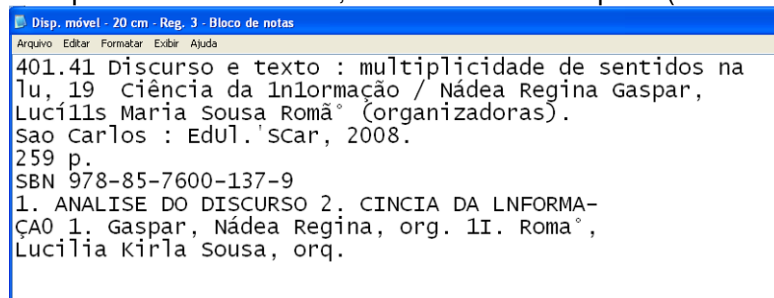
Figura 26 – Processamento da imagem do Registro 2, capturada em câmera de dispositivo móvel em distância aproximada de 20 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O teste, realizado com a imagem do Registro 3, apresenta o resultado indicado na Figura 27:

Figura 27 – Processamento da imagem do Registro 3, capturada em câmera de dispositivo móvel em distância aproximada de 20 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

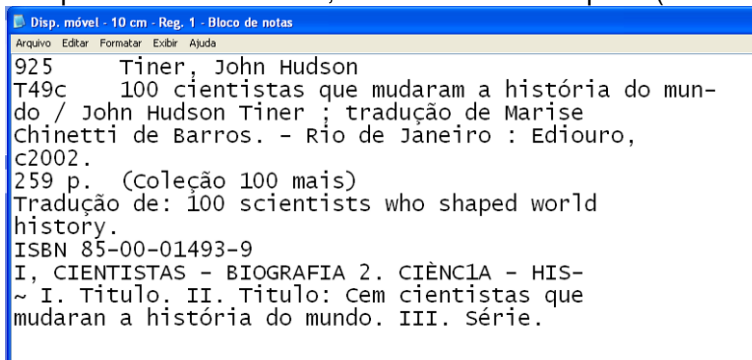
De modo geral, a análise que se pode fazer diante da multiplicidade de características específicas a cada um dos registros bibliográficos selecionados

(Registro 1, Registro 2 e Registro 3), e segundo a forma de captura das imagens, capturadas em câmera de dispositivo móvel com distância aproximada de 20 cm, permite comentar que o resultado do processamento das imagens mostra-se bastante insatisfatório quanto à interpretação dos caracteres.

Procederam-se testes a partir de imagens capturadas em distância aproximada de 10 cm.

A análise dos resultados dos testes de reconhecimento ótico de caracteres para imagens capturadas em câmera de dispositivo móvel, em distância aproximada de 10 cm, apresenta os resultados a seguir. A Figura 28 apresenta do resultado da imagem do Registro 1:

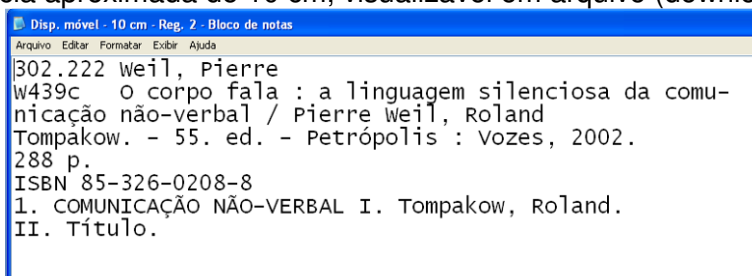
Figura 28 – Processamento da imagem do Registro 1, capturada em câmera de dispositivo móvel em distância aproximada de 10 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O teste realizado com a imagem do Registro 2, apresenta o resultado presente na Figura 29:

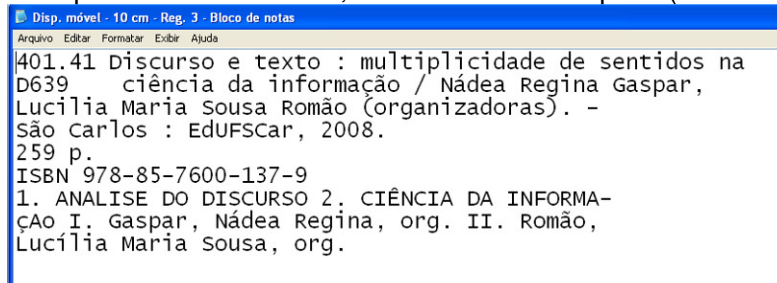
Figura 29 – Processamento da imagem do Registro 2, capturada em câmera de dispositivo móvel em distância aproximada de 10 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O teste, realizado com a imagem do Registro 3, apresenta o resultado indicado na Figura 30:

Figura 30 – Processamento da imagem do Registro 3, capturada em câmera de dispositivo móvel em distância aproximada de 10 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

De modo geral, a análise que se pode fazer diante da multiplicidade de características específicas a cada um dos registros bibliográficos selecionados (Registro 1, Registro 2 e Registro 3), e segundo a forma de captura das imagens, capturadas em câmera de dispositivo móvel com distância aproximada de 10 cm, permite os seguintes comentários:

- a) o resultado do processamento da imagem do Registro 1 mostra que houve omissão e troca de algumas informações:
  - a. omissão de caracteres (pontuação ‘-’; letra ‘Ê’);
  - b. troca de caracteres (‘1’ por ‘l’; ‘.’ por ‘;’; ‘Ê’ por ‘È’; ‘l’ por ‘1’; ‘TÓRIA’ por ‘~’; ‘í’ por ‘i’; ‘m’ por ‘n’);
- b) o resultado do processamento da imagem do Registro 2 mostra-se satisfatório, uma vez que nenhum problema quando ao reconhecimento de caracteres foi identificado;
- c) o resultado do processamento da imagem do Registro 3 mostra que houve troca de letras (‘í’ por ‘i’; ‘Á’ por ‘A’, ‘Ç’ por ‘ç’; ‘Ã’ por ‘A’; ‘O’ por ‘o’).

Diante dos testes de processamento de imagens de registros bibliográficos capturados em câmera de dispositivo móvel é possível identificar:

- a) que é necessário observar aspectos da luminosidade no momento da captura da imagem, pois se observou que estes influenciam de maneira significativa a qualidade da imagem;
- b) que aspectos quanto à distância da captura das imagens podem ser irrelevantes dado que imagens que foram capturadas a uma distância aproximada de 20 cm apresentaram resultados piores do que os testes



com imagens capturadas a uma distância aproximada de 30 cm (atenta-se aos contornos dos caracteres de forma relativamente distorcida);

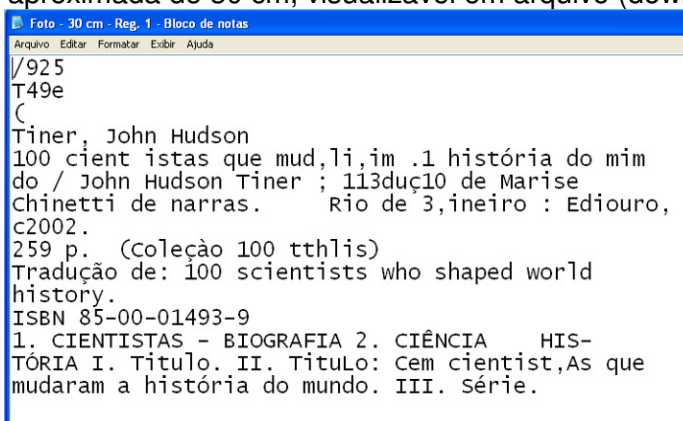
- c) que o comentário anterior pode ser decorrente da necessidade de se afixar a câmera de alguma forma para que as imagens não apresentem resultados distorcidos, tal como 'imagem tremida'.

Após a finalização dos testes de processamento de imagens de registros bibliográficos capturados em câmera de dispositivo móvel, procederam-se testes com imagens capturadas em câmera fotográfica digital.

#### 4.3.3 Processamento de imagens de registros bibliográficos capturados em câmera fotográfica digital

O teste realizado no aplicativo OnlineOCR.net, da imagem do Registro 1, capturada em câmera fotográfica digital em distância aproximada de 30 cm, apresenta o resultado presente na Figura 31:

Figura 31 – Processamento da imagem do Registro 1, capturada em câmera fotográfica digital em distância aproximada de 30 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.



O mesmo teste realizado, da imagem do Registro 2, apresenta o resultado presente na Figura 32:

Figura 32 – Processamento da imagem do Registro 2, capturada em câmera fotográfica digital em distância aproximada de 30 cm, visualizável em arquivo (download)

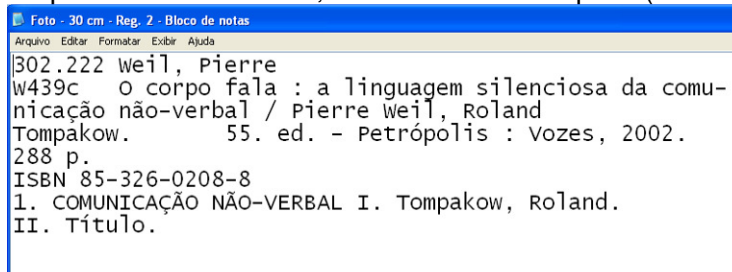


Foto - 30 cm - Reg. 2 - Bloco de notas

Arquivo Editar Formatar Exibir Ajuda

302.222 Weil, Pierre  
 W439c O corpo fala : a linguagem silenciosa da comunicação não-verbal / Pierre Weil, Roland Tompakow. 55. ed. - Petrópolis : Vozes, 2002. 288 p. ISBN 85-326-0208-8  
 I. COMUNICAÇÃO NÃO-VERBAL I. Tompakow, Roland. II. Título.

Fonte: Elaborado pelo próprio autor.

O teste, realizado com a imagem do Registro 3, apresenta o resultado indicado na Figura 33:

Figura 33 – Processamento da imagem do Registro 3, capturada em câmera fotográfica digital em distância aproximada de 30 cm, visualizável em arquivo (download)

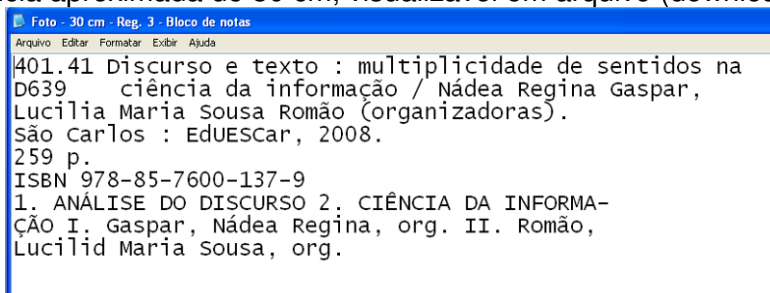


Foto - 30 cm - Reg. 3 - Bloco de notas

Arquivo Editar Formatar Exibir Ajuda

401.41 Discurso e texto : multiplicidade de sentidos na ciência da informação / Nádea Regina Gaspar, Lucilia Maria Sousa Romão (organizadoras). São Carlos : EdUEScar, 2008. 259 p. ISBN 978-85-7600-137-9  
 1. ANÁLISE DO DISCURSO 2. CIÊNCIA DA INFORMAÇÃO I. Gaspar, Nádea Regina, org. II. Romão, Lucilid Maria Sousa, org.

Fonte: Elaborado pelo próprio autor.

De modo geral, a análise que se pode fazer diante da multiplicidade de características específicas a cada um dos registros bibliográficos selecionados (Registro 1, Registro 2 e Registro 3), e segundo a forma de captura das imagens, capturadas em câmera fotográfica digital com distância aproximada de 30 cm, permite os seguintes comentários:

- a) o resultado do processamento da imagem do Registro 1 mostra-se insatisfatório, uma vez que podem ser observados muitos problemas quanto ao reconhecimento dos caracteres, com destaque para omissão, inserção e troca de algumas informações;
- b) o resultado do processamento da imagem do Registro 2 mostra-se de grande qualidade, uma vez que houve somente um caso a ser reportado: a omissão de um sinal de pontuação '-';

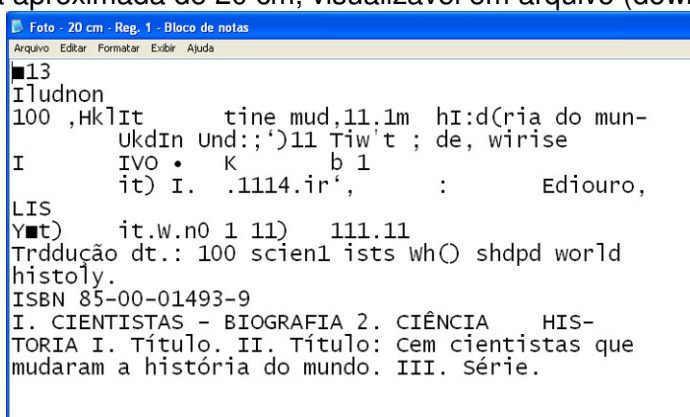
c) o resultado do processamento da imagem do Registro 3 mostra-se de qualidade considerável, uma vez que houve troca e omissão de algumas informações:

- a. troca de letras ('f' por 'i'; 'F' por 'E'; 'a' por 'd');
- b. omissão de caractere ('-').

Prosseguiu-se a análise dos testes realizados com imagens capturadas em câmera fotográfica digital, desta vez com distância aproximada de 20 cm.

O teste realizado no aplicativo OnlineOCR.net, da imagem do Registro 1, capturada em câmera fotográfica digital em distância aproximada de 20 cm, apresenta o resultado presente na Figura 34:

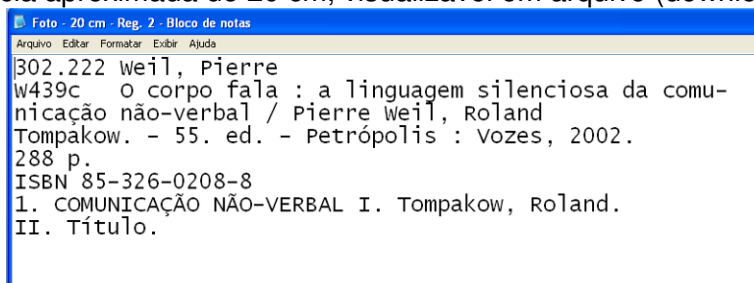
Figura 34 – Processamento da imagem do Registro 1, capturada em câmera fotográfica digital em distância aproximada de 20 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O mesmo teste realizado, da imagem do Registro 2, apresenta o resultado presente na Figura 35:

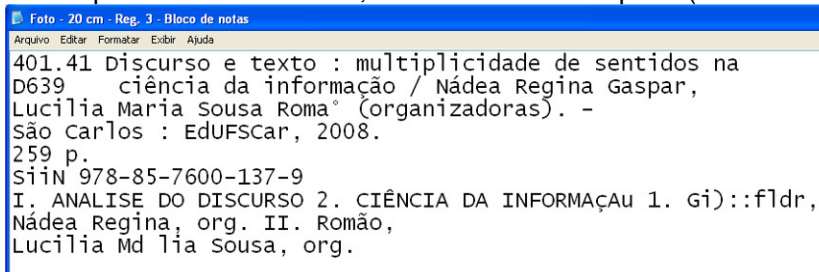
Figura 35 – Processamento da imagem do Registro 2, capturada em câmera fotográfica digital em distância aproximada de 20 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O teste, realizado com a imagem do Registro 3, apresenta o resultado indicado na Figura 36:

Figura 36 – Processamento da imagem do Registro 3, capturada em câmera fotográfica digital em distância aproximada de 20 cm, visualizável em arquivo (download)



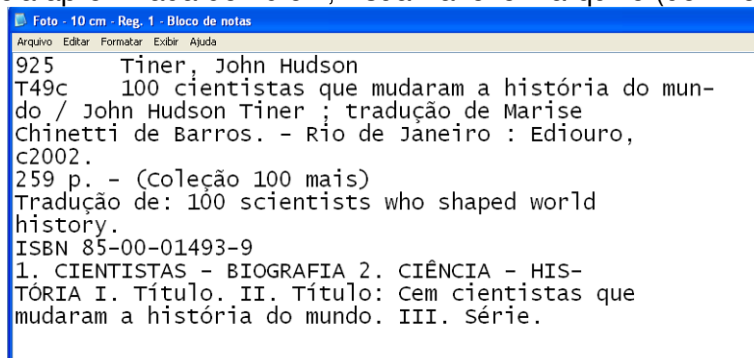
Fonte: Elaborado pelo próprio autor.

De modo geral, a análise que se pode fazer diante da multiplicidade de características específicas a cada um dos registros bibliográficos selecionados (Registro 1, Registro 2 e Registro 3), e segundo a forma de captura das imagens, capturadas em câmera fotográfica digital com distância aproximada de 20 cm, permite os seguintes comentários:

- a) o resultado do processamento da imagem do Registro 1 mostra-se insatisfatório, uma vez que podem ser observados muitos problemas quanto ao reconhecimento dos caracteres de modo a inutilizá-lo, com destaque para omissão, inserção e troca de algumas informações;
- b) o resultado do processamento da imagem do registro 2 mostra um resultado bastante satisfatório quanto à interpretação dos caracteres, uma vez que não apresenta nenhuma ocorrência de problema;
- d) o resultado do processamento da imagem do Registro 3 mostra-se de qualidade considerável, uma vez que houve troca e omissão de algumas informações:
  - a. troca de letras ('í' por 'i'; 'ã' por 'a'; 'o' por 'o'; 'Á' por 'A'; 'Ç' por 'ç'; 'Ã' por 'A'; 'o' por 'u'; 'l' por '1'; 'áspa' por 'i'):fldr; 'a' por 'd'; 'r' por 'l');
  - b. inserção de caractere em branco.

A análise dos resultados dos testes de reconhecimento ótico de caracteres para imagens capturadas em câmera fotográfica digital, em distância aproximada de 10 cm, apresenta os resultados a seguir. A Figura 37 apresenta do resultado da imagem do Registro 1:

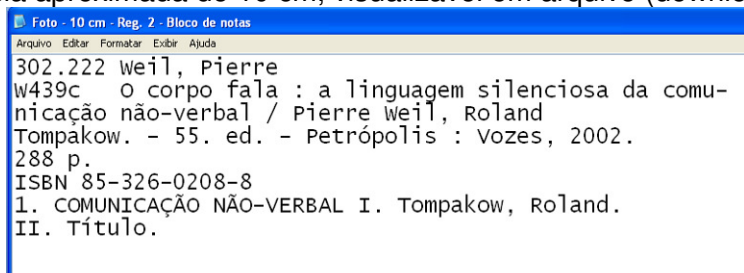
Figura 37 – Processamento da imagem do Registro 1, capturada em câmera fotográfica digital em distância aproximada de 10 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O teste realizado com a imagem do Registro 2, apresenta o resultado presente na Figura 38:

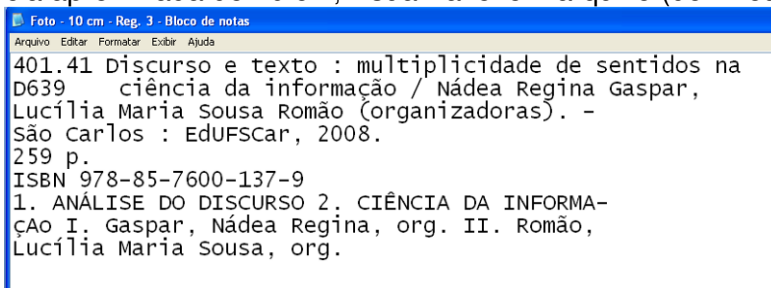
Figura 38 – Processamento da imagem do Registro 2, capturada em câmera fotográfica digital em distância aproximada de 10 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

O teste, realizado com a imagem do Registro 3, apresenta o resultado indicado na Figura 39:

Figura 39 – Processamento da imagem do Registro 3, capturada em câmera fotográfica digital em distância aproximada de 10 cm, visualizável em arquivo (download)



Fonte: Elaborado pelo próprio autor.

De modo geral, a análise que se pode fazer diante da multiplicidade de características específicas a cada um dos registros bibliográficos selecionados (Registro 1, Registro 2 e Registro 3), e segundo a forma de captura das imagens, capturadas em câmera fotográfica digital com distância aproximada de 10 cm, é possível verificar que os resultados dos processamentos das imagens dos registro 1,

2 e 3 mostraram-se altamente satisfatórios, uma vez que não há nenhum problema de reconhecimento de caractere a reportar.

Diante dos testes de processamento de imagens de registros bibliográficos capturados em câmera fotográfica digital é possível identificar:

- a) que as imagens capturadas por esse meio apresentam baixa qualidade de interpretação dos caracteres quando em distância aproximada de 20 cm ou 30 cm, fato que não ocorre quando se opta por distância aproximada de 10 cm;
- b) que há necessidade de se observar aspectos da luminosidade no momento da captura da imagem, pois se observou que estes influenciam de maneira significativa a qualidade da imagem;
- c) que aspectos quanto à distância da captura das imagens mostram-se relevantes, haja vista que os resultados de processamento com imagens capturadas em distâncias aproximadas superiores a 10 cm apresentaram resultados muito melhores (atenta-se aos contornos dos caracteres de forma relativamente distorcida).

Após a finalização dos testes de processamento de imagens de registros bibliográficos capturados em câmera de dispositivo móvel, apresentar-se-ão comentários acerca do resultado geral do processamento de imagens de registros bibliográficos capturados em scanner de impressora multifuncional, em câmera de dispositivo móvel e em câmera fotográfica digital.

#### 4.3.4 Resultado geral do processamento de imagens de registros bibliográficos

Findos os testes de processamento de imagens de registros bibliográficos capturados por três diferentes recursos é possível apresentar os seguintes resultados de análise geral:

- a) questões referentes à iluminação na captura das imagens são extremamente relevantes para o processamento em *software* de reconhecimento de caracteres pois influenciam a qualidade da imagem;

- b) os melhores índices de sucesso no processamento de imagens de registros bibliográficos foram capturados diante das imagens capturadas em scanner de impressora multifuncional;
- c) problemas com a troca de letras no processo de reconhecimento de caracteres em imagens foram identificados nos testes com imagens de registros bibliográficos capturadas nos três diferentes recursos;
- d) aspectos quanto à fixação da câmera, quer seja de dispositivo móvel quer seja de fotografia digital, mostram-se mais relevantes do que a distância a ser considerada na captura da imagem;
- e) a qualidade do reconhecimento ótico dos caracteres está intrinsecamente vinculada à qualidade da imagem e não ao método pelo qual a imagem é capturada, uma vez que pôde ser verificada semelhança no processamento de imagens capturadas em scanner de impressora multifuncional, em câmera de dispositivo móvel e em câmera fotográfica digital.

Diante dos resultados, optou-se por trabalhar com as imagens capturadas em *scanner* de impressora multifuncional, dada a sua praticidade, e o uso do OnlineOCR.net como *software* para processamento do reconhecimento de caracteres em imagens. Essas escolhas foram feitas para que sejam avaliadas as possibilidades de comprovação da tese proposta: os princípios sintáticos e semânticos de registros bibliográficos, definidos pelos esquemas de descrição e de visualização na catalogação, determinam o processo de conversão de registros bibliográficos para o Formato MARC21 Bibliográfico, apresentadas a seguir.

#### 4.4 TRATAMENTO SINTÁTICO E SEMÂNTICO DE REGISTROS BIBLIOGRÁFICOS PARA O FORMATO MARC21 BIBLIOGRÁFICO

Efetuados os testes de processamento de imagens dos registros bibliográficos analógicos selecionados, este tópico busca apresentar os procedimentos adotados para o tratamento sintático e semântico dos dados bibliográficos do arquivo, à luz da proposta teórica apresentada nesta pesquisa.

Tendo em vista a busca pela garantia da interpretação dos dados e a simplificação do processamento optou-se pelo uso da linguagem Perl que, segundo

Stockton ([2005]), é uma linguagem interpretada otimizada para examinar arquivos de textos, extrair informações de arquivos de textos, e gerar relatórios a partir destas informações. Dois pontos considerados de interesse nesta pesquisa, diante da opção pela linguagem Perl de interpretação, referem-se ao fato de estar sob os termos da Licença Pública Geral (GNU), caracterizando-se como sendo *software* livre. (PERL, [2011]), e por permitir aliar programas desenvolvidos em diversificados ambientes, tais como UNIX, MSDOS, Windows, Macintosh, OS/2.

O Perl é uma linguagem de programação de uso geral, originalmente desenvolvida para manipulação de textos, que agora é utilizada para uma infinidade de tarefas incluindo administração de sistemas, desenvolvimento web, programação de redes, desenvolvimento de interfaces gráficas, e muitos outros. A linguagem foi escrita para ser prática (fácil de utilizar, eficiente, completa) ao invés de bonita (pequena, elegante, mínima). Suas maiores características são sua facilidade de uso, que inclui o suporte tanto para programação procedural quanto para programação orientada à objetos, tem um poderoso suporte incluso na linguagem para processamento de textos, e uma das maiores coleções de módulos escritos por terceiros, o CPAN [Comprehensive Perl Archive Network]. (ROBERT, 2009).

Para o desenvolvimento dos testes, indicados a seguir, recorreu-se ao ambiente Strawberry Perl, voltado para o Microsoft Windows e projetado para ser o mais próximo possível do ambiente Perl em ambientes Unix.<sup>23</sup>

As atividades desta etapa foram desenvolvidas a partir da compreensão dos aspectos sintáticos presentes em um registro bibliográfico, determinados pelo *padrão de estrutura de metadados descritivos*. Desse modo, foram estudadas as possibilidades sintáticas advindas do padrão AACR2r, sumarizadas nas regras 1.0D1 e 1.0D2 (CÓDIGO..., 2004), que tem sua base nas ISBDs. Os aspectos semânticos são observados e desenvolvidos a partir da realização do nível subsemântico (presente na relação entre a codificação das TAGs em MARC com os valores de cada subcampo do registro), a confirmação do nível semântico e, principalmente, pela construção automatizada do nível supersemântico.

Foram estudados os sinais de pontuação, presentes no *padrão de estrutura de metadados descritivos*, que, de acordo com a tese proposta, definem a inferência de conteúdo semântico, presente nos *padrões de conteúdo*. Ressalte-se que, neste

---

<sup>23</sup> Ressalte-se que o desenvolvimento do *script* foi feito por um programador com ampla experiência em PERL.



caso, a seqüência dos sinais de pontuação apresenta, também, significado sintático inerente ao conjunto de informações de contexto no qual se insere, ou seja, o mesmo sinal, quando utilizado em duas acepções, apresenta diferentes atribuições ao conteúdo semântico.

O Quadro 5 apresenta o resultado da análise sintática e sua relação semântica nos dados de descrição de um registro bibliográfico:

Quadro 5 – Resultado da análise sintática e sua relação semântica nos dados de descrição de um registro bibliográfico

Análise sintática	Análise semântica
Primeiro parágrafo	Título
:	Precede o subtítulo ou outras informações sobre o título
/	Precede informações sobre indicação de responsabilidade
,	Precede informações sobre indicação de responsabilidade de mesmo nível que a indicação de responsabilidade anterior
;	Precede informações sobre indicação de responsabilidade de nível diferente que o nível anterior
,	Precede informações sobre indicação de responsabilidade de mesmo nível que a indicação de responsabilidade imediatamente anterior
. --	Precede informações de que há nova área a seguir. Esta sinalização ocorre em diversos momentos do registro e, em todos, a sinalização é a mesma: indica nova área a seguir. Neste caso, a informação a seguir pode ser relativa à edição ou à publicação
:	Precede informações sobre casa publicadora
;	Precede informações sobre nova ocorrência de local de publicação
:	Precede informações sobre nova ocorrência sobre casa publicadora
,	Precede informações sobre data
Novo parágrafo	Precede informações sobre área de detalhes físicos. A primeira informação refere-se à extensão do item
:	Precede informações sobre outros detalhes físicos
;	Precede informações sobre dimensões do item
. --	Precede informações de que há nova área a seguir. Esta sinalização ocorre em diversos momentos do registro e, em todos, a sinalização é a mesma: indica nova área a seguir. Neste caso, a informação a seguir pode ser relativa à série ou às notas
(	Precede informações sobre o título da série
;	Precede informações referentes à numeração na série
)	Encerra informações sobre a série
Novo parágrafo	Precede informações de notas ou de número normalizado

Fonte: Elaborado pelo próprio autor.

Para a efetiva adoção destas marcações no *script* de conversão de registros bibliográficos ao MARC21 Bibliográfico foi necessário, entretanto, cuidado quanto à pontuação que faz parte do conteúdo e não da marcação em si, presente, por



exemplo, em 3. ed., no qual o ‘.’ (ponto final) indica abreviação do conteúdo (numeral ordinal) e não pontuação de nova informação a ser processada; ou em 23 cm. – (Série), que remete o ‘.’ (ponto final) para a indicação de separação entre áreas e não de abreviação da unidade de medida indicada. Outra análise também foi requerida quanto às informações dos pontos de acesso.

Tomou-se como padrão o fato de que os pontos de acesso definidos por cabeçalhos de assunto iniciam-se por numerais indo-arábicos seguidos de ponto, e de que os outros pontos de acesso que se fizerem necessários (de indicação de responsabilidade, de título da obra ou de título da série) iniciam-se por numerais romanos.

Para o desenvolvimento da *versão 1 do script* procedeu-se avaliação sintática de um registro bibliográfico, voltado para a inferência semântica, com sua modelagem a partir de registros razoavelmente simples, no qual se observava um roteiro com as características presentes na Figura 40:

Figura 40 – Primeiro roteiro para o *script* de conversão de registros bibliográficos

```
123      Sobrenome, Nome
S789a    Título : subtítulo / Nome Sobrenome,
          Outro Nome Sobrenome. -- Local de Publicação :
          Editora, Ano.
          929 p. : il. ; 23 cm

          Inclui bibliografia.

          1. ASSUNTO I. Sobrenome, Outro Nome. II.
          Título.
```

Fonte: Elaborado pelo próprio autor.

Adotar-se um registro com tal configuração mostrou-se necessário para, em um primeiro momento, fornecer os subsídios primários para a confirmação da tese proposta. Nesse sentido, a Figura 41 apresenta o resultado obtido nos primeiros testes de validação do *script*:

Figura 41 – Resultado de conversão de registro bibliográfico conforme versão 1 do *script*

```
==== MARC =====
090 $a123 $bs789a
1001 $aSobrenome, Nome
24510$aTítulo $bsubtítulo $cNome Sobrenome, Outro Nome Sobrenome
260 $aLocal de Publicação $bEditora $cAno
300 $a929 p. $bil. $c23 cm
500 Inclui bibliografia
65014$aASSUNTO
7001 $aSobrenome, Outro Nome
=====
```

Fonte: Elaborado pelo próprio autor.

Sem dúvida alguma, um dos maiores problemas identificados para o processamento das informações nesta versão do *script* foi decorrente do tipo de codificação dos caracteres do arquivo texto de entrada (UTF-8/ISO, UFT-16/UNICODE, ASCII/ ANSI), necessários para a correta interpretação dos sinais diacríticos. Diante disso, outras alterações se fizeram necessárias.

Neste primeiro teste também houve dificuldade para se pensar no tratamento que os elementos da área de notas merecem, uma vez que a única marcação sintática presente, observável no caso acima, é a de início de parágrafo. De modo geral, e ainda incipiente, declarou-se que as notas deveriam ser agregadas ao campo 500 que não é dedicado a receber notas de Bibliografia e Índice (TAG 504).

Posteriormente, tendo sido possível identificar, ainda de modo sumarizado, a possibilidade de conversão de registros bibliográficos a partir de análise sintática e inferência semântica, optou-se pela adoção de um segundo roteiro, com certo grau de complexidade, conforme ilustra a Figura 42:

Figura 42 – Segundo roteiro para o *script* de conversão de registros bibliográficos

```
123.456 Sobrenome do Autor, Primeiro Nome
S789a      Aqui vai o título do livro : aqui o sub-
           título que, neste caso, é muito importante
           também / Primeiro Nome Sobrenome do Autor,
           Segundo Nome Sobrenome do Autor e Terceiro
           Nome Sobrenome do Autor ; traduzido por Pri-
           meiro Nome Sobrenome do Tradutor e Segundo
           Nome Sobrenome do Tradutor. -- 3. ed. --
           Local de Publicação : Editora ; Outro local
           de publicação : Outra Editora, 2012.
           7.929 p. : il. ; 23 cm. -- (Aqui entra o
           nome da série ; 23)
```

```
Bibliografia.
Índice.
ISBN: 12-345-67890-123.
```

```
1. PRIMEIRO ASSUNTO 2. ASSUNTO - SEGUNDO
3. ASSUNTO COMPOSTO - PRIMEIRO COMPOSTO 4.
ASSUNTO COMPOSTO - SEGUNDO COMPOSTO I. Autor,
Segundo Nome Sobrenome do. II. Sobrenome do
Autor, Terceiro Nome. III. Sobrenome do Tradutor,
Primeiro Nome, trad. IV. Tradutor, Segundo
Nome Sobrenome do, trad. V. Título. VI.
Título: Aqui o subtítulo que, neste caso, é
muito importante também. VII. Série.
```

Fonte: Elaborado pelo próprio autor.

Buscou-se, com este roteiro, prever amplo espectro de situações para a representação do tipo documental *livro*. O processamento deste roteiro, considerado

complexo dada à multiplicidade de possibilidades de representação presentes, e considerando-se os ajustes necessários na *versão 1 do script* de conversão, apresentou, segundo interpretação da *versão 2 do script*, o resultado presente na Figura 43:

Figura 43 – Resultado de conversão de registro bibliográfico conforme versão 2 do *script*

```

=== MARC =====
020 $a1234567890123
090 $a123.456 $b5789a
1001 $aSobrenome do Autor, Primeiro Nome
24510$aAqui vai o título do livro $baqui o subtítulo que, neste caso, é muito importante
também $cPrimeiro Nome Sobrenome do Autor, Segundo Nome Sobrenome do Autor e Terceiro Nome
Sobrenome do Autor ; traduzido por Primeiro Nome Sobrenome do Tradutor e Segundo Nome
Sobrenome do Tradutor
2463 $aAqui o subtítulo que, neste caso, é muito importante também
250 $a3. ed.
260 $aLocal de Publicação $bEditora $aoutro local de Publicação $boutra Editora $c2012
300 $a7.929 p. $bil. $c23 cm.
4901 $aAqui entra o nome da série $b23
500 Bibliografia
500 Índice
65014$aPRIMEIRO ASSUNTO
65024$aASSUNTO $xSEGUNDO
65024$aASSUNTO COMPOSTO $xPRIMEIRO COMPOSTO
65024$aASSUNTO COMPOSTO $xSEGUNDO COMPOSTO
7001 $aAutor, Segundo Nome Sobrenome do
7001 $aSobrenome do Autor, Terceiro Nome
7001 $aSobrenome do Tradutor, Primeiro Nome $etrad.

```

Fonte: Elaborado pelo próprio autor.

Observe-se que, neste resultado, a codificação em Formato MARC21 Bibliográfico está muito próxima daquela feita por profissionais catalogadores. Percebe-se, entretanto, alguns problemas a serem saldados e refinados, tais como:

- a) falta de diferenciação entre título alternativo e título adicional:
  - i. identificação: ‘2463 \$aAqui o subtítulo que, neste caso, é muito importante também’;
  - ii. descrição: neste caso não deve ser usada a TAG 246, mas 740; como diferenciar esse caso se, nos pontos de acesso, não há uma forma de identificar a diferenciação? Compreende-se que, além do título principal registrado na manifestação, há a possibilidade de serem estabelecidos pontos de acesso secundários para título, um deles como título alternativo e outro como título adicional. Para título alternativo a informação semântica deverá ser agregada na TAG 246; para título adicional, em 740;
  - iii. solução estudada: se, nos pontos de acesso, a informação estiver igual à da descrição, considerar como conteúdo do campo 740 (e não 246). Note-se que, o uso do indicador 1

também deverá ser trabalhado diante da quantidade de caracteres que devem desconsiderados na alfabetação (mesma situação presente no indicador 2, da TAG 245);

b) não houve retirada de ponto final depois de unidade de medida, quando há informação da área de série:

- i. identificação: '\$c23 cm.';
- ii. descrição: não houve retirada de ponto final depois de unidade de medida, quando há informação da área de série. Em unidades de medida não há ponto. Nesse caso, o '.' é marcação de anteposição da série: '23 cm. -- (Aqui entra o nome da série ; 23)';
- iii. solução estudada: se estiver presente a informação 'cm. –' o sinal de '.' deverá ser tratado como parte de '.' e não parte de 'cm.';

c) a informação de número de série está em subcampo incorreto:

- i. identificação: '\$b23';
- ii. descrição: a informação de número de série está em subcampo incorreto; a designação de número ou volume em uma série deverá estar no '\$v', e não no '\$b';
- iii. solução estudada: corrigir código de subcampo no *script*, de '\$b' para '\$v';

d) não procedeu a diferenciação de notas gerais de notas de bibliografia e de índice e não inseriu código de subcampo:

- i. identificação: '500 Bibliografia' e '500 Índice';
- ii. descrição: notas de bibliografia e índice devem ser registradas na TAG 504, e não 500. Faltou, também, a indicação do código de subcampo;
- iii. solução estudada: verificar se, em notas (portanto com marcação de novo parágrafo), há conteúdo semântico específico, tais como 'bibliogr', ou 'índice' ou 'índex'. No caso de encontrar tais traços semânticos, sem que haja nova ocorrência de parágrafo, marcar com TAG 504. Para cada

ocorrência deverá ser usado código de '\$a': '504 \$aBibliografia' e '504 \$aÍndice'.

Diante da análise de resultado deste roteiro observa-se que outras alterações, diante de casos em que se verifica o ponto de acesso principal definido pelo título, e não pela indicação de responsabilidade, como apresentado nos casos anteriores. Observe-se, para tanto, o roteiro de descrição presente na Figura 44:

Figura 44 – Terceiro roteiro para o *script* de conversão de registros bibliográficos

```
123.456 Aqui vai o título do livro : aqui o sub-
A789      título que, neste caso, é muito importante
          também / organizado por Primeiro Nome
          Sobrenome do Organizador, Segundo Nome
          Sobrenome do Organizador e Terceiro Nome
          Sobrenome do Organizador. -- 3. ed. --
          Local de Publicação : Editora ; Outro Local
          de Publicação : Outra Editora, 2012.
          7.929 p. : il. ; 23 cm. -- (Aqui entra o
          nome da série ; 23)
```

```
Bibliografia.
Índice.
ISBN: 12-345-67890-123.
```

```
1. PRIMEIRO ASSUNTO 2. ASSUNTO - SEGUNDO
3. ASSUNTO COMPOSTO - PRIMEIRO COMPOSTO 4.
ASSUNTO COMPOSTO - SEGUNDO COMPOSTO I. Orga-
nizador, Primeiro Nome Sobrenome do, org. II.
Organizador, Segundo Nome Sobrenome do, org.
III. Organizador, Terceiro Nome Sobrenome do,
org. IV. Título: Aqui o subtítulo que, neste
caso, é muito importante também. V. Série.
```

Fonte: Elaborado pelo próprio autor.

Este roteiro, com interpretação baseada na *versão 2 do script*, que salda, mesmo que parcialmente, os casos previstos no segundo roteiro, apresenta resultado difuso em se tratando da submissão de arquivo com informações do terceiro roteiro, conforme pode ser observado na Figura 45:

Figura 45 – Resultado de conversão de registro bibliográfico conforme *versão 2 do script*

```

=== MARC =====
020 $a1234567890123
090 $a123.456 $bA789
1001 $aAqui vai o título do livro : aqui o sub-
24510$título que, neste caso, é muito importante também $corganizador por Primeiro
Nome Sobrenome do Organizador, Segundo Nome Sobrenome do Organizador e Terceiro
Nome Sobrenome do Organizador
2463 $aAqui o subtítulo que, neste caso, é muito importante também
250 $a3. ed.
260 $aLocal de Publicação $bEditora $aOutro Local de Publicação $bEditora $c2012
300 $a7.929 p. $bil. $c23 cm.
4901 $aAqui entra o nome da série $b23
500 Bibliografia
500 Índice
65014$aPRIMEIRO ASSUNTO
65024$aASSUNTO $xSEGUNDO
65024$aASSUNTO COMPOSTO $xPRIMEIRO COMPOSTO
65024$aASSUNTO COMPOSTO $xSEGUNDO COMPOSTO
7001 $aOrganizador, Primeiro Nome Sobrenome do $org.
7001 $aOrganizador, Segundo Nome Sobrenome do $org.
7001 $aOrganizador, Terceiro Nome Sobrenome do $org.

```

Fonte: Elaborado pelo próprio autor.

Este resultado também prevê necessidade significativa de alteração no *script*, no sentido de complementá-lo para que seja possível atender às outras especificidades da representação do tipo documental *livro*. Quanto a este novo diagnóstico, observa-se que não foi possível interpretar o ponto de acesso pelo título, tendo sido interpretado como ponto de acesso por indicação de responsabilidade:

- i. identificação: '1001 \$aAqui vai o título do livro';
- ii. descrição: a escolha do ponto de acesso principal depende dos elementos dispostos na fonte principal de informação do documento. Há casos em que o ponto de acesso é definido pelo autor (casos dos roteiros 1 e 2, e outros, com ponto de acesso pelo título, roteiro 3);
- iii. solução estudada: indicação de cláusulas de verificação que possam validar a oposição *ponto de acesso por indicação de responsabilidade x ponto de acesso por título*. Eis alguns estudos previstos para saldar o problema identificado:
  1. verificação da ocorrência de ',' (vírgula);
  2. verificação de palavras com três ou mais letras e inicial maiúscula;
  3. verificação da ocorrência de um ou mais dos seguintes caracteres: '/' (barra), ':' (dois pontos), '?' (interrogação), '!' (exclamação), '(' ou ')' (parênteses),

‘[’ ou ‘]’ (colchetes), ‘{’ ou ‘}’ (chaves), ‘+’ (mais), ‘=’ (igual).

no caso de as cláusulas 1 e 2 serem verdadeiras e a 3 ser falsa, então o registro deverá ser tratado como aquele que apresentada ponto de acesso por indicação de responsabilidade; caso contrário, ponto de acesso por título.

Tendo como base o terceiro roteiro, foi feita a adequação do *script*, agora em sua *versão 3*, para o tratamento da oposição *ponto de acesso por indicação de responsabilidade x ponto de acesso por título*, o qual apresenta o resultado evidenciado na Figura 46:

Figura 46 – Resultado de conversão de registro bibliográfico conforme *versão 3 do script*

```

=== MARC =====
020 $a1234567890123
090 $a123.456 $bA789
24510$aAqui vai o título do livro $baqui o subtítulo que, neste caso, é muito
importante também $corganizador por Primeiro Nome Sobrenome do Organizador,
Segundo Nome Sobrenome do Organizador e Terceiro Nome Sobrenome do Organizador
2463 $aAqui o subtítulo que, neste caso, é muito importante também
250 $a3. ed.
260 $aLocal de Publicação $bEditora $aOutro Local de Publicação $bEditora $c2012
300 $a7.929 p. $bil. $c23 cm.
4901 $aAqui entra o nome da série $b23
500 Bibliografia
500 Índice
65014$aPRIMEIRO ASSUNTO
65024$aASSUNTO $xSEGUNDO
65024$aASSUNTO COMPOSTO $xPRIMEIRO COMPOSTO
65024$aASSUNTO COMPOSTO $xSEGUNDO COMPOSTO
7001 $aorganizador, Primeiro Nome Sobrenome do $eorg.
7001 $aorganizador, Segundo Nome Sobrenome do $eorg.
7001 $aorganizador, Terceiro Nome Sobrenome do $eorg.

```

Fonte: Elaborado pelo próprio autor.

Observe-se que, neste resultado, a codificação em Formato MARC21 Bibliográfico mantém-se muito próxima daquela feita por catalogadores. Note-se que o problema identificado quanto ao ponto de acesso pelo título foi resolvido parcialmente, uma vez que não houve interpretação correta do indicador 1 da TAG 245, que, neste caso, deverá ser igual a ‘0’ (não exige entrada adicional no índice). Entretanto, problemas identificados anteriormente não foram tratados na *versão 3 do script*.

Diante da análise de resultado deste novo roteiro observa-se a necessidade de ampliar outros casos, tais como indicação de edição nos dados de localização do recurso informacional; indicação, em área de notas, de título original; e previsão de



casos em que há a indispensabilidade de indicação de dois ISBNs. Propõe-se, então, o quarto roteiro, presente na Figura 47:

Figura 47 – Quarto roteiro para o *script* de conversão de registros bibliográficos

```
123.456 Aqui vai o título do livro : aqui o sub-
A789      título que, neste caso, é muito importante
3.ed.     também / organizado por Primeiro Nome
          Sobrenome do Organizador, Segundo Nome
          Sobrenome do Organizador e Terceiro Nome
          Sobrenome do Organizador. -- 3. ed. --
          Local de Publicação : Editora ; Outro Local
          de Publicação : Outra Editora, 2012.
          7.929 p. : il. ; 23 cm. -- (Aqui entra o
          nome da série com texto extenso para compli-
          car o comportamento do script ; 23)
```

```
Título original: This is the English title.
Bibliografia.
Índice.
ISBN: 12-345-67890-123 (paperback)
ISBN: 12-345-67890-124 (hardback)
```

```
1. PRIMEIRO ASSUNTO 2. ASSUNTO - SEGUNDO
3. ASSUNTO COMPOSTO - PRIMEIRO COMPOSTO 4.
ASSUNTO COMPOSTO - SEGUNDO COMPOSTO I. Orga-
nizador, Primeiro Nome Sobrenome do, org. II.
Organizador, Segundo Nome Sobrenome do, org.
III. Organizador, Terceiro Nome Sobrenome do,
org. IV. Título: Aqui o subtítulo que, neste
caso, é muito importante também. V. Série. VI.
Título: This is the English title.
```

Fonte: Elaborado pelo próprio autor.

Este roteiro, com interpretação baseada na *versão 3 do script*, apresenta resultado conforme a Figura 48:

Figura 48 – Resultado de conversão de registro bibliográfico conforme *versão 3 do script*

```
==== MARC =====
020 $a1234567890124 (hardback).
090 $a123.456 $bA789
24510$aAqui vai o título do livro $baqui o subtítulo que, neste caso, é muito
importante 3.ed. também $corganizado por Primeiro Nome Sobrenome do Organizador,
Segundo Nome Sobrenome do Organizador e Terceiro Nome Sobrenome do Organizador
2463 $aAqui o subtítulo que, neste caso, é muito importante também
2463 $aThis is the English title
250 $a3. ed.
260 $aLocal de Publicação $bEditora $aOutro Local de Publicação $bEditora $c2012
300 $a7.929 p. $bil. $c23 cm.
4901 $aAqui entra o nome da série com texto extenso para complicar o comportamento
do script $b23
500 $aTítulo original: This is the English title
500 $aBibliografia
500 $aÍndice
65014$aPRIMEIRO ASSUNTO
65024$aASSUNTO $xSEGUNDO
65024$aASSUNTO COMPOSTO $xPRIMEIRO COMPOSTO
65024$aASSUNTO COMPOSTO $xSEGUNDO COMPOSTO
7001 $aOrganizador, Primeiro Nome Sobrenome do $eorg.
7001 $aOrganizador, Segundo Nome Sobrenome do $eorg.
7001 $aOrganizador, Terceiro Nome Sobrenome do $eorg.
=====
```

Fonte: Elaborado pelo próprio autor.



Este resultado também prevê necessidade significativa de alteração no *script*, no sentido de complementá-lo diante dos casos previstos no novo roteiro, que devem ser somados aos anteriores, não saldados:

- a) primeira ocorrência de ISBN não foi interpretada:
  - i. identificação: 'ISBN: 12-345-67890-123 (paperback)';
  - ii. descrição: existem recursos informacionais que contam com dois números de ISBN e é indispensável o registro dos dois, uma vez que não há a opção por um deles;
  - iii. solução estudada: indicação de condições no *script* que contemplem ocorrência maior do que um;
- b) informação da terceira linha dos dados de localização está entremeada nos referentes a outras informações sobre o título:
  - i. identificação: '\$baqui o subtítulo que, neste caso, é muito importante 3.ed. também';
  - ii. descrição: a informação '3.ed.' presente na terceira linha dos dados de localização, foi tratada como parte do título;
  - iii. solução estudada: tratar como parte do primeiro bloco de informações (dados de localização), portanto, como informação de começo de parágrafo, e não como sendo parte do texto, como é o caso do conteúdo 'Sobrenome do Organizador. -- 3. ed. --'.

Em respeito aos dados de localização observou-se, no quarto roteiro, que o primeiro bloco de informações, com destaque observável na Figura 49, pode conter até três linhas.

Figura 49 – Indicação de dados de localização de um registro bibliográfico

123.456	Aqui vai o título do livro : aqui o sub-
A789	título que, neste caso, é muito importante
3.ed.	também / organizado por Primeiro Nome Sobrenome do Organizador, Segundo Nome Sobrenome do Organizador e Terceiro Nome Sobrenome do Organizador. -- 3. ed. --

Fonte: Elaborado pelo próprio autor.

Estes dados devem ser trabalhados para que façam parte de um campo destinado às informações da própria unidade, identificado, aqui como campo 090. Cada elemento dos dados de localização deve ser destinado a um código de subcampo: '123.456' (dados de classificação de assunto) deve ser interpretado

como '\$a123.456'; 'S2789a' (dados de classificação de autor) deve ser interpretado como '\$bS789a'; e, '3.ed' (dados complementares de localização) deve ser interpretado como '\$c3.ed.'.

Grande dificuldade foi observada quanto à interpretação da terceira linha deste bloco de informações uma vez que pode ocorrer de a informação estar presente junto a estes elementos do registro e também em outro, na área de edição. Outra dificuldade quanto a este conjunto foi observada diante da necessidade de se buscar padrões para interpretação dos dados, dentre eles os presentes:

- a) na classificação de assunto (pela Classificação Decimal de Dewey – CDD – e pela Classificação Decimal Universal – CDU –, sendo que esta última adota, ainda, sinais gráficos);
- b) na classificação de autor (pela Tabela Cutter-Sanborn ou pela Tabela PHA) e, por fim;
- c) na possibilidade de a terceira linha deste conjunto (ou quarta, em alguns casos) estar ou não presente. Desse modo, ressalte-se que até esta versão do *script* não foi possível trabalhar elementos de classificação de assunto pela CDU.

Tendo como base o quarto roteiro, foi feita a adequação do *script*, agora em sua *versão 4*, para o tratamento da [1] diferenciação entre título principal, título adicional e título original e a indicação nas TAGs; [2] retirada de ponto final depois de unidade de medida, quando há informação da área de série; [3] correção de código de subcampo para a informação de número de série; [4] diferenciação de notas gerais de notas de bibliografia e de índice; [5] inserção de código de subcampo na nota de bibliografia e de índice; [6] interpretação da informação de terceira e quarta linhas dos dados de localização; [7] correção do indicador 1, da TAG 245, para o tratamento da entrada adicional do título no índice; e [8] interpretação de mais uma ocorrência de ISBN.

Diante da aplicação de novos testes foram obtidos os resultados indicados na Figura 50:

Figura 50 – Resultado de conversão de registro bibliográfico conforme *versão 4 do script*

```

=====
=== MARC =====
020 $a1234567890123 (paperback)
020 $a1234567890124 (hardback)
091 $123.456 $bA789 $c3.ed.
24500$aAqui vai o título do livro $baqui o subtítulo que, neste
caso, é muito importante também $corganizado por Primeiro Nome
Sobrenome do Organizador, Segundo Nome Sobrenome do Organizador
e Terceiro Nome Sobrenome do Organizador
250 $a3. ed.
260 $aLocal de Publicação $bEditora $aOutro Local de Publicação
$bEditora $c2012
300 $a7.929 p. $bil. $c23 cm
4900 $aAqui entra o nome da série com texto extenso para complicar
o comportamento do script $v23
500 $aTítulo original: This is the English title
504 $aBibliografia
504 $aÍndice
65014$aPRIMEIRO ASSUNTO
65024$aASSUNTO $xSEGUNDO
65024$aASSUNTO COMPOSTO $xPRIMEIRO COMPOSTO
65024$aASSUNTO COMPOSTO $xSEGUNDO COMPOSTO
7001 $aOrganizador, Primeiro Nome Sobrenome do $org.
7001 $aOrganizador, Segundo Nome Sobrenome do $org.
7001 $aOrganizador, Terceiro Nome Sobrenome do $org.
7400 $aAqui o subtítulo que, neste caso, é muito importante também
7400 $aThis is the English title
=====

```

Fonte: Elaborado pelo próprio autor.

Observe-se que, neste resultado, a codificação em Formato MARC21 Bibliográfico manteve-se correlacionada com aquela feita por catalogadores. Note-se a solução dos problemas anteriormente identificados.

Diante da análise de resultado entende-se que a diante da proposta de interpretação sintática e semântica de registros bibliográficos, o resultado tenha sido de qualidade.

Outra questão atinente, e que não foi proposta nos roteiros anteriores, diz respeito ao tratamento do conteúdo de termos qualificadores nos pontos de acesso de assunto tópico (TAG 650). Essa abordagem requer cuidado especial haja vista a possibilidade de ocorrência de subdivisões nos pontos de acesso de assunto, tais como:

- a) subdivisão de forma, usada para designar dicionários, periódicos, manuais, diretórios, atlas, enciclopédias, entre outros, identificada com código de subcampo '\$v';
- b) subdivisão cronológica, com destaque para ano, período de anos, ou século, identificada com código de subcampo '\$y';

- c) subdivisão geográfica, para o registro de cidades, estados, países, ou, ainda, regiões, para o qual se recorre ao uso do código de subcampo '\$z'; e
- d) subdivisão geral, para o registro de casos que não se enquadram nas condições acima.

Para saldar estes casos, prevê-se o uso de um dicionário semântico para cada uma das possibilidades de ocorrência dos termos qualificadores dos pontos de acesso de assunto. Entretanto, por ora, não há condições de se trabalhar nessa questão.

Estudam-se, ainda, duas implementações que requerem destaque:

- a implementação de processo de teste de consistência do arquivo convertido, uma vez que isso afere caráter qualitativo ao registro bibliográfico gerado;
- a integração do script com o OCR, o que exige a adoção de um que seja em linha de comando.

Ao final das análises consagradas aos testes de processamento de imagens dos registros bibliográficos analógicos, para o qual se recorreu aos aspectos sintáticos e semânticos dos registros bibliográficos, o capítulo seguinte destina-se às considerações finais da pesquisa.

## 5 CONSIDERAÇÕES FINAIS

Diante da temática definida para esta pesquisa, a conversão de registros bibliográficos para o Formato MARC21 Bibliográfico, desenvolveu-se a teoria sintática e semântica de registros bibliográficos, definida por *padrões de estrutura de metadados descritivos* e por *padrões de conteúdo*, consubstanciados a partir das ISBDs e das AACR2r.

Reiterou-se o encadeamento temático entre a pesquisa e a linha de pesquisa Informação e Tecnologia da área de concentração Informação, Tecnologia e Conhecimento do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, dada a abordagem teórico-conceitual acerca da representação de recursos informacionais e do compartilhamento e conversão de registros bibliográficos analógicos em ambiente digital; o desenvolvimento tecnológico alcançado diante da proposta de garantir a fidedignidade de aspectos da representação do conhecimento; e, a análise e o desenvolvimento de aspectos teóricos e metodológicos que subsidiem atividades de conversão de dados, recorrendo-se a métodos inerentes às tecnologias de informação e comunicação.

Reafirmou-se a importância da adoção de normas, padrões, formatos, metodologias e critérios para a representação de recursos informacionais em unidades de informação tendo em vista a aplicação de processos, permeados por aplicações tecnológicas e midiáticas que recorrem a estruturas computacionais que avalizem produção, organização, armazenamento, gerenciamento, tratamento, preservação, distribuição, provisão, recuperação, acesso, uso, reuso e compartilhamento de registros informacionais em diversos suportes.

Acredita-se ter saldado lacunas quanto às concepções teórico-metodológicas que regem o processo de conversão de dados bibliográficos para o Formato MARC21 Bibliográfico e ter encontrado respostas aos questionamentos indicados, uma vez que se verificou a hipótese de que, a partir de princípios e esquemas sintáticos e semânticos, inerentes aos padrões de estrutura de metadados descritivos e aos padrões de conteúdo adotados na representação de registros bibliográficos, é possível traçar-se sua interferência no processo de conversão para

o Formato MARC21 Bibliográfico. Verificou-se, também, que, diante da proposta teórica de sintaxe e semântica de registros bibliográficos, é possível garantir, em um interpretador computacional identificado como Scan for MARC, voltado à conversão de registros bibliográficos, as relações estudadas, a partir de Saussure, entre obra e manifestação, representado e representante, significado e significante, e, com Hjelmslev, as estruturas semânticas inerentes à catalogação e aos catálogos.

A tese desta pesquisa, de que os princípios sintáticos e semânticos de registros bibliográficos, definidos pelos esquemas de descrição e de visualização na catalogação e presentes nos padrões de estrutura de metadados descritivos e nos padrões de conteúdo, determinam o processo de conversão de registros bibliográficos para o Formato MARC21 Bibliográfico, foi possível de ser confirmada diante das discussões teórico-filosóficas, embasadas em Saussure e Hjelmslev.

A partir de Saussure estabeleceu-se, de modo análogo à compreensão de signo na comunicação humana, o modelo de representação dos recursos informacionais calcado na relação entre significado e significante, no qual se discute a arbitrariedade da *manifestação* em relação à *obra*, bem como o desenvolvimento da linearidade da *manifestação* em relação ao conteúdo ideacional da *obra*, fator determinante para a compreensão do documento e necessária à elaboração do registro bibliográfico.

Com Hjelmslev discutiu-se, a partir dos princípios formais adotados na Linguística para o estudo da estrutura do sistema lingüístico, a representação documental na Ciência da Informação, haja vista que esta se fundamenta nos constructos teóricos e sistêmicos da análise sincrônica do documento, com traços inerentes da indissociação entre *obra* e *manifestação*, para o qual se propõe o estudo dos registros bibliográficos a partir das relações internas entre os elementos de um registro (subsemântica), entre os registros de um catálogo (supersemântica), e em relação ao documento descrito (semântica).

Entende-se que objetivo geral de desenvolver um modelo teórico-conceitual de sintaxe e semântica em registros bibliográficos, a partir de estudos lingüísticos saussureanos e hjelmslevianos das manifestações da linguagem humana e aplicável a um interpretador computacional voltado à conversão de registros bibliográficos ao formato MARC21 Bibliográfico, tenha sido alcançado, bem como os objetivos

específicos estabelecidos nesta pesquisa, tendo em vista a opção pelo percurso metodológico adotado, que se mostrou pertinente e relevante.

Em decorrência das questões apresentadas procurou-se formar o arcabouço teórico-metodológico da representação bibliográfica e seus aspectos sintáticos e semânticos dos objetos representados, com reflexo nos seguintes aspectos:

- compreensão da *obra*, enquanto *significado*, e da *manifestação*, enquanto *significante*;
- entendimento da representação bibliográfica como resultado da relação entre *significante* e *significado* e entre *obra* e *manifestação* e como definitiva para a semântica;
- percepção da sintaxe para a definição da subsemântica, além de necessária à representação do recurso informacional;
- apreensão do conceito de supersemântica, a partir de sua co-dependência com a subsemântica e a semântica, na relação, identificável nos catálogos, entre registros bibliográficos e documentos de um acervo e entre *obra* e *manifestação*, tomados como *significante* e *significado*.

Tendo em vista os resultados obtidos nos testes, embora considerados iniciais, uma vez que ainda requer ajustes e aprimoramentos, realizados em prol da definição do comportamento do interpretador sintático e semântico de registros bibliográficos, o *Scan for MARC*, entende-se que a contribuição quanto aos aspectos sociais ensejados, de contribuir para o compartilhamento de registros bibliográficos, bem como para o amplo acesso e rapidez aos recursos informacionais representados pelas diversas unidades de informação, apresentem repercussão teórico-prática significativa na área da Ciência da Informação, bem como em sua interdisciplinaridade com a Ciência da Computação e a Lingüística.

Para estudos futuros, em se tratando do *Scan for MARC* retomem-se as necessidades identificadas para melhorias e aperfeiçoamento do interpretador computacional:

- a) conjugação integrada das fases de reprodução e de representação, recorrendo-se à adoção de um OCR em linha de comando;
- b) desenvolvimento de interface gráfica;

- c) previsão de padrões para a interpretação de dados de classificação de assunto (CDU);
- d) tratamento do conteúdo de termos qualificadores (subdivisão de forma, subdivisão cronológica, subdivisão geográfica e subdivisão geral) nos pontos de acesso de assunto tópico;
- e) previsão de padrões para tratamento de pontos de acessos secundários de assuntos para nome pessoal, institucional, evento e título uniforme;
- f) previsão de padrões para tratamento de pontos de acessos principais para instituições, eventos e título uniforme;
- g) previsão para criação de dicionários semânticos para pontos de acesso e seus termos qualificadores;
- h) implementação de processo de teste de consistência do arquivo convertido, uma vez que isso afere caráter qualitativo ao registro bibliográfico gerado;
- i) integração do script com o OCR, o que exige a adoção de um que seja em linha de comando.

Enseja-se, ainda, dar continuidade aos estudos dos métodos sintáticos e semânticos de registros bibliográficos e pesquisar a validade deste método de conversão de dados bibliográficos analógicos quando aplicados à interpretação dos dados da catalogação na fonte do tipo documental livro.

Nessa oportunidade, ressalta-se (cf. Apêndice M e N) a solicitação de registro da marca Scan for MARC, bem como o de programa de computador, via Agência UNESP da Inovação (AUIN), junto ao Instituto Nacional de Propriedade Industrial (INPI).



## REFERÊNCIAS

AGÊNCIA UNESP DE INOVAÇÃO. **Scan for MARC** [depósito de pedido de registro de programa de computador]. 2012a. Mensagem de Fabíola M. Spiandorello para Zaira Regina Zafalon, Plácida Leopoldina Ventura Amorim da Costa Santos, Patrícia Villar Martins, Karizi Cristina da Silva e Agência UFSCar de Inovação, com cópia para Leopoldo C. Zuaneti, enviada em 03 abr. 2012, às 10:57 AM.

AGÊNCIA UNESP DE INOVAÇÃO. **Scan for MARC** [depósito de pedido de registro da marca]. 2012b. Mensagem de Fabíola M. Spiandorello para Zaira Regina Zafalon, com cópia para Plácida Leopoldina Ventura Amorim da Costa Santos, enviada em 25 abr. 2012, às 16:37 PM.

APPLE. **iPhone**. 2011. Disponível em:  
<http://www.apple.com/br/iphone/features/camera.html>. Acesso em: 04 nov. 2011.

BORGMAN, C. L. From acting locally to thinking globally: a brief history of library automation. **The Library Quarterly**, v. 67, n. 3, p. 215-249, jul. 1997.

BORKO, H. Information science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5, jan. 1968.

BROTHER. **Manual do utilizador**: DCP-110C, DCP-310CN. Versão B. 2004. Disponível em:  
[http://welcome.solutions.brother.com/BSC/Public/files/dlf/doc000269/Mini2\\_DCP\\_PorUsr\\_B.pdf](http://welcome.solutions.brother.com/BSC/Public/files/dlf/doc000269/Mini2_DCP_PorUsr_B.pdf). Acesso em: 04 nov. 2011.

BUCKLAND, M. **Redesigning libraries services**: a manifesto. 1997. Disponível em:  
<http://sunsite.berkeley.edu/Literature/Library/Redesigning/bibaccess.html>. Acesso em: 30 abr. 2012.

CAFÉ, L.; SANTOS, C.; MACEDO, F. Proposta de um método para escolha de *software* de automação de bibliotecas. **Ciência da Informação**, Brasília, v. 30, n. 2, p. 70-79, maio/ago. 2001. Disponível em:  
<http://revista.ibict.br/index.php/ciinf/article/view/198/175>. Acesso em: 10 jan. 2012.

CASTRO, F. F.; SANTOS, P. L. V. A. C. Conversão retrospectiva de registros bibliográficos. In: FUJITA, M. S. L. (Org.) et. al. **A dimensão social da biblioteca digital na organização e acesso ao conhecimento**: aspectos teóricos e aplicados. São Paulo: DT/SIBi : IBICT, 2005. v. 2, p. 63-87.

CÓDIGO DE CATALOGAÇÃO ANGLO AMERICANO. 2. ed. rev. ampl. São Paulo: FEBAB/IOESP, 2004.

COLEMAN, A. S. From cataloging to metadata: Dublin Core records for the library catalog. In: SMIRAGLIA, R. P. (Ed.). **Metadata**: a cataloger's primer. New York: Haworth, 2005. p. 153-181.

CÔRTE, A. R. et al. Automação de bibliotecas e centros de documentação: o processo de avaliação e seleção de *softwares*. **Ciência da Informação**, Brasília, v. 28, n. 3, p. 241-256, set./dez. 1999. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/284>. Acesso em: 10 jan. 2012.

CÔRTE, A. R. et al. **Avaliação de softwares para bibliotecas e arquivos**. 2. ed. rev. ampl. São Paulo: Polis, 2002.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Brique de Lemos/Livros, 2008.

CUNHA, M. L. M. ISBD: origem, evolução e aceitação. **Revista Brasileira de Biblioteconomia e Documentação**, v. 12, n. 1/2, p. 7-14, jan./jun. 1979.

DANSKIN, A. Cataloging. In: ZUMER, M. (Ed.). **National bibliographies in the digital age**: guidance and new directions. Munchen: K. G. Saur, 2009. IFLA Series on Bibliographic Control, v. 39.

DENTON, W. **FRBR and Fundamental Cataloguing Rules**. 2009. Disponível em: <http://www.miskatonic.org/library/frbr.html>. Acesso em: 20 jan. 2009.

DZIEKANIAK, G. V. et al. Uso do padrão MARC em bibliotecas universitárias da Região Sul do Brasil. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. 26, jul./dez. 2008. Disponível em: <http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n26p188/6645>. Acesso em: 10 jan. 2012.

FERNANDES, A. M. R. **Inteligência artificial**: noções gerais. Florianópolis: VisualBooks, 2005.

FERREIRA, J. R. et al. Redes nacionais de informação, catalogação na fonte e outras experiências. **Revista Brasileira de Biblioteconomia e Documentação**, v. 12, n. 1/2, p. 67-88, jan./jun. 1979.

FOULONNEAU, M.; RILEY, J. Choosing metadata standards for a digital library project. In: \_\_\_\_\_. **Metadata for digital resources**: implementation, systems design and interoperability. Oxford: Chandos, 2008. p.13-28.

FREEDMAN, M. J. Must we limit the catalog? **Library Journal**, n. 15, p. 322-324, Feb. 1984.

FRÈGE, G. Sobre o sentido e a referência. In: \_\_\_\_\_. **Lógica e filosofia da linguagem**. São Paulo: Cultrix, 1978.

GARBERSON, E. Libraries, memory and the space of knowledge. **Journal of the History of Collection**, v. 18, n. 2, p. 105-136, 2006.

GARCÍA MARTÍNEZ, A. T. **Sistemas bibliotecários**: análisis conceptual y estructural. Gijón: Trea, 2005.

GARRIDO ARILLA, M. R. Procedimientos automáticos de creación y transformación de los registros bibliográficos. **Anales de Documentación**, n. 4, p. 127-137, 2001. Disponível em: <http://revistas.um.es/analesdoc/article/view/2361/2351>. Acesso em: 10 jan. 2012.

GARRIDO ARILLA, M. R. **Teoría e historia de la catalogación de documentos**. Madrid: Síntesis, 1999.

GEERAERTS, D. **Theories of lexical semantics**. New York: Oxford University Press, 2010.

GONSALVES, E. P. **Conversas sobre iniciação à pesquisa científica**. 5. ed. rev. e ampl. Campinas: Alínea, 2011.

HJELMSLEV, L. **Ensaio lingüísticos**. São Paulo: Perspectivas, 1991.

HOUAISS: dicionário da língua portuguesa. 2008. Disponível em: <http://houaiss.uol.com.br/busca.jhtm>. Acesso em: 10 dez. 2009.

HÜBNER, E. Conversão retrospectiva de registros bibliográficos. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 13., Recife, 2002. **Anais...** Recife: UFPE, 2002. 1 CD-ROM.

HUNTER, E. J.; BAKEWELL, K. G. B. **Cataloguing**. 2nd. rev. exp. ed. London: Clive Bingley, 1983.

IDICIONÁRIO Aulete. 2008. Disponível em: [http://aulete.uol.com.br/site.php?mdl=aulete\\_digital&op=loadVerbete&pesquisa=1&p\\_alavra=diacritico&x=8&y=12](http://aulete.uol.com.br/site.php?mdl=aulete_digital&op=loadVerbete&pesquisa=1&p_alavra=diacritico&x=8&y=12). Acesso em: 29 dez. 2011.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Declaração de princípios internacionais de catalogação**. 2009. Disponível em: [http://www.ifla.org/files/cataloguing/icp/icp\\_2009-pt.pdf](http://www.ifla.org/files/cataloguing/icp/icp_2009-pt.pdf). Acesso em: 20 jun. 2011.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Functional requirements for bibliographic records**: final report. 1998. Disponível em: <http://archive.ifla.org/VII/s13/frbr/frbr3.htm#6>. Acesso em: 20 jun. 2011. Aprovado pelo Standing Committee of the IFLA Section on Cataloguing em setembro de 1997 e alterado e corrigido até fevereiro de 2009.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Guidelines for Online Public Access Catalogue (OPAC) displays**: final report: may 2005. München: K. G. Saur, 2005. Recommended by the Task Force on Guidelines for OPAC Displays. Approved by the Standing Committee of the IFLA Cataloguing Section.

KAKKONEN, T. **Framework and resources for natural language parser evaluation**. Joensuu: University of Joensuu, 2007. Disponível em:

[http://books.google.com.br/books/about/Framework\\_and\\_Resources\\_for\\_Natural\\_La.html?hl=pt-BR&id=X4xog4y8alkC](http://books.google.com.br/books/about/Framework_and_Resources_for_Natural_La.html?hl=pt-BR&id=X4xog4y8alkC). Acesso em: 24 maio 2012.

KARISIDDAPPA, C. R. Co-operation to consortia: present status and future perspective with reference to Indian National Library. In: NIGAM, B. S.; KATARIA, S. **Digital libraries**: a festschrift volume of Professor R. K. Rout. New Delhi: Mahamaya, 2008. p. 282-298.

KRZYZANOVSKI, R. F. Cooperação em bibliotecas no Brasil: um panorama da década de 50 até nossos dias. **Revista Brasileira de Biblioteconomia e Documentação**, Nova Série, São Paulo, v. 3, n. 1, p. 1-24, jan./jun. 2007. Disponível em: <http://www.febab.org.br/rbbd/index.php/rbbd/article/view/27/32>. Acesso em: 10 jan. 2012.

KRZYZANOVSKI, R. F. et al. Implementação do Banco Dedalus, do Sistema Integrado de Bibliotecas da Universidade de São Paulo. **Ciência da Informação**, Brasília, v. 26, n. 2, p. 168-176, maio/ago. 1997. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/397/357>. Acesso em: 10 jan. 2012.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de metodologia científica**. 4. ed. rev. e ampl. São Paulo: Atlas, 2001.

LANGKER, R. ISBD: another step in the right direction. **The Australian Library Journal**, v. 23, n. 3, p. 99-103, April, 1974.

LOPES, E. **Fundamentos da linguística contemporânea**. 20. ed. São Paulo: Cultrix, 2008.

MARCONDES, C. H. **Automação de bibliotecas**. In: CURSO Aplicação de tecnologia ao desenvolvimento de bibliotecas. Rio de Janeiro: FGV, 1997.

MAXWELL, R. L. **FRBR**: a guide for perplexed. Chicago: ALA, 2008.

METSKER, S. J. **Building Parsers with JAVA**. Boston: Addison-Wesley, 2001.

MEY, E. S. A. **Acesso aos registros sonoros**: elementos necessários à representação bibliográfica de discos e fitas. 1999. 145 f. Tese (Doutorado em Ciências da Comunicação)—Universidade de São Paulo, São Paulo, 1999. Disponível em: <<http://www.conexaorio.com/biti/mey/>>. Acesso em: 03 set. 2008.

MEY, E. S. A. **Catálogo e descrição bibliográfica**: contribuições a uma teoria. 1986. 201 f. Dissertação (Mestrado em Biblioteconomia)—Universidade de Brasília, Brasília, 1986.

MEY, E. S. A. **Introdução a catalogação**. Brasília: Briquet de Lemos, 1995.

MEY, E. S. A. The item, the work and the object of cataloging. **Cataloging & Classification Quarterly**, New York, v. 26, n. 1, p. 45-62, 1998.

MEY, E. S. A.; SILVEIRA, N. C. **Catálogo no plural**. Brasília: Briquet de Lemos, 2009.

MEY, E. S. A.; ZAFALON, Z. R. Diversidade cultural aplicada ao código de catalogação: utopia ou necessidade? In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 23., 2009, Bonito. **Anais...** São Paulo: FEBAB, 2009. Disponível em: <http://hdl.handle.net/10760/15263>. Acesso em: 10 jan. 2012.

MILLER, P. Interoperability: what is it and why should I want it? **Ariadne**, n. 24, jun. 2000. Disponível em: <http://www.ariadne.ac.uk/issue24/interoperability/intro.html>. Acesso em: 10 jan. 2012.

MORALEJO ÁLVAREZ, R.; MARQUINA GARCÍA, J. L.; ABAD HIRALDO, R. Cooperación interbibliotecaria. **Boletín de la Anabad**, Madrid, v. 39, n.3-4, p. 571-601, 1989. Disponível em: [http://dialnet.unirioja.es/servlet/fichero\\_articulo?codigo=798159&orden=68773](http://dialnet.unirioja.es/servlet/fichero_articulo?codigo=798159&orden=68773). Acesso em: 10 jan. 2012.

NOERR, P. **The digital library toolkit**. 3rd ed. Santa Clara: Sun Microsystems, 2003. Disponível em: [http://www.ncsi.iisc.ernet.in/raja/is214/is214-2005-01-04/digital\\_library\\_toolkit-ed3.pdf](http://www.ncsi.iisc.ernet.in/raja/is214/is214-2005-01-04/digital_library_toolkit-ed3.pdf). Acesso em: 10 jan. 2012.

NÖTH, W. **A semiótica no século XX**. São Paulo: Annablume, 1996.

OLIVEIRA, N. M. et al. Compact disc cataloging – CatCD: análise de um instrumento para conversão retrospectiva no Sistema de Bibliotecas da UNICAMP. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 3, n. 1, p. 41-46, jan./jun. 1998. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/608/381>. Acesso em: 10 jan. 2012.

OTHERO, G. A.; MENUZZI, S. M. **Linguística computacional: teoria e prática**. São Paulo: Parábola, 2005.

PARANHOS, W. M. M. R. Fragmentos metodológicos para projetos e execução de gestão informatizada de coleções de documentos e serviços em bibliotecas. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. esp., jul./dez. 2004. Disponível em: <http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2004v9nesp2p14/5488>. Acesso em: 10 jan. 2012.

PEREZ, D. R.; LIMA, P. O projeto de conversão retrospectiva de registros bibliográficos: uma experiência do sistema de bibliotecas da Pontifícia Universidade Católica do Rio de Janeiro. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 12., 2002, Recife. **Anais**. 2002. Disponível em: <<http://www.sibi.ufrj.br/snbu/snbu2002/oralpdf/37.a.pdf>>. Acesso em: 24 jan. 2009.

PERL Programming Documentation. [2011]. Disponível em: <http://perldoc.perl.org/perl.html>. Acesso em: 24 maio 2012.

PRADO, N. S.; ABREU, J. Informatização das bibliotecas universitárias do Estado de Santa Catarina: cenário. **Revista ACB: Biblioteconomia em Santa Catarina**, v. 7, n. 1, 2002. Disponível em: <http://dici.ibict.br/archive/00000872/01/Rev%5B1%5D.AC-2005-92.pdf>. Acesso em: 24 jan. 2009.

RANGANATHAN, S. R.; GOPINATH, M. A. **Prolegomena to library classification**. 3<sup>rd</sup> ed. New York: Asia Publishing, 1967.

RÍOS HILARIO, A. B. Análisis de la funcionalidad de la parte descriptiva de las reglas de catalogación españolas mediante la aplicación del modelo entidad-relación (E-R). **Anales de documentación**, n. 10, p. 345-359, 2007.

ROBERT, Kirrily "Skud". **Introdução ao PERL**. [Brasil]: PERL Brasil, 2009. Disponível em: <http://www.perl.org.br/Perldoc/V500807/Perlintro>. Acesso em 24 maio 2012.

RUSSELL, S.; NORVIG, P. **Inteligência artificial**. Rio de Janeiro: Elsevier, 2004. Tradução da segunda edição.

SANTOS, P. L. V. A. C.; CORRÊA, R. M. R. **Catálogo**: trajetória para um código internacional. Niterói: Intertexto, 2009.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>. Acesso em: 10 jan. 2012.

SAUSSURE, F. **Curso de lingüística geral**. São Paulo: Cultrix, 2010.

SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, v. 27, p. 379-423, 623-656, Jul., Oct. 1948.

SHANNON, C. E. The mathematical theory of communication. In: SHANNON, C. E.; WEAVER, W. **The mathematical theory of communication**. Urbana: The University of Illinois Press, 1969.

SILVA, E. L.; MENEZES, E. M. Metodologia da pesquisa e elaboração de **dissertação**. 4. ed. rev. e ampl. Florianópolis: UFSC, 2005. Disponível em: <http://soniaa.arq.prof.ufsc.br/roteirosmetodologicos/metpesq.pdf>. Acesso em: 24 maio 2012.

SILVA, F. C. C.; FAVARETTO, B. Uso de *softwares* para o gerenciamento de bibliotecas: um estudo de caso da migração do sistema Aleph para o sistema Pergamum na Universidade de Santa Cruz do Sul. **Ciência da Informação**, Brasília, v. 34, n. 2, p. 105-111, maio/ago. 2005. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/662/576>. Acesso em: 10 jan. 2012.

SILVA, F. C. L.; ALVEZ, G.; VIAPIANA, N. Informatização da rede de bibliotecas da Secretaria Municipal de Educação de Florianópolis. **Revista ACB: Biblioteconomia**



em Santa Catarina, v. 13, n. 1, p. 211-222, jan./jun. 2008. Disponível em: <http://revista.acbsc.org.br/index.php/racb/article/view/549/674>. Acesso em: 10 jan. 2012.

SMIRAGLIA, R. P. Further Reflections on the Nature of 'A Work': An Introduction. **Cataloging & Classification Quarterly**, v. 33, n. ¾, p. 1-11, 2002.

SONY. **Cyber-shot W210**: câmara digital compacta. c2011. Disponível em: <http://www.sony.pt/product/dsc-w-series/dsc-w210#pageType=TechnicalSpecs>. Acesso em: 02 jan. 2012.

SOUZA, T. B.; CATARINO, M. E.; SANTOS, P. C. Metadados: catalogando dados na Internet. **Transinformação**, Campinas, v. 9, n. 2, p. 93-105, maio/ago. 1997.

SOUZA, T. B.; MOSTAFA, S. P. Catalogação cooperativa na rede BIBLIODATA/CALCO: a questão da repetitividade dos títulos no catálogo coletivo. **Transinformação**, Campinas: PUCCAMP, v. 11, n. 2, p. 127-133, maio/ago. 1999.

SPAULDING, C. S. ISBD: its origin, rationale, and implications. **Library Journal**, n. 15, p. 121-123, Jan. 1973.

STOCKTON, R. **PERL**: practical extraction and report language. [2005]. Disponível em: <http://www.stacken.kth.se/help/perl/>. Acesso em: 24 maio 2012.

SVENONIUS, E. Bibliographic objectives. In: \_\_\_\_\_. **The intellectual foundation of information organization**. Cambridge: MIT Press, 2000. p. 15-30.

SWANSON, G. ISBD: standard or secret? **Library Journal**, n. 15, p. 124-130, Jan. 1973.

TAMBA-MECZ, I. **A semântica**. São Paulo: Parábola, 2006.

TILLET, B. **O que é FRBR?**: um modelo conceitual para o universo bibliográfico. [2007?]. Disponível em: <http://www.loc.gov/catdir/cpsa/o-que-e-frbr.pdf>. Acesso em em: 24 maio 2012.

TRASK, R. L. **Dicionário de linguagem e lingüística**. São Paulo: Contexto, 2008.

VASCONCELLOS, P. A. G. Bibliodata/CALCO: informação bibliográfica para o desenvolvimento. **Ciência da Informação**, Brasília, v. 25, n. 3, p. 450-453, set./dez. 1996. Disponível em: <<http://www.ibict.br/cionline/250396/25039623.pdf>>. Acesso em: 12 ago. 2011.

WATSON, M. **Top five reasons why library administrators should support participation in the Program for Cooperative Cataloging**. 2001. Disponível em: <http://www.loc.gov/catdir/pcc/topfive.html>. Acesso em: 10 jan. 2012.

YOUNG, H., ed. **Glosario ALA de bibliotecología e ciencias de la información**. Madrid: Ediciones Diaz de Santos, 1988.

ZAFALON, Z. R. **Formato MARC21 Bibliográfico**: estudos e aplicações para livros, folhetos, folhas impressas e manuscritos. São Carlos: EdUFSCar, 2008.

ZAFALON, Z. R.; MEY, E. S. A. Códigos nacionales de catalogación: em busca de la propuesta perdida. In: ENCUENTRO DE CATALOGACIÓN Y METADADOS, 4., 2009, Ciudad de México. **Memória...** México, DF: UNAM/CUIB, 2010. p. 65-84.

Disponível em:

[http://132.248.242.3/~publica/archivos/libros/iv\\_encuentro\\_catalogacion.pdf](http://132.248.242.3/~publica/archivos/libros/iv_encuentro_catalogacion.pdf). Acesso em: 10 jan. 2012.

ZUMER, M.; LUPOVICI, C. Functionality and interface. In: ZUMER, M. (Ed.).

**National bibliographies in the digital age**: guidance and new directions. Munchen: K. G. Saur, 2009. IFLA Series on Bibliographic Control, v. 39.



**BIBLIOGRAFIA CONSULTADA**

- AHO, A. V.; SETHI, R.; ULLMAN, J. D. **Compilers: principles, techniques and tools**. Reading: Addison-Wesley, 1986.
- BAILEY JR.; C. W. Intelligent Library Systems: Artificial Intelligence Technology and Library Automation Systems. In: **Advances in Library Automation and Networking**, Greenwich, v. 4, 1991.
- BAPTISTA, A. A.; MACHADO, A. B. Um gato preto num quarto escuro: falando sobre Metadados. **Revista de Biblioteconomia de Brasília**, v. 25, n. 1, p. 77-90, jan./jun. 2001.
- BOGDAN, R.; BIKLEN, S. Investigação qualitativa em educação: uma introdução à teoria e aos métodos. Porto: Porto Editora, 1994. Resenha de: TEIXEIRA, E. [Resenhas]. **Cadernos de Textos de Educação**, Belém, v. 2, n. 2, p. 39-52, jun. 1997.
- BOUQUET, S. **Introdução à leitura de Saussure**. São Paulo: Cultrix, 2000.
- CALDAS, S. E. da S. **Elementos necessários à representação descritiva de partituras: um estudo com as Bachianas Brasileiras n.1, 2 e 4 de Heitor Villa-Lobos**. 2007. Trabalho de conclusão de curso (Graduação em Biblioteconomia)--Pontifícia Universidade Católica de Campinas, Campinas, 2007. Disponível em: <[http://eprints.rclis.org/archive/00013930/01/TCC\\_-\\_S%C3%A9rgio\\_-\\_Final.pdf](http://eprints.rclis.org/archive/00013930/01/TCC_-_S%C3%A9rgio_-_Final.pdf)>. Acesso em: 20 ago. 2008.
- CERVO, A. L.; BERVIAN, P. A. **Metodologia científica**. 5. ed. São Paulo: Prentice Hall, 2002.
- COELHO NETTO, J. T. **Semiótica, informação e comunicação**. São Paulo: Perspectiva, 1996.
- COELHO, H. **Inteligência artificial em 25 lições**. Lisboa: Fundação Calouste Gulbenkian, 1995.
- CONTERATTO, G. B. H. Semântica e computação: uma interação necessária para o aperfeiçoamento de sistemas PLN. **Letras de Hoje**, Porto Alegre, v. 41, n. 2, p. 353-367, jun. 2006.
- CREUS, S. Q. Apresentação: a lingüística computacional e suas linhas de investigação. **Letras de Hoje**, Porto Alegre, v. 41, nº 2, p. 5-8, junho, 2006.
- DEMO, P. Inteligência e complexidade: a propósito das idéias de Hillis. **Liinc em Revista**, v. 1, n. 2, p. 83-95, set. 2005. Disponível em: <<http://www.ibicit.br/liinc>>. Acesso em: 14 fev. 2009.

DUBEY, Y. P. Application of artificial intelligence (IA) in library and information systems. In: NIGAM, B. S.; KATARIA, S. **Digital libraries: a festschrift volume of Professor R. K. Rout.** New Delhi: Mahamaya, 2008. p. 112-129.

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, Brasília, v. 35, n. 1, p. 25-30, jan./abr. 2006.

FREYRE, E. A.; MARINHO, S. M. O. X. Uso das tecnologias da informação na biblioteca de Manguinhos: relato de experiência. **Revista Brasileira de Biblioteconomia e Documentação**, Nova Série, São Paulo, v.3, n.1, p.116-125, jan-jun. 2007.

FUCHS, C.; LE GOFFIC, P. **Les linguistiques contemporaines: repères théoriques.** Paris: Hachette Livre, 1996.

GALO, F. C. M.; SANTOS, P. L. V. A. C. Conversão retrospectiva. In: CONGRESSO DE INICIAÇÃO CIENTÍFICA, 16., 2008, São Carlos. **Anais de Eventos da UFSCar.** Disponível em: <http://ict2008.nit.ufscar.br/cic/uploads/C01/C01-002.pdf>. Acesso em: 17 nov. 2008.

GROSS, M.; LENTIN, A. **Introduction to formal grammars.** New York: Springer, 1970.

HOPCROFT, J. E.; ULLMAN, J. D.; MOTWANI, R. **Introdução à teoria de autômatos, linguagens e computação.** 2. ed. Rio de Janeiro: Elsevier, 2003.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Functional requirements for authority data.** 2007. Disponível em: <http://www.ifla.org/VII/d4/franar-conceptual-model-2ndreview.pdf>. Acesso em: 20 jun. 2008.

KRZYZANOWSKI, R. F. Cooperação e compartilhamento para o aperfeiçoamento dos serviços bibliotecários em bibliotecas universitárias. **Transinformação**, Campinas, v.10, n.1, p.15-32, jan./abr. 1998.

LIMA, G. A. B. *Softwares* para automação de bibliotecas e centros de documentação na literatura brasileira até 1998. **Ciência da Informação**, Brasília, v. 28, n. 3, p. 310-321, set./dez. 1999.

LOBATO, L. M. P. **A semântica na lingüística moderna: o léxico.** Rio de Janeiro: Livr. Francisco Alves, 1977.

LOUDEN, K. C. **Compiladores: princípios e práticas.** São Paulo: Pioneira Thomson Learning, 2004.

MARC Standards. Disponível em: <http://www.loc.gov/marc/>. Acesso em: 01 mar. 2008.

MELLO, G. N. Pesquisa educacional no Brasil. **Cadernos de Pesquisa**, São Paulo, n. 46, p. 67-72, ago. 1983.

MENDES, R. D. Inteligência artificial: sistemas especialistas no gerenciamento da informação. **Ciência da Informação**, Brasília, v. 26, n. 1, p. 39-45, jan./abr. 1997.

MENEZES, P. F. B. **Linguagens formais e autômatos**. 5. ed. Porto Alegre: Bookman, 2008.

MEY, E. S. A. **FRBR**: Functional requirements for bibliographic records. São Carlos: UFSCar, 2007. Palestra ministrada no curso Estudos e aplicações do AACR-2 e do Formato MARC21, em 22 ago. 2007.

MIZRAHI, V. V. **Treinamento em linguagem C++**. São Paulo: Makron Books, c1995. v.1.

MONTEIRO, S. D. A forma eletrônica do hipertexto. **Ciência da Informação**, Brasília, v. 29, n. 1, p. 25-39, jan./abr. 2000.

OTHERO, G. A. Lingüística computacional: uma breve introdução. **Letras de Hoje**, Porto Alegre, v. 41, n. 2, p. 341-351, jun. 2006.

PITTMAN, T.; PETERS, J. **The art of compiler design**: theory and practice. Englewood Cliffs: Prentice Hall, c1992.

RESMER, M. J.; COSTA, O. M. S. Conversão de base de dados MicroISIS para a Internet. **Ciência da Informação**, Brasília, v. 26, n. 2, n. 2, maio/ago. 1997.

ROCHA, R. P. Metadados, web semântica, categorização automática: combinando esforços humanos e computacionais para a descoberta e uso dos recursos da web. **Em Questão**, Porto Alegre, v. 10, n. 1, p. 109-121, jan./jun. 2004.

SANTOS, D. L.; SANTOS, P. L. V. A. C. Análise dos cinco formatos MARC21 (bibliográfico, autoridade, comunidade, coleção, classificação). In: CONGRESSO DE INICIAÇÃO CIENTÍFICA, 16., 2008, São Carlos. **Anais de Eventos da UFSCar**. Disponível em: <<http://ict2008.nit.ufscar.br/cic/uploads/C31/C31-005.pdf>>. Acesso em: 17 nov. 2008.

SARACEVIC, T. Interdisciplinary nature of information science. **Ciência da Informação**, Brasília, v. 24, n. 1, 1995. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/530/482>. Acesso em: 12 jan. 2011.

SAYÃO, L. F. Padrões para bibliotecas digitais abertas e interoperáveis. **Enc. Bibli.: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n. esp., jan./jun. 2007.

SCHIESSL, M. Ontologia: o termo e a idéia. **Enc. Bibli.: R. Eletr. Bibliotecon. Ciência da Informação**, Florianópolis, n. 24, p. 172-181, jul./dez. 2007.

SENNE, E. L. F. **Primeiro curso de programação em C**. 2. ed. Florianópolis: Visual Books, 2006.

TEIXEIRA, J. F. **O que é inteligência artificial**. São Paulo: Brasiliense, 1990. Coleção primeiros passos, 230.

VIANA, C. L. M. V. O impacto das inteligências artificiais na formação dos bibliotecários e cientistas da informação: revisão de literatura. **Ciência da Informação**, Brasília, v. 19, n. 1, p. 72-78, jan./jun. 1990.

VIEIRA, R.; LIMA, V. L. S. Lingüística computacional: princípios e aplicações. In: ESCOLA DE INFORMÁTICA DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 11, 2001. **Anais**. Passo Fundo: SBC-Sul, 2001. Disponível em: <http://www.inf.unioeste.br/~jorge/MESTRADOS/LETRAS%20-%20MECANISMOS%20DO%20FUNCIONAMENTO%20DA%20LINGUAGEM%20-%20PROCESSAMENTO%20DA%20LINGUAGEM%20NATURAL/ARTIGOS%20INTERESSANTES/lingu%EDstica%20computacional.pdf>. Acesso em: 10 jan. 2012.

WILLER, M. (Ed.) **UNIMARC manual authorities format**. 3<sup>rd</sup> ed. München: K. G. Saur, 2009.

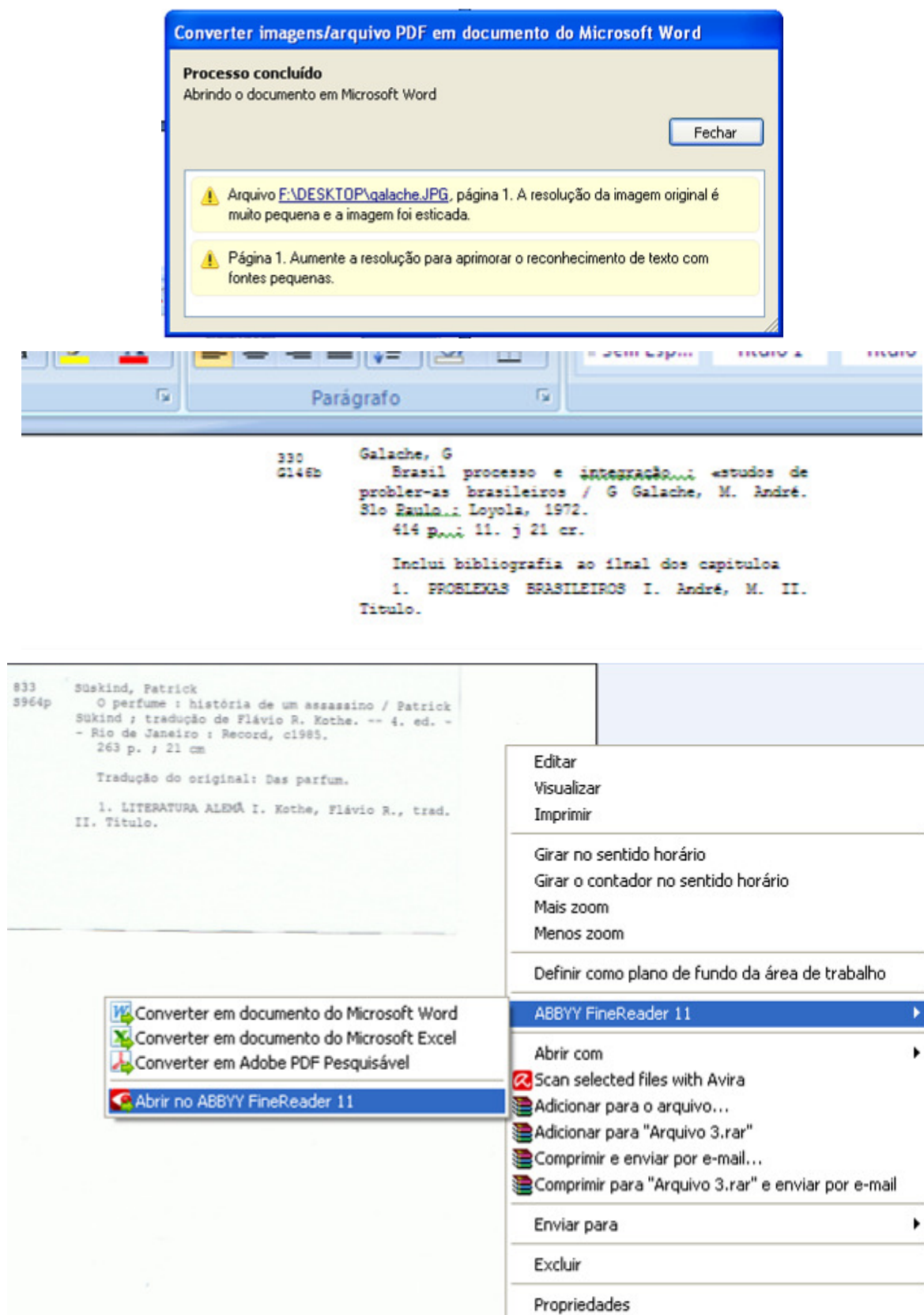
WINSTON, P. H.; HORN, B. K. P. **Lisp**. 2. ed. Reading: Addison-Wesley, 1984.

ZAFALON, Z. R.; SANTOS, P. L. V. A. C. Conversão de registros bibliográficos ao Formato MARC21 Bibliográfico a partir de análise sintática e semântica de registros descritos segundo as AACR2r e o RDA. In: ENCUENTRO DE CATALOGACIÓN Y METADADOS, 4., 2009, Ciudad de México. **Memória...** México, DF: UNAM/CUIB, 2010. p. 127-148. Disponível em: [http://132.248.242.3/~publica/archivos/libros/iv\\_encuentro\\_catalogacion.pdf](http://132.248.242.3/~publica/archivos/libros/iv_encuentro_catalogacion.pdf). Acesso em: 10 jan. 2012.

ZANDONADE, T. A cibernética aplicada ao controle bibliográfico. **UnB Contábil**, Brasília, v. 2, n. 1, p. 53-61, 1999.

## APÊNDICES

## APÊNDICE A – RESULTADO DO TESTE COM O ABBY FineReader 11



Fonte: Elaborado pelo próprio autor.

## APÊNDICE B – RESULTADO DO TESTE COM O ABBYY FineReader Online

abbyyonline.com/en/Task/Queue ★ - ↻ Google

🛒 Store **zafalon** : Profile | Log Off Eng | Pyc | Fra

🏠 Dictionary **FineReader OCR** Translation Interpreting Aligner

### ABBYY® FineReader Online

Flash on

**1. Load a file to process** ?

Upload...

The file must not be larger than 30 MB in size.

Send me a download link by e-mail

**2. Select the language(s) of your document** ?

Chinese Simplified

Chinese Traditional

Croatian

Czech

Danish

Dutch (Belgium)

Dutch (Netherlands)

English

These languages have been selected:  
**English** [x]

To select more than one language, hold the Ctrl key and click the required languages. Please do not select more than three languages.

[Click here](#) if your text is set in a black-letter typeface

**3. Select an output format** ?

Microsoft Word 97-2003 Document (\*.doc) ?

Select a format for the output file.

**Available: 3 pages**

**Recognize document**

Task history

Buy pages

---

**Information**

About service

FAQ


What is online OCR?

---

**Special Offers**

[Invite a friend](#)

Activate promo code



byyonline.com/en/Task/Queue ★ - ↻ Google

🛒 Store **zafalon** : Profile | Log Off Eng | Pyc | Fra

🏠 Dictionary **FineReader OCR** Translation Interpreting Aligner

### ABBYY® FineReader Online

Flash on

**1. Load a file to process** ?

Upload...

The file must not be larger than 30 MB in size.

Send me a download link by e-mail

**2. Select the language(s) of your document** ?

Norwegian (Bokmal)

Norwegian (Nynorsk)

Polish

Portuguese (Brazil)

Portuguese (Portugal)

Romanian

Russian

Slovak

These languages have been selected:  
**Portuguese (Brazil)** [x]

To select more than one language, hold the Ctrl key and click the required languages. Please do not select more than three languages.


[Click here](#) if your text is set in a black-letter typeface

**3. Select an output format** ?

Plain text (\*.txt) ?

Select a format for the output file.

**4. Select files to process** ?

 **Document:** Arquivo 3.jpg ( 309 KiB ) 🗑 delete

**Available: 3 pages**

**Recognize document**

Task history

Buy pages

---

**Information**

About service

FAQ


What is online OCR?

---

**Special Offers**

[Invite a friend](#)

Activate promo code



**ABBYY® FineReader 11**

Learn more

bvryonline.com/en/Task/Queue

Norwegian (Bokmål)  
Norwegian (Nynorsk)  
Polish  
Portuguese (Brazil)  
Portuguese (Portugal)

To select more than one language, hold the Ctrl key and click the required languages. Please do not select more than three languages.

Click here if your text is set in a black-letter typeface

**3. Select an output format**

Plain text (\*.txt)

Select a format for the output file.

**4. Select files to process**

Document: Arquivo 3.jpg ( 309 KiB ) [delete](#)  
Number of pages: 1


Document: Arquivo 1.JPG (1 334 KiB ) [delete](#)  
Number of pages: 1

Document: Arquivo 9.jpg ( 321 KiB ) [delete](#)  
Number of pages: 1

Estimated processing time: less than one minute [Recognize](#)

**Special Offers**

[Invite a friend](#)  
Activate promo code



**ABBY FineReader 11**

[Learn more](#)

[f](#) [t](#) [Curir](#) €175

All ABBYY products [About](#) [Terms of use](#) [Privacy](#) [Feedback](#)

bvryonline.com/en/Task/Success

[Store](#) [zafalon](#) : Profile | [Log Off](#) [Eng](#) | [Pyc](#) | [Fra](#)

[Dictionary](#) [FineReader OCR](#) [Translation](#) [Interpreting](#) [Aligner](#)

## ABBY® FineReader Online

### Congratulations!

Your task has been accepted. Once the processing is complete, the resulting document(s) can be downloaded from your [Task history](#). You will also receive an e-mail notification if you selected the corresponding option. The results will be available for download for 72 hours.

Estimated processing time: **less than one minute**

[Process another file](#)

[View available tasks](#)

**Available: 0 pages**

[Recognize document](#)  
[Task history](#)  
[Buy pages](#)

#### Information

[About service](#)  
[FAQ](#)  
[What is online OCR?](#)

#### Special Offers

[Invite a friend](#)  
[Activate promo code](#)

















byyonline.com/en/Task/History/

Store | zafalon : Profile | Log Off | Eng | Pyc | Fra

Dictionary **FineReader OCR** Translation Interpreting Aligner

## ABBYY® FineReader Online

Documents

<input type="checkbox"/>	 <a href="#">Arquivo 9.txt</a> 299 B	09:58 today		<a href="#">delete</a>	 
<input type="checkbox"/>	 <a href="#">Arquivo 1.txt</a> 261 B	09:58 today		<a href="#">delete</a>	 
<input type="checkbox"/>	 <a href="#">Arquivo 3.txt</a> 264 B	09:58 today		<a href="#">delete</a>	 

[Download all the selected files](#)

**Available: 0 pages**

[Recognize document](#)

**Task history**

[Buy pages](#)

**Information**

[About service](#)


[FAQ](#)

[What is online OCR?](#)

**Special Offers**

[Invite a friend](#)

[Activate promo code](#)



finereader.abbyyonline.com/en/Task/History/

Store | zafalon : Profile | Log Off | Eng | Pyc | Fra

Dictionary **FineReader OCR** Translation Interpreting Aligner

## ABBYY® FineReader Online

Documents

**Available: 0 pages**

**Arquivo 1 - Bloco de notas**

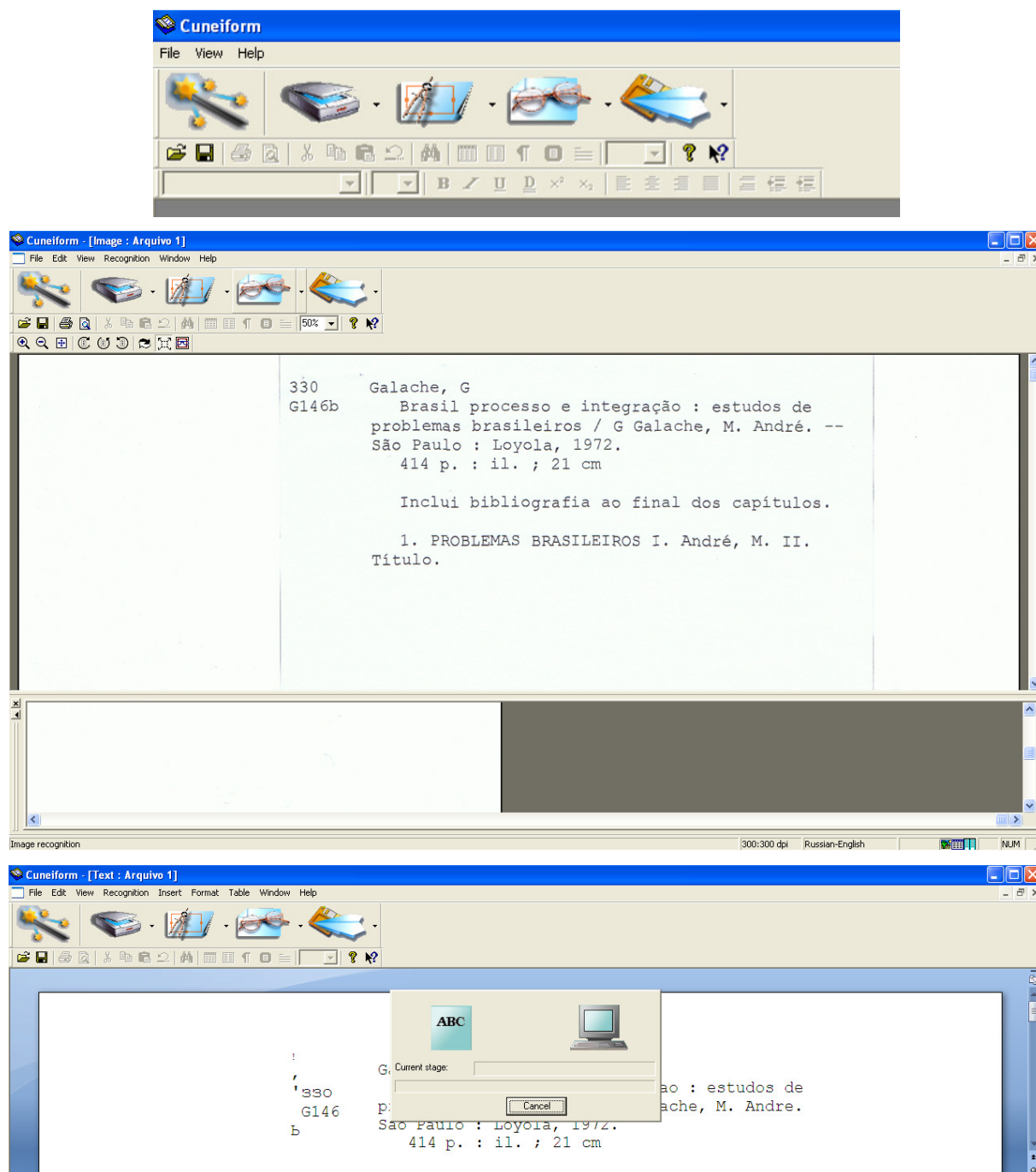
Arquivo Editar Formatar Exibir Ajuda

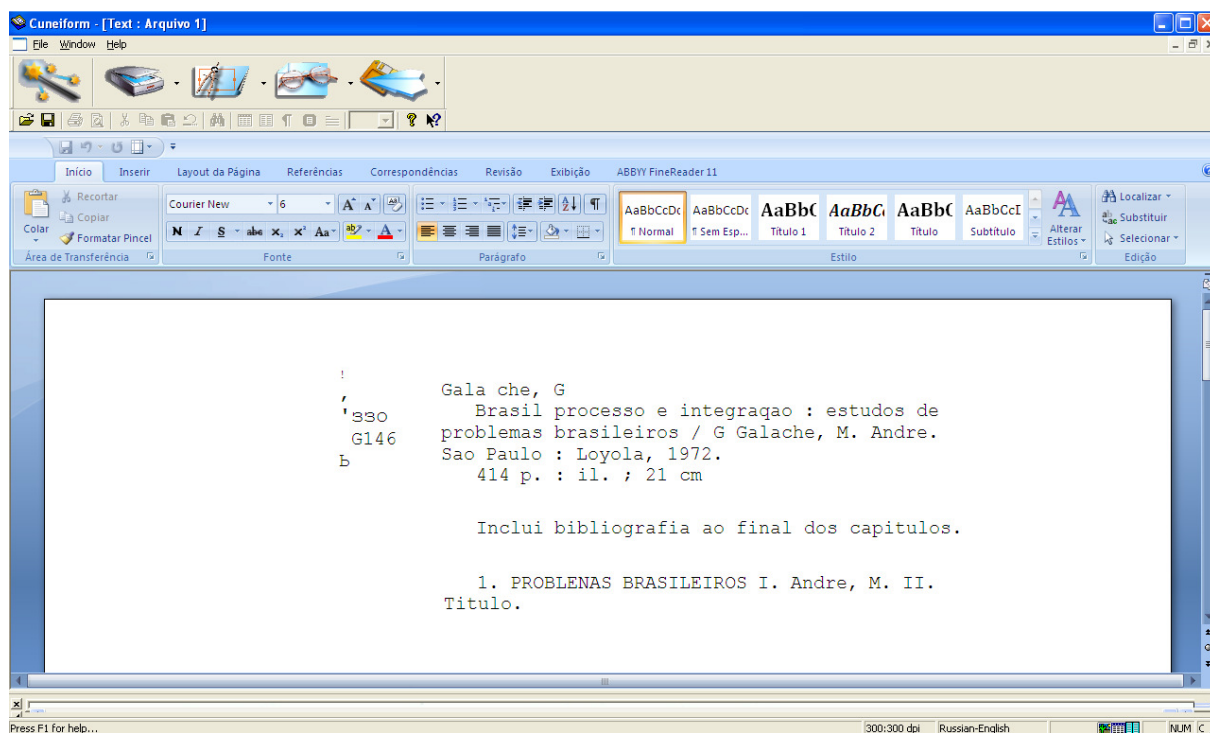
330 G14 6b  
Galache, G  
Brasil processo e integração : estudos de problemas brasileiros / G Galache, M. André. São Paulo : Loyola, 1972.  
414 p. : il. ; 21 cm  
Inclui bibliografia ao final dos capitulos  
1. PROBLEMAS BRASILEIROS I. André, M. II. Título.

Fonte: Elaborado pelo próprio autor.

## APÊNDICE C – RESULTADO DO TESTE COM O Cognitive Open OCR (Cuneiform)

## 0.1





Fonte: Elaborado pelo próprio autor.

## APÊNDICE D – RESULTADO DO TESTE COM O FreeOCR

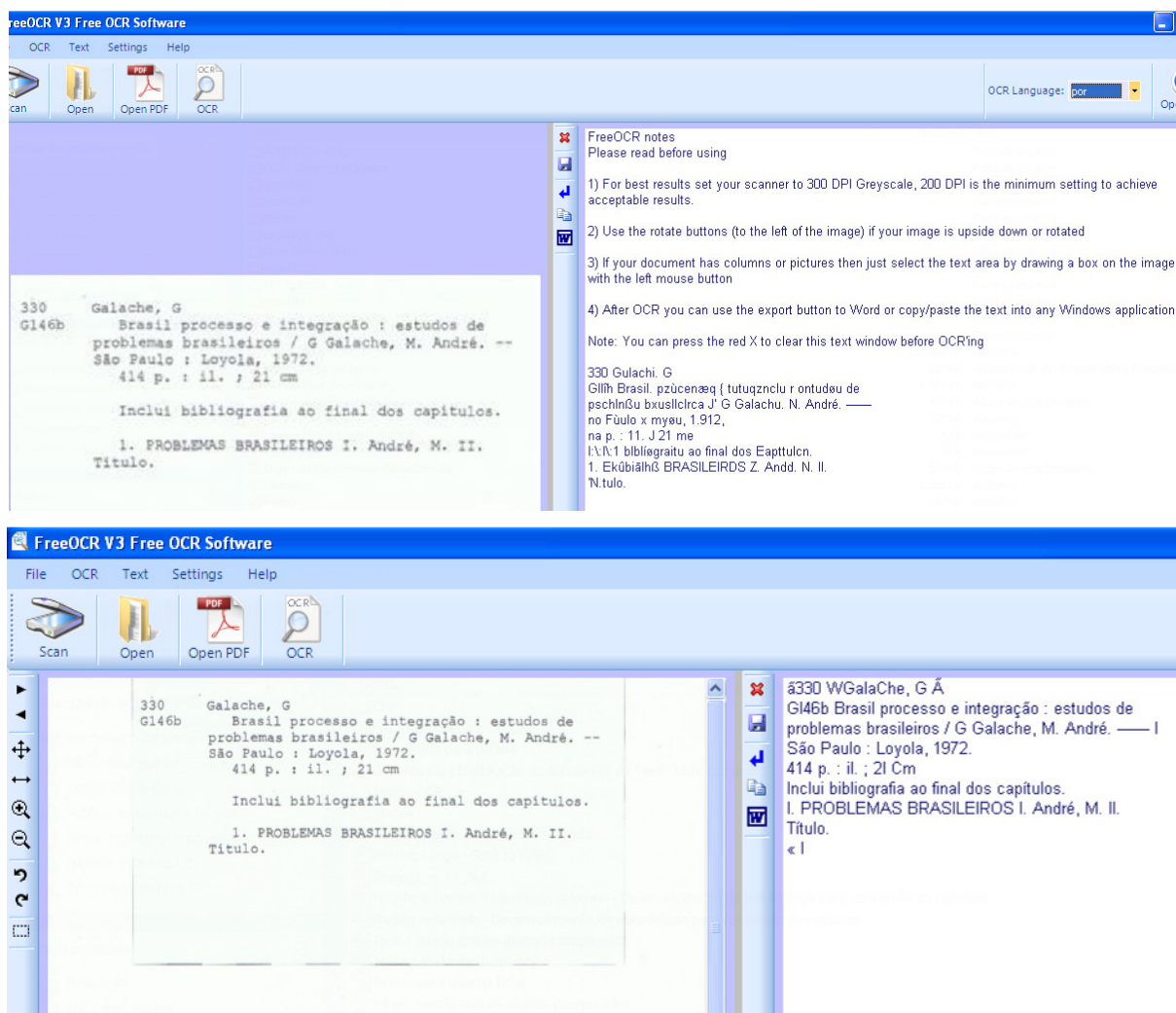


The screenshot shows a web browser window with the address bar displaying "www.free-ocr.com". The page title is "Free OCR". Below the title is a navigation menu with links for "Home", "About", "FAQ", "Privacy", and "Contact". The main content area is titled "Text recognition result:" and contains a scrollable box with the following text:

1 |  
Á330 Galache, G  
G146b Brasil processo e integração : estudos de  
problemas brasileiros / G Galache, M. André. --  
São Paulo : Loyola, 1972. `.  
414 p. : il. ; 21 cm \  
Inclui bibliografia ao final dos capítulos.  
1. PROBLEMAS BRASILEIROS I. André, M. II.  
Título.  
â

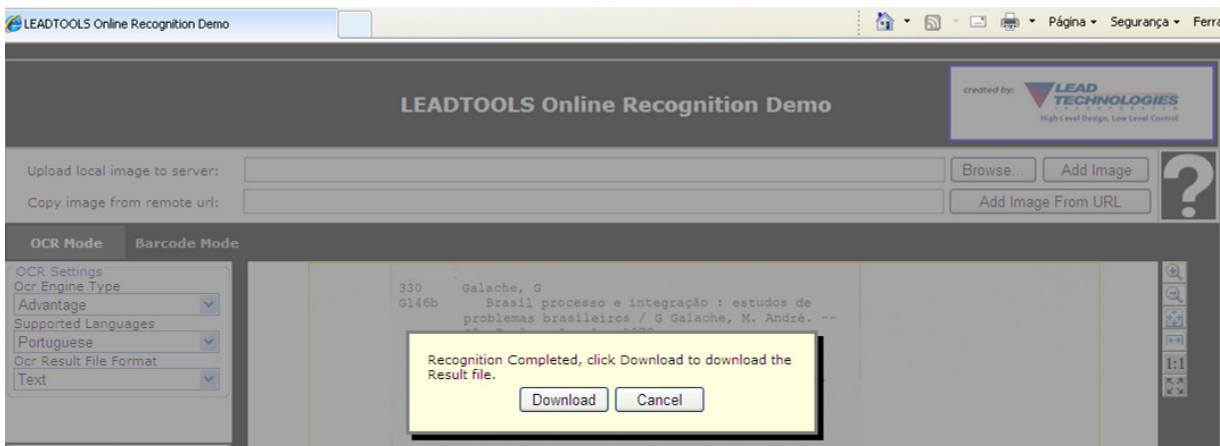
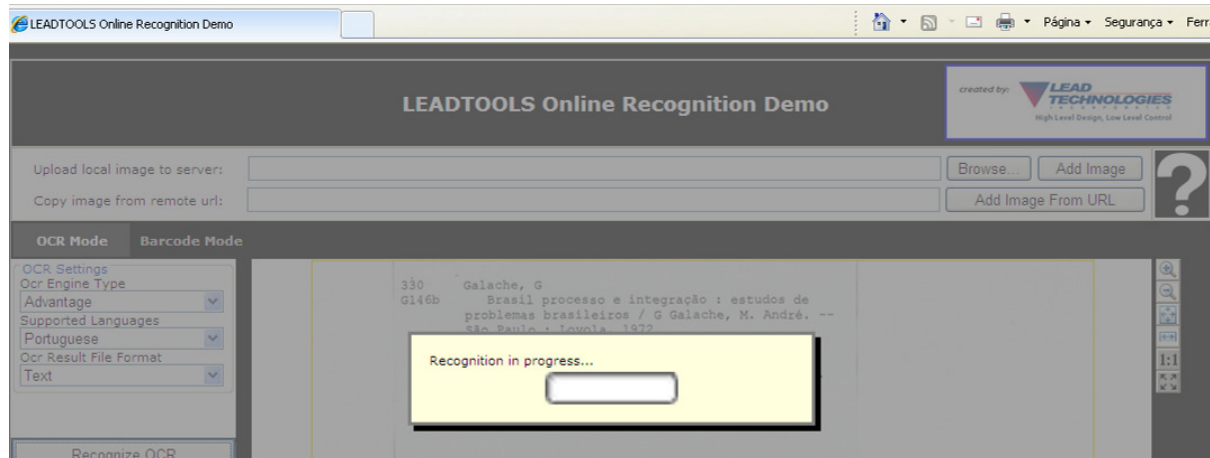
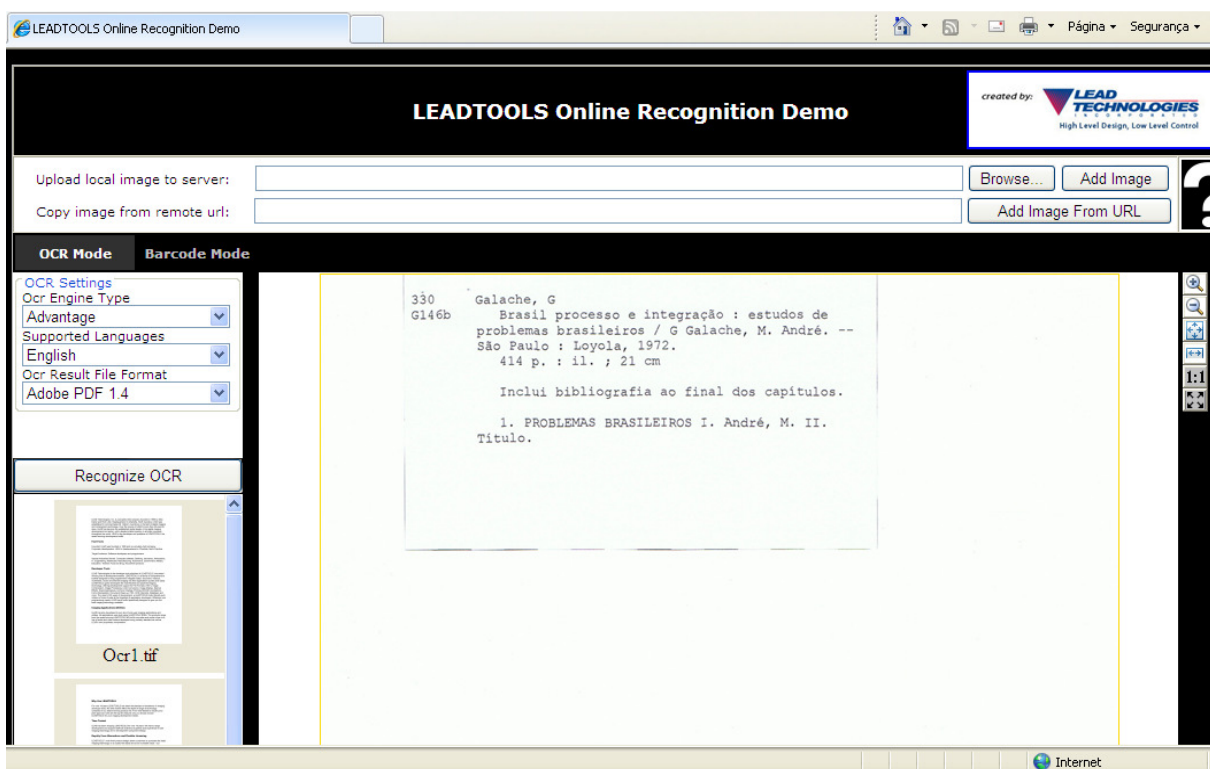
Fonte: Elaborado pelo próprio autor.

## APÊNDICE E – RESULTADO DO TESTE COM O FreeOCR 3.1



Fonte: Elaborado pelo próprio autor.

## APÊNDICE F – RESULTADO DO TESTE COM O Leadtools



The screenshot shows the LEADTOOLS Online Recognition Demo interface. The top navigation bar includes "Conectando...", "Página", "Segurança", and "Ferramentas". The main header displays "LEADTOOLS Online Recognition Demo" and the LEAD TECHNOLOGIES logo. Below the header, there are input fields for "Upload local image to server:" and "Copy image from remote url:", along with "Browse...", "Add Image", and "Add Image From URL" buttons. A "Recognize OCR" button is also visible.

The "OCR Settings" panel on the left includes:

- OCR Mode: Barcode Mode
- OCR Settings
- Ocr Engine Type: Advantage
- Supported Languages: Portuguese
- Ocr Result File Format: Text

A "Download de Arquivos" dialog box is open, asking "Deseja salvar ou abrir este arquivo?". The file details are:

- Nome: Ocr.txt
- Tipo: Documento de texto, 239 bytes
- Origem: demo.leadtools.com

The dialog also contains a warning: "Embora arquivos provenientes da Internet possam ser úteis, alguns arquivos podem danificar seu computador. Se você não confiar em sua origem, não abra nem salve este arquivo. [Qual é o risco?](#)"

The OCR result is displayed in a text area, showing the following text:

```
330 Galache, G
G146b Brasil processo e integração : estudos de
      problemas brasileiros / G Galache, M. André. --
      São Paulo : Loyola, 1972.
      414 p. : il. ; 21 cm

      Inclui bibliografia ao final dos capítulos.

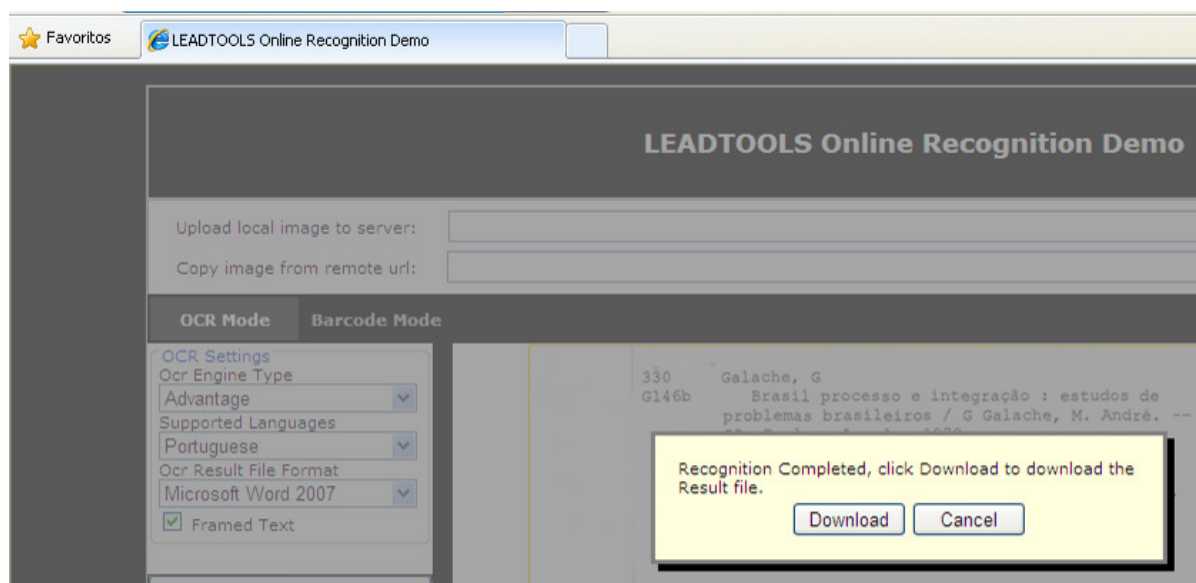
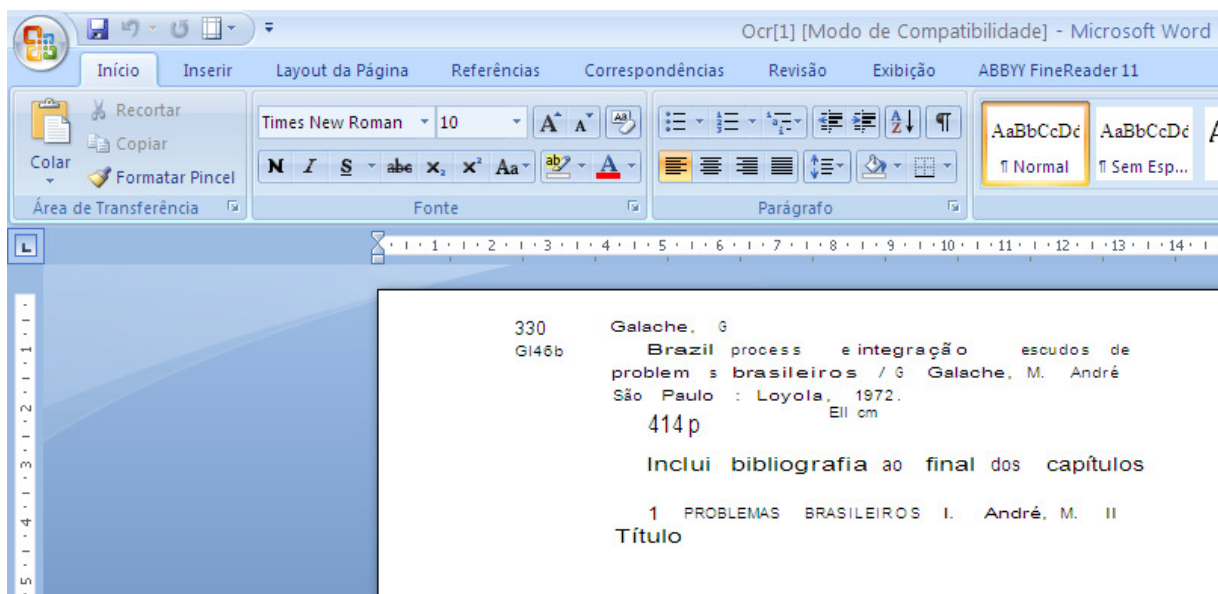
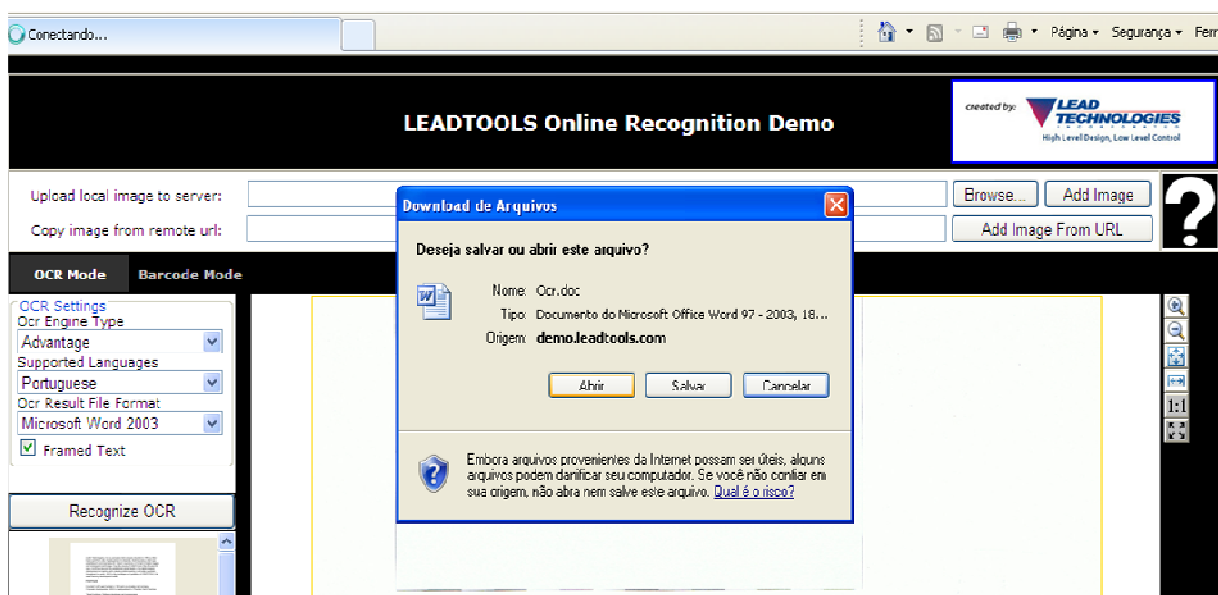
      1. PROBLEMAS BRASILEIROS I. André, M. II.
      Título.
```

## Ocr[1] - Bloco de notas

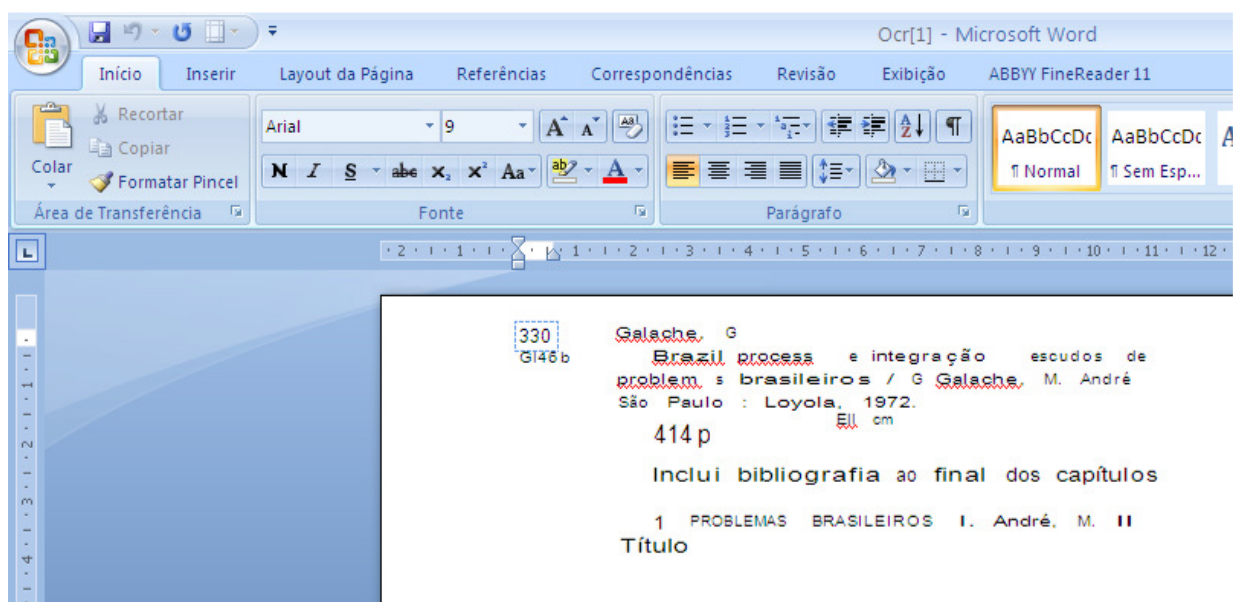
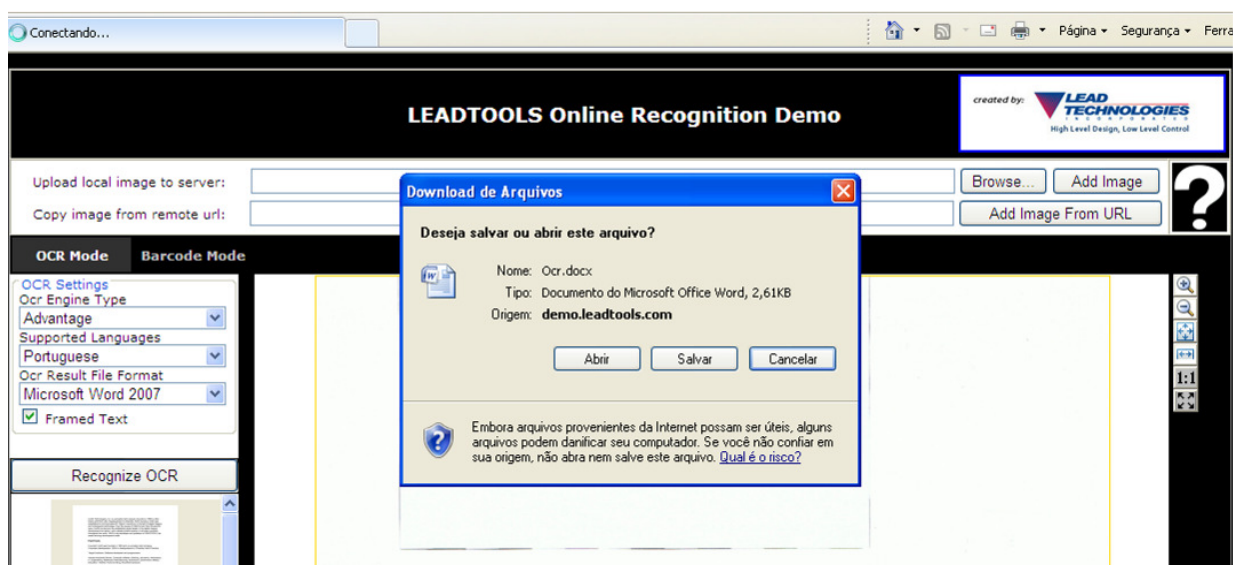
Arquivo Editar Formatar Exibir Ajuda

```
330 Galache, G
G146b Brazil process e integração escudos de
problem s brasileiros / G Galache, M. André
São Paulo : Loyola, 1972.
414 p E11 cm
Inclui bibliografia ao final dos capítulos
1 PROBLEMAS BRASILEIROS I. André, M. II
Título
```









Fonte: Elaborado pelo próprio autor.

## APÊNDICE G – RESULTADO DO TESTE COM O OnlineOCR.net

www.onlineocr.net/default.aspx

licença freeware

Login

Forgot password?  
Create Account

What is new ?

**New Feature !**  
OCR Web Service  
[learn more...](#)

Email OCR  
[learn more...](#)

OnlineOCR.net is a Free Online OCR (Optical Character Recognition software) service that allows you to convert scanned PDF and images (jpg, bmp) into editable Word, Text, Excel, PDF, Html output formats.

without installation on your computer. Recognize text and characters from PDF scanned documents (including multipage files), photographs and digital camera captured images. Free Online OCR service supports 32 recognition languages.

**Convert scanned PDF to Word, Txt, Rtf**

Extract text from image (JPG, JPEG, BMP, TIFF, GIF) and convert into editable Word, Text, Excel, PDF, Html output formats.  
Converted documents look exactly like the original - tables, columns and graphics.

**Free**

OnlineOCR.net is a free service in a "Guest mode" (without registration) that allows you to convert 15 images per hour.

Max file size 4 mb.

Selecionar Upload

Size: 1.37 MB

File: Arquivo 1.JPG

Recognition language: PORTUGUESE

Output format: Text Plain (txt)

Recognize

330 Galache, G G146b Brasil processo e integração : estudos de problemas brasileiros / G Galache, M. André. -- São Paulo : Loyola, 1972. 414 p. : il. ; 21 cm  
Inclui bibliografia ao final dos capítulos.  
1. PROBLEMAS BRASILEIROS I. André, M. II. Título.

1112290948\_CCF01072011\_00000 [Modo de Compatibilidade] - Microsoft Word

Início Inserir Layout da Página Referências Correspondências Revisão Exibição ABBYY FineReader 11

Recortar Copiar Colar Formatar Pincel Área de Transferência

Courier New 9

Fonte

Parágrafo

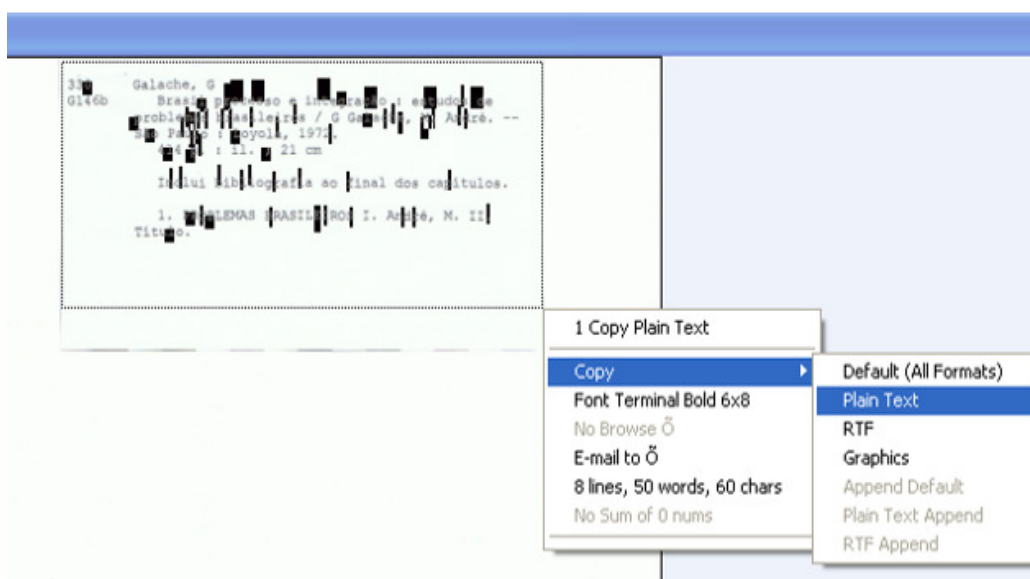
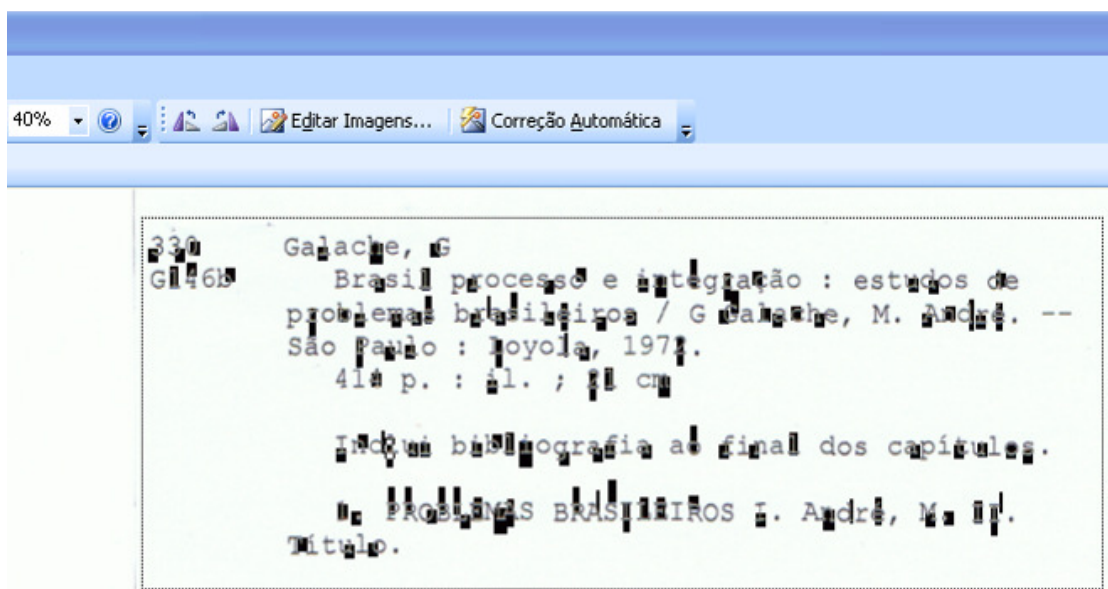
AaBbCcDd AaBbCcDd

¶ Sem Esp... Título 1

330 Galache, G  
G146b Brasil processo e integração : estudos de problemas brasileiros / G Galache, M. André. -- São Paulo : Loyola, 1972. 414 p. : il. ; 21 cm  
Inclui bibliografia ao final dos capítulos.  
1. PROBLEMAS BRASILEIROS I. André, M. II. Título.

Fonte: Elaborado pelo próprio autor.

## APÊNDICE H – RESULTADO DO TESTE COM O ScreenOCR 9.1



Õ  
 Õ -.. .đ . . . .  
 Õ . . . . . Õ . . - ^Õ. . .  
 ù ..Õ. Õ . . . .  
 Õ Õ. Õ  
 .. . . . . . . . . .  
 Õ.đ . . . . . . . . .  
 Õ

Fonte: Elaborado pelo próprio autor.

## APÊNDICE I – RESULTADO DO TESTE COM O Sci2ools (i2OCR)

Online OCR - Sciweavers

tp://www.sciweavers.org/free-online-ocr

Yahoo! Google Maps YouTube Wikipedia Notícias Popular

**Sciweavers**

i2OCR  
The Free Online OCR

- Extract Text from Document Image
- Supports (TIF, JPEG, PNG, BMP, GIF, PBM, PGM, PPM)
- 33 Languages (English, French, Italian, Chinese, +more)
- Supports Multi-column Document
- Unlimited Uploads
- No Registration, No Email

Spread the Word [Tweet](#) 183 [Curir](#) 559 [+1](#) 39 [Maximize](#)

Image Language Text

File  URL

Selecione Arquivo galache.JPG

Portuguese

**Convert**

Online OCR - Sciweavers

www.sciweavers.org/free-online-ocr

Yahoo! Google Maps YouTube Wikipedia Notícias Popular

Image Language Text

File  URL

Selecione Arquivo galache.JPG

Portuguese

**Convert**

### Download Text

<p>330 Glib Gnlchl. G Illil pzocuo 0 Intiqucl: 1 ucudou de probluu bxunciru l6 Guincho, N. Andzi. sla Paulo x myoh, 1972. IN p. : 11. 1 21 Q Xaclu Mbuoqnlln no 11:11 ua uptrulol L PMLDMS IIAQILBIMS 2. Nuká, N. XX. Titulo.</p>	<p>330 Galache, G G146b Brasil processo e integração : estudos de problemas brasileiros / G Galache, M. André. -- São Paulo : Loyola, 1972. 414 p. : il. ; 21 cm  Inclui bibliografia ao final dos capitulos.  1. PROBLEMAS BRASILEIROS I. André, M. II. Titulo.</p>
---	--

**Image**

File  
 URL

Escolher arquivo | Arquivo 1.JPG

**Language**

Portuguese

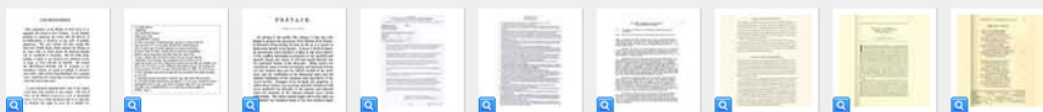
**Text**

Convert

Recognizing Text, Please Wait...

For better recognition, image must be at least 200 dpi. Screen shot is only 72 dpi | Downloaded text file may require a special editor to read if not English | Maximum image size is 10MB

**Sample OCR Test Images** Select image thumbnail then click the **Orange** button. To preview a sample file, click the zoom icon.



**Image**

File  
 URL

Escolher arquivo | Arquivo 1.JPG

**Language**

Portuguese

**Text**

Convert

## Download Text

330  
G146b  
Galache, G  
Brasil processo e integração : estudos de  
problemas brasileiros / G Galache, M. André.  
São Paulo : Loyola, 1972.  
414 p. z il. ; 21 cm  
Inclui bibliografia ao final dos capitulos  
1. PROBLEMAS BRASILEIROS I. André, M. II.  
Título.

Fonte: Elaborado pelo próprio autor.

## APÊNDICE J – RESULTADO DO TESTE COM O SimpleOCR 3.5

SimpleOCR [Machine-Print-English (U.S.)-Upper and Lower]

File View Process Tools Help

File Edit Undo Redo Find Print Page Setup Help

File Edit Undo Redo Find Print Page Setup Help

330 Galache, G

G146b Brasil processo e integração : estudos de  
problemas brasileiros / G Galache, M. André. --  
São Paulo : Loyola, 1972.  
414 p. : il. ; 21 cm

Inclui bibliografia ao final dos capítulos.

1. PROBLEMAS BRASILEIROS I. André, M. II.  
Título.

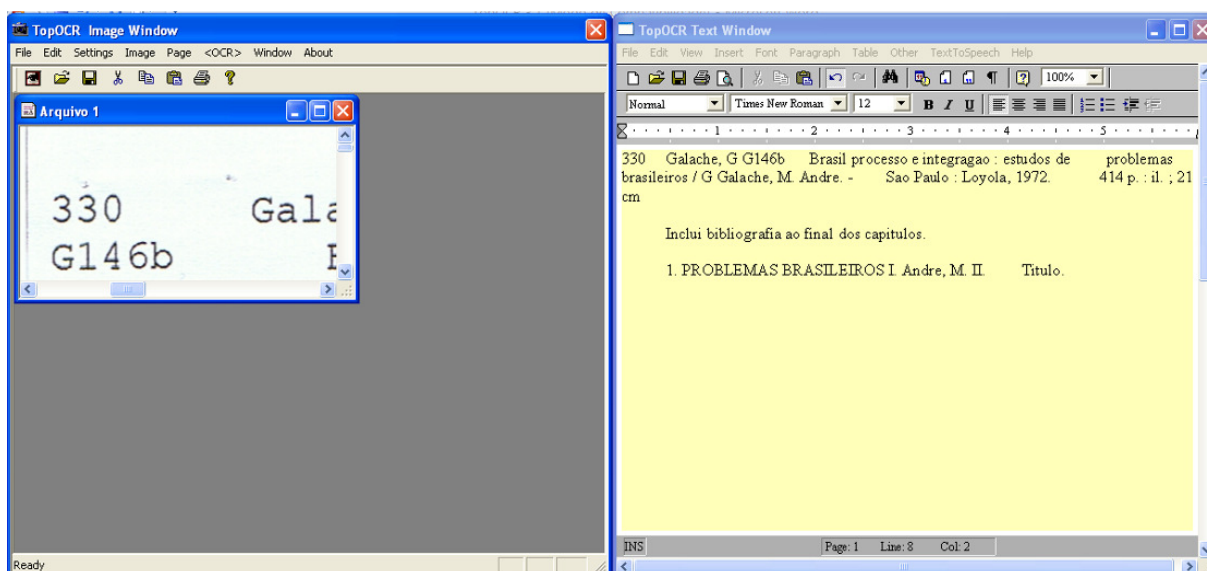
Accept Decide later Keep as image Merge Undo merge

330

G146b Did  
Bib  
apa G  
Bra Pip processo e integrable : escudos de  
pro Tit nas brasileiros / G apache, M. Andre.  
So Paulo : Loyola, 1972.  
414 p. : it. ; 21 cm  
Inclui bibliography  
1. PROBLEMAS  
Título.  
ao final dos capitulos.  
BRASILEIROS I. Andre, M. II.

Fonte: Elaborado pelo próprio autor.

## APÊNDICE K – RESULTADO DO TESTE COM O TopOCR 3.1



Fonte: Elaborado pelo próprio autor.



## APÊNDICE L – RESULTADO DO TESTE COM O WeOCR Server



The screenshot shows a web browser window with the address bar containing the URL `ocr1.sc.isc.tohoku.ac.jp/cgi-bin/weocr/submit_e1.cgi`. The main content area displays the following text:

Recognition result:

---

' 3j0  
G146b

Galache, G  
Brasil processo e integração : estudos de  
problemas brasileiros / G Galache, M. André. --  
São Paulo : Loyola, 1972.  
414 p. : il. ; 21 cm

Inclui bibliografia ao final dos capítulos.

|  
|

1. PROBLEMAS BRASILEIROS I. André, M. II.  
Título.

\_. . . . : \_ . . . . i \_  
\_ \_ . . i . . ' .

---

Download:

- [Raw recognition result \(txt\)](#)

The file will be removed from the server automatically

---

[<- back](#)

Fonte: Elaborado pelo próprio autor.



## APÊNDICE M – PEDIDO DE REGISTRO DE PROGRAMA DE COMPUTADOR

CÓPIA

**INPI****PEDIDO DE REGISTRO DE  
PROGRAMA DE COMPUTADOR**

protocolo

**IDENTIFICAÇÃO DO PEDIDO** (Para uso do INPI)

Número do Pedido

Protocolo, Data e Hora

**DADOS DO AUTOR DO PROGRAMA**Nº de Autores **3** Se mais de um, preencha a "Continuação", com todos os dados solicitados neste Quadro. Date e assine.CPF\* **167.183.508-50**Nome **ZAIRA REGINA ZAFALON**

Nome Abreviado, pseudônimo ou sinal convencional (se houver)

Data de Nascimento **13/08/1971**Nacionalidade **BRASILEIRA**Endereço **RODOVIA WASHINGTON LUÍS, KM 235, MONJOLINHO**Cidade **SÃO CARLOS**UF **SP** País **BRASIL**CEP **13.565-905**

Telefone

FAX

E-mail **zzafalon@gmail.com****DADOS DO TITULAR DOS DIREITOS PATRIMONIAIS**Nº de Titulares **2** Se mais de um, preencha a "Continuação", com todos os dados solicitados neste Quadro. Date e assine.CPF/CNPJ\* **48031918000124**Nome/Razão Social **UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"**Nome abreviado, pseudônimo ou sinal convencional (se houver) **UNESP**

Data de Nascimento

Nacionalidade/Origem

Endereço **RUA QUIRINO DE ANDRADE, 215****CENTRO**Cidade **SÃO PAULO**UF **SP** País **BRASIL**CEP **01.049-010**Telefone **1133937903**

FAX

E-mail **auin@unesp.br** **SIM**, este Titular é Pessoa Jurídica. Caso afirmativo, assinale a melhor classificação:

- Órgão Público     Sociedade com Intuito não Econômico     Microempresa     Software House  
 Instituição Pública de Ensino ou Pesquisa     Instituição Privada de Ensino ou Pesquisa     Outras

**ENDEREÇO PARA CORRESPONDÊNCIA E CONTATO** (Preencha apenas o necessário)

Toda correspondência será enviada para:

 O Procurador ou O Titular acima ou Escaninho nº Representação INPI em: O Endereço abaixo:

Nome

Endereço

Cidade

UF

País

CEP

Telefone

FAX

E-mail

Modelo I (folha 1/2) E

AUIN 010/12

**DADOS DO PROGRAMA**

Título	Scar for MARC				
Data de Criação do Programa	01/02/2012	Regime de Guarda	<input checked="" type="checkbox"/> COM SIGILO	<input type="checkbox"/> SEM SIGILO	
Linguagens	PERL				
Classificação do Campo de Aplicação	IF - 02	IF - 04	IF - 05	IF - 07	-
Classificação do Tipo de Programa	SO - 02	SO - 04	LG - 07	GI - 06	-

**SIM**, este Programa é Modificação Tecnológica ou Derivação. Caso afirmativo, informe Título do Programa Original e (se houver) Número de Registro:

Título do Programa Original \_\_\_\_\_

**SIM**, este Registro é composto por obra(s) de outra(s) natureza(s) de ordem intelectual. Caso afirmativo assinale-a(s) abaixo:

Literária     Musical     Artes Plásticas     Áudio-Visual     Arquitetura     Engenharia

**DOCUMENTOS ANEXADOS** (Informe as quantidades de documentos, não o número de páginas)

Quant	Nome	Quant	Nome
1	Guia de Recolhimento		Contrato de Trabalho/Prestação de Serviço
2	Procuração	1	Involucros/mídia eletrônica Utilizados
1	Termo de Cessão		Contrato/Estatuto Social e Alterações (ou equivalente)
	Termo de Autorização para Modificações Tecnológicas ou Derivações	1	Autorização para Cópia do CD
			Outros(especificar)

**DECLARAÇÕES****DECLARO, PARA TODOS OS FINS DE DIREITO:**

- A) que estou ciente de **TODAS AS RECOMENDAÇÕES** constantes do "Manual do Usuário de Registro de Programas de Computador", **ESPECIALMENTE NO QUE TANGE AO TÍTULO E AOS DOCUMENTOS DO PROGRAMA**, bem como da legislação pertinente ao assunto, constante dos anexos "A"; "B"; "C"; "E" e "F", do referido Manual;
- B) que se deixar de solicitar a prorrogação do sigilo, nos casos necessários, estarei desistindo desse caráter de guarda dos documentos de programa do presente depósito, na forma do art. 3º, § 2º, da Lei 9.609, de 12 de fevereiro de 1998;
- C) que, se devido à qualidade do papel ou à qualidade gráfica dos documentos sigilosos anexos ao presente, houver deterioração ou perda de seu conteúdo, nenhuma responsabilidade caberá ao INPI, desde que mantida a inviolabilidade dos involucros (ressalvadas as hipóteses de serem abertos por ordem judicial ou motivo de força maior);
- D) que em caso de perda do SIGILO ou dos documentos, por culpa exclusiva do INPI, a indenização por perdas e danos, porventura cabível, estará limitada a 20 (vinte) salários mínimos;
- E) que devo manter guardado, em segurança e inviolado, o COMPARTIMENTO "3" do involucro especial para depósito, que é restituído pelo INPI, para fins de recomposição do arquivo do Instituto, no caso de sua destruição total ou parcial por algum tipo de sinistro;
- F) que deverei manter endereço atualizado junto à Divisão de Registro de Programa de Computador, a fim de garantir o recebimento das comunicações relativas ao andamento do meu pedido/registro, ressalvando o INPI de qualquer responsabilidade decorrente da não observação deste preceito.

**DADOS DO PROCURADOR**

CPF/CNPJ*	29059847830	Código do Procurador (se houver)	
Nome	LEOPOLDO CAMPOS ZUANETI - OAB/SP 235.031		
Endereço	R. DR. BENTO TEOBALDO FERRAZ, 271, BL. 2, BARRA FUNDA		
Cidade	SÃO PAULO	UF	SP País BRASIL
CEP	01.140-070	Telefone	1133937903 FAX
E-mail	zuaneti@reitoria.unesp.br		

**DECLARO, SOB AS PENAS DA LEI, SEREM VERDADEIRAS AS INFORMAÇÕES PRESTADAS**

São Paulo, 1º de março de 2012.

Local/Data

Assinatura/Carimbo

LEOPOLDO C. ZUANETI  
Assessor Jurídico  
Agência Unesp de Inovação

Modelo 1 (folha 2/2) E

**REGISTRO DE PROGRAMA DE COMPUTADOR - CONTINUAÇÃO**

Utilize este ANEXO, em quantas folhas forem necessárias, para complementar as informações dos formulários "Pedido de Registro de Programa de Computador" e "Folha de Petição" (DIRTEC).

**CONTINUAÇÃO DADOS DO TITULAR DOS DIREITOS PATRIMONIAIS.:**

Fundação Universidade Federal de São Carlos - UFSCar  
CNPJ.: 45.358.058/0001-40  
Endereço.: Rodovia Washington Luís, Km 235, CEP 13.565-905, São Carlos – SP  
Tel.: 16 3351-9040  
e-mail.: inovacao@usfcar.br

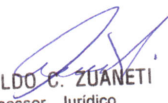
**CONTINUAÇÃO DADOS DO AUTOR.:**

PLÁCIDA LEOPOLDINA VENTURA AMORIM DA COSTA SANTOS, nascida em 24/06/1959, brasileira, casada, professora universitária, inscrita no CPF/MF sob o nº 063.360.218-31, portadora do documento de identidade RG nº 10.464.536-2, SSP/SP, domiciliada em Marília, Estado de São Paulo, na Av. Hygino Muzzi Filho, nº 737, CEP 17.525-900.

JAIRO DA SILVA, nascido em 29/05/1963, brasileiro, solteiro, professor universitário, inscrito no CPF/MF sob o nº 060.263.018-52, portador do documento de identidade RG nº 33.573.505-8, SSP/SP, domiciliado em Marília, Estado de São Paulo, na Av. Hygino Muzzi Filho, nº 737, CEP 17.525-900.

São Paulo, 1º de março de 2012.

\*\*\*fim do documento\*\*\*

  
LEOPOLDO C. ZUÑETTI  
Assessor Jurídico  
Agência Unesp de Inovação

## APÊNDICE N – PEDIDO DE REGISTRO DE MARCA



25/04/2012

850120060201  
15:24

00.000.2.3.12.0209233.2

**Pedido de Registro de Marca de Produto (Mista)****Número do Processo:** 904747972**Dados do Requerente:**

**Nome:** Universidade Estadual Paulista Julio de Mesquita Filho  
**CPF/CNPJ/Número INPI:** 48031918000124  
**Endereço:** Rua Quirino de Andrade, 215 - Centro  
**Cidade:** São Paulo  
**Estado:** SP  
**CEP:** 01049-010  
**País:** Brasil  
**Natureza Jurídica:** Lógico Público  
**e-mail:**

**Dados do Procurador/Escritório****Procurador:**

**Nome:** Leopoldo Campos Zuaneti  
**CPF:** 29059847830  
**e-mail:** zuaneti@gmail.com  
**Nº API:**  
**Nº OAB:** 235031SP  
**UF:** SP

**Dados da Marca**

**Apresentação:** Mista  
**Natureza:** Produto  
**Elemento Nominativo:** Scanfor MARC  
**Marca possui elementos em idioma estrangeiro?** Sim  
**Tradução da Marca:** Varredura para arquivos tipo MARC21



### Imagem Digital da Marca



A eventual deformação desta imagem, com relação à constante do arquivo originalmente anexado, terá sido resultado da necessária adequação aos padrões requisitados para a publicação da marca na RPI. Assim, a imagem ao lado corresponde ao sinal que efetivamente será objeto de exame e publicação, ressalvada a hipótese de substituição da referida imagem decorrente de exigência formal. Portanto, se a mesma não corresponder à imagem desejada para registro nesse órgão, substitua-a, antes de finalizar o Pedido/Petição, observando as especificações constantes do Manual do Usuário.

### Especificação de Produtos ou Serviços, segundo a Classificação de NICE e listas auxiliares

Classe escolhida: NCL(10) 9

#### Descrição da Especificação:

- Computador (Programas de -) [para download]
- Computador (Programas de -), gravados
- Computador (Programas de -), gravados [programas]
- Computador (Programas operacionais para -) [gravados]

### Declaração de Atividade

- Em cumprimento ao disposto no art. 128 da Lei 9279/96, declaro, sob as penas da Lei, que exerço efetiva e lícitamente atividade compatível com os produtos ou serviços reivindicados de modo direto ou através de empresas controladas direta ou indiretamente.

### Classificação dos Elementos Figurativos da Marca - CFE(4), segundo a Classificação de Viena

Categoria	Divisão	Seção	Descrição
27	5	1	Letras apresentando um grafismo especial
27	5	8	Letras ligadas a um elemento figurativo

**Anexos**

Descrição	Nome do Arquivo
Procuração	Procuracao_UNESP.pdf
Procuração	Procuracao_UFSCar.pdf
Termo de cessão de direitos	Cessao_Direitos.pdf
GRU	GRU_Scanfor_MARC.pdf
Esclarecimento cotitularidade	Scan_for_MARC_Esclarecimento_cotitularidade.pdf
Cópia petição depósito de programa de computador	Scan_for_MARC_Software.pdf

Declaro, sob as penas da lei, que todas as informações prestadas neste formulário são verdadeiras.

**Obrigado por acessar o e-Marcas.**

A partir de agora, o número 904747972 identificará o seu pedido junto ao INPI. Contudo, a aceitação do pedido está condicionada à confirmação do pagamento da respectiva GRU (Guia de Recolhimento da União), que deverá ter sido efetuado previamente ao envio deste formulário eletrônico, bem como ao cumprimento satisfatório de eventual exigência formal, (prevista no art. 157 da Lei 9.279/96), em até cinco dias contados do primeiro dia útil após a publicação da referida exigência na RPI (disponível em formato .pdf no portal [www.inpi.gov.br](http://www.inpi.gov.br)), sob pena do presente pedido vir a ser considerado inexistente. Portanto, acompanhe o andamento do seu processo, acessando regularmente a RPI.

**e-MARCAS** Este pedido foi enviado pelo sistema e-Marcas (Versão 2.0) em 25/04/2012 às 15:24