

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA  
E TECNOLOGIA DO RIO GRANDE DO SUL  
CÂMPUS BENTO GONÇALVES

SISTEMA DE RECOMENDAÇÃO  
PARA OBJETOS DE APRENDIZAGEM

MÁRCIO BORTOLINI DOS SANTOS

Bento Gonçalves, Dezembro de 2013.

MÁRCIO BORTOLINI DOS SANTOS

SISTEMA DE RECOMENDAÇÃO  
PARA OBJETOS DE APRENDIZAGEM

Monografia apresentada junto ao Curso de Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – Câmpus Bento Gonçalves, como requisito parcial à obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Orientadora: Dra. Júlia Marques Carvalho da Silva

Bento Gonçalves, Dezembro de 2013.

## RESUMO

Através da Computação surgiram novas tecnologias que dão apoio ao ensino e facilitam a aprendizagem online, uma delas são os objetos de aprendizagem, que são definidos como sendo quaisquer materiais que tenham objetivo didático. Estes se encontram armazenados em repositórios que, além de armazená-los, provêm métodos de recuperação e indexação dos mesmos. Dentre estes métodos de recuperação, se enquadram os sistemas de recomendação. Neste contexto, o presente trabalho tem por finalidade apresentar um projeto de pesquisa de como os sistemas de recomendação facilitam a localização dos objetos em um repositório, mais especificamente, no que está sendo construído pelo o IFRS-BG. Para tal, foi feito um levantamento bibliográfico sobre técnicas e abordagens de recomendação de conteúdo, objetos de aprendizagem, padrões de metadados e tecnologias envolvidas no encapsulamento dos objetos, bem como tecnologias que serão necessárias para o processo de desenvolvimento do sistema de recomendação que será aplicado no repositório mencionado e analisado posteriormente.

**Palavras-chave:** Sistemas de Recomendação, Objetos de Aprendizagem, Repositório de Objetos de Aprendizagem

## **ABSTRACT**

Through technological advances emerged new technologies that support teaching and facilitate student's learning. Among these technologies are the learning objects, which are defined as being any content that aims to teach something. Learning objects are stored in repositories, that also provide methods to retrieve and index them. Among these methods are the recommender systems. In this context, a survey on the literature of techniques and approaches to recommend contents, learning objects, metadata standards and technologies involved in the encapsulation of objects, as well as the technologies that were needed to the development of the recommender system, was made. The recommender algorithm was applied on the repository mentioned earlier and analysed afterwards. Thus, this paper aimed to analyze how recommender systems facilitate users to find objects in a repository, more specifically, in the repository that is being built by IFRS-BG.

**Keywords:** Recommender Systems, Learning Objects, Learning Objects Repositories.

*O maior desafio de qualquer pensador é  
enunciar o problema de tal modo que possa  
permitir uma solução.*

Bertrand Russel

## LISTA DE FIGURAS

Figura 1: Página inicial do MERLOT.....	11
Figura 2: Página inicial do BIOE.....	12
Figura 3: Listas de recomendação do site Last.fm.....	14
Figura 4: Listas de recomendação do site Skoob.....	14
Figura 5: Exemplo de avaliação dos usuários no site da Submarino.....	15
Figura 6: Disponibilização das avaliações dos usuários no site da Submarino.....	16
Figura 7: Exemplo de recomendação personalizada no site do Twitter.....	16
Figura 8: Exemplo de recomendação por relação entre itens vistos pelos usuários no site da Americanas.....	17
Figura 9: Exemplo de recomendação por relação entre itens vistos pelos usuários no site da Casas Bahia.....	17
Figura 10: Exemplo de recomendação por conteúdo no site da Saraiva.....	18
Figura 11: Características herdadas pela Filtragem Híbrida.....	23
Figura 12: Representação gráfica do resultado da expressão booleana (“modelos” and “booleanos”) or “conjuntos”.....	24
Figura 13: Espaço vetorial tri-dimensional.....	27
Figura 14: Interface de relação entre objetos de aprendizagem.....	31
Figura 15: Objetos de Aprendizagem recomendados pelo LorSys.....	32
Figura 16: Modelo Lógico das tabelas criadas.....	37
Figura 17: Tabelas utilizadas para criar a view com os campos que serão considerados no cálculo de similaridade.....	37
Figura 18: Diagrama de Classes.....	38
Figura 19: Passos para Extração da Informação.....	39
Figura 20: Processo de extração da informação.....	40
Figura 21: Passos do Stemmer para a língua portuguesa.....	44
Figura 22: Exemplo de Índice Invertido.....	46
Figura 23: Exemplo de índice sequencial.....	47
Figura 24: Lista circular duplamente encadeada dos tipos de recurso educacional.....	53
Figura 25: Lista de recomendações.....	54
Figura 26: Página inicial do Sistema de Tarefas.....	55
Figura 27: Tela para avaliação das recomendações geradas pelo sistema.....	60
Figura 28: Repositório 1 – sem as recomendações.....	61
Figura 29: Repositório 2 – com as recomendações.....	62
Figura 30: Formulário para informar os objetos de aprendizagem encontrados na busca.....	64

# SUMÁRIO

<b>1 CONTEXTUALIZAÇÃO.....</b>	<b>1</b>
1.1 PROBLEMA.....	3
1.2 OBJETIVOS.....	3
<b>1.2.1 Objetivo Geral.....</b>	<b>3</b>
<b>1.2.2 Objetivos específicos.....</b>	<b>3</b>
1.3 JUSTIFICATIVA.....	3
1.4 METODOLOGIA.....	3
<b>2 REVISÃO BIBLIOGRÁFICA.....</b>	<b>7</b>
2.1 Objetos de Aprendizagem.....	7
<b>2.1.1 Metadados.....</b>	<b>7</b>
<b>2.1.2 Repositórios.....</b>	<b>10</b>
2.2 Sistemas de Recomendação.....	13
<b>2.2.1 Estratégias de Recomendação.....</b>	<b>13</b>
2.2.1.1 Listas de recomendação.....	13
2.2.1.2 Avaliações de usuários.....	15
2.2.1.3 Suas recomendações.....	16
2.2.1.4 “Usuários que se interessaram por X também se interessaram por Y”.....	16
2.2.1.5 Associação por conteúdo.....	17
<b>2.2.2 Técnicas de Recomendação.....</b>	<b>18</b>
2.2.2.1 Baseados em conteúdo.....	18
2.2.2.2 Colaborativos.....	19
2.2.2.3 Híbridos.....	22
2.3 Modelos para recuperação de informação para sistemas de recomendação baseados em conteúdo.....	23
<b>2.3.1 Modelo booleano.....</b>	<b>23</b>
<b>2.3.2 Modelo de Espaço Vetorial (VSM) ou Modelo Vetorial.....</b>	<b>25</b>
<b>2.3.3 Modelo Probabilístico.....</b>	<b>28</b>
2.4 Trabalhos similares.....	29
<b>2.4.1 LorSys.....</b>	<b>30</b>
<b>2.4.2 DICA.....</b>	<b>32</b>
<b>2.4.3 Características consideradas para o presente trabalho.....</b>	<b>33</b>
<b>3 PROJETO.....</b>	<b>35</b>
3.1 Levantamento de requisitos.....	35
<b>3.1.1 Requisitos funcionais.....</b>	<b>35</b>
<b>3.1.2 Requisitos não funcionais.....</b>	<b>35</b>
<b>3.1.3 Regras de negócio.....</b>	<b>36</b>
3.2 Modelagem do sistema.....	36
<b>3.2.1 Modelo Lógico do Banco de Dados.....</b>	<b>36</b>
<b>3.2.2 Diagrama de classes.....</b>	<b>38</b>
3.3 Desenvolvimento.....	39
<b>3.3.1 Extração da informação.....</b>	<b>39</b>
3.3.1.1 Coleta.....	40
3.3.1.2 Pré-processamento.....	41
3.3.1.2.1 Normalização.....	41
3.3.1.2.2 Stop Words.....	41
3.3.1.2.3 Stemming ou Normalização Morfológica.....	42
3.3.1.3 Indexação.....	46
3.3.1.4 Mineração.....	47

<b>3.3.2 Cálculo de Similaridade.....</b>	<b>48</b>
3.3.2.1 <i>TextFieldsSimilarity</i> .....	48
3.3.2.2 <i>DistanceFieldsSimilarity</i> .....	50
<b>3.3.3 Visualização da recomendação.....</b>	<b>54</b>
<b>4 RESULTADOS.....</b>	<b>55</b>
4.1 Perfil dos sujeitos.....	55
4.2 Avaliação das recomendações geradas pelo sistema.....	59
4.3 Tarefas de busca.....	61
4.3.1 Tarefa de busca no repositório sem as recomendações.....	64
4.3.2 Tarefa de busca no repositório com as recomendações.....	66
4.3.3 Comparação entre as tarefas de busca.....	68
4.3.4 Considerações sobre a aplicação do experimento.....	68
<b>5 CONCLUSÕES.....</b>	<b>70</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>72</b>
Apêndice A - Cronograma.....	76
Apêndice B – Diagrama de Classes.....	77



## 1 CONTEXTUALIZAÇÃO

A Computação visa desenvolver tecnologias que permitam ser aplicadas em diversas áreas, como: saúde, engenharia e educação. Nesta última, em especial, é possível fornecer desde ambientes virtuais que oferecem ferramentas de apoio para aproximam o cotidiano da sala de aula, como softwares educacionais com o intuito de simular ou reforçar a aprendizagem. Ainda, dentro deste cenário, um dos conceitos abordados por pesquisas (IEEE LTSC, 2013; SILVA, BAVARESCO e SILVEIRA, 2008; McGREAL, 2004), são os objetos de aprendizagem (OAs).

Sob o ponto de vista pedagógico, os objetos de aprendizagem consistem em recursos utilizados para fins didáticos (SILVA e VICARI, 2010). Ou seja, qualquer que seja o material, digital ou não, desde que aplicado no processo de ensino-aprendizagem, pode ser considerado um objeto de aprendizagem. Sendo assim, podem ser considerados como sendo exemplos destes: textos, figuras, músicas e filmes. Entretanto, sob o ponto de vista computacional, tal recurso precisa conter um arquivo que descreva suas características. Essas características são denominadas metadados (Vicari *et al*, 2009).

Tal arquivo torna-se fundamental para que se conheça o conteúdo inserido no objeto de aprendizagem, e permita responder a algumas perguntas: “Como posso utilizá-lo em minha aula? Para qual nível de ensino e disciplina? É necessário algum recurso tecnológico específico para a sua execução?”. Estas informações são descritas pelos metadados, de forma com que tanto o utilizador (professor) conheça previamente o conteúdo, sem que necessite explorá-lo completamente; quanto que sistemas computacionais (repositórios) possam indexar e recuperá-lo. Ainda, atualmente existem algumas entidades as quais normatizam o conjunto de informações que um arquivo de metadados para objetos de aprendizagem deve conter, como por exemplo: IEEE LOM (2013), DublinCore (2013) e OBAA (2013).

Alguns exemplos de repositórios são: *Multimedia Educacional Resource for Learning and Online Teaching* (MERLOT, 2013), Banco Internacional de Objetos de Aprendizagem (BIOE, 2013), entre outros. Esses repositórios oferecem opção de visualizar os objetos de aprendizagem por categorias (nível de ensino, tecnologia, etc) ou realizar buscas diretas (por palavras-chave, título, descrição). Entretanto, é possível encontrar exemplos de pesquisas que visam desenvolver sistemas que também permitem a recomendação de conteúdos, de forma

personalizada, como o *Learning Object Recommender System* (LorSys) (Ferreira, 2009), *E-Learning Object Recommendation System* (E-LORS)(Vieira, 2012) e o DICA (Zaina *et al.*, 2012), os quais utilizam técnicas baseadas em recomendação de conteúdo.

Os sistemas de recomendação oferecem uma maneira inteligente e personalizada para encontrar informações fácil e rapidamente (Bradley e Smyth, 2001). Esses sistemas, segundo Basu *et al.* (1998), dão sugestões de artefatos para um usuário, utilizando-se de várias abordagens para que tais sugestões correspondam aos interesses do usuário, além de garantir que a informação seja disponibilizada de forma rápida e efetiva.

O Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - Câmpus Bento Gonçalves (IFRS-BG) está construindo um Repositório de Objetos de Aprendizagem (ROA), utilizando o padrão de metadados OBAA. O ROA permitirá, aos discentes e docentes, encontrar OAs, através de uma busca simples, e filtrar os resultados por tipo de conteúdo e nível de ensino a ser aplicado. Porém verificou-se que esse repositório não possui algoritmo de recomendação e não aplica nenhuma técnica para recomendação de conteúdos. Deste modo o presente trabalho objetivou aplicar um algoritmo de recomendação do ROA do IFRS e avaliar como a aplicação do algoritmo pode facilitar que os usuários encontrem OAs no referido repositório.

Para a construção do algoritmo recomendação, optou-se por utilizar a abordagem de recomendação baseada em conteúdo, pois os dados armazenados para a recomendação são somente sobre as descrições do OA, não levando em conta dados que possam ser utilizados para a abordagem colaborativa, e para recuperação dos dados, será utilizado o modelo espaço vetorial, pois este considera a quantidade de vezes que um termo aparece em um documento, diferente do probabilístico, e, além disso, permite recomendações mais avançadas do que o booleano, tais abordagens e técnicas são explicados na Seção 2.2.

Os metadados dos OAs utilizados para a geração das recomendações foram: palavras chave; descrição; título; cobertura; formato, descrição educacional e tipo de recurso educacional, sendo estes pertencentes a diferentes categorias do padrão OBAA. A partir destes metadados, quando um usuário seleciona um OA no repositório, serão recomendados outros objetos similares, presentes em uma lista gerada pelo algoritmo que foi desenvolvido neste trabalho. Sendo assim, o presente trabalho pertence as áreas de pesquisa de informática na educação e inteligência artificial.

## **1.1 PROBLEMA**

A partir dos metadados, serão geradas as listas de objetos similares, para que quando um usuário acesse um objeto qualquer, estas sejam disponibilizadas para ele. Neste contexto, de que forma as recomendações podem facilitar que os usuários encontrem OAs no ROA do IFRS?

## **1.2 OBJETIVOS**

Nessa seção são apresentados os objetivos geral e específicos do presente trabalho.

### **1.2.1 Objetivo Geral**

Avaliar como a aplicação de um algoritmo de recomendação de conteúdo no ROA do IFRS-BG facilita para que os usuários encontrem OAs.

### **1.2.2 Objetivos específicos**

- Investigar o funcionamento do ROA;
- Elaborar um algoritmo de recomendação de OAs para o ROA;
- Testar como as recomendações facilitam que os usuários encontrem OAs;

## **1.3 JUSTIFICATIVA**

Os sistemas de recomendação tem sido amplamente aplicados em sites de comércio eletrônico, redes sociais, dentre outros, isto demonstra a importância de sua aplicação e utilização. Além disso, os estudos sobre estes sistemas têm se abrangido cada vez mais a medida que novas técnicas e abordagens são criadas e utilizadas. Desta forma, o tema do presente trabalho se mostra relevante em sua aplicação prática, além de sua importância pelas áreas de pesquisa em que se encontra (informática na educação e inteligência artificial). O que leva a necessidade de maiores estudos na área, justificando a adequação deste trabalho em um projeto de pesquisa para um Trabalho de Conclusão de Curso (TCC).

## **1.4 METODOLOGIA**

Neste trabalho foi utilizada a abordagem de pesquisa exploratória. A pesquisa

exploratória foi escolhida, pois proporciona a primeira aproximação de um pesquisador com o tema da pesquisa e o ajuda a obter informações sobre o mesmo. Além disto, serve como uma ferramenta para auxílio na delimitação do tema e na definição de seus objetivos. Por fim, esta abordagem o ajuda a formular suas hipóteses de trabalho e também a descobrir uma forma original de desenvolver seu assunto, através da identificação de um novo aspecto do tema da pesquisa. A pesquisa exploratória pode ser construída por meio de levantamento bibliográfico, entrevistas, observações e análise de exemplos sobre o tema estudado. (Corrêa, 2003; Reis, 2008).

Dentro da pesquisa exploratória, foi utilizado o conceito de estudo de caso, que consiste na seleção de um objeto de pesquisa, o qual pode ser um fato ou fenômeno estudado em seus vários aspectos. Neste tipo de pesquisa, o pesquisador e o participante representante da situação-problema cooperam mutuamente com o estudo (Reis, 2008). Portanto, o estudo de caso foi escolhido objetivando uma melhor compreensão da forma de utilização de repositórios e para ajudar na avaliação da solução desenvolvida para a disponibilização das recomendações para o usuário.

Os participantes do estudo de caso fazem parte da comunidade escolar do IFRS-BG. Os testes foram feitos em diferentes grupos, onde foram explicados conceitos relevantes sobre sistemas de recomendação, objetos de aprendizagem e repositórios, com o intuito de familiarizá-los com o objeto desta pesquisa. Os participantes atuam em diferentes áreas e, portanto, possuem diferentes conhecimentos e níveis de experiência sobre o uso de um computador e, mais especificamente, internet. Após, os participantes acessaram um sistema que foi feito para execução dos testes, que será chamado aqui de Sistema de Tarefas, no qual, eles primeiramente preencheram um formulário contendo informações sobre o quão familiarizados os mesmos estão quanto ao uso do computador e da internet para execução de buscas e pesquisas.

Preenchido o formulário, os participantes avaliaram a precisão do algoritmo de recomendação. Para tal, no Sistema de Tarefas, lhes era mostrado um objeto de aprendizagem específico presente no repositório. Tal objeto fora selecionado de forma aleatória, de tal modo que diferentes objetos pudessem ser avaliados pelos participantes, e além do objeto de aprendizagem, eram listados cinco outros objetos de aprendizagem que o sistema recomendaria no repositório. Assim, os participantes tiveram que ler a descrição do primeiro

objeto de aprendizagem e compará-las com os da lista, e após, para cada objeto de aprendizagem recomendado, separadamente, responder se acreditavam que o mesmo tinha alguma relação ou relevância em relação ao primeiro.

Para a seleção dos objetos de aprendizagem recomendados, que os participantes iriam avaliar, foram selecionados os dez itens que o algoritmo de recomendação classificara como sendo os mais relevantes, e de forma aleatória. Além disso, o Sistema de Tarefas escolhia se seriam apresentados os itens em posições pares ou ímpares para os usuários. Deste modo ou seriam avaliados os itens nas posições 1, 3, 5, 7, 9 ou nas posições 2, 4, 6, 8, 10. Isto objetivou avaliar não somente os primeiros cinco itens da lista, e desta forma, caso fosse selecionado o mesmo objeto de aprendizagem para ser avaliado por dois ou mais participantes, haveria possibilidade de serem avaliados os 10 itens mais relevantes e não somente os cinco primeiros.

Após, através do Sistema de Tarefas, foram delegadas duas tarefas aos participantes, através de um roteiro previamente elaborado, contendo atividades relacionadas ao uso do repositório para execução de buscas por objetos de aprendizagem. Uma das tarefas seria executada no repositório que continha recomendações e a outra no repositório que não continha recomendações, o sistema selecionava de forma aleatória qual repositório seria utilizado na primeira tarefa e qual seria utilizado na segunda.

Para tais tarefas, foram selecionados três diferentes grupos de objetos de aprendizagem presentes no repositório, todos considerados como contendo alguma relação com os demais em seu grupo, e cada grupo era constituído de três objetos de aprendizagem.

O grupo utilizado para a tarefa também era selecionado de forma aleatória, dentre os três grupos definidos anteriormente, e de forma que os usuários não utilizassem mais o mesmo grupo para a próxima busca. Com isto objetivou-se diminuir a possibilidade de que todos os participantes obtivessem o mesmo grupo de objetos de aprendizagem para buscar no mesmo repositório (que contém ou não as recomendações). Além disto, acreditou-se que isto diminuiria a possibilidade de que os participantes executassem as duas tarefas da mesma forma, caso fosse realizada a mesma tarefa (busca pelos mesmos OAs) nos diferentes repositórios.

Cada busca foi executada separadamente, e após os usuários preenchiam um formulário sobre a dificuldade de execução de cada tarefa e para o repositório contendo as

recomendações, o quanto as recomendações ajudaram na execução da tarefa.

Durante a execução das duas tarefas, foi feita observação sistemática, porém de forma não participativa, ou seja, sem participação na execução das tarefas. Durante a observação, foram realizadas algumas anotações manuscritas de dúvidas e dificuldades encontradas. A observação sistemática serve para que o pesquisador possa criar o conhecimento científico do aspecto da realidade que ele deseja conhecer (Pádua, 2004). A observação permitiu compreender melhor como os usuários se comportam durante o uso do sistema.

A etapa final consistiu-se na análise dos resultados obtidos no experimento para verificação do cumprimento das metas propostas por este trabalho.

A presente pesquisa foi composta pelas seguintes etapas, seguindo o cronograma apresentado no APÊNDICE A:

1. **Revisão bibliográfica:** Revisão bibliográfica sobre sistemas de recomendação e definição da técnica que será utilizada para criação do algoritmo;
2. **Análise e projeto:** Análise e modelagem para o algoritmo de recomendação;
3. **Implementação:** Implementação do algoritmo de recomendação e testes;
4. **Integração:** Integração das recomendações geradas com o repositório;
5. **Roteiro de tarefas:** Definição roteiro de tarefas para os testes com os participantes;
6. **Questionários:** Definição dos questionários para coleta dos dados das tarefas;
7. **Oficina:** Execução das tarefas junto aos participantes;
8. **Tabulação dos dados:** Tabulação e análise dos dados provenientes das tarefas;
9. **Conclusões:** Análise dos resultados e conclusões;
10. **Escrita do TCC:** Escrita de todos os componentes envolvidos na pesquisa.

## 2 REVISÃO BIBLIOGRÁFICA

Nessa seção são apresentados os fundamentos teóricos relevantes para o desenvolvimento do presente trabalho. Dessa forma, a seção 2.1 trata dos conceitos relativos a OAs, metadados e repositórios. Por outro lado, os conceitos sobre sistemas de recomendação, abordagens para recomendação de conteúdos e trabalhos similares são apresentados nas seções 2.2, 2.3 e 2.4, respectivamente.

### 2.1 *Objetos de Aprendizagem*

Objetos de Aprendizagem são definidos como sendo qualquer entidade, digital ou não, que pode ser utilizada, reutilizada ou referenciada durante a aprendizagem apoiada pela tecnologia. Alguns exemplos de OAs incluem conteúdos multimídia (vídeos, áudios, imagens, etc), conteúdos instrutivos (como por exemplo textos), softwares ou ferramentas instrucionais (IEEE LTSC, 2013).

Um OA corresponde ao menor bloco de instrução ou informação, elaborado de forma independente, capaz de transmitir conhecimento. Eles são estruturados através da combinação de seus conteúdos didáticos, que podem ser: apostila, apresentação, tutorial, exercícios etc, e seus respectivos metadados educacionais, sendo estes: título, autor, faixa etária, etc. (Ghelman *et al.* 2007).

Os metadados são tipicamente identificados (ou descritos) através de etiquetas (*tags*) em *eXtensible Markup Language* (XML), que são definidas por um conjunto de padrões internacionais e diretrizes de especificação. OAs são igualmente catalogados através de *tags*, conforme os mesmos padrões e diretrizes, e armazenados em repositórios. As *tags* de identificação dos OAs auxiliam os mecanismos de busca na localização do objeto quando o usuário solicita uma busca em um repositório (Pereira *et al.*, 2002). Os conceitos referentes à metadados, seus padrões, e de repositórios serão descritos nas seções a seguir.

#### 2.1.1 Metadados

Os metadados são dados que descrevem dados (Ghelman *et al.* 2007), ou seja, dizem do que um dado se trata. Em complemento ALCTS (2013) afirma que os metadados são estruturados, codificados e descrevem características de entidades portadoras de informação

para ajudar na identificação, descoberta, avaliação e gerência das entidades descritas. Assim, no caso de OAs, os metadados podem ser entendidos como sendo as descrições do que se tratam os dados sobre o OA que estão sendo guardados. Como por exemplo, título ou palavras-chave.

A principal propriedade de um OA é a sua reusabilidade. Esta, pode ser alcançada através da modularidade, interoperabilidade e recuperação. Modularidade descreve o quanto um OA será separado, ou seja, o grau de separação, e consequente recombinação de seus componentes. Interoperabilidade é definida como a capacidade de operar em plataformas heterogêneas, ou seja, capacidade de operar em diferentes plataformas do mesmo modo. A recuperação está relacionada com sua capacidade de poder ser encontrado através de sua descrição de propriedades e funcionalidades. (Vicari *et al.*, 2009).

Para atender a estas características, surgiram algumas iniciativas que incentivam a criação de padrões de metadados (Gluz e Vicari, 2011; Vicari *et al.*, 2010). Nestes padrões, um OA é analisado em dois níveis: o de metadados e o de conteúdo. O nível de metadados está relacionado com as informações que serão catalogadas a respeito do OA, já o nível de conteúdo está relacionado com a organização e encapsulamento do material de aprendizado propriamente dito (Gluz e Vicari 2011). Isto, segundo Vicari *et al.* (2009), para compatibilizar tanto o desenvolvimento quanto o uso dos OAs em escala mundial.

Dentre as diversas organizações internacionais incentivadoras da criação dos padrões de metadados, algumas destacam-se, como por exemplo o organização criadora do padrão LOM (Vicari *et al.*, 2010). O padrão LOM especifica os conjuntos mínimos de atributos necessários para permitir que OAs possam ser gerenciados, localizados e avaliados em repositórios. Esses atributos incluem: tipo de objeto, autor, proprietário, termos de distribuição, formato, etc, e, onde aplicável, pode também incluir atributos pedagógicos, tais como, estilo de ensino ou interação, nível de ensino e pré-requisitos, etc. Em suma o LOM define o esquema de dados que descrevem um OA (IEEE LTSC 2013).

No Brasil, a primeira iniciativa que buscou a criação de um padrão de metadados para OAs foi o projeto Objetos de Aprendizagem Baseados em Agentes (OBAA), desenvolvido em parceria pela Universidade Federal do Rio Grande do Sul (UFRGS) e Universidade do Vale dos Sinos (UNISINOS). O padrão OBAA foi criado a partir do padrão LOM, com todas as suas categorias (geral, ciclo de vida, mete-metadados, técnicos, educacional, direitos,



relacional, anotação e classificação), sendo que as categorias técnica e educacional foram complementadas com metadados adicionais, e duas novas categorias, sendo estas relativas a aspectos de acessibilidade e segmentação (Vicari et al., 2010; Padrão OBAA - Mapa Mental, 2013). Cada uma das nove categorias pertencentes ao padrão LOM versão 1.0 (Draft Standard for Learning Object Metadata, 2012) e das outras três pertencentes ao padrão OBAA (Vicari et al, 2009), é descrita a seguir:

- a. **Geral** (*General*): agrupa as informações gerais que descrevem um objetos de aprendizagem como um todo.
- b. **Ciclo de Vida** (*Lifecycle*): agrupa as características relacionadas ao histórico e estado atual de um OA e outros que afetaram este OA durante sua evolução.
- c. **Meta-Metadados** (*Meta-Metadata*): agrupa as informações sobre as instâncias de metadados em si, mais do que as informações sobre o conteúdo que instância de metadados descreve sobre um OA.
- d. **Técnico** (*Technical*): agrupa os requisitos e características técnicos de um OA.
- e. **Educacional** (*Educational*): agrupa as características educacionais e pedagógicas de um OA.
- f. **Direitos** (*Rights*): agrupa as informações referentes aos direitos de propriedade intelectual e condições de uso de um OA.
- g. **Relação** (*Relation*): agrupa os aspectos que definem a relação entre um OA e outros OAs relacionados a ele.
- h. **Anotação** (*Annotation*): disponibiliza comentários sobre o uso educacional do OA, além de informações sobre quando e por quem os comentários foram criados.
- i. **Classificação** (*Classification*): descreve um OA em relação a um sistema de classificação em particular.
- j. **Acessibilidade** (*Accessibility*): armazena informações sobre as configurações para que o usuário possa acessar e utilizar o OA.
- k. **Segmentação** (*Segmentation*): guarda dados que permitam que um AO possa ser subdividido logicamente em outros OAs, para organização do mesmo por módulos ou assuntos que este trata.

Para encontrar e recomendar objetos similares, conforme é proposto nesse trabalho, serão utilizados os metadados apresentados na Quadro 1:

**Quadro 1: Metadados que serão utilizados para a recomendação no ROA**

<b>Categoria</b>	<b>Metadado</b>	<b>Descrição</b>
Geral	Título ( <i>Title</i> )	Nome do objeto de aprendizagem
	Descrição ( <i>Description</i> )	Descrição do objeto de aprendizagem
	Palavras-chave ( <i>Keyword</i> )	Palavras ou frases que descrevam o objeto de aprendizagem
	Cobertura ( <i>Coverage</i> )	A época, cultura ou geografia que o objeto de aprendizagem se aplica, por exemplo pode ser algo que se aplica somente em determinado período de tempo, e que já perdeu sua validade para a atualizade, ou se aplica a pessoas de uma região devido a sua cultura e não se enquadra em outras realidades.
Técnicos	Formato ( <i>Format</i> )	Formato dos conteúdos do objeto de aprendizagem
Educacional	Educacional ( <i>Educational</i> )	Características pedagógicas do objeto de aprendizagem
	Tipo de recurso ( <i>LearningResourceType</i> )	Tipo de recurso de aprendizagem utilizado, por ordem de predominância. Exemplo: vídeo, texto questionário, etc

Apesar de haver muitos outros metadados nas diferentes categorias descritas anteriormente, os metadados apresentados na Quadro 1 foram escolhidos pois se acredita que estes sejam suficientes para que se possam encontrar e recomendar objetos similares. Além disso, alguns deles (ex: palavras-chave, formato e tipo de recurso) serão utilizados como meio de filtrar previamente os OAs que serão desconsiderados para a verificação de similaridade.

### 2.1.2 Repositórios

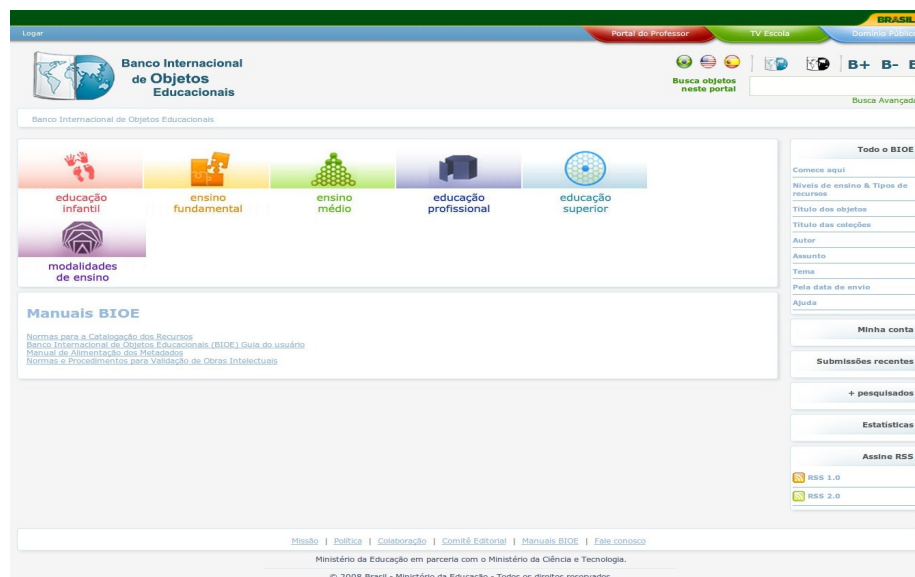
Repositório é o nome dado aos locais utilizados para armazenar OAs. Em uma analogia com uma biblioteca, repositórios armazenam OAs e seus metadados do mesmo modo que uma biblioteca armazena livros e suas descrições. Além disso, repositórios provem mecanismos para facilitar a descoberta dos OAs (Richards *et al.* 2002). Como dito anteriormente, são exemplos de ROAs o MERLOT Figura 1 e o BIOE Figura 2, os mesmos serão descritos a seguir.



Figura 1: Página inicial do MERLOT

Fonte: <http://www.merlot.org>

O MERLOT (Figura 1) é uma comunidade online de recursos projetada primariamente para que funcionários e alunos de instituições de ensino superior possam compartilhar materiais para aprendizagem. Nele é possível a pesquisa de OAs através da aplicação filtros como por exemplo língua, descrição, palavras-chave e título. Além disso ele disponibiliza informações como comentários dos usuários e avaliações dos materiais quanto a qualidade dos conteúdos, potencial de eficiência como uma ferramenta de ensino e facilidade de uso (MERLOT, 2013). Este repositório disponibiliza uma lista dos tipos de recurso e a quantidade de OAs presente em cada tipo de recurso, isto para facilitar que os usuários encontrem OAs. Quando um objeto em específico é acessado, ele também disponibiliza as categorias deste, de forma que o usuário possa navegar pelas categorias para encontrar outros OAs que pertençam a mesma categoria do objetos que está sendo acessado.



**Figura 2: Página inicial do BIOE**

**Fonte:** <http://objetoseducacionais2.mec.gov.br>

O BIOE (Figura 2) é um repositório que disponibiliza os OAs separados por nível de educação (cobertura) a serem aplicados, isto de forma similar a que o MERLOT disponibiliza a lista de tipos de recurso. Além disto, disponibiliza ferramentas para busca, que permitem filtrar por objetos individualmente ou por coleções de objetos, idioma, tipo de recurso e país. Nele também é possível classificar os OAs e visualizar uma lista com todos os seus metadados. O BIOE também disponibiliza listas de recomendação com os objetos mais acessados, mais votados e tipos mais vistos, além de destaques recentes. Este repositório será utilizado durante a primeira oficina, que será descrita na metodologia, pois este utiliza o padrão de metadados OBAA, que também é utilizado no repositório em que serão aplicadas as técnicas de recomendação.

O repositório do IFRS, também permite que o usuário navegue por nível de ensino como o BIOE, além de tipo de mídia. Apesar disso, ele disponibiliza essas opções em forma de lista, de forma semelhante a utilizada pelo MERLOT para tipos de recurso e categorias. Apesar de o BIOE fazer recomendações, o modo como as mesmas serão feitas no presente trabalho é diferente, pois utilizarão somente os conteúdos do objeto, ou seja, seus metadados, e não considerarão as contribuições dos usuários como classificação ou comentários sobre os OAs.

## **2.2 Sistemas de Recomendação**

Bradley e Smyth (2001) afirmam que sistemas de recomendação oferecem uma maneira inteligente e personalizada, para os usuários, para encontrar novas informações fácil e rapidamente. Sistemas de recomendação podem ainda, segundo Basu *et al.* (1998), dar sugestões de artefatos para um usuário. Segundo ele, estes sistemas se utilizam de diferentes abordagens para que tais sugestões correspondam aos interesses dos usuários e a seus objetivos durante o uso de sistema, além de , como dito anteriormente, garantir que a informação seja disponibilizada de forma rápida. A seguir serão descritas algumas destas abordagens ou estratégias, além de algumas técnicas utilizadas para recomendação.

### **2.2.1 Estratégias de Recomendação**

Existem algumas estratégias empregadas para efetuar as recomendações. Reategui e Cazella (2005) definem cinco estratégias mais utilizadas para recomendação: (i) listas de recomendação, (ii) avaliações de usuários, (iii) suas recomendações, (iv) Usuários que se interessaram por X também se interessaram por Y e (v) associação por conteúdo. As mesmas serão descritas nas seções a seguir.

#### **2.2.1.1 Listas de recomendação**

Listas de recomendação é uma estratégia baseada na separação de itens em listas, as quais estão organizadas por tipos de interesses. Ela pode ser encontrada em sites, onde são observados os itens mais populares. Estes, por sua vez, são agrupados e ordenados. São exemplos de listas: “Mais vendidos”, “Mais escutados”, “Mais baixados da semana”, etc. Nesta estratégia, a análise dos dados dos usuários é superficial, ou seja seus perfis não são bem definidos nesta análise, visto que os dados requeridos estão mais relacionados com os itens, de forma que se possa dizer quais são mais populares (Reategui e Cazella, 2005). As Figuras 3 e 4 apresentam exemplos de lista de recomendações.

Artistas <b>mais populares</b> desta semana	Principais <b>faixas</b> desta semana	<b>Eventos</b> populares perto de você
 <b>Fuse ODG</b> 4.937 ouvintes	 <b>Get Lucky (feat. Pharrell Williams)</b> Daft Punk 284.281 ouvintes	 <b>17ª Cultura Inglesa Festival</b> Memorial da América Latina, São Paulo 23 Jun 2013 - <b>528 vão</b>
 <b>Bosnian Rainbows</b> psychedelic rock · alternative rock 2.652 ouvintes	 <b>Give Life Back to Music</b> Daft Punk 276.329 ouvintes	 <b>Black Sabbath</b> Campo de Marte, São Paulo 11 Out 2013 - <b>486 vão</b>
 <b>Jason Isbell</b> alt-country · americana 40.225 ouvintes	 <b>Instant Crush (feat. Julian Casablancas)</b> Daft Punk 237.780 ouvintes	 <b>Planeta Terra Festival 2013</b> Campo de Marte, São Paulo 9 Nov 2013 - <b>189 vão</b>
 <b>Dawid Podsiadlo</b> polish · folk 8.785 ouvintes	 <b>Lose Yourself to Dance (feat. Pharrell Williams)</b> Daft Punk 228.896 ouvintes	 <b>Paramore</b> Espaço das Américas, São Paulo 30 Jul 2013 - <b>238 vão</b>
 <b>Baggi Begovic</b> house · deep house 13.309 ouvintes	 <b>Radioactive</b> Imagine Dragons 403.085 ouvintes	 <b>WROS Fest 2013</b> Espaço das Américas, São Paulo 12 Out 2013 - <b>214 vão</b>

**Figura 3: Listas de recomendação do site Last.fm**

Fonte: <http://www.lastfm.com.br>

No exemplo da Figura 3, que é de um site relacionado a músicas, são apresentadas três listas, a primeira referente aos artistas mais populares de uma semana específica, a segunda às principais faixas da semana e última aos eventos populares perto da localização do usuário. Já no exemplo da Figura 4, que se trata de um site para leitores, há duas listas: uma dos livros mais lidos e outra dos mais desejados para um determinado período de tempo.

### Top Mais

**Livros MAIS LIDOS**
[ver mais](#)





**Livros MAIS DESEJADOS**
[ver mais](#)





**Figura 4: Listas de recomendação do site Skoob**

Fonte: <http://www.skoob.com.br>

Ainda segundo Reategui e Cazella (2005), a principal vantagem neste tipo de estratégia está na facilidade de implementação. Pois, segundo eles, basta manter-se listas de acordo com as necessidades de marketing, de aumento de lucratividade, de disseminação de informações, etc. Eles afirmam que a desvantagem é que as recomendações não são dirigidas

a cada usuário independentemente, mas sim a todos os usuários sem distinção, ou seja ao público em geral, não sendo personalizadas para um usuário específico.

### 2.2.1.2 Avaliações de usuários

Dentre as estratégias mais utilizadas, em sistemas de recomendação, se encontram as avaliações dos usuários. Nela, o usuário pode, além de comprar um produto, deixar um comentário sobre o item adquirido. Esta estratégia é comum de ser encontrada na forma de opinião dos usuários ou na forma de avaliação dos itens, como ilustrado nas Figuras 5 e 6. Este também é um mecanismo fácil de implementar, pois não exige nenhum tipo de dispositivo inteligente. Basta somente armazenar e disponibilizar as avaliações de usuários sobre os itens tratados (Reategui e Cazella, 2005).

A Figura 5 mostra a forma de obtenção das avaliações para determinados itens em um site. Nela os usuários escolhem o número de estrelas para um item de acordo com a sua satisfação, sendo que a escala vai de um a cinco, onde quanto maior o número de estrelas, maior a satisfação do usuário. Além disso o usuário pode dar sua opinião em forma de um comentário textual.



**Figura 5: Exemplo de avaliação dos usuários no site da Submarino**

Fonte: <http://www.submarino.com.br>

A partir destes dados são geradas as recomendações, como mostrado na Figura 6, onde, de acordo com a média de estrelas escolhidas pelos usuário é calculada uma nota para cada um dos itens apresentados no site, e também é mostrado o número de usuários que classificaram cada item.



Figura 6: Disponibilização das avaliações dos usuários no site da Submarino

Fonte: <http://www.submarino.com.br>

### 2.2.1.3 Suas recomendações

Neste tipo de recomendação, as sugestões são feitas especificamente para o usuário, sendo que são possíveis dois tipos de recomendação: aquelas feitas a partir de preferências implícitas e explícitas (Reategui e Cazella, 2005). A Figura 7 apresenta um exemplo de página de recomendação onde são trazidas sugestões a partir de dados coletados de maneira implícita, pois o próprio sistema verifica o histórico de páginas acessadas pelo usuários para gerá-las. Nela, a partir da análise do perfil do usuário, são recomendados outras páginas para que o usuário se inscreva para receber atualizações.



Figura 7: Exemplo de recomendação personalizada no site do Twitter

Fonte: <http://www.twitter.com>

### 2.2.1.4 “Usuários que se interessaram por X também se interessaram por Y”

Segundo Reategui e Cazella (2005), este tipo de recomendação é obtido através de técnicas capazes de encontrar, em uma base de dados, associações entre itens avaliados ou vistos por usuários. Isto pode ser feito encontrando-se, a partir dos itens avaliados, comprados ou vistos pelo usuário, outros itens populares de mesmos autores, artistas, mesmas palavras-chave ou de assuntos similares (LINDEN, SMITH e YORK, 2003). Ela é como mostrado nos



exemplos a seguir.



**Figura 8:** Exemplo de recomendação por relação entre itens vistos pelos usuários no site da Americanas

Fonte: <http://www.americanas.com.br>

No exemplo ilustrado na Figura 8, a partir da seleção de um produto X, o sistema analisa quais outros produtos foram visualizados pelos usuários que visualizaram o produto X e os apresenta em uma lista.



**Figura 9:** Exemplo de recomendação por relação entre itens vistos pelos usuários no site da Casas Bahia

Fonte: <http://www.casasbahia.com.br>

O exemplo da Figura 9, a partir da busca por um produto X, o sistema analisa quais outros produtos os usuários que buscaram o produto X demonstraram ter interesse.

Este tipo de recomendação é a forma mais complexa de recomendação, pois ela exige uma análise bastante profunda dos hábitos dos usuários, de modo que possa identificar padrões e recomendar itens com base nestes padrões (Reategui e Cazella, 2005).

### 2.2.1.5 Associação por conteúdo

Reategui e Cazella (2005), definem que, nesta estratégia, a recomendação é feita com

base no conteúdo de um item em específico, por exemplo um autor, um compositor, um editor, etc. Para tal, se faz necessário encontrar associações entre os itens, porém estas estão num escopo mais restrito. A Figura 10 mostra um exemplo deste tipo de recomendação, onde a partir da seleção de um livro, o sistema recomenda a compra de outros livros em conjunto, normalmente estes são do mesmo autor. Segundo eles, diferentes técnicas podem ser utilizadas para se obter este tipo de recomendação, estas são apresentadas no próximo capítulo.

aproveite e compre junto



Figura 10: Exemplo de recomendação por conteúdo no site da Saraiva

Fonte: <http://www.livrariasaraiva.com.br>

## 2.2.2 Técnicas de Recomendação

Alguns autores definem três categorias principais de técnicas aplicadas em sistemas de recomendação de acordo com a sua abordagem. São elas: baseados em conteúdo, colaborativos e híbridos (Reategui e Cazella, 2005; Kim *et al.*, 2006). Segundo Reategui e Cazella (2005), estas técnicas é que fundamentam o funcionamento dos sistemas de recomendação. Elas estão descritas nas seções a seguir:

### 2.2.2.1 Baseados em conteúdo

Sistemas de recomendação baseados em conteúdo analisam as descrições de itens para identificar quais deles são de interesse particular para o usuário. Além disso, esses sistemas podem recomendar itens parecidos com outros itens que o usuário teve interesse no passado. Isto pois esta técnica parte do princípio de que os usuários tendem a interessar-se por itens similares aos que demonstraram interesse no passado (Pazzani e Billsus, 2007; Kim *et al.*, 2006; Lichtnow *et al.*, 2006).

A descrição de interesses do usuário pode ser obtida através de informações fornecidas pelo próprio usuário ou de forma automática, sendo que na última são observados os conteúdos dos itens que o usuário consumiu, através de informações mantidas em históricos de visualização e aquisição de itens (Herlocker, 2000). Uma abordagem utilizada por muitas ferramentas que aplicam esta técnica é a indexação de frequência de termos (Salton e Buckley, 1988). Neste tipo de indexação, as informações dos documentos e das necessidades dos usuários são descritas na forma de vetores de uma dimensão para cada palavra encontrada na base de dados. Cada item do vetor é a frequência que uma respectiva palavra é encontrada em um documento ou na consulta do usuário. Por fim, os vetores dos documentos que estão mais próximos aos vetores da consulta feita pelo usuário são considerados os que tem maior relevância para ele (Herlocker, 2000).

Para trabalhar com a filtragem baseada em conteúdo, o sistema pode solicitar ao usuário que o mesmo avalie se determinados itens são de seu interesse. Após a avaliação, o sistema faz a filtragem dos itens que "casam" melhor seus conteúdos com os itens que foram classificados como sendo de interesse, e desconsidera os que "casam" seus conteúdos como os que foram classificados como não sendo de interesse. Porém, sistemas deste tipo apresentam limitações como por exemplo: o conteúdo de dados pouco estruturados é difícil de ser analisado (ex: vídeo e som); o entendimento do conteúdo do texto pode ser prejudicado devido a uso de sinônimos; pode ocorrer a super especialização, pois o sistema procura se basear em avaliações positivas e negativas feitas pelo usuário, não apresentando conteúdos que não fechem com o perfil dele (Reategui e Cazella, 2005).

#### **2.2.2.2 Colaborativos**

Sistemas que utilizam filtragem colaborativa tem sido desenvolvidos para serem aplicados para tratar de limitações da técnica baseada em conteúdo. Esta técnica se diferencia da baseada em conteúdo, pois não necessita de análise do conteúdo dos itens, mas sim de se conhecer avaliações dos usuários para aqueles itens, ou seja, ela tenta prever a utilidade de um item baseada em itens previamente avaliados por outros usuários (Herlocker, 2000; Kim *et al.* 2006).

Enquanto os primeiros sistemas de filtragem colaborativa necessitavam que os usuários especificassem o relacionamento de predição entre as opiniões de outros usuários, ou explicitamente enviar ou indicar itens a outros usuários, os sistemas mais modernos

automatizaram todo este procedimento, menos a parte da coleta das avaliações dos itens pelos usuários (Goldberg *et al.*, 1992 *apud* Herlocker, 2000; Maltz e Ehrlich, 1995 *apud* Herlocker, 2000). Um usuário de um sistema de filtragem colaborativa automatizada deve portanto avaliar cada item experimentado, indicando o quanto este item casa com suas necessidades. Estas avaliações são coletadas de grupos de pessoas, o que permite que usuários se beneficie das avaliações apresentadas por outros usuários.

Sistemas de filtragem colaborativa simples apresentam para o usuário uma média das avaliações para cada item de interesse em potencial (Maltz, 1994 *apud* Herlocker, 2000). Deste modo, as avaliações permitem que o usuário conheça os itens que são considerados de interesse popular e evite os itens que são de pouco interesse. Porém as recomendações não são personalizadas, pois apresentam para todos os usuários os mesmos itens de acordo com suas avaliações.

Os sistemas mais avançados, por sua vez, descobrem de maneira automática as relações entre os usuários, baseado na descoberta de padrões comuns de comportamento (Herlocker, 2000). Sistemas deste tipo mostram para cada usuário predições personalizadas que podem ser diferentes das mostradas aos outros usuários.

A Quadro 2 mostra, na prática, o funcionamento da filtragem colaborativa. No exemplo, caso fosse desejado recomendar um produto ao usuário Antônio, outros usuários com hábitos de consumo semelhantes teriam que ser analisados. Neste caso, os usuários Juliana e Rosana já compraram produtos que Antônio também comprou (Tablet e Smartphone). Em seguida, seriam recomendados a Antônio produtos que estes dois outros usuários possuem, mas que Antônio ainda não possui (Notebook e Celular).

**Quadro 2: Exemplo prático de recomendação baseada em filtragem colaborativa**

<b>Usuário</b>	<b>Tablet</b>	<b>Desktop</b>	<b>Smartphone</b>	<b>Notebook</b>	<b>Ultrabook</b>	<b>Celular</b>
Antônio	<u>x</u>		<u>x</u>			
Jéssica					x	x
Marcos		x			x	
Juliana			<u>x</u>	x		x
Paulo		x				x
Rosana	<u>x</u>			x		

A filtragem colaborativa apresenta algumas vantagens, como por exemplo a

possibilidade recomendar aos usuários itens inesperados. O usuário poderia receber recomendações de itens que não estavam sendo pesquisados por ele de forma ativa. Esta técnica também possibilitou a formação de comunidades de usuários pela identificação de seus gostos e interesses similares (Reategui e Cazella, 2005).

Segundo Reategui e Cazella (2005), é importante observar que em sistemas colaborativos, no refere-se à coleta de informações dos usuários, podem haver alguns problemas:

1. **Problema do primeiro avaliador:** que ocorre quando um novo item aparece no banco de dados, neste caso não existe maneira deste item ser recomendado para o usuário até que mais informações sejam obtidas através da avaliação dos outros usuários.
2. **Problema de pontuações esparsas:** o objetivo dos sistemas de filtragem colaborativa é ajudar pessoas, focando em documentos lidos ou itens adquiridos. Caso o número de usuários seja pequeno em relação ao volume de informações no sistema existe um grande risco das pontuações se tornarem muito esparsas, pois os usuários podem ter preferências diferentes.
3. **Similaridade:** caso um usuário tenha gostos que variam dos padrões do grupo, este terá dificuldades para encontrar outros usuários com gostos similares, deste modo suas recomendações podem se tornar pobres.

Reategui e Cazella (2005) definem ainda, três passos para que ocorra a recomendação por esta técnica:

1. Primeiro faz-se necessário calcular o peso de cada usuário (similaridade) em relação ao usuário ativo ou que está acessando o sistema.
2. Após é selecionado um subconjunto de usuários com maiores similaridades (vizinhos mais próximos) para considerar na predição.
3. E por fim são normalizadas as avaliações e computadas as predições ponderando as avaliações dos vizinhos com seus pesos.

Segundo eles, esta técnica também é chamada de “k-nearest-neighbor” ou “user-based”. No primeiro passo, a definição da similaridade pode ser realizada através de diversas técnicas, sendo mais comumente usada a correlação de Pearson como mostra a equação (1), onde  $r_{u,i}$  é o conjunto de avaliações do usuário ativo,  $r_{u,i}$  é conjunto de avaliações do usuário similar a  $u$

e  $\omega_{a,u}$  é o nível de similaridade entre os usuários comparados.

$$\omega_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) * (r_{u,i} - \bar{r}_u)}{\sigma_a * \sigma_u}$$

**Equação 1: Correlação de Pearson**

**Fonte:** Adaptado de Reategui e Cazella, 2005

Dado o conjunto de avaliações  $r_{a,i}$  do usuário ativo  $a$ , e o conjunto de avaliações  $r_{u,i}$  do vizinho  $u$ ,  $\omega_{a,u}$  é o índice de similaridade entre os dois usuários. Porém, para tal é necessário haver mais de uma avaliação em comum para que o índice seja útil, e os resultados variam entre 1 para similaridade total, e -1 para total dissimilaridade.

Para que esta recomendação seja realizada, faz-se necessária a verificação das avaliações feitas pelos usuários aos item em questão e o grau de similaridade entre os perfis destes e o perfil do usuário que se deseja realizar a recomendação. O cálculo da predição é efetuado através da Equação 2.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * \omega_{a,u}}{\sum_{u=1}^n \omega_{a,u}}$$

**Equação 2: Cálculo da predição**

**Fonte:** Adaptado de Reategui e Cazella, 2005

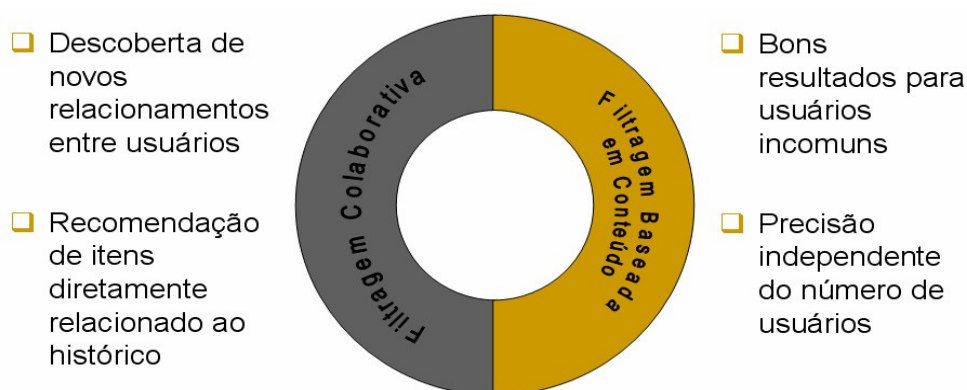
O valor da predição  $p_{a,i}$  do item  $i$  para o usuário ativo  $a$  é dada pela média ponderada das avaliações dadas à tal item  $i$ , pelos  $n$  vizinhos  $u$  do usuário ativo  $a$ . A quantidade  $n$  de melhores vizinhos, ou seja os que possuem maiores correlações, é uma escolha de cada sistema que utiliza esta técnica.

### 2.2.2.3 Híbridos

Sistemas de recomendação híbridos combinam os pontos fortes dos métodos colaborativo e baseado em conteúdo para satisfazer algumas limitações que cada um dos métodos possui. Esta combinação pode ser feita de diferentes maneiras, como descrito a seguir: implementando ambos métodos separadamente e combinando os resultados; incorporando algumas características da abordagem baseada em conteúdos na a abordagem colaborativa; ou construindo um modelo unificado que incorpora as características de ambos

(Kim *et al.* 2006).

Essa abordagem é constituída de vantagens proporcionadas pela filtragem baseada em conteúdo e pela filtragem colaborativa, unindo o melhor das duas técnicas e eliminando as fraquezas de cada uma, conforme expressado pela Figura 11 (Reategui e Cazella, 2005).



**Figura 11: Características herdadas pela Filtragem Híbrida**

Fonte: Adaptado de Reategui e Cazella, 2005

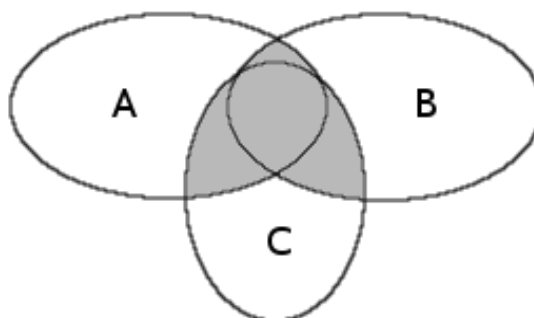
## ***2.3 Modelos para recuperação de informação para sistemas de recomendação baseados em conteúdo***

A base do processo de recomendação pode ser vista como sendo a recuperação de informação, na qual documentos relevantes aos usuários devem ser recuperados e recomendados. Para isto, podem ser utilizados modelos clássicos de recuperação de informação tais como o booleano, o espaço vetorial e o probabilístico (Salton e McGill, 1983 *apud* Lopes, 2007; Baeza-Yates e Tibeiro-Neto, 1999 *apud* Lopes, 2007; Grossman, 2004 *apud* Lopes, 2007). Estes três modelos estão descritos nas seções a seguir.

### **2.3.1 Modelo booleano**

O modelo booleano (Salton e McGill, 1983 *apud* Lopes, 2007) é um modelo de recuperação de informação simples, que se baseia na teoria de conjuntos e na álgebra booleana. Nele, os documentos (***D***) são representados como conjuntos de termos de indexação e as consultas (***Q***) são formuladas através de expressões booleanas formadas por termos e conectivos de boole (***and***, ***or*** e ***not***). Estas expressões booleanas permitem a combinação das operações de união (***or***), intersecção (***and***) e negação (***not***) de conjuntos. Um exemplo de uma expressão booleana que poderia ter sido formulada, visando recuperar documentos que

possuam informações sobre um assunto como por exemplo “modelo booleano” e/ou sobre “conjuntos” é apresentado na Figura 12. Onde **A** representa os documentos que contém o termo “modelo”; **B** os documentos que contém o termo “booleano”; e **C** os documentos que contém o termo “conjuntos”.



**Figura 12: Representação gráfica do resultado da expressão booleana (“modelos” and “booleanos”) or “conjuntos”**

**Fonte: Adaptado de Lopes, 2007**

Como resultado do processo de recuperação, são retornados somente os documentos que satisfaçam a todas as restrições lógicas representadas pela consulta, ou seja, é uma estratégia de recuperação baseada num critério de decisão binário. Um documento só pode ser dito como relevante ( $\text{Similaridade}(Q, D) = 1$ ) ou não-relevante ( $\text{Similaridade}(Q, D) = 0$ ). Caso todas as condições da expressão booleana **Q** sejam satisfeitas a similaridade será 1, a mesma será 0 em caso contrário. Neste modelo não há a noção de “casamento parcial” em relação às condições da consulta (Lopes, 2007).

Objetivando garantir a qualidade da informação que será recuperada, muitas vezes, há a necessidade de especificação de consultas complexas, o que exige um conhecimento profundo sobre a informação desejada e lógica booleana para elaboração de consultas. Outra dificuldade do uso deste modelo é a incapacidade de se representar pesos associados aos termos desejados, não havendo possibilidade de diferenciação entre a importância dos termos para consulta desejada, ou seja, dois termos, solicitados em uma consulta, são igualmente importantes para representar a consulta, e, desta forma, se um termo está presente em um documento ele é tão importante para representação daquele documento quanto qualquer outro termo que esteja presente no mesmo. Portanto, o modelo booleano não leva em consideração as diferentes importâncias dos termos para representação dos documentos (Lopes, 2007).

Segundo Baeza-Yates e Ribeiro-Neto (1999, *apud*, Lopes 2007), as principais



vantagens do modelo booleano estão na sua simplicidade e no formalismo claro envolvidos. Já as desvantagens estão associadas ao fato de que o “casamento” exato pode levar à recuperação de poucos ou muitos documentos, obtendo uma performance fraca. Em virtude de suas limitações, outro modelo foi desenvolvido como uma proposta de extensão deste, este modelo é o Modelo Booleano Estendido (Salton *et al.*, 1983 *apud* Lopes 2007), um modelo intermediário entre o sistema booleano e o modelo vetorial, apresentado na próxima seção.

A estrutura de consulta inerente ao sistema booleano é preservada no modelo estendido, porém pesos associados aos termos podem ser incorporados em consultas e documentos. Deste modo, a saída obtida pode ser classificada em ordem de similaridade com a consulta desejada. Entretanto, tal modelo se torna ainda mais complexo, pois, além da necessidade do domínio da lógica booleana, existe a necessidade da determinação do grau de importância dos termos, o que é representado através dos valores de seus pesos, que são aplicados tanto para termos da consulta quanto para representação dos documentos.

### 2.3.2 Modelo de Espaço Vetorial (VSM) ou Modelo Vetorial

O VSM (*Vector Space Model*) é definido como sendo um modelo baseado em conteúdo, com pesos associados aos termos de indexação e, cujo resultado da função de similaridade é dado na forma de ranking. Esta abordagem possui uma implementação relativamente simples e provê resultados satisfatórios. No modelo de espaço vetorial (Salton *et al.*, 1975 *apud* Lopes, 2007), documentos e consultas são representados como vetores de termos de indexação. Cada termo possui um peso atrelado a si, para permitir distinção entre os termos de acordo com sua importância. Segundo Salton e Buckley (1988), os pesos podem variar entre 0 e 1, sendo que os valores próximos a 1 correspondem aos termos mais relevantes e, os valores próximos a 0 correspondem aos termos menos relevantes.

O VSM utiliza um espaço  $n$ -dimensional para representar os termos. Neste espaço,  $n$  (ou número de dimensões) representa o número de termos distintos encontrados. Para cada vetor  $w$  de documentos ou de consulta, os pesos representam as coordenadas do vetor  $w$  na dimensão  $n$  correspondente. O princípio do VSM é baseado na correlação inversa entre a distância ou ângulo entre vetores de termos no espaço e a similaridade entre os documentos que eles representam. Para calcular o escore de similaridade, o co-seno (Equação 3) pode ser utilizado (fórmula obtida do produto escalar entre os dois vetores desejados, dividido pela multiplicação dos módulos desses vetores). O valor resultante indica o grau de relevância

entre a consulta ( $Q$ ) e o documento ( $D$ ), onde  $w$  representa os pesos dos termos contidos em  $Q$  e  $D$ , e  $t$  representa o número de termos ou tamanho do vetor. Esta equação provê uma saída, classificada com base na ordem decrescente dos valores de similaridade obtidos (Salton e Buckley, 1988).

$$\text{Similaridade}(Q, D) = \frac{\sum_{k=1}^t w_{qk} \cdot w_{dk}}{\sqrt{\sum_{k=1}^t (w_{qk})^2 \cdot \sum_{k=1}^t (w_{dk})^2}}$$

**Equação 3: Escore de similaridade**

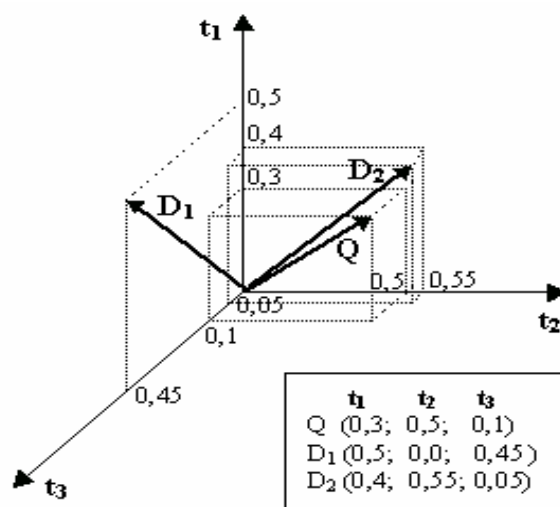
**Fonte: Adaptado de Lopes, 2007**

Na Figura 13, é apresentado um exemplo de um espaço vetorial tri-dimensional, onde cada dimensão é um dos termos de indexação, ou seja,  $t1$ ,  $t2$  e  $t3$ . Nela, são mostrados dois documentos, representados no espaço vetorial por  $D1$  (0,5; 0,0; 0,45) e  $D2$  (0,4; 0,55; 0,05), onde os números indicam as coordenadas dos termos nessas dimensões e representam os pesos associados a cada um dos termos na representação dos documentos. Também é representada a consulta  $Q$  (0,3; 0,5; 0,1) desejada, onde o peso dos termos representa a respectiva importância de cada termo para esta consulta. Sendo aplicada a Equação 3, para calcular a similaridade entre os documentos e a consulta desejada, são obtidos os seguintes valores aproximados:

- **Similaridade ( $Q, D1$ )** = 0,4899 = 48,99%

- **Similaridade ( $Q, D2$ )** = 0,9915 = 99,15%

Através deste resultado, pode ser observado que o documento  $D2$  é mais similar à consulta  $Q$ , em comparação ao documento  $D1$ . Isto pode ser visto no gráfico da Figura 13, onde  $D2$  está mais próximo à consulta  $Q$ , ou seja, o ângulo formado entre os dois vetores é menor do que o formado entre a consulta  $Q$  e  $D1$ .



**Figura 13: Espaço vetorial tri-dimensional**

Fonte: Adaptado de Lopes, 2007

De acordo com Lopes (2007), a Equação 3 é amplamente utilizada para comparar a similaridade entre documentos. Entretanto, em seu trabalho,  $Q$  representa o perfil do usuário e  $D$  os descritores dos documentos que foram obtidos através de colheita em uma biblioteca digital. Outro aspecto importante é a atribuição de pesos aos termos, isto garante um processo de recuperação de informação mais efetivo. Segundo o autor, os resultados deste processo dependem crucialmente da escolha deste sistema e, além disso, a seleção dos termos da consulta é um passo fundamental para a obtenção de um resultado de recomendação, que esteja em conformidade com as necessidades do usuário.

Vantagens e desvantagens do modelo vetorial (Baeza-Yates e Ribeiro-Neto, 1999 *apud* Lopes, 2007):

- Aplica esquema de atribuição de pesos aos termos, o que melhora a performance da recuperação;
- Utiliza estratégia de casamento parcial, o que permite recuperação de documentos, os quais se “aproximam” das condições da consulta;
- Com a fórmula de ranking pelo co-seno, documentos são ordenados de acordo com o grau de similaridade, em relação à consulta.
- Por fim, o modelo vetorial tem a desvantagem de não considerar a correlação entre os termos de indexação.

### 2.3.3 Modelo Probabilístico

O modelo probabilístico, está baseado na teoria das probabilidades como meio para modelar o processo de recuperação de informação. Nele, a função de similaridade entre um documento e uma de busca, é calculada pela probabilidade de tal documento (**D**) ser relevante a tal consulta (**Q**), caso os termos (**t<sub>i</sub>**) da consulta apareçam no documento. Presume-se que a distribuição dos termos, nos documentos da coleção, é uma informação que é capaz de determinar a relevância ou não de um documento em responder a uma dada consulta. Portanto, a ideia do modelo probabilístico é de que, quando vetores de documentos e consultas (termos envolvidos em ambas) são suficientemente similares, a probabilidade de relevância correspondente é alta o suficiente, para ser considerada a recuperação do documento em resposta à consulta (Salton e McGill, 1983 *apud* Lopes 2007).

No modelo, cada documento é representado por um vetor de termos, porém este difere do modelo de espaço vetorial apresentado na seção anterior pois não há um peso atrelado a cada termo mas sim um valor binário. Este valor binário indica apenas a presença (1) ou ausência (0) do termo no documento. Para o cálculo da similaridade, é utilizada a Equação 4 (Lopes, 2007).

Na equação:  $p(t_i | Rel)$  representa a probabilidade de um termo  $t_i$  estar presente em um documento selecionado do conjunto dos relevantes;  $p(t_i | \overline{Rel})$  representa a probabilidade de um termo  $t_i$  estar presente em um documento selecionado do conjunto dos não-relevantes  $p(\bar{t}_i | \overline{Rel})$  representa a probabilidade de um termo  $t_i$  não estar em um documento selecionado do conjunto dos não-relevantes;  $p(\bar{t}_i | Rel)$  representa a probabilidade de um termo  $t_i$  não estar presente em um documento selecionado do conjunto dos documentos relevantes.

$$Similaridade(Q, D) = \sum_{i=1}^t \left( \log \frac{p(t_i | Rel).p(\bar{t}_i | \overline{Rel})}{p(t_i | \overline{Rel}).p(\bar{t}_i | Rel)} \right)$$

Equação 4: Cálculo da similaridade

Fonte: Adaptado de Lopes, 2007

A Equação 4 é fundamental para a ordenação dos documentos neste modelo, sendo que ela foi obtida com base na aplicação do Teorema de Bayes. Esta equação pode também ser expressa como mostrado na Equação 5, caso sejam considerados os parâmetros apresentados no Quadro 3, na qual: **N** representa o número total de documentos da coleção; **n** é o número de documentos, contendo o termo desejado; **R** é o número de documentos (**D**),

relevantes para a consulta ( $Q$ ) e  $r$  é o número desses documentos relevantes que contêm o termo especificado (com o intuito de simplificar a equação, o sufixo  $i$ , que especifica o termo desejado, foi omitido; então,  $r = r_i$  e  $n = n_i$ , caso os termos sejam especificados).

$$Similaridade(Q, D) = \sum_{i=1}^t \left( \log \frac{r \cdot (N - n - R + r)}{(n - r) \cdot (R - r)} \right)$$

**Equação 5:** Cálculo da similaridade, considerando os termos do Quadro 3

**Fonte:** Adaptado de Lopes, 2007

**Quadro 3:** Lista de contingência da incidência de termos

	Relevantes	Não-Relevantes	
Contendo termo	$r$	$N - r$	$n$
Não contendo termo	$R - r$	$N - n - R + r$	$N - n$
	$R$	$N - R$	$N$

**Fonte:** Adaptado de Jones *et al.*, 1998 *apud* Lopes, 2007

A principal vantagem do modelo probabilístico, é que o ranking dos documentos é realizado em ordem decrescente da probabilidade de relevância. E, as desvantagens são: a necessidade da suposição da separação inicial dos documentos nos conjuntos relevantes ou não-relevantes; o fato de que o método não leva em conta a frequência com que um termo de indexação ocorre no documento (somente considera pesos binários); e a adoção da suposição da independência dos termos (Baeza-Yates & Ribeiro-Neto, 1999 *apud* Lopes, 2007).

## 2.4 Trabalhos similares

Nesta seção são apresentados trabalhos similares ao presente trabalho, são eles o *Learning Object Recommender System* (LorSys) e o DICA, que serão descritos nas seções a seguir. Mesmo havendo vários outros sistemas de recomendação disponíveis, estes foram escolhidos pois ambos utilizam OAs para recomendação, além disso, os outros sistemas encontrados durante a pesquisa não foram considerados relevantes para serem apresentados aqui. Vale ressaltar que estes dois sistemas, são de referencial puramente teórico, pois durante a pesquisa não se teve acesso aos sistemas como usuário para testá-los, nem tão pouco aos códigos fonte para entender a forma como cada um trabalha.

### 2.4.1 LorSys

O LorSys, segundo Ferreira (2009), é um sistema de recomendação de objetos de aprendizagem, que utiliza as técnicas de filtragem colaborativa e filtragem baseada em conteúdo para gerar recomendações. O sistema gera recomendações de objetos de aprendizagem baseado nas informações (metadados) dos OAs e com o uso de informações sobre a utilização dos mesmos, além permitir que um professor de uma disciplina defina grau de relação entre eles.

Primeiramente, o sistema utiliza informações sobre os OAs que estão armazenadas em uma base de dados, porém o LorSys também grava algumas informações sobre a utilização de um OA. Para tal, sempre que uma página contendo um OA for acessada por um usuário e este permanecer por no mínimo 20 segundos, é salvo no banco em uma tabela denominada “preferencia”, isto pois o autor estipulou que se o usuário utilizou o OA por 20 segundos ou mais, este o fez de forma intencional (Ferreira, 2009).

A partir dos dados salvos no banco, o algoritmo faz a recomendação utilizando duas técnicas: filtragem baseada em conteúdos e filtragem colaborativa. Porém elas não cooperem entre si, não permitindo que a mesma seja considerada uma abordagem híbrida (Ferreira, 2009).

Segundo (Ferreira, 2009), para implementar a técnica de filtragem colaborativa, foram utilizados dados sobre as preferências de todos os alunos que estavam matriculados em uma disciplina no sistema em que foi aplicada a técnica. O LorSys verifica então qual foi o último OA que o usuário, para quem serão feitas as recomendações, acessou. A partir disto, o sistema busca quais foram os outros usuários que acessaram este mesmo objeto. Identificados os usuários, o sistema recomenda o último objeto acessado por cada um deles.

Segundo o autor, apesar de existirem equações que possibilitem o cálculo de similaridade entre os perfis dos usuários, as mesmas não são utilizadas pelo LorSys. O mesmo optou por utilizar somente as informações sobre a utilização dos objetos de aprendizagem para gerar as recomendações.

Para implementar a técnica de filtragem baseada em conteúdo, segundo o autor, são utilizados os dados armazenados na tabela de preferência e de relação entre o conteúdo dos objetos de aprendizagem. A relação, é feita por parte do professor de forma manual. Para isto, o LorSys disponibiliza uma interface com os objetos de aprendizagem e seu conteúdo,

extraídos dos metadados e o professor deve estabelecer um grau de relacionamento entre os objetos de aprendizagem, conforme mostrado na Figura 14.

<b>Título</b>	Árvore Binária de Busca		
<b>Assunto:</b>	Estrutura de Dados, Árvore Binária de Busca, Aplicativo		
<b>Descrição:</b>	Permite ao acadêmico visualizar e interagir com a estrutura de dados de uma Árvore Binária de Busca.		
<b>Autor:</b>	Laboratório de Soluções em Software	<b>Organizador:</b>	UNIVALI
<b>Contribuição:</b>	Vinicius H. Ferreira (Produtor do Objeto), Elieser Ademir de Jesus (Desenvolvedor do Aplicativo), Rudimar Luís Scaranto Dazzi (Desenvolvedor do Aplicativo), Fabiane Vavassori (Produtora do Objeto).		
<b>Licença:</b>	(BY-NC-ND)Esta obra está licenciada sob uma Licença Creative Commons ( <a href="http://creativecommons.org/licenses/by-nc-nd/3.0/deed.pt">http://creativecommons.org/licenses/by-nc-nd/3.0/deed.pt</a> )		

Título	Relação
Busca em Árvore Binária	Sem relação ▼
Compilando Pascal	Pré-Requisito ▼
Fila	Complemento ▼
Lista	Similar ▼

Figura 14: Interface de relação entre objetos de aprendizagem

Fonte: Adaptado de Ferreira, 2005

De acordo com Ferreira (2009), o grau de relacionamento entre os OAs pode ser:

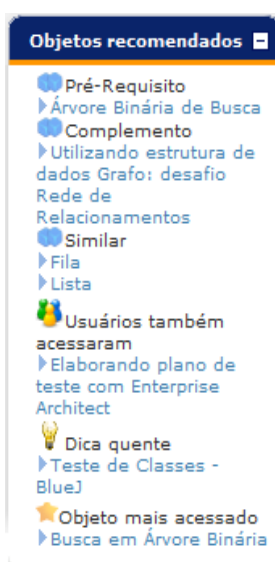
- **Pré-requisito:** para objetos de aprendizagem com conteúdo necessário para o entendimento do objeto de aprendizagem sendo acessado.
- **Complemento:** para objetos de aprendizagem com conteúdo complementar ao assunto tratado no objeto de aprendizagem sendo acessado.
- **Similar:** para objetos de aprendizagem com conteúdo similar ao assunto tratado no objeto de aprendizagem sendo acessado.
- **Sem relação:** para objetos de aprendizagem sem relação com o objeto de aprendizagem sendo acessado (esta é a opção padrão).

Quando um aluno acessa uma disciplina no sistema em que o LorSys está inserido, é gerada uma lista de objetos recomendados e também com o objeto de aprendizagem que foi mais acessado pela turma da disciplina em que o OA foi inserido. Quando um objeto de

aprendizagem aparece mais de uma vez nesta lista de recomendação, sua recomendação é reforçada e classificada como dica quente (Ferreira, 2009).

De acordo com Ferreira (2009), esta abordagem possibilita que um aluno encontre OAs que foram acessados por outros alunos com interesses similares aos dele, e, com objetos que tenham assunto relacionado ao objeto ao qual ele acessou.

Para a disponibilização das recomendações geradas pelo sistema foi utilizada uma lista como mostrado na Figura 15, onde são listados os objetos relacionados, os objetos que também foram acessados por outros usuários que acessaram o objeto sendo acessado pelo usuário, o objeto classificado como dica quente e o mais acessado:



**Figura 15: Objetos de Aprendizagem recomendados pelo LorSys**

Fonte: Adaptado de Ferreira, 2005

## 2.4.2 DICA

O DICA utiliza as informações do BIOE para recomendar objetos de acordo com as semelhanças existentes entre eles e também disponibiliza um espaço para comentários, para que os usuários do sistema opinem em relação aos OAs. O sistema utiliza as técnicas de agrupamento e a do vizinho mais próximo (Vieira e Nunes, 2012).

De acordo com os autores, o processo de coleta de dados foi realizado de forma semiautomática. Nele, foi analisado o código fonte da página do BIOE, suas *tags html* e os seus metadados para que, ao reconhecer um padrão, fosse desenvolvido um software que coletasse os dados de forma automatizada. Dentre os metadados encontrados, muitos foram considerados como não sendo importantes para a recomendação. Após análise dos mesmos,



foram escolhidos os seguintes para a recomendação: código do objeto, nome do objeto, observação, descrição, tema, visualizações e downloads.

Feita a escolha, a próxima etapa foi o processamento dos dados, para o qual, foram estudadas algumas técnicas de recomendação e, como dito anteriormente, por fim escolhidas a de agrupamento e a do vizinho mais próximo. O agrupamento tem como premissa básica a agregação de registros que possuem semelhanças em uma determinada base de dados em subconjuntos, denominados de grupos. No agrupamento realizado para o processamento dos dados foram escolhidas as informações referentes às visualizações e downloads; estes dois valores juntos foram utilizados como valor base do agrupamento (Vieira e Nunes, 2012).

Ao se abrir o sistema, são recomendados os OAs com maior taxa de visualização. Após é possível se fazer uma consulta por tema do OA, e então são listados OAs referentes aqueles temas. Caso hajam poucos OAs para o tema selecionado, são listados também os que objetos de temas que, a partir da análise da distância dos termos, sejam considerados mais próximos (Vieira e Nunes, 2012).

A recomendação propriamente dita ocorre durante a seleção de um OA específico. Onde são apresentadas as recomendações em uma lista, utilizando a técnica do vizinho mais próximo. Para tal, os autores fizeram uma pequena modificação no algoritmo TF-IDF. Para o qual são consideradas duas medidas: *tf*(*term frequency*) ou frequência do termo e o *idf*(*inverse document frequency*) ou frequência inversa do documento, a qual define a relevância de um termo em um conjunto de documentos e, quanto maior for o valor desse índice, o termo é considerado mais importante para o documento em que ele ocorre (Vieira e Nunes, 2012).

No sistema é feita uma comparação entre todos os termos do objetivo mais os termos da descrição do OA que está sendo consultado, com os 20 mais próximos a ele, que são selecionados através do algoritmo de distância entre os temas, como dito anteriormente. Através deste cálculo são apresentados os vizinhos mais próximos, que serão aqueles que possuem uma maior quantidade de termos em comum com o OA selecionado (Vieira e Nunes, 2012).

### 2.4.3 Características consideradas para o presente trabalho

Algumas das características apresentadas nestes sistemas que se pretende aplicar ao

algoritmo de recomendação que será elaborado no presente trabalho são:

- Disponibilizar a recomendação quando um usuário acessar um OA em específico;
- Recomendar somente OAs através da técnica baseada em conteúdo;
- Levar em consideração da frequência que um termo aparece;
- Levar em consideração da distância entre textos.

## 3 PROJETO

Visto que o presente trabalho objetiva avaliar como a aplicação de um algoritmo de recomendação no ROA do IFRS facilita que os usuários encontrem OAs, este capítulo tem por objetivo apresentar as definições quanto ao desenvolvimento de tal algoritmo.

Nesta etapa, primeiramente foi feito o levantamento de requisitos que o algoritmo deveria contemplar. Após foi gerado o modelo lógico do banco de dados e o diagrama de classes. Feito isto o próximo passo foi o desenvolvimento do algoritmo e implementação das técnicas utilizadas para recomendação.

### 3.1 *Levantamento de requisitos*

Nesta seção são descritas os requisitos necessários para o desenvolvimento do algoritmo de recomendação e para disponibilização das recomendações.

#### 3.1.1 Requisitos funcionais

A seguir são apresentados os principais requisitos funcionais, descrevendo as funcionalidades que o presente trabalho atende.

- **RF01:** O sistema deve gerar a lista de recomendações através dos dados armazenados sobre os objetos de aprendizagem;
- **RF02:** O sistema deve disponibilizar a lista de recomendações quando o usuário acessar um objeto em específico.

#### 3.1.2 Requisitos não funcionais

Nesta seção são apresentados os requisitos não funcionais do algoritmo.

- **RNF01:** Os metadados dos OAs devem estar no padrão OBAA;
- **RNF02:** O sistema deve ser desenvolvido na linguagem PHP.
- **RNF03:** O sistema não deve efetuar o cálculo da similaridade enquanto o usuário acessa um objeto específico, mas sim, automaticamente uma vez ao dia em horário em que hajam poucos acessos ao repositório.

- **RNF04:** O sistema deve salvar a lista de recomendações no banco de dados, para que esta seja consultada quando um usuário acessa um objeto de aprendizagem, e então se faça a recomendação.

### 3.1.3 Regras de negócio

A seguir são apresentadas as principais regras de negócio do sistema.

- **RN01:** O cálculo de similaridade entre os OAs deve ser feito de forma transparente ao usuário, e não durante seu uso do sistema, com o intuito de tornar o acesso as recomendações mais rápido.
- **RN02:** A lista com as recomendações deve conter 20 itens, somente.

## 3.2 Modelagem do sistema

Nesta seção são apresentados os diagramas e artefatos gerados durante a modelagem do sistema. Os diagramas escolhidos foram:

- Modelo ER: que objetivou entender o banco atual do repositório, bem como as novas tabelas que foram geradas.
- Diagrama de Classes: artefato utilizado para facilitar o desenvolvimento das classes que foram criadas, suas ligações e seus métodos.

### 3.2.1 Modelo Lógico do Banco de Dados

Na Figura 16 visualiza-se o modelo lógico do banco de dados contendo as tabelas que foram criadas para o funcionamento do sistema. Foram elas: a tabela “*recommendations*” que armazena os registros dos valores que são obtidos através do cálculo de similaridade, e a tabela “*learning\_objects*”, que trata-se de uma *view* criada para facilitar a busca das informações que foram necessárias para calcular a similaridade entre os OA, para tal foram utilizadas informações de outras tabelas do repositório como mostrado na Figura 17.

The image shows two database table schemas. The 'recommendations' table has columns: id INT, learning\_object\_id INT, recommendation\_id INT, keywords DOUBLE, description DOUBLE, educational\_description DOUBLE, title DOUBLE, learning\_resource\_type DOUBLE, level DOUBLE, and general DOUBLE. The 'learning\_objects' table has columns: id INT, title VARCHAR(255), description TEXT, educational\_description TEXT, keywords TEXT, learning\_resource\_type VARCHAR(255), and level INT(10).

Table	Column	Data Type
recommendations	id	INT
	learning_object_id	INT
	recommendation_id	INT
	keywords	DOUBLE
	description	DOUBLE
	educational_description	DOUBLE
	title	DOUBLE
	learning_resource_type	DOUBLE
	level	DOUBLE
	general	DOUBLE
learning_objects	id	INT
	title	VARCHAR(255)
	description	TEXT
	educational_description	TEXT
	keywords	TEXT
	learning_resource_type	VARCHAR(255)
	level	INT(10)

**Figura 16: Modelo Lógico das tabelas criadas**

As tabelas do repositório (Figura 17) que foram utilizadas são as seguintes: (i) *learningobject*, que armazena informações dos objetos cadastrados; (ii) *general*, a qual foi utilizada para recuperar o título dos objetos de aprendizagem; (iii) *keyword*, utilizada para recuperar as palavras chave de cada objeto; (iv) *learningresourcetype*: que armazena os tipos de recurso educacional que um objeto contém; (v) *educational\_level*: armazena o nível educacional do OA; e (vi) *description*: a qual armazena a descrição do objeto.

The image shows six database table schemas. The 'learningobject' table has columns: idlearningobject INT(10), ip VARCHAR(255), arquivoobaa TEXT, idperson INT(11), fl\_web TINYINT(4), fl\_dtv TINYINT(4), fl\_mobil TINYINT(4), fl\_repositorio VARCHAR(20), and dt\_envio DATETIME. The 'general' table has columns: idgeneral INT(10), idlearningobject INT(10), title VARCHAR(255), structure VARCHAR(255), and aggregationlevel VARCHAR(255). The 'educational\_level' table has columns: ideducational\_level INT(10), ideducational INT(10), and idlevel INT(10). The 'keyword' table has columns: idkeyword INT(10), idclassification INT(10), idgeneral INT(10), and value VARCHAR(255). The 'learningresourcetype' table has columns: idlearningresourcetype INT(10), ideducational INT(10), and value VARCHAR(1000). The 'description' table has columns: iddescription INT(10), idresource INT(10), ideducational INT(10), idgeneral INT(10), and value VARCHAR(1000).

Table	Column	Data Type
learningobject	idlearningobject	INT(10)
	ip	VARCHAR(255)
	arquivoobaa	TEXT
	idperson	INT(11)
	fl_web	TINYINT(4)
	fl_dtv	TINYINT(4)
	fl_mobil	TINYINT(4)
	fl_repositorio	VARCHAR(20)
	dt_envio	DATETIME
general	idgeneral	INT(10)
	idlearningobject	INT(10)
	title	VARCHAR(255)
	structure	VARCHAR(255)
	aggregationlevel	VARCHAR(255)
educational_level	ideducational_level	INT(10)
	ideducational	INT(10)
	idlevel	INT(10)
keyword	idkeyword	INT(10)
	idclassification	INT(10)
	idgeneral	INT(10)
	value	VARCHAR(255)
learningresourcetype	idlearningresourcetype	INT(10)
	ideducational	INT(10)
	value	VARCHAR(1000)
description	iddescription	INT(10)
	idresource	INT(10)
	ideducational	INT(10)
	idgeneral	INT(10)
	value	VARCHAR(1000)

**Figura 17: Tabelas utilizadas para criar a view com os campos que serão considerados no cálculo de similaridade**

### 3.2.2 Diagrama de classes

As classes criadas foram (Figura 18):

- *Similarities*: responsável preparar os dados dos objetos de aprendizagem vindos do banco, pré-processamento dos textos dos mesmos, adição dos termos ao índice e cálculo de similaridade entre os objetos de aprendizagem.
- *Terms*: Classe que guarda os termos indexados e é responsável por calcular as frequências dos termos, frequências inversas, número de objetos de aprendizagem indexados e outros valores relevantes para o cálculo de similaridade;
- *Distance*: Responsável por calcular os valores de similaridade para os campos nível e tipo de recurso educacional.
- *Recommendation*: responsável por salvar, deletar e buscar os dados das recomendações no banco;
- *StopWords*: Armazena a lista de *StopWords* bem como encontrar se uma palavra faz parte desta lista;
- *Stemmer*: responsável pelo processo de *stemming* dos termos antes dos mesmos serem adicionados ao índice.

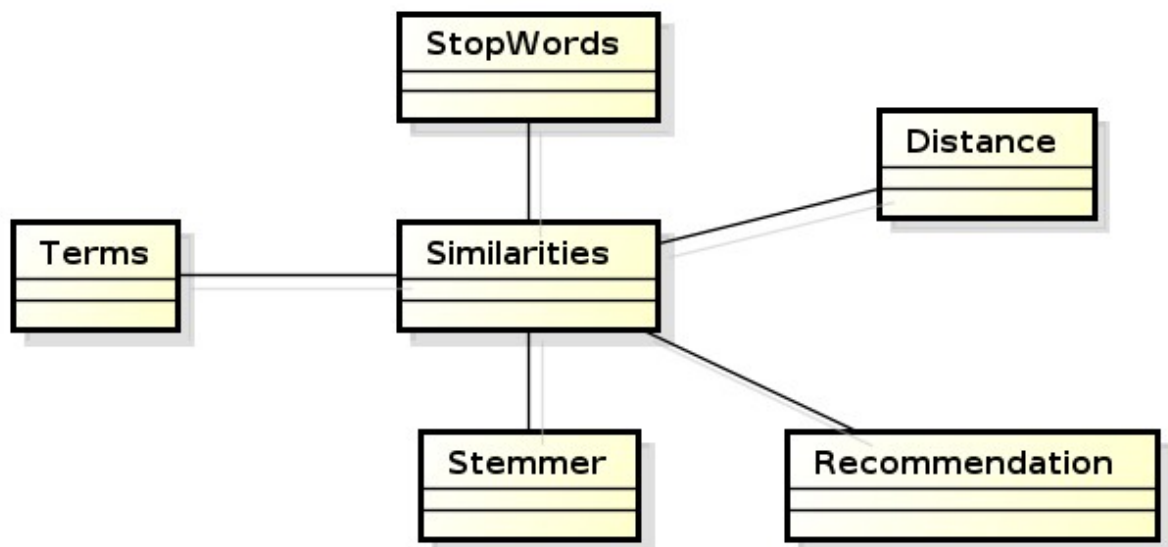


Figura 18: Diagrama de Classes

O diagrama completo encontra-se no APÊNDICE - B.

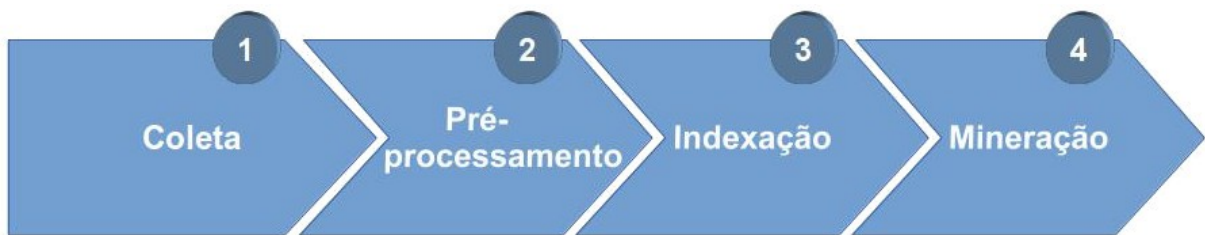
### 3.3 Desenvolvimento

O processo de recomendação está organizado da seguinte forma: (i) extração de informação; e (ii) visualização da recomendação. Estes passos, são descritos a seguir.

#### 3.3.1 Extração da informação

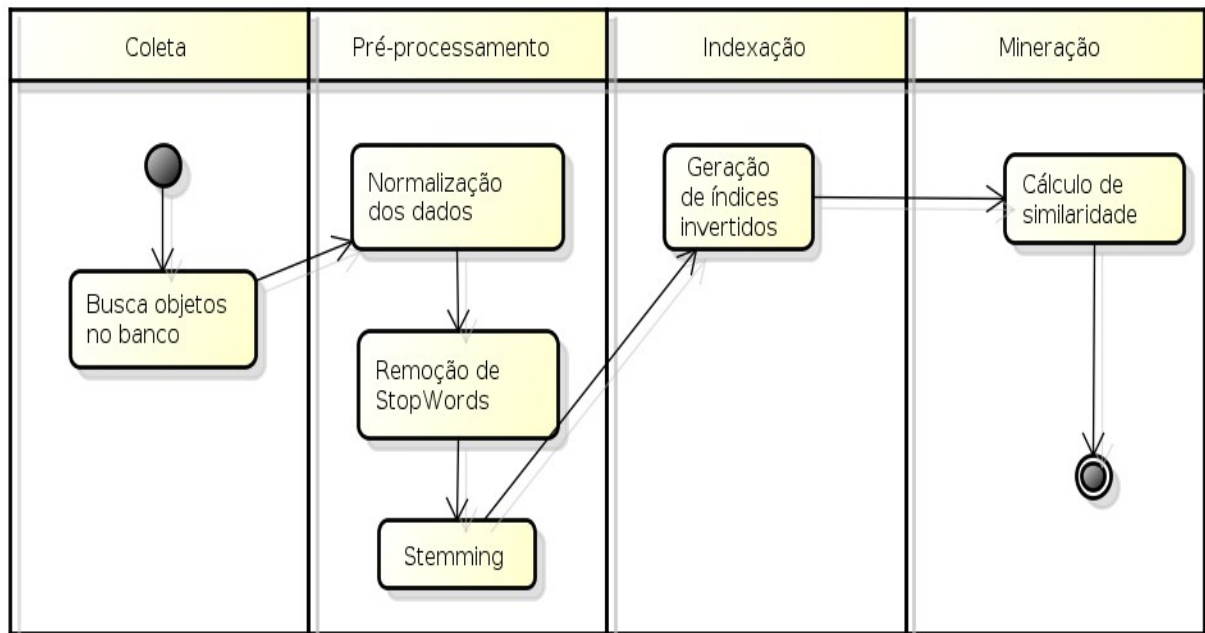
A extração da informação ou mineração dos dados foi feita seguindo-se os passos mostrados na Figura 19, tais passos têm os seguintes objetivos:

1. Coleta: Formação da base de documentos ou *Corpus*;
2. Pré-processamento: Preparação dos dados;
3. Indexação: Objetiva o acesso rápido ou busca rápida dos dados;
4. Mineração: Cálculos, inferência e extração de conhecimento.:



**Figura 19: Passos para Extração da Informação**

Em cada passo são executadas as tarefas apresentadas na Figura 20, onde primeiramente são buscados dados dos objetos de aprendizagem no banco de dados, após os dados são normalizados, removidas as *StopWords* e executa-se o processo de *stemming*, então geram-se os índices com os termos e calcula-se as similaridades entre os objetos de aprendizagem.



**Figura 20: Processo de extração da informação**

Cada um dos passos mostrados na Figura 20 é descrito nas seções a seguir.

### 3.3.1.1 Coleta

Para coleta dos dados, foi criada uma *view* no banco de dados, com o intuito de facilitar a recuperação dos textos e dados sobre os objetos, que foram escolhidos para serem analisados. Nela estão contidos:

- Título dos objetos;
- Palavras-chave: estas encontravam-se em vários registros no banco e foram unidas em um campo somente;
- Descrição;
- Descrição educacional;
- Tipo de recurso educacional;
- Nível;

O campo cobertura (*coverage*) foi removido, pois não estava preenchido na maioria dos OAs, o que não contribuiria para a análise do conteúdo dos OAs e cálculo de similaridade.



### 3.3.1.2 Pré-processamento

Conforme Silla e Kaestner (2002), esta etapa tem por objetivo tratar os documentos para que os mesmos tenham uma representação melhor estruturada, de modo a facilitar a etapa de mineração. Para tal, são executadas as seguintes tarefas: (i) normalização do texto; (ii) remoção de *stop words* e (iii) *stemming*. Estas três etapas são descritas a seguir.

#### 3.3.1.2.1 Normalização

Primeiramente foi feito o *Case Folding*, o qual, segundo Silla e Kaestner (2002), é o processo de conversão de todos os caracteres de um mesmo documento para um formato comum. Tal formato comum, pode ser todo o texto em letras maiúsculas ou em letras minúsculas. Por exemplo, as palavras “Objeto”, “oBjEtO”, “OBjEtO”, “oBEJTO” seriam todas convertidas ou para “OBJETO” ou para “objeto”. Para o presente trabalho, foi considerada a segunda opção, ou seja, todas as palavras em minúsculas.

Após, são removidos símbolos que não tenham importância para o texto do documento, como por exemplo: caracteres de pontuação, parênteses, colchetes, chaves, barras e barras invertidas, nova linha, números, etc (., ; : " ' ! ? / \ | % + { } [ ] ( ) \*). Nesta etapa também pode ser feita a remoção de acentuação, porém para o presente trabalho, esta é feita posteriormente.

Feito isso, é aplicada a técnica conhecida como *Tokenize*, a qual, segundo Amazonas (2008), consiste na separação e armazenamento de cada *token* ou termo do texto em um vetor. Nesta etapa, também é possível o registro de informações relevantes sobre cada um deles. Ao fim da etapa de *Tokenize*, obtém-se uma lista com todos os *tokens* de um documento. Por exemplo, após esta etapa, a frase “Vídeo com a definição de Função Afim.” seria o vetor  $w =$  (“vídeo”, “com”, “a”, “definição”, “de”, “função”, afim”).

Após esta etapa é feita a remoção de *Stop Words*, que é descrita a seguir.

#### 3.3.1.2.2 Stop Words

De acordo com RAJARAMAN e ULLMAN (2012), em diversas aplicações para mineração de dados existe o problema de categorização dos documentos (sequências de palavras) por seus temas. Temas são identificados encontrando-se palavras especiais que caracterizam um documento em tal tema. Por exemplo: em um documento sobre objetos de aprendizagem palavras como “Educação”, “Informática”, “Internet”, “Objetos de

Aprendizagem”, “Repositórios”, etc tenderiam a ter uma maior ocorrência. E, após termos classificado o documento por seu tema como sendo sobre objetos de aprendizagem, não é difícil notar que tais palavras aparecem de forma mais frequente. Porém, antes de termos identificado o tema, não é possível identificar estas palavras como sendo características do documento.

Deste modo, a classificação comumente começa pela análise dos documentos e identificação das palavras que tenham alguma relevância para o mesmo. O que nos leva a crer que palavras que aparecem com maior frequência são as mais significativas para o documento. Porém a realidade é o oposto disso. As palavras que aparecem com maior frequência são palavras comuns como artigos, preposições, pronomes, advérbios, conjunções, entre outras, como por exemplo “o”, “a”, “seus”, “ela”, “e”, as quais não são significativas quando consideradas isoladamente. Tais palavras são conhecidas como *Stop Words* e por sua pouca relevância são comumente removidas dos documentos, para facilitar a sua classificação. Isto é feito criando-se uma *stoplist*, na qual as são armazenadas inúmeras *stopwords*, para que estas sejam removidas ou desconsideradas ao se processar o texto de um documento. Com isto a quantidade de *tokens* é reduzida consideravelmente. (RAJARAMAN e ULLMAN, 2012; Yamada *et al.*, 2012)

Considerando o vetor de termos  $w = (\text{“vídeo”, “com”, “a”, “definição”, “de”, “função”, afim})$ , após esta etapa, o mesmo seria reduzido para  $w = (\text{“vídeo”, “definição”, “função”, afim})$ .

Feita a remoção de *Stop Words*, é feita a operação de *stemming* para cada *tokens* do vetor, tal operação é descrita a seguir.

### 3.3.1.2.3 *Stemming* ou Normalização Morfológica

*Stemming* é um processo utilizado para reduzir a quantidade de *tokens* necessários para representar uma coleção de documentos. Ele consiste em uma normalização linguística, na qual as formas variantes de um termo são reduzidas a uma forma comum denominada *stem*.

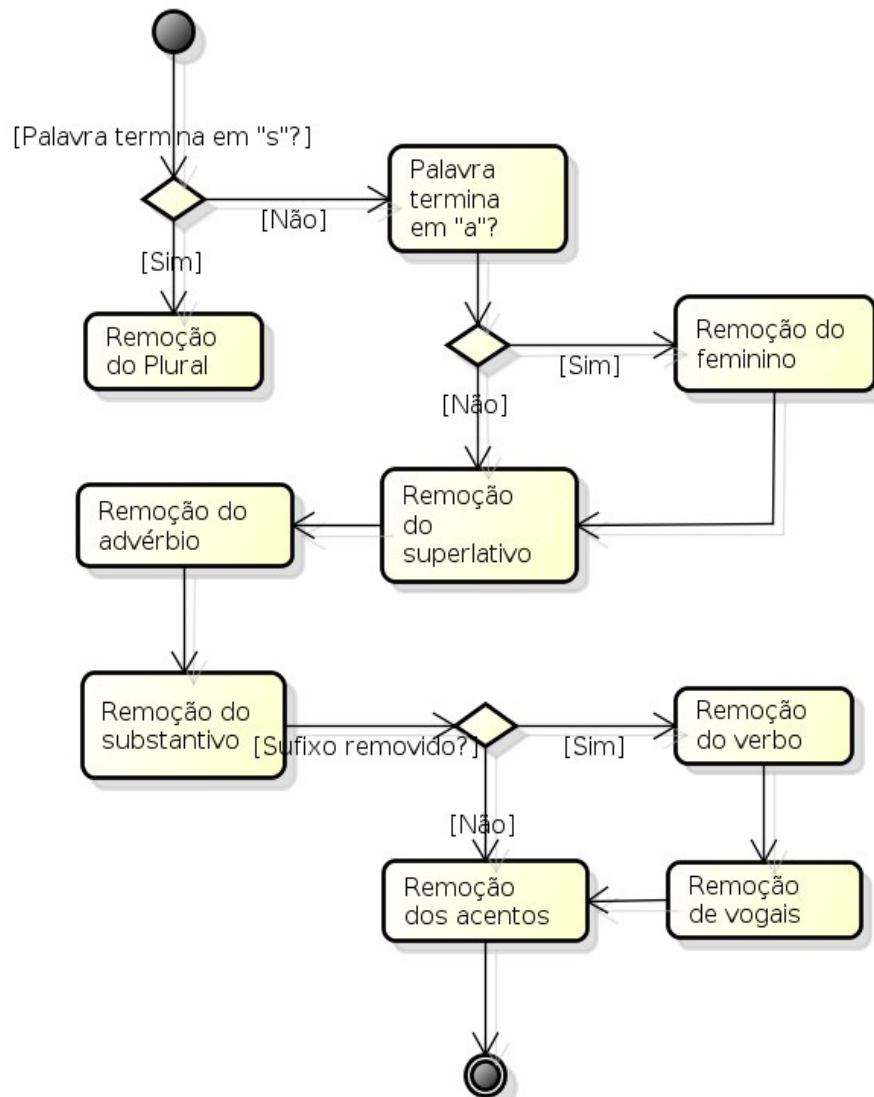
Este processo converte cada palavra para o seu radical, eliminando sufixos representados por flexões verbais, plural e gênero das palavras. Os algoritmos de *Stemming* geralmente incorporam um grande conhecimento de linguística, por essa razão, e são fortemente dependentes de linguagem (idioma) no qual os documentos estão escritos (Silla e Kaestner, 2002; Yamada *et al.*, 2012).

Por exemplo as palavras “brasileira”, “brasileiro”, “brasileirinha”, “brasileiresco”, “brasileiras”, “brasileiros”, podem ser transformadas em um mesmo *stem* “brasil”.

São exemplos de algoritmos para stemming: método do *stemmer* S; método de Porter; método de Lovins, Ebecken (2003) apud Moraes e Ambrósio (2007).

- método do *stemmer* S: Este é considerado o método mais simples, pois consiste na eliminação de apenas alguns finais de palavras. Geralmente remove somente sufixos que formam o plural.
- método de Porter : Consiste na identificação de diferentes inflexões referentes à mesma palavra e sua substituição por um radical comum. Este algoritmo remove cerca de 60 sufixos diferentes para palavras da língua inglesa e é baseado nas seguintes etapas [2]: redução do plural, troca de sufixos, retirada de sufixos, remoção de sufixos padrões e remoção da vogal “e” ao final da palavra.
- método de Lovins : Este método remove cerca de 250 sufixos diferentes para palavras da língua inglesa. Este algoritmo remove apenas um sufixo por palavra, retirando o sufixo mais longo conectado à mesma.

Outro exemplo, similar ao utilizado para no presente trabalho, foi proposto por Orengo e Huyck (2001), que segue as etapas mostradas na Figura 21.



**Figura 21: Passos do Stemmer para a língua portuguesa**

**Fonte:** Adaptado de Moraes e Ambrósio (2007)

As etapas mostradas mostradas na Figura 21 tem as seguintes finalidades (Orengo e Huyck, 2001):

1. Remoção do plural : Exceto por algumas exceções, a maior parte das palavras em português termina em “s”. Apesar disso, nem todas as palavras terminadas em “s” denotam o plural, como é o caso de “lápiz”. Sendo assim, este passo consiste basicamente em remover o “s” do final das palavras que não são consideradas exceções. Porém algumas vezes é necessário realizar correções em algumas palavras como por exemplo “bons”, na qual a terminação “ns” é substituída por “m”.
2. Remoção do feminino : Em português todos os substantivos e adjetivos possuem um gênero. Esta etapa consiste em transformar palavras no feminino em sua forma no

masculino.

3. Remoção de advérbio : Esta é a etapa mais curta de todas, uma vez que o único sufixo que denota um advérbio é “mente”. Porém para este caso também há uma lista de exceções, pois nem todas as palavras terminadas em “mente” são advérbios.
4. Remoção de aumentativo e diminutivo : No português as palavras possuem aumentativo, diminutivo e superlativo. Deste modo, neste passo é feita a remoção dos sufixos dos substantivos e adjetivos que podem ter aumentativo e diminutivo. Por exemplo, “gatinha” ou “menininha”.
5. Remoção de sufixos em nomes : Esta etapa testa as palavras contra uma lista de 61 sufixos para substantivos e adjetivos. Caso o sufixo seja removido, as etapas seis e sete não serão executadas.
6. Remoção de sufixos em verbos : Os verbos regulares da língua portuguesa possuem mais de 50 formas diferentes de conjugação. Cada forma possui seu sufixo específico. Neste caso, os verbos podem variar de acordo com o tempo, a pessoa, o número e o modo. A estrutura das formas verbais pode ser representada por: radical + vogal temática (a, e, i conforme as terminações ar, er, ir) + tempo + pessoa. Por exemplo: “and + a + ra + m”. As formas verbais são reduzidas ao seu radical.
7. Remoção de vogais : Esta etapa consiste em remover a última vogal (“a”, “e” ou “o”) das palavras que não foram reduzidas nas etapas cinco e seis. Por exemplo: “menino”, não sofreria nenhuma modificação nos dois passos anteriores, então neste passo será removido o final “o”, ficando “menin”, de modo a permanecer o mesmo *stem* que suas variações (menina, meninice, menino, menininho).
8. Remoção de acentos : Remoção de acentos se faz necessária, pois existem casos em que algumas variantes de palavras são acentuadas e outras não, como por exemplo “psicólogo” e “psicologia”. Após este passo ambos seriam representados por “psicolog”. É importante que a execução deste passo se faça no final, pois a presença de acentos é significativa em para regras, por exemplo: “óis” para “ol” (em que “sóis” se transforma em “sol”). Se a regra fosse mudada de “ois” para “ol”, causaria erros como no caso da palavra “dois” que se tornaria “dol”.

Após este processo, no caso do vetor de *tokens*  $\mathbf{w}$  = (“vídeo”, “definição”, “função”,

afim”), este passaria a ser  $w = (“vid”, “defin”, “func”, “afim”)$ .

### 3.3.1.3 Indexação

De acordo com Amazonas *et al.* (2008), o processo de indexação consiste basicamente em mapear os termos de um documento em uma estrutura de dados específica chamada de índice. Segundo eles, os tipos de índices mais comuns são o Índice Invertido e o Índice Sequencial. O Índice Invertido consiste no mapeamento de cada *token* para a lista de documentos aos quais ele pertence. Este processo exige que sejam pesquisados todos os *tokens*, de todos os textos, a serem inseridos ou alterados, sendo assim, pode gerar um aumento de tempo considerável (*overhead*) no processo de indexação.

Um exemplo de Índice Invertido pode ser observado na Figura 22, a qual apresenta uma lista de *tokens*, à esquerda, apontando para a lista de documentos que o possuem (à direita).

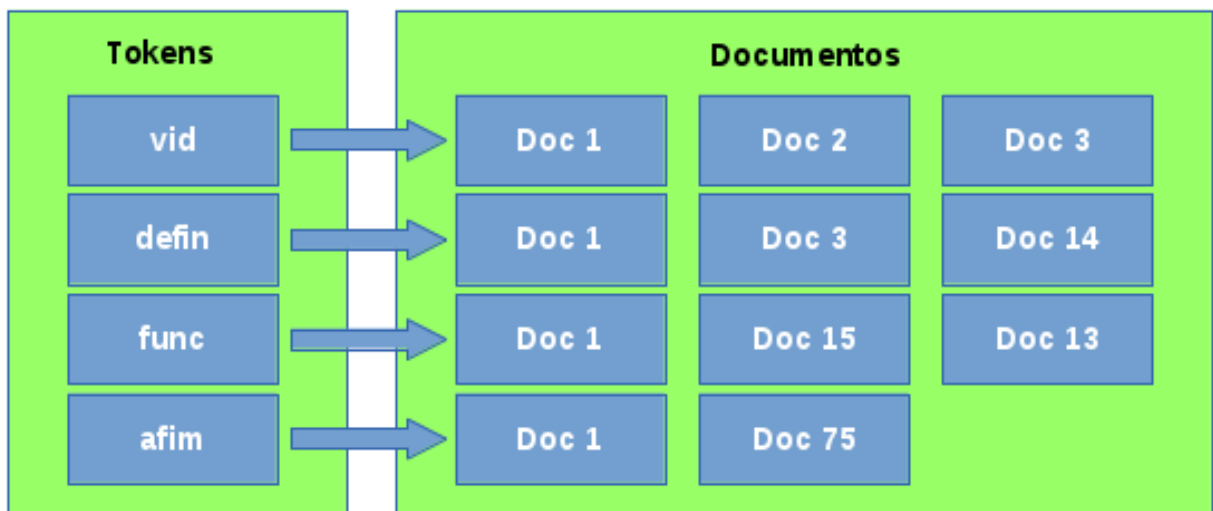
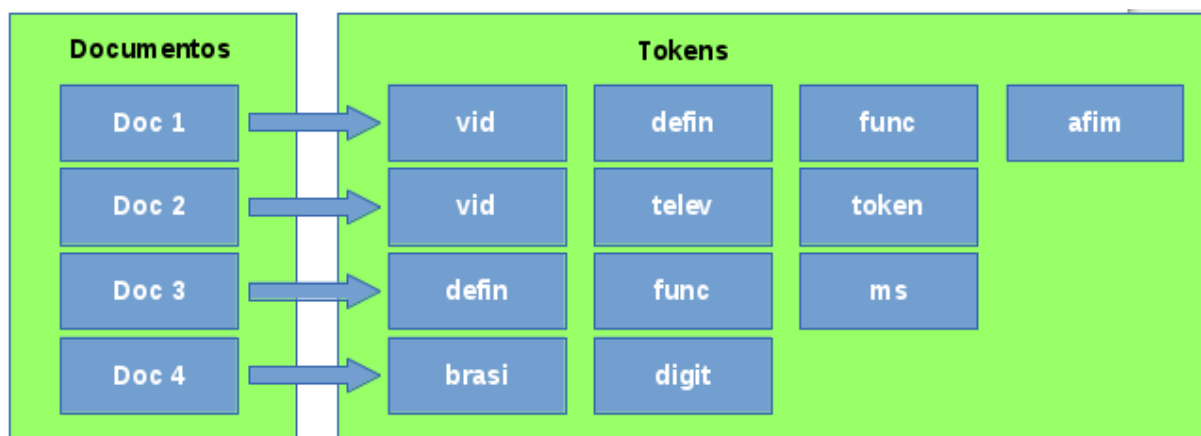


Figura 22: Exemplo de Índice Invertido

Fonte: Adaptado de Amazonas *et al.* (2008)

Ainda segundo Amazonas, o Índice Sequencial consiste de uma lista de pares (documento, *token*) ou (documento, lista de *tokens*), ordenados pelos documentos. Segundo os autores, esta abordagem faz com que seja reduzido o *overhead* na inclusão e atualização existente nos Índices Invertidos. No Índice Sequencial, as inserções sempre ocorrem no final da lista, e as atualizações ocorrem diretamente no par do documento a ser atualizado. Porém, um Índice Sequencial pode ser transformado em Índice Invertido em tempo de execução, isto com o intuito de reduzir o tempo de pesquisa. Um exemplo de Índice Sequencial é apresentado na Figura 23, a qual apresenta à esquerda uma lista de documentos indexados, os

quais apontam para suas respectivas listas de *tokens*.



**Figura 23: Exemplo de índice sequencial**

Fonte: adaptado de Amazonas *et al.* (2008)

Para o presente trabalho, o índice gerado é um Índice Invertido. Tal índice guarda as seguintes informações:

- Lista dos *tokens*;
- O número total de OAs;
- A frequência que cada *token* aparece em cada documento;
- O campo (palavras-chave, título, descrição, ou descrição educacional) em que o *token* aparece no OA;
- A frequência que o *token* aparece em cada campo de cada OA;
- O número total de *tokens* em cada objeto OA;
- O número total de OAs em que o *token* aparece.

Após a geração dos índices, vem a etapa de mineração de dados, que é descrita no tópico a seguir.

#### **3.3.1.4 Mineração**

A mineração de dados é um processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis disponíveis nos dados. O principal objetivo desta etapa é a descoberta de co-relacionamentos entre os dados e seus documentos (Morais e Ambrosio, 2007). Tal processo é também conhecido como extração de conhecimento.

Com os documentos estruturados em índices, o conhecimento que é objetivado por este trabalho é a similaridade entre dois objetos de aprendizagem distintos. Tal processo é feito através de cálculos, e é descrito na próxima seção .

### 3.3.2 Cálculo de Similaridade

O algoritmo de recomendação criado é classificado, por sua abordagem, como sendo baseado em conteúdos. Além disso, ele utiliza o Modelo de Espaço Vetorial (VSM) para o cálculo de similaridade nos campos título, palavras-chave, descrição e descrição educacional, e, calcula a distância entre os OAs para os campos tipo de recurso educacional e nível.

O cálculo se dá através da Equação 6, onde *textFieldsSimilarity* representa a similaridade para os campos textuais e tem peso 0.7 na média final, enquanto que *distanceFieldsSimilarity* representa a distância entre OAs para os demais campos e possui peso 0.3. O processo para obtenção dos mesmos é detalhado a seguir.

$$\text{Similarity}(Q, D) = 0.7 * \text{textFieldsSimilarity} + 0.3 * \text{distanceFieldsSimilarity}$$

Equação 6: Similaridade entre dois OAs

Nas seções a seguir é apresentado como os valores de *textFieldsSimilarity* e *distanceFieldsSimilarity* são obtidos.

#### 3.3.2.1 TextFieldsSimilarity

Estes são calculados baseados no VSM, o qual os trata como sendo vetores de valores reais contendo os pesos de cada termo ( $w_{ij}$ ) dos documentos. A matriz é escrita na forma  $d_i = (w_{i1}, w_{i2}, \dots, w_{in})$ , onde  $w_{ij}$  corresponde ao peso do  $j$ -ésimo ( $j = 1, 2, \dots, n$ ) termo do  $i$ -ésimo documento ( $i = 1, 2, \dots, n$ ). Sendo assim, a coleção de objetos é representada como sendo uma matriz ( $D$ ) de termos de tamanho  $n \times m$  onde  $n$  é o número de documentos e  $m$  o número de termos.

Para o cálculo do peso de um termo  $j$  em um documento  $i$  é utilizado o método  $tf*idf$ . Nele,  $tf$  representa a frequência do termo  $j$  (*term frequency*) no documento  $i$ , ou seja, o número de vezes que o termo aparece no documento, e,  $idf$  representa a frequência inversa do documento (*inverse document frequency*), ou seja, o número de documentos em que o  $j$  termo aparece. O Cálculo para obtenção do  $idf$  é feito utilizando-se a Equação 7. Nela,  $n$  representa o número de objetos (documentos) na coleção e  $df_j$  o número de documentos em que o termo  $j$



aparece.

$$idf = \log_2\left(\frac{n}{df_j}\right)$$

**Equação 7: Cálculo da frequência inversa do documento**

**Fonte: Adaptado de Korenius, Laurikkala e Juhola (2007)**

Após o cálculo de frequência do termo e de frequência inversa do documento é calculado o peso  $w_{ij}$  de cada termo. Este, para um termo  $j$  em um documento  $i$  é calculado multiplicando-se a  $tf_j$  por  $idf_j$ , como é mostrado na Equação 8.

$$w_{ij} = tf_{ij} \cdot idf_{ij}$$

**Equação 8: Cálculo do peso dos termos**

**Fonte: Adaptado de Korenius, Laurikkala e Juhola (2007)**

Após obtido o peso de cada termo é calculada a similaridade entre os dois OAs. Isto pode ser feito utilizando-se a Equação 3, apresentada anteriormente. Porém a equação considera somente um texto único, e, no caso dos OAs são utilizados vários campos textuais, os quais acredita-se possuírem diferentes relevâncias para o conteúdo do OA, desta forma foram adicionados pesos para cada campo, como forma de avaliar de forma mais correta o conteúdo do mesmo. Os pesos são apresentados no Quadro 4.

**Quadro 4: Pesos dos campos**

<b>Campo</b>	<b>Peso</b>
Palavras-chave	0.4
Descrição	0.3
Título	0.2
Descrição educacional	0.1

O cálculo da similaridade com os pesos é feito separadamente para cada campo, utilizando-se a Equação 9, onde, para calcular a similaridade entre os diferentes campos de dois objetos  $Q$  e  $D$ , é feita a somatória do produto entre os pesos ( $w_{qfk} * w_{dfk}$ ), obtidos em ambos objetos, de todos os termos ( $fk$ ) no campo  $f$ . E após divide-se este valor, pela raiz quadrada do produto das somatórias dos pesos dos de cada termo elevado ao quadrado ( $w_{qfk}^2$ ).

$$FieldSim(Q_f, D_f) = \frac{\sum_{k=1}^{ft} w_{qfk} \cdot w_{dfk}}{\sqrt{\sum_{k=1}^{ft} (w_{qfk})^2 \cdot \sum_{k=1}^{ft} (w_{dfk})^2}}$$

**Equação 9: Similaridade entre os campos**

**Fonte: Adaptado de Lopes, 2007**

Após encontrados os valores da similaridade de cada campo, estes são multiplicados pelos seus respectivos pesos (Quadro 4), e, por fim, somados para representar a similaridade entre os dois OAs.

### **3.3.2.2 DistanceFieldsSimilarity**

Para os campos nível (*level*) e tipo de recurso educacional (*learning resource type*) são calculadas as distâncias entre os valores dos mesmos para calcular a similaridade, sendo o resultado obtido com a Equação 10. Na equação, é possível observar que a similaridade para o campo nível (*lvlSimilarity*) tem peso 0.8 enquanto que tipo de recurso educacional (*lrtSimilarity*) tem peso 0.2. Isto pois acredita-se que o tipo de recurso educacional é um atributo que tenha menor relevância que o nível. Os cálculos para cada campo são descritos separadamente mas adiante nesta seção.

$$DistanceFieldsSimilarity = lvlSimilarity * 0.8 + lrtSimilarity * 0.2$$

**Equação 10: Similaridade para os campos utilizando a distância**

Para execução do cálculo foram seguidos os seguintes passos:

1. Identificação dos possíveis valores para os campo nível e tipo de recurso educacional;
2. Ordenação de seus valores, conforme sua relação entre si;
3. Recuperação dos valores para cada objeto;
4. Cálculo da distância.

Os valores identificados para o campo nível são apresentados no Quadro 5.

**Quadro 5: Valores que um OA pode conter para o campo nível**

Nível	valor
infantil	1
fundamental - séries iniciais	2
fundamental - séries finais	3
EJA	4
médio	5
técnico	6
superior - social aplicada	7
superior – humanas	8
superior – comunicação	9
superior – jurídica	10
superior – exatas	11
superior – biológicas	12
superior – terra	13

Para o campo nível, o cálculo da similaridade é feito utilizando-se a Equação 11, onde o primeiro passo a ser executado é o cálculo da distância entre os valores dos dois objetos ( $val_a - val_b$ ), então calcula-se o módulo deste valor, para que o valor da distância seja sempre positivo. Após diminui-se este valor de **maxDistance**, que representa o maior valor de distância possível para o campo, este é igual ao número de casos possíveis menos 1, para o caso do campo nível o valor de **maxDistance** é igual a 12. Por fim, divide-se o valor obtido pelo **maxDistance**, a fim de normalizar o resultado. O valor obtido através desta equação será entre 0 e 1, onde, 0 representa a menor similaridade possível e 1 a maior similaridade.

$$DistanceSimilarity(val_a, val_b) = \frac{maxDistance - |val_a - val_b|}{maxDistance}$$

**Equação 11: Cálculo da similaridade utilizando a equação de distância**

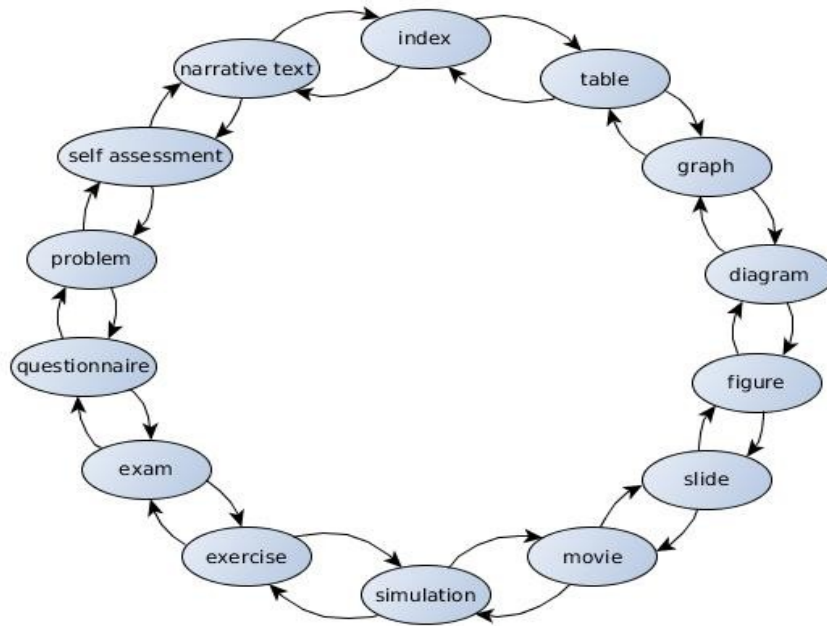
Os valores identificados para o campo tipo de recurso educacional são apresentados no Quadro 6.

**Quadro 6: Valores que um OA pode conter para o campo tipo de recurso educacional**

<b>Tipo de recurso educacional</b>	<b>valor</b>
index	1
table	2
graph	3
diagram	4
figure	5
slide	6
movie	7
simulation	8
exercise	9
exam	10
questionnaire	11
problem	12
self assessment	13
narrative text	14

Diferente do campo nível que possui uma lista sequencial de casos, para tipo de recurso educacional foi considerado que o campo possui uma lista de casos circular, ou seja, o primeiro e o último itens não são os mais distantes, mas sim um encontra-se como sendo o próximo item da lista em relação ao outro. Por exemplo, considerando-se dois objetos A e B, caso A possua como valor “*index*” e B “*narrative text*”, a distância entre os mesmos não será 13 e sim 1, pois “*index*” é o próximo item da lista em relação a “*narrative text*”.

Além disso para percorrer a lista, é considerada uma estrutura de dados duplamente encadeada, ou seja, é possível percorrer a lista em ambas as direções. Deste modo a distância considerada é a com menor valor, dentre as duas encontradas percorrendo-se a lista em ambas as direções. Sendo assim, caso A e B, possuam como valor “*movie*” e “*exam*”, respectivamente as distâncias encontradas seriam 3 e 11, logo, considera-se como distância o valor 3. A Figura 24 mostra uma representação da lista circular duplamente encadeada contendo os tipos de recurso educacional.



**Figura 24:** Lista circular duplamente encadeada dos tipos de recurso educacional

Porém, um objeto de aprendizagem pode possuir nenhum ou muitos tipos de recurso educacional, e deste modo não é possível calcular distância da mesma forma que para o campo nível. O cálculo da distância entre dois objetos com mais de um tipo de recurso educacional é feito utilizando-se a Equação 12.

$$lrtSimilarity(Q, D) = \frac{\sum_{i=1}^{lrt} maxSimQ_i D}{lrt}$$

**Equação 12:** Cálculo da similaridade para o campo tipo de recurso educacional

Considerando um objeto **Q** = (“slide”, “graph”), e um objeto **D** = (“slide”, “figure”, “simulation”). Para cada tipo de recurso presente em **Q**, este é primeiramente comparado com todos os itens presentes em **D**, obtendo-se os valores das distâncias. Para o caso de *slide*, comparando-se com os itens em **D** (“slide”, “figure”, “simulation”), obtém-se como valores de distância 0, 1 e 2, já para *graph*, obtém-se 3, 2 e 5, respectivamente. Após é calculada a similaridade, utilizando a Equação 12. Obtendo-se 1, 0.92307692307692 e 0.84615384615385 para *slide* e 0.76923076923077, 0.84615384615385 e 0.61538461538462 para *graph*. Então para cada item em **D**, considera-se como similaridade o maior valor dentre os encontrados (**maxSimQ<sub>i</sub>D**), ou seja 1 e 0.84615384615385. Estes são somados resultando em 1.8461538461538 e após divididos pelo número de itens em **D** que é de 2, sendo o valor

final  $lrtSimilarity(Q, D) = 0.92307692307692$ .

### 3.3.3 Visualização da recomendação

O processo de recomendação ocorre quando um usuário acessa a página de detalhes de algum objeto específico. Para a visualização das recomendações optou-se por utilizar a estratégia de listas de recomendação, como mostrado na Figura 25, onde são exibidos os 20 itens mais similares ao item que está sendo visualizado pelo usuário. O nome dado a lista foi de “Objetos similares a este”.

Objetos similares a este
<a href="#">Graficando</a>
<a href="#">Traffic Light System</a>
<a href="#">Ponto comum de duas funções de 1º Grau</a>
<a href="#">Explorando Funções Polinomiais de 1º Grau</a>
<a href="#">Introdução á Disciplina de Informática</a>
<a href="#">Cyclotron</a>
<a href="#">A Estrutura Espacial das Moléculas</a>
<a href="#">Decole do Porta Aviões</a>
<a href="#">Ley de Coulomb, Física Divertida</a>
<a href="#">Aprendendo Matrizes - através de campeonatos de futebol</a>
<a href="#">Navegando em Alto Mar</a>
<a href="#">Embalagens para Sorvete</a>
<a href="#">Mamíferos</a>
<a href="#">Movimento Unidimensional: Deslocamento, Velocidade e Aceleração</a>
<a href="#">Moléculas</a>
<a href="#">Cartazes de Dom Casmurro</a>
<a href="#">Quebra-Cabeça das Frações</a>
<a href="#">Tabela de Dados - Classificar e Filtrar</a>
<a href="#">Potência</a>
<a href="#">Soma de frações</a>

Figura 25: Lista de recomendações

## 4 RESULTADOS

Nesta seção são apresentados os resultados do experimento feito com os participantes. No experimento foi testada a precisão do algoritmo de recomendação e além disso, foram avaliados como os usuários utilizam o sistema com e sem recomendações.

Para a execução dos testes, os participantes acessaram um sistema que foi construído com o intuito de facilitar a coleta dos dados necessários para o presente trabalho, tal sistema será chamado aqui de Sistema de Tarefas. A Figura 26 mostra a página inicial do Sistema de Tarefas, onde, foi apresentado aos participantes o convite para participação nos testes que foram executados. Como apresentado na figura, as tarefas foram divididas em 4 passos: (i) formulário de perfil dos participantes; (ii) avaliação das recomendações geradas pelo sistema; (iii) tarefa de busca 1; (iv) e tarefa de busca 2.

Bem vindo

Início Passo 1 Passo 2 Passo 3 Passo 4

**Caro aluno / professor / colaborador**

Você está sendo convidado para auxiliar na Avaliação de uma ferramenta construída durante o desenvolvimento do meu TCC.

O objetivo deste trabalho é entender como os usuários interagem com a ferramenta, que se trata de um sistema de recomendação (sistema que recomenda conteúdos para usuários).

Tal sistema foi aplicado em um repositório de objetos de aprendizagem (recursos educacionais).

Você é livre para participar ou não da atividade proposta.

Suas informações não serão divulgadas em nenhum meio. Somente serão utilizadas para entender e avaliar a ferramenta.

[Clique aqui caso você aceite participar da Avaliação](#)

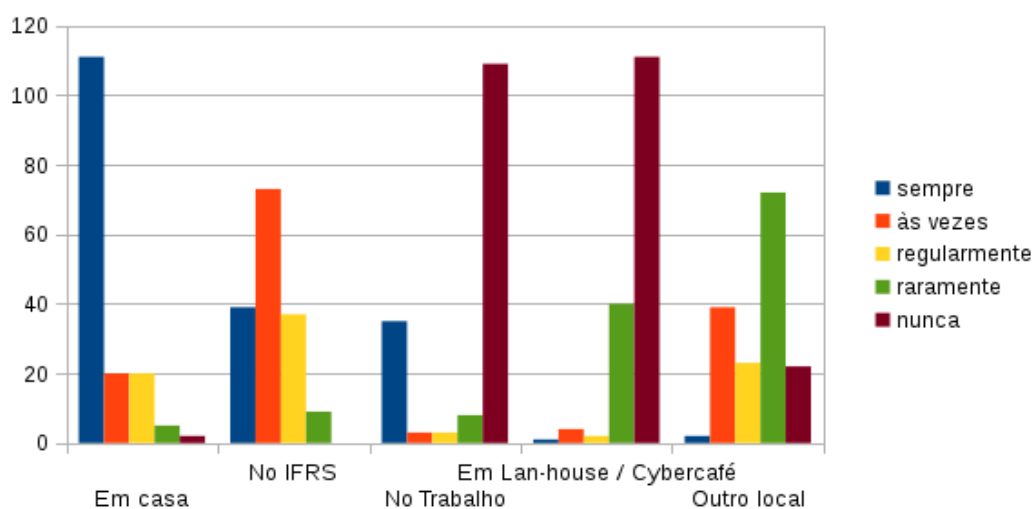
Figura 26: Página inicial do Sistema de Tarefas

### 4.1 Perfil dos sujeitos

No Sistema de Tarefas, os participantes que aceitaram participar dos testes, primeiramente preencheram um formulário contendo informações sobre os diferentes locais em que costumam utilizar o computador e acessar a internet, bem como a frequência que costumam utilizar ambos para execução de diferentes tarefas (como estudar, pesquisas para trabalho, etc), suas faixas etárias e seus níveis de escolaridade. Isto com o objetivo de

entender o quão habituados os mesmos estão com a utilização do computador, da internet e mais especificamente com a execução de buscas por diferentes tipos de conteúdos. Como dito anteriormente, os participantes fazem parte da comunidade escolar do IFRS-BG e, além disso, atuam em diferentes áreas e, portanto, possuem diferentes conhecimentos e níveis de experiência sobre o uso de um computador e, mais especificamente, internet. Os dados obtidos são apresentados a seguir.

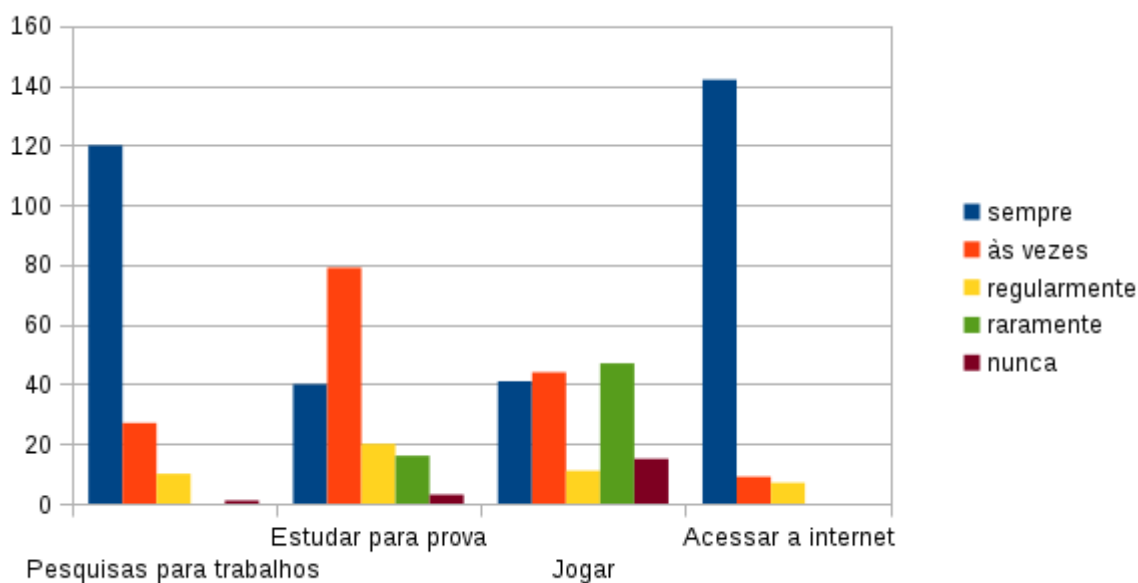
No Gráfico 1, são apresentadas as frequências com que os usuários utilizam o computador em diferentes locais. Nele é possível perceber que os locais onde os participantes mais utilizam o computador é em casa e no próprio IFRS-BG.



**Gráfico 1: Frequência com que os participantes utilizam o computador nos locais definidos**

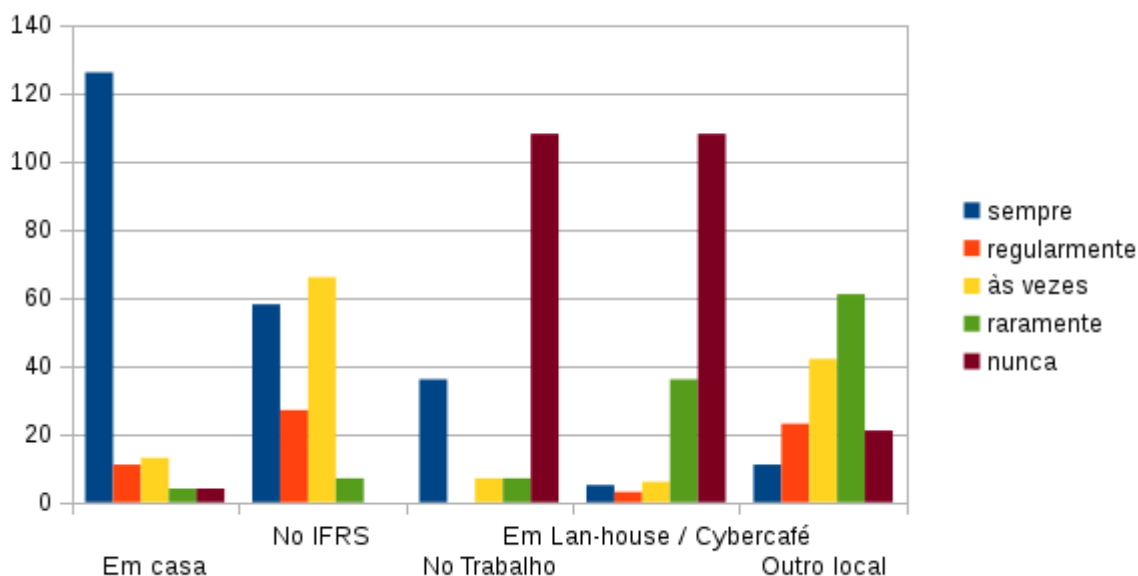
No Gráfico 2 são apresentadas as frequências com que os usuários utilizam o computador para execução de diferentes tarefas. Neste gráfico, é possível constatar que a tarefa que os participantes, em sua maioria, mais utilizam o computador para fazer é acessar a internet e em segundo lugar vem pesquisas para trabalhos, o que leva a crer que os mesmos estejam habituados a executar buscas por materiais de apoio a aprendizagem.





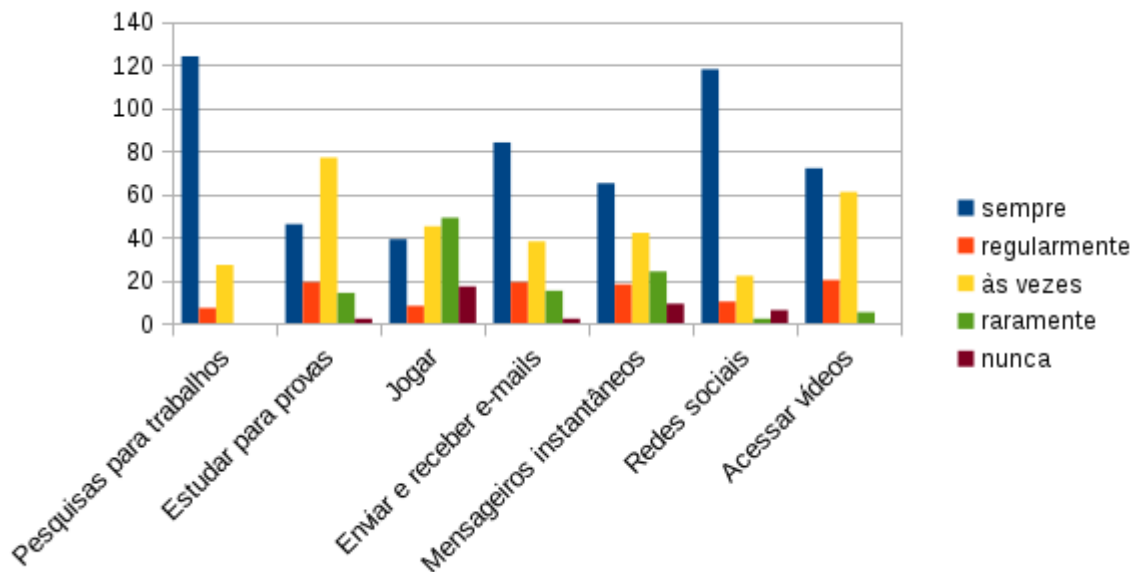
**Gráfico 2: Frequência com que os participantes utilizam o computador para executar as tarefas definidas**

Nos gráficos 3 e 4 são apresentados dados quanto ao local de acesso a internet e as tarefas mais executadas. No Gráfico 3, é possível perceber que os locais mais utilizados para acessar a internet são em casa e no IFRS-BG.



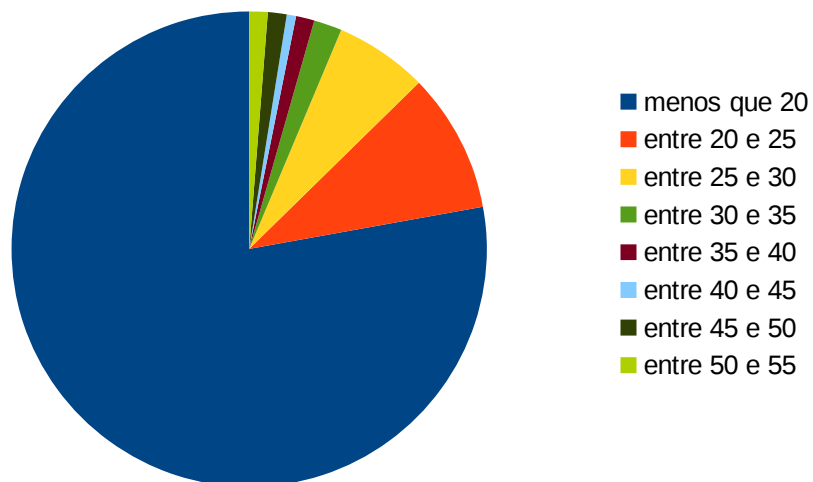
**Gráfico 3: Frequência com que os participantes acessam a internet nos locais definidos**

No Gráfico 4, é possível perceber que os usuários acessam a internet com maior frequência para executar pesquisas para trabalhos e para acesso a redes sociais.



**Gráfico 4: Frequência com que os participantes acessam a internet para executar as tarefas definidas**

Quanto a faixa etária, a maioria dos participantes possui menos do que vinte anos de idade, como mostrado no Gráfico 5.



**Gráfico 5: Faixas etárias dos participantes**

Quanto a escolaridade, a maior parte dos participantes se encontra no ensino médio (cursando também o ensino técnico) ou está cursando o ensino superior.

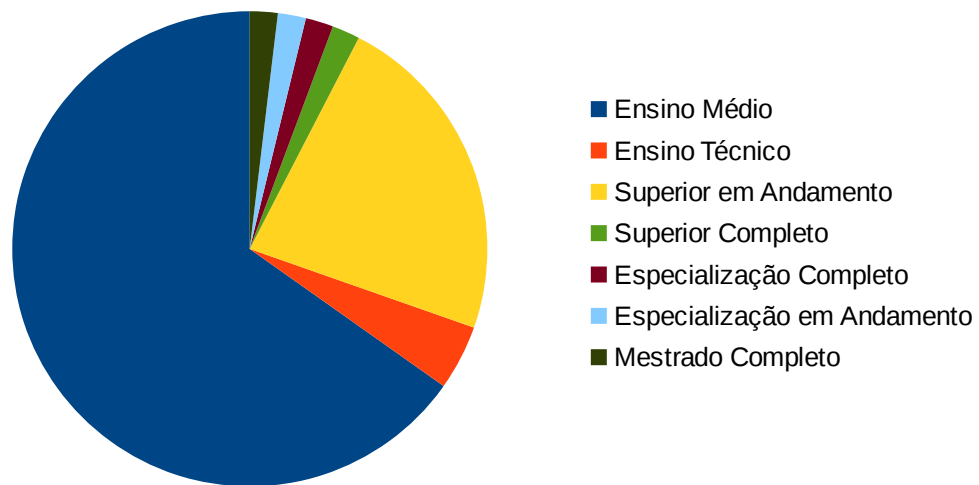


Gráfico 6: Escolaridades dos participantes

#### ***4.2 Avaliação das recomendações geradas pelo sistema***

O segundo passo durante os testes foi a avaliação da relevância das recomendações geradas pelo algoritmo de recomendação por parte dos participantes. Para tal, como dito anteriormente na seção 1.4, lhes eram mostrados os dados de um objeto de aprendizagem e uma lista com as descrições de cinco recomendações geradas pelo algoritmo, sendo assim, os participantes tiveram que ler a descrição do primeiro e compará-las com os da lista, e após, para cada objeto de aprendizagem recomendado, separadamente, responder se acreditavam que o mesmo tinha alguma relação ou relevância em relação ao primeiro, como mostrado na Figura 27. Na tela apresentada na figura, são mostrados somente dois objetos de aprendizagem por vez, isto para facilitar a avaliação por parte dos participantes. Deste modo, assim que os mesmos avaliassem a primeira recomendação, eles clicariam no *link* “Próximo”, e o próximo OA lhes seria apresentado, e assim por diante, até que os cinco objetos escolhidos fossem avaliados.

Compare os recursos educacionais abaixo e responda se você acredita que os mesmos têm ou não alguma relação.

**Título:** Geometria Analítica I

**Descrição:**  
Vídeo aula de como desenvolver os cálculos de geometria analítica.

**Objetivo pedagógico:**  
Vídeo aula que substitui uma aula explanativa.

**Tipos de recurso utilizados:**  
-

**Onde aplicar:**  
Ensino Médio

**Título:** Geometria Analítica II

**Descrição:**  
Vídeo aula que vai rever como se faz uma equação reduzida da reta, coeficientes angulares e lineares e as posições relativas entre duas retas

**Objetivo pedagógico:**  
Vídeo aula que substitui uma aula explanativa.

**Tipos de recurso utilizados:**  
-

**Onde aplicar:**  
Ensino Médio

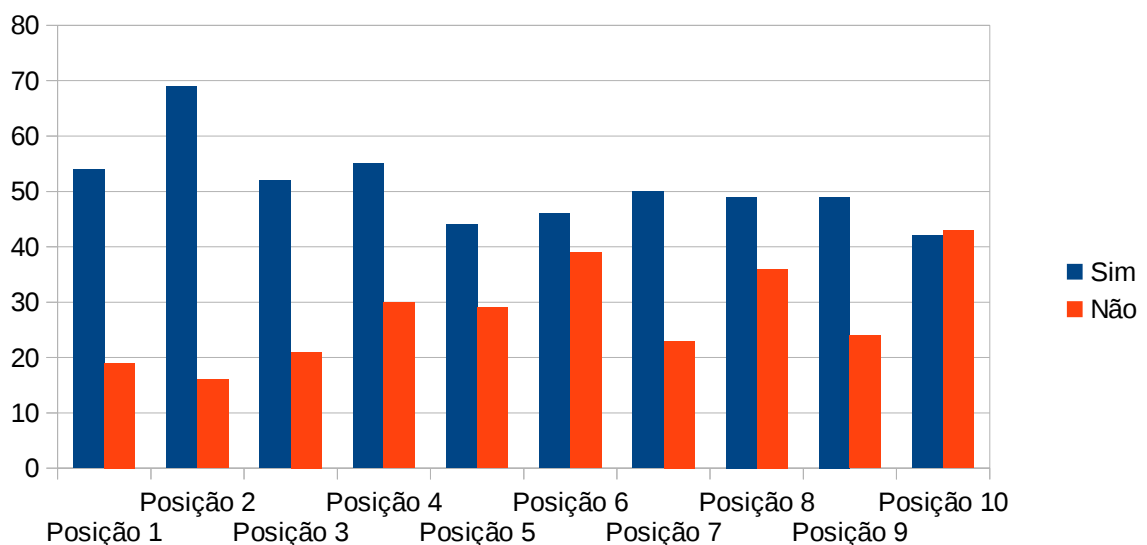
Você considera que este recurso tem alguma relação com o primeiro ?\*

Sim ▼

Próximo

**Figura 27: Tela para avaliação das recomendações geradas pelo sistema**

Os resultados das avaliações das recomendações geradas pelo algoritmo de recomendação são mostradas no Gráfico 7. No processo de avaliação os participantes leram informações sobre um objeto de aprendizagem específico e responderam se acreditavam que, para cada item da lista recomendações gerada o item tinha alguma relação com o objeto citado anteriormente. Desta forma foi possível saber se o algoritmo gera recomendações condizentes quanto a análise do conteúdo dos objetos.



**Gráfico 7: Avaliações dos participantes para as recomendações geradas**

No gráfico, é possível ver a lista de recomendações geradas para os dez itens considerados como sendo os mais parecidos com o objeto de aprendizagem que os participantes acessaram. Os itens estão ordenados do mais parecido (Posição 1) para o menos parecido (Posição 10).

Os participantes avaliaram ao todo 790 itens que foram recomendados pelo sistema, sendo que estes podem ser em alguns casos repetidos. Destes, 510 foram considerados como tendo alguma relação com objeto acessado e 280 como não tendo nenhuma relação. Isto significa que, aproximadamente 64,56% das recomendações seriam consideradas relevantes ou satisfatórias para os usuários do repositório e que 35,44% das recomendações não seriam úteis para os usuários do repositório.

### 4.3 Tarefas de busca

Após, foram delegadas duas tarefas de busca aos participantes, as quais continham três objetos de aprendizagem cada. Uma das tarefas seria executada no repositório que não continha recomendações (aqui chamado de Repositório 1) (Figura 28) e a outra no repositório que continha recomendações (aqui chamado de Repositório 2) (Figura 29).

**Repositório sem recomendações (Repositório 1)**

**Detalhe do Objeto de Aprendizagem**

**Título:** O Desastre

**Descrição:** Aplicativo onde deve-se calcular a velocidade final de algumas coisas ao caírem de certas alturas e com pesos variados.

**Idioma:** Português

**Palavras-chave:** 1. física  
2. movimento de queda livre

**Local:** [http://www2.fe.usp.br/~mef-pietro/ived/sim\\_cinematica\\_desastre.html](http://www2.fe.usp.br/~mef-pietro/ived/sim_cinematica_desastre.html)

**Requisitos de instalação:** Para usar este objeto de aprendizagem, é necessário um navegador de internet, por exemplo: Internet Explorer, Mozilla Firefox, Google Chrome.

**Plataformas suportadas:** Computador  
TV Digital  
Dispositivos Móveis

**Requisitos de acessibilidade:** Tem recursos visuais  
Tem textos

**Tipo de recurso:** simulation

**Tempo de aprendizagem estimado:**

**Nível de Ensino:**

**Áreas de ensino:**

**Como usar:** **O que diz o professor:**  
Material complementar ao estudo do movimento de queda livre.

**Relações com outros objetos de aprendizagem:**

**Acessar**

**Nome do Usuário**

**Senha**

**Acessar**

**Cadastrar-se!**

Figura 28: Repositório 1 – sem as recomendações

Por Nível de Ensino		Acessar
Ensino Infantil Ensino Fundamental - Séries Iniciais Ensino Fundamental - Séries Finais Ensino Médio Ensino Técnico Ensino Superior	<h2>Detalhe do Objetos de Aprendizagem</h2> <hr/> <p><b>Título:</b> O Desastre</p> <p><b>Descrição:</b> Aplicativo onde deve-se calcular a velocidade final de algumas coisas ao caírem de certas alturas e com pesos variados.</p> <p><b>Idioma:</b> Português</p> <ol style="list-style-type: none"> <li>física</li> <li>movimento de queda livre</li> </ol> <hr/> <p><b>Palavras-chave:</b></p> <hr/> <p><b>Local:</b> <a href="http://www2.fe.usp.br/~mf-pietro/rved/sim_cinematica_desastre.html">http://www2.fe.usp.br/~mf-pietro/rved/sim_cinematica_desastre.html</a></p> <p><b>Requisitos de instalação:</b> Para usar este objeto de aprendizagem, é necessário um navegador de internet, por exemplo: Internet Explorer, Mozilla Firefox, Google Chrome.</p> Computador	<input type="text"/> <b>Nome do Usuário</b> <input type="password"/> <b>Senha</b> <a href="#">Acesso!</a> <a href="#">Esqueceu sua senha?</a> <a href="#">Cadastre-se!</a>
<b>Por Tipo de Mídia</b> Texto Imagem Página Web Audio Video Software	TV Digital Dispositivos Móveis <hr/> <p><b>Plataformas suportadas:</b></p> <hr/> <p><b>Requisitos de acessibilidade:</b></p> <ul style="list-style-type: none"> <li> Tem recursos visuais</li> <li> Tem textos</li> </ul> <hr/> <p><b>Tipo de recurso:</b> simulation</p> <p><b>Tempo de aprendizagem estimado:</b></p> <p><b>Nível de Ensino:</b></p> <p><b>Áreas de ensino:</b></p> <hr/> <p><b>Como usar:</b></p> <p>O que diz o professor: Material complementar ao estudo do movimento de queda livre.</p> <pre>SimpleXMLElement Object ([0] =&gt; SimpleXMLElement Object ([0] =&gt; SimpleXMLElement Object ()) [1] =&gt; SimpleXMLElement Object ([kind] =&gt; isAlternative [resource] =&gt; SimpleXMLElement Object ()) [2] =&gt; SimpleXMLElement Object ([kind] =&gt; isAlternative [resource] =&gt; SimpleXMLElement Object ()) [3] =&gt; SimpleXMLElement Object ([kind] =&gt; isAlternative [resource] =&gt; SimpleXMLElement Object ()) [4] =&gt; SimpleXMLElement Object ([kind] =&gt; isAlternative [resource] =&gt; SimpleXMLElement Object ()) [5] =&gt; SimpleXMLElement Object ([kind] =&gt; isAlternative [resource] =&gt; SimpleXMLElement Object ()) [6] =&gt; SimpleXMLElement Object ([kind] =&gt; isAlternative [resource] =&gt; SimpleXMLElement Object ()) [7] =&gt; SimpleXMLElement Object ([kind] =&gt; isAlternative [resource] =&gt; SimpleXMLElement Object ())) Alternativo: Alternativo: Alternativo: Alternativo: Alternativo:</pre> <hr/> <p><b>Relações com outros objetos de aprendizagem:</b></p> <hr/> <p align="right"> Baixar</p>	

Objetos similares a este

A Queda Desesperadora
As Bexigas
Cadê o ovo que estava aqui?
Movimento Unidimensional: Deslocamento, Velocidade e Aceleração
Free Rolling and Circular Motion
Movimento de Projéteis
Movimento Relativo (Sistemas de Referência)
Frictional force
Decole da Porta Avões
Juízo Final
Circular Motion and Centripetal Force
Pulley
Show de Rock
Colisões Unidimensionais e Conservação do Movimento em Diferentes Referenciais Inerciais
Física - Dinâmica: Energia e quantidade de movimento
Colors
Motociclista Maluco
Balança matemática
Acende Aqui Apaga Ali
Escova Perfeta?

**Figura 29: Repositório 2 – com as recomendações**

Em cada busca foi sorteado para os participantes um conjunto com três objetos de aprendizagem (aqui chamados de Objeto 1, Objeto 2 e Objeto 3), dentre os três diferentes conjuntos (aqui chamados de Tarefa 1, Tarefa 2 e Tarefa 3) escolhidos para a tarefa, assim, os participantes liam as descrições dos três OAs que deveriam ser encontrados em cada repositório, os conjuntos de OAs escolhidos foram os seguintes:

- **Tarefa 1**
  - **Objeto 1** - Jogo educativo de geografia, com os continentes e oceanos do planeta Terra;
  - **Objeto 2** - Material complementar ao aprendizado das localizações dos estados brasileiros, com exercício para encontrar no mapa o estado brasileiro indicado;
  - **Objeto 3** - Jogo educativo de geografia com os estados brasileiros e suas capitais.
- **Tarefa 2**
  - **Objeto 1** - Jogo educativo interativo que aborda os conceitos de trigonometria;
  - **Objeto 2** - Conceito de ângulos, tipos (côncavo, raso, de uma volta, nulo). Grau. Tipos de ângulos (agudo, obtuso, reto, complementares, suplementares, adjacentes, opostos pelo vértice). Demonstração de ângulos;
  - **Objeto 3** - Jogo educativo sobre os nomes e as formas das figuras geométricas: círculo, triângulo, quadrado, retângulo, trapézio e pentágono.
- **Tarefa 3**
  - **Objeto 1** - Apresentação de slides composta pelas diretrizes da disciplina de Informática, bem com pelos primeiros e fundamentais conceitos da mesma (o que é, como funciona, tipos e componentes de um computador...);
  - **Objeto 2** - Apresentação do sistema operacional Windows XP, contendo também imagens dos programas da guia "Acessórios", exercícios e uma breve introdução a internet;
  - **Objeto 3** - Apresentação de slides que ilustra as várias maneiras de combinar educação e informática, permitindo assim uma maximização do rendimento em sala de aula ou fora dela (EAD).

Através do sorteio, alguns participantes executaram a Tarefa 1 no Repositório 1, enquanto que outros a executaram no Repositório 2, o mesmo vale para as demais tarefas. Além disso, cada busca foi executada separadamente, e após os participantes preencherem um formulário contendo informações quanto a dificuldade de execução da tarefa e, para o repositório contendo recomendações, o quanto as recomendações ajudaram na execução da

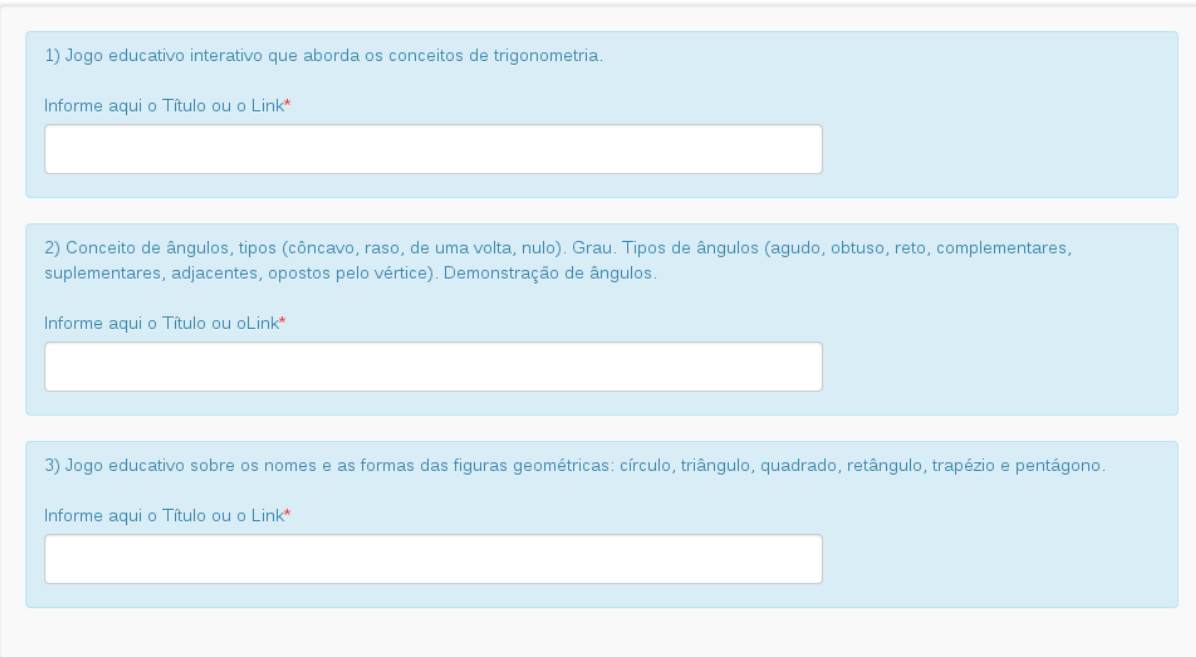
tarefa.

Vale ressaltar que, tais objetos de aprendizagem foram escolhidos pois acreditou-se que devido aos grupos de participantes escolhidos, todos estariam familiarizados com os assuntos abordados em cada OA, de forma a facilitar que os mesmos soubessem do que o objeto de aprendizagem se tratava, não interferindo nas buscas.

Os resultados obtidos após a execução de cada tarefa são apresentados nas seções a seguir.

#### 4.3.1 Tarefa de busca no repositório sem as recomendações

Nesta tarefa, era apresentada aos participantes uma tela contendo os seguintes dizeres: “Imagine que você está em busca de materiais de apoio para seus estudos. Desta forma você tem que encontrar alguns recursos educacionais para auxiliá-lo. Então, leia com atenção as três descrições de recursos educacionais abaixo, elas descrevem os materiais que você precisa encontrar. Após lê-los, vá até o repositório indicado no link abaixo, encontre cada um deles e informe seu título ou link no campo abaixo deles”, e logo após era apresentado o *link* para acesso ao sistema. Assim, os participantes deveriam acessar o repositório, buscar os objetos que eram descritos no formulário e informar seu título ou *link*, como mostrado na Figura 30.



1) Jogo educativo interativo que aborda os conceitos de trigonometria.

Informe aqui o Título ou o Link\*

2) Conceito de ângulos, tipos (côncavo, raso, de uma volta, nulo). Grau. Tipos de ângulos (agudo, obtuso, reto, complementares, suplementares, adjacentes, opostos pelo vértice). Demonstração de ângulos.

Informe aqui o Título ou o Link\*

3) Jogo educativo sobre os nomes e as formas das figuras geométricas: círculo, triângulo, quadrado, retângulo, trapézio e pentágono.

Informe aqui o Título ou o Link\*

**Figura 30: Formulário para informar os objetos de aprendizagem encontrados na busca**

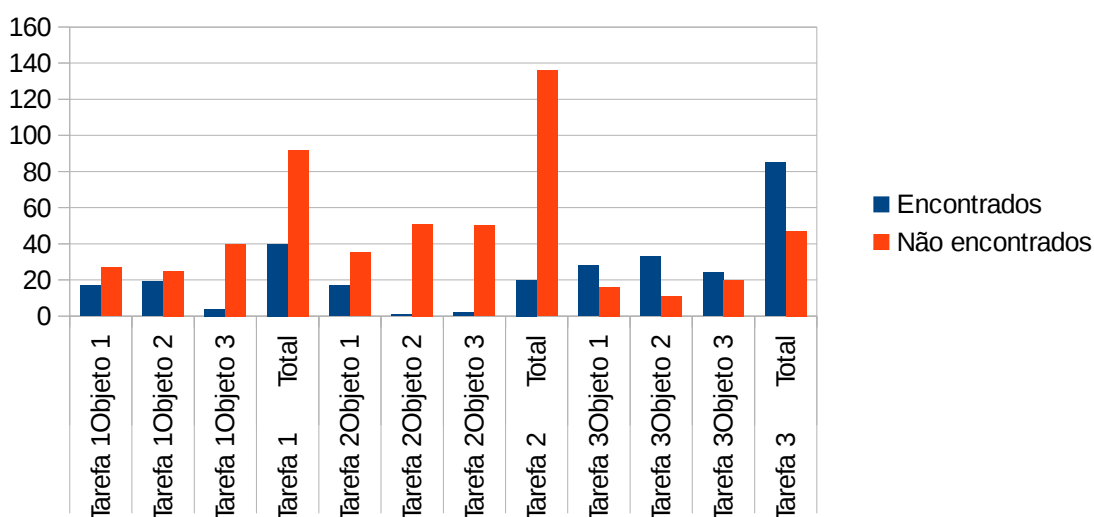
Os resultados das buscas no Repositório 1, podem ser vistos no Gráfico 8. Nele, é



possível observar que para os objetos propostos para a busca:

- **Na Tarefa 1:** 40 (30,3%) foram encontrados com sucesso, enquanto que 92 (69,7%) não foram encontrados.
- **Na Tarefa 2:** 20 (12,82%) foram encontrados com sucesso, enquanto que 136 (87,18%) não foram encontrados.
- **Na Tarefa 3:** 85 (64,40%) foram encontrados com sucesso, enquanto que 47 (35,6%) não foram encontrados.

A avaliação dos participantes quanto á dificuldade de execução das tarefas é apresentada nos Gráficos 10 e 11, sendo o primeiro para o Repositório 1 e o segundo para o Repositório 2.



**Gráfico 8: Resultados da primeira tarefa de busca no sistema sem as recomendação**

Quanto a dificuldade de execução da tarefa (Gráfico 9), é possível perceber, que segundo os participantes, a execução da tarefa é classificada pela maioria como estando entre moderadamente difícil e difícil. No gráfico as classificações estão separadas para cada uma das tarefas (conjuntos de OAs que deveriam ser buscados).

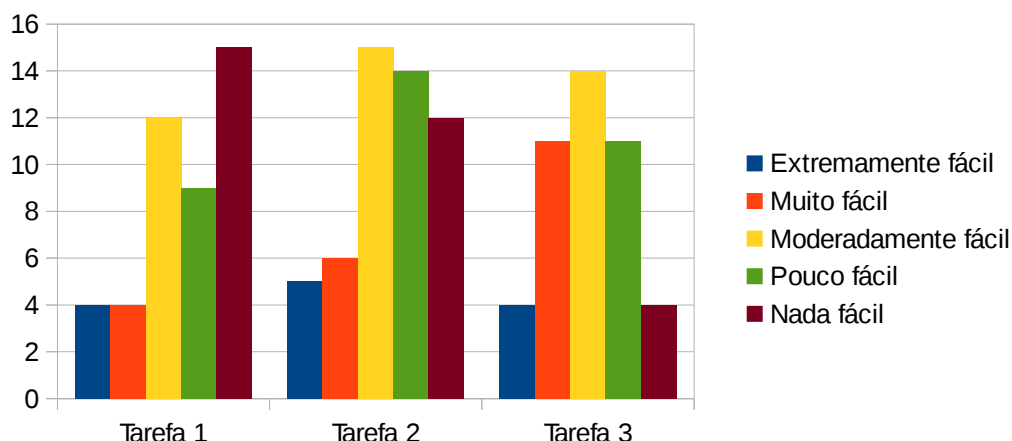


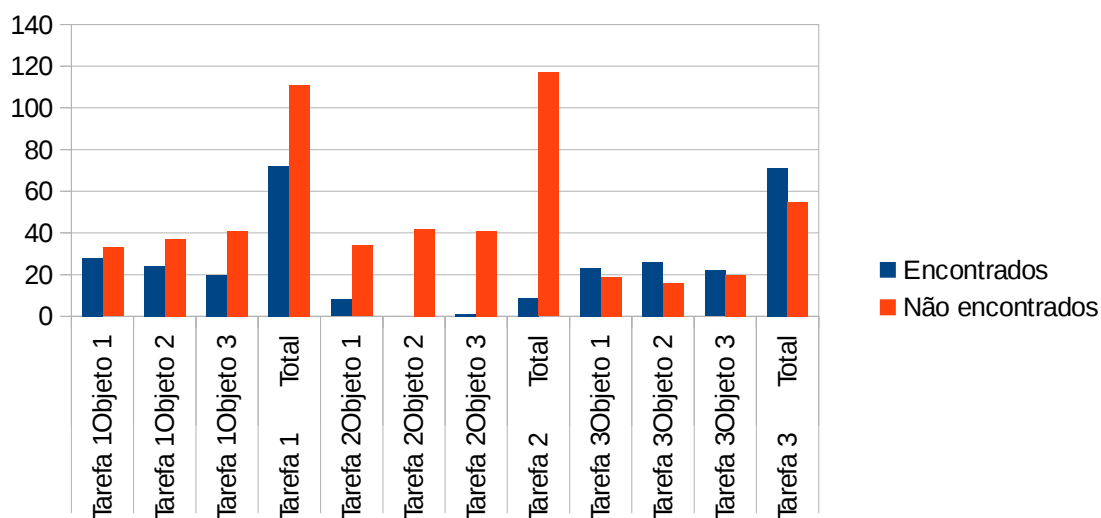
Gráfico 9: Avaliação da dificuldade de execução das tarefas para o repositório sem recomendações

#### 4.3.2 Tarefa de busca no repositório com as recomendações

Da mesma forma que na primeira tarefa de busca, os usuários tinham as seguintes instruções “Imagine agora que você precisa de outros recursos educacionais. Os mesmos estão descritos abaixo, leia-os com atenção e vá até o repositório indicado no link abaixo, tente encontrar cada um deles e informe seu título ou link no campo abaixo deles.”. E então os mesmos preenchiam o formulário com as informações dos objetos de aprendizagem encontrados.

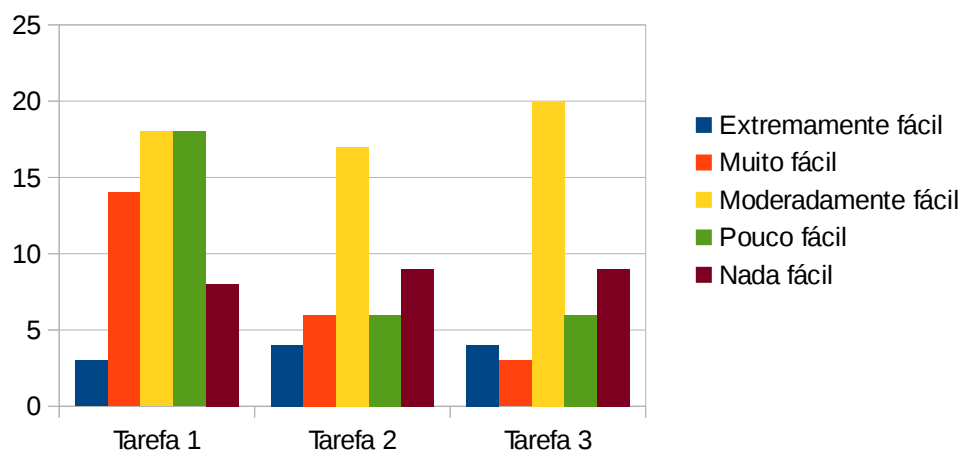
Para as buscas no Repositório 2, os resultados são apresentados no Gráfico 1, onde é possível observar que, para os objetos propostos para a busca:

- **Na Tarefa 1:** 72 (39,34%) foram encontrados com sucesso, enquanto que 111 (60,66%) não foram encontrados.
- **Na Tarefa 2:** 9 (7,14%) foram encontrados com sucesso, enquanto que 117 (92,86%) não foram encontrados.
- **Na Tarefa 3:** 71 (56,35%) foram encontrados com sucesso, enquanto que 55 (43,65%) não foram encontrados.



**Gráfico 10: Resultados da segunda tarefa de busca no sistema com as recomendações**

Quanto a dificuldade de execução da tarefa, o mesmo que ocorreu na Tarefa 1 se sucede para o repositório contendo as recomendações (Gráfico 11), ou seja, segundo os participantes, a execução da tarefa é classificada pela maioria como estando entre moderadamente difícil e difícil.



**Gráfico 11: Avaliação da dificuldade de execução das tarefas para o repositório com recomendações**

Além disso, para o caso deste repositório (Repositório 2), os participantes avaliaram o quão úteis as recomendações foram no processo de busca pelos objetos, como mostrado no Gráfico 11. Nesta avaliação, a maior parte dos participantes classificou as recomendações como estando entre moderadamente úteis e extremamente úteis, como mostrado do Gráfico 12.

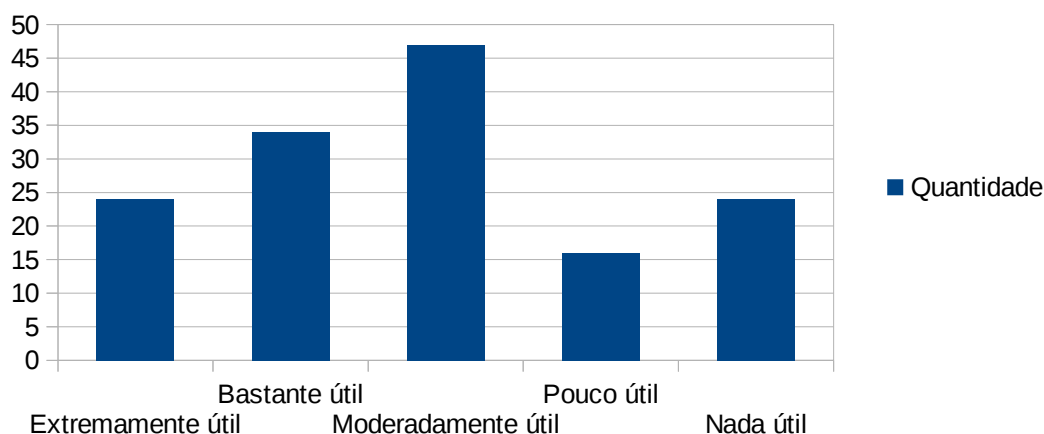


Gráfico 12: Utilidade das recomendações para o Repositório 2

### 4.3.3 Comparação entre as tarefas de busca

Em suma, é possível perceber que, mesmo no repositório que possui as recomendações, o número de objetos de aprendizagem encontrados permanece baixo (Tabela 1). Na tabela são mostrados os números percentuais de objetos de aprendizagem encontrados e também dos não encontrados para cada uma das tarefas ou conjuntos de OAs.

Tabela 1: Comparação entre resultados da Tarefa 1 e 2

	Encontrados		Não encontrados	
	Repositório 1	Repositório 2	Repositório 1	Repositório 2
Tarefa 1	30,30%	39,34%	69,70%	60,66%
Tarefa 2	12,82%	7,14%	87,18%	92,86%
Tarefa 3	64,39%	56,35%	35,61%	43,65%

### 4.3.4 Considerações sobre a aplicação do experimento

Nesta seção são apresentadas algumas considerações relevantes quanto a aplicação do experimento, como por exemplo: clareza das tarefas, problemas ocorridos durante a aplicação do experimento, questionamentos, etc.

Primeiramente, durante a preparação do roteiro, quando se escolheram os objetos de aprendizagem que iriam fazer parte das buscas, acreditou-se que todos os participantes estariam familiarizados com os termos e assuntos apresentados nas descrições de cada OA. Porém durante a aplicação, ocorreram questionamentos quanto ao significado de termos, e a partir da observação, que algumas descrições não se fizeram totalmente claras a todos os participantes, o que pode, de certa forma, ter interferido nos resultados finais das buscas.

Além disso, o modo como a tarefa foi apresentada aos participantes também não se fazia claro a todos, e sendo assim, a interpretação do que deveria ser feito nas buscas e como

isto deveria ser feito, também teve influência na execução das tarefas.

Outro ponto importante a ser considerado, é o fato de que, por problemas com a conexão com o Sistema de Tarefas, que por vezes ficava inacessível através da rede, nem todos os participantes conseguiram executar os últimos dois passos (tarefas de busca). Vale ressaltar que em um primeiro momento os participantes preencheram o formulário com seus dados pessoais, e após avaliaram as recomendações, e então viriam as tarefas de buscas, porém caso os participantes não executassem uma das buscas, ou ambas, o que mudaria seria somente o número considerado nos resultados, pois nenhum dos passos (avaliação das recomendações, busca 1 e busca 2) tinha dependência dos demais. Além da conexão, por problemas técnicos algumas máquinas (computadores) utilizados apresentaram problemas o que também acarretou em alguns participantes não conseguirem terminar todas as tarefas.

Nas tarefas de buscas, para o sistema que continha as recomendações a taxa de OAs encontrados se mantiveram parecidas com as do sistema sem as recomendações. Isto pode ter ocorrido por dois motivos principais: primeiramente ao fato de as recomendações serem apresentadas na página de detalhes dos um objeto de aprendizagem, o que levava os participantes a terem que acessar um OA e utilizar a rolagem de página para chegar nas recomendações, pois as mesmas estavam localizadas no final da página (como apresentado na Figura 28), e deste modo, através da observação, percebeu-se que muitas vezes as recomendações passaram por despercebidas; segundo pela própria dificuldade de execução das tarefas, que requeria que os participantes encontrassem um objeto de aprendizagem específico através de sua descrição, o que foi considerado como sendo algo de certa forma difícil.

Apesar da taxa de acertos obtida, através da verificação dos resultados, percebeu-se que alguns objetos encontrados possuíam assuntos relacionados as descrições fornecidas aos participantes, o que poderia levar os mesmos à ser considerados como acertos, o mesmo poderia ocorrer com os objetos de aprendizagem que nas respostas apareciam trocados entre si, ou seja o que deveria ser, por exemplo informado para a descrição 1 foi informado para a descrição 2, e como dito anteriormente os três OAs presentes em cada grupo eram relacionados. Porém uma análise mais detalhada quanto aos acertos, fugiria ao objetivo do presente trabalho.

## 5 CONCLUSÕES

Como dito anteriormente, este trabalho objetivou avaliar como a aplicação de um algoritmo de recomendação de conteúdos no ROA do IFRS-BG facilitaria que os usuários encontrassem informações (objetos de aprendizagem). Desta forma alguns outros objetivos tiveram que ser alcançados antes. O primeiro foi o investigar o funcionamento do ROA, onde foram considerados alguns pontos relevantes para a construção do algoritmo de recomendação, que era o próximo objetivo deste trabalho. Os pontos considerados foram quais informações seriam úteis para gerar as recomendações, como descrições dos OAs e quais menus do ROA os usuários teriam disponíveis para filtrar conteúdos, pois estes também são informações gravadas nos OAs. Deste modo tal análise foi importante pois ajudou na criação do algoritmo.

O algoritmo desenvolvido gerou recomendações para todos os objetos de aprendizagem presentes no repositório, os quais, mas não em sua totalidade, foram avaliados pelos participantes durante os experimentos de testes. Os resultados obtidos em tais avaliações levaram concluir que as recomendações foram bem aceitas pelos usuários, ou seja, os conteúdos que seriam recomendados eram relevantes, desta forma o algoritmo atingiu seu objetivo.

Após a aplicação do experimento, o qual inicialmente acreditou-se ser suficiente para a avaliação de como a aplicação do algoritmo facilitaria que os usuários encontrassem OAs, os dados obtidos não foram totalmente concludentes. Porém, desta forma percebeu-se que os usuários chegavam mais rapidamente as informações, desde que estes conhecessem o assunto de que o OA se tratava, pois sem as recomendações, os mesmos navegavam através dos menus, ou utilizavam a busca e tinham que abrir OA por OA nas listas de OAs, até encontrar o desejado, já no repositório com recomendações, desde que encontrando algum OA relacionado ao desejado, o mesmo estaria na lista recomendados, diminuía o número de OAs que os usuários deveria acessar para verificar se este se tratava do que eles deveriam encontrar, ou seja a busca era facilitada de certa forma.

Como futuros trabalhos, podem ser agregadas também outras técnicas de recomendação, como a colaborativa, onde os usuários avaliariam os objetos de aprendizagem dando notas, de forma que quando alguém acessasse o repositório, este poderia saber o quão

úteis os OAs são. Além desta, também podem ser criadas listas de recomendações como: objetos mais acessados, últimos cadastrados, melhor avaliados, etc. O que facilita ainda mais que os usuários encontrassem OAs relevantes. Outro trabalho futuro seria a definição de outro local mais apropriado para a localização das recomendações geradas pelo algoritmo, o que também facilita a busca no repositório.

## REFERÊNCIAS BIBLIOGRÁFICAS

Amazonas *et al.* - ORDER, INTEGRATING DATA INDEX ENGINES IN. Integrando motores de indexação de dados para a construção de sistemas de recuperação de informação em ambientes heterogêneos. **Journal of Information Systems and Technology Management**, v. 5, n. 2, p. 193-222, 2008.

BASU, Chumki et al. **Recommendation as classification: Using social and content-based information in recommendation**. In: AAAI/IAAI. 1998. p. 714-720.

BRADLEY, Keith; SMYTH, Barry. **Improving recommendation diversity**. In: Proceedings of the Twelfth National Conference in Artificial Intelligence and Cognitive Science (AICS-01). 2001.

CORRÊA, Marilda Ciribelli. **Como Elaborar uma dissertação de mestrado através da pesquisa científica** – Rio de Janeiro: 7Letras, 2003.

**Draft Standard for Learning Object Metadata**. Disponível em:

<[http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf)>. Acesso em 17 jul. 2013.

FERREIRA, Vinicius Hartmann. **Sistema de Recomendação de Objetos de Aprendizagem**. TCC (graduação em Ciência da Computação) - Universidade do Vale do Itajaí, Itajaí, 2009. Disponível em : <[http://Siaibib01.univali.br/pdf/Vinicius\\_Hartmann\\_Ferreira.pdf](http://Siaibib01.univali.br/pdf/Vinicius_Hartmann_Ferreira.pdf)>.

GHELMAN, Raphael *et al.* **Reconhecimento e Agrupamento de Objetos de Aprendizagem Semelhantes**. Revista Brasileira de Informática na Educação, v. 15, n. 3, 2007.

GLUZ, João Carlos; VICARI, Rosa Maria. **Uma Ontologia OWL para Metadados IEEELOM, DublinCore e OBAA**. In: Anais do Simpósio Brasileiro de Informática na Educação. 2011.

HERLOCKER, Jonathan Lee. **Understanding and improving automated collaborative filtering systems**. 2000. Tese de Doutorado. University of Minnesota.

IEEE LTSC. Disponível em: <<http://www.ieeeltsc.org:8080/Plone/working-group/learning-object-metadata-working-group-12/learning-object-metadata-lom-working-group-12>>. Acesso em 23 abr. 2013.



- KIM, Byeong Man et al. **A new approach for combining content-based and collaborative filters**. *Journal of Intelligent Information Systems*, v. 27, n. 1, p. 79-91, 2006.
- KORENIUS, Tuomo; LAURIKKALA, Jorma; JUHOLA, Martti. On principal component analysis, cosine and Euclidean measures in information retrieval. *Information Sciences*, v. 177, n. 22, p. 4893-4905, 2007.
- LICHTNOW, Daniel *et al.* **O uso de técnicas de recomendação para apoio à aprendizagem colaborativa. Revista Brasileira de Informática na Educação**. vol. 14, n.3, set. 2006.
- LINDEN, Greg; SMITH, Brent; YORK, Jeremy. **Amazon. com recommendations: Item-to-item collaborative filtering**. *Internet Computing, IEEE*, v. 7, n. 1, p. 76-80, 2003.
- LOPES, Giseli Rabello - **Sistema de Recomendação para Bibliotecas Digitais sob a Perspectiva da Web Semântica**. Porto Alegre: Programa de Pós-Graduação em Computação, 2007.
- MCGREAL, Rory. **Learning objects: A practical definition**. *International Journal of Instructional Technology and Distance Learning*, v. 1, n. 9, p. 21-32, 2004.
- MERLOT. Disponível em: <<http://www.merlot.org/>>. Acesso em 16 jun. 2013.
- MORAIS, Edison Andrade Martins; AMBRÓSIO, Ana Paula L. Mineração de Textos. **Relatório Técnico–Instituto de Informática (UFG)**, 2007.
- OBAA, Portal. Disponível em <<http://www.portalobaa.org/>> Acesso em 3 jul. 2013.
- ORENGO, Viviane Moreira; HUYCK, Christian R. A Stemming Algorithm for the Portuguese Language. In: **spire**. 2001. p. 186-193.
- PÁDUA, Elisabete Matallo Marchesini de – **Metodologia da pesquisa: Abordagem teórico-prática** 0 10. ed. rev. E atual. - Campinas, SP: Papirus, 2004.
- Padrão OBAA - Mapa Mental. Disponível em <<http://www.portalobaa.org/padrao-obaa/OBAA-v2.mm/>>. Acesso em 03 jul. 2013.
- PAZZANI, Michael J.; BILLSUS, Daniel. **Content-based recommendation systems**. In: **The adaptive web**. Springer Berlin Heidelberg, 2007. p. 325-341.
- PEREIRA *et al.*. **Objetos de Aprendizado Re-Utilizáveis (RLOs): Conceitos, Padronização, Uso e Armazenamento**. Monografia Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro, 2002.

RAJARAMAN, Anand; ULLMAN, Jeffrey David. **Mining of massive datasets**. Cambridge University Press, 2012.

REATEGUI, Eliseo Berni; CAZELLA, Sílvio César. **Sistemas de recomendação**. In: XXV Congresso da Sociedade Brasileira de Computação. 2005. p. 306-348.

REIS, Linda G. - **Produção de monografia da teoria à prática** - 2. ed. - Brasília : Senac-DF, 2008.

RICHARDS, Griff; MCGREAL, Rory; FRIESEN, Norman. **Learning object repository technologies for telelearning: The Evolution of POOL and CanCore**,. In: Proceedings of the IS2002, Informing Science - IT Education Conference, 2002.

SALTON, Gerard; BUCKLEY, Christopher. Term-weighting approaches in automatic text retrieval. **Information processing & management**, v. 24, n. 5, p. 513-523, 1988.

SALTON, Gerard; BUCKLEY, Christopher. **Term Weighting Approaches** in Automatic Text Retrieval. *Information Processing & Management* 24 (5), 513-523, 1988.

SANTOS, Vanice dos; CANDELORO, RosanaJ. – **Trabalhos acadêmicos: uma orientação para a pesquisa e normas técnicas** – Porto Alegre: RS: AGE, 2006.

SILLA JR, Carlos N.; KAESTNER, Celso AA. Estudo de métodos automáticos para sumarizaç ao de textos. **Simpósio de Tecnologias de Documentos**, p. 45-49, 2002.

SILVA, Júlia Marques Carvalho Da; VICARI, Rosa Maria. **Ampliando as possibilidades de uso do elemento relation nos objetos de aprendizagem**. In: Anais do Simpósio Brasileiro de Informática na Educação. 2010.

SILVA, Júlia Marques Carvalho da; BAVARESCO, Natanael; SILVEIRA, Ricardo Azambuja. **Projeto e desenvolvimento de um sistema multi-agentes para objetos inteligentes de aprendizagem baseado no padrão scorm**. *Revista Brasileira de Informática na Educação*, v. 16, n. 01, 2008.

VICARI, Rosa Maria *et al.* **Proposta brasileira de metadados para objetos de aprendizagem baseados em agentes (obaa)**. RENOTE: revista novas tecnologias na educacao [recurso eletrônico]. Porto Alegre, RS, 2010.

VICCARI, R. *et al.* **Projeto OBAA - Relatório Técnico RTOBAA-01 – Proposta de Padrão para Metadados de Objetos de Aprendizagem Multiplataforma**.

UFRGS/CINTED. Disponível em: <<http://www.portalobaa.org/obaac/padrao-obaa/relatorios-tecnicos/RT-OBAA-01.pdf/view>>, 2009. Acesso em :17 jul. 2013.

VIEIRA, Felipe José Rocha; NUNES, Maria Augusta Silveira Netto. **DICA: Sistema de Recomendação de Objetos de Aprendizagem Baseado em Conteúdo**. Scientia Plena, v. 8, n. 5, 2012.

YAMADA, ADRIANO K. et al. Uma solução flexível para a etapa de pré-processamento em mineração de textos. In: **CONGRESSO INTERINSTITUCIONAL DE INICIAÇÃO CIENTÍFICA**. 2012. p. 1-12.

ZAINA, L., BRESSAN, G., CARDIERI, M., RODRIGUES JÚNIOR, J.. **e-LORS: Uma Abordagem para Recomendação de Objetos de Aprendizagem**. Revista Brasileira de Informática na Educação, 20, abr. 2012. Disponível em: <<http://www.br-ie.org/pub/index.php/rbie/article/view/1289>>. Acesso em: 15 Mai. 2013.

## Apêndice A - Cronograma

**Table 2: Cronograma das etapas de desenvolvimento do presente trabalho**

[illegible]

## Apêndice B – Diagrama de Classes

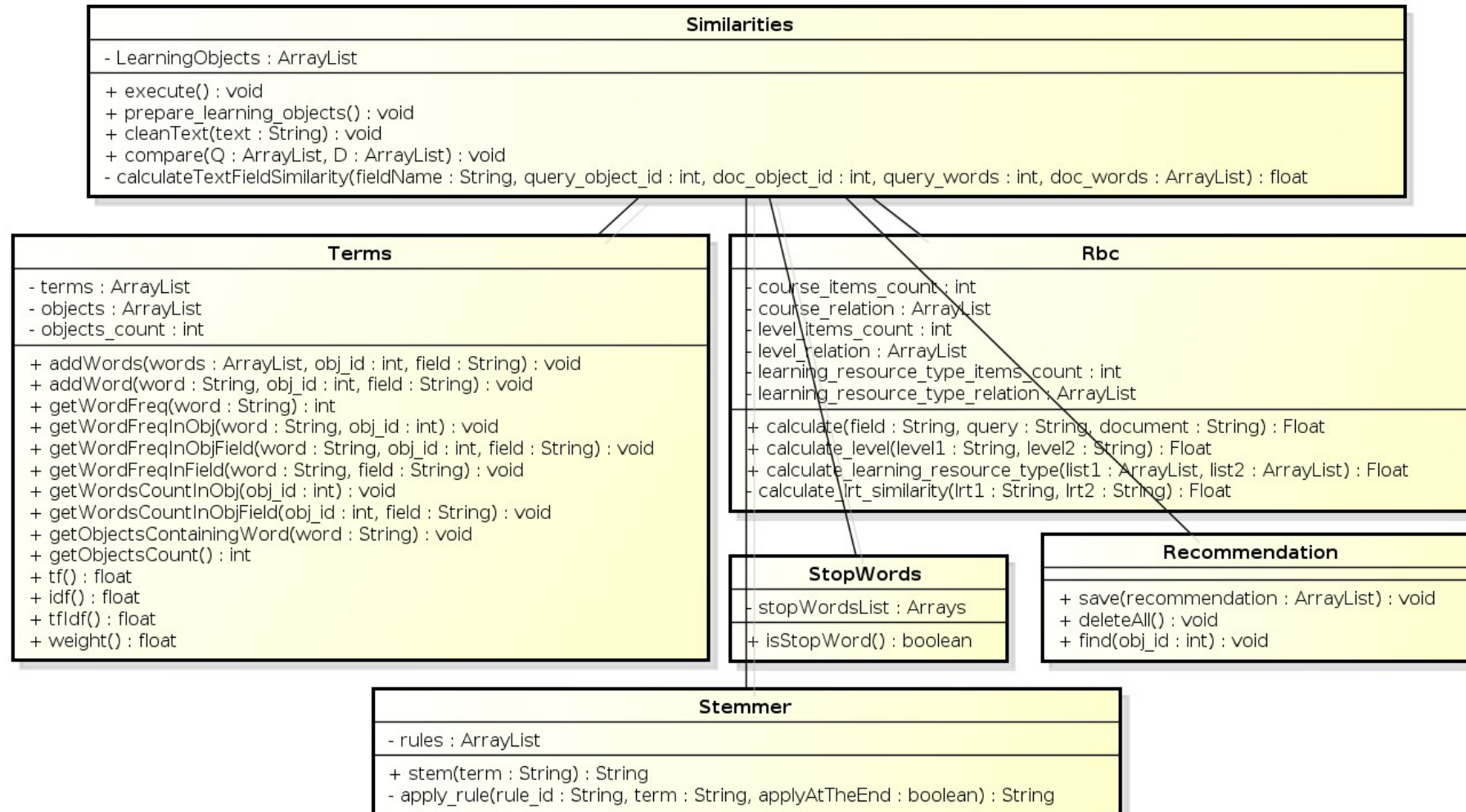


Illustration 31: Diagrama de Classes do Algoritmo de recomendações