

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-8 – Informação e Tecnologia

ARQUITETURA PARA RECUPERAÇÃO DE METADADOS DE OBJETOS DE APRENDIZAGEM

Luciana Maria Vieira Pöttker (Instituto Federal do Paraná - IFPR)

Edberto Ferneda (Universidade Estadual Paulista - UNESP)

ARCHITECTURE FOR LEARNING OBJECTS METADATA RETRIEVAL

Modalidade da Apresentação: Comunicação Oral

Resumo: Desde a sua criação, a Web tem crescido a um ritmo sem precedentes, situação esta que exigiu mudanças na forma como a sociedade busca e tem acesso à informação. O incremento informacional foi observado em todas as áreas do conhecimento e, desde então, problemas relacionados à recuperação de informação são investigados. O objetivo deste trabalho é propor a unificação da representação dos objetos de aprendizagem disponibilizados em diferentes repositórios e permitir que o usuário realize buscas qualificadas para encontrar os recursos educacionais adequados para sua necessidade de informação, por meio de uma arquitetura para recuperação de metadados de objetos de aprendizagem. Esta pesquisa é classificada como qualitativa e de natureza aplicada, uma vez que se relaciona com o problema prático de recuperação de metadados de objetos de aprendizagem disponíveis em repositórios da Web. Os procedimentos metodológicos consistiram em uma análise exploratória, descritiva e aplicada sobre os padrões e os instrumentos que orientam o desenvolvimento dos sistemas de recuperação de informação, auxiliando na compreensão do problema proposto e na concepção dos requisitos para a proposição da arquitetura para recuperação de metadados de objetos de aprendizagem baseado no princípio de *relevance feedback*. O principal diferencial desta proposta foi de valorizar a inferência do usuário no processo de recuperação de informação, por meio do processo de *relevance feedback*. Neste processo, o usuário estabelece um diálogo com o sistema de recuperação de informação realizando refinamentos nos resultados que lhe foram retornados. Como esse processo é cíclico, ele pode ser executado até que o usuário esteja satisfeito com os resultados que lhe foram retornados. Conclui-se que um sistema de recuperação de informação é mais eficiente quando amplia seu escopo de recuperação a partir de diferentes fontes de dados e permite a inferência do usuário no julgamento da informação que lhe foi retornada.

Palavras-Chave: Recuperação de Informação; Objetos de Aprendizagem; Metadados; Agentes; Realimentação por relevância.

Abstract: The Web has been growing in a record speed since its creation and, therefore, such prospect has demanded changes in the way society seeks for and accesses information. Informational increment was evident in all fields of knowledge and since then, the relevant

information retrieval issues have been investigated. The main purpose to propose the unification of the representation of the learning objects available in different repositories and allow the user to perform qualified searches to find the appropriate educational resources for their information needs through an architecture for retrieving metadata of learning objects. This is a qualitative and applied research once it relates to the practical problem of learning objects retrieval available on the Web. The methodological procedures consisted of an exploratory, descriptive and applied analysis on the standards and instruments that guide the development of information retrieval systems, helping to understand the proposed problem and the design of the requirements for the architecture proposition for the retrieval of metadata from learning objects based on the principle of relevance feedback. The main difference of this suggestion was to value – via relevance feedback – the importance of the user's inference in the process of such retrieval, in which the user establishes a dialog with the information retrieval system as to enhance the obtained results, and thus – being a cyclical process – it can be executed until he is pleased them. The conclusion is that an information retrieval system more efficient when its scope is enlarged from the different sources of data and allows the inference of the user when judging what he was presented with.

Keywords: Information Retrieval; Learning Objects; Metadata; Agents; Relevance Feedback.

1 INTRODUÇÃO

Os objetos de aprendizagem (OA) são recursos educacionais planejados para incrementar as estratégias de ensino e aprendizagem para as mais variadas áreas do conhecimento. Esses recursos, frequentemente disponibilizados na *Web*, são desenvolvidos em diversos formatos e podem ser adaptados para serem aplicados em diferentes contextos de aprendizagem. Conforme a IEEE LTSC LOM (2002) os objetos de aprendizagem são qualquer tipo de recurso (digital ou não digital) que podem ser utilizados, reutilizados ou referenciados durante todo o processo de ensino e aprendizagem apoiado pelo computador.

Nas últimas décadas, com o advento da *Web* e a incorporação das tecnologias de informação e comunicação (TIC) na esfera educacional, torna-se difícil dimensionar a quantidade de repositórios e/ou objetos de aprendizagem existentes na *Web*. Recuperar essas informações utilizando os motores de busca de uso geral (como por exemplo, o Google), os metabuscadores ou os recursos de busca específicos de cada repositório de objetos de aprendizagem (ROA) são um fator limitante ao acesso desses recursos informacionais, uma vez que o conjunto de resultados alcançados pode ser muito amplo, no caso dos buscadores de uso geral ou dos metabuscadores, ou muito restrito, no caso das buscas em um único repositório.

Esse cenário propõe uma reflexão a respeito da recuperação de informação (RI) na *Web* como um problema desafiador para a sociedade da informação, pois a maneira como a sociedade busca e tem acesso à informação tem sido alterada bruscamente tanto pela forma como ocorre o incremento informacional quanto pela dinamicidade e volatilidade das informações nesse ambiente. Todos os problemas e dificuldades relacionados à recuperação de informação na *Web* se refletem no contexto particular dos objetos de aprendizagem, acrescido da complexidade relacionada às características inerentes a esses tipos de objetos digitais.

De acordo com Baeza-Yates e Ribeiro-Neto (2013, p.4), o principal propósito de um sistema de recuperação de informação (SRI) é auxiliar os usuários a encontrar informações de seu interesse, buscando sempre maximizar a proporção de satisfação do usuário e, conseqüentemente, minimizar seu esforço na busca pela informação desejada. No escopo de um sistema de recuperação de informação, normalmente, o usuário desconhece o *corpus* a ser recuperado e nem sempre consegue formular uma expressão de busca que represente

adequadamente sua necessidade de informação. Entretanto, ao avaliar os itens retornados se torna mais fácil identificar o que é ou não pertinente à sua busca naquele momento.

Quando se trata da recuperação de objetos de aprendizagem descritos por um padrão de metadados, parte-se do princípio de que a descrição desses recursos pode facilitar a inferência do usuário na análise de relevância das características dos itens que lhe foram retornados. Como os objetos de aprendizagem possuem diversas especificidades em sua descrição é possível estabelecer um diálogo do usuário com o SRI para que os resultados sejam melhor classificados. Esse processo dialógico permite que o usuário realize refinamentos sobre os resultados que lhe foram retornados, por meio de um processo denominado de *relevance feedback* (RF), com isso é possível aprimorar os resultados recuperados sem que seja necessário que o usuário refaça sua expressão de busca inicial.

O RF é um processo que permite obter informações complementares sobre a necessidade de informação do usuário, por meio da seleção de itens relevantes recuperados de um *corpus*. Esse processo é utilizado para melhorar a qualidade e o ranqueamento dos resultados obtidos pelo motor de recuperação de informação. O propósito do RF é incrementar o desempenho do SRI, por meio da inferência do usuário no refinamento dos resultados de uma busca.

Dessa forma, o objetivo deste trabalho é propor a unificação da representação dos objetos de aprendizagem disponibilizados em diferentes repositórios e permitir que o usuário realize buscas qualificadas para encontrar os recursos educacionais adequados para sua necessidade de informação, por meio de uma arquitetura para recuperação metadados de objetos de aprendizagem.

A metodologia utilizada ao longo deste trabalho consistiu em uma análise exploratória, descritiva e aplicada sobre recuperação de objetos de aprendizagem, o que permitiu a construção de um embasamento teórico sobre os princípios, padrões e instrumentos que orientam o desenvolvimento dos sistemas de recuperação de informação, auxiliando na compreensão do problema proposto e na concepção dos requisitos para a proposição da arquitetura para recuperação de metadados de objetos de aprendizagem baseado no princípio de RF. Com base na abordagem do problema, essa pesquisa pode ser classificada como qualitativa e de natureza aplicada, uma vez que se relaciona com um problema prático de recuperação de objetos de aprendizagem disponíveis em repositórios da *Web*.

2 OBJETOS DE APRENDIZAGEM

Wiley (2001) define os objetos de aprendizagem como elementos instrucionais, baseados no paradigma da orientação a objetos da Ciência da Computação, que valorizam a criação de componentes (objetos) que podem ser reutilizados em múltiplos contextos de aprendizagem. Sendo assim, parte-se do princípio de que os *designers* tenham condições de produzir pequenos componentes pedagógicos (unidades de ensino) que sejam reutilizados para gerar novas experiências de aprendizagem, adaptadas às necessidades de aprendizagem de cada educando.

Na concepção de Ip, Morrison e Currie (2001), um objeto de aprendizagem é percebido como uma estrutura que deve ter, no mínimo, quatro subcomponentes que são: o conteúdo, as funções pedagógicas, os objetivos de aprendizagem e, por último, serem visuais e permitirem sua experimentação (*look and feel*). Esses componentes indicados pelo autor garantem que o objeto de aprendizagem cumpra seu papel de facilitar e enriquecer o processo de ensino e aprendizagem.

O domínio de objetos de aprendizagem é amplo e não há como dimensionar a quantidade desses recursos que estão espalhados por repositórios ou iniciativas individuais na *Web*. Esse domínio ainda apresenta outras características relacionadas à dinamicidade e volatilidade desses recursos, uma vez que novos objetos de aprendizagem podem ser adicionados, excluídos e modificados a qualquer tempo. Diante deste contexto, o processo de recuperação desses recursos se torna uma tarefa complexa visto que o acervo a ser recuperado é distribuído e composto por uma multiplicidade de formatos muito ampla.

2.1 Repositórios de Objetos de Aprendizagem – ROA

Os objetos de aprendizagem são mantidos em repositórios específicos para esse fim. Esses repositórios são ambientes digitais que têm como propósito disponibilizar recursos informacionais, descritos em um padrão de metadados educacional, a fim de facilitar a busca e o acesso a esses objetos, bem como para garantir sua interoperabilidade e reuso.

Um ROA pode ser criado por uma iniciativa isolada ou envolver um consórcio de instituições acadêmicas. Possui sua própria infraestrutura e organização, que contempla desde a adoção de um padrão de metadados educacional para a descrição dos recursos informacionais até um motor de busca próprio para viabilizar a recuperação de informação, por meio de uma interface de busca. Essa situação limita a recuperação de informação ao seu

próprio *corpus* de recursos educacionais, uma vez que os resultados estarão limitados pela base de dados suportada pelo ROA.

Nesses ambientes, os objetos de aprendizagem são classificados, descritos, preservados e disponibilizados para que sejam reusados em diferentes cenários. Os ROA são mantidos por iniciativas privadas ou governamentais e podem ser temáticos (referem-se a uma única área de conhecimento) ou institucionais (compreendem várias áreas do conhecimento). De acordo com Santarém Segundo (2010, p.154), os repositórios digitais, independente se são temáticos ou institucionais, possuem características comuns como, por exemplo: a facilidade de acesso à informação e a interoperabilidade entre os recursos informacionais.

Segundo Afonso et al. (2011), os repositórios de objetos de aprendizagem são sistemas de informação que têm como propósito facilitar a interoperabilidade e reutilização de recursos educacionais, por meio da disponibilização de um acervo dinâmico e que contribui para a realização de diversas práticas educacionais. Assim, a finalidade principal dos ROA é permitir que os usuários localizem os recursos educacionais de diferentes conteúdos, níveis de granularidade, qualidades e formatos em um mesmo ambiente (MOHAN, 2007, p.527).

2.2 Metadados de Objetos de Aprendizagem

Um padrão de metadados é um modelo de dados que tem sua própria sintaxe e semântica, com características muito similares a uma linguagem de marcação como XML (*eXtensible Markup Language*) ou ao RDF (*Resource Description Framework*), por exemplo. Diante deste contexto, percebe-se que ocorre uma convergência de diferentes comunidades em desenvolver seu próprio padrão de metadados, com a intenção de descrever de forma mais completa possível seu recurso informacional para facilitar o processo de recuperação de informação.

Conforme Chowdhury (2010, p.61) as funções primárias de um padrão de metadados são facilitar a: identificação, localização, recuperação, manipulação e o uso objetos digitais na *Web*. Diante desse contexto, percebe-se que existem várias iniciativas, tais como: ISO (*International Organization for Standardization*), IEEE (*Institute of Electrical and Electronic Engineers*), LTSC (*Learning Technology Standards Committee*) e DCMI (*Dublin Core Metadata Initiative*) que têm reunido esforços no desenvolvimento de padrões de metadados

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

educacionais com a intenção de garantir a interoperabilidade dos objetos de aprendizagem e permitir uma recuperação mais relevante desses recursos informacionais.

Os padrões de metadados educacionais mais utilizados são os apresentados no Quadro 1, que sintetiza as principais características suportadas por cada padrão de metadados.

Quadro 1 – Síntese das características dos padrões de metadados educacionais.

METADADOS	COMPATIBILIDADE	IDIOMA	ELEMENTOS	PADRÃO
Ariadne (ARIADNE, 2015)	Serviu de base para o LOM, embora atualmente não seja adotado como padrão de metadados, caracteriza-se como ROA.	Inglês	44 elementos divididos em seis categorias (<i>General, Semantics, Pedagogical, Technical, Indexation, Annotations</i>)	Aberto / Internacional
AICC – <i>Aviation Industry CBT Committee</i> (BIANCO, De MARSICO; TEMPERINI, 2005)	Específico para capacitação na área de aviação. Propõe uma estruturação de cursos e do ambiente de execução (<i>e-learning</i>) para treinamentos baseados em computador e na <i>Web</i> .	Inglês	Não define nenhuma categoria/ elemento para o padrão. Identifica sete arquivos (<i>Course Description File, Assignable Unit Table, Descriptor Table, Course Structure Table, Objectives Relationships File, Prerequisite Listing, Completion Requirements</i>) que são usados para descrever a estrutura e conteúdo de um curso.	Aberto / Internacional
DC – <i>Dublin Core</i> (DCMI, 2016)	Descrição de qualquer tipo de recurso informacional. Garante os princípios extensibilidade, modularidade e interoperabilidade semântica.	Inglês	15 elementos divididos em três categorias (<i>Content, Intellectual Property, Instance</i>). Também é possível utilizar o DC em nível qualificado para permitir que os elementos sejam refinados e apresentem valores padronizados por esquemas de codificação externos, a fim de tornar a descrição mais precisa. Os qualificadores são organizados em: elementos de refinamento e esquemas de codificação.	Aberto / Internacional
IMS-LRM – <i>Instructional Management</i>	Baseado em DC e LOM	Inglês	Composto por 5 partes: (<i>Meta-data, Content Packing, Question & Test</i>)	Aberto / Internacional

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

<i>Systems - Learning Resource Metadata (IMS LRM, 2015)</i>			<i>Interoperability, Learning Design e Simple Sequencing)</i>	
ISO-MLR – ISO <i>Metadata for Learning Resources</i> (ISO MLR, 2015)	Baseado em DC e LOM	Inglês	Consiste de 11 partes (<i>Framework, Dublin Core elements, Basic application profile, Technical elements, Educational elements, 'Availability, distribution and intellectual property elements', Bindings, Data elements for MLR records, Data elements for people, Application profile for access, distribution and intellectual property (WIPO compliant) elements, Migration from LOM to MLR</i>)	Aberto / Internacional
LOM – <i>Learning Object Metadata</i> (IEEE LTSC LOM, 2015)	Baseado em Ariadne	Inglês	76 elementos divididos em nove categorias (<i>General, Lyfe Cicle, Meta – Metadata, Technical, Educational, Rights, Relation, Annotation, Classification</i>)	Aberto / Internacional
OBAA – Objetos de Aprendizagem Baseado em Agentes (VICCARI et al., 2009)	Compatível com LOM	Inglês / Português	Contempla todas as categorias do LOM e adiciona novos elementos para as categorias <i>technical</i> e <i>educational</i> , totalizando 108 elementos.	Aberto / Nacional

Fonte: elaboração própria.

Percebe-se que todos os padrões de metadados possibilitam a descrição dos objetos de aprendizagem em seus respectivos repositórios, independente do nível de descrição que cada padrão suporta. Um fator que deve ser destacado é que os padrões DC e LOM são os que possuem uma maior compatibilidade com os demais e são os mais utilizados para a descrição dos objetos de aprendizagem nos ROA.

3 RECUPERAÇÃO DE INFORMAÇÃO

A área de Recuperação de Informação é considerada como sendo uma disciplina interdisciplinar, compreendida tanto pela Ciência da Informação quanto pela Ciência da Computação e tem como premissa básica disponibilizar o acesso facilitado às informações de acordo com o interesse do usuário.

Sob o olhar da Ciência da Informação, a recuperação de informação é investigada por meio de uma interpretação mais centrada no usuário, na maneira como as pessoas compreendem, buscam e utilizam a informação. Na perspectiva da Ciência da Computação, a recuperação de informação é pesquisada através do desenvolvimento de algoritmos e técnicas de recuperação de informação que visam aumentar o desempenho dos sistemas de recuperação de informação.

Neste trabalho, parte-se do enfoque da recuperação de informação sob o olhar da Ciência da Informação, que se preocupa com o grau de satisfação do usuário na busca e recuperação de informação que lhe interessa. Entretanto, a visão computacional não pode ser ignorada uma vez que fornece as estratégias para tratar, representar e recuperar grandes volumes informacionais.

Meadow et al. (2007, p.3) destacam que a recuperação de informação é um processo de comunicação que inclui o usuário, que formula sua expressão de busca por meio de uma interface e perpassa por todo processamento computacional pertinente ao processo de RI. Portanto, a recuperação de informação pode ser encarada como um sistema comunicacional que busca identificar em um *corpus* quais itens (a princípio) que atendem à necessidade de informação do usuário.

Na visão de Salton e McGill (1983, p.7), os sistemas de recuperação de informação podem ser conceituados como um sistema que trata da representação, do armazenamento, da organização e do acesso aos itens de informação. Harter (1986, p.2) complementa essa definição quando afirma que um SRI é um dispositivo que se interpõe entre os usuários e o *corpus* a ser recuperado.

Os modelos clássicos de recuperação de informação são:

- Booleano: tem seus fundamentos baseado na Teoria dos Conjuntos e na Álgebra Booleana (FERNEDA, 2012, p.2);
- Espaço vetorial: foi proposto por Salton (1971) e utiliza o sistema de atribuição de pesos para definir o ranqueamento dos documentos recuperados;

- Probabilístico: foi proposto por Maron e Kuhns (1960) com o propósito de recuperar informação por meio dos fundamentos da probabilidade para estimar a relevância e considerar o grau de incerteza no julgamento entre um *corpus* documental e a expressão de busca do usuário.

Todo sistema de recuperação de informação precisa se basear em um dos modelos clássicos que são responsáveis por determinar como será o funcionamento computacional do SRI.

3.1 Relevance Feedback

Segundo Rocchio (1971), o *relevance feedback* faz parte do domínio de recuperação de informação e é definido como um processo no qual as consultas de um usuário de um SRI são incrementadas automaticamente, a partir de consultas iniciais formuladas pelo usuário com base em julgamentos de relevância para os documentos recuperados de um *corpus*. Contribuindo com esta mesma visão Buckley e Salton (1995), afirmam que *relevance feedback* é um processo automático de refinamento a uma busca (inicial), que utiliza informações fornecidas pelo usuário sobre a relevância dos documentos previamente recuperados buscando apresentar resultados cada vez mais apurados ao usuário.

A relevância é um dos conceitos centrais de um SRI e está diretamente relacionada com o ponto de vista do usuário na seleção da informação desejada naquele momento. Neste contexto, para avaliar a relevância da informação recuperada por um SRI é necessário que ocorra a participação do usuário no julgamento das informações recuperadas.

O *corpus* a ser recuperado pelo SRI, geralmente, é desconhecido pelo usuário. Por isso, considera-se que as informações coletadas durante o processo de descrição do objeto de aprendizagem, quando apresentadas ao usuário, podem auxiliá-lo no julgamento dos itens que tenham uma maior probabilidade de serem relevantes para sua busca.

Desta maneira, acredita-se que o processo de *relevance feedback* quando aplicado ao domínio dos OA pode facilitar a compreensão do usuário de sua necessidade de informação. Com esse processo é possível estabelecer um diálogo do usuário com o SRI que visa melhorar o ranqueamento dos itens retornados sem que ele precise alterar sua expressão de busca inicial. Este processo permite selecionar os itens julgados como relevantes e realizar uma nova submissão ao SRI para que um novo ranqueamento seja realizado com base na inferência do usuário.

3.2 Agentes

Os agentes inteligentes são inseridos neste trabalho, com o intuito de potencializar a busca por informações estruturadas (metadados) em diferentes repositórios de objetos de aprendizagem para, conseqüentemente, melhorar a eficácia de um SRI.

No escopo da inteligência artificial (IA), os agentes inteligentes (ou racionais, como também são denominados) são considerados como sistemas computacionais que simulam o comportamento humano agindo em conformidade com determinadas situações e estímulos, ou seja, pode-se defini-los como uma entidade (semi ou completamente autônoma) que atua coerentemente de acordo com suas percepções do exterior e de seu estado de conhecimento para escolher as ações em função de suas percepções para atingir suas metas.

De acordo com Russel e Norvig (2004, p.33), “um agente é tudo o que pode ser considerado capaz de perceber seu ambiente por meio de sensores e de agir sobre esse ambiente por intermédio de atuadores”. Assim, os agentes têm a função de interagir com o ambiente de tal forma que tenham capacidade de reunir informações sobre o meio em que estão inseridos, para que estejam aptos a tomar decisões baseados nestas informações e iniciar execuções específicas conforme a ação percebida neste ambiente.

4 ARQUITETURA PARA RECUPERAÇÃO DE METADADOS DE OBJETOS DE APRENDIZAGEM

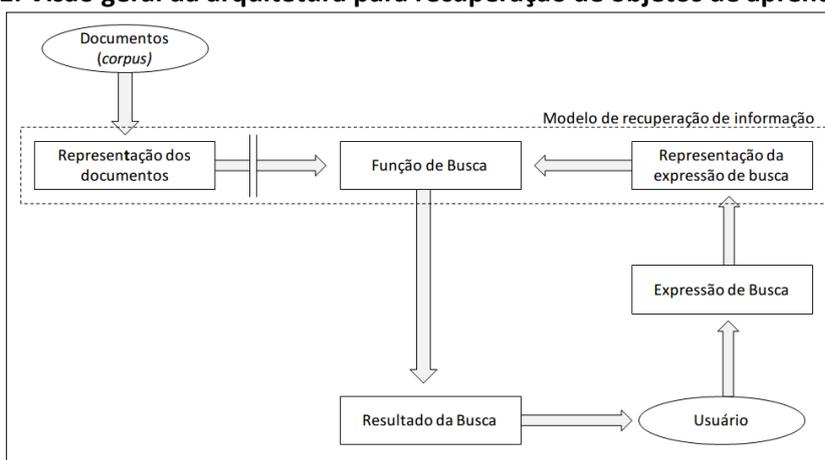
O propósito fundamental da arquitetura para recuperação de metadados de objetos de aprendizagem é unificar a representação desses recursos educacionais que são descritos por diferentes padrões de metadados e disponibilizados em repositórios específicos para este fim. Com isso, é possível que o usuário realize buscas qualificadas e encontre os OA mais adequados para sua necessidade de informação.

Todo o processo de RI tem por premissa básica descobrir itens de um *corpus* para apresentar ao usuário os resultados que mais se aproximam de sua necessidade de informação. Essa necessidade é formalizada por meio de uma função de busca. A representação do processo de RI proposto neste trabalho é baseado no modelo recomendado por Fereda (2012, p.14), conforme Figura 1. Neste modelo, são consideradas as seguintes etapas: representação dos documentos, a representação da expressão de busca e a função de busca.

No caso desta investigação, o *corpus* documental (representação dos documentos) é composto pelos registros de metadados que foram recuperados dos repositórios de objetos

de aprendizagem. Os metadados são recuperados por meio de um *crawler* parametrizado que extrai as URLs válidas (aquelas que apresentam informações de metadados) em um arquivo de *log*. A partir da lista de URLs válidas, retornada pelo *crawler*, identifica-se o padrão de metadados e os metadados são extraídos. Verifica-se a existência do objeto de aprendizagem, caso o mesmo não esteja persistido, executa-se a correlação e armazena as informações dos metadados que são persistidos em um modelo abstrato de metadados padrão XML.

Figura 1: Visão geral da arquitetura para recuperação de objetos de aprendizagem.



Fonte: adaptado de Ferneda (2012. p.14).

Para que o usuário se comunique com o sistema de recuperação de informação é necessário que ele acesse uma interface de busca e demonstre sua necessidade informacional, por meio da inserção de uma expressão de busca que será submetida ao motor de recuperação de informação para localizar os itens que sejam do provável interesse do usuário.

A representação da expressão de busca está relacionada com a forma como o usuário comunica sua necessidade informacional. Essa necessidade deve ser expressa de forma similar ao da representação dos documentos que compõe o *corpus* que, neste caso, ambos são demonstrados em linguagem natural. Quanto mais homogênea for à expressão de busca com os documentos do *corpus* melhor serão os resultados retornados ao usuário.

Assim sendo, o usuário inicia o processo de comunicação no momento em que informa uma expressão de busca para que o SRI processe a requisição e devolva um conjunto de itens recuperados com base na similaridade entre a expressão de busca informada e os registros de metadados.

Como a arquitetura para recuperar objetos de aprendizagem se baseia no modelo de recuperação de informação probabilístico, parte-se do entendimento de que no início do

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

processo de recuperação de informação o *corpus* de documentos relevantes é desconhecido. Portanto a probabilidade da expressão de busca informada pelo usuário estar presente ou não nos itens a serem recuperados é a mesma.

A partir da análise dos itens recuperados o usuário pode informar ao SRI quais itens são relevantes, em sua concepção, e o sistema de recuperação de informação recalcula os valores de relevância presumida e apresenta um novo ranqueamento com base no *feedback* do usuário com vistas a adequar os resultados à necessidade de informação do usuário.

A função de busca que está situada no centro do processo de recuperação de informação. Tem por finalidade comparar o grau de similaridade entre a expressão de busca informada pelo usuário e os elementos de metadados de objetos de aprendizagem que foram recuperados dos repositórios.

Para ranquear os itens recuperados pelo grau de similaridade, neste trabalho, optou-se em atribuir pesos aos elementos de metadados, conforme pode ser observado na Tabela 1. Os pesos foram distribuídos a partir de uma análise de relevância dos elementos de metadados. Foi atribuído um peso maior aos três primeiros elementos, pois estes estão diretamente relacionados com a descrição do conteúdo informacional do OA, fator este, que contribui para a recuperação mais relevante da informação, uma vez que existe a necessidade de ranqueamento dos metadados de OA.

Tabela 1 – Definição de pesos aos elementos de metadados.

ELEMENTO DE METADADOS	PESO
arq_ROA:title	1
arq_ROA:subject	0.8
arq_ROA:description	0.6
arq_ROA:types	0.2
arq_ROA:source	0.4
arq_ROA:relation	0.2
arq_ROA:coverage	0.4
arq_ROA:creator	0.4
arq_ROA:publisher	0.4
arq_ROA:contributor	0.4
arq_ROA:rights	0.2
arq_ROA:date	0.2

arq_ROA:identifier	0.0
arq_ROA:language	0.4

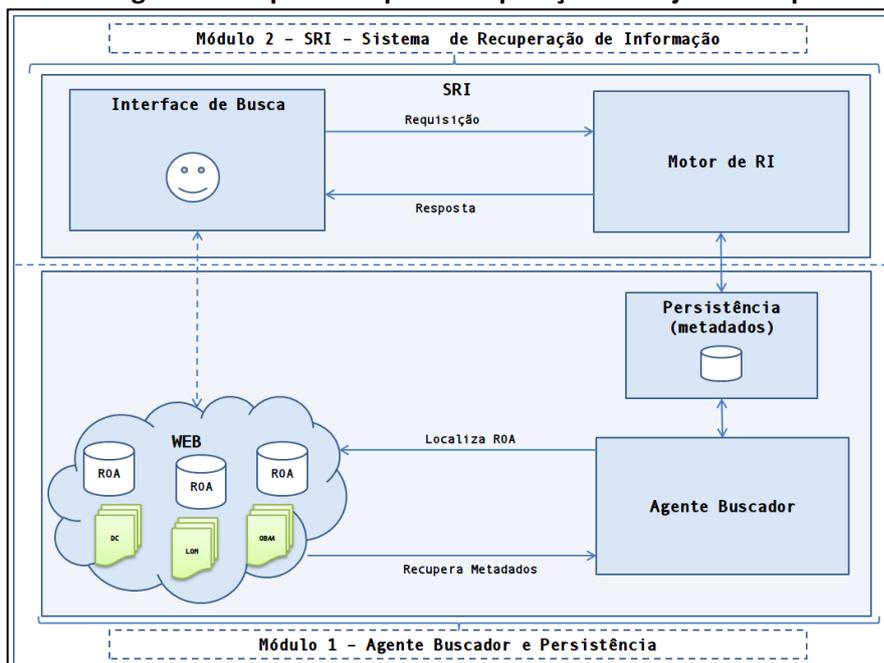
Fonte: elaboração própria.

Com base nos pesos estabelecidos é calculado o grau de relevância para cada item recuperado. Desta maneira, é possível ranquear os itens recuperados em ordem decrescente de uma relevância presumida.

4.1 Funcionamento da arquitetura

Para compreender o funcionamento desta arquitetura é apresentado o modelo desenvolvido ao longo deste trabalho, conforme Figura 2. A arquitetura foi projetada, em dois módulos: o primeiro é denominado “Agente Buscador e Persistência” dos metadados de OA que foram recuperados a partir dos repositórios. O segundo módulo é o “Sistema de Recuperação de Informação” que engloba a interface com o usuário e o motor de recuperação de informação. Este módulo é responsável pela comunicação do usuário com o SRI.

Figura 2: Visão geral da arquitetura para recuperação de objetos de aprendizagem.



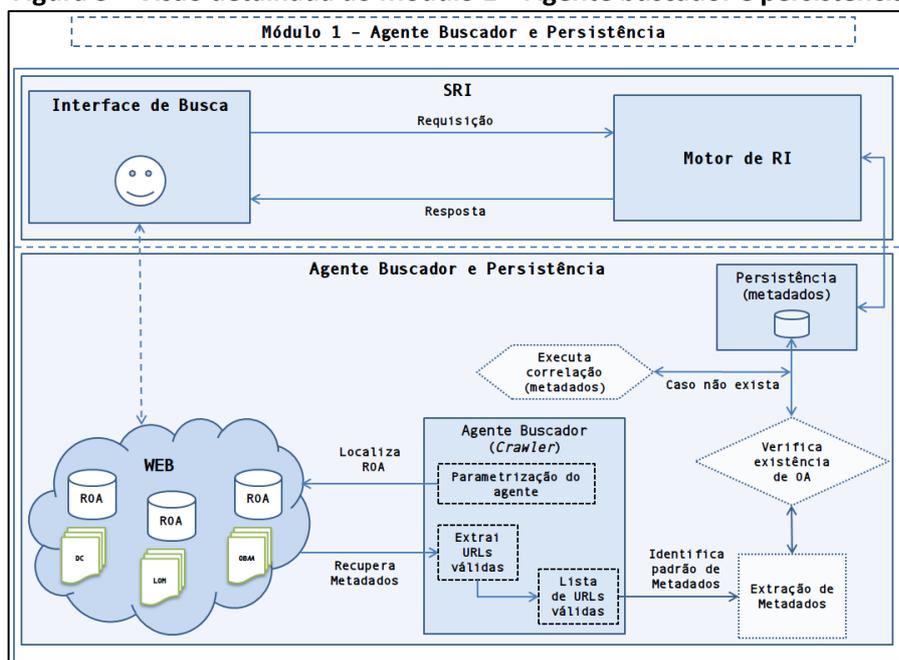
Fonte: elaboração própria.

Observa-se, na Figura 2, que os dois módulos que compõem a arquitetura interagem por meio dos metadados que foram persistidos quando da recuperação pelo agente buscador. Nesta arquitetura, pode-se considerar que o módulo 2 é executado *on-line*, ou seja, no

momento em que o usuário solicitar a busca, enquanto o módulo 1 pode ser executado *off-line*, isto é, em um tempo distinto ao da demanda do usuário.

O módulo 1 tem como objetivo buscar repositórios de objetos de aprendizagem e localizar os recursos informacionais por meio da identificação de seu padrão de metadados. Atua, em um primeiro momento, como um *crawler* baseado em agentes que é parametrizado para descobrir, buscar e extrair informações sobre os metadados de objetos de aprendizagem nos repositórios da *Web*. A Figura 3 apresenta a arquitetura proposta para esse módulo.

Figura 3 – Visão detalhada do Módulo 1 – Agente buscador e persistência



Fonte: elaboração própria.

O agente buscador de OA é parametrizado com informações referentes aos padrões de metadados, ao repositório e a quantidade máxima de URLs a serem recuperadas. Após esta parametrização, inicia-se a busca pelas URLs que contêm os metadados de objetos de aprendizagem. Essa busca é realizada pelo cálculo de similaridade entre as informações de metadados e o conteúdo da URL.

A extração dos metadados é realizada por meio de um *parsing* sobre o HTML recuperado. Para auxiliar no processo de extração foi utilizado o HTML *parser* JSOUP (JSOUP, 2016) que é uma biblioteca Java para manipulação e extração de informação no formato HTML. Optou-se em extrair os metadados diretamente do HTML, uma vez que nem todos os ROA disponibilizam seus metadados em XML. Os metadados extraídos são persistidos,

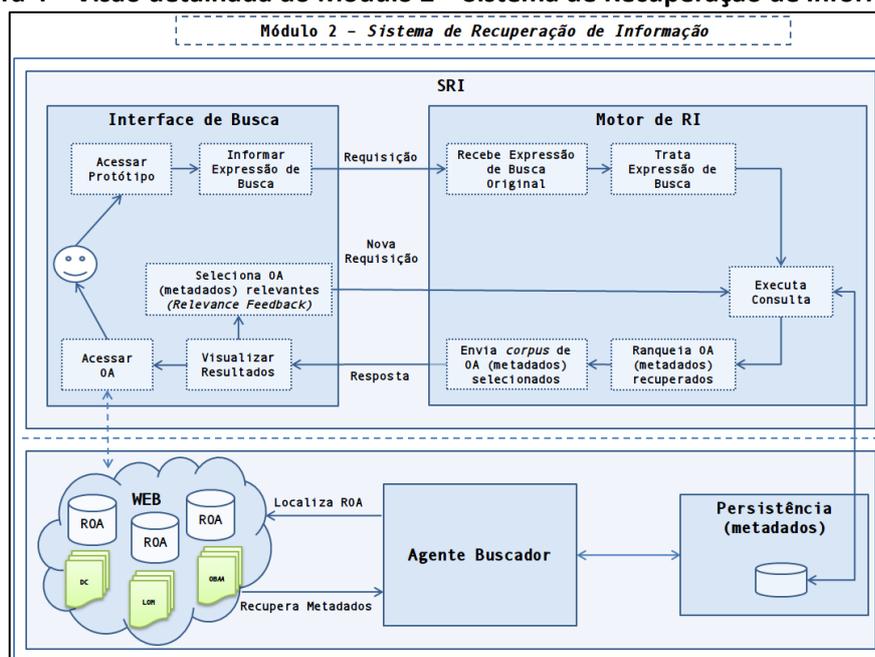
inicialmente, em um modelo de metadados (padrão XML) que possuem propriedades e valores associados a cada elemento da correlação de metadados, conforme Tabela 1.

O Módulo 2 (Figura 4) refere-se ao processo de recuperação de informação. O usuário acessa a interface de recuperação de objetos de aprendizagem e traduz sua necessidade de informação por meio de uma expressão de busca. O usuário submete sua requisição ao motor de recuperação de informação que, por sua vez, retorna um conjunto de itens recuperados e ranqueados.

O usuário pode analisar e avaliar os itens recebidos e selecionar (a partir das informações que lhe são apresentadas, como por exemplo: título, palavras-chave, autor, formato etc.) os que lhe parecem como mais relevantes naquele momento. A partir desta seleção, o usuário faz uma nova requisição ao motor de recuperação de informação baseada nos itens que ele julgou como relevantes. Este processo dialógico entre o usuário e o sistema de recuperação de informação é denominado de *relevance feedback*.

O motor de RI recebe a informação, calcula uma nova representação de sua necessidade de informação e retorna novos resultados baseado em melhores taxas de recuperação. Esse processo é cíclico e pode ser executado inúmeras vezes até que o usuário esteja satisfeito com a resposta que lhe foi retornada. A Figura 4 apresenta uma visão detalhada das fases desse processo.

Figura 4 – Visão detalhada do Módulo 2 – Sistema de Recuperação de Informação



Fonte: elaboração própria.

Ressalta-se que o processo de *relevance feedback* não pode ser aplicado na formulação da expressão de busca inicial do usuário, uma vez que não existe nenhum resultado disponível para indicar a relevância. O sucesso deste processo está diretamente relacionado com a espontaneidade do usuário em informar seu posicionamento em relação ao conjunto de itens que lhe foi apresentado.

De acordo com Jin e Hauptmann (2001), essa interação do usuário com o SRI possibilita uma RI mais eficiente. Os trabalhos de Lavrenko e Croft (2001), Robertson e Jones (1976), Rocchio (1971), Salton e Buckley (1990) e Zhai e Lafferty (2001) comprovaram que o processo de RF é eficiente e incrementa precisão do processo de recuperação de informação. Assim, esta técnica visa melhorar a efetividade na recuperação de informação baseado na avaliação de relevância dos documentos pelo usuário.

Portanto, o principal objetivo desta arquitetura foi de disponibilizar um modelo capaz de permitir que um usuário obtenha um *corpus* de OA (por meio de metadados) mais significativo para sua necessidade de informação e espera-se que a partir da interação do usuário com o SRI seja possível melhorar a relevância do retorno das buscas para que o usuário possa acessar o recurso a partir do repositório no qual ele está sendo mantido.

5 CONSIDERAÇÕES FINAIS

Percebeu-se, no decorrer desta investigação, que os desafios inerentes à recuperação de informação relevante na *Web* são constantes e incrementados pelas características intrínsecas deste ambiente, tais como a dinamicidade e a volatilidade das informações. No âmbito educacional estas características também são observadas e acrescidas da complexidade e das peculiaridades dos objetos de aprendizagem que são idealizados para atender a diferentes objetivos pedagógicos e desenvolvidos em diversos tipos de formatos.

A arquitetura para recuperação de metadados de objetos de aprendizagem permitiu unificar recursos educacionais descritos em diferentes padrões de metadados e mantidos por diferentes repositórios. Os agentes desempenharam um papel fundamental na descoberta de informação disponível na *Web* e potencializaram a busca por informações estruturadas (metadados) em diferentes repositórios de objetos de aprendizagem. O modelo de *crawler* proposto baseado em agentes inteligentes foi capaz de descobrir, recuperar e mapear objetos de aprendizagem armazenados em repositórios que descrevem seus recursos por diferentes padrões de metadados.

O principal diferencial desta proposta foi de valorizar a inferência do usuário no processo de recuperação de informação, por meio do processo de *relevance feedback*. Neste processo, o usuário estabelece um diálogo com o sistema de recuperação de informação realizando refinamentos nos resultados que lhe foram retornados. Como esse processo é cíclico, ele pode ser executado até que o usuário esteja satisfeito com os resultados que lhe foram retornados.

A arquitetura proposta neste trabalho não foi implementada na íntegra, apenas o módulo referente ao agente buscador foi desenvolvido. Esta limitação possibilita visualizar perspectivas futuras de continuidade desta investigação, com o desenvolvimento de um protótipo para validar a arquitetura proposta. A partir do protótipo concluído será possível propor um sistema de recomendação de objetos de aprendizagem, baseado no conceito de relevância do usuário.

Conclui-se, que um sistema de recuperação de informação é mais eficiente quando amplia seu escopo de recuperação a partir de diferentes fontes de dados (ROA, no caso deste trabalho) e permite a inferência do usuário no julgamento da informação que lhe foi retornada. Assim, acredita-se que um sistema de recuperação de informação retorna resultados mais relevantes quando possibilita estabelecer um processo dialógico entre o usuário e o sistema de recuperação de informação.

REFERÊNCIAS

AFONSO, Maria da Conceição Lima *et al.* **Banco Internacional de Objetos Educacionais (BIOE):** tratamento da informação em um repositório educacional digital. *Perspect. Ciênc. Inf.*, Belo Horizonte, v. 16, n. 3, p.148-158, set. 2011.

ARIADNE. The Alliance of Remote Instructional Authoring and Distribution Networks for Europe. Disponível em: <<http://www.ariadne-eu.org>>. Acesso em: 24 ago. 2015.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação de Informação:** conceitos e tecnologia das máquinas de busca. 2. Ed. – Porto Alegre: Bookman, 2013.

BIANCO, A. M.; DE MARSICO, M.; TEMPERINI, M. **Standards for e-learning.** Project QUIS: (QUality, Interoperability and Standards in e-learning). The TISIP Research Foundation, Trondheim, 2005. ISBN 82-8055-018-6.

BUCKLEY, Christopher; SALTON, Gerard Optimization of Relevance Feedback Weights. In: **Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.** Washington: USA. 1995. p.9-13.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

CHOWDHURY, Gobinda. **Introduction to Modern Information Retrieval**. 3rd ed, Facet Publishin, 2010.

DCMI. Dublin Core Metadata Initiative. Disponível em: <<http://dublincore.org/>>. Acesso em: 15 dez. 2016.

FERNEDA, Edberto. **Introdução aos Modelos Computacionais de Recuperação de Informação**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2012.

HARTER, Stephen P. **Online Information Retrieval: Concepts, principles and techniques**. Orlando: Academic Press. Inc., 1986. 259 p.

IEEE LTSC LOM. IEEE Learning Technology Standards Committee (LTSC) – Standard for Learning Object Metadata (LOM). 2002. Disponível em: <http://grouper.ieee.org/groups/ltsc/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf>. Acesso em: 25 ago. 2015.

IMS LRM. Disponível em: <https://www.imsglobal.org/> – Acesso em: 27 jul. 2015.

IP, Albert; MORRISON, Iain; CURRIE, Mike. **What is a learning object, technically?** In: WebNet, pp 580–586. 2001. Disponível em: <[http://users.tpg.com.au/adslfrcf/lo/learningObject\(WebNet2001\).pdf](http://users.tpg.com.au/adslfrcf/lo/learningObject(WebNet2001).pdf)>. Acesso em: 28 jul. 2017.

ISO MLR. ISO/IEC 19788. 2014. Disponível em: <<https://www.iso.org/obp/ui/#iso:std:iso-iec:19788:-4:ed-1:v1:en>>. Acesso em: 05 set. 2015.

JIN, Rong; HAUPTMANN, Alexander G. **A New Probabilistic Model for Title Generation**. Carnegie Mellon University, Pittsburgh, USA, 2001.

JSOUP. Java HTML Parser. Disponível em: <<http://jsoup.org/>>. Acesso em: 30 set. 2016

LAVRENKO, Victor; CROFT, W. Bruce. Relevance based language models. In: **Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval**, SIGIR '01, New York, NY, USA: ACM, 2001. Disponível em: <<http://doi.acm.org/10.1145/383952.383972>>. Acesso em: 02 out. 2016.

MARON, Melvin Earl; KUHNS, John Lary. **On relevance, probabilistic indexing and information retrieval**. Journal of the ACM, v. 7, n. 3, 1960, p.216-244.

MEADOW, Charles T. *et al.* **Text Information Retrieval Systems**. 3rded. London UK: Elsevier, 2007.

MOHAN, Permanand. Learning Objects Repositories. In: KOOHANG, Alex.; HARMAN, Keith **Learning Objects: theory, praxis, issues, and trends**. Santa Rosa: Informing Science Press, p.527-548, 2007.

ROBERTSON, Stephen; JONES, Karen Sparck. Relevance weighting of search terms. **Journal of the American Society for Information Science**, v. 27, n. 3, p.129-146, 1976.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

ROCCHIO, Joseph Junior. Relevance feedback in information retrieval. In G. Salton (Ed.), **The Smart Retrieval System** – Experiments in Automatic Document Processing. Englewood Cliffs, NJ, USA: Prentice Hall, Inc. p.313 – 323, 1971.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**. Segunda Edição, Ed. Campus, 2004.

SALTON, Gerard. **The SMART Retrieval System**: experiments in automatic document processing. Englewood Cliffs, NJ: Prentice Hall, 1971.

SALTON, Gerard.; BUCKLEY, Christopher. Improving Retrieval Performance by Relevance Feedback. In K. Sparck Jones & P. Willet (Eds.), **Readings in Information Retrieval**. San Francisco, California: Morgan Kauffmann, 1997.

SALTON, Gerard.; MCGILL, Michael. **Introduction to Modern Information Retrieval**. New York, McGraw-Hill, 1983.

SANTARÉM 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.

VICCARI, R.; GLUZ, J.; SANTOS, E.; et al. **Projeto OBAA Relatório Técnico RTOBAA01** – Proposta de Padrão para Metadados de Objetos de Aprendizagem Multiplataforma. UFRGS/CINTED, 2009. Disponível em:
<<http://www.portalobaa.org/obaac/padraoobaa/relatoriostecnicos/RTOBAA01.pdf/view>>. Acesso em: 28 nov. 2016.

WILEY, David. **Connecting learning objects to instructional design theory**: A definition, a metaphor, and a taxonomy. The Instructional Use of Learning Objects. Wiley, D. (ed.) 2001. Disponível em: <<http://www.reusability.org/read/chapters/wiley.doc>>. Acesso em: 21 jul. 2017.

ZHAI, Chengxiang; LAFFERTY, John. Model-based feedback in the language modeling approach to information retrieval. In: **Proceedings of CIKM '01**, 2001, p.403–410. Disponível em: <<http://sifaka.cs.uiuc.edu/czhai/pub/cikm2001-fb.pdf>>. Acesso em: 20 out. 2016.