



UNIVERSIDAD
COMPLUTENSE
MADRID

edicic¹⁵

VII Encuentro Ibérico

**DESAFÍOS
Y OPORTUNIDADES
DE LAS CIENCIAS
DE LA INFORMACIÓN
Y LA DOCUMENTACIÓN
EN LA ERA DIGITAL**

**16-17
NOV
2015
MADRID**

Jon Zabala Vázquez
Rodrigo Sánchez Jiménez
María Antonia García Moreno
(coords.)

Desafíos y oportunidades de las Ciencias de la Información y la Documentación en la era digital:

actas del VII Encuentro Ibérico EDICIC 2015
(Madrid, 16 y 17 de noviembre de 2015)



UNIVERSIDAD
COMPLUTENSE

MADRID

2015

© Los respectivos autores

© De la presente edición:
Universidad Complutense de Madrid
Facultad de Ciencias de la Documentación
Departamento de Biblioteconomía y Documentación
C/ Santísima Trinidad, 37
28010 · Madrid. España
Teléfono: +(34) 913 946662
<http://documentacion.ucm.es/>

ISBN: 978-84-608-3330-7

Diseño de portada: Departamento de Estudios e Imagen Corporativa. UCM
Maquetación: Jon Zabala
Fuente tipográfica utilizada: EHU · www.ehu.es/tipografia



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-CompartirIgual (BY-NC-SA) v.3.0 España de Creative Commons. Por tanto, la obra se puede copiar, reproducir, distribuir, remezclar, transformar o comunicar públicamente en cualquier medio o formato, siempre que se cite al autor y a la fuente (UCM. Universidad Complutense de Madrid), y siempre que la obra derivada quede sujeta a la misma licencia y que se haga sin fines comerciales o ánimo de lucro.

La licencia completa se puede consultar en: <http://creativecommons.org/licenses/by-nc-sa/3.0/es/>



Perspectivas de investigación

Scan for MARC: conversor de dados analógicos para o formato MARC21 bibliográfico

Zaira Regina Zafalon

Universidade Federal de São Carlos
Brasil · zaira@ufscar.br

Plácida L. V. A. da Costa Santos

Universidade Estadual Paulista
Brasil · placida@marilia.unesp.br

Jairo da Silva

Universidade Federal de São Carlos
Brasil · jairo.silva@gmail.com

Resumo: Estuda-se a conversão de registros bibliográficos e delimita-se o objeto de estudo na conversão de dados 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. Os princípios sintáticos e semânticos de registros bibliográficos, definidos pelos esquemas de descrição e de visualização na catalogação, estão presentes nos padrões de estrutura de metadados descritivos e nos padrões de conteúdo e determinam o processo de conversão de registros bibliográficos para o Formato MARC21 Bibliográfico. Propõe-se desenvolver um estudo teórico que subsidie a conversão de registros bibliográficos analógicos para o Formato MARC21 Bibliográfico em um interpretador computacional. Definem-se como objetivos específicos: examinar métodos e processos de conversão de registros bibliográficos; apresentar princípios característicos da sintaxe e da semântica em registros bibliográficos; apresentar o interpretador computacional de conversão de dados analógicos para o Formato MARC21 Bibliográfico. O percurso metodológico baseia-se na abordagem qualitativa, com objetivos de caráter exploratório e descritivo. Adota-se a pesquisa bibliográfica para a compreensão e elaboração do arcabouço teórico. O caráter experimental está presente por referir-se a um fenômeno reproduzido de forma controlada, com experimentação, verificação e estudo das evidências e relações entre fatos e as teorias. Como resultado apresenta-se o Scan for MARC, um interpretador computacional de acesso web para conversão de dados bibliográficos analógicos para o Formato MARC21 Bibliográfico. Entende-se que, assim, seja possível saldar demandas de 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; Sintaxe e semântica de registros bibliográficos; ISBD; AACR2r; Metadados.

Abstract: We study the conversion of bibliographic records and limits to the object of study in the conversion of analog data for the MARC 21 Bibliographic format from the parsing and semantics described according to records patterns of descriptive metadata structure and content standards. The syntactic and semantic principles of bibliographic records as defined by description schemes and visualization cataloging, are present in patterns of descriptive metadata structure and content standards and determine the process of converting bibliographic records for the MARC21 Format Library. It is proposed to develop a theoretical study that subsidizes the conversion of analog bibliographic records for the MARC21 Format Library in a computer interpreter. Are defined as specific objectives: to present characteristic principles of syntax and semantics in bibliographic records; examine conversion methods and

processes of bibliographic records; set image processing method bibliographic records in analog optical character recognition software (OCR); present the computational interpreter conversion of analog data to the MARC21 Format Library. The methodological approach is based on the qualitative approach with exploratory and descriptive purposes. Adopts to literature for understanding and developing the theoretical framework. The experimental character is present by referring to a phenomenon reproduced in a controlled manner, with experimentation, verification and study the evidence and relationships between facts and theories. As a result presents the Scan for MARC, a computational interpreter of web access for converting analog bibliographic data for the MARC21 Format Library. It is understood that, so you can pay off demands from institutions who want to use up conversion of bibliographic records database procedure to the MARC 21 Bibliographic format from the description schemes and bibliographic records display (AACR2R and ISBD), character that is considered innovative research.

Keywords: Conversion of bibliographic records; MARC 21 Bibliographic format; Syntax and semantics of bibliographic records; ISBD; AACR2R; Metadata.

Introdução

A pesquisa apresenta como tema nuclear a conversão de registros bibliográficos, e refere-se, principalmente, à conversão retrospectiva. O processo de conversão proposto vincula-se à análise sintática e semântica estabelecida pelos metadados e seus esquemas e diz respeito, portanto, ao estudo da representação de recursos informacionais, do compartilhamento e do reuso de registros descritivos em ambiente digital. Inclui, sobremaneira, o desenvolvimento tecnológico e os aspectos teóricos e metodológicos que envolvem tais processos, utilizando-se de ferramentas e métodos inerentes às tecnologias da informação e comunicação.

Os recursos tecnológicos e midiáticos, por meio de estruturas computacionais e de métodos de definição de metadados, permeiam a produção, a organização, a distribuição, o acesso, o armazenamento, a preservação, o uso e o reuso dos recursos informacionais. Recorrer a mecanismos computacionais nas atividades do dia-a-dia das comunidades de patrimônio cultural (arquivos, bibliotecas e museus), como substitutos das atividades manuais e mecânicas, já foram indicados por Ranganathan e Gopinath (1967), para quem esse processo exige viabilidade econômica e, em última instância, depende do estágio de desenvolvimento de um país. No âmbito técnico-funcional-administrativo destas instituições o emprego de tecnologias de informação e comunicação tornou-se constante, o que favorece tanto o surgimento quanto a inovação de várias atividades cooperativas.

Com as novas tecnologias de informação e comunicação surgem, também dentre as atividades cooperativas, as facilidades de compartilhamento de registros descritivos entre as mais variadas unidades de informação. Nota-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.

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 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.

Na literatura brasileira, 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 no 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. Côrte et al. (1999) destacam como opções para as bibliotecas que desejam incluir seus registros em fichas em um catálogo automatizado que procedam a conversão retrospectiva, para o que recorrem à digitação de todos os registros no novo sistema, à digitalização destas fichas, ou à busca do registro em outras bases. Para registros eletrônicos recorre-se à conversão de dados. Em face de uma solução mais eficiente para a conversão retrospectiva de dados bibliográficos propõe-se, a partir de análises sintáticas e semânticas dos dados bibliográficos, o aproveitamento integral do registro da própria unidade de informação, uma vez que nos métodos indicados é requerido, de algum modo, tratamento técnico dos dados.

Diante do exposto propõe-se apresentar estudo teórico-conceitual baseado em metadados e estruturas descritivas que subsidie a conversão de registros bibliográficos analógicos para o Formato MARC21 Bibliográfico em um interpretador computacional¹ identificado como Scan for MARC. Para o alcance do objetivo proposto definem-se como objetivos específicos: examinar métodos e processos de conversão de registros bibliográficos; apresentar princípios característicos da sintaxe e da semântica em registros bibliográficos; apresentar o interpretador computacional de conversão de dados analógicos para o Formato MARC21 Bibliográfico. Compreende-se, portanto, que princípios sintáticos e semânticos de registros bibliográficos, definidos pelos esquemas de descrição e de visualização na catalogação, estão presentes nos padrões de estrutura de metadados descritivos e nos padrões de conteúdo e determinam o processo de conversão de registros bibliográficos para o Formato MARC21 Bibliográfico.

Consideram-se, como ideias iniciais, para o desenvolvimento da pesquisa, tanto nas questões teóricas quanto naquelas inerentes ao Scan for MARC, o fato de que, no domínio bibliográfico, os registros descritivos são elaborados a partir do estabelecimento de metadados e de estruturas descritivas; de que há esquemas tradicionais e internacionais de descrição e de visualização de registros descritivos, identificados nos padrões de estrutura de metadados descritivos e nos padrões de conteúdo; de que os metadados são estabelecidos com o propósito de promover a identificação e a localização de um documento e estabelecem-se como instrumentos de recuperação da informação; e de que os metadados promovem a comunicação e a encontrabilidade entre comunidades e objetos documentais.

A definição dos metadados e a adoção de estruturas descritivas tornam-se essenciais em processos de recuperação da informação, tanto por humanos quanto por máquinas, por veicularem mensagens codificadas sobre os recursos informacionais de modo a atender e a se nutrir das informações demandadas pelas comunidades. Dada a relação entre itens documentais e demandas informacionais, o esquema de comunicação proposto por Shannon (1948; 1969) pode ser (re)configurado, conforme Figura 1.

FIG. 1. Esquema do processo comunicativo entre recursos informacionais e comunidades



¹ 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. Não será adotado o termo *parsing* nesta pesquisa tendo como base a afirmação de Othero e Menezzi (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."

Tendo em vista os objetivos definidos e as alegações indicadas, o percurso metodológico nesta pesquisa baseia-se na abordagem qualitativa, com objetivos de caráter exploratório e descritivo. Adota-se a pesquisa bibliográfica para a compreensão e elaboração do arcabouço teórico. O caráter experimental está presente por referir-se à validação da proposta teórica consubstanciada no Scan for MARC

Tal estudo, a nosso ver, deflagra contribuições tanto no plano 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 Linguística, de modo a reafirmar a proposta de Borko (1968) e de Saracevic (1996); quanto no prático, por abranger o desenvolvimento de interpretador computacional que possa ser adotado por quaisquer instituições que queiram se utilizar de procedimento de conversão retrospectiva 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.

Conversão retrospectiva como requisito para a cooperação e o 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, contempla não só a entrada de dados, mas a catalogação cooperativa e o recebimento de dados do sistema legado (Marcondes, 1997). A catalogação cooperativa pode ser 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.

No Brasil, 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)², 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.

A partir do caráter de cooperação interbibliotecária, a eficiência e a conectividade entre unidades bibliotecárias estão garantidas por padrões, dado que evitam o retrabalho, facilitam ações futuras de migração de *softwares* de gerenciamento integrado de unidades de informação e promovem o reaproveitamento dos registros bibliográficos. A catalogação cooperativa permite que não seja catalogado o que já está registrado em algum ponto da rede e serve de canal de transmissão de documentos entre certo número de bibliotecas. A catalogação cooperativa, entretanto, está resguardada às unidades de informação que contam com o gerenciamento automatizado, uma vez que 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. 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. Dziekaniak et al. (2008), Paranhos (2004), Café, Santos e Macedo (2001), Côrte et al. (1999) e Marcondes (1997), indicam que, para fins de intercâmbio de registros bibliográficos, os padrões a serem adotados, indicados por privilegiarem o compartilhamento de dados, envolvem a adoção de um padrão de representação do

² Atualmente denominado Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

conteúdo do documento, pelas AACR2r, por exemplo; 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.

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 é economicamente inaceitável criar novos registros que já estão codificados em outro ponto da rede, torna-se fundamental garantir que os dados do acervo sejam (re)utilizados. Desse modo, está posta a necessidade de conversão retrospectiva do acervo. Oliveira et al. (1998) indicam que, no âmbito do Brasil, 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.” (p. 41).

Oliveira et al. (1998) e Côte et al. (2002) 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.

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 indicada por Oliveira et al. (1998), que ressaltam que o método de conversão retrospectiva a ser adotado deve adequar-se às especificidades de prazo, custos e qualidades de registros de cada biblioteca.

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.

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.

Em face de solução para o que se mostra como problema, sob o ponto de vista desta pesquisa – a necessidade de adequações do registro bibliográfico advindos do sistema-base da catalogação cooperativa, de modo que o novo registro deva refletir informações locais –, apresenta-se a proposta de análises sintáticas e semânticas do registro bibliográfico como base para a conversão retrospectiva.

Aproximações teóricas da sintaxe e da semântica na catalogação

Instituições de patrimônio cultural, com destaque para as bibliotecas, representam o papel de agentes sociais de comunicação e, por conseguinte, intermediam recursos informacionais para uso pela humanidade. Desse modo, 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, devem pautar suas atividades na conveniência do usuário. Para tanto, a representação documental 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. O constructo da representação dos recursos informacionais em todos os seus aspectos, tanto descritivos quanto de conteúdo, em outras palavras, a representação documental, visa, portanto, 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. 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. Tais conjuntos são responsáveis, respectivamente, pela individualização de diferentes manifestações de uma obra, pela reunião de suas semelhanças, e pela obtenção do documento.

Por considerar-se que os aspectos sintáticos e semânticos dos registros bibliográficos concentram-se no processo de catalogação e na extração de elementos representativos dos suportes documentais, apresentam-se, como princípios característicos de tal abordagem, a semântica e a sintaxe de registros bibliográficos. Zafalon (2014) explicita que, no estudo estruturalista dos registros bibliográficos, percebe-se a intenção de se identificar a estrutura do registro, sua relação com outros registros e a relação com o documento e indica que, desse modo, não é a sintaxe do registro bibliográfico que dá conta do catálogo, mas a semântica, dado o fato de que indica o contexto e as sinapses entre os vários registros bibliográficos. 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. Semântica, por sua acepção etimológica, apresenta aspectos da significação. É a semântica que dá conta dos processos mentais segundo os quais se produz, constitui, compreende e descreve a representação de um recurso informacional.

O papel semântico do registro bibliográfico está posto nas relações estabelecidas entre os registros bibliográficos, entre o registro bibliográfico e o recurso informacional, e entre os elementos do próprio registro bibliográfico. Zafalon (2014) considera, assim, o estudo da 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 vários registros bibliográficos, a partir de suas semelhanças e diferenças), passando pelo nível semântico (objeto que está sendo descrito e a descrição em si).

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 bibliográfico³ que, em um catálogo, faz com que o disperso e o aparente sejam marcados por identidades que lhe são próprias. 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. 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 por uma metasemântica, no sentido das relações de metalinguagem que podem ser estabelecidas entre diferentes registros. 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. Compreende-se, portanto, que 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.

É assim que se propõe a conversão retrospectiva de registros bibliográficos a ser consubstanciada no Scan for MARC: considerar os aspectos sintáticos em um registro bibliográfico analógico para, a partir do tratamento computacional, demarcar a presença de aspectos semânticos. Compreende-se, portanto, que 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*.

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. Configuram-se como padrões de estrutura de metadados descritivos o MARC Bibliográfico, o MARCXML, o MODS e o Dublin Core. 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.

³ 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.

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.

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. 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*. 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 pontuação⁴ 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 (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-se pelo estudo do Formato MARC21 Bibliográfico. O Formato MARC21 Bibliográfico abrange aspectos de leitura e

⁴ 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."

de interpretação de dados disponíveis em registros bibliográficos, por meio computacional, cujo processo de leitura e interpretação de um registro bibliográfico é facilitado pelas marcações inerentes a ele. 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.

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, promovem, cada uma a seu modo, o desenvolvimento de catálogos, o atendimento aos objetivos bibliográficos e tornam o registro, o gerenciamento e a recuperação dos dados bibliográficos mais eficientes.

A questão que se apresenta é: como se pode entender a conversão retrospectiva de registros bibliográficos 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, pela inferência semântica, prevista em uma aplicação computacional.⁵ 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 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 parte-se para a apresentação dos procedimentos metodológicos adotados e para a apresentação do Scan for MARC, o interpretador de registros bibliográficos analógicos para o Formato MARC21 Bibliográfico.

Scan for MARC: dados analógicos para o formato MARC21 bibliográfico

A partir do pressuposto que considera os esquemas sintáticos e semânticos dos registros descritivos, discute-se o interpretador computacional sintático e semântico de registros descritivos, identificado como Scan for MARC, por aludir ao método de escaneamento dos registros descritivos e sua conseqüente conversão ao Formato MARC21 Bibliográfico.

O desenvolvimento do interpretador, em sua versão beta, envolveu, sumariamente, ações de captura de imagens de registros descritivos analógicos, apresentados em fichas catalográficas em meios eletrônicos; análise de resultados de processamentos de imagem em software de reconhecimento de caracteres, o que permite a conversão de imagem para texto editável; processamento das imagens de registros descritivos selecionados para teste (processo inicial da construção da representação); tratamento sintático e semântico dos registros descritivos digitais, checagem dos resultados e avaliação de ajustes.

A fase de testes para a captura de imagens dos registros descritivos dividiu-se em três etapas: em scanner de impressora funcional, em câmera de dispositivo móvel e em câmera fotográfica digital. Finda esta fase procederam-se testes para o processamento de imagens com reconhecimento de caracteres de registros descritivos em formato analógico, para o qual se adotou um software de reconhecimento de caracteres (OCR). Nesta fase de testes, dividida em dois momentos (o de análise de softwares freeware ou free software, e o de softwares proprietários, com análise em versões trial), analisou-se os seguintes softwares: ABBY FineReader 11, Cognitive Open OCR (Cuneiform) 0.1, FreeOCR, FreeOCR 3.1, Leadtools,

⁵ 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).

OnlineOCR.net, ScreenOCR 9.1, Sci2ools (i2OCR), SimpleOCR 3.5, TopOCR 3.1, WeOCR Server.⁶

Os testes foram desenvolvidos a partir de cotejo e análise quanto à pontuação, aos sinais diacríticos, à troca de letras, ao espaçamento entre informações e margens. Dentre os aplicativos identificados, o software OnlineOCR.net ofereceu melhores resultados. Seguiu-se nova fase de testes na qual se buscava definir o método de processamento de imagens que oferecesse melhor resultado (diante das diversas formas de captura de imagens).

Como resultado geral compreende-se que questões referentes à iluminação na captura das imagens são extremamente relevantes para o processamento em software de OCR dada a influência na qualidade da imagem; os melhores índices de sucesso no processamento de imagens de registros descritivos foram capturados diante das imagens capturadas em scanner de impressora multifuncional; problemas com a troca de letras no processo de reconhecimento de caracteres em imagens foram identificados nos testes dos três diferentes recursos; aspectos quanto à fixação da câmera, quer seja de dispositivo móvel quer seja de fotografia digital, mostraram-se mais relevantes do que a distância a ser considerada na captura da imagem; a qualidade do OCR está intrinsecamente vinculada à qualidade da imagem e não ao método pelo qual a imagem é capturada. Na versão beta, disponível para testes locais, optou-se por trabalhar com as imagens capturadas em scanner de impressora multifuncional e com o OnlineOCR.net. Efetuados os testes de processamento de imagens dos registros descritivos analógicos selecionados, apresentam-se os procedimentos adotados para o tratamento sintático e semântico dos dados bibliográficos do arquivo, à luz da proposta teórica apresentada nesta pesquisa.

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). Foram estudados os sinais de pontuação, presentes no *padrão de estrutura de metadados descritivos*, que, de acordo com a proposta, definem a inferência de conteúdo semântico, presente nos *padrões de conteúdo*. Ressalte-se que, neste 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. 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. Para a efetiva adoção destas marcações no *script* de conversão de registros descritivos ao MARC21 Bibliográfico foi necessário, entretanto, cuidado quanto à pontuação que faz parte do conteúdo e não da estrutura de metadados descritivos. Outra análise, em busca da definição de padrões, foi requerida quanto às informações dos pontos de acesso para o que se determinou que os pontos de acesso definidos por cabeçalhos de assunto iniciam-se por numerais arábicos seguidos de ponto, e de que os outros pontos de acesso que se fizerem necessários iniciam-se por numerais romanos.

Sem dúvida alguma, um dos primeiros problemas identificados para o processamento das informações 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. A fase de tratamento sintático e semântico dos registros descritivos digitais foi feita a partir de testes de processamento de roteiros de registros descritivos. Foram estabelecidos quatro roteiros, com diferentes graus de complexidade, que subsidiassem as versões de

⁶ Endereço para acesso: ABBYY FineReader 11 (<http://www.abbyy.com.br/finereader/>); Cognitive Open OCR (Cuneiform) 0.1 (<http://cognitive-openocr-cuneiform.en.softonic.com/download>); FreeOCR (<http://www.free-ocr.com/>); FreeOCR 3.1 (<http://www.paperfile.net/freeocr.exe>); Leadtools (<http://www.leadtools.com/sdk/ocr/default.htm>); OnlineOCR.net (<http://www.onlineocr.net/default.aspx>); ScreenOCR 9.1 (<http://www.screenocr.com>); Sci2ools (i2OCR) (<http://www.sciweavers.org/free-online-ocr>); SimpleOCR 3.5 (<http://www.characterll.com/scanstore/>); TopOCR 3.1 (<http://www.brothersoft.com/topocr-download-47055-s1.html>); WeOCR Server (<http://ocr1.sc.isc.tohoku.ac.jp/e1/>).

scripts (desenvolveram-se quatro versões, cada qual com pequenos ajustes necessários após a checagem dos resultados). O quadro 1 concentra os roteiros 1 a 4.

QUADRO 1 – Roteiros utilizados na configuração do interpretador computacional

<i>Primeiro roteiro para o script de conversão de registros bibliográficos</i>
<p>123 Sobrenome, Nome S789a Título : subtítulo / Nome Sobrenome, Outro Nome Sobrenome. -- Local de Publicação : Editora, Ano. 929 p. : il. ; 23 cm</p> <p>Inclui bibliografia.</p> <p>1. ASSUNTO I. Sobrenome, Outro Nome. II. Título.</p>
<i>Segundo roteiro para o script de conversão de registros bibliográficos</i>
<p>123.456 Sobrenome do Autor, Primeiro Nome S789a Aqui vai o título do livro : aqui o subtítulo 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 Primeiro 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)</p> <p>Bibliografia. Índice. ISBN: 12-345-67890-123.</p> <p>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.</p>
<i>Terceiro roteiro para o script de conversão de registros bibliográficos</i>
<p>123.456 Aqui vai o título do livro : aqui o subtítulo 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)</p> <p>Bibliografia. Índice. ISBN: 12-345-67890-123.</p> <p>1. PRIMEIRO ASSUNTO 2. ASSUNTO - SEGUNDO 3. ASSUNTO COMPOSTO - PRIMEIRO COMPOSTO 4. ASSUNTO COMPOSTO - SEGUNDO COMPOSTO I. Organizador, 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.</p>
<i>Quarto roteiro para o script de conversão de registros bibliográficos</i>
<p>123.456 Aqui vai o título do livro : aqui o subtítulo que, neste caso, é muito importante também / organizado por Primeiro Nome Sobrenome do Organizador, Segundo Nome Sobrenome do Organizador e Terceiro Nome</p>

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.

Os resultados da conversão dos roteiros indicados podem ser observados no quadro 2, no qual estão destacadas as inconsistências observadas.

QUADRO 2 – Resultados de conversão de registros bibliográficos a partir dos modelos definidos

Resultado de conversão de registro bibliográfico conforme roteiro 1	
	<pre> ===== 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 Includ bibliografia 65014\$aASSUNTO 7001 \$aSobrenome, Outro Nome ===== </pre>
Resultado de conversão de registro bibliográfico conforme roteiro 2	
	<pre> ===== MARC ===== 020 \$a1234567890123 090 \$a123.456 \$bs789a 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 Setrad. </pre>
Resultado de conversão de registro bibliográfico conforme roteiro 3	
	<pre> ===== 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. </pre>
Resultado de conversão de registro bibliográfico conforme roteiro 4	

```

===== MARC =====
020 $a1234567890124 (hardback).
090 $a123-456 $bA789
24500$aAqui vai o título do livro $baqui o subtítulo que, neste caso, é muito
importante $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. $b1l. $c23 cm
4901 $aAqui entra o nome da série com texto extenso para complicar o comportamento
do script $v23
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.
=====

```

Observou-se a necessidade de ajustes no quarto de modo a contemplar os problemas identificados. Desse modo, adequou-se o *script* 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. O resultado dos ajustes no quarto roteiro podem ser visualizados na figura 2.

FIG 2 - Resultados de conversão de registros bibliográficos a partir dos ajustes necessários no quarto roteiro



```

===== MARC =====
020 $a1234567890123 (paperback)
020 $a1234567890124 (hardback)
090 $a123-456 $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. $b1l. $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 $eorg.
7001 $aOrganizador, Segundo Nome Sobrenome do $eorg.
7001 $aOrganizador, Terceiro Nome Sobrenome do $eorg.
7400 $aAqui o subtítulo que, neste caso, é muito importante também
7400 $aThis is the English title
=====

```

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

O Scan for MARC, validado em ambiente de acesso local, com as ações em linha de comando e com processamento com interação humana na captura das imagens, serviu como base para que o desenvolvimento do interpretador em ambiente de acesso web ocorresse. Tal interface está disponível no para acesso no endereço: www.scanformarc.ufscar.br. Neste endereço é possível conferir duas versões do interpretador: uma delas para a digitação ou colagem do texto, utilizada para verificação da adequabilidade de script; outra, com a integração do script com o OCR, para o qual adotou-se o Tesseract. Nesta versão é possível tratar a variação sintática do texto nos idiomas português, inglês, espanhol, francês e alemão; fazer o upload de imagem a ser processada, com opção de uso ou não de filtro para o tratamento da imagem por OCR; e opção para download do registro.

Considerações finais

Diante da temática definida para esta pesquisa, a conversão retrospectiva 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.

Apresentou-se 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 preservação das estruturas nas propostas de construção, de compartilhamento, de conversão e de reuso de registros descritivos por meio da adoção de normas, regras, 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.

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: [1] compreensão da obra, enquanto significado, e da manifestação, enquanto significante; [2] entendimento da representação bibliográfica como resultado da relação entre significante e significado e entre obra e manifestação e como definitiva para a semântica; [3] percepção da sintaxe para a definição da subsemântica, além de necessária à representação do recurso informacional; [4] apreensão do conceito de supersemântica, a partir de sua co-dependência com a subsemântica e a semântica, na relação, identificável nos catálogos, entre registros descritivos 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 por requererem, ainda, ajustes e aprimoramentos, realizados em prol da definição do comportamento do Scan for MARC, entende-se que a contribuição quanto aos aspectos sociais ensejados apresentam 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: [1] aprimoramento da interface gráfica; [2] proposta para

processamento de registros em lote, bem como o download; [3] previsão de padrões para a interpretação de dados de classificação de assunto (CDU); [4] tratamento do conteúdo de termos qualificadores (subdivisão de forma, subdivisão cronológica, subdivisão geográfica e subdivisão geral) nos pontos de acesso de assunto tópico; [5] previsão de padrões para tratamento de pontos de acessos secundários de assuntos para nome pessoal, institucional, evento e título uniforme; [6] previsão de padrões para tratamento de pontos de acessos principais para instituições, eventos e título uniforme; [7] previsão para criação de dicionários semânticos para pontos de acesso e seus termos qualificadores; [8] implementação de processo de teste de consistência do arquivo convertido, uma vez que isso afere caráter qualitativo ao registro bibliográfico gerado.

Enseja-se, ainda, dar continuidade aos estudos dos métodos sintáticos e semânticos de registros descritivos 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.

Referências

Borko, Harold. (1968). Information science: what is it? *American Documentation*, 19 (1): 3-5.

Café, L.; Santos, C.; Macedo, F. (2001). 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. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/198/175>. Acesso em: 10 jan. 2012.

Chong-De la Cruz, I. *Directrices para la descripción y catalogación del libro antiguo*. México, DF: UNAM, 2014.

Código de catalogação anglo americano. (2004). 2. ed. rev. ampl. São Paulo: FEBAB/IOESP.

Côrte, A. R. et al. (1999). 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. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/284>. Acesso em: 10 jan. 2012.

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

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

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

Dziekaniak, G. V. et al. (2008). 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. 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. (2005). *Inteligência artificial: noções gerais*. Florianópolis: VisualBooks.

Ferreira, J. R. et al. (1979). 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.

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

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

Hubner, E. (2002). Conversão retrospectiva de registros bibliográficos. In: *Seminário Nacional de Bibliotecas Universitárias*, 13., Recife, 2002. Anais... Recife: UFPE. 1 CD-ROM.

Langker, Robert. (1974). ISBD: another step in the right direction. *The Australian Library Journal*, 23 (3): 99-103.

Marcondes, C. H. (1997). Automação de bibliotecas. In: CURSO Aplicação de tecnologia ao desenvolvimento de bibliotecas. Rio de Janeiro: FGV.

Oliveira, N. M. et al. (1998). 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. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/608/381>. Acesso em: 10 jan. 2012.

Paranhos, W. M. M. R. (2004). 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. 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. (2002). 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. Disponível em: <<http://www.sibi.ufrj.br/snbu/snbu2002/oralpdf/37.a.pdf>>. Acesso em: 24 jan. 2009.

Ranganathan, Shiyali Ramamrita, Maluer Aji Gopinath. (1967). *Prolegomena to library classification*. 3rd ed. New York: Asia Publishing.

Saracevic, Tefko. (1996). Ciência da informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, Belo Horizonte, 1 (1): 41-62.

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

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

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

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

Swanson, Gerald. (1973). ISBD: standard or secret? *Library Journal*, 15: 124-130.

Zafalon, Z. R. (2014). *Scan for MARC: conversão de registros bibliográficos analógicos para o Formato MARC21 Bibliográfico*. São Paulo: Ed. UNESP, 2014.