

**FUNDAÇÃO DE ENSINO “EURÍPIDES SOARES DA ROCHA”  
CENTRO UNIVERSITÁRIO EURÍPIDES DE MARÍLIA – UNIVEM  
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

**ANA CAROLINE SOUZA PAULO FACHINI**

**PROPOSTA DE PLATAFORMA DE RECUPERAÇÃO DA  
INFORMAÇÃO EM FONTES INFORMACIONAIS ACADÊMICAS**

**MARÍLIA  
2015**

**ANA CAROLINE SOUZA PAULO FACHINI**

**PROPOSTA DE PLATAFORMA DE RECUPERAÇÃO DA  
INFORMAÇÃO EM FONTES INFORMACIONAIS ACADÊMICAS**

Trabalho de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação da Fundação de Ensino “Eurípides Soares da Rocha”, mantenedora do Centro Universitário Eurípides de Marília – UNIVEM, como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Orientador  
Prof<sup>o</sup>. Dr. Elvis Fusco

**MARÍLIA  
2015**

FACHINI, Ana Caroline Souza Paulo

**Proposta de Plataforma de Recuperação da Informação em fontes Informacionais Acadêmicas/** Ana Caroline Souza Paulo; orientador: Profº. Dr. Elvis Fusco. Marília, SP: [s.n.], 2015.

75 folhas

Monografia (Bacharelado em Sistemas de Informação): Centro Universitário Eurípides de Marília.



CENTRO UNIVERSITÁRIO EURÍPIDES DE MARÍLIA - UNIVEM  
MANTIDO PELA FUNDAÇÃO DE ENSINO "EURÍPIDES SOARES DA ROCHA"

BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Ana Caroline Souza Paulo Fachini

Proposta de Plataforma de Recuperação da Informação em Fontes Informacionais Acadêmicas.

Banca examinadora da monografia apresentada ao Curso de Bacharelado em Sistemas de Informação do UNIVEM/F.E.E.S.R., para obtenção do Título de Bacharel em Sistemas de Informação.

Nota: 6,5 (seis e meio)

Orientador: Elvis Fusco [assinatura]

1º. Examinador: Fabio Piola Navarro [assinatura]

2º. Examinador: Caio Saraiva Coneglian [assinatura]

Marília, 01 de dezembro de 2015.

*Dedico este trabalho a Deus, meus pais, Valdir e Suzy, e meu esposo, Felipe Fachini, pela  
confiança e apoio durante toda essa jornada.*

## **AGRADECIMENTOS**

Agradeço, primeiramente a Deus que permitiu que tudo isso acontecesse, ao longo de minha vida, e não somente nestes anos como universitária, mas que em todos os momentos, principalmente por ter concedido a mim uma família mais que abençoada e um esposo maravilhoso que sempre acreditaram em mim e incentivaram-me.

Aos meus pais, Valdir e Suzy, que apesar de todas as dificuldades me fortaleceram e investiram em mim o tempo todo, que para mim foi muito importante para continuar firme nessa caminhada.

Obrigada amor, Felipe Fachini, por estar ao meu lado me incentivando e sonhando comigo, obrigada mesmo pela paciência, pelo companheirismo e principalmente pelo seu amor.

Ao meu orientador, Dr. Elvis Fusco, pelo empenho dedicado à elaboração deste trabalho.

Agradeço a todos os professores por me proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional, por tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender. A palavra mestre, nunca fará justiça aos professores dedicados aos quais sem nominar terão os meus eternos agradecimentos.

A todos que direta ou indiretamente fizeram parte da minha formação, muito obrigada.

*“Que os vossos esforços desafiem as impossibilidades, lembrai-vos de que as grandes coisas do homem foram conquistadas do que parecia impossível.”*

*Charles Chaplin*

## RESUMO

Com o aumento da produção de informações no mundo surge a necessidade de transformar um grande volume de dados estruturados e não estruturados em informações úteis para tomada de decisão. Estas informações dentro do meio acadêmico podem estar espalhadas por repositórios digitais, sistema de gestão de aprendizagem, revistas e periódicos e a coleção do acervo de Bibliotecas digitais e não digitais. Neste contexto, esta pesquisa tem como objetivo propor uma arquitetura que realize em um ambiente centralizado, a recuperação das informações que estão inseridas nestas diversas fontes de dados.

**Palavras-Chave:** Repositórios Digitais, Recuperação da Informação, Ambientes Informacionais Digitais.



## **ABSTRACT**

With the increase in the production of information in the world arises the need to transform a large volume of structured and unstructured data into useful information for decision-making. This information inside the academia can be spread across digital repositories, learning management system, magazines and periodicals and collection of digital and non-Digital Libraries. In this context, this research aims to propose an architecture that perform in a centralized environment, the recovery of information that are included in these data sources.

**Keywords:** Digital Repositories, Information Retrieval, Digital Information Environments.

## LISTA DE ILUSTRAÇÕES

Figura 1: Componentes de um sistema de recuperação de informação.....	19
Figura 2: Registro do repositório digital Univem Aberto do Centro Universitário Eurípides de Marília .....	27
Figura 3: Repositório digital Univem Aberto.....	27
Figura 4: Moodle Univem. ....	29
Figura 5: Portal de Revistas Eletrônicas do Univem.....	30
Figura 6: Estrutura Tecnológica de um Acervo Digital. ....	32
Figura 7: Correspondência entre padrões de metadados (CROMWELL; KESSLER, 1998). .	39
Figura 8: Padrões de Metadados.....	40
Figura 9: Metadata Harvesting (GARCIA; SUNYE (2007,p.4)) .....	41
Figura 10: Proposta de Plataforma de Recuperação da Informação.....	43
Figura 11: Recuperação da Informação utilizando a ferramenta VuFind. (Fonte: VuFind, wiki.ibict.br) .....	44
Figura 12: Tela de Autoconfiguração do VuFind.....	47
Figura 13: Configuração ILS - VuFind. ....	48
Figura 14: Tela de Busca VuFind.....	49

## LISTA DE ABREVIATURAS E SIGLAS

DC	Dublin Core
DCMI	Dublin Core Metadata Initiative
DP	Data Providers
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
ISO	International Standard Organization
LSM	Linux Security Modules
MFC	Meta Content Framework
NEAD	Núcleo de Educação à Distância
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OJS	Open Journal Systems
RDs	Repositórios Digitais
RDF	Resource Description Framework
SELinux	Security-Enhanced Linux
SP	Service Providers
SRI	Sistema de Recuperação de Informação
XML	eXtensible Markup Language
W3C	World Wide Web Consortium

# SUMÁRIO

INTRODUÇÃO .....	14
1 RECUPERAÇÃO DA INFORMAÇÃO .....	17
1.1 Definição .....	18
1.2 Modelos Clássicos.....	19
1.3 Modelo Quantitativo .....	20
1.3.1 Modelo Booleano .....	20
1.3.2 Modelo Vetorial .....	21
1.3.3 Modelo Probabilístico .....	21
1.4 Modelo Dinâmico.....	22
1.4.1 Algoritmos Genéticos e Redes Neurais.....	22
2 FONTES INFORMACIONAIS ACADÊMICAS.....	24
2.1 Definição .....	24
2.2 Repositórios Digitais .....	24
2.2.1 DSpace .....	26
2.3 Sistema de Gestão de Aprendizagem .....	27
2.3.1 Moodle .....	28
2.4 Revistas Eletrônicas .....	29
2.4.1 Open Journal Systems .....	29
2.5 Biblioteca Digital .....	30
2.5.1 Acervo Digital .....	31
2.6 Biblioteca Não Digital.....	32
2.6.1 Acervo não Digital .....	32
3 METADADOS.....	33
3.1 Definição .....	33
3.2 Padrão de Metadados .....	33
3.2.1 Dublin Core.....	34
4 Interoperabilidade e Integração .....	36
4.1 Definição .....	37

4.1.1	Níveis de Interoperabilidade .....	37
4.2	Coleta de Metadados .....	38
4.2.1	Open Archives Initiative Protocol for Metadata Harvesting .....	40
5	PROPOSTA DE PLATAFORMA DE RECUPERAÇÃO DA INFORMAÇÃO EM FONTES INFORMACIONAIS ACADÊMICAS .....	42
5.1	FERRAMENTA VUFIND.....	44
5.1.1	Funções VuFind .....	45
5.1.2	Componentes.....	45
5.2	Instalação do Vufind .....	47
5.3	Testes Efetuados.....	47
	CONCLUSÃO .....	50
	REFERÊNCIAS .....	51
	Apêndice A.....	55
	VuFind no Fedora Linux .....	55
	Apêndice B.....	62
	VuFind no Ubuntu Linux .....	62
	Apêndice C.....	69
	VuFind no Windows .....	69

## INTRODUÇÃO

Com o crescimento do volume de informações, ao longo dos anos, gerou-se a necessidade de desenvolver técnicas de recuperação de informação para atender às demandas dos usuários. Estas informações dentro do meio acadêmico podem estar distribuídas por repositórios digitais, sistema de gestão de aprendizagem, revistas, periódicos e coleção do acervo de Bibliotecas Digitais e não Digitais.

Outrossim, o desenvolvimento do volume de publicações e também das necessidades de informações dos usuários, sejam elas em papel ou em formato eletrônico, é importante que os ambientes informacionais digitais possuam sistemas de informações capazes de armazenar e indexar informações bibliográficas de forma a facilitar a recuperação e disseminação aos usuários (CARDOSO, 2000). Desta maneira, há a necessidade de fontes acadêmicas mais robustas, visando facilitar o armazenamento e conseqüentemente a recuperação dessas informações, evitando assim que se percam.

Atualmente, grande parte dos alunos possuem acesso à rede de Internet, gerando uma constante troca de informações e de conhecimento, destarte, a busca por informações tem aumentado em ambientes acadêmicos brasileiros, em especial no nível superior (SANTARÉM SEGUNDO, 2010).

O acesso aos grandes sistemas de recuperação de informação, também denominados de bancos de dados e, às suas bases de dados ampliou a qualidade das buscas bibliográficas, visto que essas bases proporcionam diversificados pontos de acesso à informação (LOPES, 2002).

Neste contexto, uma solução seria uma arquitetura que realize em um ambiente centralizado, com intuito de suprir as necessidades do usuário de fontes informacionais acadêmicas, baseando-se em ambientes informacionais digitais abertos, como, repositórios digitais, sistema de gestão de aprendizagem, revistas e periódicos e a coleção do acervo de Bibliotecas digitais e não digitais, de forma que seja possível realizar uma extração das informações mais completa, tendo assim, maiores chances de satisfazer as necessidades informacionais do usuário.

A ferramenta Vufind é um software livre, que permite a integração de ferramentas isoladas, de forma a facilitar o acesso à informação, através do apoio a criação de portais integrados de informações geridas pelas bibliotecas.

### **Motivação e Justificativa**

Com o aumento do número de documentos e usuários na Web, modelos para recuperação de informações passaram a ser de suma importância, devido à heterogeneidade das diversas fontes informacionais, buscando a padronização da recuperação. Nesse contexto, o grande volume de dados exige processos de recuperação cada vez mais sofisticados, deste modo, faz-se o questionamento de como o processo de recuperação de informação possa trazer informações mais relevantes ao usuário, de forma que, este consiga ter suas necessidades informacionais atendidas, sem ter a necessidade de o usuário ficar realizando diversas buscas em fontes distintas.

A problemática principal desse trabalho está relacionada ao fato de que o usuário de fontes informacionais acadêmicas realiza a mesma pesquisa diversas vezes em fontes, isso acarreta em possíveis perdas de informações e tempo do usuário.

Portanto, visa-se suprir essa problemática com a proposta da plataforma de Recuperação de Informação baseado em ambientes informacionais digitais abertos, de forma que seja possível realizar uma extração das informações mais completa e variada, tendo assim, maiores chances de satisfazer as necessidades informacionais do usuário.

### **Objetivo**

O objetivo desta pesquisa é propor uma plataforma de Recuperação de Informação baseada em ambientes informacionais digitais abertos, como repositórios digitais, ambientes de aprendizagem a distância, sendo possível realizar uma extração das informações mais completa e variada, pois as fontes buscadas serão diversas, tendo assim, maiores chances de satisfazer as necessidades informacionais do usuário.

### **Metodologia**

A metodologia para pesquisa e desenvolvimento deste projeto pode ser dividida em duas etapas principais. Estas que serão descritas a seguir:

1. Pesquisa.
  - 1.1. Estudo aprofundado do tema abordado.
  - 1.2. Estudo de pesquisas e artigos correlatos.
  - 1.3. Pesquisas de padrões de metadados.
  - 1.4. Estudo aprofundado da ferramenta VuFind.
2. Escrita da Monografia abordando os resultados obtidos.

### **Trabalhos Correlatos**

Denton e Coysh aplicaram teste de usabilidade e pesquisas online com alunos graduandos e graduados na interface VuFind. No estudo com a mesma ferramenta, identificou que a busca pode ser oferecida em único local, como outros buscadores, podendo-se desmembrar palavras e ou assuntos, criando hiperlinks para documentos. O autor destaca que o resultado do teste ajudou na melhoria da interface, uma vez que teve a aprovação dos usuários. Como resultado final, é possível constatar que a interface do VuFind é intuitiva e amigável.

### **Organização do Trabalho**

Esta monografia está organizada da seguinte forma:

Nas primeiras páginas esta aborda a visão geral do trabalho, a motivação e justificativa, seus objetivos gerais, metodologia, trabalhos correlatos e a organização do trabalho.

O primeiro capítulo apresenta a fundamentação teórica sobre Recuperação da Informação com seus modelos clássicos.

O segundo capítulo traz a fundamentação teórica sobre Fontes Informacionais e suas principais fontes.

O terceiro capítulo exhibe a fundamentação teórica sobre Metadados e seus padrões.

O quarto capítulo cita a fundamentação teórica sobre Interoperabilidade, seus níveis e Coleta de Metadados.

O quinto capítulo expõe as etapas de teste. Por fim são apresentadas as conclusões e referências.



## 1 RECUPERAÇÃO DA INFORMAÇÃO

O termo Recuperação da Informação possui diversas definições, para uns, consiste no fornecimento dos elementos de informação documentária correspondentes, que por sua vez, ocorre a partir de uma demanda fornecida pelo usuário. Para outros, Recuperação da Informação significa a operação pela qual se seleciona documentos, a partir de acervos, em função de demanda de usuário. Há autores que associam recuperação de informação ao tratamento de informação, ou seja, catalogação, indexação, classificação.

Com o passar dos anos e o aumento dos números de documentos e usuários na Web, modelos para recuperação de informações passaram a ser de suma importância, devido à heterogeneidade das diversas fontes informacionais, buscando a padronização da recuperação.

Este trabalho optou por uma abordagem que proponha uma plataforma de Recuperação de Informação baseado em ambientes informacionais digitais abertos, como repositórios digitais, ambientes de aprendizagem a distância, de forma que seja possível realizar uma extração das informações mais completa e variada, pois as fontes buscadas serão diversas, tendo assim, maiores chances de satisfazer as necessidades informacionais do usuário.

O processo de recuperação de informação gera preocupações, nos pesquisadores da área, em como organizar e representar as inserções de novos documentos em acervos, bem como facilitar a busca aos documentos, em uma posterior recuperação (MIORELLE, 2001). Um dos motivos dessas preocupações está relacionado ao fato de que atualmente grande parte das informações armazenadas pelo homem se encontra em formato digital, porém, uma das ferramentas mais importante para auxiliar esse processo é denominada índice, que se trata de uma coleção de termos que indicam o local onde a informação desejada pode ser localizada (CARDOSO, 2000).

No que tange à finalidade do índice é facilitar todo o processo de busca de informações, bem como a automação do processo de indexação, que só se torna possível através de uma simplificação onde considera que os assuntos de um documento podem ser derivados de uma estrutura textual através de métodos algorítmicos. Por sua vez, a principal vantagem da automação é seu baixo custo. Segundo Ferneda (2003) nos métodos automáticos de indexação são utilizados “filtros” para eliminar palavras de pouca significação (*stop words*), além de normalizar os termos reduzindo-os a seus radicais, processo conhecido como *stemming*. Ainda que esta forma de indexação seja bem utilizada, suas falhas e limitações se evidenciam pela simplificação da dimensão semântico da linguagem.

O processo de recuperação de informação tem a função de busca como centro do processo, que compara as representações dos documentos com a expressão de busca dos usuários e recupera os itens que supostamente fornecem a informação que o usuário procura. Em relação à eficiência de um sistema de recuperação de informação está diretamente associado ao modelo que os utiliza, onde o modelo, uma vez que, influencia diretamente no modo de operação do sistema (FERNEDA,2003).

## 1.1 Definição

O termo Recuperação de Informação surgiu em 1951, com o cientista americano, Calvin Mooers, em sua pesquisa para desenvolver um mecanismo automatizado para acessar informações armazenadas em grande escala. Mooers definiu os seguintes problemas a serem abordados:

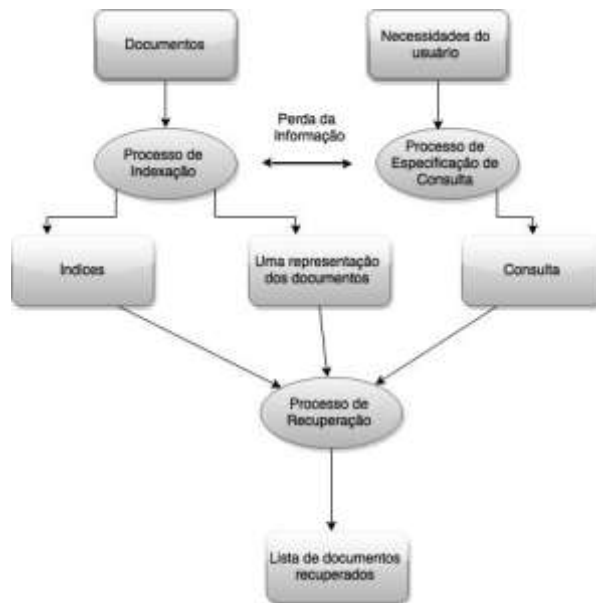
“A Recuperação de Informação trata dos aspectos intelectuais da descrição da informação e sua especificação para busca, e também de qualquer sistema, técnicas ou máquinas que são empregadas para realizar esta operação.” (MOOERS, 1951).

Por sua vez, essa teoria do autor está relacionada à comunicação humana, que é verbal, dificultando o mapeamento de forma euclidiana.

“O problema da recuperação é um problema digital, pois de fato toda a comunicação humana é digital. Recuperação de Informação é um problema não numérico em parte porque a maior parte da comunicação humana é verbal, mas mais importante porque a maioria das ideias ou dos conceitos não podem ser mapeados em um espaço euclidiano tridimensional, ou espaço superior. Enquanto há valores de escalas para a representação de alguma informação, esses são relativamente poucos e sem importância. Conceitos espaciais e métricos não são aplicáveis à maioria das informações, ao menos não de simples formas. Ainda, embora o problema da recuperação de informação seja não numérico, não parece haver qualquer alternativa para uso de técnicas digitais para sua solução.” (MOOERS, 1950,p.3).

Outrossim, existem várias definições para Recuperação de Informação, para Baeza-Yates e Ribeiro-Neto (1999, p. 1) a recuperação da informação está diretamente ligada à representação, armazenamento, organização e acesso aos itens de informação. Dizem também que a representação e a organização dos itens de informação deveriam prover o uso e o fácil acesso a informação necessária ao usuário.

Segundo Cardoso (2007), a Recuperação da Informação é uma área da Ciência da Computação que lida com armazenamento automático e recuperação de documentos, que são de grande importância devido ao uso universal da linguagem para comunicação. Um sistema de Recuperação de Informação (SRI) pode ser estruturado conforme a Figura 1.



**Figura 1: Componentes de um sistema de recuperação de informação**

**Fonte: Gey (1992).**

A Figura 1 representa uma estrutura de um sistema de recuperação da informação, tendo como componente desse sistema os documentos e as necessidades do usuário, gera-se uma consulta formulada e por fim o processo de recuperação da informação que, através das estruturas dos dados e da consulta em si recuperam uma lista de documentos que são considerados relevantes a consulta.

Cada processo possui sua função dentro do sistema. O processo de indexação refere-se à criação de estruturas dos dados associados à parte textual do documento como, por exemplo, as estruturas dos arranjos dos sufixos e arquivos invertidos. O processo de especificação da consulta é considerado uma das tarefas mais difíceis, pelo fato de que há uma distância grande da semântica original, daquilo que o usuário necessita e de como expressa sua necessidade na consulta. Já o processo de recuperação é responsável por apresentar a lista de documentos que responda a consulta realizada pelo usuário.

Existem vários modelos de recuperação de informação, entre eles serão abordados, Modelo Clássico, Modelo Quantitativo e Modelo Dinâmico.

## 1.2 Modelos Clássicos

Os modelos de recuperação de informação foram desenvolvidos há muitas décadas e embora sejam muito antigos, diversos conceitos relacionados à recuperação de informação

são utilizados até os dias atuais, com o intuito de buscar cada vez mais a sofisticação e ser cada vez mais eficientes e atentados com a qualidade das informações que serão recuperadas, para evitar perda de informação.

Nesta seção serão abordados os modelos clássicos de Recuperação da Informação, embora vários autores sugeriram diversos modelos para a realização da recuperação da informação, foram divididos em modelos quantitativos e modelos dinâmicos.

### **1.3 Modelo Quantitativo**

Grande parte dos sistemas de recuperação de informação é de natureza quantitativa, que são arquitetados em conceitos de lógica, estatística e teoria dos conjuntos. Em um estudo sobre os modelos de recuperação de informação, Robertson (1977) justifica que esse predomínio pelo fato de que a determinação de um modelo matemático geralmente pressupõe uma cuidadosa análise formal do problema e especificações de hipóteses, além de uma formulação explícita.

Quiçá a principal tarefa para os sistemas de recuperação da informação seja definir a importância de um termo para a descrição do conteúdo de um documento, pois, neste tipo de modelo, os documentos são representados por um conjunto termos de indexação. Um termo de indexação comumente representa um significado ou um conceito de um documento. Porém, uma das dificuldades é de qual será a representatividade dos termos para um determinado documento, ou seja, se aquele termo de fato conseguirá representar o documento dando-lhe a real representatividade, portanto, é de suma importância levar em considerações diversos aspectos para conseguir suprir as necessidades. (FERNEDA, 2003).

A seguir são descritos alguns dos modelos mais utilizados dentro dos métodos quantitativos.

#### **1.3.1 Modelo Booleano**

O modelo booleano baseia-se na teoria de conjuntos e álgebra booleana. Álgebra booleana é uma variante de álgebra ordinária, ou seja, um sistema binário, aonde os dados podem adotar somente dois valores possíveis: zero ou um, verdadeiro ou falso.

Esse modelo está presente em quase todos os sistemas de recuperação de informação, pois, oferece certo controle sobre o sistema. Contudo, uma das desvantagens desse modelo é a sua imperícia com os documentos de uma busca, se tornando inapropriado para sistemas

modernos de busca de texto integral, como mecanismos de buscas da Web, onde o ordenamento dos documentos é essencial mediante a grande demanda de dados que são recuperados (FERNEDA, 2003).

### **1.3.2 Modelo Vetorial**

O modelo vetorial foi instituído em 1968, por Salton, foi motivado principalmente pelas limitações apresentadas no modelo booleano (SALTON, 1988). Esse modelo sugere um ambiente no qual é possível obter documentos que respondem parcialmente a uma expressão de busca, ou seja, o processo é feito agregando pesos tanto aos termos de indexação dos documentos como aos termos utilizados na expressão de busca, obtendo um conjunto de documentos ordenados pela similaridade de cada documento em relação à expressão de busca.

O uso de similaridade promove o processo de definição de um ranking para os resultados da consulta (SANTAREM SEGUNDO, 2010, p. 33). Segundo Souza (2006, p. 167):

O modelo vetorial é a base da grande maioria de sistemas de recuperação de informações, mais notadamente os que têm como objeto a Internet, embora estes utilizem também outras técnicas para determinar o ranking de documentos como resposta a uma consulta.

A principal característica do modelo vetorial, está relacionada a simplicidade e a facilidade que permite calcular a similaridade entre as informações, ademais da execução de comparações.

O modelo vetorial tem como desvantagem o não uso de expressões booleanas, ao mesmo tempo, este modelo se caracteriza por aproximar muito as combinações, podendo encontrar relações que não possui de fato afinidade.

### **1.3.3 Modelo Probabilístico**

Na matemática, a teoria das probabilidades estuda os experimentos aleatórios que, repetidos em condições idênticas, podem apresentar resultados diferentes e imprevisíveis. Isso ocorre, por exemplo, quando se observa a face superior de um dado após o seu lançamento. Por apresentarem resultados imprevisíveis, é possível apenas estimar a possibilidade ou a chance de um determinado evento ocorrer (FERNEDA,2003)

O modelo probabilístico apresenta documentos levando em consideração pesos binários, que representam a presença ou falta de termos. Outrossim, é baseado no princípio

probabilístico de ordenação, ou seja, na hipótese de que a relevância de um documento para uma consulta é independente de outros, portanto é situado que este modelo pode ser usado de forma ótimo.

Logo pode-se definir que este modelo tem como vantagem, o princípio probabilístico de ordenação, que procede em uma conduta ótima do método. Todavia, sua desvantagem está relacionada à dependência da precisão das estimativas de probabilidade.

## **1.4 Modelo Dinâmico**

O modelo dinâmico de recuperação da informação tem o intuito de proporcionar uma participação mais ativa na representação dos documentos. De tal modo que surgiram devido um de esgotamento das funções e fórmulas matemáticas relacionadas aos modelos clássicos.

Nesta seção serão apresentados alguns modelos, tais como as Redes Neurais e os Algoritmos Genéticos, como forma de aclarar o conceito de modelos dinâmicos de recuperação da informação.

### **1.4.1 Algoritmos Genéticos e Redes Neurais**

As redes neurais artificiais buscam a simulação do funcionamento das células do cérebro humano, onde, o processamento das redes de neurônios é responsável pela capacidade de aprendizagem do ser humano. O intuito deste modelo é tentar a proximidade de emular sua forma de inteligência ao processar uma informação.

As redes neurais artificiais se destacam por sua arquitetura que restringe o tipo de problema no qual a rede poderá ser utilizada e pela forma como os pesos agregados às conexões que são ajustados durante o processo de aprendizado.

Os Algoritmos Genéticos, assim como as Redes Neurais, é uma técnica baseada em uma analogia da natureza, tendo como base a genética, onde garante que todo ser herda características de seus genitores, podendo herdar mais ou menos características.

Esse modelo é muito eficiente para busca de soluções ótimas, ou aproximadamente ótimas, em uma grande variedade de problemas, pois não impõem muitas das limitações encontradas nos métodos de busca tradicionais. Ferneda (2009) atesta que os algoritmos genéticos aplicados aos sistemas de Recuperação da Informação representam um novo modelo de ver este processo, pois a representação dos documentos será modificada conforme o que o sistema for aprendendo com o usuário.

A evolução tecnologia e conseqüentemente o aumento da comunicação acarretaram grande avanço da produção de dados na Internet. Nesse processo evolutivo inclui-se a produção científica que vem passando por momentos de transição, onde os documentos não se limitam apenas a publicação de livros e revista, estão passando a fazer parte de mídias digitais espalhadas pela rede. Assim surgem também as fontes informacionais acadêmicas, onde são armazenados documentos digitais em diversos formatos e junto com o surgimento vem a preocupação da recuperação destes documentos de maneira eficiente e eficaz, que atenda as necessidades informacionais do usuário.

## **2 FONTES INFORMACIONAIS ACADÊMICAS**

Neste capítulo serão descritos conceitualmente fontes informacionais acadêmicas.

### **2.1 Definição**

O estudo do comportamento informacional é uma das áreas de investigação da Ciência da Informação que visa, de forma geral, identificar os fatores que geram a necessidade de informação; as etapas do processo de busca; os elementos que influenciam este comportamento e para que fim o usuário utilize a informação obtida. (BARRET, 2005).

As fontes informacionais obtêm grande influência no nosso conhecimento e aprendizado. Atualmente podem ser classificadas em diversos tipos, como por exemplo, orais, impressos, digitais e multimídia, onde cada uma possui sua função, diferenciando-se por seu conteúdo e o público que é direcionado.

Esta seção irá apresentar algumas fontes informacionais acadêmicas, destacando seus objetivos e funcionalidades, como, Repositórios Digitais, Sistemas de Gestão de Aprendizagem, Revistas Eletrônicas, Biblioteca Digital e Não Digital.

### **2.2 Repositórios Digitais**

Como o crescimento da internet aumenta o desenvolvimento de recursos de informação digital diversificada, ocasionando uma enorme proliferação de computadores, redes, e recursos de informação em rede, tornando o acesso à informação relativamente fácil, barato, generalizada e democrática (GANDEL; KATZ; METROS, 2004). Essa disponibilidade generalizada de conteúdo digital é a criação de oportunidades para novas formas de pesquisa e ensino, qualitativamente diferente da maneira tradicional do uso de publicações acadêmicas e dados de pesquisa (ARMS; LARSEN, 2007).

Neste contexto, gerou-se a necessidade de divulgação independente de trabalhos desenvolvidos por instituições, que tomaram iniciativa e começaram a divulgar na internet, dando origem a diversas ferramentas denominadas repositórios digitais, permitindo assim que as instituições começassem a fazer o papel de editora.

Viana, Márdero Adellano e Shintaku (2005) evidenciam que os repositórios digitais:



[...] incentivam a publicação na Rede gerenciada pelo pesquisador (self archiving), utilizam novas tecnologias abertas (open source), e as informações ficam disponíveis para serem acessadas permanentemente por diversos provedores de serviços a nível nacional e internacional.

Na perspectiva de Lawrence (2003), entender os repositórios institucionais é perceber a “[...] manifestação visível da importância emergente da gestão do conhecimento na educação superior”. Nesse sentido, é válido ressaltar, conforme enfatiza Crow (2002), que os repositórios institucionais constituem “[...] um sistema global de repositórios distribuídos e interoperáveis que fundamentam um novo modelo de publicação científica.”.

Um repositório digital pode ser qualquer coleção de material digital hospedado, propriedade controlada ou divulgados por uma faculdade ou Universidade, independentemente da sua finalidade ou proveniência, têm as principais funções de armazenamento, disseminação e durabilidade de arquivos digitais, com o intuito de facilitar o acesso aos arquivos que são submetidos nessas ferramentas.

Os RDs armazenam arquivos de diversos formatos, além de resultar em uma série de benefícios tanto para os pesquisadores quanto às instituições ou sociedades científicas, proporcionam maior visibilidade aos resultados de pesquisas e possibilitam a preservação da memória científica de sua instituição.

Tratando-se de repositórios institucionais de universidades Lynch (2003, p.02) os define como sendo um conjunto de serviços que uma instituição de ensino oferece aos membros de sua comunidade, objetivando o gerenciamento e disseminação eficiente dos materiais digitais gerados pela instituição e pelos membros de sua comunidade. Os repositórios temáticos com a produção científica de uma determinada área, sem limites institucionais.

Atualmente, os repositórios digitais são usados como partes dos recursos de informação curricular, perpetuam o desenvolvimento dos alunos em relação ao conhecimento e habilidades para encontrar, organizar e usar conteúdo digital e muitas vezes para integrar esse conteúdo em sistema de gestão de aprendizagem. Um exemplo de Repositório Digital é o DSpace.

## 2.2.1 DSpace

O DSpace é um software livre, que foi desenvolvido pelo *Massachusetts Institute of Technology* (MIT) e Laboratórios *Hewlett-Packard*, com o fito de permitir a criação de repositórios digitais e é um dentre vários projetos, atualmente em operação, orientados à criação de repositórios institucionais e à preservação digital, sendo responsável pelo armazenamento, disseminação e durabilidade de arquivos digitais. A plataforma é voltada para a área acadêmica, com a finalidade de funcionar como base para o desenvolvimento futuro para atingir a preservação dos arquivos e suprir problemas de acesso (TANSLEY,2003).

Lewis (2008) afirma que sua estrutura provê um modelo de informação organizacional baseado em “comunidades” e coleções, o qual pode ser configurado de modo a refletir todo o conjunto de unidades administrativas de uma instituição. Permite a configuração do processo editorial nos moldes dos periódicos tradicionais, incluindo a possibilidade de revisão pelos pares.

DSpace suporta diferentes tipos de formatos de arquivos, como por exemplo, digitais, incluindo textos, som e imagem, portanto este software é apropriado para criação de repositórios institucionais, por permitir interoperabilidade com outros sistemas.

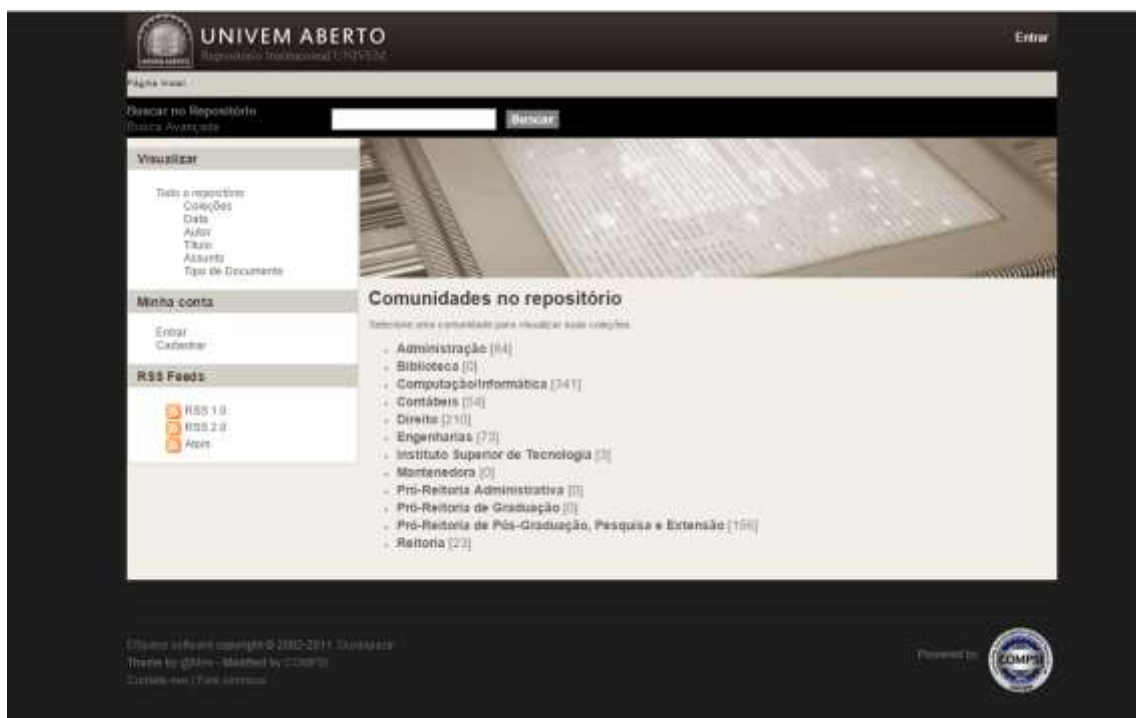
Atualmente existem 2751 repositórios registrados no site de registro de provedores de repositórios *OPEN ARCHIVES* (Fonte: REGISTERED DATA PROVIDERS, <http://www.openarchives.org/Register/BrowseSites>).

Neste site é exibida uma tabela com informações dos repositórios digitais registrados, o Univem Aberto é um dos repositórios registrados, conforme a Figura 2

Record	Identify	UNI ScholarWorks	<a href="http://scholarworks.uni.edu/do/oa/">http://scholarworks.uni.edu/do/oa/</a>	scholarworks.uni.edu
Record	Identify	UnICA Eprints	<a href="http://eprints.unica.it/cgi/oa2">http://eprints.unica.it/cgi/oa2</a>	eprints.unica.it
Record	Identify	UnipiEprints: University of Pisa institutional archive of e-prints	<a href="http://eprints.adm.unipi.it/peri/oa2">http://eprints.adm.unipi.it/peri/oa2</a>	eprints.adm.unipi.it
Record	Identify	UnS Scholarship Online	<a href="http://epubs.surrey.ac.uk/cgi/oa2.cgi">http://epubs.surrey.ac.uk/cgi/oa2.cgi</a>	epubs.surrey.ac.uk
Record	Identify	UnisaResearch	<a href="http://eprints.unisa.it/cgi/oa2">http://eprints.unisa.it/cgi/oa2</a>	eprints.unisa.it
Record	Identify	Unitec ResearchBank	<a href="http://unitec.researchbank.ac.nz/oa/request">http://unitec.researchbank.ac.nz/oa/request</a>	
Record	Identify	Unim-eprints PhD - University of Trento	<a href="http://eprints-phd.biblio.unim.it/cgi/oa2">http://eprints-phd.biblio.unim.it/cgi/oa2</a>	eprints-phd.biblio.unim.it
Record	Identify	Univ Provence repository	<a href="http://orion.univ-provence.fr/oti-oi-repository/OAIHandler">http://orion.univ-provence.fr/oti-oi-repository/OAIHandler</a>	univ-provence.fr
Record	Identify	Univem Aberto	<a href="http://aberto.univem.edu.br/oa/request">http://aberto.univem.edu.br/oa/request</a>	
Record	Identify	Universidad Autonoma del Estado de Hidalgo - Biblioteca Digital-	<a href="http://dgsa.uaeh.edu.mx:8080/oa/request">http://dgsa.uaeh.edu.mx:8080/oa/request</a>	
Record	Identify	Universidad de Chile - Tesis Electronicas	<a href="http://www.cybertesis.cl:8080/idx/idx/oa/uchile/documents">http://www.cybertesis.cl:8080/idx/idx/oa/uchile/documents</a>	
Record	Identify	Universidad de las Americas Puebla, Electronic Thesis and Dissertation Archive	<a href="http://catarina.udlap.mx/u_dl_a/tales/oa/requestETD.jsp">http://catarina.udlap.mx/u_dl_a/tales/oa/requestETD.jsp</a>	tesis.udlap.mx
Record	Identify	Universidad de las Americas, Puebla: Digital Theses	<a href="http://ict.udlap.mx:9090/Tales/Oai_tesis">http://ict.udlap.mx:9090/Tales/Oai_tesis</a>	tesis.UDLAP
Record	Identify	Universidad de Murcia	<a href="http://www.tesisdigital.es/ojs/index.php/journal/oa">http://www.tesisdigital.es/ojs/index.php/journal/oa</a>	ojs.www.tesisdigital.es
Record	Identify	UNIVERSIDAD NACIONAL DE COLOMBIA - Repositorio Institucional Dspace	<a href="http://www.digital.unal.edu.co/dspace-oa/request">http://www.digital.unal.edu.co/dspace-oa/request</a>	

**Figura 2: Registro do repositório digital Univem Aberto do Centro Universitário Eurípides de Marília**

O Univem Aberto foi implantado no Centro Universitário Eurípides de Marília – UNIVEM em 2012, com a finalidade de divulgar os trabalhos científicos, monografias, teses de mestrado e doutorado do corpo docente e trabalhos realizados pelos seus alunos e pesquisadores. O repositório contém atualmente 944 trabalhos que são distribuídos em suas comunidades, que são divididas pelos cursos oferecidos pela universidade, como, Administração, Computação/Informática, Contábeis, Direito, Engenharia, e alguns setores da mesma, como, Biblioteca, Instituto Supervisor de Tecnologia, Mantedora, Pró-Reitoria Administrativa, Pró-Reitoria de Graduação, Pró-Reitoria de Pós-Graduação, Pesquisa e Extensão e Reitoria, conforme apresenta a Figura 3.



**Figura 3: Repositório digital Univem Aberto**

### 2.3 Sistema de Gestão de Aprendizagem

Sistema de Gestão da Aprendizagem é uma plataforma que oferece diversos recursos, que proporcionam suporte ao processo de aprendizagem, ou seja, permitindo seu planejamento, implantação e avaliação do processo de aprendizagem, oferecendo um conjunto de funcionalidade que ajudam no armazenamento e distribuição de informações, gerenciando todos os conteúdos de aprendizagem, bem como registros e relatos das atividades, e até

mesmo o desempenho do estudante.

Com a evolução do mundo tecnológico, os ambientes de aprendizado também têm sofrido algumas adaptações que facilitam a vida do estudante, hoje além do ambiente tradicional, que são lugares constituídos de salas, mesas, cadeiras, materiais didáticos, alunos e professores o ambiente virtual de aprendizagem tem adquirido um grande destaque entre os usuários, pelo fato de ser um espaço relacional criado pelas redes digitais e no qual é preciso aprender a perceber a interação por meio de sinais na interface web das aplicações.

A evolução desses ambientes virtuais é crescente, e rapidamente têm se tornado cada vez mais popular, embora ainda existam muitos déficits nos ambientes atuais, como por exemplo, a falta de um acompanhamento mais adequado para facilitar as tarefas do professor. Um sistema de gestão de aprendizagem muito utilizado pelas universidades é o Moodle.

### **2.3.1 Moodle**

O Moodle é um ambiente de aprendizagem a distância baseada em software livre. Embora continue sendo desenvolvido de maneira contínua por uma “comunidade” de vários programadores do mundo, o moodle foi desenvolvido por Martin Dougiamas e lançado em 2001 com o intuito de propor em um sistema de administração de atividades educacionais voltado à criação de comunidades on-line, em ambientes virtuais para a aprendizagem colaborativa, além de permitir a interação entre um estudante e o professor de forma simples e eficaz, ou seja, é um aplicativo desenvolvido para ajudar os educadores a criar cursos on-line, ou suporte on-line a cursos presenciais, de alta qualidade e com muitos tipos de recursos disponíveis.

Um exemplo de implementação do Moodle é o sistema de Educação a Distância do UNIVEM funciona sobre a plataforma Moddle, onde são oferecidas várias atividades, como por exemplo, Disciplinas de Dependência, Disciplinas de Adaptação, Treinamentos, Cursos de Curta Duração e Acompanhamento de trabalhos acadêmicos como iniciação científica e trabalho de conclusão. Na página inicial da plataforma são fornecidos aos usuários opções de acesso à mesma até outras funcionalidades, como, Informações a respeito do NEAD (Núcleo de Educação à Distância), Guia para iniciantes na plataforma Moodle, Material de apoio a plataforma. Forma de contato com o Suporte do ambiente, Link alternativo de acesso, Página com as novidades institucionais e Link para acesso da área do aluno, conforme a Figura 4.



**Figura 4: Moodle Univem.**

## 2.4 Revistas Eletrônicas

As Revistas Eletrônicas na área acadêmica estão associadas ao campo de publicações, dedicada à promoção do avanço da ciência, ou seja, investindo e divulgando novas pesquisas. Portanto, essas revistas são compostas de artigos que são submetidos a revisão, com o intuito de certificar que atendam todos os padrões válidos, desta forma cada artigo da revista se torna parte de um registro científico permanente.

### 2.4.1 Open Journal Systems

O *Open Journal System* (OJS) também é um software *opensource* de nível mundial e foi desenvolvido pelo *Public Knowledge Project* por intermédio do Governo Federal Canadense, com a finalidade de criar um acesso aberto, fornecendo uma opção viável para diversas revistas, aumentando os leitores, bem como gera uma grande contribuição aos leitores a nível global. OJS é um sistema de gerenciamento e publicação de revistas, tem o intuito de auxiliar todas as fases do processo de publicação. Por fim, o *Open Journal System* visa à melhoria da qualidade acadêmica e pública em geral.

O Centro Universitário Eurípides de Marília - UNIVEM, com o objetivo de estimular a publicação científica de seus professores e alunos e também de outras instituições, publica

anualmente três revistas eletrônicas. A Revista Eletrônica do UNIVEM é um periódico multidisciplinar destinado à publicação de artigos científicos baseados em trabalhos de conclusão de curso e de iniciação científica dos cursos de graduação e tecnólogos da Instituição.



Figura 5: Portal de Revistas Eletrônicas do Univem.

## 2.5 Biblioteca Digital

A biblioteca digital é constituída por documentos primários em formato digital, com acesso via rede de computadores. De acordo com Borgman (1996, p. 6),

Bibliotecas digitais são um conjunto de fontes eletrônicas e serviços técnicos associados para a criação, pesquisa e uso da informação, que possibilitam uma extensão e um aumento do armazenamento da informação e dos sistemas de recuperação de informação, manipulando dados digitais em qualquer meio (texto, imagens, sons, imagens dinâmicas e estáticas) em redes distribuídas de trabalho. O conteúdo das bibliotecas digitais inclui dados e metadados que descrevem vários aspectos do dado (representação, criador, dono, direitos de reprodução) e metadados que consistem em ligações ou relacionamentos com outros dados ou metadados,

sejam esses externos ou internos à biblioteca digital.

No que tange aos benefícios das bibliotecas digitais, pode-se articular que eliminam as barreiras físicas e a distância em relação às bibliotecas físicas, entretanto, em relação à estrutura tecnológica que deve possuir, o planejamento e o desenvolvimento devem estar pautados de acordo com o modelo estipulado em uma biblioteca tradicional, ou seja, todo o tratamento informacional (representação, organização e preservação da informação), formas de aquisição, recuperação, disseminação, serviço de referência, devem estar contemplados de acordo com uma política institucional que dirigirá a funcionalidade da biblioteca digital. Neste aspecto, Wainwright ressaltou que (1996, p. 2):

[...] as bibliotecas digitais, ainda permanecem sendo bibliotecas, [sustentadas] com os mesmos propósitos, funções e metas das bibliotecas tradicionais [convencionais]. A parte digital do termo indica meramente que os materiais estão armazenados e acessados digitalmente.

Portanto, a internet fornece às bibliotecas formas mais concretas de atuação no ciberespaço. Atualmente, as bibliotecas digitais se tornaram extremamente importantes no tange a ambientes informacionais, devido a facilidade referente à disseminação da informação que proporciona, além de gerar confiabilidade e integridade dos recursos informacionais. E através do acervo digital as buscas se tornaram mais rápidas, facilitando a realização de pesquisas do usuário.

### **2.5.1 Acervo Digital**

Acervo digital é um conjunto de obras disponíveis na Internet para consulta, onde sua coleção é compreendida como um processo de aquisição, sendo composta por recursos de multimídias, imagem, som, texto.

Com o surgimento dos computadores e da internet, começaram as modificações na organização dos acervos e serviços das bibliotecas, ou seja, usaram dessa nova tecnologia para automatizar os processos, como, catalogação, classificação, indexação, proporcionando o acesso on-line aos bancos de dados e a facilitando o processo de recuperação da informação.

De acordo com a figura 6, pode-se melhor compreender os elementos que devem estar contemplados numa estrutura tecnológica de um acervo digital.



**Figura 6: Estrutura Tecnológica de um Acervo Digital.**

## **2.6 Biblioteca Não Digital**

A Biblioteca Não Digital nada mais é do que a biblioteca tradicional, onde grande parte do seu acervo é composto por documento de papel, como, livros e revistas por exemplo.

Todavia, com a evolução da tecnologia a partir do século XIX a forma de catalogação das fichas vem sofrendo alteração, passando a ser informatizado, de forma digital, consequentemente abandonando os catálogos em forma de livro, em busca de algo mais prático e eficaz, gerando otimização nos processos de busca e evitando perda de informações que podem acarretar outros problemas.

### **2.6.1 Acervo não Digital**

Acervo denota de uma grande quantidade de algo, em geral, acervo é utilizado para fazer referência a uma coleção de obras ou bens que fazem parte de um patrimônio, seja de propriedade privada ou pública, podendo ser de âmbito artístico, bibliográfico, científico, documental, genético, iconográfico, histórico.

O Acervo não Digital é constituído por documentos em papel. Porém com os avanços tecnológicos a catalogação, classificação, indexação do acervo passou a ser automatizada, de maneira digital, para evitar perdas de informações e facilitar o leitor.

Uma fonte informacional acadêmica reúne uma coleção de documentos digitais, que por sua vez adotam padrões de metadados, protocolos de comunicação e processos de extração como o projeto OAI (*Open Archive Initiative*).



### **3 METADADOS**

Neste capítulo serão descritos conceitualmente metadados e padrões de metadados. Também serão mostrados conceitos específicos do Padrão Dublin Core.

#### **3.1 Definição**

Metadados são definidos como “dados sobre dados” (WEIBEL, 1995), facilitam o entendimento dos relacionamentos e a utilidade das informações dos dados. Ademais, esse termo possui diferentes conceitos e significados, porém seu foco é a descrição da informação visando sua recuperação.

Segundo Niso (2004, p.01), “metadados são informações estruturadas que descrevem, explicam localizam ou senão facilitam a recuperação, uso ou o gerenciamento de um recurso informacional”.

Metadado é determinado como dado que descrevem atributos de um recurso, caracteriza suas relações, possibilitando sua recuperação e uso efetivo. Metadados normalmente consistem em um jogo de elementos de dados onde cada elemento descreve um atributo do recurso, sua administração ou uso. (FUSCO, 2010, p.62. VELLUCCI, 1998, p.192). Pode-se dizer que os metadados basicamente auxiliam na descoberta da existência dos dados por seus usuários, transformando dados em conhecimento.

O conceito de metadados não é novo, entretanto a utilização desse termo em ambientes digitais sim. Os registros bibliográficos que tem sido criado por profissionais da informação a muito tempo devem ser considerados essencialmente como metadados, pois proporcionam informação descritiva e analítica sobre um objeto de informação. (FUSCO, 2010).

#### **3.2 Padrão de Metadados**

Os padrões de dados permitem a melhoria na manipulação, armazenamento, recuperação e troca de informações de forma ágil e eficaz. De acordo com Nascimento (2008) a padronização é necessária para que haja entendimento das estruturas, da sintaxe e do significado dos metadados, e com isso melhorar a comunicação, reutilização e compartilhamento das informações, resultando em redução de trabalho e custos.

Por meio dos padrões de metadados é possível garantir que informações sejam

comuns, tornando fácil a compreensão, integração e compartilhamento de informações entre usuários. Alguns aspectos que tornam a padronização dos metadados importantes são (NASCIMENTO, 2008):

- a) Auxílio na automatização da recuperação das informações;
- b) Estabelecimento da comunicação consistente entre bancos de dados;
- c) Garantia do registro de informações que sejam importantes;
- d) Facilidade no uso e no compartilhamento das informações e conhecimentos; e
- e) Auxílio na troca entre informações de diferentes bancos de dados.

Dentre os vários padrões existentes pode-se destacar o Padrão Dublin Core, como exemplo, pois, este padrão define um grupo de atributos que pode ser utilizado por autores para descrever seus próprios recursos na web, onde de acordo com DCMI o padrão Dublin Core se desta pela Simplicidade, Interoperabilidade Semântica, Consenso Internacional, Extensibilidade e Modularidade de Metadados na Web.

### 3.2.1 Dublin Core

O padrão de metadados *Dublin Core Metadata Initiative* (DCMI) é uma organização aberta que apoia a inovação em design e a busca de melhores práticas de metadados, teve origem em 1994, em Chicago, Estados Unidos, durante a segunda Conferência Internacional sobre a Web. Suas atividades estão associadas também a trabalhos sobre arquitetura e modelagem. Ainda o DCMI cultiva um número de ligações e relações formais e informais com organismos de normalização e outras organizações de metadados.

Dublin Core é um padrão de metadados que tem como fins apresentar objetos digitais, bem como, vídeos, sons, imagens, textos e sites na web, suas aplicações utiliza RDF (*Resource Description Framework*). DCMI é composto de elementos para descrever uma vasta quantidade de recursos eletrônicos e, por sua vez, compreende quinze elementos semânticos que foram estabelecidos através do consenso de grupos interdisciplinares internacionais de bibliotecários, cientistas da computação, comunidade de museus, e outros estudiosos deste campo.

Para Souza, (2000), o padrão de metadados *Dublin Core* pode ser definido como um conjunto de elementos descritivos com a intenção de facilitar a descrição de recursos digitais disseminados na Internet. Marcondes (2005), aponta que o padrão de metadados para descrição de documentos eletrônicos proposto pela DCMI foi pensado para ser simples o

suficiente e autoexplicativo de modo que o próprio autor do documento possa descrevê-lo ao publicá-lo eletronicamente.

O padrão *Dublin Core* inclui dois níveis: Simples e Qualificado. O *Dublin Core* Simples contém quinze elementos (Título, Autor, Assunto, Descrição, Editor, Contribuidor/colaborador, Data, Tipo, Formato, Identificador, Fonte, Idioma, Relação, Abrangência e Gerenciamento de Direitos Autorais), já o Qualificado contém três elementos adicionais (Audiência, Proveniência e Detentor de Direitos), deste modo como um grupo de refinamentos de elementos, denominados qualificadores, que refinam a semântica dos elementos de maneira que sejam úteis nas descobertas de recursos.

A *Dublin Core Metadata Initiative* (DCMI) (Iniciativa Dublin Core Metadados) é tem o objetivo de promover a adoção de padrões de interoperabilidade de metadados e desenvolver vocabulários especializados para descrever fontes e recursos da Web para tornar o mecanismo de busca e recuperação da informação mais rápido e flexível. Metadados são definidos como mecanismos que ajudariam a resolver o problema da interoperabilidade, pois permitem estabelecer um acesso aos dados disponíveis bem como a definição do significado da informação (GRÁCIO, 2002).

Estudos sobre padrões de metadados têm surgido como uma solução para a integração de dados disponíveis em fontes de informações distribuídas fisicamente e heterogêneamente, estabelecendo a interoperabilidade entre essas fontes e especialmente entre os padrões que as descrevem. Para Cromwell-Kessler (1998), a correspondência entre os padrões é um dos segredos para que o usuário possa buscar as informações desejadas na Web, pois é estabelecida uma associação entre esses padrões, permitindo que através de uma única interface de busca, o usuário possa pesquisar e acessar as informações desejadas.

## 4 INTEROPERABILIDADE E INTEGRAÇÃO

A heterogeneidade semântica é um obstáculo para prover interoperabilidade entre diversas fontes de dados, pois podem ter diversas interpretações da mesma informação, causando conflitos de sinonímia (dados com o mesmo conteúdo semântico, mas com nomes diferentes) e de homonímia (dados com o mesmo nome, mas com conceitos diferentes) (MARINO, 2001)

A interoperabilidade pode ser definida como uma característica que se refere à capacidade de vários sistemas e organizações trabalharem em conjunto (interoperar) de tal forma que garanta que pessoas, organizações e sistemas computacionais troquem informações de maneira eficaz e eficiente. Um dos aspectos importantes para a interoperabilidade de informações é a organização do conhecimento e representação da informação (OLIVEIRA, 2005).

A interoperabilidade é buscada por diversos fatores, são sistemas antes destinados basicamente para recuperação de referências bibliográficas em bases de dados isoladas, exigem uma recuperação distribuída de objetos digitais publicados na Internet em fontes heterogêneas e globalmente distribuídas. Ademais, existe a necessidade de divulgação de trabalhos científicos, buscando maior impacto possível dos resultados da pesquisa. E outros fatores que levam a busca da interoperabilidade são o aumento das citações de publicações eletrônicas e o desenvolvimento de mecanismos de publicação eletrônica para a comunidade acadêmica aumentando sua visibilidade, questões essenciais para o desenvolvimento e maturidade da pesquisa científica.

A complexidade do compartilhamento de informações foi ampliada nos últimos anos com a utilização de ambientes altamente distribuídos e heterogêneos como a Web, e a busca por normas, padrões, regras e protocolos que permitam uma organização e representação da informação de modo que se possa alcançar a interoperabilidade.

Para que bases de dados ou instituições possam trocar informações de forma eficaz é necessário o entendimento dos códigos utilizados por ambos e que eles tenham o mesmo entendimento quanto ao significado destes códigos. Para isto é necessária a implementação de padrões e normas que possibilitem o entendimento entre eles, desenvolvendo e reforçando trabalhos cooperativos.

## 4.1 Definição

De acordo com a *International Standard Organization* (ISO) interoperabilidade é a habilidade de dois ou mais sistemas, que podem ser computadores, meios de comunicação, redes de software e outros componentes de tecnologia de informação, de interagir e de intercambiar dados a partir de um método definido, objetivando obter os resultados esperados.

Interoperabilidade é a capacidade de operar em conjunto, ou seja, a capacidade que sistemas ou produtos têm de trabalhar com outros sistemas ou produtos sem necessitar de esforço especial por parte do cliente.

Segundo Arms (2002), o objetivo da interoperabilidade é desenvolver serviços coerentes para os usuários, a partir de recursos informacionais que são tecnicamente diferentes e gerenciados por diferentes organizações. Isto requer acordos de cooperação em três níveis:

- O nível técnico proporciona a interoperabilidade tecnológica;
- O nível de conteúdo remete a interoperabilidade semântica, na qual a representação e organização do conhecimento são áreas chave a serem estudadas;
- O nível organizacional se refere à interoperabilidade política, quando organizações se reúnem com o intuito de alcançar a interoperabilidade, implementando padrões e tecnologias que cooperem com este objetivo.

### 4.1.1 Níveis de Interoperabilidade

Para entender afundo a interoperabilidade de informações é necessário a compreensão de três níveis de características no intercâmbio das informações: semântico, estrutural e sintático.

- a) Nível Semântico: Possibilita compreender o significado de cada elemento descritor do recurso, juntamente com as associações nele embutidas. O uso de vocabulários específicos, ontologias e/ou padrões de metadados são essenciais para assegurar esse tipo de interoperabilidade (MOURA 2002).

Existem dois subníveis no nível semântico (MARINO, 2001):

- Epistemológico: refere-se ao significado dos elementos descritores do formato e das relações nele existentes;
- Ontológico: refere-se ao uso de ontologias, vocabulários controlados e

padrões de metadados para o estabelecimento dos significados dos dados representados.

Como exemplo desse nível de interoperabilidade têm-se os campos que têm o mesmo significado entre elementos descritivos de padrões de representação como o MARC21.

- b) **Nível Estrutural:** Define cada elemento componente de um padrão de metadados, descreve os seus tipos, a escala de valores possíveis para esses elementos e os mecanismos utilizados para se agrupar (ou relacionar) esses elementos de modo a que possam ser processados de forma automática. Quanto mais complexa for à estrutura do padrão de metadados, mais complexo deve ser o modelo de dados empregado para descrevê-la (BARRETO, 1999).

Exemplos: Padrão de metadados *Dublin Core* (DC) e MARCXML.

- c) **Sintática:** de acordo com Barreto (1999), a sintaxe fornece uma linguagem comum para representar a estrutura dos metadados. Para Moura (2002), a interoperabilidade sintática estabelece como os metadados devem ser codificados para a transferência de informações.

Exemplo: A linguagem XML (*eXtensible Markup Language*) na gerência da troca de informações.

## 4.2 Coleta de Metadados

Como dito na seção 4.1, interoperabilidade é a capacidade de compartilhamento de informações em diferentes ambientes informacionais e, através de tecnologias como linguagens de marcação associadas ao uso de padrões de metadados, faz com que informações persistidas e em plataformas heterogêneas possam ser trocadas possibilitando a colaboração entre diferentes sistemas de informação.

Por sua vez, os metadados permitem a busca de elementos de informação por critérios relevantes, identificação de recursos, agrupamento de recursos similares, diferenciação de recursos não similares e a obtenção de informação de localização. A descrição de recursos por meio de metadados permite que eles sejam semanticamente compreendidos por sistemas de informação possibilitando a interoperabilidade entre aplicações e o compartilhamento de dados entre sistemas.

Os metadados são um mecanismo que ajuda a resolver este problema da interoperabilidade, os dados podem estar descritos por um único padrão de metadados, para

tanto, há a necessidade de correspondência entre padrões de metadados, mesmo dentro de uma mesma área.

A figura 7 mostra parte da tabela de “Correspondências entre padrões de metadados” (CROMWELL-KESSLER, 1998), exemplificando a construção de relações entre diferentes padrões de metadados, requisito para interoperar sistemas baseados nesses tipos de bases heterogêneas.

CDWA	Object ID	FDA	VRA Core	REACH	USMARC	DC
Titles or names	Title	Group/Item Identification-Repository Title Group/Item Identification-Descriptive Title Group/Item Identification-Inscribed Title	W2. Title	Field #4: Object Name/Title	24Xa Title and Title-Related Information	Title
State					562c Copy and Version Identification Note-Version Identification	Description
Classification						Subject
Copyright/Restrictions	Copyright restriction	Internal Documentation Restriction			540a Terms Governing Use and Reproduction	Rights

**Figura 7: Correspondência entre padrões de metadados (CROMWELL; KESSLER, 1998).**

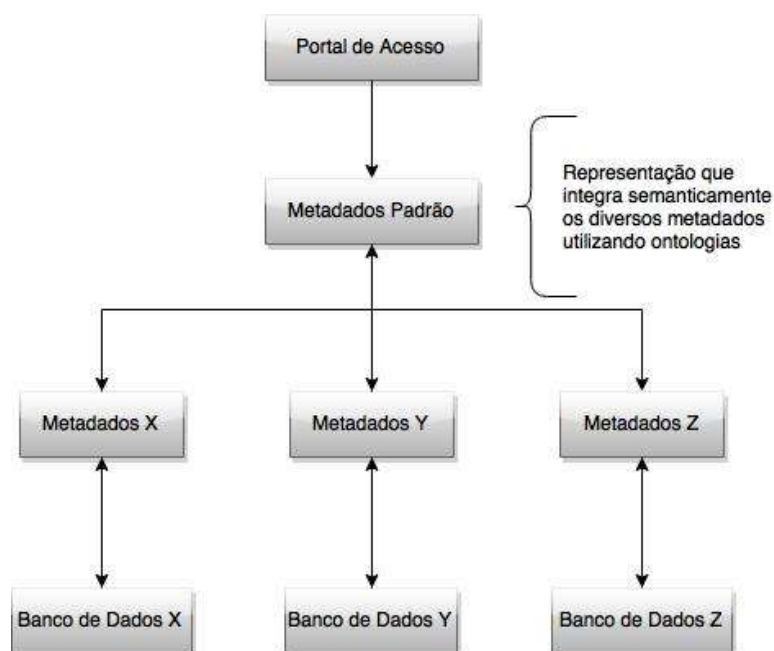
Pode-se verificar nessa correspondência que nem todos os elementos têm correspondência nos diversos padrões.

As arquiteturas de metadados foram desenvolvidas para garantir a interoperabilidade entre diversos padrões de metadados, com o propósito de representar e dar suporte ao transporte de uma variedade de esquemas de metadados em ambiente distribuído, promovendo interoperabilidade nos níveis sintático, estrutural e semântico (MOURA, 2002).

Pode-se destacar entre os tipos de arquitetura de metadados na Web a arquitetura Warwick, MFC – Meta Content Framework, a RDF – Resource Description Framework e a arquitetura de modelagem de quatro níveis. De acordo com Alves (2005), a RDF vem sendo indicada e recomendada pelo W3C (*World Wide Web Consortium*) para a interoperabilidade na rede, uma vez que, unida aos metadados, ao uso da linguagem XML e às ontologias,

possibilita de modo flexível, a interoperabilidade nos três níveis.

A Figura 8 esquematiza a definição de um esquema para interoperar diversos sistemas utilizando padrões de metadados distintos.



**Figura 8: Padrões de Metadados.**

Nota-se, na figura, que deve ser definida uma arquitetura de metadados padrão, em que deve haver uma correspondência entre os diversos metadados que participam da integração das bases com o metadados padrão, a esses metadados são feitos os acessos às diversas bases de dados do ambiente integrado. Qualquer instituição que queira se interoperar a esse ambiente deve fazer a correspondência do metadados utilizado em seu sistema legado com o metadados padrão.

#### **4.2.1 Open Archives Initiative Protocol for Metadata Harvesting**

O *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) é um mecanismo de baixa barreira para a interoperabilidade de repositório, ou seja, é um conjunto de seis verbos ou serviços que são chamados dentro do HTTP (*Hypertext Transfer Protocol*).

Em conformidade com Garcia e Sunye (2007), sites como o Google, Alta Vista, Cadê etc. fazem a busca convencional comumente retornando uma pesquisa vaga, atingindo apenas as páginas *HyperText Markup Language* (HTML) ignorando as bases de dados que se encontram por trás de algumas destas páginas.



O protocolo OAI-PMH vem consolidando-se com a base para a interoperabilidade entre bibliotecas/repositórios digitais acadêmicos e científicos no mundo todo, onde é possível proporcionar visibilidade e integração de informações (metadados) com custos acessíveis. Por sua vez, os metadados do protocolo OAI-PMH seguem o padrão Dublin Core onde os participantes da Iniciativa são divididos em:

- Provedores de Dados (*Data Providers* ou DP): Servidor de Dados que mantém um ou mais repositórios, ou seja, o DP contém o documento ou arquivo que será disponibilizado na rede;
- Provedores de Serviços (*Service Providers* ou SP): Provedor de Serviços realiza a coleta da descrição dos documentos para dentro de sua base.

O protocolo OAI-PMH introduz o conceito de *Metadata Harvesting* (colheita de metadados) como um processo unilateral, no qual os provedores de serviços, a partir da lista de repositórios (provedores de dados) registrados na OAI, realizam periodicamente uma busca a estes provedores de dados “colhendo” os metadados para exibição sob a forma de consultas efetuadas pelos usuários, conforme a Figura 9.



Figura 9: Metadata Harvesting (GARCIA; SUNYE (2007,p.4))

Percebe-se na Figura 9 o Provedor de Serviços devidamente registrado no OAI fazendo a solicitação (*request*) ao Provedor de Dados que retorna os metadados aos usuários.

## **5 PROPOSTA DE PLATAFORMA DE RECUPERAÇÃO DA INFORMAÇÃO EM FONTES INFORMACIONAIS ACADÊMICAS**

Com o estudo sobre Recuperação da Informação foi possível verificar a existência de diversos meios, onde para cada um deles há técnicas específicas a serem utilizadas.

Este projeto tem como proposta uma arquitetura de Recuperação da Informação que concretize em um ambiente único, a recuperação das informações que estão dentro destas diversas fontes de dados.

Para o desenvolvimento da Plataforma será utilizada a ferramenta VuFind, que possibilita a integração de variadas ferramentas, criando um portal de busca sólido. Este contexto visa propor um padrão de metadados para integração de das fontes informacionais, como por exemplo, moodle, Open Journal Systems (OJS), Repositórios - DSpace e Biblioteca física.

Para realizar a coleta de registros de metadados em repositórios será utilizado o protocolo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). OAI-PMH é um protocolo que define um mecanismo de coleta de registros de metadados. O VuFind opera por meio de coleta (harvesting) automática de metadados, utilizando o protocolo Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH).

Portanto, o intuito da presente pesquisa é propor uma plataforma de Recuperação de Informação baseado em ambientes informacionais digitais abertos, como repositórios digitais, sistema de gestão de aprendizagem, revistas e periódicos e a coleção do catálogo e do acervo de Bibliotecas Digitais e não digitais, visando melhorar os processos de recuperação trazendo melhor experiência ao usuário.

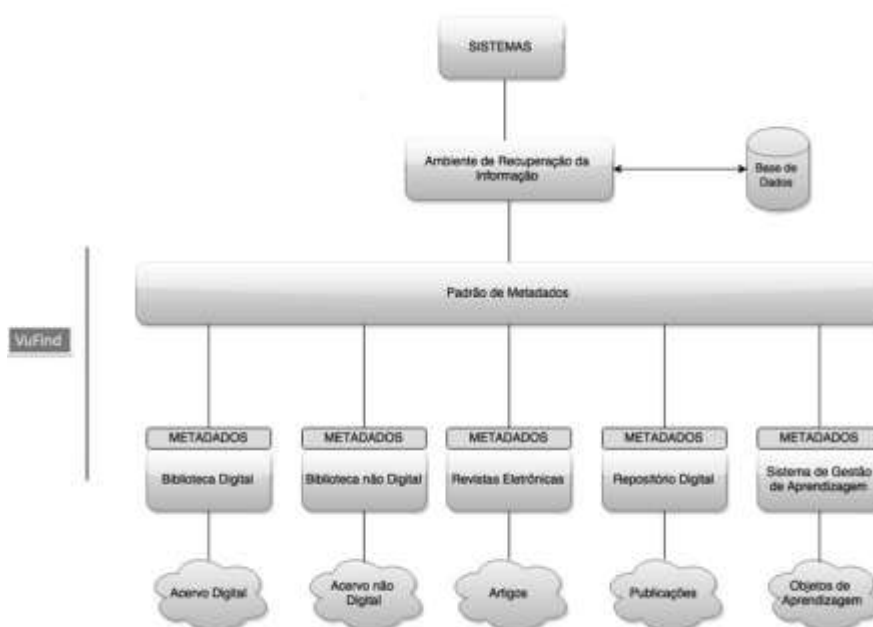
O levantamento bibliográfico possibilitou conhecer os conceitos que envolvem os repositórios digitais, sistema de gestão de aprendizagem, revistas e periódicos e a coleção do catálogo e do acervo de Bibliotecas Digitais e não digitais, assim como os modelos clássicos da recuperação da informação e seus conceitos, e a ferramenta VuFind que facilitará o processo de desenvolvimento da Plataforma de Recuperação da Informação.

Sugere-se que a implementação de uma plataforma centralizada de recuperação da informação de ambientes acadêmicos aberto possibilita uma visão não linear dos elementos descritivos dos produtos de informação e permite uma eficiente transmissão das mensagens contidas nas representações e a otimização do acesso e uso de ambientes catalográficos, por meio da intersecção entre os itens bibliográficos e as necessidades informacionais dos usuários (FUSCO, 2010).

Por meio desta pesquisa pôde-se verificar que a utilização da ferramenta VuFind ainda é pouco explorada, porém de grande valia quando utilizada adequadamente, pois facilita o acesso à informação, bem como a criação de portais integrados, de forma a criar um ambiente centralizado de recuperação da informação.

Ao concluir este projeto espera-se obter uma solução para otimizar os processos da recuperação da informação, ou seja, uma plataforma de recuperação da informação que tenha a capacidade de atender todas as demandas dos usuários, através da criação um padrão de metadados para moodle, Open Journal Systems (OJS), Repositórios e Biblioteca física.

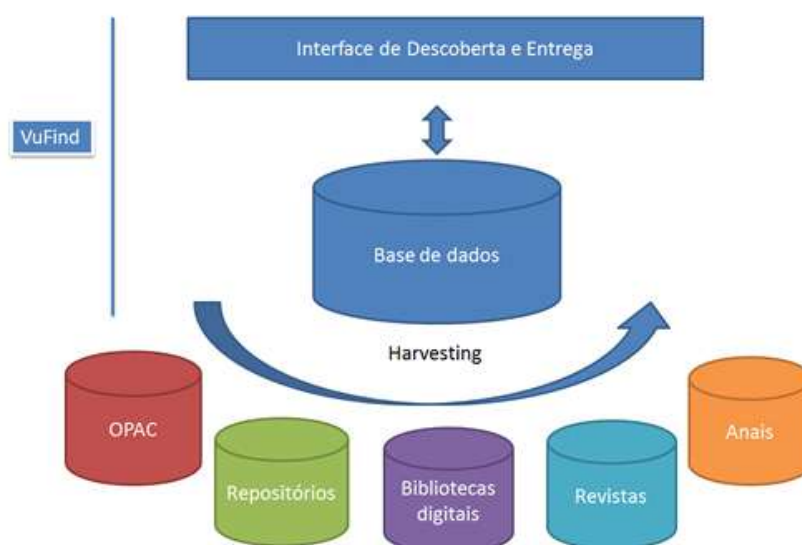
Conforme a Figura 10, a plataforma integrará a biblioteca digital, biblioteca não digital, revistas eletrônicas, repositório digital e sistema de gestão de aprendizagem. Essa integração será feita através da ferramenta VuFind, para isso, será criado um padrão de metadados para que o usuário consiga realizar a pesquisa apenas uma vez, pois o resultado recuperado será de todas as fontes integradas, ou seja, com a proposta da plataforma toda a necessidade de um usuário de fontes informacionais acadêmicas serão supridas, bem como facilitará todo processo de busca e recuperação da informação, pois, como trará o resultado da busca de uma forma unificada diminuirá possíveis perdas de informação.



**Figura 10: Proposta de Plataforma de Recuperação da Informação.**

## 5.1 FERRAMENTA VUFIND

A ferramenta VuFind é uma descoberta em software livre, sustentada pela biblioteca Falvey Memorial Libray pertencente à Universidade Villa Nova, com o intuito de integrar ferramentas isoladas, de forma a facilitar o acesso à informação, através do apoio a criação de portais integrados de informações geridas pelas bibliotecas. O VuFind possibilita a integração de variadas ferramentas, de forma a criar um portal de busca consolidado, conforme a figura 11.



**Figura 11: Recuperação da Informação utilizando a ferramenta VuFind. (Fonte: VuFind, wiki.ibict.br)**

VuFind opera de forma modular, podendo ser implementado de forma completa ou apenas alguns módulos. Essa ferramenta é altamente customizável e possibilita ajustes posto que seu código fonte é aberto, permitindo assim ajustes conforme as necessidades da instituição. Parte do VuFind é desenvolvido em PHP e outra em Java. O banco de dados utilizado é o MySQL e utiliza o apache como servidor de aplicação.

O VuFind atua nas seguintes etapas:

- Coleta de Metadados: utiliza diferentes formas de coleta, tanto nos padrões de metadados, como por exemplo, Dublin Core, quanto nos protocolos, exemplo, OAI\_PMH; além de integrar-se com muitos catálogos e outros serviços;
- Indexação: utiliza ferramentas como Apache SOLR para indexar os metadados coletados. Possui aspectos mais técnicos, que fazem com que

todas as palavras tornam-se ponto de recuperação. Podem-se criar índices exclusivos para alguns campos, como título e autor.

- Recuperação: a recuperação pode ser feita através de ferramenta de busca simples ou avançada, com a possibilidade de realizar diversas configurações na busca e facetar.
- Apresentação: a apresentação é semelhante a catálogos de bibliotecas, entretanto permite integração com ferramentas externas, como por exemplo, o google books;
- Interação: permite interação social com inserção de comentários, favoritos e Tags; personalização dos formatos de citações para os documentos.

### 5.1.1 Funções VuFind

VuFind é um portal de recursos para bibliotecas, que tem o objetivo de facilitar a capacitação de seus usuários para realização de pesquisas e navegação através de todos os seus recursos, substituindo o tradicional OPAC (*online public access catalog*) para incluir: Catálogo dos Registros, Itens Biblioteca Digital Repositório Institucional, Institucional Bibliografia e Outras coleções e recursos de biblioteca.

A ferramenta é *open source*, permitindo a personalização da mesma de acordo com a necessidade de cada um, ou seja, pode-se implementar o sistema básico ou todos os componentes configuráveis, tendo assim uma personalização sem alteração no código.

VuFind é executado no *Solr Energy*, utilizando o *Apache Solr*, um motor de busca *open source*, que oferece um desempenho e escalabilidade incríveis para permitir que o VuFind responda as consultas de pesquisa em milissegundos.

### 5.1.2 Componentes

Para realizar a configuração do VuFind é necessário realizar a instalação de alguns componentes fundamentais para seu funcionamento, como, Apache, Solr, PHP, Sistema de Gestão Integrado de Bibliotecas (ILMS, ILS) e MySQL.

#### 5.1.2.1 Apache

O Apache é um servidor Web livre e foi criado em 1995. O é totalmente compatível com o protocolo HTTP. Suas funcionalidades são mantidas através de uma estrutura de

módulos, permitindo que o usuário escreva seus próprios módulos utilizando a API do software e é disponibilizado em versões para os sistemas Windows, Novell Netware, OS/2 e diversos outros do padrão POSIX (Unix, Linux, FreeBSD, etc.).

VuFind precisa do Apache para que suas páginas sejam visíveis para os usuários que acessá-los.

#### **5.1.2.2 Solr**

Solr é um motor de busca altamente confiável, escalável e tolerante a falhas, fornecendo indexação, replicação e consultas com balanceamento de carga, possui poderosos recursos de busca e navegação que são características de muitos dos maiores sites da internet do mundo. É um servidor de procura corporativa independente com uma API REST-like, sua indexação ocorre através de JSON, XML, CSV ou binário através de HTTP.

A plataforma é escrita em Java, a partir do Apache Lucene. Suas principais características incluem pesquisa de texto completo, pesquisa facetada, indexação em tempo real, agrupamento dinâmico, integração de banco de dados, ainda fornece pesquisa distribuída e replicação do índice, sendo projetado para escalabilidade e tolerância a falhas. Solr se tornou o motor de busca mais popular.

VuFind usa Solr para indexar seus registros e pesquisar através deles para os usuários. Interage com Solr exatamente da mesma maneira que seu navegador fala com um servidor web. Para tornar isso possível, Solr é executado dentro de seu próprio software de servidor web chamado Jetty. Felizmente, Solr e Jetty são incluídos com a própria VuFind, assim não precisa instalá-los separadamente.

#### **5.1.2.3 PHP**

O PHP é uma das linguagens mais utilizadas na web. Milhões de sites no mundo inteiro utilizam PHP. A principal diferença em relação às outras linguagens é a capacidade que o PHP tem de interagir com o mundo web, transformando totalmente os websites que possuem páginas estáticas. [...] Diversos bancos de dados são suportados pelo PHP, ou seja, o PHP possui código que executa funções de cada um. Entre eles, temos MySQL, PostgreSQL, Sybase, Oracle, SQL Server e muitos outros. Cada um dos bancos de dados suportados pelo PHP possui uma série de funções que você poderá usar em seus programas para aproveitar todos os recursos. Os bancos de dados não suportados diretamente pelo PHP podem ser acessados via ODBC (NIEDERAUER, 2011, p. 23-26).

PHP é a linguagem de programação utilizada no desenvolvimento do VuFind. O Apache usa PHP para transformar código do VuFind em páginas da web personalizadas para responder a solicitações de usuários, ou seja, o PHP é o responsável por impulsionar toda a interatividade do VuFind.

#### **5.1.2.4 Sistema de Gestão Integrado de Bibliotecas (ILMS, ILS)**

Um Sistema de Gestão Integrado de Bibliotecas (ILMS, ILS) é um software que

controla tradicionalmente pesquisas de catálogo, assim como a circulação e administração. VuFind é projetado para falar com um ILS de algum tipo.

### 5.1.2.5 MySQL

O MySQL é um sistema de gerenciamento de banco de dados (SGBD), que utiliza a linguagem SQL (Linguagem de Consulta Estruturada, do inglês Structured Query Language) como interface. MySQL abriga o banco de dados do aplicativo local do VuFind para seus metadados sociais, ou seja, quando os usuários adicionam tags ou deixam comentários, as informações são armazenadas no banco de dados MySQL.

## 5.2 Instalação do VuFind

VuFind é arquitetado para ser instalado em muitas plataformas diferentes, como, Fedora Linux, Ubuntu Linux e Windows.

## 5.3 Testes Efetuados

Os testes foram efetuados no sistema operacional Ubuntu, versão 14.4. Após, executar a instalação de todos os componentes com sucesso, vem a etapa da configuração final, ao abrir um navegador da web é inserido a URL (192.168.1.105/vufind/Install/Home), que direcionará a tela de configuração, conforme a figura 12.

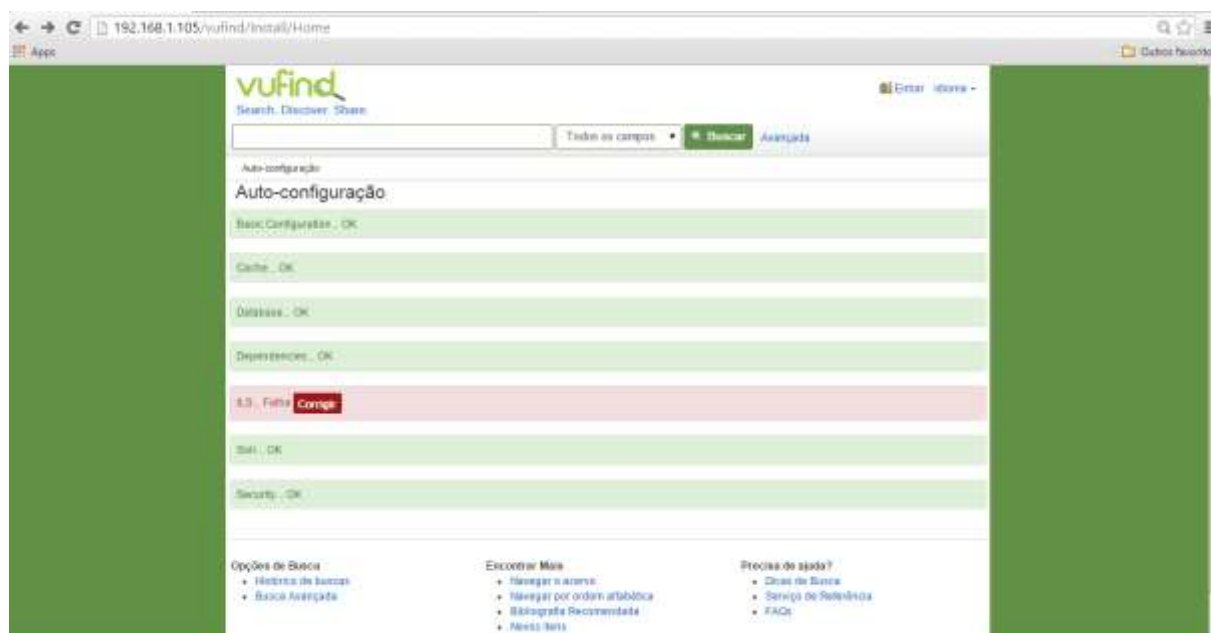


Figura 12: Tela de Autoconfiguração do VuFind.

A figura 12 apresenta a tela de autoconfiguração do VuFind. Essa tela traz uma lista com os itens de configuração marcando as que estão configuradas corretamente e os itens da lista que estão com falha, apresentando um link "Corrigir" ao lado deles. Neste caso a única configuração com falha é a ILS, ou seja, não foi configurado ainda nenhuma biblioteca, catalogo. O VuFind oferece algumas bibliotecas, como visto na figura 13, além de permitir a configuração do próprio driver do usuário.



**Figura 13: Configuração ILS - VuFind.**

VuFind aparentemente é uma ferramenta completa, oferece diversos recursos aos seus usuários, como, Opções de Busca, incluindo Histórico de buscas e Busca avançada, Navegação pelo acervo, Navegação em ordem alfabética, Bibliografias recomendadas, além de diversos recursos de ajuda. A ferramenta ainda é disponibilizada em diversos idiomas.



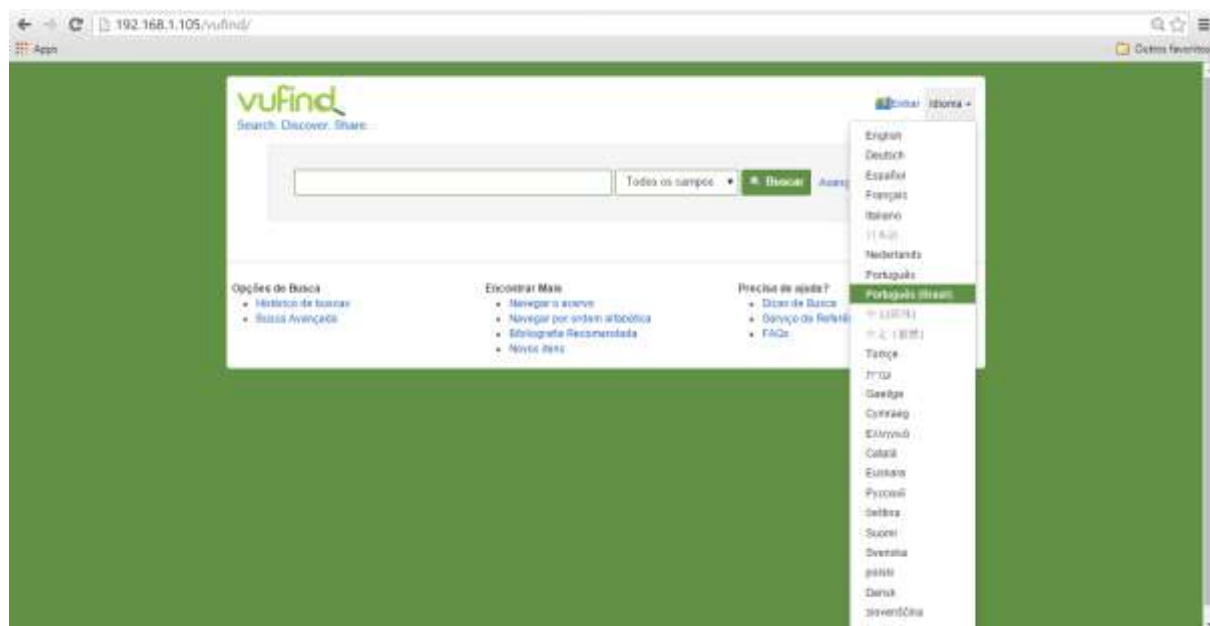


Figura 14: Tela de Busca VuFind.

## CONCLUSÃO

Este trabalho apresenta o uso da ferramenta VuFind para integração de fontes informacionais acadêmicas, buscando a melhoria do processo de recuperação da informação dessas fontes.

O objetivo desta pesquisa é propor uma Plataforma de Recuperação da Informação, baseado em ambientes informacionais digitais abertos, como repositórios digitais, sistema de gestão de aprendizagem, revistas e periódicos e a coleção do catálogo e do acervo de Bibliotecas Digitais e não digitais, visando melhorar os processos de recuperação trazendo melhor experiência ao usuário.

Para atender este objetivo, foi proposto a utilização do software livre, disponibilizado pela biblioteca Falvey Memorial Libray pertencente à Universidade Villa Nova, VuFind, que possibilita a integração de ferramentas isoladas, de forma a facilitar o acesso à informação, através do apoio a criação de portais integrados de informações geridas pelas bibliotecas. O VuFind possibilita a integração de variadas ferramentas, de forma a criar um portal de busca consolidado

Foram realizadas pesquisas referente a ferramenta, bem como a instalação da mesma no sistema operacional Ubuntu Linux, explorando suas funcionalidades. Embora seja uma ferramenta nova pode-se concluir que a integração proposta será possível realizar, o VuFind disponibiliza diversos recursos que facilitarão a integração das fontes informacionais acadêmicas.

Portanto, apresentam como trabalhos futuros a implementação da integração das fontes informacionais acadêmicas do UNIVEM, como, repositórios digitais, sistema de gestão de aprendizagem, revistas e periódicos e a coleção do acervo de Bibliotecas digitais e não digitais, bem como sua implantação, fornecendo assim, uma arquitetura que realize em um ambiente centralizado, a recuperação das informações que estão inseridas nestas diversas fontes de dados, facilitando a busca dos usuários ao realizarem uma busca, evitando perda de informação e proporcionando uma melhor visualização dos resultados.

## REFERÊNCIAS

- ALVES, R. C. V. Web semântica: uma análise focada no uso de metadados. 2005. 180 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências - Universidade Estadual Paulista, Marília, 2005.
- ALVES, M. das D. R.; SOUZA, M. I. F. Estudo de correspondência de elementos metadados: Dublin Core e MARC 21. Revista Digital de Biblioteconomia e Ciência da Informação. v. 4, n. 2, p. 20-38, 2007. Disponível em: <<http://server01.bc.unicamp.br/seer/ojs/viewarticle.php?id=94>>.
- ARMS, W. A spectrum of interoperability: the site for science prototype for the NSDL. Dlib Magazine, v. 8, n. 1, jan. 2002.
- ARMS, W.; LARSEN, R., “The Future of Scholarly Communication: Building the Infrastructure for Cyberscholarship”, 2007. [Online]. Available: <http://www.sis.pitt.edu/~repwkshop/SIS-NSFReport2.pdf>
- BAEZA- YATES, R. e RIBEIRO- NETO, B. (1999) Modern Information Retrieval. Addison-Wesley.
- BARRETT, A. The Information-Seeking Habits of Graduate Student Researchers in the Humanities. Journal of Academic Librarianship, v. 31, n. 4, p. 324-331, jul. 2005.
- BARRETO, C. M. Modelo de metadados para a descrição de documentos eletrônicos na web. 1999. 189 f. Dissertação (Mestrado em Ciências em Sistemas de Computação)–Instituto Militar de Engenharia, Rio de Janeiro, 1999.
- BORGMAN, C. L. Social aspects of digital libraries. In: WORKSHOP ON SOCIAL ASPECTS OF DIGITAL LIBRARIES, 1996, Los Angeles. Final Report. Los Angeles: UCLA/NSF, 1996.
- CARDOSO, Olinda Nogueira Paes; Recuperação da Informação. Universidade Federal de Lavras, 2000.
- CARDOSO, Olinda Nogueira Paes. Recuperação de Informação. Lavras, 2007.
- CROW, R. The Case for Institutional Repositories: A SPARC Position Paper. Washington, DC: Scholarly Publishing & Academic Resources Coalition, 2002.

- CROMWELL-KESSLER W. Correspondencias entre metadatos e interoperabilidad: qué significa todo esto?. In: BACA, M. (Ed.). Introducción a los metadatos vias a la información digital. Traducido al español por Marisol Jacas-Santoll. Los Angeles, CA: J. Paul Getty Trust, 1998.
- DENTON W, COSH SJ. Usability testing of vufind at an academic library. Library Hi Tech [Internet]. 2011.
- FERNEDA, E. Recuperação da Informação: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003.
- FERNEDA, E. Aplicando algoritmos genéticos na recuperação de informação, DataGramZero: Revista de Ciência da Informação, Rio de Janeiro, v. 10, n. 1, fev. 2009.
- FUSCO, E. Modelos conceituais de dados como parte do processo de catalogação: perspectiva de uso dos FRBR no desenvolvimento de catálogos bibliográficos digitais. Tese (Doutorado em Ciência da Informação)-Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.
- GANDEL, P. B.; R. N. KATZ; S. E. METROS. 2004. The "weariness of the flesh": Reflections on the life of the mind in an era of abundance. Educause Review 39 (2): 40–51.
- GARCIA, P. de A. B. e SUNYE M. S. O protocolo OAI-PMH para Interoperabilidade em Bibliotecas Digitais. 2007. Disponível em: <[http://conged.deinfo.uepg.br/~iconged/Artigos/artigo\\_09.pdf](http://conged.deinfo.uepg.br/~iconged/Artigos/artigo_09.pdf)>.
- GEY, F. “Models in Information Retrieval”. Folders of Tutorial Presented at the 19th ACM Conference on Research and Development in Information Retrieval (SIGIR), 1992.
- GRÁCIO, José Carlos Abbud. Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade, 2002.
- LAWRENCE,S. Institucional repositories: enhancing teaching, learning and research. EDUCAUSE Envolving Technologies Committee.
- LEWIS, S.; YATES, C. The DSpace Course – Introduction to Dspace. CADAIR, 2008
- LOPES, Ilza Leite. Estratégia de busca na recuperação da informação: revisão da literatura. Brasília, 2002.

- LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. Association of Research Libraries, n. 226, 2003. Disponível em: <<http://www.arl.org/storage/documents/publications/arl-br-226.pdf>>.
- MARC21. MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media. Library of Congress. Network Development and MARC Standards Office. 2000. Disponível em: <<http://www.loc.gov/marc/specifications/>>.
- MARCONDES, C. H. Metadados: descrição e recuperação na Web In: MARCONDES, C. H. et al. (Org.). Bibliotecas digitais: saberes e práticas. Salvador, BA : EDUFBA; Brasília; IBICT, 2005. p. 77-143
- MARINO, M. T. Integração de informações em ambientes científicos na web: uma abordagem baseada na arquitetura RDF. 2001. 122 f. Dissertação (Mestrado em Informática)– Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001. Disponível em: <[http://genesis.ncl.ufrj.br/dataware/Metadados/Teses/Teresa/pagina\\_tese.htm](http://genesis.ncl.ufrj.br/dataware/Metadados/Teses/Teresa/pagina_tese.htm)>.
- MIORELLI, Sandra Teresinha. ED-CER: Extração do Sintagma nominal em Sentenças em Português. 2001. 98f. Dissertação de Mestrado em Ciência da Computação - Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul.
- MOOERS, C. (1951). Zetocoding applied to mechanical organization of knowledge. American Documentation, v.2, n.1, p.20-32.
- MOURA, A. M. A. de C. A Web Semântica: fundamentos e tecnologias. [S. l.: S. n.], 2002.
- NASCIMENTO, L.U. Um Padrão de Metadados para Indexação e Recuperação de Objetos Multimídia, 2008. Disponível em: <<http://dspace.c3sl.ufpr.br/dspace/bitstream/handle/1884/24279/Um%20Padrao%20de%20Metadados%20Para%20Indexacao%20e%20Recuperacao%20de%20Objetos%20Multimidia.pdf?sequence=1>>.
- NISO (National Information Standards Organization). Understanding Metadata. NISO Press, 2004. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>.
- OLIVEIRA, V. S. Buscando interoperabilidade entre diferentes bases de dados: o caso da Biblioteca do Instituto Fernandes Figueira. 2005. 109 f. Dissertação (Mestrado em Saúde Pública)- Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública Sérgio Arouca, Rio de Janeiro, 2005. Disponível em: <<http://www.arca.fiocruz.br/handle/icict/1318>>.

- GANDEL, P. B; KATZ, R. N. ; METROS, S. E.. “The "Weariness of the Flesh: Reflections on the Life of the Mind in an Era of Abundance”, *EDUCAUSE Review Magazine*, vol. 39, no. 2, pp. 40–51, 2004.
- ROBERTSON, S.E. (1977) Theories and models in information retrieval. *Journal of Documentation*, 33, p.126-148.
- SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, Oxford v. 24, n. 5, p. 513 – 523, 1988.
- SANTAREM SEGUNDO, J. E. Representação Iterativa: um modelo para Repositórios Digitais. 2010. 224 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2010.
- SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. *Ciência da Informação*, v. 29, n. 1, p. 93-102, jan./abr. 2000.
- TANSLEY, Robert; BASS, Mick; SMITH, MacKenzie. *DSpace as an Open Archival Information System: Current Status and Future Directions* (2003).
- VIANA, C. L. M.; MÁRDERO ADELLANO, M. A.; SHINTAKU, M. Repositórios institucionais em ciência e tecnologia: uma experiência de customização do DSpace. 2005.
- VELLUCCI, S. L. Bibliographic relationships. *International Conference On The Principles And Future Development Of AACR*. Toronto: American Library Association: Library Association Publishing, 1998. p.105-147.
- WAINWRIGHT, E. *Digital libraries: some implications for government and education from the Australian development experience*. Canberra, Australia: National Library of Australia, 1996.
- WEIBEL, Stuart. Metadata the foundations of resources description. *D-Lib Magazine*, july 1995.

## APÊNDICE A

### **VuFind no Fedora Linux**

As instruções a seguir foram testadas no Fedora 22, porém também devem funcionar com versões recentes. As versões antigas do Fedora (16 ou anteriores) não são recomendadas, pois, incluem versões do PHP ultrapassadas o que podem se tornar incompatíveis com Vufind2.

Fedora se difere de outros sistemas operacionais, de distribuição Linux, é que o mesmo utiliza o módulo de segurança SELinux. SELinux (Security-Enhanced Linux) é uma implementação da arquitetura MAC para o sistema operacional Linux, foi integrado no Linux usando o framework LSM (Linux Security Modules). Tem o intuito de fornecer uma política de segurança sobre os processos e objetos do sistema, fundamentando suas decisões contendo informações relevantes à segurança.

Outrossim, o SELinux pode bloquear um monte de coisas, é necessário configurar a segurança de maneira adequada para que não impeça O Vufind de se comunicar com Apache Solr e suas ILS. Caso os arquivos de biblioteca tem as permissões erradas, o SELinux não permitirá carregamento do código ou sua execução, se esses problemas ocorrerem durante a instalação, deve-se adicionar SELinux a lista de solução de problemas e desativá-lo temporariamente se necessário.

- ***Instale o Fedora***

Para realizar a instalação do Fedora é possível obter uma cópia gratuita do sistema, bem como as instruções de como instalá-lo no site oficial do Fedora, disponível em [fedoraproject.org](http://fedoraproject.org).

- ***Ir para o Terminal do Sistema***

Por padrão Fedora apresenta um desktop gráfico. Para introduzir todos os comandos que serão descritos abaixo, é necessário ir para um prompt de comando, que pode ser acessado selecionando "Terminal" em Aplicativos (em versões anteriores, este foi encontrado no Ferramentas do Sistema submenu).

- **Fazer o login como usuário root**

Os comandos a seguir necessita acesso root ao sistema. Caso estiver conectado como uma conta de usuário, para ter acesso root basta digitar:

```
su
```

Após executar o comando será solicitado para a senha root do sistema. Por sua vez, o acesso root pode ser perigoso, por isso após realizar todo o trabalho na conta root é recomendável que sai da mesma, evitando maiores problemas.

- **Atualizar o Sistema**

É importante verificar se o sistemas possui todos os patches mais recentes instalados.

```
dnf update
```

O gerenciador de pacotes dnf esta disponível apenas para Fedora 22, substituindo o yum que ainda esta disponível para versões anteriores do Fedora. Então caso executar um comando dnf em uma versão mais antiga Fedora e ele não funciona, deverá substituir yum.

Posteriormente a instalação das atualizações é importante reiniciar o sistema.

```
shutdown -r now
```

- **Instalar o Apache HTTP**

Instalar o servidor web Apache, isto facilitará a comunicação entre Vufind e navegadores web. Em algumas distribuições, este é pré-instalado, mas a tentativa de instalá-lo novamente é inofensiva.

```
dnf install httpd
```

Vufind depende que o mod\_rewrite seja habilitado, embora este é ativado por padrão, pode ocorrer foi esteja desativado em uma instalação do Fedora pré-existente, por isso é preciso reativá-lo para Vufind.

Acaso a instância do Vufind incluir registros com barras em seus IDs, é preciso adicionar "AllowEncodedSlashes on" na seção <VirtualHost> da configuração do Apache.



- **Instalar MySQL/MariaDB**

Vufind utiliza o banco de dados MySQL para armazenar os comentários dos usuários, tags e outras informações, por isso deve-se instalar este componente. Algumas versões do Fedora têm substituído MySQL pelo o MariaDB, funcionalmente equivalente.

Fedora 22 ou mais recente:

```
dnf install mariadb-server
```

As versões mais antigas do Fedora:

```
dnf install httpd
```

Após a instalação do MySQL / MariaDB, carece de iniciá-lo:

Fedora 22 ou mais recente:

```
service mariadb start
```

As versões mais antigas do Fedora:

```
service mysqld start
```

Por razões de segurança, também é recomendável definir uma senha de administrador para o banco de dados; este comando guia através do processo (independentemente de ser MariaDB ou MySQL):

```
/usr/bin/mysql_secure_installation
```

- **Instalar PHP**

A maioria do código do Vufind é escrito em linguagem PHP. Além de instalar o PHP, é necessário ter a certeza que todos os módulos estão habilitados para as atender as principais tecnologias utilizadas pelos Vufind (MySQL, LDAP, etc.)

```
dnf install php php-devel php-intl php-ldap php-mysql php-xsl php-gd php-mbstring php-mcrypt
```

Se utilizar a biblioteca Voyager, é preciso instalar o PHP OCI driver para Oracle. Se você vai estar acessando um banco de dados Sybase (por exemplo, para a Horizon LMS),

também instalar php-mssql.

```
dnf install php-mssql
```

- **Instalar o Java Jdk**

Em seguida, instalar JDK (Java Development Kit) no servidor - Vufind de procura back-end.

```
dnf install java-*-openjdk-devel
```

- **Download do Vufind**

Para realizar o download do Vufind em formato tar.gz ou zip, é necessário a instalação da ferramenta wget para recuperar arquivos sobre HTTP.

```
cd /tmp
wget http://downloads.sourceforge.net/vufind/vufind-2.5.tar.gz?use_mirror=osdn -O vufind-2.5.tar.gz
tar xzvf vufind-2.5.tar.gz
mv vufind-2.5 /usr/local/vufind2
```

- **Instalar o Vufind**

Após relizar todos os processos acima, pode-se executar VuFind e instalar o script para definir sua configuração básica, mesmo aceitando os padrões por agora é permitido executado novamente mais tarde, caso precise fazer alterações.

```
cd /usr/local/vufind2
php install.php
```

Permissões de segurança apropriadas precisam ser configurado para que Apache possa acessar o código VuFind. A linha chcon define as permissões de arquivo para que Apache tem permissão para carregar os arquivos. O primeiro comando setsebool permite Apache para se comunicar com o próprio internamente, que é necessário para acessar o back-

end de pesquisa. O segundo comando `setsebool` permite que o Apache enviar e-mail, que é necessário para funções de envio de mensagem do VuFind.

```
chcon -R unconfined_u:object_r:httpd_sys_content_t:s0 /usr/local/vufind2/
setsebool -P httpd_can_network_relay=1
setsebool -P httpd_can_sendmail=1
```

A instrução `chcon` acima permite o acesso de leitura Apache de mais arquivos do que o estritamente necessário. Esta é a simplicidade, mas ao fornecer uma longa lista de diretórios, as configurações de segurança poderiam ser feitas mais focalizadas.

Também é necessário fazer alguns diretórios graváveis pelo Apache para que VuFind possa gerar arquivos de cache e configurações básicas:

```
chown -R apache:apache /usr/local/vufind2/local/cache
chcon -R unconfined_u:object_r:httpd_sys_rw_content_t:s0 /usr/local/vufind2/local/cache
chown -R apache:apache /usr/local/vufind2/local/config
chcon -R unconfined_u:object_r:httpd_sys_rw_content_t:s0 /usr/local/vufind2/local/config
```

Se planejar usar ferramentas de linha de comando do VuFind, também é preciso um cache separado para isso:

```
mkdir /usr/local/vufind2/local/cache/cli
chmod 777 /usr/local/vufind2/local/cache/cli
```

- ***Link VuFind para Apache***

Apache precisa de algumas configurações extras do VuFind carregadas. Deve-se executar esses comandos para definir o contexto de segurança apropriadas para a configuração do VuFind e, em seguida, vinculá-lo com Apache.

```
chcon system_u:object_r:httpd_config_t:s0 /usr/local/vufind2/local/httpd-vufind.conf
ln -s /usr/local/vufind2/local/httpd-vufind.conf /etc/httpd/conf.d/vufind2.conf
```

Apache precisa ser reiniciado para que as alterações entrem em vigor:

```
apachectl restart
```

- **Configurar Variáveis de Ambiente**

Algumas variáveis de ambiente precisam ser definidas para que scripts relacionados ao VuFind podem encontrar os recursos necessários. Se executar o VuFind sob uma conta de usuário específico, deve defini-los somente para esse usuário. Se quiser fazer as configurações globais para todas as contas (o mais fácil, mas não necessariamente o melhor, abordagem), apenas execute este código para criar um script no diretório `/etc/profile.d`:

```
echo export VUFIND_HOME="/usr/local/vufind2" > /etc/profile.d/vufind.sh
echo export VUFIND_LOCAL_DIR="/usr/local/vufind2/local" >> /etc/profile.d/vufind.sh
```

Depois de criar esse arquivo, deve carregá-lo manualmente para que as alterações entrem em vigor imediatamente.

```
source /etc/profile.d/vufind.sh
```

- **Configurar e Iniciar o Vufind**

Para iniciar Vufind:

```
echo export VUFIND_HOME="/usr/local/vufind2" > /etc/profile.d/vufind.sh
echo export VUFIND_LOCAL_DIR="/usr/local/vufind2/local" >> /etc/profile.d/vufind.sh
```

Para configurar o Vufind precisa-se abrir um navegador web e acessar esta URL:

```
source /etc/profile.d/vufind.sh
```

Substituir "seu-servidor-name" com o endereço que deseja usar para acessar Vufind; substituir "Vufind" com o seu caminho de base personalizado se alterou a configuração padrão durante a instalação.

- **Soluções de Problemas**

Caso receba um erro 403 proibido, isso pode ser causado por causa do Apache 2.3 ou posterior com Vufind 2.1.1 ou anterior, pode ser corrigido ajustando a configuração do Apache do Vufind e reiniciar Apache.

Se aparecer uma tela em branco, algo está errado. Verificar o log de erro do Apache (normalmente /var/log/httpd/error.log) para mensagens.

Depois que um problema é resolvido com sucesso, pode-se clicar no "Autoconfiguração" breadcrumb para voltar à lista principal e prosseguir para o próximo problema.

Para configurar o banco de dados do Vufind, precisa-se ter a senha de root que você definiu quando instalar o MySQL.

Uma vez que todos os problemas de configuração são resolvidos com êxito, terá um link "Desativar Autoconfiguração" na página "Configuração automática". Se precisar acessar novamente no futuro, pode reativá-lo manualmente editando seu config.ini arquivo.

Depois de desabilitar a configuração automática, também deve desativar a capacidade do Apache para escrever para seu diretório de configuração:

```
chown -R root: root / usr / local / vufind2 / local / config
```

```
chcon -R unconfined_u: object_r: httpd_sys_content_t: s0 / usr / local / vufind2 / local / config
```

## APÊNDICE B

### VuFind no Ubuntu Linux

As instruções a seguir foram testadas atualmente no Ubuntu 15.04, mas também deve funcionar com outras versões recentes com pouca ou nenhuma modificação. Evitar versões antigas do Ubuntu (12.04 ou anteriores) que incluem versões do PHP desatualizados que podem ser incompatíveis com Vufind 2.

Se tiver um servidor Ubuntu, será capaz de pular algumas etapas, mas você pode ter que reconfigurar alguns software existente.

- **Instalar o Ubuntu**

Para realizar a instalação do Ubuntu é possível obter uma cópia gratuita do sistema, bem como as instruções de como instalá-lo no site oficial do Ubuntu, disponível em [www.ubuntu.com](http://www.ubuntu.com).

- **Atualizar o Sistema**

O primeiro passo é certificar-se que você tem os patches mais recentes instalados.

```
sudo apt-get update  
sudo apt-get dist-upgrade
```

Após a instalação de patches, você deve reiniciar seu sistema para que tudo possa ter efeito:

```
sudo shutdown -r now
```

- **Instalando o Vufind do pacote DEB**

A maneira mais fácil de obter VuFind em funcionamento é instalá-lo do pacote DEB. Esta seção fornece instruções sobre como fazer isso. Rapidamente, ele irá carregar 2 VuFind no diretório `/usr/local/vufind2` do padrão, fazendo a suposição de que suas configurações locais pertencem em `/usr/local/vufind2/local`.

- Obtendo o pacote:

```
wget http://downloads.sourceforge.net/vufind/vufind_2.5.deb?use_mirror=osdn -O vufind_2.5.deb
```

- Instalando o pacote:

```
sudo dpkg -i vufind_2.5.deb
```

Se você não tem todas as dependências do VuFind já instaladas, o dpkg irá falhar com uma mensagem de erro. Você pode corrigir este problema instalando os requisitos ausentes usando o apt-get:

```
sudo apt-get install -f
```

- **Notas importantes**

Caso precise instalar o MySQL, será solicitado uma senha de root durante a instalação, para garantir maior segurança, será necessário lembrar da senha para configurar VuFind para acessar o banco de dados.

Se a biblioteca Voyager for utilizada, também precisará instalar o Driver OCI PHP para Oracle.

Se acessar o banco de dados Sybase (por exemplo, para o horizonte LMS), também deve instalar o pacote php5-sybase usando o apt-get.

Se, necessário remover o pacote, pode emitir este comando (Observe que - P irá remover os arquivos de configuração, bem como os componentes executáveis; usam - r para uma remoção mais cautelosa): sudo dpkg -P vufind2

Uma vez que tudo está configurado, deverá ter uma cópia do trabalho do VuFind em /usr/local/vufind2.

Para reiniciar o sistema, para ter certeza de todas as novas configurações estão no lugar ou pelo menos certificar-se de configurações das variáveis de ambiente, basta executar:

```
source /etc/profile
```

- **Instruções detalhadas de instalação**

As instruções abaixo mostram o passo-a-passo da instalação de todos os componentes fundamentais para o funcionamento do Vufind, mas acaso já realizou a

instalação do Vufind pelo pacote DEB, deverá desconsiderar esta seção.

- ***Instalar o servidor HTTP Apache***

Primeiramente deverá instalar o servidor web Apache, o que facilitará a comunicação entre os navegadores web e o VuFind. Os comandos a seguir realizam três coisas: a primeira linha instala o Apache, a segunda linha ativa o módulo exigido pelo VuFind de regravação de URL, e a terceira linha reinicia o servidor para ativar o módulo recém instalado.

```
sudo apt-get -y install apache2
sudo a2enmod rewrite
sudo /etc/init.d/apache2 force-reload
```

Se a instância do VuFind incluir registros com barras em suas identificações, será preciso adicionar "AllowEncodedSlashes on" para a seção de <VirtualHost>apropriado da sua configuração do Apache.

- ***Instalar MySQL***

VuFind usa o banco de dados MySQL para armazenar comentários de usuários, tags e outras informações. Você deve instalar este componente em seguida:

```
sudo apt-get -y install mysql-server
```

Durante a instalação, será solicitado uma senha de root do MySQL, para garantir maior segurança, mas é fundamental lembra-la mais para que possa configurar o VuFind para acessar o banco de dados mais tarde.

- ***Instalar PHP***

A maioria do código do Vufind é escrito em linguagem PHP. Além de instalar o PHP, é necessário ter a certeza que todos os módulos estão habilitados para atender as principais tecnologias utilizadas pelos Vufind (MySQL, LDAP, etc.)

```
sudo apt-get -y install php5 php5-dev php-pear php5-json php5-ldap php5-mcrypt php5-mysql
php5-xsl php5-intl php5-gd
```



Se utilizar a biblioteca Voyager, é preciso instalar o PHP OCI driver para Oracle. Se você vai estar acessando um banco de dados Sybase (por exemplo, para a Horizon LMS), também instalar php-mssql.

A biblioteca de php5-ldap só é necessária se estiver usando autenticação LDAP; podendo excluir este pacote se quiser. O pacote php5-gd também é opcional, apesar de incluí-lo irá garantir o melhor suporte para imagens de capa. Biblioteca de mbstring do PHP é altamente recomendável, mas como este escrito, ele aparece para ser automaticamente incluído no pacote php5-dev.

Caso seja uma biblioteca Voyager, também precisará instalar o Driver OCI PHP para Oracle.

Se acessar banco de dados Sybase (por exemplo LMS), também instale o php5-sybase:

```
sudo apt-get -y install php5-sybase
```

- **Instalar o Java Jdk**

Em seguida, instalar JDK (Java Development Kit) no servidor - Vufind de procura back-end.

```
sudo apt-get -y install default-jdk
```

Para versões anteriores do Ubuntu (12.04 ou anterior) é necessário especificar o "sudo apt-get - y instalar openjdk-7-jdk" em vez do padrão para obter uma versão suficientemente recente de Java.

- **Download do Vufind**

Para realizar o download do Vufind em formato tar.gz ou zip, é necessário a instalação da ferramenta wget para recuperar arquivos sobre HTTP.

```
cd /tmp
wget http://downloads.sourceforge.net/vufind/vufind-2.5.tar.gz?use_mirror=osdn -O vufind-2.5.tar.gz
tar xzvf vufind-2.5.tar.gz
sudo mv vufind-2.5 /usr/local/vufind2
```

Após realizar todos os processos acima, pode-se executar VuFind e instalar o script para definir sua configuração básica, mesmo aceitando os padrões por agora é permitido executado novamente mais tarde, caso precise fazer alterações.

```
cd /usr/local/vufind2  
php install.php
```

Também deve-se definir algumas permissões para que o Apache grave configurações e arquivos de cache no disco:

```
sudo chown -R www-data:www-data /usr/local/vufind2/local/cache  
sudo chown -R www-data:www-data /usr/local/vufind2/local/config
```

Caso pretenda usar ferramentas de linha de comando do VuFind, também precisa de um cache separado para isso:

```
mkdir /usr/local/vufind2/local/cache/cli  
sudo chmod 777 /usr/local/vufind2/local/cache/cli
```

- ***Link Vufind para Apache***

O Apache precisa de algumas configurações extras:

Apache 2.2 (Ubuntu 13,04 ou anterior):

```
sudo ln -s /usr/local/vufind2/local/httpd-vufind.conf /etc/apache2/conf.d/vufind2
```

Apache 2.4 (Ubuntu 13.10 ou posterior):

```
sudo ln -s /usr/local/vufind2/local/httpd-vufind.conf /etc/apache2/conf-enabled/vufind2.conf
```

Apache precisa ser reiniciado para que as alterações possam ser ativadas:

```
sudo /etc/init.d/apache2 reload
```

- **Configurar Variáveis de Ambiente**

Algumas variáveis de ambiente precisam ser definidas para que scripts relacionados ao VuFind possam encontrar o Java e o VuFind em si. Caso planeje executar VuFind sob uma conta de usuário específico, deve defini-los somente para esse usuário. Se quiser fazer as configurações globais para todas as contas, apenas execute este código para adicionar as linhas necessárias para um novo arquivo `/etc/profile.d`:

```
sudo sh -c 'echo export JAVA_HOME="/usr/lib/jvm/default-java" > /etc/profile.d/vufind.sh'
sudo sh -c 'echo export VUFIND_HOME="/usr/local/vufind2/" >> /etc/profile.d/vufind.sh'
sudo sh -c 'echo export VUFIND_LOCAL_DIR="/usr/local/vufind2/local/" >>
/etc/profile.d/vufind.sh'
```

Se estiver usando o Sun JVM em vez do padrão da JVM, pode precisar especificar um caminho diferente do `JAVA_HOME`, como `/usr/lib/jvm/java-x-sun`.

Depois de criar o arquivo, deve carregá-lo manualmente para que as alterações sejam efetivadas sem forçá-lo a logout:

```
source /etc/profile.d/vufind.sh
```

- **Configurar e Iniciar o Vufind**

Independentemente do método utilizado para configurar o VuFind, precisará seguir estas etapas para configurar detalhes finais e obter o código em execução.

Para iniciar o VuFind:

```
cd /usr/local/vufind2/
./vufind.sh start
```

Para configurar o Vufind precisa-se abrir um navegador web e acessar esta URL:

```
source /etc/profile.d/vufind.sh
```

Substituir "seu-servidor-name" com o endereço que deseja usar para acessar Vufind; substituir "Vufind" com o seu caminho de base personalizado se alterou a configuração padrão durante a instalação.

- ***Soluções de Problemas***

Se aparecer uma tela em branco, algo está errado. Verificar o log de erros do Apache (normalmente `/var/log/apache2/error.log`) para mensagens.

Caso não ajude, tentar editar `/usr/local/vufind2/local/httpd_vufind.conf` e eliminar o comentário da linha "SetEnv VUFIND\_ENV do desenvolvimento" – após um reinício Apache, isto irá colocar VuFind em modo de desenvolvimento (que irá exibir mensagens de erro mais detalhadas).

Há um problema conhecido no Ubuntu 13.10 que impede que o módulo de PHP mcrypt de instalá-lo corretamente. Consulte esta página para solucionar esse problema.

Ubuntu 14.04, 14.10 e 15.04 também sofrem com o problema de instalação do módulo mcrypt, que a solução é mais simples: `sudo php5enmod mcrypt`; reiniciar o apache2 serviço `sudo`. Se persistir o problema, tente uma das listas de discussão na página de suporte.

- ***Configuração Automática***

Se a instalação foi bem sucedida, agora deve aparecer uma tela de Autoconfiguração. Alguns itens da lista serão marcados como "Falha" com "Corrigir" os links ao lado deles.

Clicar em cada link de correção, e seguir as instruções.

Depois de um problema resolvido com êxito, pode-se clicar em "Configurar" para voltar à lista principal e prosseguir para a correção do próximo problema.

Para configurar o banco de dados do VuFind, é preciso ter a senha de root que foi definida durante a instalação do MySQL.

Uma vez que todos os problemas de configuração são resolvidos com êxito, terá um link "Desativar Autoconfiguração" na página "Configurar o Auto". Se necessário acessá-lo novamente, poderá reativá-lo editando manualmente o arquivo `config.ini`.

Depois de desativar a configuração automática, também deve-se desativar a capacidade do Apache para escrever para o diretório de configuração:

```
sudo chown -R root:root /usr/local/vufind2/local/config
```

Substituir o "root: root" com um usuário/grupo caso tenha configurado um determinado usuário de Linux para fins de execução VuFind; substituir `/usr/local/vufind2` com seu caminho de base VuFind 2 se tiver personalizado o local de instalação).

## APÊNDICE C

### VuFind no Windows

Para satisfazer as dependências do VuFind, é fundamental certifique-se de instalar as versões de software recomendadas na página de requisitos. Essas instruções assumem que está instalando todos os componentes separadamente. Muitos desenvolvedores usam o XAMPP para instalar o Apache, MySQL e PHP como um pacote. XAMPP não deve ser usado em um ambiente de produção, mas pode poupar algum tempo para um servidor de teste, no entanto, ele funciona de forma diferente do que a instalação do componente separado.

- ***Download do Vufind***

O download do Vudfnd pode ser realizado através do link <https://vufind-org.github.io/vufind/downloads.html>, sendo que o formato .zip, geralmente é a opção mais conveniente para uso em Windows. Essas instruções presumem a utilização de um caminho de destino de c: / vufind.

- ***Instalar o servidor HTTP Apache***

O servidor está disponível para download no site oficial do Apache (<http://httpd.apache.org/download.cgi>), bem como as instruções e opções de instalação.

Certificar-se que o módulo `mod_rewrite` é ativado, editar o arquivo de configuração do Apache (algo como `c:\Program programas\Apache Software Foundation\Apache#\conf\httpd.conf`, facilmente disponível no grupo do Menu Iniciar do Apache sob o Apache HTTP Server, configurar o Apache Server) e verifique se esta linha está descomentada (não tem um # na frente dela):

```
LoadModule rewrite_module modules/mod_rewrite.so
```

Também certifique-se de que o comando `DirectoryIndex` inclui o `index.php`:

```
DirectoryIndex index.html index.php
```

Se a instância VuFind incluir registros com barras em suas identificações, precisa adicionar "AllowEncodedSlashes on" para a seção de <VirtualHost>apropriado da configuração do Apache.

- **Instalar MySQL**

O MySQL está disponível para download no seu site oficial (<http://dev.mysql.com/downloads/mysql/5.0.html#downloads>), com todas as instruções da instalação padrão da ferramenta. Acaso seja definida uma senha de root, que é recomendável para garantir maior segurança, certificar-se de tomar nota disso.

Uma parte importante da instalação do MySQL é o arquivo de biblioteca, libmysql.dll. Se o Windows não consegue encontrar este arquivo, pode ocorrer erros sobre o driver de banco de dados ausente quando tentar visualizar páginas do VuFind. Dependendo de quais versões de software do sistema operacional, pode não precisa fazer nada especial para fazer este trabalho. No entanto, se tiver problemas depois de completar as outras etapas do processo de instalação VuFind, tentar estas opções:

- 1) Certifique que o libmysql.dll esta no diretório do PHP e que este diretório está incluído no caminho do sistema (como descrito abaixo, na etapa de Java JDK).

- 2) se libmysql.dll não está no seu diretório do PHP, pode ser no diretório do MySQL (normalmente algo como C:\Program Files\MySQL\MySQL Server 5.0\bin). Para adicionar este diretório de caminho do sistema também pode ajudar. Se o arquivo estiver em ambos os diretórios do PHP e MySQL, a versão do PHP é provavelmente a correta, usar o MySQL versão pode causar problemas. Em alguns cenários, pode-se realmente precisar remover o diretório MySQL do caminho para resolver este problema.

- 3) como um último recurso, tentar copiar libmysql.dll para a pasta c:\windows\system32, embora não seja o mais recomendado, pois pode dificultar futuras atualizações.

- **Instalar PHP**

O PHP para está disponível para download no site <http://windows.php.net/download/>. A escolha da versão deve se a segura do segmento, atentando para qual versão do Apache foi instalada, pois, o PHP e Apache se integram melhor se eles são construídos para a mesma arquitetura (32 bits ou 64 bits), usando o mesmo compilador. Por exemplo, se baixou uma

versão de 32 bits (x86) do Apache, não se deve tentar instalar uma versão de 64 bits (x64) do PHP.

Escolha a versão apropriada do Apache e forneça o diretório de configuração (por exemplo, `c:\Program programas\Apache Software Foundation\Apache#.#\conf\`). Se receber uma mensagem de erro, pode corrigi-lo manualmente mais tarde

Quando chegar ao estágio de "Escolher itens para instalar" do processo de instalação, certifique-se de instalar o PEAR (encontrado sob Extras) e essas extensões (encontrado em PHP / Extensions): GD2 (opcional, mas recomendada para melhor visualização de imagem de capa), Internacionalização (intl), LDAP (opcional), Sequência de caracteres de múltiplos Byte (opcional, mas recomendado), MySQL, MySQLi, DOP (com MySQL) e XSL.

Certifique-se de que a variável de ambiente PATH do Windows inclui o local do PHP (esse local foi especificado durante a instalação). Isso permitirá que digite "php" no prompt de comando para executar scripts.

Caso algum componente não tenha sido instalado na primeira instalação, é possível executar o instalador novamente para atualizar a instalação com módulos adicionais.

É possível que o PHP irá instalar algumas extensões sem todos os arquivos necessários de biblioteca. Para corrigir o problema, basta editar o arquivo `php.ini` (geralmente encontrado em `c:\Program Files\PHP\` e também está disponível através do atalho do Menu Iniciar) e comente fora (colocando uma ponto e vírgula na frente) os módulos ofensivos (encontrado em versos como extensão = `php_*.dll`). Se realmente precisar de alguns dos módulos afetados, pode ter que rastrear cópias dos arquivos ausentes.

Se estiver usando o XAMPP em vez de componentes separados, o processo de instalação não vai perguntar sobre extensões. Em vez disso, terá que manualmente, descomentar as linhas relevantes no seu arquivo `php.ini`.

Se utilizar uma biblioteca Voyager, você também precisará instalar o Driver OCI PHP para Oracle.

- **Instalar o Java Jdk**

Para obter o instalador do JDK, basta acessar o site da oracle (<http://www.oracle.com/technetwork/java/index.html>) e realiza o download, lá também estão disponíveis todas as opções de instalação padrão.

Definir uma variável de ambiente do sistema chamada `JAVA_HOME` apontando para o caminho para instalação do Java (como `c:\Program Files\Java\jdk1.x.y_z`).

- ***Executar o Script de instalação do Vufind***

Após realizada a instalação de todos os componentes pode-se executar VuFind instalar o script para definir sua configuração básica. Para executar o script, é necessário ir em Iniciar &gt; executar e digitar "cmd" sem aspas (para chegar a um prompt de comando).

No Windows 7, não há nenhuma opção de "Executar", clique no comando "cmd" e execute como administrador para garantir que tudo fique instalado com permissões apropriadas. Deverá ainda, alternar para o diretório VuFind (ou seja, "cd c:\vufind"), digitar "install. php" (sem as aspas) e tecla Enter, será solicitado vários detalhes, pode aceitar os padrões inicialmente, caso necessário é possível instalar o instalador novamente e fazer as alterações necessárias.

O passo final do script de instalação é configurar as variáveis de ambiente VUFIND\_HOME e VUFIND\_LOCAL\_DIR.

- ***Link Vufind para Apache***

Edite o arquivo de configuração do Apache (como c:\Program programas\Apache Software Foundation\Apache#.#\conf\httpd.conf, facilmente disponível no grupo do Menu Iniciar do Apache sob o Apache HTTP Server, configurar o Apache Server) e adicione:

```
Include c:/vufind/local/httpd-vufind.conf
```

Certifique-se de ajustar o caminho, se seu arquivo está localizado em algum lugar que não seja c: / vufind/local.

Uma vez que os arquivos de configuração são editados, precisa reiniciar o Apache para eles funcionem, isso pode ser feito no menu iniciar sob o Apache HTTP Server, Control Apache Server, Restart.

A fim de tratar adequadamente os tempos limite de rede, deve modificar máximo tempo de execução e padrão soquete de configurações de tempo limite do PHP. Pode ser feito no php. ini do servidor Web ou adicionando as seguintes linhas ao arquivo http-vufind.conf:

```
php_value default_socket_timeout 25  
php_value max_execution_time 30
```

Pode-se escolher valores inferiores ou superiores (ambos em segundos), dependendo



do seu ambiente. O tempo limite de soquete deve ser sempre menor do que o tempo máximo de execução. O tempo para se conectar a um host remoto (`default_socket_timeout`) conta para o tempo máximo que é permitido executar um script PHP. Em \* nix plataformas não existir.

Problemas de configuração de sistema específica podem impedir algumas coisas. Uma configuração incorreta pode impedir o Apache de iniciar. Se a instalação do PHP não pôde adicionar automaticamente o PHP para a configuração do Apache, pode precisar adicionar estas linhas para configuração do Apache, em algum lugar acima o "incluir" `http-vufind.conf`:

```
LoadModule php5_module "c:\Program Files\PHP\php5apache2_2.dll"  
  
AddType application/x-httpd-php .php  
  
PHPIniDir "c:\Program Files\PHP\"
```

Nesse caso pode ser preciso substituir um caminho ligeiramente diferente para PHP, dependendo da versão do Windows, no Windows 7, será "c:\Program Files (x86)\PHP\").

- **Configurar Solr**

Os arquivos de configuração para cada índice Solr atualmente são configurados para usar caminhos relativos, pois, no Windows os serviços executados em um contexto de usuário diferente com diferentes diretórios de trabalho, Solr pode ter problemas para encontrar seus arquivos de índice. Para solucionar problemas como esse deve reconfigurar Solr e referir explicitamente onde se encontram os arquivos de índice.

Nos seguintes arquivos:

```
c:/vufind/solr/biblio/conf/solrconfig.xml  
c:/vufind/solr/authority/conf/solrconfig.xml  
c:/vufind/solr/stats/conf/solrconfig.xml
```

Localizar o nó:

```
\config\dataDir
```

Mudar os caminhos relativos atualmente para:

```
c:/vufind/solr/biblio  
c:/vufind/solr/authority  
c:/vufind/solr/stats
```

Por exemplo, mudar isto:

```
<dataDir>${solr.data.dir:./solr/biblio}</dataDir>
```

Para isto:

```
<dataDir>c:/vufind/solr/biblio</dataDir>
```

- ***Iniciar o Vufind***

Agora que o VuFind está instalado e com sua configuração básica completa, pode iniciar o servidor executando.

- ***Configuração Final***

Abra um navegador da web e navegue para este URL:

```
http://your-server-name/vufind/Install/Home
```

Substituir "seu--nome do servidor" com o endereço que deseja usar para acessar VuFind; substituir "vufind" com caminho de base personalizado se alterado durante a instalação da configuração padrão.

- ***Soluções de Problemas***

Se você ainda está preso, tente uma das listas de discussão na página de suporte.

Se aparecer uma tela em branco, algo está errado. Verificar o log de erros do Apache para mensagens, solucionando problemas do Apache.

Caso não ajude, tentar editar `c:/vufind/local/httpd_vufind.conf` e eliminar o comentário da linha "SetEnv VUFIND\_ENV do desenvolvimento", após um reiniciar o Apache, isto irá colocar VuFind em modo de desenvolvimento (que irá exibir mensagens de erro mais detalhadas, se o código é capaz de correr).

- ***Configuração Automática***

Se a instalação foi bem sucedida, agora deve aparecer uma tela de Autoconfiguração.

Alguns itens da lista serão marcados como "Falha" com "Corrigir" os links ao lado deles.

Clicar em cada link de correção, e seguir as instruções.

Depois de um problema resolvido com êxito, pode-se clicar em "Configurar" para voltar à lista principal e prosseguir para a correção do próximo problema.

Para configurar o banco de dados do VuFind, é preciso ter a senha de root que foi definida durante a instalação do MySQL.

Uma vez que todos os problemas de configuração são resolvidos com êxito, terá um link "Desativar Autoconfiguração" na página "Configurar o Auto". Se necessário acessá-lo novamente, poderá reativá-lo editando manualmente o arquivo `config.ini`.