

---

# HEREDOGRAMA INFORMACIONAL: um caminho para combate à desinformação

*INFORMATIONAL HEREDOGRAM: A way to combat disinformation*

---

**Oswaldo de Souza (1)**

(1) Universidade Federal do Ceará, Brasil, osvsouza@gmail.com



## Resumo

Neste texto discutimos questões relacionadas à organização e representação do conhecimento no sentido de propor uma contribuição teórico-metodológica com possíveis impactos positivos na qualidade da informação. Caracterizando-se como uma pesquisa de natureza qualitativa do tipo exploratória com levantamento bibliográfico, na qual foram explorados padrões de codificação, modelos conceituais e de estruturação de dados bibliográficos, destacando-se o BIBFRAME, RDA, RDF e o OWL, os quais foram combinados para definir uma nova abordagem teórico-metodológicas denominada de Heredograma Informacional, com aplicação em documentos e textos provenientes de fontes de informações da WEB e também provenientes de conhecimento registrado, típico de unidades de informação. Foi obtido como resultado uma proposta que possui propriedades de uma estrutura taxonômica que também produz informação verificável a respeito dos documentos e seus respectivos conteúdos. Nela considera-se a atribuição de um Identificador Semântico de Objeto Digital (ISOD) a ser utilizado em associação ao DOI ou em substituição a este. O Heredograma Informacional prevê a incorporação de elementos funcionais ativos para a produção automática, com ou sem supervisão humana, de elementos semânticos extraídos dos conteúdos dos documentos. A proposta produzida como resultado dos trabalhos introduz elementos para a mitigação dos problemas relacionados à desinformação e a *fake news* e adiciona elementos provenientes do processamento de linguagem natural (PLN), inteligência computacional (IC) e inteligência artificial (IA).

**Palavras-chave:** heredograma informacional. desinformação. inteligência computacional. inteligência artificial. recuperação da informação. ecossistema informacional.

## Abstract

In this text we discuss issues related to the organization and representation of knowledge to propose a theoretical-methodological contribution to positively impact the quality of information. Characterized as qualitative research of exploratory type with a bibliographic survey about coding, conceptual models and structuring patterns of bibliographic data, highlighting BIBFRAME, RDA, RDF and OWL, were explored, which were combined to define a new theoretical-methodological approach called Informational Heredogram with application in documents and texts coming from WEB information sources and recorded knowledge typical of library documents. Results obtained has properties of a taxonomic structure which simultaneously produces trustable information about documents and their respective contents. It considers the attribution of a Semantic Digital Object Identifier (SDOI) to be used in association with the Digital Object Identifier (DOI) or replacing it. The Informational Heredogram foresees the incorporation of active functional elements for the automatic production, with or without human supervision, of semantic elements regarding the contents of the texts. The proposal obtained in the works introduces elements to mitigate problems related to disinformation and fake news and adds elements from natural language processing (NLP), computational intelligence (CI) and artificial intelligence (AI).

**Keywords:** informational heredogram. disinformation. computational intelligence. artificial intelligence. information retrieval. informational ecosystem.

## 1 Introdução

---

O aumento da disponibilidade de informações na internet, compreendido como um fenômeno informacional, tem sido amplamente estudado e discutido em literatura científica e acadêmica, nos quais são explorados aspectos relacionados, abordando desde considerações sobre a quantidade de dados disponíveis até os efeitos sociais e cognitivos decorrentes dele.

Um exemplo desses estudos pode ser visto em (Boyd; Crawford, 2012, p. 662-679) no qual podemos observar uma preocupação quanto às questões relacionadas a ética, a privacidade e ao poder, quando afetados ou quando afetam esse fenômeno.

A percepção do aumento da disponibilidade de informações na internet não é um acontecimento novo, de fato, já se observam trabalhos discutindo a questão desde que houve a popularização da *Word Wide Web* (WEB) e com maior intensidade a partir do fato do surgimento dos dispositivos móveis denominados de *smartphones*, o que revolucionou a forma como passamos a acessar e interagir com a internet (Deursen; Helsper, 2015, p.29-52). Tais dispositivos móveis, associados a um cenário de conectividade ubíqua à internet, permitem o acesso constante, independentemente de nossa localização geográfica, a uma vasta quantidade de informações que estão disponíveis online. Esses dispositivos alteraram a forma das pessoas obterem acesso à

informação, bem como afetaram a forma como os autores passaram a produzi-la. Sobre isso nos falam Vidotti et al. (2019, p. 12):

[...] verifica-se que o ciberespaço se torna um elemento que perpassa as diversas atividades das organizações e dos indivíduos, tendo uma relação direta com a virtualidade, o que aprofunda e torna complexo os processos inerentes aos ambientes como a Web. Tendo como foco tal complexidade, é necessário compreender o desenvolvimento da Web posteriormente a sua concepção. Neste sentido, nos seus primeiros anos, a Web se tornou um ambiente com um volume de dados muito elevado, sem ter uma descrição ou um controle sobre as informações ali contidas.

Se por um lado temos acesso a um volume crescente de fontes de informações e documentos, também podemos observar o surgimento e aumento de categorias de problemas relacionadas à qualidade e propósito da informação, problemas tais como os vistos durante a pandemia da COVID-19, discutidos em trabalhos como os de Ana Carla e Elisabete Souza (Mazzeto; Souza, 2022, p. 2-23). Dentre os problemas destacamos a desinformação, particularmente a *fake news*, que tem se revelado algo relevante a afligir a sociedade, como podemos ver no trabalho de Anna Cristina e Arthur Bezerra (Brisola; Bezerra, 2018) no qual os autores discutem a informação falsa, posta como um ato intencional a prejudicar a credibilidade do conhecimento científico.

Outro aspecto a ser considerado em relação ao crescimento da disponibilidade de informações na internet é o correspondente aumento no esforço que o sujeito informacional deve empreender para estar em contato com as novas publicações em sua área de interesse. De fato, o volume de informações disponíveis na internet passa a requerer uma elevada disponibilidade de tempo para uma simples ação de pesquisa bibliográfica abrangente e correspondente seleção de material e leituras iniciais dos materiais de interesse. Tal disponibilidade de tempo pode se tornar proibitiva se não houver alguma solução tecnológica que diminua o volume da leitura a ser realizada, como nos falam De Souza e Fernandes:

Há, portanto, um aumento do tempo necessário para que um potencial usuário de uma informação a encontre; obviamente isso é proporcional ao volume de informação disponível. Acreditamos, contudo, que é possível alterar essa relação com o tempo através do uso de tecnologia que minimize o volume de informações a serem apreendidas, quer seja pela diminuição do volume de texto a ser lido, ou pela alteração do tipo de código usado no suporte. Se um determinado documento puder ser transformado em outro tipo de documento, todavia preservando-se o

valor semântico, poderíamos ter, por exemplo, uma condensação dos textos, criando versões de denso valor semântico que requereriam menor tempo para seu consumo. (De Souza; Fernandes, 2020, p. 146).

Podemos perceber problemas relacionados ao volume de informações disponíveis, ~~e também~~ quanto a qualidade dessas informações e até mesmo em alguns casos, problemas quanto a definição da autoria, pois em alguns casos a verdadeira autoria de um texto pode ser omitida intencionalmente ou por acidente. Some-se a isso questões de textos, imagens e vídeos que vêm sendo produzidos por aplicativos de computador construídos com recursos de Inteligência Artificial (AI). Todos esses problemas dificultam a seleção de material e podem levar o sujeito informacional a selecionar um material com informação de baixa qualidade.

Há, portanto, a necessidade de discutir meios para mitigar os impactos da desinformação ou informações imprecisas e do volume de informações disponíveis, procurando caminhos metodológicos relacionados à organização e a representação do conhecimento, objetivando o desenvolvimento de contribuições teórico-metodológicas que impactem positivamente a recuperação do conhecimento. Na busca de soluções devemos avaliar a possibilidade de uso de aportes teóricos provenientes de outras áreas, especialmente aquelas relacionadas ao Processamento de Linguagem Natural (PLN), da IA e da Inteligência Computacional (IC). Tudo no sentido de propor novas abordagens, criando e adicionando metadados a padrões já existentes ou convergindo na criação de novos padrões. Usar metadados produzidos automaticamente por recursos computacionais ou por humanos, de forma que combinados possam diminuir os impactos negativos dos problemas mencionados.

A partir da compreensão dessa necessidade o objetivo deste trabalho é definir elementos informacionais estruturados em um esquema de metadados. Esse esquema deve estabelecer uma estrutura taxonômica, referencial e verificável a respeito dos documentos e seus respectivos conteúdos, origens e autores, constituindo assim um modelo de representação que se possa denominar de heredograma informacional, cujo conceito teórico e explicação do termo são apresentados adiante.

Para o alcance desse objetivo é realizada uma investigação em torno dos padrões de metadados existente e aplicáveis ao registro de dados extraídos de material informacional, sejam

eles bibliográficos, textuais ou multimídia, combinando os padrões investigados em uma proposta de contribuição para a área.

## **2 Métodos e materiais da pesquisa**

---

Este trabalho caracteriza-se como uma pesquisa de natureza qualitativa do tipo exploratória com levantamento bibliográfico, os procedimentos foram iniciados com a leitura dos trabalhos de (Boyd; Crawford, 2012, p. 662-679), (Deursen; Helsper, 2015, p.29-52) e (Mazzeto; Souza, 2022, p. 2-23). A partir dessas leituras iniciais tornou-se necessário aprofundar os estudos em torno dos conceitos relativos aos metadados, sendo eleito o termo “*metadata*” como expressão de busca para este fim.

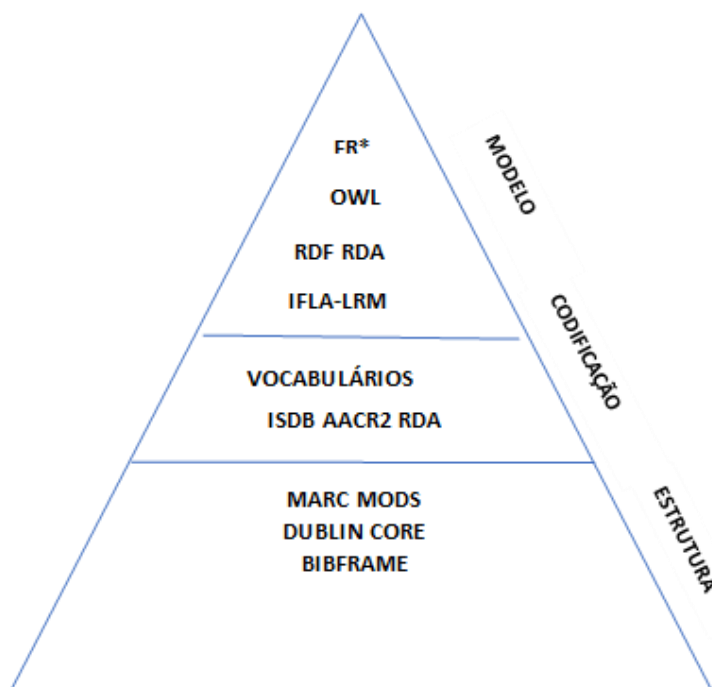
Foi então realizada uma pesquisa no Portal de Periódicos da CAPES (PPC) (CAPES, 2023), os resultados foram classificados em ordem cronológica, do mais antigo para o mais recente visando identificar os documentos originais sobre o tema. Como resultado da pesquisa no PPC foi selecionado o primeiro texto no qual fossem abordados conceitos teóricos primordiais sobre o termo “*metadata*”, o que resultou na coleta do texto de título *Understanding metadata and metadata schemes* (Greenberg, 2005, p. 17-36).

Também foi necessário visitar os conceitos relativos aos esquemas ou padrões de representação e organização de metadados de registros bibliográficos. Portanto, no sentido de conhecer pesquisas recentes sobre tais conceitos foram eleitas as expressões de busca “*bibliographic records representation*” e “*bibliographic data representation*”, na primeira expressão de busca foi utilizado a palavra “*records*” em substituição ao termo “*data*” pois na época das primeiras publicações sobre metadados o termo amplamente aceito era “*records*”. Na pesquisa com a primeira expressão de busca foram obtidos 2269 registros e 3062 com a segunda. Os resultados foram organizados por ordem de relevância e dentre eles foram selecionados os primeiros textos cujos resumos revelassem alguma pesquisa recente sobre padrões para registros bibliográficos, sendo selecionados:

- *Exploration of Subject Representation and Support of Linked Data in Recently Created Library Metadata: Examination of Most Widely Held WorldCat Bibliographic Records* (Zavalin; Zavalina; Miksa, 2021);
- *Implementing Bibliographic Enhancement Data in Academic Library Catalogs: An Empirical Study* (Wu, 2023);
- *Representing and integrating bibliographic information into the Semantic Web: A comparison of four conceptual models* (Zapounidou; Sfakakis; Papatheodouru, 2017).

A partir da leitura e compreensão desses textos foram relacionados os formatos de estruturação, codificação e modelos conceituais para dados bibliográficos. Tais formatos são discutidos nas seções 2.1, 2.2 e 2.3, os quais podem ser vistos na Figura 1. Nas discussões, todavia, não nos preocupamos com a cronologia de criação e eventual sucessão entre eles, nos limitaremos aos aspectos de alcance de representação dos mesmos e a particularidades relativas ao inter-relacionamento entre documentos e outros elementos informacionais.

Figura 1 - Padrões de estruturação, codificação e modelos de dados bibliográficos



Fonte: Dados da pesquisa (2023)

Na Figura 1 incluímos uma menção a vocabulários controlados, ISDB e AACR2, que pertencem a codificação de dados, apenas para haver coesão entre elementos presentes na literatura analisada, todavia, tendo em vista que no escopo do presente trabalho não dedicaremos esforços no sentido de fazer proposições quanto a codificação de dados, não detalharemos esses relevantes elementos da linguagem documentária. Quanto ao RDA ele foi desenvolvido como resposta do *Steering Committee* a um planejamento estratégico para o período de 2005-2009, no qual havia o objetivo de substituir o padrão AACR2 (SC, 2005) cuja primeira publicação data de 1978. O RDA é um padrão que define um formato e conjunto de regras usadas para padronizar a descrição bibliográfica visando o acesso e compartilhamento em todo o mundo, por exemplo, através de serviços na WWW (Wu, 2023).

## 2.1 Estruturação de dados

---

O MARC 21 é atualmente o formato dominante de representação e comunicação de dados bibliográficos, impõe uma estrutura fortemente definida com possibilidades de expansão, sendo destinada ao intercâmbio de dados entre máquinas (LOC, 1996). MODS é um formato especificado em XML e deriva do MARC 21, todavia, é baseado em marcação textual – tags – ao invés da codificação numérica típica do MARC 21. A criação do MODS (LOC, 2022) foi justificada como uma iniciativa para aprimorar o MARC 21 pela adição das facilidades existentes no Dublin Core (DC, 1998) o qual é um padrão de estruturação de dados que foi criado em 1995 e formalizado em 1998 como um documento RFC 5791. É importante evidenciar que o MARC 21, MODS e Dublin Core são fortemente direcionados ao registro unitário de item bibliográfico, não sendo sua utilização idealizada para interrelação entre outros itens do catálogo. O uso de apenas esses três tipos de estruturação de dados bibliográficos implicará na inexistência de recursos para criar uma estrutura taxonômica, bem como na inexistência de metadados que permitam correlacionar diretamente os itens do acervo, tanto internamente quanto externamente. A correlação que eventualmente poderia ser estabelecida seria indireta e dependeria de semelhança de assunto ou palavra-chave. Uma questão sobre o Dublin Core que deve ser considerada a parte, trata-se de que sua especificação é intencionalmente direcionada para a simplicidade e essa simplicidade também é uma fraqueza, pois em uma conversão de Dublin Core para outros formatos de estruturação existirão muitas lacunas, visto que por exemplo, o MARC 21 e o MODS possuem um número

maior de metadados em suas especificações. Embora o Dublin Core tenha evoluído ao longo dos anos desde seu lançamento, quando possuía inicialmente um conjunto de 15 metadados e atualmente possa ser ampliado com o uso de esquemas de codificação, uma decisão pela sua adoção deve considerar que o MODS tem um leque de metadados mais amplo, como os do MARC 21, sendo que a ele foram incorporadas facilidades do Dublin Core.

O BIBFRAME foi criado como um padrão para evoluir a descrição de dados bibliográficos em um modelo de dados conectados, justamente no sentido de tornar os dados aptos a serem úteis dentro ou fora da unidade de informação de sua origem. No entanto, não é possível estabelecer algum tipo de correção referencial ou hierárquica de qualquer tipo entre diferentes dados bibliográficos. Iniciativas na Biblioteca do Congresso dos EUA avaliam a migração de uma estruturação de dados de MARC 21 para BIBFRAME, como se vê em (Zavalin; Zavalina; Miksa, 2021, p. 154) trabalho no qual também utilizam o padrão *Resource Description and Access* (RDA) na codificação dos dados.

Na seção 3 retornaremos a discussão sobre alguns dos padrões de estruturação de dados.

## 2.2 Modelos conceituais

---

Os modelos conceituais estabelecem visões de como os dados bibliográficos podem ser representados de maneira que a representação favoreça recuperações futuras. Nas subseções seguintes passamos a analisar aspectos desses modelos que são necessários para as discussões que ocorrem na seção 3.

### 2.2.1 A família de modelos *Functional Requirements - FR*

A família de modelos FR compreende o *Functional Requirements for Bibliographic Records* (FRBR), o *Functional Requirements for Authority Data* (FRAD) e o *Requirements for Subject Authority Data* (FRSAD) (IFLA, 2009) e foi desenvolvida utilizando o modelo de representação de Entidade-Relacionamento (ER) definidos por Chen (Chen, 1976, p. 9-36). O objetivo principal no desenvolvimento do FRBR foi organizar uma estrutura que reunisse as diversas partes que compõem os dados bibliográficos. No FRBR foram incluídas entidades tratadas



como objetos de potencial interesse dos sujeitos informacionais. Nele os objetos podem ser abstratos ou concretos e ao todo foram incluídas 10 entidades, distribuídas em três grupos:

- Grupo 1 - que traz a identidade do que está sendo descrito, possuindo as seguintes entidades: Obra, Expressão, Manifestação e item;
- Grupo 2 – que traz dados sobre os agentes relacionados as entidades do Grupo 1, responsáveis intelectuais ou artísticos, etc. pela produção do que é descrito no Grupo 1. O Grupo 2 possui as seguintes entidades: Pessoas e Entidades coletivas;
- Grupo 3 - que agrega elementos extras de descrição dos dados do Grupo 1 e inclui as seguintes entidades: Conceito, Objeto, Lugar e Evento.

Cada uma das entidades é expressa utilizando-se o modelo RDA desde 2009 e continua assim até os dias atuais, como se vê em (Wu, 2023). Todavia, em 2017 a IFLA propôs uma nova estruturação, denominada de IFLA-LRM, na qual todos os padrões da família FR foram integrados em uma única especificação. Dentre as justificativas para essa decisão destacou-se o fato de que cada um dos três padrões teve desenvolvimentos que seguiram isolados, o que culminou por acarretar algumas sobreposições e problemas entre os padrões da mesma família.

### 2.2.2 A IFLA-LRM

A necessidade de consolidação das especificações da família FR tornou-se evidente quando a IFLA reconheceu dificuldades no uso das três especificações em conjunto:

Inevitavelmente os três modelos de FR, embora todos criados em uma estrutura de modelagem de relacionamento entre entidades, adotaram pontos de vista diferentes e soluções distintas para 6 problemas comuns. Ainda que todos os três modelos se façam necessários em um sistema bibliográfico completo, a tentativa de adotá-los em um único sistema exigia a solução de problemas complexos de maneira ad hoc, com orientação mínima dos modelos. (IFLA, 2017, p1).

Assim a partir desse entendimento foi criada a especificação IFLA-LRM (IFLA, 2017) com foco direcionado a dados bibliográficos dos acervos de bibliotecas:

O *IFLA Library Reference Model* pretende ser um modelo de referência conceitual de alto nível desenvolvido dentro de uma estrutura aperfeiçoada de modelagem entidade-relacionamento. O modelo abrange dados bibliográficos como entendidos em um sentido amplo e geral. Em termos de abordagem geral e

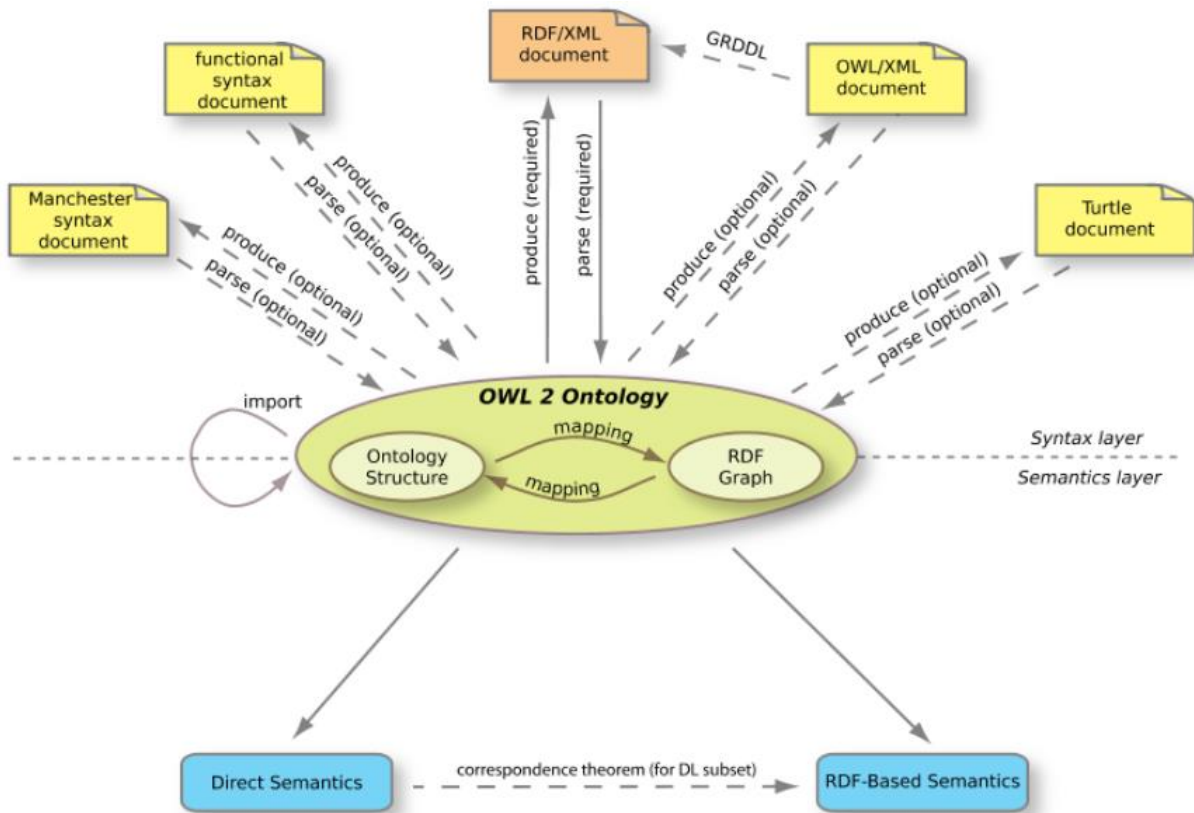
metodologia, o processo de modelagem que resultou no modelo IFLA LRM incorporou a abordagem adotada no estudo original do FRBR. (IFLA, 2017, p9).

Desta forma, quanto ao fundamental de sua concepção, não se vislumbram diferenças irreconciliáveis em comparação ao FRBR, sendo válidas as considerações quanto à possibilidade de expressão dos atributos das entidades em RDA diferindo, no entanto, na divisão e organização de grupos de objetos e suas respectivas entidades.

### 2.2.3 O modelo *Web Ontology Language - OWL*

O OWL (W3C, 2004) idealizado pelo W3C é um modelo de linguagem semântica para WEB, direcionado ao auxílio da representação das complexas relações do conhecimento que permeiam elementos abstratos ou concretos presentes no ecossistema informacional (De Souza; Fernandes, 2020), compreendendo dados sobre elementos individuais, grupos de elementos e relações entre elementos individuais. Nas discussões deste texto consideraremos apenas a versão 2 do OWL (W3C, 2012). A estruturação do modelo OWL direciona para um modelo de representação que pode ser facilmente processado por programas de computador, tendo em vista que sua notação é baseada na construção de uma linguagem lógica.

Figura 2 - Ilustração da OWL 2.0



Fonte: Adaptado de (W3C, 2012)

OWL não é um esforço isolado para o controle de dados semânticos para a WWW, na verdade ele é parte do conjunto de especificações produzidas pela W3C para esse fim, conjunto esse que inclui: o RDF, o RDFS, o SPARQL e o OWL. RDFS e SPARQL estão fora do escopo deste trabalho e não serão abordados. Segundo a W3C, a especificação do OWL (W3C, 2012) é mais abrangente e poderosa do que a especificação do RDF (W3C, 2004).

Na Figura 2 podemos observar o relacionamento entre os elementos da especificação da OWL 2.0, os quais são detalhados e definidos por cinco documentos que descrevem sua estrutura: (a) **uma** estrutura conceitual; (b) **uma** sintaxe primária de troca (RDF/XML); (c) **duas** alternativas de descrições semântica (*Direct e RDF-Based*); (d) **um** conjunto de requisitos de conformidade. Também há três especificações adicionais que descrevem características que podem ser suportadas por algumas implementações: (a) **uma** para os perfiz de linguagem; (b) **duas** alternativas de

especificações de sintaxes concretas (OWL/XML e Manchester). Deste ponto em diante usaremos a sigla OWL para qualquer referência a OWL 2.0.

A partir das especificações desses documentos é possível estabelecer complexos mapeamentos de domínios, inclusive conectando diferentes domínios. Esta é uma característica relevante da OWL para o presente estudo, pois pode ser utilizada para prover o estabelecimento de relações entre dados de diferentes elementos informacionais, quer sejam bibliográficos ou não, o que pode ser aproveitado para a construção de estruturas complexa que abranjam um ou mais domínios semânticos. Inclusive suportando inúmeros relacionamentos entre os conceitos pertencentes a eles, contemplando até elementos informacionais de diferentes suportes, tais como: textuais, iconográficos, imagens, sons ou multimídia. Obvio compreender que a OWL, como modelo conceitual de dados, apenas estabelece uma forma para correlacionar, precisaremos ainda de um padrão que defina uma estrutura adequada para usar a OWL como parte de dados informacionais relativos a itens bibliográficos ou itens da WWW.

#### 2.2.4 O modelo Resource Description Framework - RDF

O modelo RDF (W3C, 2014) é um modelo para intercâmbio de dados na WWW que possui características que permitem que documentos em RDF possam ser combinados, mesmo quando os documentos não forem produzidos com o mesmo esquema RDF. A especificação do modelo é evolutiva, o que significa que novos dados podem ser anexados ou correlacionados sem que isso implique necessariamente em mudanças em todos os dados previamente existentes. A interrelação entre especificações em RDF ocorre através da utilização de recurso semelhante ao existente no modelo WWW, através de *Uniform Resources Identifier* (URI) ou *links*. Todavia o RDF amplia os *links* tradicionais da WWW ao associar nomes as relações entre os diferentes elementos (documentos, por exemplo) bem como para nomear as duas pontas da relação. Desta forma, por exemplo, no básico da especificação estarão presentes a identificação do ponto inicial, do ponto final e do tipo de relação que há entre esses pontos. Usando os dados básicos da especificação já se torna possível a construção de um grafo que expresse algo como um modelo mental do domínio que esteja sendo modelado. Esse modelo mental de domínio retrata a mesma possibilidade de construção de estruturas complexas discutidas na especificação OWL. Portanto, usando-se RDF ou OWL torna-se possível o controle de domínios semânticos, suportando inúmeros

relacionamentos entre os conceitos pertencentes a eles. Todavia a OWL possui vantagem sobre o RDF por nativamente permitir a expressão de relações típicas de ontologias, enquanto o RDF não tenha sido criado para esse fim. Na seção 3 discutiremos essa particularidade e como ela pode ser utilizada na proposta contida neste documento.

### 2.3 Processamento de Linguagem Natural e Inteligência Computacional

---

Se considerarmos a problemática abordada no presente texto certamente perceberemos que uma possível solução deverá transcender o conhecimento registrado das bibliotecas. Tal solução precisa ser aplicável aos acervos das Unidades de Informação, mas, principalmente ao conteúdo presente em recursos informacionais da WEB. A proposta de qualquer solução deve compreender documentos e fragmentos informacionais que possam ser disponibilizados em jornais, revistas, redes sociais, blogs e outros recursos que possam povoar o ecossistema informacional (De Souza; Fernandes, 2020, p. 145). Isso implica que o volume de documentos será muito maior do que o volume disponível em acervos tradicionais. Portanto, considerando-se esse volume, um potencial leitor deverá realizar um grande esforço e dedicar muito tempo para consultar e selecionar documentos de seu potencial interesse. Sabemos que o tempo necessário para que um sujeito informacional supra sua necessidade informacional afeta seu comportamento quanto a isso. Se no passado a pesquisa limitava-se aos domínios do acervo da biblioteca, hoje a pesquisa delimita-se pela dimensão do tempo disponível pelo interessado, visto que o acervo se tornou imenso. Segundo De Souza e Fernandes (De Souza; Fernandes, 2020, p. 146):

Há, portanto, um aumento do tempo necessário para que um potencial usuário de uma informação a encontre; obviamente isso é proporcional ao volume de informação disponível. Acreditamos, contudo, que é possível alterar essa relação com o tempo através do uso de tecnologia que minimize o volume de informações a serem apreendidas, quer seja pela diminuição do volume de texto a ser lido, ou pela alteração do tipo de código usado no suporte. Se um determinado documento puder ser transformado em outro tipo de documento, todavia preservando-se o valor semântico, poderíamos ter, por exemplo, uma condensação dos textos, criando versões de denso valor semântico que requereriam menor tempo para seu consumo. (De Souza; Fernandes, 2020, p. 146).

Uma solução nesse sentido foi proposta em De Souza et al. (2017) na qual é descrita uma solução baseada em PLN e IC que produz resumos de textos automaticamente, a solução proposta foi testada e os resultados positivos dos testes podem ser vistos em Tabosa et al. (2020). O PLN

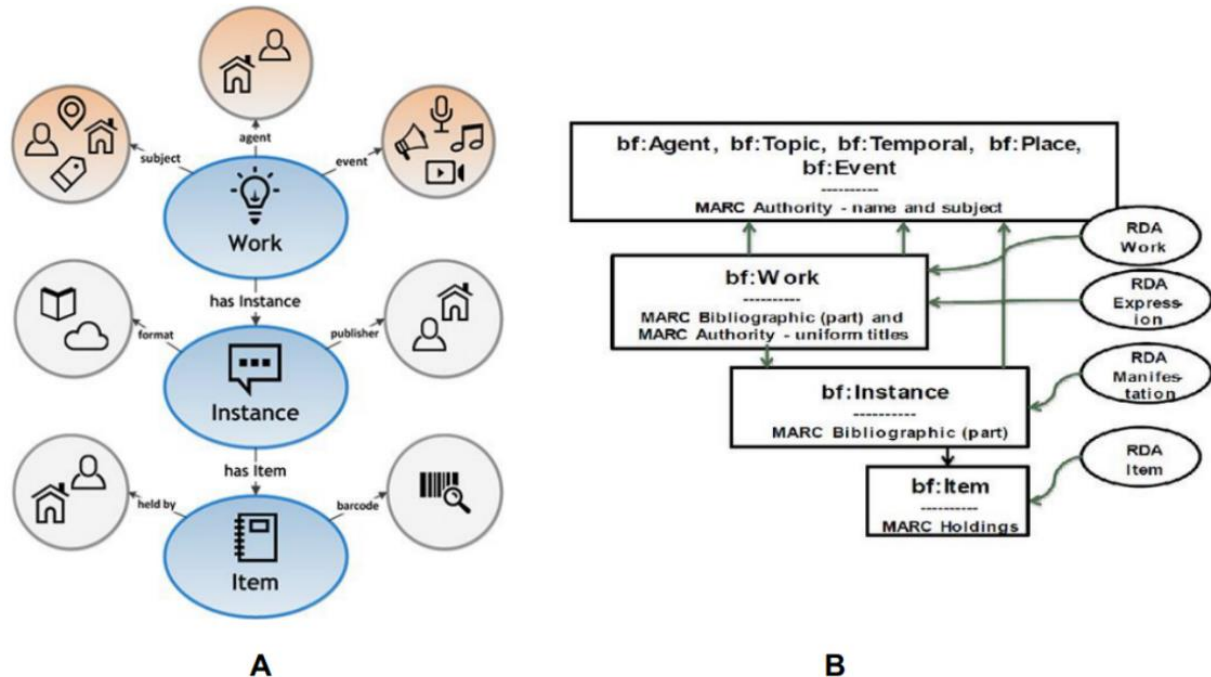
tem sido utilizado em casos nos quais deseja-se automatizar alguma tarefa de processamento de textos, quer seja uma trivial extração de dados ou sofisticados processamentos para obtenção dos valores semânticos presentes. Também tem sido aplicado para permitir diferentes tipos de interação entre pessoas e equipamentos e nesse tipo de solução, quase sempre o foco é direcionado ao usuário final, que é o potencial interessado na informação. A exemplo de casos nos quais podemos utilizar ferramentas baseadas em PLN, citamos a construção de sistemas de tradução automática (Peña; Bucheli; Gutiérrez, 2022), a criação de resumos de textos (De Souza et al, 2017), a análise de sentimentos (Silva; De Souza; Souza, 2020), a extração de informações (Peña; Bucheli; Gutiérrez, 2022) e por fim, a construção de assistentes pessoais, tais como a Alexa da empresa Amazon. Produtos baseados em soluções que utilizem o PLN são exemplos de como a área da IC contribui para a diminuição do esforço e tempo necessários para que um sujeito informacional possa realizar uma tarefa relacionada à compreensão ou obtenção de dados de textos. Além de soluções com foco em facilidades na seleção e consumo de documentos, há também esforços para trazer facilidades para autores. Um exemplo disso é o modelo teórico do *Generative Pre-trained Transformer* (GPT) em sua versão 3 (Brown et al, 2020), o qual tornou-se popular por facilitar tarefas de criação de conteúdo em formato textual ou multimídia. Ferramentas baseadas em GPT têm sido frequentes no apoio a redação de textos em diversas áreas, por exemplo, a jornalística, como se vê no trabalho de Noreen et al. (2023), bem como na área de administração de negócio no auxílio ao gerenciamento de processos de negócio, como se vê em Beheshti et al. (2023).

### **3 Convergências entre padrões: estrutural, representação, codificação**

Passamos a analisar algumas propriedades dos modelos discutidos na seção 2 para caminharmos em direção a uma proposta de uma estrutura que, como mencionamos anteriormente, consolide e relacione elementos informacionais estruturados em um denso esquema de metadados. O qual estabeleça uma estrutura taxonômica, referencial e verificável a respeito dos documentos e seus respectivos conteúdos, origens e autores. Precisaremos de uma estrutura extensível para o registro dessas relações entre documentos, de forma que ela suporte a construção de relações

taxonômicas entre diferentes documentos em diferentes suportes e em diferentes acervos, bases de dados ou outros tipos de recursos disponíveis na WWW.

Figura 3 - Níveis de abstração do BIBFRAME e a correspondência em RDA codificando MARC 21



Fonte: Parte (A) BIBFRAME - adaptado de Biblioteca do Congresso Americano (LOC, 2023), parte (B) RDA - adaptado de McCallum (McCallum, 2017)

Tal estrutura deve correlacionar textos, suas origens e seus valores semânticos e deve permitir um controle referencial sobre todos os conteúdos de maneira a estabelecer sua ancestralidade e ou descendência. Portanto, passamos a examinar os aspectos desses padrões que são relevantes para esse fim.

Analisando-se a estrutura do BIBFRAME, Figura 3 (A) podemos perceber que existem 3 níveis de abstrações relativas à representação de um elemento informacional. Tais abstrações permitem relacionar elementos do acervo a outras informações internas ou externas, quer sejam objetos reais ou intangíveis, como por exemplo: agente, evento ou assunto. Tal relação, no entanto, estará limitada aos dados que forem inseridos no cadastro do item, sem haver necessariamente, uma conexão entre o item do acervo e por exemplo, uma página na internet relativo a um assunto, agente ou evento que esteja correlacionado por semelhança textual, semântica, autoral ou de área

pertinente ao item. A semelhança, se houver, é definida e limitada pelos esforços empreendidos no processamento técnico de tais itens e que sejam coincidentes com os termos de uma expressão de busca eventualmente formulada por um interessado. Pela estruturação do BIBFRAME percebemos que embora o mesmo ofereça muitas vantagens, ele opera com o mesmo princípio estrutural do MARC. O BIBFRAME organiza as informações em três níveis principais de abstração: Trabalho, Instância e Item (Zavalin; Zavalina; Miksa, 2021, p. 154) os quais são codificados utilizando-se um conjunto de propriedades na notação *Resource Description Framework* (RDF) (W3C, 2014). Um exemplo ilustrativo de como o BIBFRAME é organizado pode ser visto na Figura 3 (A). Nela a abstração *work* refere-se a um recurso do catálogo: autores, idiomas, objeto. Um *work* que exista em um acervo será representado por uma ou mais *instance*, a qual reflete um material em si. Um *work* pode existir em mais de um suporte e cada suporte será representado por sua *instance* específica. Por sua vez, para uma *instance* pode haver vários exemplares e cada um desses exemplares será um item.

No formato BIBFRAME podemos representar todos os conceitos relativos a obras que possam estar presentes no acervo ou dispersos em fontes de informações na WEB.

Todos esses dados estão estruturados em classes RDF, as quais incluem os três níveis principais de abstração do BIBFRAME, além de várias subclasses dos níveis principais, necessárias ao detalhamento da especificação de um recurso do catálogo.

Quanto à integração com as especificações anteriores ou de finalidade semelhante, a exemplo, como se pode ver na Figura 3 (B) em comparação visual com a Figura 3 (A), na especificação do BIBFRAME pode haver o mapeamento entre os níveis principais de abstração do BIBFRAME com as transcrições dos dados bibliográficos codificados em MARC 21 e estruturados, por exemplo, em RDA. Essa característica do BIBFRAME permite uma migração gradativa da codificação do MARC 21 para o modelo de dados RDA e eventualmente deste para o modelo de dados RDF, sendo este último, o padrão de modelo de dados para o BIBFRAME, ao passo que também se aproxima de uma integração com uma especificação baseada no par FRBR/RDA (Mccallum, 2017), aumentando a compatibilidade com outras especificações.



Portanto, pela análise, podemos concluir que o BIBFRAME pode ser utilizado como modelo de estruturação de um denso esquema de metadados, sendo possível integrar dados em diferentes estruturas, mantendo ampla possibilidade de integração. No entanto ele não possui, a priori, todos os metadados necessários para permitir uma relação hierárquica baseada em valores semânticos, sendo necessária a inclusão de mais metadados. Tais metadados devem permitir a inclusão de valores semânticos que sejam produzidos manualmente ou automaticamente, ao passo que, com esses novos metadados, seja criado um modelo conceitual de estrutura de domínios conectados, algo que foge ao propósito do BIBFRAME.

Quanto ao modelo conceitual de dados, considerando-se que OWL é mais abrangente do que o RDF, embora ambos sejam da mesma família de modelos e possam ser mesclados, o OWL desponta como a escolha mais promissora. Resta agora o projeto de uma arquitetura que seja baseada na estrutura do BIBFRAME e no modelo conceitual do OWL.

#### **4 O Heredograma Informacional para textos**

---

Os modelos discutidos nas seções anteriores apresentam características positivas para o registro de dados bibliográficos e para a integração do uso interno a uma unidade de informação, bem como para o uso externo em outras unidades. Mesmo considerando-se a integração proposta pelos modelos eles não são suficientes para comporem, isoladamente, ou mesmo em conjunto, uma solução que unifique os dados e venha a ser uma solução para o problema discutido neste texto.

É necessária a criação de uma nova arquitetura para integração entre os modelos discutidos, realizando isso pela inclusão de elementos estruturantes e elementos que atribuam características de unicidade aos conteúdos semânticos dos documentos e, ao mesmo tempo, também permitam a rastreabilidade de seus respectivos valores semânticos, tanto para seus ancestrais quanto para seus descendentes, estabelecendo uma família de documentos relacionados. Passamos a denominar a arquitetura resultante dessa nova integração pelo termo Heredograma Informacional (HI). A palavra heredograma em si tem seu uso inicial na área da genética e uma definição do termo pode ser vista em Michaelis (Michaelis, 2023), sendo utilizado para expressar a presença ou semelhança

ou ausência de características dentro de um conjunto de indivíduos da mesma ancestralidade. Estamos interessados apenas na noção de ancestralidade e descendência. Esse é justamente o conceito que buscamos com a definição da arquitetura proposta neste trabalho. Cada elemento informacional será visto como um elemento individual com características próprias, as quais são influenciadas pela sua ancestralidade e que influenciam sua própria descendência. Não se trata de uma aplicação de técnicas da genética no tratamento de textos, portanto, definiu-se como Heredograma Informacional ao invés de apenas Heredograma, justamente para afastar o termo do conceito pertinente a genética.

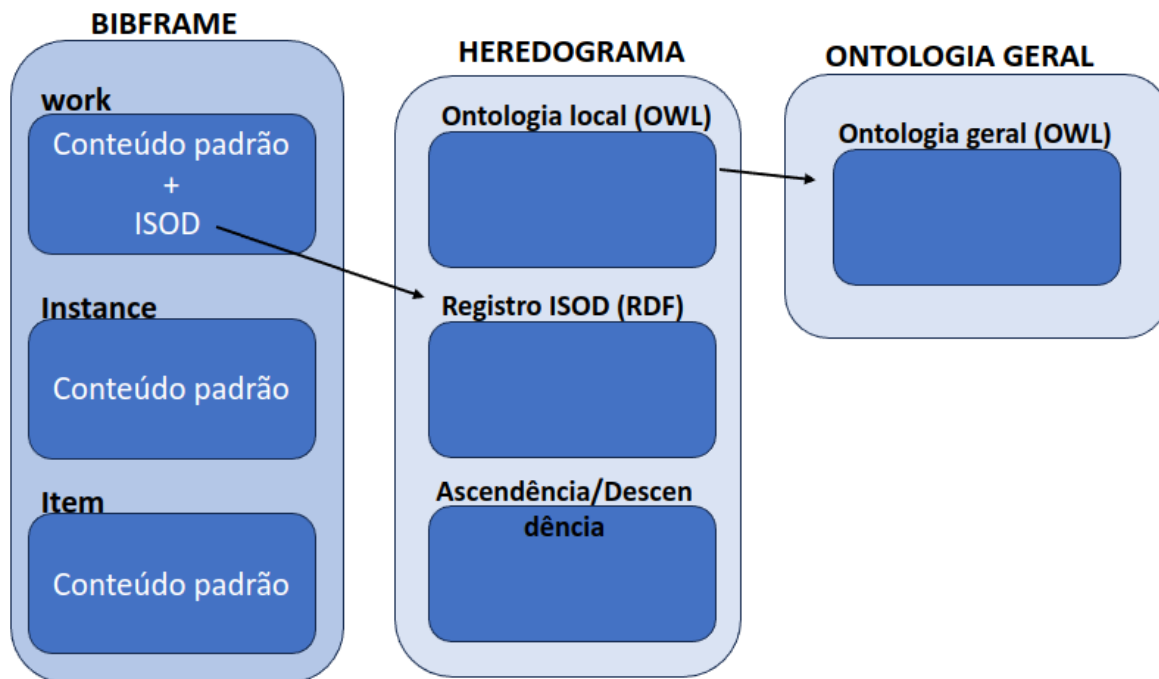
O HI deve permitir que para um determinado documento, denotado por  $D$ , possam ser encontrados todos os documentos,  $d_i$  que componham sua família, denotado por  $D_f$ . De tal forma que  $D_f = \{d_1, d_2, d_3, \dots, d_i, d_n\}$  possa ser estabelecido.

No estabelecimento de uma família os documentos que pertencerão a ela são selecionados a partir da presença de características comuns. Essas características são definidas em termos da análise dos valores semânticos contidos nos metadados que fazem parte da especificação do HI. Um determinado documento pode fazer parte de inúmeras famílias, o que garante a relação entre documentos de um mesmo domínio e de domínios diferentes. A definição de anterioridade ou posterioridade, dentro e fora das famílias deverá, obviamente, ser relacionada a temporalidade dos documentos analisados.

A ascendência e a descendência estão implícitas na definição da família, pois  $D_f$  trata apenas do agrupamento dos documentos que pertencem a ela, todavia, estão explícitas na especificação do HI, como podemos ver mais adiante na Figura 5.

Na Figura 4 podemos ver a ilustração de uma arquitetura de composição, estruturada em BIBFRAME, com ontologia para o controle semântico modelada em OWL, com uma proposta de registro de Identificação de unicidade Semântica de Objeto Digital (ISOD) encapsulado no formato proposto do HI, o qual será modelado em RDF. Ela também sugere que no registro BIBFRAME de um determinado documento será incorporado um metadado relativo a um identificador ISOD, cuja finalidade é permitir o estabelecimento de uma integração entre dados estruturados em BIBFRAME e dados contidos no HI.

Figura 4 – Arquitetura de composição entre BIBFRAME, HI, OWL e RDF



Fonte: Autor (2023)

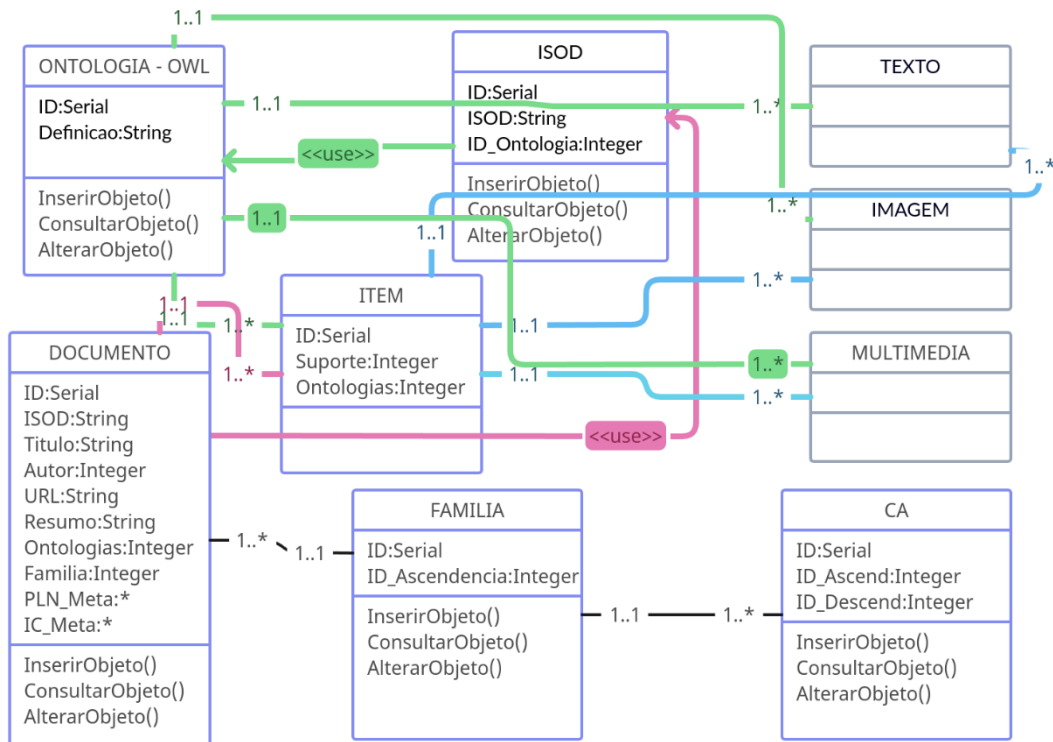
No HI o documento terá uma descrição no formato de uma ontologia local (OL), que descreve os elementos semânticos (ES) percebidos no documento. Esses ES são obtidos e mapeados em uma ontologia geral (OG), a qual reúne todos os conceitos semânticos que vão sendo incorporados a cada novo documento que venha a ser inserido no domínio. Isto produzirá um conjunto denso de domínios conectados a seus respectivos valores semânticos. Os metadados da OL auxiliarão no controle de ascendência (e de descendência) (CA). Documentos que possuam OL semelhantes tem grande probabilidade de pertencerem a um mesmo domínio semântico.

Neste ponto é importante enfatizar que, embora o BIBFRAME seja flexível e permita uma abrangência adequada para os dados bibliográficos, ele é voltado para os metadados produzidos manualmente no processamento técnico. A Figura 5, vista adiante, ilustra uma visão mais detalhada da parte da arquitetura relacionada ao HI.

A inclusão dos metadados contidos no HI justificam-se ao ampliar as possibilidades de novos valores semânticos, constituindo assim uma especificação que possa evoluir a partir da incorporação de ações autônomas de aquisição de dados, provenientes de recursos da IC, da IA e

do PLN, ou de outros esforços da inteligência artificial, tais como o GPT, citado aqui apenas como um exemplo. Os resultados das ações autônomas que são relacionados a parte semântica serão inseridos no registro ISOD do HI, que é modelado em RDF. Cada documento que seja registrado no HI será inserido no diagrama DOCUMENTO, como se vê na Figura 5. Observe que um documento pode ser composto de diversos itens, tais como textos, imagens e conteúdo multimídia, como se vê na Figura 5, nos diagrama ITEM, TEXTO e MULTIMIDIA.

Figura 5 - Detalhamento da arquitetura da parte do HI



Fonte: Autor (2023)

Os valores semânticos que indicam os domínios dos documentos serão utilizados na determinação das famílias as quais um documento pertence e os metadados relativos a isso serão inseridos no diagrama FAMÍLIA, como ilustrado na Figura 5. Uma vez que o documento seja inserido em uma família serão determinadas as suas ascendências e descendências, metadados relativos a isso serão inseridos no HI, como se vê na Figura 5 no diagrama CA. Desta forma o HI também poderá ser visualizado como um grafo ou modelo (mapa) mental, tendo em vista que essa é uma das propriedades do RDF. O CA mencionado anteriormente neste texto é visto na Figura 5

no diagrama CA e permite a construção das relações entre os documentos, construindo e mantendo sua respectiva família, denotado por  $D_f$  e utilizando para esse fim os metadados do registro ISOD, ilustrado na Figura 5 no diagrama ISOD e da OL que é ilustrado no diagrama ONTOLOGIA - OWL da mesma figura. Na Figura 5 as linhas de cor verde sinalizam o fluxo de informações da parte semântica, as de cor azul sinalizam o fluxo extrutural dos documentos, enquanto a parte rosa sinaliza o fluxo de identificação única de cada documento e seus respectivos itens.

Elementos semânticos que podem ser produzidos automaticamente e que podem ser registrados no ISOD destacamos:

- Detecção de similaridade entre os textos avaliando-se os seguintes critérios:
  - autoria e direito de propriedade expressas de maneira explícita no texto, ou inferidos através de relações de semelhança estatística de escrita (De Souza et al, 2017) e atribuição de grau de certeza na inferência;
  - elementos semânticos tais como: a) equivalência ou semelhança estatística de escrita obtidos a partir da extração subtropical; b) equivalência ou semelhança estatística de conteúdo do texto, obtida através de técnica de PLN profunda; c) equivalência ou semelhança estatística de elementos morfológicos predominantes, obtida através de técnica de PLN superficial; d) análise estatística de simples correlação textual pela intercorrência morfo-sintática entre os sintagmas.

É importante compreender que o HI proposto foi projetado de maneira a ser amplamente aplicável ao acervo de unidades de informação e seus conhecimentos registrados, não se limitando a eles e, de fato, estando preparado para contemplar documentos e dados textuais provenientes de outras fontes de informações na WEB, inclusive com fenômenos informacionais negativos, em parte respondendo a questões previamente discutidas aqui como as vistas em (Vidotti et al, 2019, p. 12), bem como os problemas que ocorreram durante a pandemia da COVID-19, caracterizados por atos de desinformação como já vimos em (Mazzeto; Souza, 2022, p. 2-23) e em (Brisola; Bezerra, 2018). A atribuição de um ISOD garantirá a existência de propriedades de rastreabilidade, sendo verificável e referenciável, o que habilita o modelo a prover informação sobre a historicidade

do texto, sua ancestralidade e descendência, bem como informação autoral, ainda que implícita e limitada ao grau de certeza definido estatisticamente. Além disso, as ontologias locais e globais e a estrutura referencial do heredograma é extensível, possibilitando a construção de uma estrutura flexível.

A organização do conhecimento proveniente dos documentos através do HI proposto neste documento poderá contribuir para a desinformação, pois ao promover a organização hierárquica do CA, os textos poderão ser analisados com maior rapidez, verificando-se se os mesmos possuem informações validadas pelos documentos ancestrais. Também contribui na medida em que facilitará a identificação da autoria dos documentos, pois o estabelecimento do CA pode-se saber exatamente o ponto no tempo em que o um determinado documento foi publicado. Cria-se assim um conjunto de recursos de análise dos documentos que permitirão agilidade na classificação de um documento quanto a fidedignidade, autenticidade e completude de um documento em relação a toda sua história.

A Figura 6 ilustra um mapa mental para capacidade de extensão e de rastreabilidade providas pelo ISOD. Assim haverá a possibilidade de identificar a autoria e as pegadas digitais dos textos, permitindo uma compreensão mais profunda desses fenômenos informacionais.

Figura 6 - O encadeamento de números ISOD



Fonte: Autor (2023)

A partir da rastreabilidade será possível a verificação e referência, além da visualização rápida do HI de um documento. Visualizações poderão ser textuais ou gráficas, permitindo uma navegação confortável para o potencial usuário dos dados do heredograma.

## 5 Considerações finais

---

Na presente investigação avaliamos os principais modelos conceituais e padrões de estruturação de registros bibliográficos e assim pudemos perceber alguns caminhos para contribuir na mitigação dos problemas relacionados à desinformação, inclusive as não intencionais.

Acreditamos que a presente proposta pode contribuir para discussões no tema ao apontar caminhos metodológicos relacionados à organização e representação do conhecimento.

Ao agregar elementos informacionais em um esquema de metadados baseado no BIBFRAME, RDA e RDF os quais foram combinados para definir uma nova abordagem denominada de HI, foi possível a proposta de desenvolvimento de uma contribuição teórico-metodológica para impactar positivamente na recuperação do conhecimento. A proposta possui propriedades de uma estrutura taxonômica, verificável e referenciável sobre os documentos e seus respectivos conteúdos. Nela considera-se a atribuição de um ISOD para dar unicidade aos conteúdos dos documentos, recurso esse que pode ser usado em associação com o DOI ou em substituição desse. A proposta foi projetada para ser utilizada com documentos e textos provenientes da WEB, mesmo que não sejam documentos registrados.

A integração com sistemas de informações de gerenciamento de registros bibliográficos, mesmo que tais sistemas sejam antigos, foi contemplada, o que potencializa as possibilidades de impacto positivo da proposta de heredograma presente neste trabalho.

## Referências

---

- Beheshti, Amin; Yang, Jian; Sheng; Benatallah, Boualem; Casati, Fabio; Dustdar, Schahram; Motahari, Hamid Reza Nezhad; Zhang Xuyun; Shan Xue. Processgpt: Transforming Business Process Management With Generative Artificial Intelligence. In Ieee International Conference On Web Services (Icws). 2023. Disponível em <https://arxiv.org/abs/2306.01771> Acesso em 10 Set. 2023.
- Boyd, Daith; Crawford, Keneth. Critical Questions For Big Data: Provocations For A Cultural, Technological, And Scholarly Phenomenon. *Information, Communication & Society*, 15(5), P. 662-679. 2012. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>. Acesso em: 19 mar 2023.

- Brisola, Anna Cristina; Bezerra, Arthur Coelho. Desinformação E Circulação De “Fake News”: Distinções, Diagnóstico E Reação. Encontro Nacional De Pesquisa E Pós-graduação Em Ciência Da Informação, 19, 2018. Anais Eletrônicos [...]. Londrina, Pr: Ancib/Uel, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/102819>. Acesso em: 20 mar. 2023.
- Brown, Tom; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared; Dhariwal, Prafulla; Neelakantan, Arvind; Shyam, Pranav; Sastry, Girish; Askell, Amanda; Agarwal, Sandhini; Hebervoss, Ariel; Krueger, Gretchen; Henighan, Tom. Language Models ARE Few-shot Learners. Openia. 2020. Disponível em: <https://arxiv.org/pdf/2005.14165.pdf>. Acesso em 19 mar. 2023.
- CAPES, Portal de Periódicos da Capes. 2023. Disponível em: <https://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br/index.php?> Acessado em: 20 mar. 2023.
- Chen, Peter. The Entity-relationship Model - Toward A Unified View OF Data. *Acm Transactions ON Database Systems*. 1 (1): P. 9–36. 1976. Disponível em: <https://citeseerx.ist.psu.edu/doc/10.1.1.523.6679>. Acesso em 21 mar. 2023.
- Deursen, Van; Helsper. (2015). The Third-level Digital Divide: Who Benefits Most From Being Online?. *Communication & Information Technologies Annual*, 10, P. 29-52. 2015. Disponível em: <http://10.1108/S2050-206020150000010003>. Acesso em 19 mar. 2023.
- DC. The Dublin Core specification, 1998. <https://www.dublincore.org/specifications/dublin-core/>. Acesso em 21 mar. 2023
- Greenberg, Jane. Understanding metadata and metadata schemes. *Taylor & Francis Journals Complete Cataloging & classification quarterly*. v.40, p.17-36. 2005. Disponível em: [https://capes-primoz.ezl.periodicos.capes.gov.br/primoz-explore/fulldisplay?docid=TN\\_cdi\\_pascalfrancis\\_primary\\_17286158&context=PC&vid=CAPES\\_V3&lang=pt\\_BR&search\\_scope=default\\_scope&adaptor=ate2&offset=0](https://capes-primoz.ezl.periodicos.capes.gov.br/primoz-explore/fulldisplay?docid=TN_cdi_pascalfrancis_primary_17286158&context=PC&vid=CAPES_V3&lang=pt_BR&search_scope=default_scope&adaptor=ate2&offset=0). Acesso em 20 mar. 2023.
- IFLA. Functional Requirements for Bibliographic Records (FRBR), 2009. Disponível em: [https://cdn.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr\\_2008.pdf](https://cdn.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr_2008.pdf). 2009. Acesso em: 21 mar. 2023.
- IFLA. Library Reference Model (IFLA-LRM), 2017. Disponível em: [https://cdn.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr\\_2008.pdf](https://cdn.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr_2008.pdf). 2009. Acesso em: 21 mar. 2023.
- LOC, The MARC21 specification. 1996. Disponível em: <https://www.loc.gov/marc/96princip.html>. Acesso em 20 mar. 2023.
- LOC, The MODS specification. 2022. Disponível em: <https://www.loc.gov/standards/mods/userguide/introduction.html> Acesso em 20 mar. 2023.



LOC, Overview of the BIBFRAME 2.0 Model. 2023. Disponível em:

<https://www.loc.gov/bibframe/docs/bibframe2-model.html> Acesso em: 20 mar. 2023.

Mazzeto, Ana Carla Epitácio; Souza, Elisabete Gonçalves De. Infodemia E Desinformação No Contexto Da Pandemia Da Covid-19: Reflexões À Luz Da Noção De Competência Em Informação. Ponto de Acesso, [S. l.], v. 16, n. 2, p. 2–23, 2022. DOI: <https://doi.org/10.9771/rpa.v16i2.49151>. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/49151> Acesso em 20 mar. 2023.

Mccallum, Sally. Bibframe Development. JLIST.it. 2017. Disponível em:

[https://www.researchgate.net/publication/320142396\\_BIBFRAME\\_development/fulltext/60b824dc\\_a6fdccb96f4d88f8/BIBFRAME-development.pdf?origin=figuresDialog\\_download&rtd=e30%3D&tp=eyJwYWdlIjoicHVibGljYXRpb25EZXRhaWwifQ](https://www.researchgate.net/publication/320142396_BIBFRAME_development/fulltext/60b824dc_a6fdccb96f4d88f8/BIBFRAME-development.pdf?origin=figuresDialog_download&rtd=e30%3D&tp=eyJwYWdlIjoicHVibGljYXRpb25EZXRhaWwifQ). Acesso em 20 mar. 2023.

Michaelis. Dicionário Online multi idiomas, 2023. Disponível em:

<https://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=heredograma>. Acesso em 21 mar. 2023.

Noureen, Fatima; Muhammad, Sher ; Zenun , Kastrati; Shariq , Imran, Ali; Saif, Hassan; SABRI, Elmitwally, Nouh. Improving news headline text generation quality through frequent POS-Tag patterns analysis. Engineering Applications Of Artificial Intelligence. V.125 P.106718. 2023. Disponível em

[https://link.periodicos.capes.gov.br/sfxlcl41/cgi/core/sfxresolver.cgi?request\\_id=123669395&tmp\\_ctx\\_svc\\_id=1&service\\_id=111026921949001&tmp\\_ctx\\_obj\\_id=1](https://link.periodicos.capes.gov.br/sfxlcl41/cgi/core/sfxresolver.cgi?request_id=123669395&tmp_ctx_svc_id=1&service_id=111026921949001&tmp_ctx_obj_id=1). Acesso em 10 Set. 2023

Peña, Jefferson Anderson Torres; Bucheli, Victor; Gutiérrez DE Piñerez Reyes, Raúl. Traducción automática de un conjunto de entrenamiento para extracción semántica de relaciones. in Cuadernos de Linguística. 2022. Disponível em:

[https://revistas.uptc.edu.co/index.php/linguistica\\_hispanica/article/view/13436](https://revistas.uptc.edu.co/index.php/linguistica_hispanica/article/view/13436). Acesso em: 20 mar. 2023.

SC. About RDA. Steering Committee. 2005. Disponível em: <http://rda-rsc.org/content/about-rda>. Acesso em 23 mar. 2023.

Silva, Arnaldo Nunes da; Souza, Osvaldo de; Souza, Jose Neuman de. Sentiment Parser based on X-Bar Theory to Brazilian Portuguese. in Conference: 2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE). 2020. Disponível em:

[https://www.researchgate.net/publication/346388177\\_Sentiment\\_Parser\\_based\\_on\\_X-Bar\\_Theory\\_to\\_Brazilian\\_Portuguese](https://www.researchgate.net/publication/346388177_Sentiment_Parser_based_on_X-Bar_Theory_to_Brazilian_Portuguese). Acesso em 20 mar. 2023.

- Souza, Osvaldo de; Fernandes, Joana D'arc Páscoa Bezerra. Discutindo a acessibilidade informacional no ecossistema digital. In: *Ciência da Informação em Contextos*. Fortaleza: Imprensa Universitária, 2020. p. 141-163.
- Souza, Osvaldo de; Tabosa, Hamilton Rodrigues.; Oliveira, Maria De; Oliveira, M. H. De. Um Método de Sumarização Automática de Textos Através de Dados Estatísticos e Processamento de Linguagem Natural. *Informação & Sociedade: Estudos*, João Pessoa, v. 27, n. 3, p. 307-320, Set./Dez. 2017.
- Tabosa, Hamilton Rodrigues.; De Souza, Osvaldo.; Cândido, José Carlos; Melo, Ana Cristina Azevedo Ursulino; Reis, Keila Giulliana Braga. Avaliação do desempenho de um software de sumarização automática de textos. *INF. INF. Londrina*. v. 25. p. 189-210. 2020. Disponível em: [https://www.researchgate.net/publication/342727034\\_Avaliacao\\_do\\_desempenho\\_de\\_um\\_software\\_de\\_sumarizacao\\_automatica\\_de\\_textos](https://www.researchgate.net/publication/342727034_Avaliacao_do_desempenho_de_um_software_de_sumarizacao_automatica_de_textos). Acesso em 20 mar. 2023.
- Vidotti, Silvana Aparecida Borsetti Gregorio; Coneglian Caio Saraiva; Roa-Martínez Sandra Milena; Vechiato Fernando Luiz; Santarém Segundo José Eduardo. WEB, WEB SEMÂNTICA E WEB PRAGMÁTICA: um posicionamento da Arquitetura da Informação. *Inf. & Soc.:Est.*, João Pessoa, v.29, n.1, p. 195-214, jan./mar. 2019. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/44358/22386>. Acesso em 5 mar. 2024.
- Wu, Singuth. (2023). Implementing Bibliographic Enhancement Data in Academic Library Catalogs: An Empirical Study [Data set]. Taylor & Francis. 2023. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/01639374.2023.2224781?scroll=top&needAccess=true&role=tab>. Acesso em 25 jun. 2023.
- W3C, W3C Web Ontology Language. 2004. Disponível em: <https://www.w3.org/OWL/>. Acesso em: 23 mar. 2023.
- W3C, W3C Web Ontology Language 2.0. 2012. Disponível em: <https://www.w3.org/TR/2012/REC-owl2-overview-20121211/>. Acesso em: 23 mar. 2023.
- W3C, RDF 1.2 Schema specification. 2014. Disponível em: <https://www.w3.org/RDF/>. Acesso em: 20 mar. 2023.
- Zavalin, Vyacheslav; Zavalina, Oksana L; Miksa, Shawne. Exploration of Subject Representation and Support of Linked Data in Recently Created Library Metadata: Examination of Most Widely Held WorldCat Bibliographic Records. *Library resources & technical services*. Vol.65 (4), p.154. 2021. Disponível em: [https://capex-primo.ezl.periodicos.capes.gov.br/primo-explore/fulldisplay?docid=TN\\_cdi\\_proquest\\_journals\\_2604707199&context=PC&vid=CAPES\\_V3&lang=pt\\_BR&tab=default\\_tab&query=any%2Ccontains%2Cbibliographics%20records%20representation&offset=0](https://capex-primo.ezl.periodicos.capes.gov.br/primo-explore/fulldisplay?docid=TN_cdi_proquest_journals_2604707199&context=PC&vid=CAPES_V3&lang=pt_BR&tab=default_tab&query=any%2Ccontains%2Cbibliographics%20records%20representation&offset=0). Acesso em: 21 mar. 2023.

Zapounidou, Sofia; Sfakakis, Michalis; Papatheodorou, Christos. Representing and integrating bibliographic information into the Semantic Web: A comparison of four conceptual models. 2017.

Disponível em:

[https://www.researchgate.net/publication/304070371\\_Representing\\_and\\_integrating\\_bibliographic\\_information\\_into\\_the\\_Semantic\\_Web\\_A\\_comparison\\_of\\_four\\_conceptual\\_models](https://www.researchgate.net/publication/304070371_Representing_and_integrating_bibliographic_information_into_the_Semantic_Web_A_comparison_of_four_conceptual_models). Acesso em 20

mar. 2023.

---

Copyright: © 2024 SOUZA, Osvaldo de. This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

---

Received: 27/10/2023

Accepted: 05/03/2023