

UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico
Programa de Pós-Graduação em Computação



Dissertação

**Recomendação de Objetos de Aprendizagem de Línguas baseada em
Inteligência de Enxames**

Paulo de Almeida Afonso

Pelotas, 2016

Paulo de Almeida Afonso

**Recomendação de Objetos de Aprendizagem de Línguas baseada em
Inteligência de Enxames**

Dissertação apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Mestre em Ciência da Computação

Orientador: Prof. Dr. Paulo Roberto Ferreira Jr.
Coorientador: Prof. Dr. Rafael Vetromille-Castro

Pelotas, 2016

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação na Publicação

A257r Afonso, Paulo de Almeida

Recomendação de objetos de aprendizagem de línguas baseada em inteligência de enxames / Paulo de Almeida Afonso ; Paulo Roberto Ferreira Júnior, orientador ; Rafael Vetromille-Castro, coorientador. — Pelotas, 2016.

99 f. : il.

Dissertação (Mestrado) — Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2016.

1. Inteligência de enxames. 2. Objetos de aprendizagem. 3. Objetos de aprendizagem de línguas. 4. Sistemas de recomendação. I. Ferreira Júnior, Paulo Roberto, orient. II. Vetromille-Castro, Rafael, coorient. III. Título.

CDD : 005

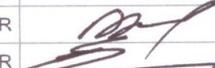
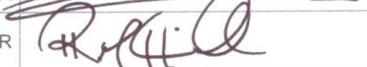
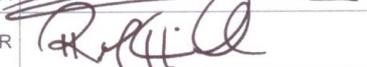
Elaborada por Aline Herbstrith Batista CRB: 10/1737

	UNIVERSIDADE FEDERAL DE PELOTAS PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO DEPARTAMENTO DE PÓS-GRADUAÇÃO		5
---	---	---	---

DEFESA DE DISSERTAÇÃO

NOME DO ESTUDANTE	MATRICULA
PAULO DE ALMEIDA AFONSO	13101708

CURSO OU PROGRAMA	NÍVEL
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO	MESTRADO

MEMBROS DA BANCA EXAMINADORA	TÍTULO	ASSINATURA
Paulo R. Ferreira Jr. (PPGC-UFPel)	DOUTOR	
Cristian Cechinel (PPGC-UFPel)	DOUTOR	
Raymundo Carlos Machado Ferreira Filho (Instituto Federal Sul-rio-grandense)	DOUTOR	

APRECIÇÃO SOBRE A DISSERTAÇÃO

NÃO SIGILOSA

Em 24 de Agosto de 2016, os membros acima nomeados para a defesa da Dissertação do(a) estudante **PAULO DE ALMEIDA AFONSO**, matriculado no Programa de Pós-Graduação em Computação, consideraram a Dissertação **aprovada**, estabelecendo o título definitivo como sendo **"Recomendação de Objetos de Aprendizagem baseada em Inteligência de Enxames"** e estabelecem um prazo máximo de 30 dias para as correções e entrega da versão definitiva.

DADOS PESSOAIS DOS MEMBROS DA BANCA EXAMINADORA					
NOME COMPLETO	CPF	ANO NASCIMENTO	TITULAÇÃO		
			Área	Local	Ano
Paulo R. Ferreira Jr. (PPGC-UFPel)	616.990.980-34	1976	Ciência da Computação	UFRGS	2008
Cristian Cechinel (PPGC-UFPel)	454.628.449-72	1977	Ciência da Computação	Universidad de Alcalá	2012
Raymundo Carlos Machado Ferreira Filho (Instituto Federal Sul-rio-grandense)	696.387.990-72	1972	Doutorado em Informática na Educação	Universidade Federal do Rio Grande do Sul	2008

1ª Via – Coordenador do Curso; 2ª Via – Orientador; 3ª Via – PRPPG.
DISTRIBUIÇÃO A CARGO DA COORDENAÇÃO DO PROGRAMA.

RESUMO

AFONSO, Paulo de Almeida. **Recomendação de Objetos de Aprendizagem de Línguas baseada em Inteligência de Enxames**. 2016. 99 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2016.

Este trabalho propõe a aplicação de técnicas de inteligência de enxames para a recomendação de objetos de aprendizagem de línguas. No melhor de nosso conhecimento, não se pode encontrar na literatura trabalhos específicos de recomendação deste tipo de objeto. As abordagens existentes são voltadas para a recomendação de objetos de aprendizagem de modo genérico, sem considerar os aspectos que norteiam o aprendizado de línguas. Nesse sentido, foi realizado um estudo sobre os elementos essenciais que definem um objeto de aprendizagem de línguas e as principais abordagens utilizadas em sistemas de recomendação de uso geral. O estudo considerou a hipótese de aplicação em um repositório desenvolvido com base em princípios pedagógicos comunicativos e levou em consideração o caráter complexo da aprendizagem de línguas. A tarefa de recomendação nesse contexto prevê a existência de múltiplas conexões conceituais entre os objetos, que são construídas à medida que o sistema é utilizado, e ocorre através da aplicação do algoritmo de Otimização por Colônia de Formigas. Como forma de validar a solução proposta foram implementados, além da abordagem proposta, dois dos algoritmos de recomendação de uso geral mais empregados atualmente. Considerando que um objeto de aprendizagem de línguas tem como objetivo principal ser um facilitador do processo de integração de competências gramatical, sociolinguística e estratégica, dois aspectos essenciais devem ser considerados durante o processo de recomendação: a capacidade de combinar um objeto com outro e a reusabilidade, de forma mais específica, a possibilidade de sua aplicação em diferentes contextos de aprendizagem. Dessa forma, os resultados obtidos a partir dos experimentos demonstraram que a solução proposta mostra-se eficiente para o contexto de aplicação a que se propõe.

Palavras-chave: Inteligência de enxames, Objetos de Aprendizagem, Objetos de aprendizagem de Línguas, Sistemas de Recomendação.

ABSTRACT

AFONSO, Paulo de Almeida. **Recommendation of Learning Objects Language based on Swarm Intelligence**. 2016. 99 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2016.

This work proposes the application of swarm intelligence to recommend language learning objects. To the best of our knowledge, one cannot find in the literature specific works addressing the recommendation of this kind of object. Existing approaches are focused to recommend general learning objects without considering the aspects that guide the learning of languages. In this sense, a study of the essential elements that define a language learning object and the main approaches used in the mentioned general-use recommender systems was conducted. The study considered the application of our approach in a repository developed based on communicative teaching principles and took into account the complex nature of language learning. The recommendation in this context considers the existence of multiple conceptual connections among objects, which are built as the system is used, and are driven by the application of an algorithm based on Ant Colony Optimization. In order to validate our approach, we also implemented two of the general recommendation algorithms most widely used currently. Whereas a language learning object aims to be a facilitator of grammatical skills integration process, sociolinguistic and strategic, two essential aspects must be considered during the recommendation process: the ability to combine an object with another and its reusability, more specifically, the possibility of its application in different learning contexts. Thus, the results obtained from the experiments demonstrated that the proposed approach is efficient for the application context that it is proposed.

Keywords: Swarm Intelligence, Learning objects, Language Learning objects, Recommender Systems.

LISTA DE FIGURAS

Figura 1	OAL - Metadados em uma ontologia comunicativa	20
Figura 2	SWARM - Experimento da ponte com ramos de tamanhos iguais	36
Figura 3	SWARM - Experimento da ponte com ramos de tamanhos diferentes	37
Figura 4	ACO - Exemplo de construção de uma solução por uma formiga	39
Figura 5	ACO - Seleção do próximo caminho - processo estocástico	39
Figura 6	DICA - Comparação entre temas	48
Figura 7	Exemplo de um grafo de citações	52
Figura 8	ACO - Exemplo de construção de uma solução por um usuário.	57
Figura 9	ACO - Seleção do próximo caminho pelo usuário.	58
Figura 10	ACO - Representação do relacionamento entre OALs.	58
Figura 11	ACO - Construção de uma solução.	59
Figura 12	BALL - Estrutura da tabela Objetos	65
Figura 13	BALL - Estrutura da tabela conexões entre objetos	66
Figura 14	BALL - Conexões entre objetos	66
Figura 15	BALL - Estrutura de tabelas para armazenamento de recomendações	67
Figura 16	BALL - Tela inicial do sistema	68
Figura 17	BALL - Resultado da busca por palavra-chave	68
Figura 18	BALL - Tela de download	69
Figura 19	Gráfico resultante da 1ª etapa de simulação	72
Figura 20	Gráfico resultante da 2ª etapa de simulação	73
Figura 21	Gráfico resultante da 3ª etapa de simulação	75
Figura 22	Exemplo de código utilizado para popular a interface DataModel	76
Figura 23	Exemplo de código para recomendação baseada em usuário	77
Figura 24	Exemplo de conexões resultantes dos experimentos	78
Figura 25	Intensidade das conexões	78
Figura 26	Gráfico de recomendações geradas pelo sistema	80
Figura 27	Percentual de recomendações, geradas pelos três algoritmos, consideradas relevantes para os usuários I, J e K	82
Figura 28	Percentual de recomendações consideradas relevantes, para os usuários I, J e K, em relação ao número total de recomendações esperadas	83
Figura 29	Percentual de recomendações inadequadas para os usuários I, J e K	83
Figura 30	Percentuais obtidos a partir da análise dos três algoritmos	84

LISTA DE TABELAS

Tabela 1	Recomendação baseada em filtragem colaborativa	29
Tabela 2	e-LORS - Categorias descritoras	46
Tabela 3	e-LORS - Campos do LOM e dimensões de preferências	46
Tabela 4	Ex. de recomendações e valores utilizados no cálculo da medida swarm	61
Tabela 5	Número de objetos acessados na 1ª simulação	72
Tabela 6	Número de objetos acessados na 2ª etapa de simulação	73
Tabela 7	Dados utilizados na simulação final	74
Tabela 8	Número de objetos acessados na 3ª etapa de simulação	75
Tabela 9	Ex. de dados de acesso, resultantes das 3 etapas de simulação . .	76
Tabela 10	Valores associados às conexões	79
Tabela 11	Exemplo de recomendações para os usuários I, J e K	79
Tabela 12	Recomendações realizadas a partir da interação dos usuários I, J e K	80
Tabela 13	Classificação do resultado da recomendação de um item para um usuário	82
Tabela 14	Valores obtidos a partir da análise das recomendações	82
15	Lista de objetos acessados	94

LISTA DE ABREVIATURAS E SIGLAS

BALL	Banco de Atividades para a Aprendizagem de Línguas.
BIOE	Banco Internacional de Objetos Educacionais
CLT	Communicative Language Teaching
CA	Classification Accuracy
IEEE	Institute of Electrical and Electronics Engineers
LE	Língua Estrangeira
LOM	Learning Object Metadata
MEC	Ministério da Educação
MIT	Massachusetts Institute of Technology
MAE	Mean Absolute Error
MSE	Mean Squared Error
NMAE	Normalized Mean Absolute Error
OA	Objetos de Aprendizagem
OE	Objetos Educacionais
OAL	Objetos de Aprendizagem de Línguas
OIE	Organização dos Estados Ibero-americanos
PA	Prediction Accuracy
RELPE	Rede Latino-americana de Portais Educacionais
RMSE	Root Mean Squared Error
SI	Swarm Intelligence
SR	Sistemas de Recomendação
SEV	Standard Error Variance
TF-IDF	Term Frequency - Inverse Document Frequency
TSP	Travelling Salesman Problem

SUMÁRIO

1	INTRODUÇÃO	12
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	Objetos de Aprendizagem	16
2.2	Objetos de Aprendizagem de Línguas	19
2.3	Sistemas de Recomendação	21
2.3.1	Recomendação Baseada em Conteúdo	23
2.3.2	Recomendação Colaborativa	27
2.3.3	Recomendação Híbrida	31
2.4	Inteligência de Enxames	33
2.4.1	Otimização por Colônia de Formigas	35
2.4.2	Algoritmo ACO (Ant Colony Optimization)	37
2.5	Apache Mahout	40
2.5.1	Técnicas de Recomendação	41
2.5.2	Clusterização	41
2.5.3	Classificação	42
2.5.4	Recomendações com Mahout	42
3	TRABALHOS RELACIONADOS	44
3.1	Recomendação de Objetos de Aprendizagem	44
3.2	Recomendação de Objetos Associados à Diferentes Contextos	49
4	ACO APLICADO À RECOMENDAÇÃO DE OBJETOS DE APRENDIZAGEM DE LÍNGUAS	56
4.1	Abordagem utilizada	56
4.2	Aplicação do algoritmo ACO	57
5	ESTUDO DE CASO	63
5.1	BALL - Banco de Atividades para Aprendizagem de Línguas	63
5.2	Protótipo	64
6	EXPERIMENTOS E RESULTADOS	71
6.1	Descrição dos experimentos	71
6.2	Resultados obtidos	77
6.3	Avaliação das recomendações	81
7	CONCLUSÕES	85
7.1	Trabalhos futuros	86

REFERÊNCIAS	87
ANEXO A LISTA DE OBJETOS ACESSADOS	94
ANEXO B RECOMENDAÇÕES REALIZADAS A PARTIR DA INTERAÇÃO DOS USUÁRIOS I, J E K	99

1 INTRODUÇÃO

A utilização de recursos online tem se tornado cada vez mais presente na educação, quer seja no âmbito do ensino à distância ou presencial. Diante disso novas formas de ensino e aprendizagem surgiram, assim como diferentes possibilidades de desenvolvimento de conteúdos por meio de objetos de aprendizagem (CAZELLA et al., 2012). Objetos de aprendizagem (OA) podem ser definidos como componentes de um tipo de instrução baseada em computador, fundamentada no paradigma orientado a objetos (WILEY, 2003). Essa característica é evidenciada pela possibilidade de criação de pequenos componentes (em relação a um curso completo) que podem ser reutilizados em diferentes contextos de aprendizagem. Em uma definição mais abrangente, segundo o mesmo autor, um OA pode ser conceituado como qualquer entidade, digital ou não-digital, que pode ser utilizada, reutilizada ou referenciada durante o processo de aprendizagem apoiado pela tecnologia. Como exemplo podemos citar conteúdos multimídia, conteúdos instrucionais, objetivos de aprendizagem, softwares e ferramentas de software instrucionais, pessoas, organizações ou eventos. Com base nesses conceitos, pode-se perceber que esses recursos encontram-se disponíveis na Internet, porém de forma descentralizada, espalhados por toda a rede, o que dificulta o processo de localização e recomendação. Como consequência desta descentralização, deixam de lado seu principal objetivo: auxiliar o processo de ensino e aprendizagem.

A fácil localização, recomendação adequada desses materiais e a disponibilização de recursos em diferentes formatos, além de facilitar a pesquisa, proporciona melhor compreensão do tema que está sendo pesquisado. Um dos principais desafios do educador é a seleção e organização desses materiais digitais.

Os repositórios de objetos de aprendizagem foram criados para esta finalidade, possibilitando a busca e agregação de objetos em um mesmo sistema. Tal característica proporciona maior facilidade à realização de consultas através da divisão por áreas, conteúdos e tipos de objetos (CAZELLA et al., 2012).

Diante deste contexto, através da presente proposta espera-se contribuir para a produção e localização de Objetos de Aprendizagem de Línguas (OAL) a partir da

constituição de um repositório, desenvolvido com base em uma ontologia fundamentada pelos princípios da CLT, voltado não somente para o desenvolvimento linguístico mas também para a formação (continuada) de professores.

Sabe-se no entanto que os repositórios de aprendizagem, apesar de apresentarem a informação de forma centralizada, muitas vezes não possuem sistemas que recomendem recursos educacionais semelhantes ou complementares (VIEIRA; NUNES, 2012). Em função disso, acabam por sobrecarregar o usuário com informações que na maioria das vezes não possuem conteúdos relevantes para o tema que está sendo pesquisado.

Os Sistemas de Recomendação (SR) surgiram como proposta para solução deste problema. Eles funcionam como filtros de informação, auxiliando o usuário na busca e seleção de conteúdo com base em seu perfil. As recomendações são realizadas a partir da análise de diferentes aspectos, como: preferências do usuário, semelhança entre conteúdos já acessados, avaliação dos usuários, entre outros. Um dos grandes desafios deste tipo de sistema é conceber a combinação apropriada entre as expectativas dos usuários e os itens a serem recomendados. A definição e descoberta do tipo de relacionamento de interesses é o problema em questão (CAZELLA; NUNES; REATEGUI, 2010).

No melhor de nosso conhecimento, não se pode encontrar na literatura trabalhos específicos relacionados à recomendação de Objetos de Aprendizagem de Línguas. As abordagens existentes são voltadas para a recomendação de objetos de aprendizagem de modo genérico, sem considerar os aspectos que norteiam a aprendizagem de línguas. Dessa forma, a partir do desenvolvimento do repositório de OAL, tem-se como principal objetivo deste trabalho propor uma estratégia para recomendação, levando em consideração os aspectos pedagógicos ligados a essa classe de objetos e o caráter complexo da aprendizagem de línguas.

A Inteligência Artificial oferece uma série de técnicas que podem ser aplicadas na busca pela solução de problemas complexos em diversas áreas. Entre as quais estão, por exemplo, o estabelecimento de rotas de veículos, o escalonamento de produção na indústria e o projeto de circuitos integrados. Tais técnicas podem oferecer soluções para a localização e recomendação adequada de objetos de aprendizagem quando aplicadas a SR. Nesse contexto, o presente trabalho aborda a aplicação de técnicas de Inteligência de Enxames para a recomendação de Objetos de Aprendizagem de Línguas (OAL).

O modelo proposto baseia-se na combinação de técnicas colaborativas e baseadas em conteúdo em um único sistema de recomendação híbrido baseado em regras, onde a recomendação propriamente dita se dá através da aplicação do Algoritmo de Otimização por Colônia de Formigas.

Conforme já descrito, o estudo considerou a hipótese de aplicação da solução aqui

proposta em um repositório desenvolvido com base em princípios pedagógicos comunicativos, atentando para o caráter complexo da aprendizagem de línguas. Durante o desenvolvimento desta dissertação foi concluída parte da implementação deste repositório, disponibilizando recursos suficientes para a inserção e catalogação de objetos de aprendizagem. Entretanto, a tarefa de popular o repositório com uma quantidade de OAL adequada para realização dos experimentos, demandaria muito tempo. Como solução para o problema, foi desenvolvido um protótipo de sistema web populado com dados obtidos do Banco Internacional de Objetos Educacionais (BIOE).

Apesar dos objetos utilizados nos experimentos não terem sido construídos com base na ontologia aqui prevista, a proposta do presente trabalho não foi inviabilizada. Os experimentos foram realizados de forma controlada e contaram com o auxílio de um professor especialista da área de Letras, e três professoras de inglês em formação, bolsistas de iniciação científica, vinculadas ao projeto do Programa de Pós-Graduação em Letras (mestrado) da Universidade Federal de Pelotas.

Os resultados obtidos a partir dos experimentos demonstraram que a solução proposta se mostra eficiente para o contexto de aplicação a que se propõe, uma vez que retornou resultados mais relacionados ao termo consultado pelos usuários do que outros modelos. Outrossim, proporcionou o estabelecimento de rotas, construídas através da aplicação do conhecimento obtido de forma implícita por meio da interação dos usuários com o sistema.

O conteúdo do presente trabalho segmenta-se em seis partes distintas e complementares, além desta introdução: O Capítulo 2 oferece uma visão geral sobre os conceitos relacionados ao tema da pesquisa, abordando conceitos relacionados aos Objetos de aprendizagem, Objetos de Aprendizagem de Línguas, bem como suas principais características e definições. Também são abordados conceitos relacionados aos Sistemas de Recomendação, apresentando sua arquitetura, classificações e as principais técnicas utilizadas no processo de recomendação. Por fim, são apresentados os conceitos relacionados ao Apache Mahout, à Inteligência de Enxames e o Algoritmo de Otimização por Colônia de Formigas, cujas técnicas são aplicadas na busca pela solução proposta no presente trabalho; O capítulo 3 apresenta os trabalhos relacionados que serviram de embasamento para análise a viabilidade de aplicação da solução proposta. Nesse contexto, foram analisados dois grupos de trabalhos: Recomendação de Objetos de Aprendizagem e Recomendação de Objetos associados à diferentes contextos; O capítulo 4 descreve o Sistema de Recomendação proposto por este trabalho. A abordagem utilizada para desenvolvimento do sistema é apresentada na primeira seção do capítulo. São definidos os dados de entrada a serem utilizados, as técnicas e o tipo de filtragem adotados, os algoritmos implementados e a forma como os dados de entrada e os dados armazenados são combinados pelos algoritmos no processo de recomendação, resultando na saída do sistema; O

capítulo 5 trata da aplicação do Sistema de Recomendação. Apresenta as características do protótipo desenvolvido, o local e a forma de coleta dos dados utilizados para realização dos experimentos, bem como a estrutura criada para armazenamento das informações. São apresentadas as telas e a forma de utilização do sistema, a metodologia utilizada no processo de recomendação e exemplos de códigos utilizados na implementação; O capítulo 6 apresenta a descrição dos experimentos realizados para verificar a viabilidade do modelo proposto. São apresentados os procedimentos realizados, as pessoas envolvidas e a interpretação dos resultados obtidos; Finalmente, o capítulo 7 apresenta as principais conclusões acerca do que foi realizado durante o desenvolvimento da solução proposta e os trabalhos futuros previstos.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo oferece uma visão geral sobre os conceitos relacionados ao tema da pesquisa. As seções que o compõem abordam conceitos relacionados aos Objetos de aprendizagem, Objetos de Aprendizagem de Línguas, bem como suas principais características e definições. Também são abordados conceitos relacionados aos Sistemas de Recomendação, apresentando sua arquitetura, classificações e as principais técnicas utilizadas no processo de recomendação. Por fim, são apresentados os conceitos relacionados à Inteligência de Enxames e o Algoritmo de Otimização por Colônia de Formigas, cujas técnicas são aplicadas na busca pela solução proposta no presente trabalho.

2.1 Objetos de Aprendizagem

Objetos de Aprendizagem são elementos essenciais de um modelo conceitual para a criação e distribuição de conteúdo. Seu principal objetivo é melhorar o desempenho humano em relação ao processo de ensino e aprendizagem (HODGINS, 2002). Compreendidos como entidades digitais ou recursos educacionais, podem ser acessados via internet, armazenados, distribuídos, compartilhados, modificados e recombina- dos com outros, formando módulos independentes de instrução, para serem utilizados on- line ou off-line (SOARES, 2009). Trata-se de um tipo de instrução baseada em com- putador, fundamentada no paradigma orientado a objetos. Essa característica pode ser percebida pela possibilidade de criação de pequenos componentes chamados de objetos, que podem ser utilizados e reutilizados em diferentes contextos de aprendi- zagem (WILEY, 2003).

Na análise realizada por MCGREAL (2004), um estudo das diferentes terminolo- gias utilizadas para conceituar um OA revela que quatro significados distintos podem ser percebidos:

- **Qualquer objeto** - Um Objeto de Aprendizagem pode ser caracterizado como qualquer entidade, digital ou não digital, que pode ser utilizada, reutilizada ou referenciada durante o processo de ensino e aprendizagem. Essa definição,

extremamente ampla, considera que o uso que se faz de um objeto é que o torna, ou não, um OA. Nesse contexto, do ponto de vista teórico, qualquer objeto utilizado para aprendizagem pode ser considerado um OA.

- **Qualquer objeto digital** - A restrição imposta pela necessidade de um OA ser digital deve-se às características inerentes a estes objetos, que são particularmente associadas à ambientes virtuais. Apesar de ser mais restrita, essa definição deve considerar os objetivos para os quais são elaborados os objetos.
- **Qualquer objeto com propósito educacional** - Essa definição não faz distinção entre um objeto digital ou não-digital. Nesse caso, um objeto pode ser considerado um OA desde que possa ser utilizado para apoiar atividades de aprendizagem: um livro, uma página na web, um microscópio ou até mesmo uma calculadora. QUINN; HOBBS (2000), prevê ainda a utilização de blocos ou pedaços de conteúdo educacional, considerando qualquer tipo de mídia, digital ou não, que possam ser agrupados e utilizados para compor uma disciplina ou até mesmo um curso completo.
- **Qualquer objeto digital com objetivo educacional** - Um OA pode ser caracterizado como um arquivo digital cuja aplicação é voltada para fins pedagógicos. Tal arquivo deve incluir internamente, ou de forma associada, sugestões para sua utilização de forma adequada (SOSTERIC; HESEMEIER, 2002). Nesse contexto, um objeto torna-se um OA quando é projetado para ser utilizado de forma individual ou combinado com outros objetos de mídia, tendo como propósito principal promover e facilitar a aprendizagem.

Apesar das discussões, os dois conceitos, digital ou não, tem se perpetuado ao longo dos anos, dividindo as preferências dos autores. Apesar disso, existe uma tendência maior para determinação do conceito de um OA como uma entidade digital (LEFFA, 2006).

Para o propósito deste trabalho será adotada a definição de (MCGREAL, 2004), que caracteriza um OA como qualquer recurso digital reutilizável. Este OA pode ser encapsulado em uma aula, ou conjunto de aulas, agrupados em unidades, módulos, cursos ou programas e deve ter um propósito educacional. Para MCGREAL (2004), a realidade consiste em aceitar a limitação de que os Objetos de Aprendizagem devem ser recursos digitais. O autor observa que, embora isso possa comprometer sua usabilidade, tal limitação não impede um OA de referenciar objetos não-digitais externos.

Além das terminologias acima descritas, os diversos estudos realizados têm destacado características inerentes aos objetos que contribuem para a definição de um conceito mais claro sobre um OA. Estas características são granularidade, reusabili-

dade, interoperabilidade e recuperabilidade e serão descritas a seguir segundo a visão de (LEFFA, 2006):

- **Granularidade** - Refere-se ao termo utilizado para definir o tamanho das estruturas que compõem um OA. Por exemplo, quanto menor o objeto (maior granularidade), mais fácil será sua combinação com diferentes Objetos de Aprendizagem. Esta definição tem sido um dos grandes desafios para os pesquisadores, pois impacta diretamente na capacidade de combinar um objeto com outro (WILEY, 2003). Apesar de não existirem padrões ou critérios para definir a granularidade, uma forma de determiná-la é através do tempo utilizado para execução de uma atividade LEFFA (2006). Para VETROMILLE-CASTRO et al. (2013), um OA pode ser considerado como tendo a granularidade adequada, quando esta permitir a reutilização e interoperabilidade do recurso a um nível considerado relevante para o processo de aprendizagem.
- **Reusabilidade** - Significa reaproveitar o objeto construído em diferentes contextos de aprendizagem, combinando-o com outros objetos e recriando-o à medida que for necessário. Nesse contexto, a reusabilidade sugere a reduplicação do mesmo objeto tornando-o mais aprimorado e eficiente ao processo de aprendizagem. Para que haja reusabilidade, o objeto deve evoluir à medida do tempo, de forma a adaptar-se à constante evolução tecnológica.
- **Interoperabilidade** - É a capacidade de resposta e funcionalidade do OA independente do ambiente digital em que se encontra. A evolução do objeto e sua capacidade de ser adaptável é que o levarão a interoperabilidade.
- **Recuperabilidade** - Refere-se à facilidade de acesso ao OA. Significa proporcionar ao usuário, no menor tempo possível, informação útil e que atenda às suas necessidades. Este acesso é realizado através de uma espécie de catalogação do objeto. A catalogação é realizada utilizando-se descritores, também chamados de metadados¹ do objeto.

Durante o desenvolvimento de Objetos de Aprendizagem têm-se atribuído maior atenção aos aspectos tecnológicos do que à aspectos pedagógicos. Dessa forma, para o ensino e aprendizagem de línguas, faz-se necessário um embasamento teórico diferenciado, que possua a relação adequada com a aprendizagem de línguas estrangeiras e com ênfase na comunicação (VETROMILLE-CASTRO et al., 2013). Nesse contexto, surgem os Objetos de Aprendizagem de Línguas (OAL), descritos na próxima seção.

¹Metadados são comumente descritos como dados sobre dados. No contexto de aplicação deste trabalho eles funcionam como descritores com informações relacionadas aos objetos de aprendizagem.

2.2 Objetos de Aprendizagem de Línguas

Um Objeto de Aprendizagem de Línguas (OAL) pode ser definido como um OA preparado especificamente para o ensino e aprendizagem de línguas. Sua implementação tem por objetivo promover e desenvolver a competência comunicativa. Para este fim, deve ser fundamentado nas teorias do Ensino Comunicativo de Línguas (CLT²) (VETROMILLE-CASTRO et al., 2013).

A definição de um OAL deve ser norteada por aspectos de design de usabilidade e usabilidade pedagógica (VETROMILLE-CASTRO, 2003). Deve ter como objetivo principal a capacidade de facilitar a integração de competências gramatical, sociolinguística e estratégica, características particularmente ligadas ao ensino e à aprendizagem de idiomas (VETROMILLE-CASTRO et al., 2012). Tais aspectos consideram características relacionadas à instrução, orientação e feedback de atividades. Essas características, além de proporcionarem a interação com o aluno, são componentes essenciais no desenvolvimento da competência comunicativa.

A competência comunicativa pode ser definida como uma característica emergente da combinação entre o conhecimento e a habilidade necessária para a comunicação. O conhecimento, que pode ser consciente ou inconsciente, refere-se à informação que dispõe um indivíduo sobre dada linguagem e sobre outros aspectos de sua utilização, enquanto a habilidade refere-se a forma como um indivíduo pode usar o conhecimento em comunicação real (CANALE; SWAIN, 1980; CANALE, 1983).

O trabalho de CANALE; SWAIN (1980) desencadeou uma série de discussões voltadas para o desenvolvimento da competência comunicativa. Foram apresentados cinco princípios essenciais necessários para guiar um programa de ensino de línguas estrangeiras através de uma abordagem comunicativa. Estes princípios foram adaptados aos OAL, e são descritos a seguir, segundo a visão de (VETROMILLE-CASTRO et al., 2013):

- Um Objeto de Aprendizagem de Línguas deve ter como objetivo principal a capacidade de proporcionar a integração de competências gramaticais, sociolinguísticas e estratégicas;
- Objetos de Aprendizagem de Línguas devem focar-se na necessidade de comunicação dos alunos;
- Devem proporcionar a interação dos alunos em situações reais e significativas de comunicação, sejam elas de produção e/ou compreensão oral e/ou escrita. Considera-se como real, situações que, apesar de não estarem realmente ocorrendo no processo de ensino, representam ou retratam possíveis eventos de uso comunicativo da língua fora do contexto da sala de aula;

²CLT refere-se ao termo em inglês Communicative Language Learning.

- Devem atentar para a forma da Língua Estrangeira (LE) em situações de comunicação, conduzindo o aluno do implícito ao explícito no uso da língua;
- E por fim, devem possuir aspectos socio-linguístico-culturais da LE e dos alunos, afim de facilitar a aprendizagem e uso da língua.

Objetos de Aprendizagem fundamentados com base nos princípios acima descritos, podem ser considerados mais do que OAs voltados para o ensino e aprendizagem de línguas. Dada sua especificidade, podem ser caracterizados como artefatos elaborados com o propósito de desenvolver a competência comunicativa dos estudantes (VETROMILLE-CASTRO et al., 2013).

Em seu trabalho, VETROMILLE-CASTRO et al. (2013) aborda a elaboração de uma ontologia fundamentada pelos princípios da CLT. Esta ontologia foi constituída através de metadados elaborados a partir da definição de uma estrutura de linguagem técnica e principalmente comunicativa (Figura 1). Segundo o autor, a construção de um repositório de Objetos de Aprendizagem de Línguas, norteado por esta ontologia, além de ser um recurso importante para estudantes de línguas, significa a concepção de uma ferramenta para a formação continuada de professores de línguas.

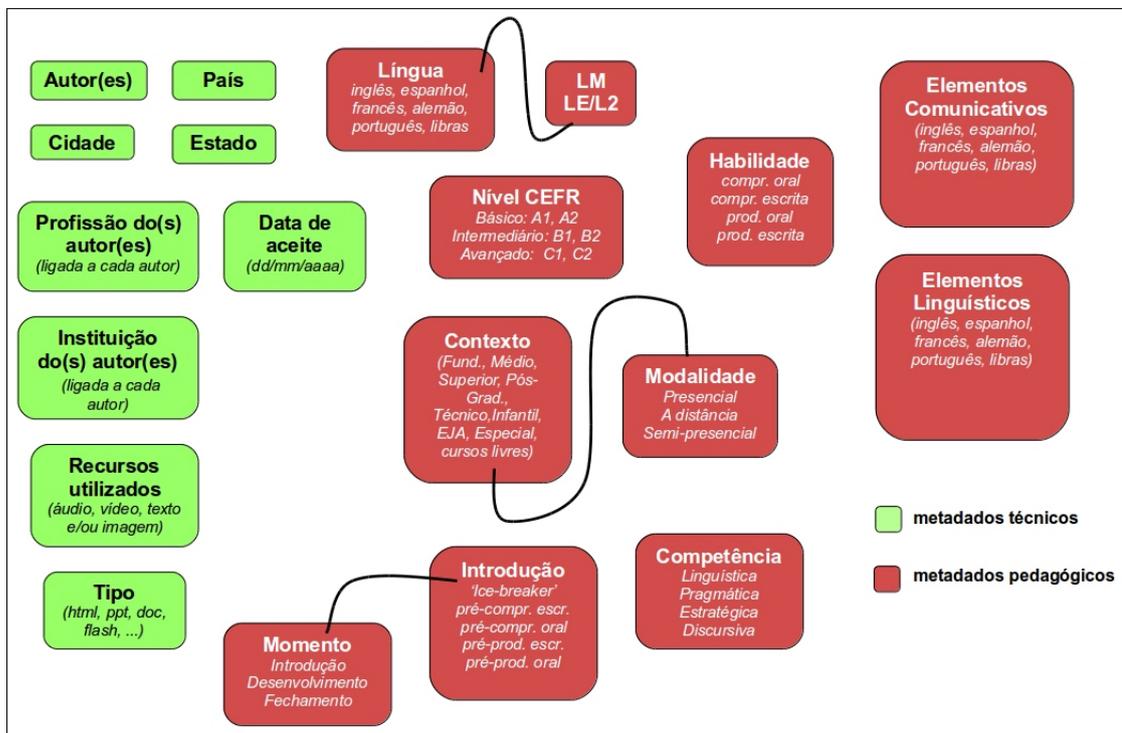


Figura 1: OAL - Metadados em uma ontologia comunicativa
Fonte: VETROMILLE-CASTRO R.; PEREIRA; RAMIRES (2013)

Os metadados apresentados na Figura 1 funcionam como descritores com informações relacionadas aos OAL, utilizados para facilitar sua indexação e

identificação em repositórios. Dessa forma, percebe-se que a adoção de uma ontologia poderá ser utilizada como um recurso importante durante o processo de desenvolvimento de um sistema para recomendação de OAL.

2.3 Sistemas de Recomendação

A revisão da literatura demonstra que as opiniões de diferentes autores convergem para o mesmo conceito acerca de um Sistema de Recomendação: uma ferramenta ou conjunto de técnicas que, independente do escopo ao qual está inserido, tem como propósito principal, através da tarefa de recomendação, prover ao usuário informações e serviços que sejam relevantes, úteis e atendam às suas necessidades (RICCI et al., 2010; ZHOU et al., 2008; BURKE, 2002, 2007; MAHMOOD; RICCI, 2009; RESNICK; VARIAN, 1997).

Os Sistemas de Recomendação surgiram em meados de 1990, a partir de uma ideia cujo propósito era auxiliar na busca de conteúdo online mais útil e relevante, utilizando para tanto, a opinião das pessoas (JANNACH et al., 2010).

O sistema de correio experimental Tapestry introduziu a ideia de filtragem colaborativa e mostrou como a utilização de dados, adquiridos de forma implícita ou explícita, baseados no comportamento dos usuários, poderiam ser aplicados para a criação de filtros pessoais. Alguns anos depois, diferentes sistemas como GroupLens (RESNICK et al., 1994), Ringo do MIT (SHARDANAND; MAES, 1995) e Bellcore Video Recommender (HILL et al., 1995), adotaram técnicas semelhantes para a recomendação de itens relacionados a diferentes contextos, como filmes, músicas, artistas e notícias, tomando por base as preferências dos usuários. No ambiente de aprendizagem online, por exemplo, os SR são aplicados de forma a recomendar materiais de aprendizagem de interesse dos alunos. Podem ser caracterizados como agentes de software inteligentes, cujas recomendações são baseadas nas atividades, ações ou estilos de aprendizagem, descobertos a partir da análise de padrões de navegação (ZHUHAR et al., 2009).

Em aplicações voltadas para o comércio eletrônico e websites, diferentes técnicas são empregadas na tentativa de otimizar a tarefa de recomendação e ofertar produtos de interesse aos clientes, objetivando, desse modo, sua fidelização e consequente aumento da lucratividade (CAZELLA; NUNES; REATEGUI, 2010). Como exemplos de websites bastante conhecidos que utilizam SR, podemos destacar Amazon.com, YouTube, Netflix, Yahoo, Tripadvisor, Last.fm, e IMDb (RICCI; ROKACH; SHAPIRA, 2011).

A tarefa de recomendação consiste em estimar classificações para itens que ainda não foram vistos ou avaliados pelos usuários. Normalmente, essa estimativa se dá através de classificações atribuídas à outros itens, além de outras informações re-

lacionadas ao conjunto de dados disponíveis para avaliação. Dessa forma, dada a possibilidade de classificar itens que ainda não foram avaliados e, considerando que tal classificação baseia-se em avaliações de itens relacionados ao grupo de interesses de um determinado usuário, a recomendação pode ser realizada através da seleção de itens que obtiveram a melhor classificação estimada.

De maneira formal, segundo ADOMAVICIUS; TUZHILIN (2005), o problema de recomendação pode ser representado através da seguinte equação:

$$\forall c \in C, s'_c = \arg \max_{s \in S} u(c, s) \quad (1)$$

Onde C representa o conjunto de todos os usuários e S o conjunto de todos os itens possíveis que podem ser recomendados.

Considerando que o conjunto de dados, tanto em S quanto em C , pode ser significativamente grande, u representa uma função que mede a utilidade do produto s para o usuário c , ou seja, $u : c \times s \rightarrow R$, onde R é um conjunto totalmente ordenado (por exemplo, números inteiros ou números reais não-negativos em um determinado intervalo). Então, para cada usuário $c \in C$, queremos escolher tal item $s' \in S$ que maximize a utilidade para o usuário.

A medida de utilidade u de um item para o usuário depende do escopo da aplicação, podendo ser obtida através de classificações feitas pelo usuário, ou através de uma função de utilidade, calculada de forma arbitrária, como uma função de utilidade baseada no lucro, por exemplo.

Cada característica do usuário $c \in C$, como sexo, idade, renda, entre outras, pode ser utilizada para a definição de um perfil. Em uma abordagem mais simples, o perfil do usuário pode conter apenas um único elemento, como seu identificador na base de dados, por exemplo. De forma análoga, cada elemento $s \in S$, é definido por um conjunto de características ou atributos que representam o objeto. Assim, em um SR de livros, onde S é uma coleção de livros, cada exemplar pode ser representado por diferentes características ou atributos como título, descrição, formato, editora, etc.

O principal problema em um SR, deve-se ao fato de que a utilidade u não é definida em todo o espaço de $C \times S$, mas sim em apenas um subconjunto dele. Ou seja, a utilidade baseia-se apenas em itens previamente classificados pelos usuários, como um subconjunto de filmes já vistos, por exemplo. Sendo assim, a utilidade u deve ser propagada para todo o espaço de $C \times S$, através da aplicação de métodos de aprendizagem de máquina, teoria da aproximação e/ou heurísticas (ADOMAVICIUS; TUZHILIN, 2005).

De maneira geral, em um SR as recomendações são realizadas tomando-se por base características relevantes ao escopo do sistema (ZAINA, 2010). Consequentemente, os métodos utilizados na implementação e sua interface gráfica, também

devem ser desenvolvidos de forma personalizada. Dessa forma, deve-se considerar os itens a serem recomendados para gerar recomendações úteis e eficazes aos usuários (RICCI et al., 2010). Nesse contexto, os SR podem ser classificados de acordo com a abordagem de recomendação utilizada, constituindo três categorias distintas: recomendação baseada em conteúdo, recomendação colaborativa e sistemas híbridos, descritos a seguir.

2.3.1 Recomendação Baseada em Conteúdo

A recomendação baseada em conteúdo consiste em recomendar itens semelhantes àqueles que um determinado usuário gostou no passado (BALABANOVIĆ; SHOHAM, 1997). Através da análise de descrições de itens busca-se identificar aqueles de maior interesse para o usuário.

A abordagem baseada em conteúdo originou-se das pesquisas na área de recuperação da informação (CAZELLA; NUNES; REATEGUI, 2010). Embora necessite de informações adicionais sobre os itens e as preferências do usuário, não exige um grande número de usuários ou histórico de classificações, sendo possível gerar recomendações mesmo se houver apenas um único utilizador (JANNACH et al., 2010).

Devido aos significativos avanços realizados pelas comunidades de recuperação e filtragem da informação e, considerando a importância de várias aplicações baseadas em texto, atualmente diversos sistemas baseados em conteúdo são focados na recomendação de itens que contém informações textuais, como documentos, websites, mensagens e notícias (ADOMAVICIUS; TUZHILIN, 2005). Segundo o mesmo autor, a melhoria associada às tradicionais abordagens de recuperação da informação deu-se através da utilização do perfil do usuário. Essa abordagem possibilitou realizar recomendações com base em informações relacionadas aos gostos, preferências e necessidades do usuário. As informações podem ser obtidas de duas formas:

- Implicitamente - com base na análise de comportamento, como uma pesquisa realizada em busca de algum item, padrões de navegação ou itens que o usuário já adquiriu, por exemplo.
- Explicitamente - através de questionários ou atribuição de notas associadas aos itens recomendados.

Sistemas de Recomendação baseados em conteúdo geralmente são constituídos de métodos (PAZZANI; BILLSUS, 2007):

- para descrição dos itens a serem recomendados;
- para a criação de um perfil do usuário que representa os itens que são de seu interesse;

- para a comparação entre os itens e o perfil do usuário a fim de determinar o que deve ser recomendado.

De maneira formal, a utilidade $u(c, s)$ do item s para o usuário c é estimada com base na utilidade $u(c, s_i)$ atribuída pelo usuário c à itens $s_i \in S$ que são similares ao item s (ADOMAVICIUS; TUZHILIN, 2005).

Assim, considerando $Content(s)$ o perfil de um item, ou seja, um conjunto de atributos que caracterizam o item s , este valor é utilizado para determinar a importância do item para fins de recomendação, calculado de forma geral através da avaliação de um conjunto de características do item s . A partir da obtenção do perfil do usuário, associado aos itens de sua preferência, o item ou conjunto de itens, cujo perfil $Content(s)$ melhor atende às suas necessidades é utilizado na recomendação.

Uma técnica bastante utilizada em SR baseados em conteúdo é a indexação de frequência de termos. Nessa abordagem, informações relacionadas aos documentos e necessidades dos usuários são descritas na forma de vetores com uma dimensão para cada ocorrência de uma palavra na base de dados (CAZELLA; NUNES; REATEGUI, 2010).

Uma das medidas mais conhecidas para especificar pesos de palavras-chave em recuperação da informação é a medida TF-IDF (Term Frequency - Inverse Document Frequency), que pode ser representada da seguinte maneira:

Supondo que N é o número total de documentos que podem ser recomendados para os usuários e que a palavra k_i aparece em n_i documentos.

Considerando que $f_{i,j}$ é o número de vezes em que a palavra k_i aparece no documento d_j . Então $TF_{i,j}$, a frequência do termo (ou frequência normalizada) da palavra k_i no documento d_j , pode ser definida como:

$$TF_{i,j} = \frac{f_{i,j}}{\max_z f_{z,j}} \quad (2)$$

onde o máximo é calculado sobre a frequência $f_{z,j}$ de todas as palavras k_z que aparecem no documento d_j .

Entretanto, palavras que aparecem em muitos documentos não são úteis na distinção de relevância entre um documento e outro. Dado esse fato, a medida de frequência inversa do documento (IDF_i) é muitas vezes utilizada em combinação com a frequência simples do termo ($TF_{i,j}$). A medida de frequência inversa para a palavra k_i pode ser definida como:

$$IDF_i = \log \frac{N}{n_i} \quad (3)$$

Assim, o peso TF-IDF para a palavra k_i no documento d_j pode ser definido da seguinte forma:

$$w_{i,j} = TF_{i,j} \times IDF_i \quad (4)$$

O conteúdo do documento d_j pode ser definido como:

$$Content(d_j) = (w_{1j}, \dots, w_{kj})$$

A frequência que uma determinada palavra ocorre em um documento ou em uma consulta realizada pelo usuário, corresponde a um componente do vetor. Dessa forma, os vetores de documentos mais relevantes para o usuário são aqueles mais próximos aos vetores de consulta.

A definição de similaridade pode ser formalizada estabelecendo *ContentBasedProfile(c)* como sendo o perfil do usuário c , que pode ser obtido através da análise de conteúdo dos itens previamente avaliados pelo usuário. *ContentBasedProfile(c)* pode ser definido como um vetor de pesos $(w_{c_1}, \dots, w_{c_j})$ onde cada peso w_{c_i} denota a importância do termo k_i para o usuário c utilizando-se a medida TF-IDF.

Em sistemas baseados em conteúdo a função utilidade $u(c, s)$ é geralmente definida conforme a Equação 5 (ADOMAVICIUS; TUZHILIN, 2005):

$$u(c, s) = score(ContentBasedProfile(c), Content(s)) \quad (5)$$

A representação na forma de vetores (TF-IDF) de pesos e termos \vec{w}_c e \vec{w}_s pode ser adotada tanto para *ContentBasedProfile(c)* como para *Content(s)*. Do mesmo modo, a função utilidade $u(c, s)$ geralmente é representada na literatura de recuperação de informação por algum tipo de pontuação heurística sobre vetores \vec{w}_c e \vec{w}_s , como por exemplo, a medida de similaridade do cosseno, representada na Equação 6, onde K é o número total de palavras no sistema:

$$u(c, s) = \cos(\vec{w}_c, \vec{w}_s) = \frac{\vec{w}_c \cdot \vec{w}_s}{\|\vec{w}_c\|_2 \times \|\vec{w}_s\|_2} = \frac{\sum_{i=1}^K w_{i,c} w_{i,s}}{\sqrt{\sum_{i=1}^K w_{i,c}^2} \sqrt{\sum_{i=1}^K w_{i,s}^2}} \quad (6)$$

Assim, o cálculo da medida de similaridade do cosseno pode ser obtido considerando o cosseno do ângulo formado pelos vetores que representam os documentos (termos e frequências). A descrição de interesses do usuário ocorre através da aquisição de informações fornecidas pelo próprio usuário ou por meio de ações, como aquisição, visualização e seleção de itens (CAZELLA; NUNES; REATEGUI, 2010).

Um exemplo adotado por ADOMAVICIUS; TUZHILIN (2005) aborda o perfil de um usuário cuja característica principal é a leitura de artigos online relacionados ao tema bioinformática. Técnicas de recomendação baseadas em conteúdo podem recomen-

dar outros artigos relacionados ao tema bioinformática para o usuário c . Uma vez que estes artigos terão mais termos relacionados ao tema bioinformática (por exemplo genoma, sequenciação, proteômica) do que os artigos sobre outros assuntos, *ContentBasedProfile(c)*, conforme definido pelo vetor \vec{w}_c , irá representar tais termos k_i com pesos elevados w_{ic} . Consequentemente, um SR utilizando o cosseno ou uma medida de similaridade relacionada atribuirá maior utilidade $u(c, s)$ à artigos que tem termos em \vec{w}_s com maior peso ponderado e menor utilidade àqueles em que o termo bioinformática possui um peso menor.

Dada a necessidade de estabelecer uma similaridade entre os itens, como no exemplo acima descrito, a filtragem baseada em conteúdo tende a ser indicada para a recomendação de itens textuais, como artigos ou documentos, onde a análise de termos em comum é mais facilmente executada.

Com base na análise dos trabalhos de SHARDANAND; MAES (1995) e BALABANOVIĆ; SHOHAM (1997), ADOMAVICIUS; TUZHILIN (2005) destaca as seguintes limitações associadas aos SR baseados em conteúdo:

- **Análise de conteúdo limitada:** Abordagens baseadas em conteúdo são limitadas pelos recursos explicitamente associados aos objetos a serem recomendados. Conteúdo multimídia, por exemplo, como imagens gráficas, áudio e vídeo são difíceis de serem analisados de forma automática. Da mesma forma, a recomendação de itens como roupas ou brinquedos também poderia ser realizada através desta abordagem. Entretanto, é necessário identificar atributos associados aos objetos, como peso, preço, marca, entre outros, o que na prática pode ser trabalhoso ou dificultado pela limitação de recursos.

Documentos baseados em texto são mais fáceis de serem analisados, porém, dois itens distintos que são representados pelo mesmo conjunto de características ou palavras-chave não podem ser distinguidos pelo sistema, como por exemplo, um artigo bem escrito e um mal escrito que utilizam os mesmos termos.

- **Super especialização:** Considerando que SR baseados em conteúdo tendem a recomendar itens que o usuário gostou ou avaliou positivamente no passado, o usuário está limitado a receber apenas recomendações semelhantes àquelas que já foram classificadas e possuem uma pontuação elevada em relação ao seu perfil. Outro problema a ser considerado é que, dependendo do contexto de aplicação, a recomendação de itens muito semelhantes aos que o usuário já visualizou pode não ser positiva, como uma notícia diferente descrevendo o mesmo evento, por exemplo.

Para ADOMAVICIUS; TUZHILIN (2005) uma característica desejável em um SR é a diversidade de recomendações, ou seja, o sistema deve ser capaz de

apresentar uma variedade de opções ao usuário, não apenas um conjunto homogêneo de alternativas.

- **Problema do novo usuário:** Tendo em vista a necessidade de conhecer as preferências do usuário, para que a recomendação seja realizada de forma confiável, um SR baseado em conteúdo necessita que os itens a serem recomendados possuam um certo número de classificações. Portanto, um novo usuário, com poucas ou nenhuma avaliação sobre os itens, provavelmente não receberá recomendações precisas.

2.3.2 Recomendação Colaborativa

A recomendação colaborativa foi desenvolvida para complementar a recomendação baseada em conteúdo (BALABANOVIĆ; SHOHAM, 1997). Seu diferencial é caracterizado pelo fato de não estar ligada às características do objeto, ou seja, compreender ou reconhecer o conteúdo dos itens, torna-se dispensável (CAZELLA; NUNES; REATEGUI, 2010).

Ao contrário dos métodos de recomendação com base em conteúdo, SR colaborativos tentam prever a utilidade de itens para um determinado usuário. Para tanto, deve-se tomar por base itens previamente avaliados por outros usuários. O sistema reconhece semelhanças entre os usuários a partir de um conjunto de avaliações ou recomendações de objetos, gerando novas recomendações baseadas neste relacionamento (BURKE, 2002). Para CAZELLA; NUNES; REATEGUI (2010) a ideia principal consiste em explorar a troca de experiências entre pessoas que possuem interesses em comum. Esta técnica permite lidar com qualquer tipo de conteúdo, não apenas aqueles observados no passado. Ela possibilita que os itens sejam avaliados por diferentes usuários, o que pode melhorar o desempenho do sistema em relação ao processo de recomendação (BALABANOVIĆ; SHOHAM, 1997).

Sistemas de recomendação colaborativos geram previsões ou recomendações para um determinado usuário de um ou mais itens. De maneira formal, a recomendação colaborativa pode ser representada da seguinte maneira (ADOMAVICIUS; TUZHILIN, 2005):

a utilidade $u(c_j, s)$ do item s para o usuário c é estimada com base na utilidade $u(c_j, s)$ por aqueles usuários $c_j \in C$ que são semelhantes ao usuário c .

O termo usuário pode ser atribuído à qualquer pessoa responsável pelo provimento de classificações para um sistema. Geralmente este termo é utilizado para referenciar pessoas que utilizam um sistema para receber informações, como recomendações, por exemplo. Entretanto pode referir-se também àqueles que fornecem dados (classificações) utilizados na produção de informações.

O termo item refere-se a qualquer coisa para a qual um ser humano pode atribuir uma classificação, como livros, CDs, artigos de jornal ou destinos de férias. As classificações por sua vez, podem assumir diferentes formas (SCHAFER et al., 2007):

- **Classificações escalares** - podem ser classificações numéricas, como as 5 estrelas previstas no MovieLens ou classificações ordinais, como concordo, neutro, discordo, discordo totalmente.
- **Classificações Binárias** - modelo de classificações com opções entre concordo/discordo ou bom/ruim.
- **Classificações Unárias** - pode indicar que um usuário tenha observado ou comprado um item, ou não avaliou o item de forma positiva.

Assim, a tarefa principal consiste na previsão de utilidade de um item para um determinado usuário. Esta previsão dá-se através da utilização de uma base de dados, constituída a partir de avaliações de uma amostra ou população de outros usuários.

A ausência de uma classificação indica que não se tem informações sobre o usuário para o item. Sendo assim, o usuário deve atribuir uma pontuação para cada item experimentado de forma a classificá-lo em um conjunto de itens de seu interesse. Esta pontuação pode ser atribuída de forma explícita ou implícita (BREESE; HECKERMAN; KADIE, 1998):

- **Explícita** - ocorre através da interação direta do usuário com o sistema. O usuário expressa sua preferência por um item, geralmente através de uma escala numérica discreta, como a classificação de um filme como bom ou ruim em uma escala de um a cinco, por exemplo.
- **Implícita** - dá-se através da interpretação de comportamento, ou seleções do usuário para imputar um voto ou preferência. Para tanto, pode basear-se em dados de navegação, histórico de compras, entre outros tipos de padrões de acesso à informação. O sistema infere preferências do usuário a partir de suas ações, não sendo exigido que o mesmo avalie de forma explícita o item.

O sistema de recomendação de filmes MovieLens (GOOD et al., 1999) é um exemplo de aplicação baseado em filtragem colaborativa. A partir de pontuações atribuídas à filmes já assistidos, o sistema encontra usuários com gostos semelhantes para recomendar filmes que ainda não tenham visto.

Um usuário do MovieLens atribui classificações à filmes através de uma pontuação que varia de uma (1) a cinco (5) estrelas, onde uma (1) significa que o filme é muito ruim e cinco (5) é a melhor classificação atribuída, indicando que o usuário gostou e recomenda o filme. Em seguida o sistema utiliza as classificações para recomendar

outros filmes que podem ser de interesse do usuário, ou ainda, para realizar previsões acerca de como o usuário poderia avaliar outros filmes.

A classificação consiste na associação entre os usuários e os itens, geralmente por meio de algum valor. Uma maneira de visualizar classificações é como uma matriz. Esta matriz pode ser representada por uma tabela onde cada linha representa um usuário, cada coluna representa um filme específico e o valor na intersecção de uma linha e uma coluna representa o valor de classificação atribuído pelo usuário. A ausência de uma pontuação de classificação nesta intersecção significa que o usuário ainda não avaliou o item (SCHAFER et al., 2007).

A Tabela 1 reproduz de forma prática o funcionamento da filtragem colaborativa. Neste exemplo, para recomendar um filme ao usuário seis (6), deve-se procurar outros usuários com preferências semelhantes. Neste caso, os usuários um (1) e dois (2) já assistiram filmes que o usuário seis (6) também assistiu (Filme 2). Portanto, pode-se recomendar ao usuário seis (6) filmes que estes dois outros usuários já assistiram, mas que o usuário seis (6) ainda não assistiu, como os Filmes 1 e 5.

A decisão sobre a recomendação destes itens deve ainda basear-se no histórico de avaliações comuns e o valor de predição calculado a partir das classificações atribuídas pelos usuários.

Tabela 1: Recomendação baseada em filtragem colaborativa

Usuário	Filme 1	Filme 2	Filme 3	Filme 4	Filme 5	Filme 6
1		x			R(x)	
2	R(x)	x				
3			x	x	x	
4			x			
5	x			x		
6	?	x			?	

Fonte: Adaptada de CAZELLA; NUNES; REATEGUI (2010)

A técnica de filtragem colaborativa pode ser definida em três passos, descritos conforme a seguir, segundo a visão de (CAZELLA; NUNES; REATEGUI, 2010):

- Calcular o peso de cada usuário em relação à similaridade ao usuário alvo (métrica de similaridade).
- Classificar um subconjunto de usuários com maior similaridade (vizinhos) para utilizar na predição.
- Normalizar as avaliações e computar as predições ponderando as avaliações dos vizinhos com seus pesos.

A definição da similaridade pode ser obtida através da aplicação de diferentes coeficientes de similaridade como Coseno, Pearson, entre outros. A Equação 7 apresenta o coeficiente de correlação de Pearson (amplamente abordado na literatura sobre SR) e adotado também pelo GroupLens (RESNICK et al., 1994), um dos primeiros sistemas de filtragem colaborativa:

$$cor_{ab} = \frac{\sum_i (r_{ai} - \bar{r}_a)(r_{bi} - \bar{r}_b)}{\sqrt{\sum_i (r_{ai} - \bar{r}_a)^2 \sum_i (r_{bi} - \bar{r}_b)^2}} \quad (7)$$

Onde:

- cor_{ab} é a correlação do usuário alvo \underline{a} com um dado usuário \underline{b} ;
- r_{ai} é a avaliação que o usuário ativo \underline{a} atribuiu para o item i ;
- r_{bi} é a avaliação que o usuário ativo \underline{b} atribuiu para o item i ;
- \bar{r}_a é a média de avaliações do usuário ativo \underline{a} , em comum com o usuário \underline{b} ;
- \bar{r}_b é a média de avaliações do usuário ativo \underline{b} , em comum com o usuário \underline{a} .

Segundo CAZELLA; NUNES; REATEGUI (2010) é preciso mais de uma avaliação em comum para que o índice seja útil, com resultados variando entre um (1) para similaridade total, e menos um (-1) para total dissimilaridade.

A predição é calculada independentemente do coeficiente empregado para o cálculo de similaridade. Dá-se através de uma média ponderada das avaliações dos vizinhos que apresentaram um coeficiente de similaridade aceitável, tomando-se por base o limiar adotado. A equação a seguir apresenta o cálculo da predição segundo a visão do mesmo autor:

$$p_{ai} = \bar{r}_a + \frac{\sum_{b=1}^n (r_{bi} - \bar{r}_b) * cor_{ab}}{\sum_{b=1}^n |cor_{ab}|} \quad (8)$$

Onde:

- cor_{ab} é a correlação do usuário alvo \underline{a} com um determinado usuário \underline{b} ;
- p_{ai} é a predição de um item i para um usuário alvo \underline{a} ;
- \bar{r}_a é a média de avaliações do usuário alvo \underline{a} aos itens que foram pontuados por todos os seus usuários similares;
- r_{bi} é a avaliação que o usuário alvo \underline{b} atribuiu para o item i ;
- \bar{r}_b é a média de avaliações do usuário \underline{b} , em comum com o usuário \underline{a} .

O algoritmo de filtragem colaborativa mais conhecido e que faz uso das técnicas acima descritas é o algoritmo do vizinho mais próximo, que pode ser dividido em duas classes ou abordagens distintas: baseado em usuário e baseado em item (SCHAFER et al., 2007).

- **Baseado em usuário:** Gera recomendações para um determinado usuário com base em avaliações de usuários semelhantes. Se um usuário n é semelhante a um usuário u , diz-se que n é um vizinho de u .
- **Baseado em item:** Gera recomendações com base na similaridade entre os itens. A previsão para um item deve ser baseada em avaliações de um usuário para itens semelhantes.

A principal vantagem associada à utilização da filtragem colaborativa é a possibilidade de apresentar aos usuários recomendações inesperadas, ou seja, recomendações de itens que não foram ativamente pesquisados. Além disso, proporciona a formação de comunidades de usuários através da análise de similaridade entre seu gostos e interesses (CAZELLA; NUNES; REATEGUI, 2010).

Apesar das vantagens anteriormente citadas, em sistemas colaborativos, uma questão importante a ser observada é a coleta de informações dos usuários, que pode apresentar as seguintes limitações (ADOMAVICIUS; TUZHILIN, 2005):

- **Problema do novo item:** Sistemas colaborativos geram recomendações baseados nas preferências dos usuários. Deste modo, até que um novo item tenha um número significativo de avaliações, o sistema não será capaz de recomendá-lo de forma eficiente.
- **Problema de pontuações esparsas:** dá-se quando o número de avaliações é muito inferior em relação ao número de classificações necessárias. Além disso, para o usuário com gostos incomuns em comparação ao resto da população, não existirá outros usuários com gostos semelhantes, levando à recomendações pobres.
- **Problema do novo usuário:** Assim como em sistemas baseados em conteúdo, para fazer recomendações precisas, o sistema depende de avaliações prévias para aprender as preferências do usuário. A maioria das técnicas utilizadas para contornar este problema utiliza a abordagem híbrida, combinando a filtragem baseada em conteúdo e a filtragem colaborativa.

2.3.3 Recomendação Híbrida

Um sistema híbrido baseia-se na combinação de técnicas de recomendação, objetivando suprir limitações particularmente associadas a cada técnica (RICCI; ROKACH; SHAPIRA, 2011), (ADOMAVICIUS; TUZHILIN, 2005).

A combinação de técnicas para otimização de resultados não é algo novo. BURKE (2002) já atentava para o fato de que as técnicas de recomendação conhecidas possuíam pontos fortes e fracos. Desta forma, sua combinação de diferentes maneiras, constitui em uma abordagem bastante utilizada por muitos pesquisadores na tentativa de otimizar o processo de recomendação.

As diferentes formas de combinar métodos colaborativos e baseados em conteúdo em um SR Híbrido podem ser classificadas da seguinte forma segundo ADOMAVICIUS; TUZHILIN (2005):

- **Implementação de métodos de colaboração e baseados em conteúdo separadamente, combinando suas previsões:** Consiste na combinação de recomendações, adquiridas a partir de SR individuais, em uma recomendação final. A principal vantagem na utilização desta abordagem é a possibilidade de realizar ajustes em ambas as técnicas de forma separada, beneficiando a recomendação final (CLAYPOOL et al., 1999). Outra forma, seria avaliar as recomendações com base em alguma métrica de qualidade ou, de maneira individual, utilizando a que melhor atende às necessidades do usuário (TRAN; COHEN, 2000).
- **Incorporando algumas características baseadas em conteúdo em uma abordagem colaborativa:** Nesta abordagem, diferente de uma abordagem puramente colaborativa, baseada em classificações feitas pelo usuário, o vetor de recomendações é aumentado tomando-se por base também o perfil do usuário (BALABANOVIĆ; SHOHAM, 1997). Assim, utiliza-se também o histórico relacionado aos itens de sua preferência, e não apenas as classificações, para calcular a similaridade entre dois usuários.
- **Incorporando algumas características colaborativas em uma abordagem baseada em conteúdo:** Consiste na utilização de técnicas para redução de dimensionalidade em um grupo de perfis com base em conteúdo. Um exemplo pode ser visto em (SOBOROFF; NICHOLAS, 1999), que utiliza a indexação semântica latente para criar uma visão colaborativa de uma coleção de perfis do usuário.
- **Construção de um modelo unificador geral que incorpora tanto características baseadas em conteúdo e colaborativas:** Esta abordagem sugere a utilização de características peculiares à cada técnica, combinando recomendações colaborativas e baseadas em conteúdo em único recomendador baseado em regras, (BASU et al., 1998). A recomendação propriamente dita, dá-se através da utilização de alguma heurística ou regra probabilística (AN-

SARI; ESSEGAIER; KOHLI, 2000; POPESCU; PENNOCK; LAWRENCE, 2001; SCHEIN et al., 2002).

2.4 Inteligência de Enxames

O termo inteligência de enxames surgiu no final de 1980 como referência à sistemas robóticos celulares nos quais um conjunto de agentes simples em um determinado ambiente interagem com base em regras locais (CASTRO, 2007). Atualmente, diversas abordagens têm sido propostas através da modelagem de comportamento de enxames de animais e insetos como formigas, cupins, abelhas, pássaros e peixes.

A Inteligência de Enxames (SI) é uma técnica para o desenvolvimento de algoritmos de otimização projetados com base no comportamento coletivo e na auto-organização dos insetos sociais (BONABEAU; DORIGO; THERAULAZ, 1999). O termo enxame é utilizado para designar um grupo de animais, tais como um cardume de peixes, um bando de pássaros ou colônias de insetos, como formigas, cupins, abelhas, entre outros, que executam comportamento coletivo.

A auto-organização pode ser caracterizada como um agrupamento dinâmico de regras, cujas respostas às interações de baixo nível de seus elementos menos complexos, resultam em padrões complexos em um nível global (BOFFO; FERREIRA JR; BAZZAN, 2007). A aplicação destas regras deve tomar por base apenas informações locais, sem demais referências para o nível global ou macroscópico (FERREIRA JR; BOFFO; BAZZAN, 2007). Esta emergência de novos padrões, sem supervisão, é a característica fundamental de sistemas complexos.

A auto-organização consiste em quatro princípios básicos (BONABEAU; DORIGO; THERAULAZ, 1999)

1. **Feedback positivo** - são regras comportamentais capazes de promover a criação de estruturas. Recrutamento e reforço tais como assentamentos de trilhas a seguir (característicos de algumas espécies de formigas), ou danças (em abelhas) podem ser mostrados como exemplos de respostas positivas.
2. **Feedback negativo** - Utilizado para contrabalançar o feedback positivo. Ajuda a estabilizar o padrão coletivo, podendo assumir a forma de saturação, exaustão, ou competição. Tomando-se o forrageamento como exemplo, o feedback negativo pode decorrer do número limitado de forrageiras disponíveis, saciedade, aglomeração, concorrência, ou exaustão da fonte de alimentos.
3. **Flutuações** - A aleatoriedade além de proporcionar o surgimento de estruturas emergentes, é muitas vezes considerada crucial. Assim como permitem a descoberta de novas soluções, as flutuações, tais como passeios aleatórios, erros ou alternância de tarefas aleatórias entre indivíduos do enxame, são vitais para

a criatividade e inovação. Elas podem atuar como sementes das quais as estruturas de núcleos emergem. Como exemplo, podemos citar forrageiras que se perdem em uma colônia e encontram novas fontes de alimentos ainda não exploradas.

4. **Variedade de interações** - A variedade de interações é fundamental para a auto-organização. Os agentes devem ser capazes de fazer uso de suas próprias informações, bem como de informações provenientes de outros agentes. Dessa forma, as informações e os dados poderão ser propagados pela colônia.

Os agentes em um enxame se comportam sem supervisão e cada um deles tem um comportamento estocástico relacionado à sua percepção do ambiente (KARABOGA; AKAY, 2009). Uma colônia de insetos, por exemplo, com centenas de milhares de membros, opera sem qualquer coordenação explícita. Um agente possui apenas informação local simples, sendo incapaz de avaliar as necessidades da colônia.

A principal característica do comportamento emergente é a plasticidade na divisão do trabalho dentro da colônia (ROBINSON, 1992). O ambiente e seus recursos são utilizados de maneira eficaz através do uso da inteligência coletiva, respondendo às mudanças de condições internas e externas.

As características principais que compõem um sistema de inteligência de enxame são (MILLONAS, 1994):

- **Proximidade** - deve haver interação entre os agentes do enxame;
- **Qualidade** - os agentes devem ser capazes de avaliar seus comportamentos, respondendo a fatores de qualidade no ambiente, tais como segurança local ou qualidade de produtos alimentares;
- **Diversidade** - permite ao sistema reagir a situações inesperadas, utilizando seus recursos de maneira adequada;
- **Estabilidade** - o comportamento de um agente não deve mudar a cada flutuação do ambiente;
- **Adaptabilidade** - capacidade de adequação a variações ambientais, alterando seu comportamento quando a economia ou o investimento de energia for conveniente.

Entre as abordagens mais populares em inteligência de enxames, pode-se destacar o algoritmo de otimização por colônias de formigas, descrito na seção a seguir.

2.4.1 Otimização por Colônia de Formigas

Apesar da simplicidade de seus agentes, colônias de formigas, assim como outros grupos de insetos sociais, constituem sistemas distribuídos que apresentam uma organização altamente estruturada (DORIGO; DI CARO; GAMBARDELLA, 1999). Essa organização emergente do comportamento coletivo resulta na realização de tarefas complexas que na maioria das vezes não poderiam ser executadas por um único agente. Como exemplo, podemos observar o alto nível de estruturação que uma colônia de formigas pode atingir.

Outro comportamento a ser observado é o forrageamento. Particularmente a forma como as formigas encontram o caminho mais curto entre seu ninho e a fonte de alimentos. Tal característica está relacionada ao conceito de estigmergia - uma forma de comunicação indireta, utilizada por insetos sociais para coordenar suas atividades (DORIGO; BONABEAU; THERAULAZ (2000).

Estigmergia foi o termo adotado por GRASSÉ (1959) para definir o estímulo dos agentes através da performance alcançada. Ele observou que os insetos são capazes de responder aos chamados estímulos significativos, que ativam uma reação geneticamente codificada (GRASSÉ, 1946). Em insetos sociais como cupins e formigas os efeitos dessas reações podem atuar como novos estímulos, tanto para o agente que os produziu, como para outros indivíduos da colônia.

A estigmergia possui duas características principais (DORIGO; BIRATTARI; STÜTZLE, 2006):

- É uma forma indireta, não-simbólica de comunicação, mediada pelo ambiente: insetos trocam informações modificando seu meio ambiente;
- As informações são locais: elas só podem ser acessadas por aqueles agentes que visitam o local em que a informação está disponível (ou sua vizinhança imediata).

O princípio básico da estigmergia pode ser descrito como um rastro no ambiente deixado pelo agente durante a realização de alguma tarefa. Esse rastro torna-se um estímulo para a realização de tarefas subsequentes, tanto para o agente que o deixou quanto para outros agentes da colônia. Esta mediação através do ambiente garante que as tarefas sejam executadas de forma ordenada, sem qualquer tipo de controle, planejamento, ou interação direta entre os agentes (HEYLIGHEN, 2011). Nesse contexto, uma classe de mecanismos estigmérgicos, particularmente, os chamados algoritmos de otimização por colônia de formigas, vem sendo amplamente explorados na tentativa de encontrar soluções para uma variedade de problemas computacionais.

Em algoritmos de formigas os agentes são denominados formigas artificiais. A coordenação entre os agentes é obtida através da exploração do mecanismo de comunicação estigmérgico (DORIGO; BONABEAU; THERAULAZ, 2000).

A implementação de algoritmos de formigas dá-se através da utilização das chamadas variáveis estigmérgicas. Variáveis que contém a informação utilizada por formigas artificiais para comunicar-se de forma indireta.

Em formigas da espécie *I. humilis* (GOSS et al., 1989), *Linepithema humile*, e *Lasius niger* (BONABEAU et al., 1997), por exemplo, o comportamento de forrageamento é baseado na comunicação indireta mediada por feromônios. Enquanto caminham do ninho para a fonte de alimentos e vice-versa, as formigas depositam feromônios no chão, formando um rastro ao longo do percurso. Outras forrageiras podem sentir o cheiro do feromônio depositado e tendem a escolher, probabilisticamente, caminhos com maior concentração de feromônio (DORIGO; BIRATTARI; STÜTZLE, 2006). Analisando esse comportamento, DENEUBOURG et al. (1990) demonstrou através de um experimento como as formigas da espécie *Linepithema humile*, conseguem encontrar o caminho mais curto entre uma fonte de alimento e seu ninho. O experimento utilizou uma fonte de alimentos ligada a um ninho de formigas por uma ponte com dois ramos de mesmo tamanho. Inicialmente, as formigas selecionam aleatoriamente, com igual probabilidade, um dos ramos (Figura 2). Devido à flutuações estatísticas, um dos ramos tende a ser escolhido por mais formigas do que o outro, recebendo uma quantidade superior de feromônio. Dessa forma, a quantidade de feromônio depositada sobre este ramo estimula as formigas à escolhê-lo (DENEUBOURG; GOSS, 1989). Após algum tempo, este processo auto-catalítico³ leva a colônia de formigas a convergir para a utilização de apenas um dos ramos (DORIGO; BONABEAU; THERAULAZ, 2000).

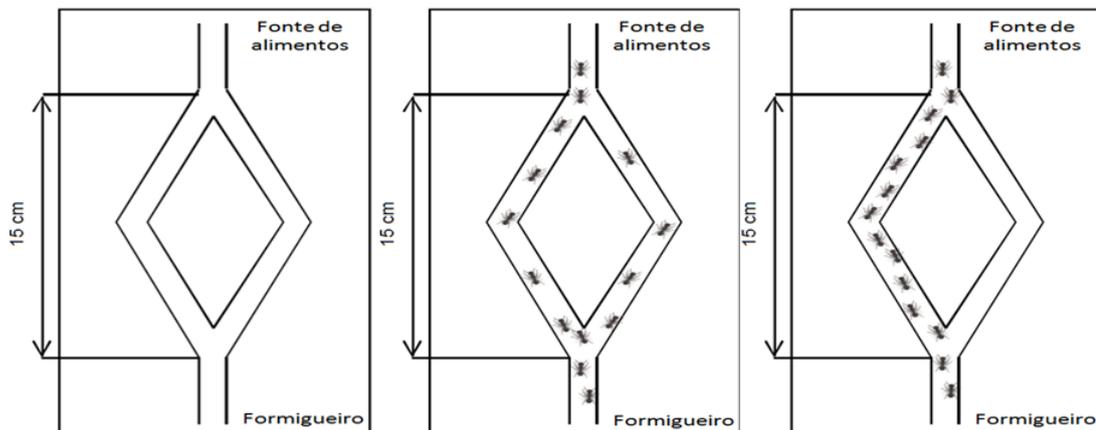


Figura 2: SWARM - Experimento da ponte com ramos de tamanhos iguais
Fonte: DORIGO; BONABEAU; THERAULAZ (2000)

Ao modificar o experimento, utilizando dois ramos de comprimentos diferentes (Figura 3), as primeiras formigas a retornar para o ninho foram aquelas que escolheram

³Um processo auto-catalítico é um processo de auto-reforço decorrente da utilização do feromônio depositado pelas formigas.

o caminho mais curto. Isto porque, ao deslocar-se para a fonte de alimentos e retornar para o ninho, estas formigas executam duas vezes o percurso. Conseqüentemente, uma quantidade maior de feromônio é depositada no ramo mais curto, estimulando outras formigas a seguirem por este caminho.

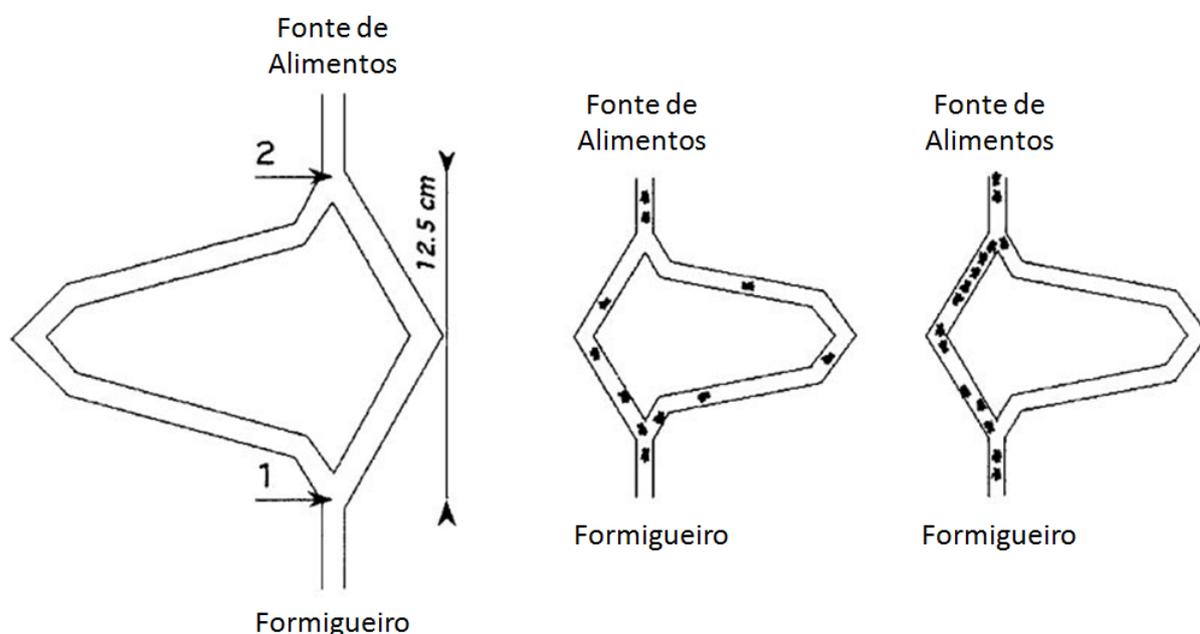


Figura 3: SWARM - Experimento da ponte com ramos de tamanhos diferentes
Fonte: DORIGO; BONABEAU; THERAULAZ (2000)

Ant System foi proposto por (DORIGO et al., 1991), sendo o primeiro algoritmo fundamentado no comportamento das formigas a ser desenvolvido e aplicado à um problema de otimização discreta. As primeiras experiências foram aplicadas ao problema do caixeiro viajante (Travelling Salesman Problem - TSP) (DORIGO et al., 1991). No TSP, o objetivo principal consiste em encontrar o caminho mais curto, realizando todo o percurso e passando uma única vez por cada cidade, utilizando para tanto, as trilhas de feromônios deixadas por um grupo de formigas artificiais.

Ant System deu origem à diversas aplicações bem sucedidas, assim como extensões que foram unificadas em uma nova metaheurística ⁴ denominada Ant Colony Optimization (ACO) (DORIGO; BONABEAU; THERAULAZ, 2000).

2.4.2 Algoritmo ACO (Ant Colony Optimization)

Ant Colony Optimization - ACO é um termo de uso geral utilizado para determinar uma classe de procedimentos metaheurísticos fundamentados no comportamento de

⁴Uma metaheurística pode ser formalmente definida como um processo de geração iterativa que orienta uma heurística subordinada a regiões promissoras do espaço de busca na obtenção de soluções de alta qualidade. Trata-se de um conjunto de métodos aproximados projetados para a solução de problemas complexos de otimização onde métodos heurísticos clássicos não atingem bons resultados (OSMAN; LAPORTE, 1996).

formigas (SERAPIÃO, 2009).

Em um algoritmo ACO os agentes se comunicam indiretamente através da estigmergia, cooperando na tentativa de encontrar soluções para problemas de otimização discreta difíceis. Esta comunicação indireta, mediada pelo ambiente, pode levar a boas soluções, que são uma propriedade emergente da interação entre os agentes. Os agentes agem simultaneamente e individualmente e, apesar de cada um possuir a capacidade necessária para encontrar a solução do problema em questão, boas soluções são obtidas apenas através da interação coletiva entre eles.

De maneira informal um algoritmo ACO pode ser representado através da interação de três procedimentos básicos (DORIGO et al., 2008):

- Construir soluções com formigas: consiste na utilização de um método construtivo e aleatorizado para obtenção da população inicial (possíveis soluções iniciais) utilizando a informação heurística e a quantidade de feromônio de forma combinada;
- Aplicar busca local: método opcional utilizado para melhorar a solução através da aplicação de busca local com base em soluções obtidas por um ou mais agentes;
- Atualizar feromônio: procedimento utilizado para atualizar o feromônio, levando em consideração o índice de evaporação.

Algorithm 1 Pseudocódigo ACO

```

1: procedure ACO
2:   while criterioDeParadaNaoSatisfeito do
3:     ProgramarAtividades
4:       ConstruirSoluções()
5:       AtualizarFeromônio()
6:       AplicarBuscaLocal()                                ▷ opcional
7:     EndProgramarAtividades
8:   end while
9: end procedure

```

A metaheurística ACO tem como princípio básico coordenar o escalonamento dos três procedimentos anteriormente descritos. A execução destes procedimentos não possui uma regra específica, podendo ser executados de forma sincronizada, paralelamente ou independentemente, de acordo com as características do problema em questão (DORIGO et al., 2008).

Em otimização por colônia de formigas, o problema é representado através da simulação de um número de formigas artificiais que se deslocam em um grafo. As

formigas percorrem o grafo construindo a solução, passando pelos vértices e pelas arestas, depositando feromônio nas arestas (Figura 4).

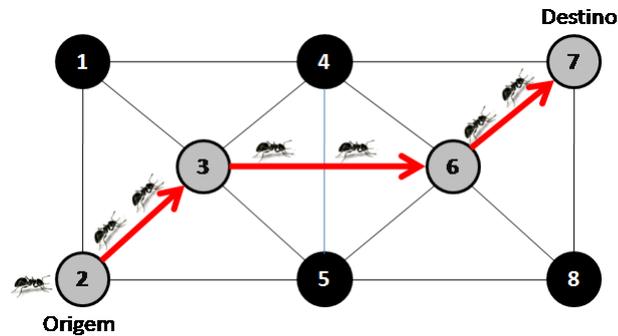


Figura 4: ACO - Exemplo de construção de uma solução por uma formiga
Fonte: Elaborada pelo autor

Cada aresta (i, j) do grafo possui uma variável τ_{ij} que representa uma trilha de feromônio artificial. Essa variável pode ser modificada pelas formigas, sendo incrementada de acordo com o percurso realizado por elas. Assim, quanto maior o nível de feromônio, maior a chance da aresta ser visitada por outras formigas.

Conforme o exemplo apresentado pela figura 4, dado um grafo $G = (i, j)$, o caminho mais curto entre um determinado par de vértices pode ser encontrado. Nesse exemplo, a formiga percorre o grafo partindo inicialmente do vértice 2 e chegando até o vértice 7, obtendo como uma das possíveis soluções o caminho 2-3-6-7.

Em cada passo da construção da solução, a formiga seleciona o próximo vértice a ser visitado de acordo com um mecanismo estocástico que é influenciado pelo feromônio: quando no vértice i , o seguinte vértice é selecionado estocasticamente entre os anteriormente visitados (Figura: 5).

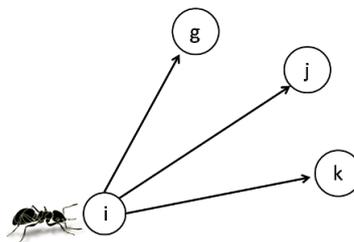


Figura 5: ACO - Seleção do próximo caminho - processo estocástico
Fonte: DORIGO; BIRATTARI; STUTZLE (2006)

Se j ainda não foi visitado, ele pode ser selecionado, com uma probabilidade que é proporcional ao índice de feromônio associado à aresta (i, j) .

Através de uma fórmula probabilística, cada formiga k constrói uma solução movendo-se por uma sequência de locais vizinhos. Ao final de cada iteração o fe-

romônio é atualizado tomando-se por base a qualidade das soluções produzidas pelas formigas.

A equação é dada por:

$$p_{ij}^k = \frac{[\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{l \in J_i^k} [\tau_{il}]^\alpha \cdot [\eta_{il}]^\beta} \quad (9)$$

onde:

p_{ij}^k : é a probabilidade da formiga k , que se encontra no vértice i , escolher o vértice j como o próximo a ser visitado;

τ_{ij} : quantidade de feromônio existente na aresta (i, j) . Inicialmente, adota-se um mesmo valor τ_0 para todas as arestas do grafo;

η_{ij} : heurística que representa a atratividade da aresta (i, j) ;

$l \in J_i^k$: conjunto de vértices ainda não visitados pela formiga k , que se encontra no vértice i ;

α : parâmetro utilizado para determinar a importância da trilha de feromônio τ_{ij} ;

β : parâmetro que pondera a influência relativa da variável η_{ij} para os vértices i e j no processo de decisão;

Para que não haja uma intensificação demasiadamente elevada do nível de feromônio, a cada iteração da fase de construção as formigas atualizam o feromônio da aresta pela qual acabaram de atravessar, de acordo com a expressão:

$$\tau_{ij} = (1 - \rho) \cdot \tau_{ij} + \rho \cdot \tau_0 \quad (10)$$

O parâmetro ρ é utilizado para que os caminhos menos frequentados sejam esquecidos com o passar do tempo. ρ é um parâmetro definido entre 0 e 1 e τ_0 é o valor inicial do feromônio das arestas.

Hoje em dia muitas implementações bem sucedidas da metaheurística ACO estão disponíveis e têm sido aplicadas à diferentes problemas de otimização combinatória. Para maiores informações, em DORIGO et al. (2008) estas aplicações estão resumidas em um quadro e são discutidas ao longo do livro.

2.5 Apache Mahout

Apache Mahout é uma biblioteca java de aprendizagem de máquina, de código aberto, iniciada em 2008 como um subprojeto da Apache Lucene, outra ferramenta de código aberto gerenciada pela Apache Software Foundation, amplamente utilizada em problemas de busca e recuperação da informação (OWEN et al., 2011), (GIACOMELLI, 2013). Trata-se de uma ferramenta altamente escalável que tem como principal objetivo atuar como uma máquina de recomendação (de filtragem colaborativa), clusterização e classificação.

2.5.1 Técnicas de Recomendação

O módulo de recomendação do Mahout divide-se em quatro variações principais (OWEN et al., 2011):

- **Baseada em usuário:** Consiste na descoberta da semelhança entre os diferentes usuários do sistema. O objetivo é recomendar itens bem avaliados por pessoas semelhantes ao usuário alvo.
- **Baseada em conteúdo:** Objetiva filtrar os itens que receberam melhor avaliação por parte do usuário para posteriormente identificar itens similares a serem recomendados.
- **Slope-One:** Uma técnica de recomendação baseada em item bastante rápida e simples. Estima preferências para novos itens com base na diferença média do valor de preferência entre um novo item e outros itens bem avaliados pelo usuário. Pode ser aplicada quando os usuários atribuem determinadas classificações aos itens (e não apenas preferências de variáveis booleanas).
- **Baseado em modelo:** Tem por objetivo construir um modelo de preferências do usuário com base em preferências existentes e, em seguida, inferir novas preferências.

Ambos os modelos contam com uma métrica de similaridade, ou noção de semelhança entre usuários ou itens. Apache Mahout fornece várias implementações de algoritmos para o cálculo de similaridade, entre as quais pode-se destacar a Correlação de Pearson, Distância Euclidiana, Coeficiente de Tanimoto e Correlação de Spearman. Um detalhamento maior acerca da lógica de cada uma das classes pode ser encontrado em (TIWARY, 2015).

2.5.2 Clusterização

Clusterização refere-se à organização de itens de uma determinada coleção em grupos de itens semelhantes. Estes grupos podem ser caracterizados como um conjunto de itens semelhantes um ao outro em alguns aspectos, porém, diferentes de elementos pertencentes a outros grupos (OWEN et al., 2011).

Apache Mahout suporta diversas implementações de algoritmos para armazenamento em cluster, como K-Means, Fuzzy K-Means, Canopy e Spectral Clustering, cada um com características próprias e critérios específicos, conforme descrito a seguir.

- **k-Means:** Um algoritmo simples, mas bem conhecido para agrupamento de objetos em clusters. Todos os objetos precisam ser representados como um conjunto

de características numéricas. Além disso, o utilizador deve especificar o número de clusters (referidos como K) que pretende identificar. Os itens são armazenados nos k clusters tomando-se por base a distância entre os itens e o centroide, ou centro, da iteração anterior.

- **Fuzzy K-Means:** Uma extensão do algoritmo k-Means. Permite descobrir pontos que podem pertencer a mais de um cluster.
- **Canopy:** Um algoritmo de pré-clusterização não-supervisionado, utilizado para criar pontos de partida para outros algoritmos, como k-means e Fuzzy K-Means.
- **Spectral Clustering:** O algoritmo faz uso do espectro (ou valores próprios) da matriz de semelhança para analisar a conexão entre os dados, ao passo que outros algoritmos de agrupamento, como k-Means, usam a densidade para atribuição de clusters.

2.5.3 Classificação

A classificação pode ser caracterizada como um processo de utilização de informações específicas (entrada) para escolha de uma única seleção (saída), a partir de uma lista de potenciais respostas pré-determinadas (OWEN et al., 2011). Os algoritmos implementados pelo Mahout para classificação/categorização são descritos a seguir.

- **Naive Bayes** - atualmente são implementadas duas abordagens relacionadas à classificação de conteúdo com base em estatísticas bayesianas. A primeira é um classificador padrão Naive Bayes e a segunda é uma extensão desse padrão, denominada de CBayes (Navi Bayes Complementar).
- **Modelo oculto de Markov** - abordado em várias áreas de aprendizagem de máquina para o reconhecimento de padrões, como processamento de linguagem natural, reconhecimento de voz, reconhecimento de escrita à mão, entre outros.
- **Regressão logística** - um modelo utilizado para a previsão da probabilidade de ocorrência de um evento. São utilizadas múltiplas variáveis de previsão que podem ser tanto numéricas ou categorias.

2.5.4 Recomendações com Mahout

Atualmente a construção de um motor de recomendação rápido e flexível pode ser realizado através da biblioteca Taste. Essa biblioteca suporta recomendações baseadas em usuário ou em conteúdo e consiste de cinco componentes principais:

- **DataModel**: contém as informações a serem analisadas pelo Mahout. Essas informações devem estar em um formato específico: UserId, ItemId, Avaliação. Neste trabalho o DataModel foi populado com dados obtidos diretamente da base de dados através da utilização da classe JDBCDataModel.
- **UserSimilarity**: interface que possui várias implementações de algoritmos para o cálculo de similaridade entre usuários.
- **ItemSimilarity**: interface utilizada para o cálculo de similaridade entre itens.
- **UserNeighborhood**: utilizada para determinação da vizinhança, ou seja, o número de usuários similares que serão utilizados no processo de recomendação.
- **Recommender**: interface que deve ser instanciada para obtenção das recomendações.

Esses componentes possibilitam a construção de sistemas de recomendação complexos capazes de realizar recomendações em tempo real ou offline. Mahout possui ainda implementações específicas para avaliação da eficiência do algoritmo escolhido para o processo de recomendação.

Maiores informações acerca de cada um dos algoritmos implementados podem ser encontradas na página do projeto em <https://mahout.apache.org/>. Nessa página são disponibilizados também tutoriais, e-books e exemplos de implementação e aplicação dos algoritmos de recomendação.

3 TRABALHOS RELACIONADOS

A recomendação de Objetos de Aprendizagem voltados para o ensino e aprendizagem de línguas é algo novo. No melhor de nosso conhecimento, não se pode encontrar na literatura trabalhos específicos de recomendação deste tipo de objeto. Através do estudo bibliográfico em trabalhos voltados para a área de Sistemas de Recomendação e Objetos de Aprendizagem, percebe-se que a literatura fornece uma série de trabalhos relacionados. Estes trabalhos permitem analisar a viabilidade de aplicação da solução proposta, através do desenvolvimento de novas soluções ou agregando melhorias às abordagens existentes. Dessa forma, foram analisados dois grupos de trabalhos: Recomendação de Objetos de Aprendizagem e Recomendação de Objetos associados à diferentes contextos, descritos a seguir.

3.1 Recomendação de Objetos de Aprendizagem

Em CAZELLA et al. (2009) é apresentado o desenvolvimento de um sistema para recomendação de Objetos de Aprendizagem baseado em competências a serem desenvolvidas por alunos e suas preferências por determinados OA.

O Sistema de Recomendação aborda a utilização da filtragem colaborativa e constitui-se de três etapas principais: cálculo de similaridade entre os usuários, predição e aplicação de regras de competência. Para o cálculo de similaridade é utilizado o coeficiente de Pearson, anteriormente descrito e representado na Equação (7). Dessa forma, tomando-se por base as avaliações feitas pelos alunos, o sistema busca identificar um conjunto de usuários, denominados vizinhos mais próximos, utilizados na etapa de predição.

A partir da obtenção da correlação de similaridade entre as avaliações sobre determinados OA, o sistema busca prever quanto o aluno apreciaria receber determinada recomendação. Esta predição objetiva descobrir qual nota o aluno daria ao objeto se tivesse acesso ao mesmo. O cálculo é realizado com base em uma média ponderada das avaliações fornecidas pelos alunos, identificados como vizinhos mais próximos (indivíduos que obtiveram um coeficiente de similaridade igual ou superior a um limiar

pré-estabelecido). Por último, são aplicadas as regras de competência, cujo objetivo é filtrar os OA. A avaliação de competências se dá através da utilização do plano de aula do professor da disciplina. São avaliados os conteúdos pré-programados para serem ministrados, considerando as competências que devem ser desenvolvidas pelo aluno em dado momento. Nesta etapa são descartados aqueles OA que têm um bom fator de predição, porém, não desenvolvem as competências apontadas pelos professores como relevantes em determinados momentos da disciplina.

Diferente da abordagem proposta no presente trabalho, o sistema limita-se a recomendar objetos relacionados ao desenvolvimento de competências específicas. Para a tarefa de recomendação é explorada a similaridade entre usuários, enquanto na abordagem proposta aqui as recomendações são baseadas na relação existente entre os OAL. Assim como na abordagem proposta, o SR adota uma abordagem híbrida, porém, utiliza classificações dos usuários obtidas de forma explícita, enquanto que na abordagem proposta as classificações são obtidas de forma implícita.

No modelo proposto por ZAINA et al. (2012), denominado de e-LORS (e-Learning Object Recommendation System), a recomendação de Objetos de Aprendizagem é realizada a partir da análise de três requisitos principais: tema de estudo a ser apresentado ao aluno, perfil de aprendizagem e possíveis restrições tecnológicas que caracterizam o ambiente em que o aluno interage.

Para a construção do perfil de aprendizagem o modelo divide as preferências do aluno em dimensões baseadas na proposta de FELDER; SILVERMAN (1988). As dimensões utilizadas são: percepção, formato/apresentação e participação do aluno. São armazenadas também informações pessoais do aluno, como identificação no sistema, nome, grau de escolaridade, curso de formação e também características tecnológicas relacionadas ao seu acesso.

A determinação de preferências de aprendizagem ocorre através da aplicação de um questionário respondido diretamente pelo aluno, permitindo que o mesmo expresse suas preferências de acordo com cada dimensão.

A Restrição Tecnológica é uma característica opcional utilizada para descrever a tecnologia no modelo do aluno. Tem por objetivo avaliar questões restritivas à utilização de um OA, como por exemplo, o formato de vídeo ou aplicações necessárias para utilização do OA.

Para a armazenagem e o gerenciamento dos OA o e-LORS utiliza o padrão LOM (Learning Object Metadata), um padrão de metadados desenvolvido pelo Institute of Electrical and Electronics Engineers (IEEE) para aplicação em OA. O padrão LOM descreve características importantes relacionadas ao OA ao qual se aplica. Estas características podem ser agrupadas em categorias gerais, educacionais, técnicas e de classificação (SILVA; CAFÉ; CATAPAN, 2011). Embora o padrão apresente um grande número de campos de especificação, o sistema e-LORS utiliza apenas alguns,

conforme apresentado na Tabela 2.

Tabela 2: e-LORS - Categorias descritoras

Categoria LOM	Campos do LOM	Categorização
Geral	Identificador; Tipo; Título; Língua; Descrição e Palavras-chaves.	Faz a descrição geral do OA.
Técnico	Formato: vídeo, som, etc.; Tamanho digital; Localização física e requisitos para uso do objeto (Ex.: versão de software necessário ao formato)	Características técnicas.
Educacional	Tipo de interatividade (ativo, expositivo); Tipo de recurso (exercício, simulação, questionário); Grau de dificuldade.	Descrição das funções educacionais e características pedagógicas do OA.
Relacionamento	Tipo de relacionamento entre os objetos e identificação do relacionamento.	Descreve o relacionamento entre OAs.

Fonte: ZAINA et al. (2012)

No E-LORS, a recomendação se dá através da associação dos objetos descritos pelo padrão LOM com as categorias de perfil do usuário. Para isso o sistema utiliza como parâmetros o tema relacionado ao OA e a identificação do usuário que receberá a recomendação. Assim, iniciado o processo de recomendação, são executadas três fases distintas: filtragem por conceitos, filtragem por perfil de aprendizagem e filtragem por tecnologia.

A filtragem por conceitos busca objetos que possuam o tema relacionado a uma ou mais palavras-chaves utilizadas na consulta. Para tanto, o sistema compara um conjunto de campos (título, descrição e palavras-chaves), associados ao OA, com a palavra-chave pesquisada.

Tabela 3: e-LORS - Campos do LOM e dimensões de preferências

Campo LOM	Valor do campo	Característica do perfil	Dimensão de preferência
Tipo de Interatividade	Ativo	Sensorial	Percepção
	Expositivo	Intuitivo	
Tipo de Recurso de Aprendizagem	Figura, vídeo, filme, etc.	Visual	Formato / Apresentação
	Texto, som, enter outros	Auditivo	
	Exercício prático, experimento	Ativo	Participação
	Questionário e leitura de textos	Reflexivo	

Fonte: ZAINA et al. (2012)

Realizada a filtragem por conceitos, é iniciada à filtragem por perfil de aprendizagem. Nesta etapa é realizada a comparação entre as preferências do aluno (descritas

em seu perfil de aprendizagem) com os campos *Tipo de interatividade* e *Tipo de recurso de aprendizagem* da categoria Educacional do padrão LOM. A Tabela 3 retrata o relacionamento entre as preferências e os campos do LOM utilizados para a filtragem por perfil de aprendizagem.

A última etapa do processo de recomendação é a filtragem de objetos por tecnologia. Esta etapa é opcional, podendo não ser considerada no modelo do aluno. Ela utiliza o subconjunto de OA obtidos a partir da filtragem por perfil de aprendizagem. Assim, as características tecnológicas dos objetos pertencentes ao subconjunto são confrontadas com as restrições cadastradas no modelo do aluno. Caso não existam restrições, todos os OA pertencentes ao subconjunto são recomendados. Caso contrário, o resultado será um novo subconjunto correspondente às questões tecnológicas evidenciadas pelo modelo do aluno.

Diferente da abordagem aqui proposta, o e-LORS não utiliza nenhuma estratégia para classificação e apresentação dos objetos a serem recomendados. Após o processo de recomendação eles são apresentados de forma aleatória, sem que seja determinada a relevância de um determinado OA para o aluno em relação ao subconjunto de objetos recomendado. Na etapa de filtragem por conceitos não é abordada nenhuma técnica para comparação entre os temas e a palavra-chave especificada, como a medida TF-IDF, por exemplo. Além disso, o processo de recomendação prevê uma relação direta entre o perfil de aprendizagem do usuário e os objetos a serem recomendados, porém, não prevê uma relação entre OA, nem entre recomendações já realizadas.

O trabalho de VIEIRA; NUNES (2012), apresenta um Sistema de Recomendação de Objetos de Aprendizagem baseado em conteúdo com dados obtidos do Banco Internacional de Objetos Educacionais (BIOE). O SR, denominado DICA, aborda a utilização de três técnicas para realizar recomendações de acordo com a semelhança entre os OA: agrupamento, medida de similaridade entre temas e técnica do vizinho mais próximo.

A primeira recomendação realizada pelo DICA se dá através do agrupamento e utiliza o algoritmo K-Means. Nesta abordagem os objetos da base de dados são divididos em dois grupos, representados no DICA como os OA mais populares e os OA menos populares, identificados de acordo com o número de visualizações e downloads associados a cada objeto.

Além da recomendação inicial, o sistema permite realizar a busca por um tema específico. Os temas aos quais os OA estão relacionados, foram utilizados para organizar uma estrutura capaz de auxiliar no processo de recomendação. Assim, ao selecionar um determinado tema, são listados todos os OA relacionados à pesquisa. Havendo poucos OA, é realizada uma nova busca na base de dados, utilizando a medida de similaridade entre temas, que ocorre da seguinte forma: os temas foram

armazenados em uma sequência de caracteres separados por ":", o primeiro atributo relacionado ao tema corresponde a um subconjunto mais abrangente dentre todos os OA e o último é um subconjunto mais específico contido nos anteriores (Ex.: Educação Básica::Ensino Médio::Matemática::Análise de dados e probabilidade). Assim, um objeto que pertença a esse tema, fará parte desses quatro subconjuntos. Para cada tema consultado, o algoritmo compara o título do subconjunto com outro tema, caso não encontre, a distância é acrescida de 1. Esse procedimento é representado na figura 6.

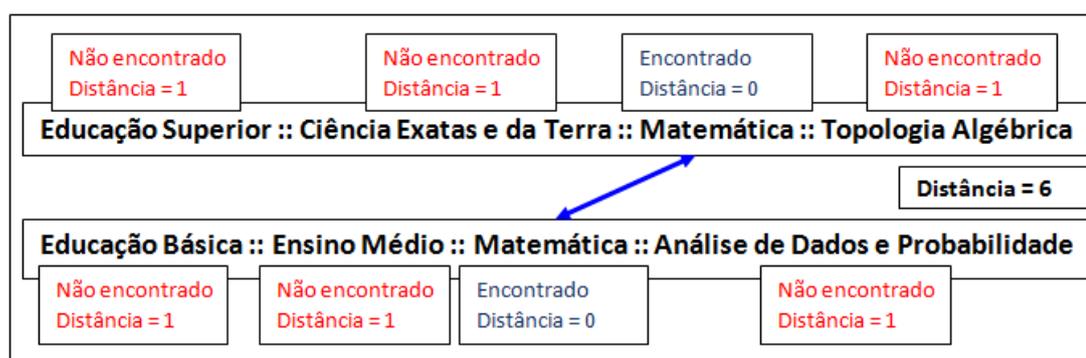


Figura 6: DICA - Comparação entre temas

Fonte: VIEIRA; NUNES (2012)

A recomendação final se dá ao clicar em um objeto da lista resultante da aplicação das duas técnicas anteriormente citadas. Através do cálculo da medida TF-IDF, são comparados todos os termos do objetivo mais a descrição do OA que está sendo consultado com os 20 mais próximos a ele, selecionados, novamente, através do algoritmo que calcula a distância entre os temas. Através desse cálculo são selecionados aqueles que possuem maior número de termos em comum, denominados de vizinhos mais próximos.

Assim como os demais trabalhos anteriormente descritos, no DICA não é explorada nenhuma relação entre os objetos. O SR faz recomendações considerando a medida TF-IDF associada a cada objeto. Essa recomendação é complementada através da medida de similaridade entre temas. Diferente da abordagem proposta aqui, trata-se de um sistema puramente baseado em conteúdo, sem que haja nenhuma interferência do usuário capaz de influenciar nas recomendações realizadas.

Além dos trabalhos ligados ao contexto de recomendação de objetos de aprendizagem, foram analisados também trabalhos relacionados à recomendação de objetos em diferentes contextos. Apesar de não recomendarem OA, as técnicas e abordagens utilizadas nestes trabalhos podem contribuir para o desenvolvimento da solução proposta e são apresentadas a seguir.

3.2 Recomendação de Objetos Associados à Diferentes Contextos

Em MONTES-GARCÍA et al. (2013), é apresentado um sistema de recomendação sensível ao contexto, voltado para a área de Jornalismo. O sistema denominado Wesomender, é capaz de recomendar notícias a partir da identificação de tópicos semelhantes em diferentes fontes de notícias. Para tanto, utiliza a avaliação de especialistas e considera características contextuais relacionadas às notícias, como: tempo, interesses dos utilizadores, local ou tendências existentes. Esses dados são combinados com as técnicas tradicionais de recomendação para obtenção de uma estrutura adaptativa, capaz de lidar com dados heterogêneos. Essa estrutura, segundo os autores, pode ser caracterizada como um sistema de filtragem colaborativa reforçada.

Wesomender é composto por dois módulos principais: um módulo de filtragem baseada em conteúdo e um módulo de filtragem colaborativa. Cada módulo produz recomendações de forma independente, tomando por base a análise de notícias que o usuário não tenha visto ou avaliado. Como resultado, para cada notícia analisada, o sistema retorna duas variáveis com valores entre 0 e 1, denominadas *CF* (collaborative-filtering) e *CB* (content-based), onde valores mais próximos a 1 significam que a notícia poderá interessar ao usuário. A análise de notícias é realizada da seguinte forma: segundo os autores, existe uma regra não escrita no mundo do jornalismo, chamada de *The Five W* (os cinco Ws), que determina a importância de uma notícia através de cinco perguntas básicas que devem ser respondidas: *who?* (quem?), *what?* (o quê?), *why?* (porquê?), *when?* (quando?) e *where?* (onde?). Considerando essas questões, o valor da variável *CB*, resultante da recomendação baseada em conteúdo, é obtido tomando-se por base os seguintes parâmetros:

1. Uma notícia é mais interessante se o que está sendo informado acaba de acontecer;
2. Uma notícia sobre algo que acontece próximo de um usuário, é mais relevante para ele;
3. Se um usuário lê uma notícia, envolvendo determinadas entidades ou cobrindo tópicos específicos, ele está interessado nessas entidades ou tópicos neste momento. Esta hipótese considera notícias antigas também, visto que o usuário pode estar à procura de referências.

Além dos parâmetros acima descritos, o sistema considera a confiabilidade da fonte de dados associada à notícia. Esse parâmetro prevê que, ao analisar um artigo, por exemplo, uma referência a outro autor pode ser considerada mais confiável do que a referência a um blog. Assim, através da utilização dos parâmetros: tempo,

localização, conteúdo e confiabilidade, o valor de CB é obtido através da seguinte fórmula:

$$CB = \alpha \cdot ER + \beta \cdot T + \gamma \cdot P + \delta \cdot R \quad (11)$$

Onde:

ER - refere-se às entidades relacionadas com a notícia. Isto é, as últimas notícias que o usuário tenha demonstrado interesse e que devem ser verificadas. Para cada entidade associada à recomendação em curso, o percentual correspondente é adicionado à variável ER ;

R - refere-se à recência. É calculada com base no tempo atual e o momento da publicação, retornando valores entre 0 e 1;

P - refere-se à proximidade. Calculada com base na distância máxima entre dois pontos, nesse caso o usuário e a notícia, retornando valores entre 0 e 1;

T - refere-se à confiabilidade. O valor de T é obtido a partir da avaliação realizada por uma equipe de jornalistas para cada mídia.

A importância de cada variável é um percentual determinado por uma equipe de jornalista profissionais, onde:

- R é o fator mais importante, sendo $\alpha = 0.5$;
- P é o segundo em importância, com $\gamma = 0.2$;
- T e ER são igualmente relevantes, sendo $\alpha = \beta = 0.15$;

Com o objetivo de projetar um SR adaptativo e híbrido, Wesomender conta com um componente que avalia o conjunto de dados existente e escolhe o melhor algoritmo de filtragem colaborativa. Este componente é aplicado sobre os algoritmos do Apache Mahout, escolhido por implementar um conjunto considerável de técnicas que podem ser facilmente avaliadas por meio de uma API para testar a precisão dos algoritmos.

Os algoritmos utilizados nos testes realizados pelo Wisomender foram:

- Generic User Based Recommender - Produz recomendações considerando itens relacionados à usuários com gostos semelhantes. Existem várias implementações em Mahout para avaliar a correlação de similaridade. Em Wesomender foram implementadas as seguintes medidas: Coeficiente de Correlação de Pearson, Distância Euclidiana, Coeficiente de Similaridade do Cosseno, Distância de Manhattan, Log-Likelihood, Coeficiente de Correlação de Spearman e Coeficiente de Tanimoto;
- Item Based Recommender - Semelhante ao Generic User Based Recommender, essa técnica também utiliza o conceito similaridade, porém aplicados aos itens, ao invés de usuários;

- Tree Clustering Recommender - Constrói recomendações tomando por base grupos de usuários com gostos semelhantes;
- SVD Recommender - Trata-se de uma das formas de fatoração de matrizes usadas em filtragem colaborativa;
- Slope One Recommender - Produz recomendações baseado na previsão de classificação que um determinado usuário daria a um item.

Assim que o melhor algoritmo é selecionado, o valor de CF pode ser calculado e aplicado na fórmula 12, para realizar a recomendação final:

$$r = w.CB + (1 - w).CF \quad (12)$$

Onde w é a ponderação utilizada para avaliações baseadas em conteúdo, sendo $0 \leq w \leq 1$.

Wesomender, difere-se da abordagem apresentada por este trabalho, não só pelo contexto de aplicação, mas pelo fato de utilizar classificações adquiridas de forma explícita e combinar recomendações, obtidas a partir de SR individuais, para realizar a recomendação final. Apesar de ser um sistema híbrido, pode-se perceber que é atribuída maior importância às características contextuais relacionadas aos itens disponíveis para recomendação. Dessa forma, dá-se maior ênfase à recomendação baseada em conteúdo.

Em ZHOU et al. (2008) é abordado o uso de propagação de rótulos para recomendação de documentos em uma biblioteca digital. A propagação de rótulos é um tipo de aprendizado transdutivo ¹ da categoria de aprendizado semi-supervisionado. Sua utilização objetiva estimar os rótulos de dados não-rotulados utilizando outros dados rotulados ou parcialmente rotulados e suas similaridades.

Na proposta apresentada pelos autores o cenário é representado por uma rede de documentos interligados através de citações e associados a outras entidades por várias relações, como autores, local de publicação e referencial bibliográfico.

O trabalho tem como foco principal a aplicação de filtragem colaborativa para itens ligados em rede, combinando vários grafos para calcular a similaridade entre itens. Os itens podem ser representados pelo conjunto completo de vértices de um grafo ou por um subconjunto de vértices desse grafo.

¹O aprendizado transdutivo objetiva rotular dados não-rotulados conhecidos sem expandir o modelo para um novo conjunto de dados desconhecido. Dessa forma, todos os exemplos dos quais o rótulo pode ser predito são conhecidos (ZHU; GOLDBERG, 2009)

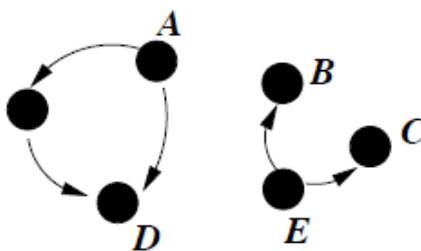


Figura 7: Exemplo de um grafo de citações
Fonte: ZHOU et al. (2008)

Conforme ilustrado na Figura 7, os documentos podem ser representados como vértices de um grafo direcionado onde as arestas indicam as respectivas citações. A similaridade entre os documentos pode ser medida em termos de co-citações² (co-citando os mesmos documentos ou sendo co-citado por outros). No exemplo apresentado na figura 7, os documentos *B* e *C* podem ser considerados semelhantes, pois são co-citados por *E*.

Além da propagação de rótulos são utilizados métodos de aprendizagem para combinar os vários grafos e medir a similaridade entre itens. Nesse contexto, a aprendizagem de três tipos de grafos são formuladas como três sub-problemas, cada um usando uma estratégia de fatoração de acordo com as características de cada tipo de grafo. Os três sub-problemas abordados consideram as citações, local de publicação e autores dos documentos. A partir do resultado obtido, um novo quadro de recomendação é desenvolvido através da aplicação do aprendizado semi-supervisionado em grafos.

Apesar de explorar a relação existente entre os itens a serem recomendados, o trabalho de ZHOU et al. (2008) difere-se da proposta aqui apresentada por não basear-se na interação do usuário com o ambiente para realizar recomendações. Percebe-se ainda que, apesar de enfatizar o uso da filtragem colaborativa (por meio das citações e co-citações realizadas), o trabalho possui características mais próximas a um sistema baseado em conteúdo. Pois, além de não utilizar avaliações realizadas pelos usuários, infere relações entre itens utilizando informações relacionadas aos autores e local de publicação dos documentos a serem recomendados.

Em PRABHA; RATHIPRIYA (2013) é apresentado um sistema de recomendação com base em um modelo de navegação. O trabalho aborda a construção de perfis de usuários obtidos a partir da análise de dados de navegação. O perfil é obtido por uma função: $perfil(A, J)$, onde *A* representa o usuário e *J* o item acessado pelo usuário *A*. A recomendação se dá através da identificação de um conjunto de perfis

²A análise de co-citação é uma das principais técnicas quantitativas em estudos científicos para mapear a estrutura e dinâmica da pesquisa científica. Esta técnica pode ser utilizada para identificar focos de pesquisa e suas relações, em particular o nível de especialização da pesquisa (BRAAM; MOED; VAN RAAN, 1991)

similares para um dado usuário, utilizando para tanto um método baseado em inteligência de enxames. Trata-se de uma função de avaliação para identificação de um conjunto de usuários, denominados de vizinhos mais próximos. Através dessa função a semelhança entre os perfis de usuários é calculada e o perfil mais próximo é escolhido. O sistema então seleciona os usuários cuja distância euclidiana está acima de um determinado valor limite. O conjunto de dados obtidos a partir dessa seleção é então aplicado a um método baseado em Inteligência de Enxames. O estudo aponta que o modelo proposto pode ser aplicado através da utilização de diferentes técnicas, tais como: Otimização por Colônia de Formigas (*Ant Colony Optimization*), *Bat Algorithm*, *Cuckoo Search*, *Firefly Algorithm* e Otimização por Enxame de Partículas (*Particle swarm optimization*).

Diferente da proposta deste trabalho o sistema não prevê uma relação entre os itens a serem recomendados. É utilizada uma abordagem puramente colaborativa, onde a recomendação se dá com base na análise do perfil de acesso dos usuários, obtido de forma implícita.

Em WANG et al. (2012) é apresentada uma proposta de algoritmo de recomendação com o objetivo de ser escalável e capaz de lidar com dados de treinamento escassos e que mudam ao longo do tempo. A abordagem apresentada prevê a utilização de um novo tipo de esquema de filtragem colaborativa baseada no comportamento das formigas e se dá através da aplicação do Algoritmo de Otimização por Colônia de Formigas. Através de um mecanismo de transmissão de feromônio o modelo proposto é capaz de identificar a similaridade entre usuários e objetos, mesmo frente ao problema da dispersão. O processo de transmissão de feromônio ocorre da seguinte forma: dado um índice de feromônio representando um usuário ou um grupo de usuários, o item compartilha feromônio do usuário ao receber uma classificação. Da mesma forma, o item transfere o feromônio nele existente para o usuário que realizou a classificação. Depois de algum tempo, itens semelhantes recebem uma quantidade semelhante de feromônio e, em seguida, os usuários com gostos semelhantes se tornam iguais no que diz respeito à quantidade de feromônio à eles associada. O sistema utiliza a evaporação do feromônio existente para identificar a evolução de preferências dos usuários ao longo do tempo. Através do índice de feromônio a abordagem utilizada permite calcular 3 formas de similaridade: similaridade entre itens, similaridade entre usuários e similaridade entre usuários e itens. Com o objetivo de melhorar o desempenho e a qualidade das recomendações realizadas, a solução utiliza ainda uma técnica para o agrupamento de usuários de acordo com o índice de feromônio associado a cada um deles.

Diferente da proposta do presente trabalho, na abordagem de WANG et al. (2012) o índice de feromônio utilizado na tarefa de recomendação pode estar associado aos usuários, aos itens ou à ambos, podendo ser transmitido de forma simultânea a cada

interação do usuário. O sistema prevê a atualização do índice de feromônio associado aos itens e aos usuários, como forma de comparar o valor associado para realizar recomendações. Porém, não considera a criação de conexões entre os itens a serem recomendados.

Em SOBECKI; TOMCZAK (2010) é abordado o desenvolvimento de um sistema para recomendação de cursos para estudantes universitários. A tarefa de recomendação se dá através de filtragem de informações integrada a uma abordagem colaborativa, baseada em ACO. O sistema utiliza três tipos básicos de filtragem: Filtragem demográfica, filtragem baseada em conteúdo e filtragem colaborativa. Os filtros são aplicados a uma série de informações associadas aos alunos, como número de identificação do aluno, curso, tipo de curso, semestre de referência, entre outras. A partir desses dados, são calculados valores médios para cada tipo de curso e para o conjunto de informações associados a cada estudante. A etapa seguinte consiste na escolha de uma das medidas obtidas para aplicação do Algoritmo de Otimização de Colônia de Formigas. A ideia principal consiste em recomendar para cada estudante uma previsão de qualidade para todos os cursos, para os quais ele tem a oportunidade de se inscrever. Os experimentos utilizaram três modelos de algoritmos baseados em ACO, constituindo três métodos de aplicação e seleção: i) Escolha da probabilidade máxima; ii) Com uma formiga e caminhada aleatória, considerando a probabilidade das arestas; iii) Com k formigas e caminhada aleatória, utilizando a probabilidade das arestas. Os resultados obtidos a partir da aplicação desses três algoritmos são então comparados com cinco métodos de filtragem. Para o processo de avaliação de desempenho são utilizadas várias medidas de comparação como: *Mean Absolute Error* (MAE), *Normalized Mean Absolute Error* (NMAE), *Prediction Accuracy* (PA), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Standard Error Variance* (SEV) e *Classification Accuracy* (CA).

A proposta apresentada por SOBECKI; TOMCZAK (2010) difere-se da abordagem utilizada neste trabalho, não só pelo contexto de aplicação, mas também por utilizar dados referentes aos cursos e aos usuários obtidos de forma explícita. O sistema faz recomendações com base nos resultados obtidos a partir da aplicação de filtros sobre o conjunto de dados disponíveis para avaliação.

No trabalho apresentado por BEDI; SHARMA (2012) são produzidas recomendações a partir da incorporação de uma medida de confiança entre os usuários. Com base nos valores obtidos são criados grupos baseados na metáfora biológica de colônias de formigas. Esses grupos, por sua vez, representam as melhores opções de recomendação para o usuário ativo. O sistema funciona em duas etapas: a primeira etapa consiste na obtenção do gráfico de confiança dirigido para cada usuário. Sua criação se dá através da matriz de classificação usuário-item e informações de confiança sobre os usuários. As informações de confiança são obtidas

a partir do grau de similaridade entre os perfis de parceiros. O valor obtido é utilizado para determinar o índice de feromônio, que representa a força de conexão ou seja, a intensidade de confiança entre os dois parceiros de recomendação (recomendador e usuário ativo) no tempo t . A segunda etapa é o processo de recomendação. Nesta etapa é escolhido o melhor grupo de usuários e aplicada a solução proposta para o problema do novo usuário. Nesse contexto, a solução apresentada sugere a criação de uma lista de recomendadores mais "populares", com base no alto valor de feromônio associado (nível de confiança global).

O trabalho de BELLAACHIA; ALATHEL (2016) propõe a aplicação do Algoritmo de Otimização por Colônia de Formigas, em um sistema de recomendação baseado em redes de confiança, objetivando resolver o problema de previsão de avaliações para itens que possuem poucas classificações por parte dos usuários. Como usuários novos possuem poucas ou nenhuma classificação relacionada aos itens, o algoritmo utiliza a confiança estabelecida entre os usuários para guiá-los em sua exploração no espaço de soluções. Nesse sistema, o usuário ativo representa o ninho e os usuários com uma classificação para o item alvo são considerados fontes de alimento. Representando o problema através de um grafo, a solução se dá através da atribuição de um nível de feromônio inicial para cada aresta que representa uma possível solução. Esse nível de feromônio reflete a importância associada à aresta (confiança). As boas soluções, neste caso, são os nós que representam os usuários com uma classificação para o item alvo. A cada iteração uma formiga k localizada no ponto x calcula a probabilidade de cruzar as bordas de ligação a um usuário y de acordo com o nível de confiança associado. Em seguida, eventualmente, cruza a borda que produziu a maior probabilidade. Nesta etapa ocorre a atualização local do índice de feromônio (atualização local). O sistema utiliza um modelo de atualização de feromônio de dois níveis: atualização local e atualização global. A atualização global é realizada após todas as formigas terminarem a construção de suas soluções.

Diferente da abordagem utilizada por este trabalho, em BEDI; SHARMA (2012) e BELLAACHIA; ALATHEL (2016), as recomendações são realizadas com base na ligação existente entre os usuários (confiança) e não entre os itens a serem recomendados. Além disso, abordagens baseadas em confiança sugerem avaliações explícitas por parte dos usuários como forma de identificar a confiança entre eles. Tal característica tem por objetivo a criação de uma lista de "amigos" (de confiança) de cada usuário, juntamente com os seus níveis de confiança associados, para posterior utilização no processo de recomendação.

4 ACO APLICADO À RECOMENDAÇÃO DE OBJETOS DE APRENDIZAGEM DE LÍNGUAS

Este capítulo descreve o Sistema de Recomendação de Objetos de Aprendizagem de Línguas aqui proposto. A abordagem utilizada para desenvolvimento do sistema é apresentada na primeira seção do capítulo. São definidos os dados de entrada a serem utilizados, as técnicas e o tipo de filtragem adotados, os algoritmos implementados e a forma como os dados de entrada e os dados armazenados são combinados pelos algoritmos no processo de recomendação, resultando na saída do sistema.

4.1 Abordagem utilizada

O Sistema de Recomendação implementado neste trabalho tem por objetivo recomendar Objetos de Aprendizagem de Línguas, classificados segundo sua relevância para o usuário. Dessa forma, a saída desejada consiste em uma lista de "*n*" OAL, ordenados de acordo a predição realizada pelo sistema, ficando a critério do usuário selecionar ou não aquele que melhor atende ao objetivo da sua consulta. O modelo proposto incorpora tanto características baseadas em conteúdo e colaborativas, combinando ambas as técnicas em único recomendador baseado em regras, constituindo dessa forma um SR híbrido. A recomendação propriamente dita ocorre através da aplicação do algoritmo de Otimização por Colônia de Formigas. Nesta etapa são consideradas as ações do usuário, o valor e o número de *downloads* associado a cada objeto, bem como a conexão existente entre os possíveis objetos a serem recomendados.

Neste trabalho foi adotada a medida TF-IDF para especificação de pesos de palavras-chave, uma técnica bastante utilizada em SR baseados em conteúdo para indexação de frequência de termos. Através dessa abordagem é possível fazer uma recomendação inicial com base na consulta realizada pelo usuário. Além disso, tendo em vista que o sistema não prevê a atribuição de avaliações explícitas para os objetos recomendados, o valor da medida TF-IDF associado a cada objeto é utilizado para o cálculo do índice de feromônio. Esse índice pode aumentar ou diminuir de acordo com

as ações realizadas pelo usuário.

A descrição de interesses do usuário ocorre por meio de ações como busca, aquisição, visualização e seleção de OAL. Essas ações são utilizadas no contexto de aplicação da filtragem colaborativa. A conexão entre os objetos é criada de forma implícita, através da interação do usuário com o ambiente e os metadados associados aos OAL. Dessa forma, um objeto pode ser vinculado à diferentes contextos, de acordo com as necessidades e o conhecimento de cada usuário, e não através de um voto atribuído de forma explícita.

4.2 Aplicação do algoritmo ACO

Na abordagem proposta por este trabalho, possíveis soluções emergem a partir da interação dos usuários. Nesse contexto, uma solução pode ser caracterizada como o estabelecimento de múltiplas conexões conceituais entre os objetos, formando um conjunto de OAL a serem recomendados, ordenados em uma lista de acordo com a predição realizada pelo sistema.

A Figura 8 apresenta o exemplo da construção de uma solução representada através de um grafo. Nesse exemplo cada vértice do grafo representa um OAL e cada aresta representa uma conexão entre dois OAL.

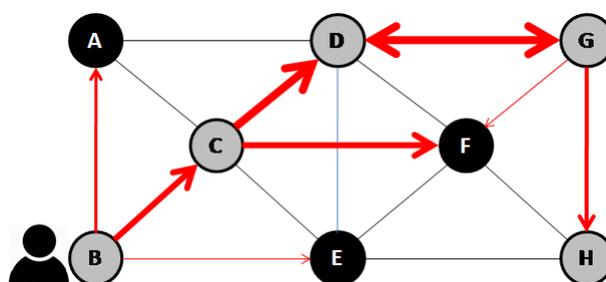


Figura 8: ACO - Exemplo de construção de uma solução por um usuário.

Fonte: Elaborada pelo autor

Cada aresta (i, j) do grafo possui uma variável τ_{ij} que representa uma trilha de feromônio artificial. Essa variável pode ser modificada pelo usuário, sendo incrementada de acordo com o percurso realizado por ele.

Através de uma fórmula probabilística, possíveis soluções são apresentadas para o usuário a cada nova interação. Em cada passo da construção da solução, o usuário seleciona o próximo vértice a ser visitado, através de uma lista, ordenada de acordo com um mecanismo estocástico que é influenciado pelo valor da variável τ_{ij} .

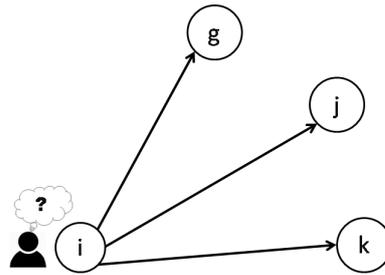


Figura 9: ACO - Seleção do próximo caminho pelo usuário.

Fonte: Elaborada pelo autor

De forma análoga às soluções construídas pelas formigas, o usuário percorre o grafo construindo a solução, passando pelos vértices e pelas arestas, modificando o valor da variável τ_{ij} associada a cada aresta. Ao final de cada iteração o feromônio é atualizado, podendo ser incrementado ou decrementado, conforme a ação realizada pelo usuário. Quando o nível de feromônio é decrementado a conexão entre os objetos é enfraquecida, podendo até mesmo ser desfeita ao longo do tempo. Dessa forma, quanto maior o nível de feromônio, maior a chance da aresta ser visitada por outros usuários. Essa característica é evidenciada pela figura 10, onde pode-se perceber que a conexão entre os vértices $B \rightarrow C$ é mais forte que $B \rightarrow A$ e $B \rightarrow E$, sendo denotada pela espessura da aresta.

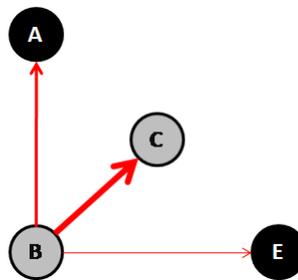


Figura 10: ACO - Representação do relacionamento entre OALs.

Fonte: Elaborada pelo autor

A figura 11 demonstra as etapas percorridas pelo usuário para construção de uma solução. Tomando-se como exemplo o grafo representado pela figura 8, ao final da interação do usuário tem-se a solução dada pelo caminho B-C-D-G-H (figura 11e).

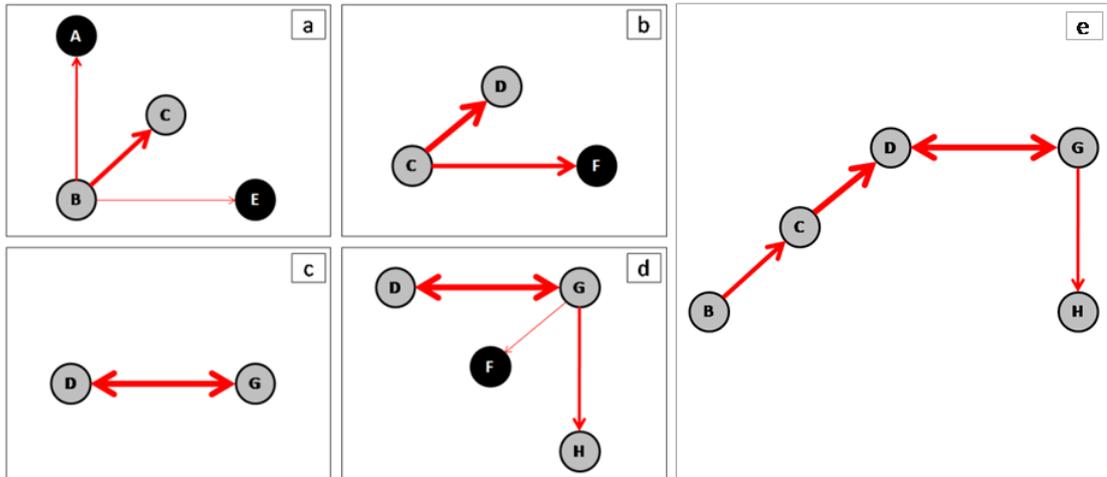


Figura 11: ACO - Construção de uma solução.

Fonte: Elaborada pelo autor

O exemplo apresentado considera que o usuário selecionou o primeiro objeto recomendado pelo SR em todas as interações. Cabe ressaltar que outros objetos do grafo poderiam ser selecionados, dando origem à novas soluções, através da criação de novas conexões ou reforçando conexões já existentes. Os passos para construção da solução representada pela figura 11e são descritos a seguir.

1. O usuário faz uma busca por palavra-chave;
2. O SR retorna um conjunto de objetos, ordenados em uma lista, de acordo com a importância do objeto para a consulta realizada, tomando-se por base a medida TF-IDF associada a cada objeto;
3. Ao clicar no objeto *B* o SR retorna uma nova lista, ordenada de acordo com o valor da variável p_{ij}^k ; (objetos *C*, *A* e *E*, respectivamente (figura 11a));
4. Ao clicar no objeto *C* são apresentados os objetos *D* e *F*, respectivamente (figura 11b);
5. Ao clicar no objeto *D* é apresentado apenas o objeto *G*, pois não existem outros objetos relacionados. Pode-se perceber através da figura 11c que existe uma forte conexão entre os dois objetos, podendo um referenciar o outro no processo de recomendação.
6. Por fim, ao clicar no objeto *G*, são apresentados os objetos *H* e *F*, respectivamente, resultando na solução representada pela figura 11e.

Através desse exemplo, pode-se perceber que o usuário constrói possíveis soluções a cada interação com o sistema. A conexão entre os objetos é reforçada de acordo com suas preferências.

A equação utilizada na solução proposta é a mesma representada pela equação 13, sendo reescrita logo a seguir para melhor compreensão em relação aos valores associados às variáveis.

$$p_{ij}^k = \frac{[\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{l \in J_i^k} [\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}$$

Onde:

p_{ij}^k : é o valor utilizado para ordenar os objetos que serão recomendados, segundo a predição calculada pelo sistema. Dessa forma, se j ainda não foi visitado, ele pode ser selecionado, com uma probabilidade que é proporcional ao índice de feromônio associado à aresta (i, j) .

τ_{ij} : variável que controla o feromônio depositado nas arestas. Na solução proposta o valor utilizado para representar o índice de feromônio associado à variável τ_{ij} é a medida TF-IDF. A aplicação da atualização de feromônio é executada a cada interação do usuário com um OAL. Se o usuário visualizou o objeto e não efetuou o download o feromônio é decrementado de acordo com a equação 10. Caso contrário, o feromônio é incrementado de acordo com a medida TF-IDF correspondente ao OAL selecionado.

η_{ij} : é um valor heurístico associado ao contexto do problema em questão, sendo representado neste trabalho pelo número de visitas ou interações, realizadas por usuários distintos, à uma conexão existente entre determinados objetos.

Três parâmetros α , β e ρ são utilizados para controlar a intensidade do feromônio (τ_{ij}), e a qualidade da aresta (η_{ij}):

- α - Utilizado para determinar a importância do feromônio. Um valor muito alto para α pode causar um efeito de intensificação do feromônio. Este efeito pode levar os usuários a escolherem arestas que outros usuários escolheram no passado, causando a estagnação precoce do algoritmo. Outrossim, valores muito baixos também podem levar a soluções ruins.
- β - Trata-se de um parâmetro importante para uma boa diversidade de soluções. Um valor baixo demais para β pode levar à estagnação precoce do algoritmo, assim como um valor alto demais pode aproximá-lo de uma construção gulosa.
- ρ - Utilizado para evitar uma grande acumulação de feromônio, o que pode prejudicar a diversidade de soluções, este parâmetro deve ser ajustado para um valor menor que 1. ρ é utilizado na equação 10, responsável pela evaporação do feromônio. Neste trabalho a evaporação é aplicada sempre que o usuário visualiza um objeto recomendado e não efetua o download. Dessa forma, conexões que não são consideradas relevantes por muito usuários, vão perdendo a intensidade ao longo do tempo, podendo até mesmo serem desfeitas.

Nos experimentos realizados para o TSP, verificou-se que diferentes combinações de parâmetros (ou seja, $(\alpha = 1, \beta = 1)$, $(\alpha = 1, \beta = 2)$, $(\alpha = 1, \beta = 5)$, $(\alpha = 0.5, \beta = 5)$) resultaram no mesmo nível de desempenho (DORIGO; MANIEZZO; COLORNI, 1996). Assim, os valores utilizados na abordagem proposta foram aqueles sugeridos pelo autor para o algoritmo Ant cycle, que obteve melhor desempenho em comparação aos outros dois: Ant density e Ant quantity. Sendo $\alpha = 1$, $\beta = 2$ e $\rho = 0.5$.

A seguir é apresentado um exemplo do cálculo realizado para obtenção do valor utilizado para ordenar os objetos recomendados (variável p_{ij}^k).

Tabela 4: Ex. de recomendações e valores utilizados no cálculo da medida swarm

Objeto selecionado: 155 - Typical breakfast in Great Britain				
Exemplo de Recomendações				
Código	Título	Visitas	Valor (conexão)	p_{ij}^k
356	What I eat for breakfast lunch...	4	40.59700848181423	0,9998572
439	Stereotypes about...	3	0.010284217713809999	0,0001424747

Objeto: 356 - What I eat for breakfast lunch and dinner

$\tau_{ij} = 40,59700848181423$ (Índice de feromônio associado à aresta);

$\eta_{ij} = 4$ (Número de visitas únicas à conexão existente entre os objetos 155 e 356);

$$[\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta = [40,59700848181423]^1 \cdot [4]^2 = 649,5522$$

Objeto: 439 - Stereotypes about English people

$\tau_{ij} = 0,010284217713809999$ (Índice de feromônio associado à aresta);

$\eta_{ij} = 3$ (Número de visitas únicas à conexão existente entre os objetos 155 e 356);

$$[\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta = [0,010284217713809999]^1 \cdot [3]^2 = 0,09255796$$

$$\sum_{l \in J_i^k} [\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta = 649,5522 + 0,09255796 = 649,6448$$

$$p_{ij}^k = \frac{[\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{l \in J_i^k} [\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}$$

$$p_{ij}^k (\text{Objeto 356}) = 649,552 / 649,6448 = 0,9998572$$

$$p_{ij}^k (\text{Objeto 439}) = 0,09255796 / 649,6448 = 0,0001424747$$

Exemplo de cálculo para atualização do índice de feromônio (τ_{ij}). Quando o usuário visualiza o objeto e não efetua o download a taxa de evaporação é decrementada do índice de feromônio associado à conexão.

O parâmetro ρ é utilizado para que os caminhos menos frequentados sejam esquecidos com o passar do tempo. ρ é um parâmetro definido entre 0 e 1 e τ_0 é o valor inicial do feromônio das arestas.

0,01635727 (Valor associado ao objeto 439). Trata-se de um valor obtido a partir do cálculo da medida TF-IDF. Esse valor é armazenado na base de dados e incrementado a cada visualização ou download de um objeto.

$$\text{Taxa de evaporação} = (1 - \rho) \cdot \tau_{ij} + \rho \cdot \tau_0;$$

$$\text{Taxa de evaporação} = (1 - 0,5) * 0,01635727 = 0,008178635$$

$$\tau_{ij} = 0,010284217713809999 - 0,008178635 = 0,002105583;$$

5 ESTUDO DE CASO

Este capítulo trata da aplicação do sistema de recomendação de Objetos de Aprendizagem de Línguas. Apresenta as características do protótipo desenvolvido, o local e a forma de coleta dos dados utilizados para realização dos experimentos, bem como a estrutura criada para armazenamento das informações. São apresentadas as telas do sistema, a forma de utilização, a metodologia utilizada no processo de recomendação e exemplos de códigos utilizados na implementação.

5.1 BALL - Banco de Atividades para Aprendizagem de Línguas

A proposta deste trabalho tem sua aplicação nas ações do grupo Elaboração de Materiais e Práticas Pedagógicas na Aprendizagem de Línguas do Programa de Pós-Graduação em Letras/Mestrado da Universidade Federal de Pelotas.

As ações do grupo, as quais abrangem os objetivos do projeto de pesquisa Línguas estrangeiras e TICs: Aprendizagem de Línguas e Elaboração de Materiais na Complexidade e no Caos, demandam a produção de Objetos de Aprendizagem de Línguas (OAL) de acordo com princípios teórico-metodológico-pedagógicos propostos pelos pesquisadores do grupo. Dessa forma, uma das metas a serem alcançadas é a constituição de um repositório de OAL, desenvolvido com base em princípios pedagógicos comunicativos e levando em consideração o caráter complexo da aprendizagem de línguas. Nesse contexto, constitui um dos objetivos do projeto materializar, na forma dos OAL e do repositório BALL (Banco de Atividades para Aprendizagem de Línguas), as conclusões acerca da especificidade da aprendizagem de línguas em meio digital, às quais os pesquisadores têm chegado.

Durante o desenvolvimento desta dissertação foi concluída parte da implementação do repositório BALL, disponibilizando recursos suficientes para a inserção e catalogação de objetos de aprendizagem. Entretanto, a tarefa de popular o repositório com uma quantidade de OALs adequada para realização dos experimentos, demandaria muito tempo. Como solução para o problema, foi desenvolvido um protótipo de sistema web populado com dados obtidos do Banco Internacional de

Objetos Educacionais (BIOE), conforme descrito a seguir.

5.2 Protótipo

O protótipo aqui especificado utilizou apenas dados textuais relacionados aos objetos de aprendizagem e foi disponibilizado para um grupo restrito de pessoas que participaram da realização dos experimentos. Os downloads são fictícios, contabilizados a partir dos cliques realizados pelos usuários, não utilizando nenhum objeto do BIOE em seu formato digital.

O BIOE (<http://objetoseducacionais2.mec.gov.br/>) é um repositório de objetos educacionais (OEs) criado em 2008 pelo Ministério da Educação (MEC) em parceria com o Ministério da Ciência e Tecnologia, Rede Latino-americana de Portais Educacionais (RELPE), Organização dos Estados Ibero-americanos (OIE) e algumas universidades brasileiras (MACHADO; FARIAS, 2012). Caracterizado como uma ferramenta digital de cunho pedagógico, o BIOE disponibiliza OEs de diferentes países e línguas, nas mais variadas áreas de conhecimento e em diferentes formatos e níveis de ensino.

Os objetos no BIOE seguem o padrão de organização Dublin Core¹, constituindo uma boa fonte de informações devido à diversidade de metadados disponíveis. Os metadados são catalogados de forma organizada, buscando prover informações relacionadas à recuperação e identificação dos objetos, assim como outras relativas a direitos autorais, datas, restrições de acesso, entre outros.

Através da utilização de um web crawler², foram adquiridas informações para popular uma base de dados com mais de quinze mil (15.000) objetos de aprendizagem. Desse grupo de objetos foram extraídos apenas aqueles relacionados ao tema língua estrangeira, totalizando seiscentos e sessenta e cinco (665) objetos de aprendizagem.

Para a constituição da base de dados foram selecionadas apenas as informações consideradas necessárias para a realização dos experimentos. A estrutura da tabela de objetos (representada pela figura 12) ficou composta por oito campos: id, título, tipo de recurso, objetivo, descrição do recurso, observações, componente curricular e tema. Esses descritores foram escolhidos por disponibilizarem informações sobre o conteúdo associado aos objetos, suficientes para sua recuperação e identificação.

¹O Dublin Core é um padrão para catalogação/descrição de recursos digitais que prevê um conjunto mínimo de elementos para auxiliar na definição da estrutura, semântica e sintaxe dos metadados.

²Um crawler é um rastreador web, capaz de coletar informações de forma automatizada com base nos metadados associados ao conteúdo da página.

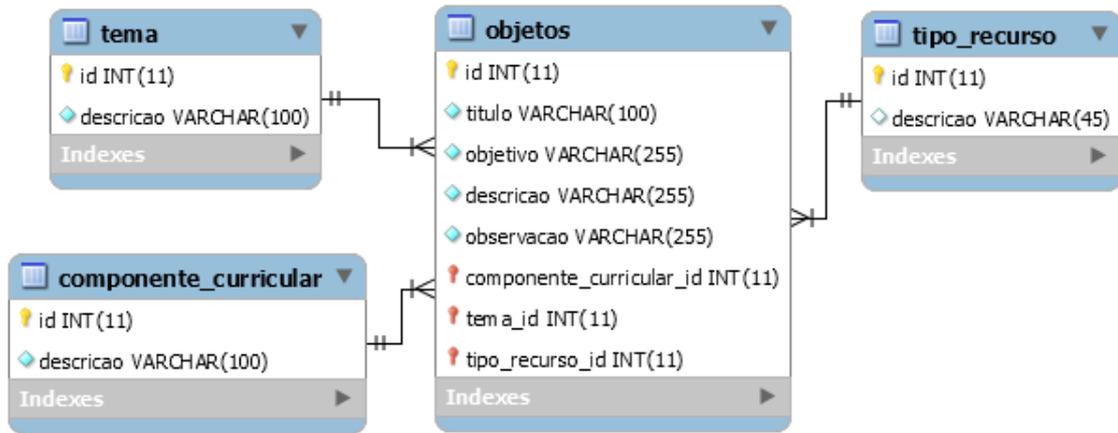


Figura 12: BALL - Estrutura da tabela Objetos

Fonte: Elaborada pelo autor

Na abordagem utilizada foi explorada uma relação entre os objetos, tomando-se por base a conexão entre eles e as visualizações e downloads. Esta conexão é criada de forma implícita, a partir das preferências do usuário, e serve como parâmetro para determinar a atualização do feromônio. Assim, a cada download de um OAL, é realizada uma verificação considerando o usuário ativo, o OAL anterior e o OAL alvo. Caso a relação exista, ou seja, o percurso já tenha sido realizado, o feromônio não é atualizado e o valor da visita é zero. Caso contrário, ambos os valores são atualizados. A figura 13 demonstra a estrutura criada para armazenar essas informações.

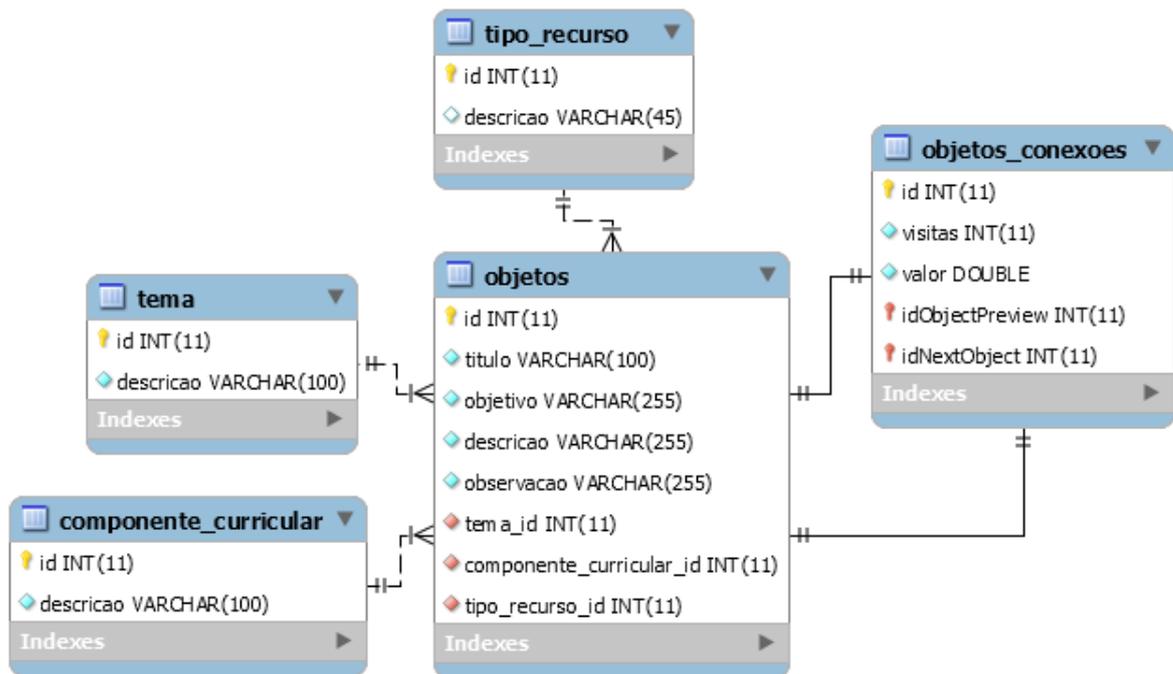


Figura 13: BALL - Estrutura da tabela conexões entre objetos

Fonte: Elaborada pelo autor

A figura 14 demonstra a forma como as informações são armazenadas. As colunas preview e next object representam a conexão entre dois OALs. A coluna visitas refere-se ao número de usuários que reforçaram esta conexão e a coluna valor representa o índice de feromônio que é incrementado a cada nova visita.

preview_object	next_object	visitas	valor
9649	9651	2	0.574654503272537
9649	9648	3	1.041238902496
9649	7825	1	0.24355708943028
9649	9645	1	0.26451972447555
9649	9650	1	0.21945593549682
9649	9647	1	0.22682140946202
9649	9652	1	0.22713379395621
9649	9653	1	0.20146962748054
9649	9646	1	0.18003416268001
9649	3425	3	0.35941054910166703
3425	9652	1	0.19086738858695
9652	9649	1	0.34258249233555

Figura 14: BALL - Conexões entre objetos

Fonte: Elaborada pelo autor

Além da solução proposta, foram implementados dois algoritmos de recomendação do Apache Mahout. As recomendações realizadas por estes algoritmos são executadas em segundo plano e utilizadas como base para avaliar a eficiência das

recomendações realizadas pelo sistema. Assim, para análise das consultas realizadas pelos usuários e recomendações geradas pelo sistema, foi criada uma estrutura na base de dados capaz de armazenar essas informações. Essa estrutura é composta por cinco tabelas, conforme representado pela figura 15.

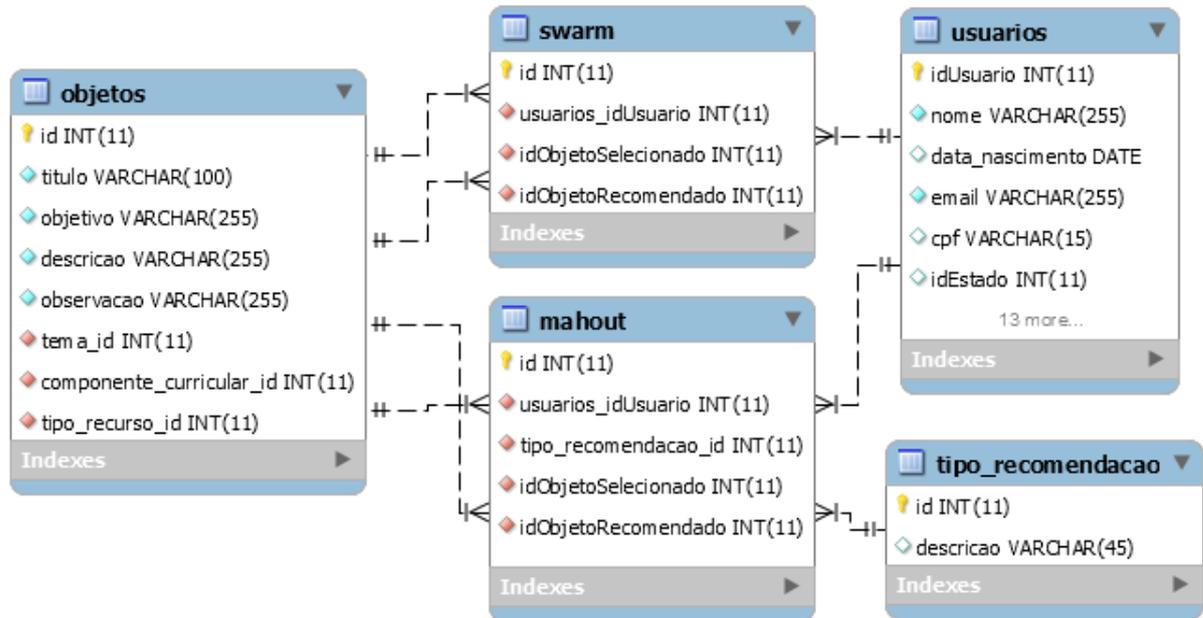


Figura 15: BALL - Estrutura de tabelas para armazenamento de recomendações

Fonte: Elaborada pelo autor

A tabela "swarm" (figura 15) é utilizada para armazenamento das recomendações realizadas através da abordagem proposta. As tabelas "tipo de recomendação" e "mahout" são utilizadas para armazenamento das recomendações realizadas pelos algoritmos do Apache Mahout. O tipo de recomendação refere-se à recomendação baseada em usuário e recomendação baseada em conteúdo.

Ao abrir o sistema é realizada uma recomendação inicial e são apresentados os objetos de aprendizagem mais acessados, ordenados de acordo com o peso TFI-IDF associado a cada objeto (Figura 16). Nesta etapa, o usuário pode visualizar e consultar objetos sem estar registrado, porém, para realizar um download deverá registrar-se e efetuar login. Da mesma forma, as estatísticas e o relacionamento entre objetos só são computados quando acessados por usuários registrados e logados no sistema.

BALL - Bank of Activities for Language Learning			
Home	Login	Logout	Hello Visitor
			Search
Most visited			
Id	Title	Description	Objective
9649	An alphabet of celebrities	This audio presents the alphabet in a different way. The letters of alphabet are showed according to the name of the celebrities who are important personalities in the world of literature	Teach the alphabet in a different way making associations between the name of some celebrities and the letters of the alphabet
9648	Footsteps on the road to learning	It's a poem about the alphabet. This poem helps students to learn how to use the alphabet. During the poem there are some words for each letter of alphabet and the rhythm gives more emphasize to the...	This audio helps children to learn the alphabet
9652	An alphabet of old friends	This is a poem which shows the alphabet throw the lines of a rhythm text	With this media the student can improve the vocabulary and assimilate the english alphabet with the words in english too

Figura 16: BALL - Tela inicial do sistema

Fonte: Elaborada pelo autor

Ao pesquisar por uma palavra-chave é realizada uma consulta na base de dados e calculada a medida TF-IDF para obtenção de uma lista de objetos relacionados à palavra-chave pesquisada (Figura 17). Essa abordagem objetiva fazer uma recomendação simples, tomando-se por base a importância da palavra-chave pesquisada em relação ao conjunto de objetos disponíveis para recomendação. Assim, objetos que não possuem nenhuma relação com a palavra pesquisada não são apresentados, enquanto outros são apresentados em ordem decrescente, de acordo com a medida TF-IDF associada a cada objeto.

Search results

Id	Title	Description	Objective	TF-IDF
3407	Typical breakfast in Great Britain	The audio presents Jennifer talking about a typical breakfast in Great Britain	The audio aims to work on listening comprehension and practice of vocabulary related to food habits	0.035985404093752
2915	My yearbook	The audio presents Gabrielle and Lindsey two american girls describing the importance of a yearbook	The audio aims to work on listening comprehension and practice of vocabulary related to past experiences and description of an yearbook	0.035038419775496
10968	What I eat for breakfast lunch and dinner	The audio presents Sally talking about food and her preferences for breakfast lunch and dinner	The audio aims to work on listening comprehension and practice of vocabulary related to food preferences and habits	0.035038419775496

Figura 17: BALL - Resultado da busca por palavra-chave

Fonte: Elaborada pelo autor

O cálculo envolve também um pré-processamento para remoção das chamadas "palavras de parada" ou "*stop words*", que se referem às palavras mais comuns em um determinado idioma, como por exemplo: "*a*", "*able*", "*about*", "*above*", "*abroad*", entre outras. São palavras consideradas irrelevantes para a consulta e que podem influenciar no processamento, alterando o resultado final.

A recomendação principal ocorre ao clicar em um objeto da lista. O usuário é direcionado para uma nova página, onde o objeto é disponibilizado para download e apresentado no topo com sua descrição completa (figura 18). O usuário tem então a opção de realizar o download ou selecionar outro objeto pertencente à lista de recomendações.

9645 - THE ILLUSTRATED ALPHABET OF BIRDS				Download	
Description	Objective	Theme	Curricular component		
This is a poem which shows the alphabet through the lines of a rhythm text. It is a good way to teach the alphabet in an English pronunciation and expanding the vocabulary	Improve vocabulary and assimilate the English alphabet	Educação Básica::Ensino Fundamental Final::Língua Estrangeira::Textos orais com marcas entonacionais e pronúncia	Ensino Fundamental::Séries Finais::Língua Estrangeira		

Recommended for you

Id	Title	Description	Objective	TF-IDF	SWARM
9652	An alphabet of old friends	This is a poem which...	With this media the...	0.00133862932080322	0.54098360655738
9653	The absurd ABC	This is a poem which ...	Improve the vocabulary...	0.00453631150272342	0.45901639344262

Other objects based on your search

Id	Title	Description	Objective	TF-IDF
9649	An alphabet of celebrities	This audio presents the alphabet in a different...	Teach the alphabet in a different way making...	0.001201334005849
9648	Footsteps on the road to learning	It's a poem about the alphabet. This poem helps...	This audio helps children to learn the alphabet	0.00089241954720214

Figura 18: BALL - Tela de download

Fonte: Elaborada pelo autor

Nesta etapa também são realizadas as recomendações pelos algoritmos do Apache Mahout. Essas recomendações, conforme já descrito, não são percebidas pelo usuário. Elas ocorrem em segundo plano e são armazenadas em uma base de dados para posterior avaliação. Ao fazer o download de um objeto, os dois algoritmos são executados, realizando recomendações baseadas em usuário e recomendações baseadas em conteúdo.

Além da recomendação realizada pelo algoritmo utilizado na solução proposta o sistema recomenda uma lista complementar de objetos pertencentes ao mesmo tema do objeto selecionado. Esses objetos são ordenados a partir do cálculo da medida TF-IDF aplicada ao grupo de objetos pertencentes ao tema em questão. Dessa forma, são apresentadas outras opções para o usuário quando não houver recomendações a serem realizadas, ou os objetos recomendados não forem de seu interesse.

Ao realizar o download de um objeto, e selecionar outro, recomendado a partir deste, o sistema estabelece uma conexão entre esses objetos. Essa conexão possui um valor que é incrementado a partir do acesso de outros usuários para o mesmo conjunto de objetos. Da mesma forma, se o usuário clica em um objeto, que possui uma conexão já estabelecida, e não realiza o download, essa conexão tem seu valor decrementado, podendo desfazer-se ao longo do tempo, de acordo com a variedade de interações.

A conexão entre os objetos é criada de forma implícita, através da interação do usuário com o sistema e os metadados associados aos OALs. Dessa forma, um objeto pode ser vinculado à diferentes contextos, de acordo com as necessidades e o conhecimento de cada usuário e não através de um voto atribuído de forma explícita.

6 EXPERIMENTOS E RESULTADOS

Neste capítulo são descritos os experimentos realizados para verificar a viabilidade do modelo proposto. São apresentados os procedimentos realizados, as pessoas envolvidas e a interpretação dos resultados obtidos.

6.1 Descrição dos experimentos

Para validação do sistema proposto foram cadastrados e disponibilizados por meio do protótipo desenvolvido 665 objetos de aprendizagem (OAs). Os experimentos foram realizados através da simulação de 10 usuários pré-cadastrados no sistema e contaram com a participação de um Professor (Doutor), especialista da área de Letras, e três professores de inglês em formação, bolsistas de iniciação científica vinculados ao projeto do Programa de Pós-Graduação em Letras (Mestrado) da Universidade Federal de Pelotas.

As primeiras duas etapas de simulação foram realizadas pelo professor especialista e tiveram como principal objetivo a criação de conexões entre os objetos. A etapa final contou com a participação das três bolsistas e serviu de instrumento para avaliação das recomendações realizadas pelo sistema. Os usuários foram nomeados de A, B, C, D, E, F, G, H, I e J, respectivamente, enquanto os objetos foram mapeados e numerados de 1 a 665. Os experimentos foram realizados conforme descrito a seguir.

Os primeiros acessos foram realizados através da simulação dos usuários A e B. Para a consulta inicial foi utilizado o termo *"listening comprehension"*. A partir dos resultados obtidos, foi explorada a recomendação de itens afins com o usuário A, fazendo downloads e reforçando relacionamento de itens recomendados com o usuário B. Posteriormente, a ordem dos usuários foi invertida, repetindo-se o mesmo procedimento, porém com o termo *"Food"*.

Para os usuários C e D, foram pesquisados os termos *"plural"*, *"colors"*, *"writing"*, *"speaking"*, *"vocabulary"*, *"animals"*, *"pronunciation"* e *"USA"*, fazendo downloads e associações, estabelecendo e reforçando conexões entre objetos, de acordo com as

recomendações geradas pelo sistema. A Tabela 5 apresenta o número de objetos acessados na primeira simulação com esses quatro usuários.

Tabela 5: Número de objetos acessados na 1ª simulação

Usuário	Número de objetos acessados
A	22
B	30
C	31
D	17

Os objetos acessados, assim como as relações resultantes da interação dos quatro usuários, podem ser visualizados na forma de um gráfico de rede. Esse tipo de gráfico assemelha-se a um multigrafo, onde os objetos e os usuários são representados na forma de nodos e a ligação entre ambos, decorrente dos acessos realizados, correspondem aos vértices do grafo.

O gráfico resultante da primeira simulação é demonstrado através da figura 19. Nesse gráfico, os usuários correspondem aos nodos A,B,C e D, enquanto cada nodo, representado por um número, corresponde a um objeto acessado pelo usuário. Os vértices representam a ligação entre os objetos e os usuários, formando uma espécie de rede, conectada a partir dos interesses de cada usuário.

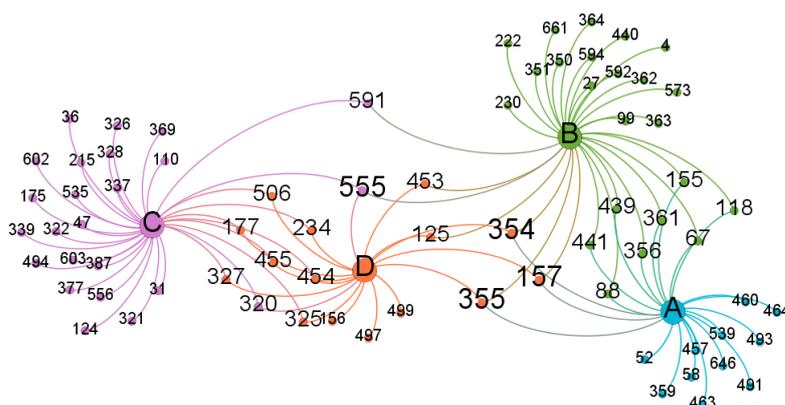


Figura 19: Gráfico resultante da 1ª etapa de simulação

Através da figura 19 é possível perceber que dos 30 objetos acessados pelo usuário B, 11 também foram acessados pelo usuário A. Essa relação de preferência entre objetos é caracterizada pela semelhança entre os termos pesquisados nos primeiros acessos e as recomendações realizadas a partir da interação do usuário B. Da mesma forma, percebe-se que dos 31 objetos selecionados pelo usuário C, 8 também foram acessados pelo usuário D, apenas 2 pelo usuário B e nenhum pelo usuário A. Essa relação entre C e D é maior por terem pesquisado pelos mesmos termos, assim

como os usuários A e B. Porém, apesar da diferença entre os termos da pesquisa, percebe-se que o usuário D, mesmo acessando uma quantidade menor de objetos, está conectado a todos os outros usuários. Isso porque, através da solução proposta por este trabalho, essa relação vai sendo ampliada ao longo do tempo, mediante o reforço das conexões entre objetos, originadas a partir da interação dos usuários.

Os acessos seguintes foram realizados através da simulação dos usuários E, F, G e H. O procedimento adotado foi o mesmo utilizado para os usuários A, B, C e D. Os dados obtidos a partir da interação desses usuários são apresentados na tabela 6.

Tabela 6: Número de objetos acessados na 2ª etapa de simulação

Usuário	Número de objetos acessados
E	10
F	15
G	18
H	17

O gráfico resultante da 2ª etapa de simulação, representado pela figura 20, demonstra um aumento da similaridade entre as preferências dos usuários. Percebe-se que os usuários D, E, F, G e H estão localizados mais ao centro do gráfico, isso porque os objetos escolhidos por esses usuários, ou pelo menos grande parte deles, também estão conectados à outros usuários. Essa similaridade é denotada pela formação de clusters, originados a partir dos acessos realizados pelos usuários e representados pelas diferentes cores do gráfico. Os objetos que aparecem em maior destaque no gráfico, são aqueles que tiveram um número maior de acessos durante a interação dos usuários.

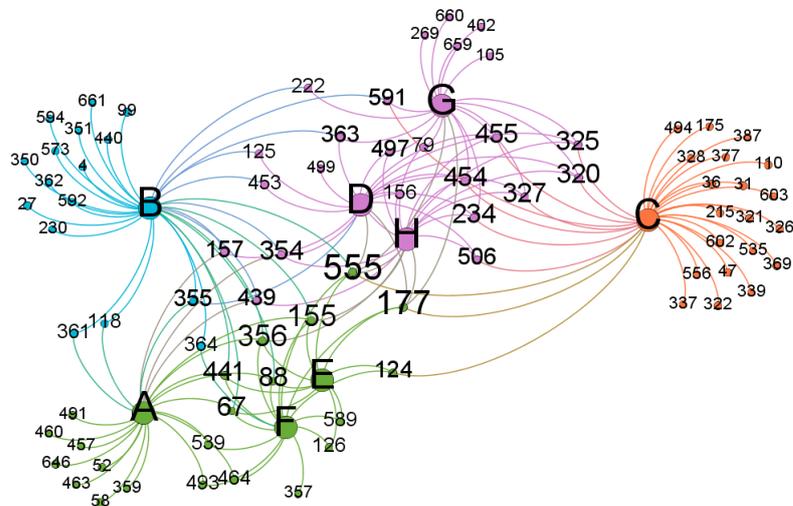


Figura 20: Gráfico resultante da 2ª etapa de simulação

Os últimos experimentos foram realizados pelas três professoras de inglês em formação, bolsistas de iniciação científica vinculados ao projeto do Programa de Pós-Graduação em Letras. As professoras foram convidadas pelo professor especialista a participar dos experimentos e receberam instruções para acesso e realização de buscas no sistema. A simulação foi orientada pelo professor especialista e deu-se através dos seguintes passos:

- Fornecimento de dados necessários para login e acesso ao sistema;
- Cada usuário foi orientado a pesquisar por termos específicos, em dois dias diferentes, conforme ordem apresentada na tabela 7.

Tabela 7: Dados utilizados na simulação final

Usuário	Simulação	Termos
I	1º dia	Listening comprehension, USA, Color
	2º dia	Animal, Plural, Stereotypes
J	1º dia	Food, Pronunciation, Cat
	2º dia	Listening comprehension, USA, Color
K	1º dia	Animal, Plural, Stereotypes
	2º dia	Food, Pronunciation, Cat

- Após o sistema retornar os resultados, os usuários foram orientados a escolher aquele que melhor atende à sua pesquisa;
- Ao ser direcionado para a página com a descrição completa do objeto (fictício), o usuário tem a opção de realizar o download, ou não, devendo considerar nesse caso a relevância do objeto em questão para a consulta realizada.
- Na mesma página, abaixo da descrição do objeto, após o título *“Recommended for you”*, são apresentados os objetos recomendados pelo SR. Havendo mais de um objeto listado, o usuário deve escolher o que mais atende sua busca (hipoteticamente) e repetir o passo 4. Caso não haja nenhum objeto listado, deve então repetir a busca na caixa *“Search”*, acessar a nova listagem de objetos recomendados e repetir o passo 4.

O número de objetos acessados pelos 3 usuários é apresentado na tabela 8.

Tabela 8: Número de objetos acessados na 3ª etapa de simulação

Usuário	Número de objetos acessados
I	11
J	12
K	12

O gráfico resultante da 3ª etapa de simulação é representado pela figura 21.

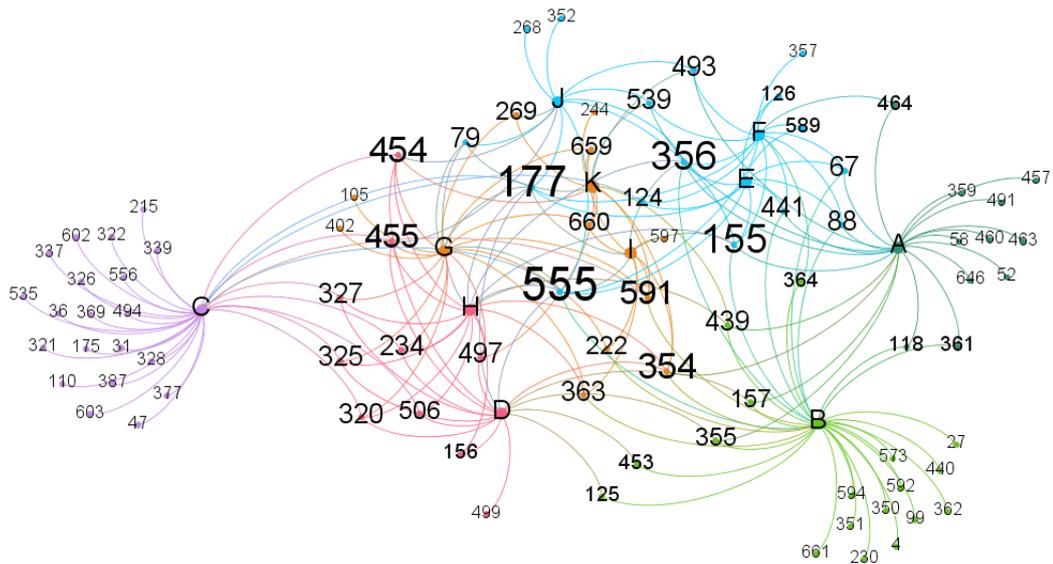


Figura 21: Gráfico resultante da 3ª etapa de simulação

O gráfico apresentado pela figura 21 demonstra que os usuários I, J e K estão localizados mais ao centro da rede. Isso porque, com exceção de dois objetos (268 e 352), selecionados pelo usuário J, todos os demais objetos também foram escolhidos por outros usuários. Essa característica demonstra que a maioria das escolhas realizadas pelos usuários foram influenciadas pelas recomendações do sistema. Os resultados obtidos a partir dos experimentos comprovam esse fato, e são descritos na próxima seção.

Durante as três etapas de simulação, foram armazenados dados relacionados às estatísticas de acesso aos objetos. No total foram acessados 86 objetos de aprendizagem, resultando em 269 visualizações e 252 downloads. A tabela 9 apresenta um exemplo do conteúdo armazenado.

Tabela 9: Ex. de dados de acesso, resultantes das 3 etapas de simulação

Usuário	Objeto	Visualizações	Downloads	Valor
H	497	2	2	0.21498369737568002
K	497	1	1	0.01306330384692
F	539	1	1	0.046560078490399996
J	539	2	1	0.09177990811627999
K	539	1	1	0.01096989519034
B	555	1	1	0.01146268931654

O campo valor apresentado na tabela 9 é dado pela medida TF-IDF, utilizada neste trabalho para atribuição de avaliações implícitas relacionadas aos objetos. A cada interação do usuário o campo valor é incrementado de acordo com a ação realizada pelo usuário. Durante os experimentos essas informações foram acessadas diretamente na base de dados para gerar recomendações através dos algoritmos do Apache Mahout. O acesso se deu através da utilização da classe JDBCDataModel. A figura 22 apresenta um exemplo do código utilizado.

```

public class RecommenderDAO extends DAO{
    public JDBCDataModel dataModel(){
        MySQLDataSource datasource = new MySQLDataSource();
        datasource.setServerName("host");
        datasource.setUser("user");
        datasource.setPassword("password");
        datasource.setDatabaseName("database");
        JDBCDataModel dataModel = new MySQLJDBCDataModel(
            datasource, "usuarios_objetos",
            "id_usuario", "id_objeto",
            "valor", null);
        return dataModel;
    }
}

```

Figura 22: Exemplo de código utilizado para popular a interface DataModel

Fonte: Elaborada pelo autor

Para implementação desses algoritmos foi utilizada a biblioteca Taste, cujos componentes foram descritos na seção 2.5. A figura 23 apresenta um exemplo do código utilizado para recomendações baseadas em usuário.

```

public List<RecommendedItem> UserBased(int user) throws TasteException{
    JDBCDataModel dataModel = new RecommenderDAO().dataModel();

    UserSimilarity similarity =
        new PearsonCorrelationSimilarity(dataModel);

    UserNeighborhood neighbourhood =
        new NearestNUserNeighborhood(3, similarity, dataModel);

    Recommender recommender =
        new GenericUserBasedRecommender(dataModel,
            neighbourhood, similarity);

    List<RecommendedItem> recommendations;
    recommendations = recommender.recommend(user, 10);
    return recommendations;
}

```

Figura 23: Exemplo de código para recomendação baseada em usuário

Fonte: Elaborada pelo autor

A medida de similaridade utilizada nos algoritmos do Apache Mahout foi o coeficiente de correlação de Pearson, representado pela Equação 7 (amplamente abordado na literatura sobre SR) e adotado também pelo GroupLens (RESNICK et al., 1994), um dos primeiros sistemas de filtragem colaborativa.

6.2 Resultados obtidos

Conforme já descrito, a abordagem proposta prevê uma recomendação mediada pelo ambiente, através das conexões entre objetos, estabelecidas a partir da interação dos usuários. Através da solução apresentada por este trabalho, as três etapas de simulação resultaram na criação de 72 conexões entre objetos de aprendizagem. Essas conexões, quando representadas através de um grafo, assemelham-se aos caminhos construídos pelas formigas durante o forrageamento. A figura 29 demonstra essa característica.

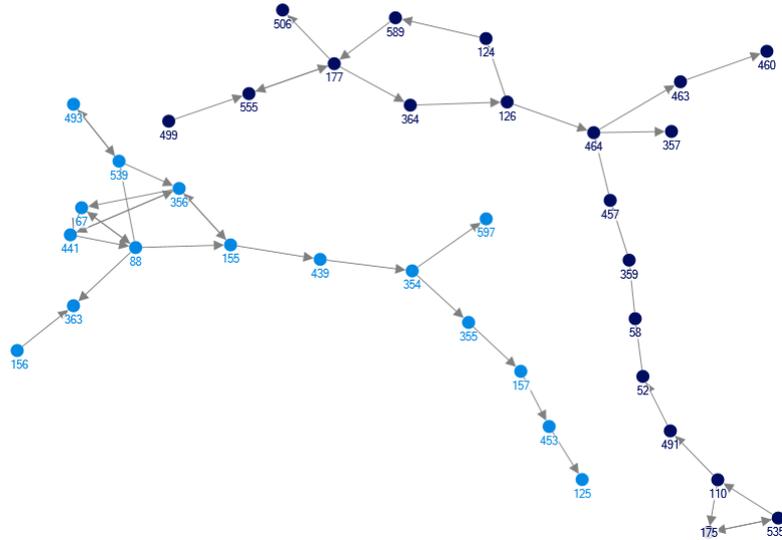


Figura 24: Exemplo de conexões resultantes dos experimentos
 Fonte: Elaborada pelo autor

A cada interação podem surgir novas conexões, ocorrer o reforço ou serem desfeitas conexões já existentes. A intensidade das conexões, resultantes dos experimentos, evidencia essa característica e pode ser visualizada através da figura 25.

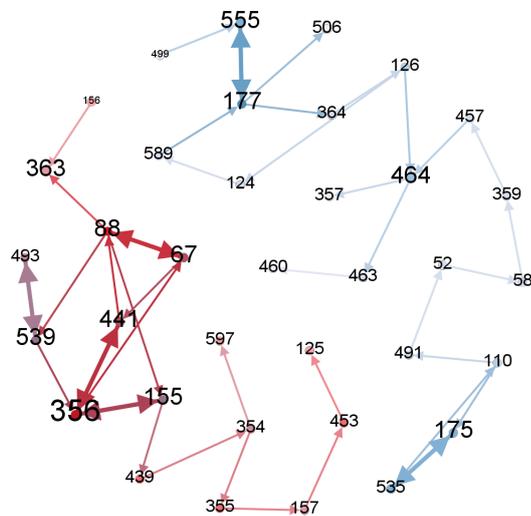


Figura 25: Intensidade das conexões

Tomando-se como exemplo o objeto 177 da figura 25, é possível perceber que ele está conectando e referenciando os objetos 555, 364 e 506. A espessura da aresta denota a intensidade da conexão entre os objetos, sendo mais evidente entre os objetos 177 e 555. Essa intensidade é determinada pelas ações do usuário e calculada de acordo com a fórmula 13.

A tabela 10 apresenta os valores resultantes das conexões estabelecidas entre os

objetos 177 e 88, que aparecem em destaque na figura 25. Através desse exemplo, é possível perceber que, apesar de haver apenas 1 visita em cada conexão do objeto 88, ele também é referenciado pelo objeto 67, reforçando a conexão entre ambos, conforme apresentado pelo gráfico da figura 25.

Tabela 10: Valores associados às conexões

Objeto 01	Objeto 02	nº de visitas	valor
177	506	3	0.02181577381956
177	555	6	5.623404391201620
177	364	1	0.00130210145424
88	363	1	0.00203450848225
88	155	1	0.00242404106039
88	539	1	0.00539399568028
88	67	1	0.0050215004294
67	88	1	0.00250540171723

A fim de facilitar a visualização e compreensão acerca do exposto, neste exemplo foram considerados apenas dois conjuntos de conexões. Apesar disso, através da comparação dos resultados apresentados pelas figuras 21 e 25, é possível perceber que os objetos mais acessados são aqueles que estão mais fortemente conectados. Isso porque o valor associado a uma conexão é utilizado para determinar a relevância de um determinado objeto em relação ao conjunto de objetos a serem recomendados.

As duas primeiras etapas de simulação tiveram como objetivo principal relacionar objetos a partir da experiência do professor especialista. Dessa forma, para análise das recomendações realizadas pelo sistema, foram considerados os acessos realizados pelos usuários I, J e K. A tabela 11 apresenta parte da lista de objetos recomendados a partir da interação desses usuários.

Tabela 11: Exemplo de recomendações para os usuários I, J e K

Objeto selecionado	Objeto recomendado	Valor
155	356	4.059700848181420
155	439	0.010284217713809999
177	555	5.623404391201620
177	506	0.02181577381956
177	364	0.00130210145424
354	355	0.00355263866103

Através dos dados apresentados na tabela 11, é possível perceber que as

recomendações são realizadas com base nas conexões estabelecidas entre os objetos de aprendizagem. Os objetos são recomendados, ordenados de acordo com o valor associado a cada conexão, calculado pela equação 13.

A figura 26 apresenta o gráfico de recomendações resultante das 3 etapas de experimentos.

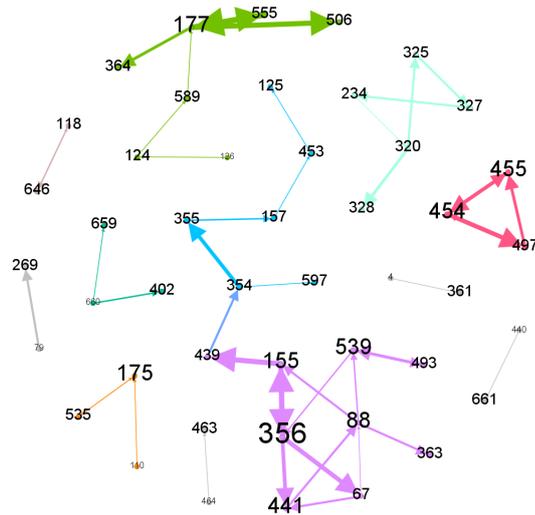


Figura 26: Gráfico de recomendações geradas pelo sistema

Através dos algoritmos do Apache Mahout foram realizadas 17 recomendações baseadas em item e 159 recomendações baseadas em usuário. O total de recomendações realizadas para os três usuários é apresentado na tabela 12.

Tabela 12: Recomendações realizadas a partir da interação dos usuários I, J e K

Objeto selecionado	Recomendações		
	BALL	Mahout (Baseada em item)	Mahout (baseada em usuário)
155	356, 439	589, 355, 455	157, 354, 355, 439, 441
177	555, 506, 364	67	67, 354, 363, 441, 555, 88, 157, 355
354	355, 597	–	234, 325, 454, 455, 497, 88, 118, 155, 356, 555
356	155, 441, 67	441	67, 88, 177, 441, 555
555	177	589, 355	234, 325, 454, 455, 497, 157, 354, 355
660	402, 659	–	234, 325, 454, 455, 497
79	269	–	67, 88, 177, 441, 555, 155, 356, 455
454	497, 455	506, 493	67, 155, 177, 356, 555, 455
455	454	–	67, 88, 155, 356, 441
493	539	464	67, 88, 155, 356, 441, 454, 455, 497
539	493, 356	–	67, 88, 155, 356, 441, 454, 455, 497
439	354	–	177, 222, 354, 356, 363
497	455, 454	506	88, 118, 356, 441, 555

6.3 Avaliação das recomendações

Diferentes parâmetros de avaliação tem sido sugeridos na literatura de Sistemas de Recomendação. Cada métrica possui propriedades importantes que devem ser consideradas como forma de avaliar à qual tarefa sua aplicação é mais adequada (GUNAWARDANA; SHANI, 2009). Para a tarefa de prever classificações, por exemplo, o sistema deve fornecer um conjunto de classificações previstas, sendo avaliado de acordo com a precisão dessas classificações. Nesse contexto destaca-se o Root of the Mean Square Error (RMSE), um método bastante utilizado para a análise da pontuação obtida a partir de diferentes algoritmos. A equação do RMSE é dada por:

$$\sqrt{\frac{\sum_{(i,j) \in K} (p_{i,j} - v_{i,j})^2}{n}} \quad (13)$$

Onde:

$p_{i,j}$ é a classificação prevista para o usuário i sobre o item j , $v_{i,j}$ é a avaliação correta e $K = \{(i, j)\}$ é o conjunto de classificações usuário-item escondidas.

Segundo o mesmo autor, o RMSE é adequado para análise da tarefa de predição porque mede imprecisões em todas as avaliações, sejam positivas ou negativas.

Outras medidas bastante utilizadas para a análise de predições em SR são Mean Square Error, Mean Average Error (MAE) e Normalized Mean Average Error (NMAE). Comparando as três abordagens, o autor observa que o RMSE tende a penalizar os maiores erros de forma mais severa do que as outras métricas, enquanto NMAE normaliza MAE através do conjunto de avaliações para facilitar a comparação de erros em vários domínios.

Uma outra forma de avaliação consiste na análise de conjuntos de dados de seleção binária, que podem ser representados por meio de itens visitados ou não visitados. Como por exemplo, o histórico de cliques em notícias, onde pode-se atribuir um valor de 1 para cada notícia visitada e 0 para notícias não visitadas. Essa forma de avaliação é aplicável à sistemas onde a tarefa de recomendação consiste em fornecer, dada uma lista existente de itens que foram vistos, uma lista de itens adicionais que o usuário pode querer visitar (GUNAWARDANA; SHANI, 2009).

Dadas as características da aplicação proposta pelo presente trabalho e o conjunto de dados obtido a partir das três etapas de simulação, a segunda forma de avaliação é a que melhor se aplica para análise das recomendações realizadas pelos três algoritmos aqui implementados. Dessa forma, pode-se classificar os resultados de tais recomendações utilizando a Tabela 13.

Tabela 13: Classificação do resultado da recomendação de um item para um usuário

	Recomendado	Não recomendado
Selecionado	Verdadeiro-Positivo (vp)	Falso-Negativo (fn)
Não selecionado	Falso-Positivo (fp)	Verdadeiro-Negativo (vn)

A tabela 14 apresenta os valores resultantes da análise das recomendações realizadas pelos três algoritmos implementados.

Tabela 14: Valores obtidos a partir da análise das recomendações

	vp	fp	vn	fn
Swarm	14	6	637	8
Mahout (baseada em item)	2	6	637	20
Mahout (baseada em usuário)	12	8	635	10

Com base nos valores da tabela 14 podemos calcular as seguintes medidas:

Precision: Utilizada para descrever a quantidade de recomendações realizadas pelo sistema que foram adequadas para o usuário alvo.

$$Precision = \frac{vp}{vp + fp} \quad (14)$$

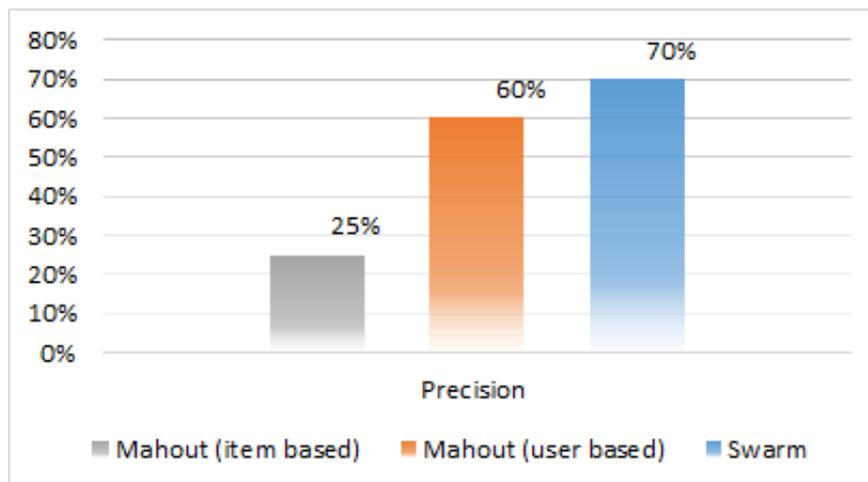


Figura 27: Percentual de recomendações, geradas pelos três algoritmos, consideradas relevantes para os usuários I, J e K

Recall (True Positive Rate): Utilizada para medir a proporção de recomendações corretamente realizadas (verdadeiros positivos) sobre o número total de

recomendações esperadas (verdadeiros positivos e falsos negativos).

$$Recall(TruePositiveRate) = \frac{vp}{vp + fn} \quad (15)$$

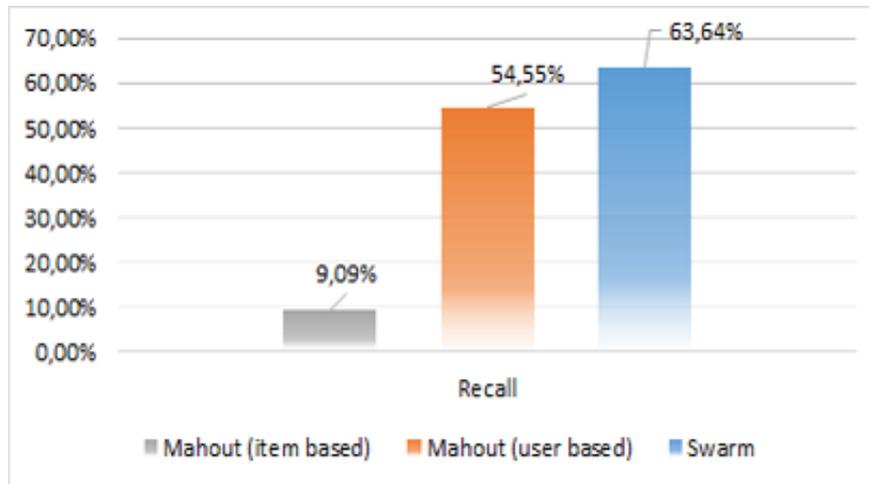


Figura 28: Percentual de recomendações consideradas relevantes, para os usuários I, J e K, em relação ao número total de recomendações esperadas

False Positive Rate: A taxa de falsos positivos descreve a proporção de recomendações realizadas pelo sistema que foram inadequadas para usuário alvo.

$$FalsePositiveRate = \frac{fp}{fp + vn} \quad (16)$$

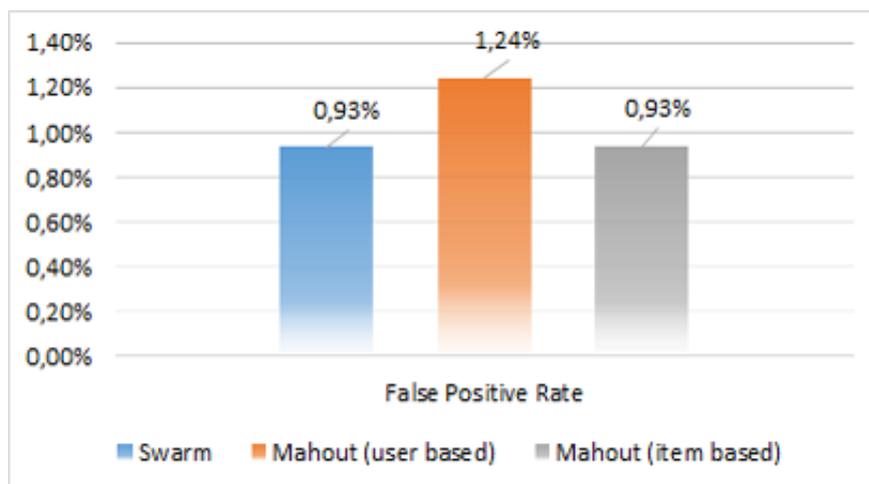


Figura 29: Percentual de recomendações inadequadas para os usuários I, J e K

A aplicação do algoritmo do Apache Mahout, baseado em conteúdo, foi limitada ao tema do objeto selecionado dentre aqueles resultantes da lista de objetos recomen-

dados após a consulta por palavra-chave. O algoritmo foi implementado de forma a recomendar itens com maior similaridade ao objeto selecionado, resultando em um número menor de recomendações em comparação à abordagem proposta por este trabalho. O algoritmo do Apache Mahout (baseado em usuário) e o algoritmo baseado em ACO, não tiveram sua aplicação limitada ao tema do objeto selecionado. Dessa forma, os resultados apresentados na tabela 12 demonstram que o algoritmo baseado em usuário retornou uma quantidade maior de itens durante o processo de recomendação. Isto porque foram avaliados e recomendados todos os objetos que possuem algum valor associado na tabela 9, utilizada para popular o DataModel cujos dados são avaliados no processo de recomendação do Apache Mahout.

O conjunto de resultados obtidos a partir da aplicação dos três algoritmos, demonstra que as recomendações realizadas para os usuários I, J e K, através da abordagem proposta, obtiveram um melhor desempenho em comparação aos algoritmos do Apache Mahout (tabela 30).

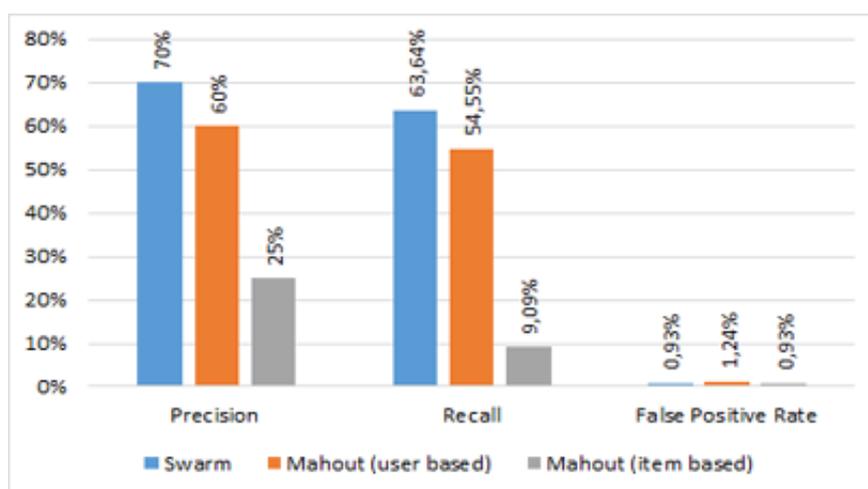


Figura 30: Percentuais obtidos a partir da análise dos três algoritmos

Além dos resultados obtidos a partir do cálculo das medidas *Precision*, *Recall* e *False Positive Rate*, através da análise realizada pelo professor especialista, pode-se concluir que a solução apresentada por este trabalho se mostra eficiente para o contexto de aplicação a que se propõe, uma vez que retornou resultados mais relacionados ao termo consultado pelos usuários do que outros modelos. Outrossim, proporcionou o estabelecimento de rotas, construídas através da aplicação do conhecimento obtido de forma implícita por meio da interação dos usuários com o sistema. Essas rotas originam-se das conexões conceituais estabelecidas entre os objetos, oferecendo dessa forma, segundo o professor especialista, maior potencial de aprendizagem e formação docente (continuada).

7 CONCLUSÕES

Este trabalho apresentou uma proposta para aplicação de técnicas de inteligência de enxames para a recomendação de objetos de aprendizagem de línguas. Os resultados obtidos a partir dos experimentos demonstraram que a solução proposta mostrou-se eficiente para o contexto de aplicação a que se propõe.

Como forma de validar a solução proposta foram implementados, além da abordagem apresentada, dois dos algoritmos de recomendação de uso geral mais empregados atualmente. Os resultados obtidos a partir dos experimentos demonstraram que esses algoritmos apresentam melhores resultados quando aplicados a um tema específico. Caso contrário, acabam por sobrecarregar o usuário com informações que na maioria das vezes não possuem conteúdo relevante para o tema que está sendo pesquisado.

Considerando ainda o contexto de aplicação, estudos preliminares realizados demonstraram que um objeto de aprendizagem de línguas tem como objetivo principal ser um facilitador do processo de integração de competências gramatical, sociolinguística e estratégica. Dessa forma, dois aspectos fundamentais a serem considerados durante o processo de recomendação, foram contemplados através da abordagem proposta por este trabalho, quais sejam: a capacidade de combinar um objeto com outro e a reusabilidade, isto é, a possibilidade de sua aplicação em diferentes contextos de aprendizagem. Para tanto, a tarefa de recomendação prevê a existência de múltiplas conexões conceituais entre os objetos, que são construídas à medida que o sistema é utilizado, e ocorre através da aplicação do algoritmo de Otimização por Colônia de Formigas.

Além da recomendação realizada através da solução proposta, o sistema recomenda uma lista complementar de objetos pertencentes ao mesmo tema do objeto alvo. Esses objetos são ordenados a partir do cálculo da medida TF-IDF. Dessa forma, são apresentadas outras opções para o usuário quando não houver recomendações a serem realizadas, ou os objetos recomendados não forem de seu interesse. Essa funcionalidade ajuda a lidar com dois problemas conhecidos em sistema de recomendação: o problema do novo usuário e o problema do novo item. Além

disso, proporciona maior facilidade de acesso aos objetos disponíveis, atendendo a mais uma característica desejável durante o processo de recomendação de objetos de aprendizagem: a recuperabilidade.

Conforme já descrito, a tarefa de recomendação prevê a existência de múltiplas conexões conceituais entre os objetos. Essas conexões são construídas à medida que o sistema é utilizado e ocorrem de acordo com as ações dos usuários. Dessa forma, ao longo do tempo podem ser estabelecidas novas conexões, ocorrer o reforço ou a extinção de conexões existentes. Essa característica adaptativa ajuda a lidar com a super especialização, outro problema bastante comum em sistemas de recomendação.

7.1 Trabalhos futuros

Durante o desenvolvimento desta dissertação foi concluída parte da implementação do repositório BALL (Banco de Atividades para a Aprendizagem de Línguas), disponibilizando recursos suficientes para a inserção e catalogação de objetos. Esse repositório foi desenvolvido com base em princípios pedagógicos comunicativos e levando em consideração o caráter complexo da aprendizagem de línguas. Os metadados são baseados em uma ontologia fundamentada pelos princípios da CLT, elaborados a partir da definição de uma estrutura de linguagem técnica e principalmente comunicativa. Certamente o uso desses metadados terá grande influência durante o estabelecimento de conexões e, conseqüentemente, no processo de recomendação. Dessa forma, como trabalhos futuros pretende-se implementar e testar a solução proposta neste repositório, utilizando um maior número de usuários, com aplicação em situações reais, para verificar se seu desempenho se mantém satisfatório.

REFERÊNCIAS

- ADOMAVICIUS, G.; TUZHILIN, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. **Knowledge and Data Engineering, IEEE Transactions on**, [S.l.], v.17, n.6, p.734–749, 2005.
- ANSARI, A.; ESSEGAIER, S.; KOHLI, R. Internet recommendation systems. **Journal of Marketing research**, [S.l.], v.37, n.3, p.363–375, 2000.
- BALABANOVIĆ, M.; SHOHAM, Y. Fab: Content-based, Collaborative Recommendation. **Commun. ACM**, New York, NY, USA, v.40, n.3, p.66–72, Mar. 1997.
- BASU, C.; HIRSH, H.; COHEN, W. et al. Recommendation as classification: Using social and content-based information in recommendation. In: AAAI/IAAI, 1998. **Anais...** [S.l.: s.n.], 1998. p.714–720.
- BEDI, P.; SHARMA, R. Trust based recommender system using ant colony for trust computation. **Expert Systems with Applications**, [S.l.], v.39, n.1, p.1183–1190, 2012.
- BELLAACHIA, A.; ALATHEL, D. Improving the Recommendation Accuracy for Cold Start Users in Trust-Based Recommender Systems. **International Journal of Computer and Communication Engineering**, [S.l.], v.5, n.3, p.206, 2016.
- BOFFO, F.; FERREIRA JR, P. R.; BAZZAN, A. L. A comparison of algorithms for task allocation in robocup rescue. In: EUROPEAN WORKSHOP ON MULTIAGENT SYSTEMS, 5., 2007. **Proceedings...** [S.l.: s.n.], 2007. p.537–548.
- BONABEAU, E.; DORIGO, M.; THERAULAZ, G. **Swarm intelligence: from natural to artificial systems**. [S.l.]: Oxford university press, 1999. n.1.
- BONABEAU, E.; THERAULAZ, G.; DENEUBOURG, J.-L.; ARON, S.; CAMAZINE, S. Self-organization in social insects. **Trends in Ecology & Evolution**, [S.l.], v.12, n.5, p.188–193, 1997.
- BRAAM, R. R.; MOED, H. F.; VAN RAAN, A. F. Mapping of science by combined co-citation and word analysis I. Structural aspects. **Journal of the American Society for information science**, [S.l.], v.42, n.4, p.233, 1991.

BREESE, J. S.; HECKERMAN, D.; KADIE, C. Empirical analysis of predictive algorithms for collaborative filtering. In: FOURTEENTH CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, 1998. **Proceedings...** [S.l.: s.n.], 1998. p.43–52.

BURKE, R. Hybrid recommender systems: Survey and experiments. **User modeling and user-adapted interaction**, [S.l.], v.12, n.4, p.331–370, 2002.

BURKE, R. Hybrid Web Recommender Systems. In: BRUSILOVSKY, P.; KOBASA, A.; NEJDL, W. (Ed.). [S.l.]: Springer Berlin Heidelberg, 2007. p.377–408. (Lecture Notes in Computer Science, v.4321).

CANALE, M. From communicative competence to communicative language pedagogy. **Language and communication**, [S.l.], v.1, p.1–47, 1983.

CANALE, M.; SWAIN, M. Theoretical bases of communicative approaches to second language teaching and testing. **Applied linguistics**, [S.l.], v.1, n.1, p.1–47, 1980.

CASTRO, L. N. de. Fundamentals of natural computing: an overview. **Physics of Life Reviews**, [S.l.], v.4, n.1, p.1–36, 2007.

CAZELLA, S. C.; BEHAR, P.; SCHNEIDER, D.; SILVA, K. K. da; FREITAS, R. Desenvolvendo um Sistema de Recomendação de Objetos de Aprendizagem baseado em Competências para a Educação: relato de experiências. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2012. **Anais...** [S.l.: s.n.], 2012. v.23, n.1.

CAZELLA, S. C.; NUNES, M.; REATEGUI, E. B. A Ciência da Opinião: Estado da arte em Sistemas de Recomendação. In: XXX CONGRESSO DA SBC JORNADA DE ATUALIZAÇÃO DA INFORMÁTICA, PORTO ALEGRE, BRASIL, 52P, 2010. **Anais...** [S.l.: s.n.], 2010.

CAZELLA, S. C.; REATEGUI, E.; MACHADO, M.; BARBOSA, J. Recomendação de objetos de aprendizagem empregando filtragem colaborativa e competências. **Simpósio Brasileiro de Informática na Educação (SBIE)**, [S.l.], 2009.

CLAYPOOL, M.; GOKHALE, A.; MIRANDA, T.; MURNIKOV, P.; NETES, D.; SARTIN, M. Combining content-based and collaborative filters in an online newspaper. In: ACM SIGIR WORKSHOP ON RECOMMENDER SYSTEMS, 1999. **Proceedings...** [S.l.: s.n.], 1999. v.60.

DENEUBOURG, J.-L.; ARON, S.; GOSS, S.; PASTEELS, J. M. The self-organizing exploratory pattern of the argentine ant. **Journal of insect behavior**, [S.l.], v.3, n.2, p.159–168, 1990.

DENEUBOURG, J.-L.; GOSS, S. Collective patterns and decision-making. **Ethology Ecology & Evolution**, [S.l.], v.1, n.4, p.295–311, 1989.

DORIGO, M.; BIRATTARI, M.; BLUM, C.; CLERC, M.; STÜTZLE, T.; WINFIELD, A. **Ant Colony Optimization and Swarm Intelligence**: 6th International Conference, ANTS 2008, Brussels, Belgium, September 22-24, 2008, Proceedings. [S.l.]: Springer, 2008. v.5217.

DORIGO, M.; BIRATTARI, M.; STÜTZLE, T. Ant colony optimization. **Computational Intelligence Magazine, IEEE**, [S.l.], v.1, n.4, p.28–39, 2006.

DORIGO, M.; BIRATTARI, M.; STUTZLE, T. **Ant colony optimization**: artificial ant as a computational intelligence technique. University Libre De Bruxelles. [S.l.]: IRIDIA Technical report Series, Belgium, 2006.

DORIGO, M.; BONABEAU, E.; THERAULAZ, G. Ant algorithms and stigmergy. **Future Generation Computer Systems**, [S.l.], v.16, n.8, p.851–871, 2000.

DORIGO, M.; DI CARO, G.; GAMBARDELLA, L. M. Ant algorithms for discrete optimization. **Artificial life**, [S.l.], v.5, n.2, p.137–172, 1999.

DORIGO, M.; MANIEZZO, V.; COLORNI, A. Ant system: optimization by a colony of cooperating agents. **Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on**, [S.l.], v.26, n.1, p.29–41, 1996.

DORIGO, M.; MANIEZZO, V.; COLORNI, A.; DORIGO, M.; DORIGO, M.; MANIEZZO, V.; MANIEZZO, V.; COLORNI, A.; COLORNI, A. **Positive Feedback as a Search Strategy**. [S.l.: s.n.], 1991.

FELDER, R. M.; SILVERMAN, L. K. Learning and teaching styles in engineering education. **Engineering education**, [S.l.], v.78, n.7, p.674–681, 1988.

FERREIRA JR, P. R.; BOFFO, F. S.; BAZZAN, A. L. Using Swarm-GAP for distributed task allocation in complex scenarios. In: **Massively Multi-Agent Technology**. [S.l.]: Springer, 2007. p.107–121.

GIACOMELLI, P. **Apache mahout cookbook**. [S.l.]: Packt Publishing Ltd, 2013.

GOOD, N.; SCHAFER, J. B.; KONSTAN, J. A.; BORCHERS, A.; SARWAR, B.; HERLOCKER, J.; RIEDL, J. Combining collaborative filtering with personal agents for better recommendations. In: AAAI/IAAI, 1999. **Anais...** [S.l.: s.n.], 1999. p.439–446.

GOSS, S.; ARON, S.; DENEUBOURG, J.-L.; PASTEELS, J. M. Self-organized shortcuts in the Argentine ant. **Naturwissenschaften**, [S.l.], v.76, n.12, p.579–581, 1989.

GRASSÉ, P. Les Insectes Dans Leur Univers, Paris, France, Ed. **Du Palais de la découverte**, [S.l.], 1946.

GRASSÉ, P.-P. La reconstruction du nid et les coordinations interindividuelles chez *Bellicositermes natalensis* et *Cubitermes* sp. la théorie de la stigmergie: Essai d'interprétation du comportement des termites constructeurs. **Insectes sociaux**, [S.l.], v.6, n.1, p.41–80, 1959.

GUNAWARDANA, A.; SHANI, G. A survey of accuracy evaluation metrics of recommendation tasks. **Journal of Machine Learning Research**, [S.l.], v.10, n.Dec, p.2935–2962, 2009.

HEYLIGHEN, F. Stigmergy as a generic mechanism for coordination: definition, varieties and aspects. **Available at: {<http://pespmc1.vub.ac.be/Papers/Stigmergy-WorkingPaper.pdf>} (accessed 04.2014)**, [S.l.], 2011.

HILL, W.; STEAD, L.; ROSENSTEIN, M.; FURNAS, G. Recommending and evaluating choices in a virtual community of use. In: 1995. **Anais...** [S.l.: s.n.], 1995. p.194–201.

HODGINS, W. H. The future of learning objects. In: LEARNING OUTCOMES PROVIDING FUTURE POSSIBILITIES, 2002. **Anais...** ECI Conference on e-Technologies in Engineering Education, 2002. p.76–82.

JANNACH, D.; ZANKER, M.; FELFERNIG, A.; FRIEDRICH, G. **Recommender Systems: An Introduction**. 1st.ed. New York, NY, USA: Cambridge University Press, 2010.

KARABOGA, D.; AKAY, B. A survey: algorithms simulating bee swarm intelligence. **Artificial Intelligence Review**, [S.l.], v.31, n.1-4, p.61–85, 2009.

LEFFA, V. J. Nem tudo o que balança cai: objetos de aprendizagem no ensino de línguas. **Polifonia**, [S.l.], v.12, n.12 (2), 2006.

MACHADO, C.; FARIAS, M. A. Banco Internacional de Objetos Educacionais. **Educação, Formação & Tecnologias-ISSN 1646-933X**, [S.l.], v.5, n.1, p.89–91, 2012.

MAHMOOD, T.; RICCI, F. Improving Recommender Systems with Adaptive Conversational Strategies. In: ECI CONFERENCE ON E-TECHNOLOGIES IN ENGINEERING EDUCATION, 2009, New York, NY, USA. **Anais...** ACM, 2009. p.73–82. (HT '09).

MCGREAL, R. Learning objects: A practical definition. **International Journal of Instructional Technology and Distance Learning (IJITDL)**, [S.l.], v.9, n.1, 2004.

MILLONAS, M. M. Swarms, phase transitions, and collective intelligence. In: SANTA FE INSTITUTE STUDIES IN THE SCIENCES OF COMPLEXITY-PROCEEDINGS VOLUME-, 1994. **Anais...** [S.l.: s.n.], 1994. v.17, p.417–417.

MONTES-GARCÍA, A.; ÁLVAREZ-RODRÍGUEZ, J. M.; LABRA-GAYO, J. E.; MARTÍNEZ-MERINO, M. Towards a journalist-based news recommendation system: The Wesomender approach. **Expert Systems with Applications**, [S.l.], v.40, n.17, p.6735–6741, 2013.

OSMAN, I. H.; LAPORTE, G. Metaheuristics: A bibliography. **Annals of Operations research**, [S.l.], v.63, n.5, p.511–623, 1996.

OWEN, S.; ANIL, R.; DUNNING, T.; FRIEDMAN, E. **Mahout in action**. [S.l.]: Manning Shelter Island, 2011.

PAZZANI, M. J.; BILLSUS, D. Content-based recommendation systems. In: **The adaptive web**. [S.l.]: Springer, 2007. p.325–341.

POPESCU, A.; PENNOCK, D. M.; LAWRENCE, S. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In: SEVENTEENTH CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, 2001. **Proceedings...** [S.l.: s.n.], 2001. p.437–444.

PRABHA, V. D.; RATHIPRIYA, R. A Study on Swarm Intelligence Techniques in Recommender System. In: IJCA PROCEEDINGS ON INTERNATIONAL CONFERENCE ON RESEARCH TRENDS IN COMPUTER TECHNOLOGIES 2013, 2013. **Anais...** [S.l.: s.n.], 2013. n.4, p.32–34.

QUINN, C.; HOBBS, S. Learning objects and instruction components. **Educational Technology & Society**, [S.l.], v.3, n.2, p.13–20, 2000.

RESNICK, P.; IACOVOU, N.; SUCHAK, M.; BERGSTROM, P.; RIEDL, J. GroupLens: an open architecture for collaborative filtering of netnews. In: 1994. **Anais...** [S.l.: s.n.], 1994. p.175–186.

RESNICK, P.; VARIAN, H. R. Recommender Systems. **Commun. ACM**, New York, NY, USA, v.40, n.3, p.56–58, mar 1997.

RICCI, F.; ROKACH, L.; SHAPIRA, B. Introduction to recommender systems handbook. In: **Recommender systems handbook**. [S.l.]: Springer, 2011. p.1–35.

RICCI, F.; ROKACH, L.; SHAPIRA, B.; KANTOR, P. B. **Recommender Systems Handbook**. 1st.ed. New York, NY, USA: Springer-Verlag New York, Inc., 2010.

ROBINSON, G. E. Regulation of division of labor in insect societies. **Annual review of entomology**, [S.l.], v.37, n.1, p.637–665, 1992.

SCHAFER, J. B.; FRANKOWSKI, D.; HERLOCKER, J.; SEN, S. Collaborative filtering recommender systems. In: **The adaptive web**. [S.l.]: Springer, 2007. p.291–324.

SCHEIN, A. I.; POPESCUL, A.; UNGAR, L. H.; PENNOCK, D. M. Methods and metrics for cold-start recommendations. In: ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 25., 2002. **Proceedings...** [S.l.: s.n.], 2002. p.253–260.

SERAPIÃO, A. B. d. S. Fundamentos de otimização por inteligência de enxames: uma visão geral. **Sba: Controle & Automação Sociedade Brasileira de Automatica**, [S.l.], v.20, n.3, p.271–304, 2009.

SHARDANAND, U.; MAES, P. Social information filtering: algorithms for automating "word of mouth". In: OF THE 25TH ANNUAL INTERNATIONAL , 1995. **Anais...** [S.l.: s.n.], 1995. p.210–217.

SILVA, E. L. da; CAFÉ, L.; CATAPAN, A. H. Os objetos educacionais, os metadados e os repositórios na sociedade da informação. **Ciência da Informação**, [S.l.], v.39, n.3, 2011.

SOARES, D. d. A. Objetos de aprendizagem eo ensino de língua inglesa para fins específicos a distância–. **Anais do 7º Encontro de Educação e Tecnologias de Informação e Comunicação-Universidade Estácio de Sá, Setembro**, [S.l.], p.1–16, 2009.

SOBECKI, J.; TOMCZAK, J. M. Student courses recommendation using ant colony optimization. In: ASIAN CONFERENCE ON INTELLIGENT INFORMATION AND DATABASE SYSTEMS, 2010. **Anais...** [S.l.: s.n.], 2010. p.124–133.

SOBOROFF, I.; NICHOLAS, C. Combining content and collaboration in text filtering. In: IJCAI, 1999. **Proceedings...** [S.l.: s.n.], 1999. v.99, p.86–91.

SOSTERIC, M.; HESEMEIER, S. When is a Learning Object not an Object: A first step towards a theory of learning objects. **The International Review of Research in Open and Distance Learning**, [S.l.], v.3, n.2, 2002.

TIWARY, C. **Learning Apache Mahout**. [S.l.]: Packt Publishing, 2015.

TRAN, T.; COHEN, R. Hybrid recommender systems for electronic commerce. In: KNOWLEDGE-BASED ELECTRONIC MARKETS, PAPERS FROM THE AAAI WORKSHOP, TECHNICAL REPORT WS-00-04, AAAI PRESS, 2000. **Proceedings...** [S.l.: s.n.], 2000.

VETROMILLE-CASTRO, R. A usabilidade e a elaboração de materiais para o ensino de inglês mediado por computador. **Rev. Brasileira de Lingüística Aplicada**, [S.l.], v.3, n.2, p.9–23, 2003.

VETROMILLE-CASTRO, R.; MOOR, A. M.; DUARTE, G. B.; SEDREZ, N. H. Objetos de Aprendizagem de Línguas: uma proposta. **Aprendizagem de Línguas-a Presença na Ausência: CALL, Atividade e Complexidade. Uma Homenagem aos**, [S.l.], v.70, 2012.

VETROMILLE-CASTRO, R.; MOOR, A. M.; DUARTE, G. B.; SEDREZ, N. H. From Learning Objects to Language Learning Objects: Communicative Language Teaching Principles in CALL Material. **International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)**, [S.l.], v.3, n.2, p.82–96, 2013.

VETROMILLE-CASTRO R.; PEREIRA, G. J.; RAMIRES, H. R. **Ontologias pedagógicas no ensino de libras**. In: III Jornada de Elaboração de Materiais, Tecnologia e Aprendizagem de Línguas - JETAL, 2013, Pelotas. III Jornada de Elaboração de Materiais, Tecnologia e Aprendizagem de Línguas - JETAL, 2013. v. 1. p. 28-28.

VIEIRA, F. J. R.; NUNES, M. A. S. N. DICA: Sistema de Recomendação de Objetos de Aprendizagem Baseado em Conteúdo. **Scientia Plena**, [S.l.], v.8, n.5, 2012.

WANG, Y.; LIAO, X.; WU, H.; WU, J. Incremental collaborative filtering considering temporal effects. **arXiv preprint arXiv:1203.5415**, [S.l.], 2012.

WILEY, D. A. **Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy**. [S.l.: s.n.], 2003.

ZAINA, L. Uma abordagem para recomendação de objetos de aprendizagem em ambientes educacionais. **Revista de Computação e Tecnologia (ReCeT). ISSN 2176-7998**, [S.l.], v.2, n.1, p.pg–23, 2010.

ZAINA, L.; BRESSAN, G.; CARDIERI, M.; JÚNIOR, J. F. R. e-lors: Uma abordagem para recomendacao de objetos de aprendizagem. **Revista Brasileira de Informática na Educação**, [S.l.], v.20, n.1, p.04, 2012.

ZHOU, D.; ZHU, S.; YU, K.; SONG, X.; TSENG, B. L.; ZHA, H.; GILES, C. L. Learning multiple graphs for document recommendations. In: WORLD WIDE WEB, 17., 2008. **Proceedings...** [S.l.: s.n.], 2008. p.141–150.

ZHU, X.; GOLDBERG, A. B. Introduction to semi-supervised learning. **Synthesis lectures on artificial intelligence and machine learning**, [S.l.], v.3, n.1, p.1–130, 2009.

ZHUHADAR, L.; NASRAOUI, O.; WYATT, R.; ROMERO, E. Multi-model ontology-based hybrid recommender system in e-learning domain. In: IEEE/WIC/ACM INTERNATIONAL JOINT CONFERENCE ON WEB INTELLIGENCE AND INTELLIGENT AGENT TECHNOLOGY-VOLUME 03, 2009., 2009. **Proceedings...** [S.l.: s.n.], 2009. p.91–95.

ANEXO A LISTA DE OBJETOS ACESSADOS

Tabela 15: Lista de objetos acessados

Código	Título	Objetivo
4	What would you bring on a desert island	The audio aims to work on listening comprehension and practice of vocabulary related to hypothetical situations in English. It also emphasizes the use of structures with 'would' and 'could'
27	Dance dance doll of mine!	This resource is aimed at developing students reading and comprehension skills
31	The Grimms Fairy Tales 11 - Willow wren bear	To practice the trivial English students skills such as listening text comprehension and oral skills. The audio also helps to improve the students vocabulary and to increase the knowledge in literature
36	The Grimms Fairy Tales 52 - Twelve huntsmen	To practice the trivial English students skills such as listening text comprehension and oral skills. The audio also helps to improve the students vocabulary and to increase the knowledge in literature
47	The Grimms Fairy Tales 20 - The mouse the bird and the sausage	To practice the trivial English students skills such as listening text comprehension and oral skills. The audio also helps to improve the students vocabulary and to increase the knowledge in literature
52	Recycling	To work on listening as one of the four skills in English language and to get familiar to the culture of the countries in which English is spoken
58	Kevin: Football and me	To work on listening as one of the four skills in English language and to get familiar to the culture of the countries in which English is spoken
67	French food from a Trinidadian point of view	To work on listening as one of the four skills in English language and to get familiar to the culture of the countries in which English is spoken
79	The cat and the cock	To improve students' vocabulary and to help teachers to initiate discussions in the classroom about the moral of the story
88	Carla: Traditional food in Trinidad	To work on listening as one of the four skills in English language and to get familiar to the culture of the countries in which English is spoken.
99	Spanish Lesson - presente tense: regular -ir verbs	Demonstrar una explicación de los verbos de tercera conjugación en el presente de indicativo
105	The swollen fox	Improve students vocabulary and help teachers to initiate discussions in classe about the moral of the story
110	My family	To work on listening as one of the four abilities in the foreign language learning
118	Short poetry collection 004 - When I have fears that I may cease to be	The main goal of this collection is to practice the trivial English students skills such as listening text comprehension and oral skills. The audio also helps to improve the students vocabulary and to increase the knowledge in literature
124	My yearbook	The audio aims to work on listening comprehension and practice of vocabulary related to past experiences and description of an yearbook
125	What English think about the French	The audio aims to work on listening comprehension and practice of vocabulary related to cultural aspects. It provides a good opportunity to discuss about stereotypes and multicultural themes

Continua na página seguinte

Tabela 15 – Continuação da página anterior

Código	Título	Objetivo
126	University ring	The audio aims to work on listening comprehension and practice of vocabulary related to cultural aspects in the US
155	Typical breakfast in Great Britain	The audio aims to work on listening comprehension and practice of vocabulary related to food habits
156	My pets	To work on listening as one of the four abilities in the foreign language learning
157	What french think about the english	The audio aims to work on listening comprehension and practice of vocabulary related to cultural aspects. It provides a good opportunity to discuss about stereotypes and multicultural themes
175	Siobhan: Christmas with my family	To work on listening as one of the four abilities in the foreign language learning
177	High school in the USA	The audio aims to work on listening comprehension with a real situation. It provides an adequate opportunity to improve english language skills and to allow students to get familiar with the culture of the countries in which english is spoken
215	Alice's adventures in wonderland 05 - Advice from a caterpillar	To practice the trivial English students skills such as listening text comprehension and oral skills. The audio also helps to improve the students vocabulary and to increase the knowledge in literature
222	Venezuela: parte 3 [Habla América]	Aprender la lengua española conociendo la cultura de Venezuela
230	The indiscreet letter - chapter 2	The goal of the audio is show a new perspective of literature. The spoken literature can draw the attention of the students and the teacher. It doesn't matter is the audio is used in an English class for students who are learning English as a second language or at university where professors and students can deal better with the language and try to see the literature with "ears"
234	Dame Wonder's picture alphabet	Improve the vocabulary and assimilate the english alphabet
244	Venus and the cat	To improve students' vocabulary and to help teachers to initiate discussions in class about the moral of the story
268	The Cat and the Mice	Improve the vocabulary and the spoken English of the students and start some discussions about the story
269	The goods and the ills	Improve students vocabulary and help teachers to initiate discussions in classe about the moral of the story
320	The illustrated alphabet of birds	Improve vocabulary and assimilate the English alphabet
321	The anti-slavery alphabet	The goal of the audio is combine the slaving history describing slavers lives and culture and also shows the application of the initial letters of the alphabet in words that are related with the theme
322	The Peter Pan alphabet	The goal of the text is to combine a Peter Pan's story but in other situation describing the characters and make the associations with the letters of the alphabet
325	Uncle Frank's series: funny alphabet	With this media the student can improve the vocabulary and assimilate the english alphabet with the words in english too
326	Little people - an alphabet	Teach the alphabet in a different way making associations between the names of nationalities and the letters of the alphabet
327	An alphabet of old friends	With this media the student can improve the vocabulary and assimilate the english alphabet with the words in english too
328	The absurd ABC	Improve the vocabulary and assimilate the english alphabet
337	Easter	To work on listening as one of the four skills in English language and to get familiar to the culture of the countries in which English is spoken
339	I am from Oklahoma	To work on listening as one of the four skills in English language and to get familiar to the culture of the countries in which English is spoken
350	The quack frog	Help teachers and students since each short story brings a moral which can be discussed in the classroom and contribute to the improvement of spoken english for students

Continua na página seguinte

Tabela 15 – Continuação da página anterior

Código	Título	Objetivo
351	The mouse the frog and the hawk	Help teachers and students since each short story brings a moral which can be discussed in the classroom and contribute to the improvement of spoken english for students
352	The eagle the cat and the wild sow	Improve students vocabulary and help teachers to initiate discussions in classroom about the moral of the story
354	What is a stereotype	The audio aims to work on listening comprehension and practice of vocabulary related to cultural aspects. It provides a good opportunity to discuss about stereotypes and multicultural themes
355	What English people think of the French	The audio aims to work on listening comprehension and practice of vocabulary related to cultural aspects. It provides a good opportunity to discuss about stereotypes and multicultural themes
356	What I eat for breakfast lunch and dinner	The audio aims to work on listening comprehension and practice of vocabulary related to food preferences and habits
357	San Francisco	The audio aims to work on listening comprehension with a real situation. It provides an adequate opportunity to improve english language skills and to allow students to get familiar with the culture of the countries in which english is spoken
359	A painting	To work on listening as one of the four abilities in the foreign language learning
361	If I were (pre-intermediate)	To work on listening as one of the four skills in english language and to get familiar with the culture of the countries in which english is spoken
362	If I were... (elementary)	The audio aims to work on listening comprehension and structure of imaginary situations with 'if I were' and 'would'. Besides that some cultural aspects are presented in order to provide not only examples of use of english but contextualized situations
363	If I were...	The audio aims to work on listening comprehension and practice of vocabulary related to description of personal preferences
364	If I were...	The audio aims to work on listening comprehension and structure of imaginary situations. Another goal is to broaden student's knowledge on vocabulary about seasons numbers and some adjectives
369	The Grimms Fairy Tales 28 - Little peasant	To practice the trivial English students skills such as listening text comprehension and oral skills. The audio also helps to improve the students vocabulary and to increase the knowledge in literature
377	The Grimms Fairy Tales 38 - Queen bee	To practice the trivial English students skills such as listening text comprehension and oral skills. The audio also helps to improve the students vocabulary and to increase the knowledge in literature
387	The Grimms Fairy Tales 25 - Rumpelstiltskin	To practice the trivial English students skills such as listening text comprehension and oral skills. The audio also helps to improve the students vocabulary and to increase the knowledge in literature
402	Estoy en contra de la corrida	Además de la comprensión oral obtener una competencia cultural relacionada con el comportamiento juvenil y relación con los animales
439	Stereotypes about English people	The audio aims to work on listening comprehension and practice of vocabulary related to cultural aspects. It provides a good opportunity to discuss about stereotypes and multicultural themes
440	My Chinese portrait (1)	The audio aims to work on listening comprehension and practice of vocabulary related to cultural aspects. It provides a good opportunity to discuss about stereotypes and multicultural themes
441	Food in New York	The audio aims to work on listening comprehension with a real situation. It provides an adequate opportunity to improve English language skills and to allow students to get familiar with the culture of the countries in which English is spoken
453	Stereotypes about French people	The audio aims to work on listening comprehension and practice of vocabulary related to cultural aspects. It provides a good opportunity to discuss about stereotypes and multicultural themes
454	American versus british	To work on listening as one of the four skills in english language and to work on the culture of the countries in which english is spoken

Continua na página seguinte

Tabela 15 – Continuação da página anterior

Código	Título	Objetivo
455	English versus irish pronunciation	To work on listening as one of the four skills in english language and to get familiar to the culture of the countries in which english is spoken
457	Johnny Depp is my favourite actor	The audio aims to work on listening comprehension with a real situation. It provides an adequate opportunity to improve english language skills and to allow students to get familiar with the culture of the countries in which english is spoken
460	My studies	To work on listening as one of the four abilities in the foreign language learning
463	Ghosts	To work on listening as one of the four skills in english language and to get familiar to the culture of the countries in which english is spoken
464	Edinburgh	The audio aims to work on listening comprehension with a real situation. It provides an adequate opportunity to improve english language skills and to allow students to get familiar with the culture of the countries in which english is spoken
491	Holly: Sci-Fi in the UK	To work on listening as one of the four skills in English language and to get familiar to the culture of the countries in which English is spoken.
493	Food	To work on listening as one of the four skills in english language and to get familiar to the culture of the countries in which english is spoken
494	Independence Day	The audio aims to work on listening comprehension with a real situation. It provides an adequate opportunity to improve English language skills and to allow students to get familiar with the culture of the countries in which English is spoken
497	American pronunciation	To work on listening as one of the four skills in english language and the culture of the countries in which english is spoken
499	A few words about me	To work on listening as one of the four skills in english language learning and to get familiar to the culture of the countries in which english is spoken
506	The Second World War - part 2	To work on listening as one of the four skills in english language and to get familiar to the culture and history of the United States
535	David Beckham is my favourite celebrity	The audio aims to work on listening comprehension with a real situation. It provides an adequate opportunity to improve English language skills and to allow students to get familiar with the culture of the countries in which English is spoken
539	Food!	To work on listening as one of the four skills in English language and to get familiar to the culture of the countries in which English is spoken
555	Mascots and colors in US schools	The audio aims to work on listening as one of the four skills in english language and to get familiar with the culture of the countries in which english is spoken
556	Manchester	The audio aims to work on listening as one of the four skills in English language and to get familiar with the culture of the countries in which English is spoken
573	The ass in the lion's skin	Help teachers and students since each short story brings a moral which can be discussed in the classroom and contribute to the improvement of spoken english for students
589	A typical school day in the US	The audio aims to work on listening comprehension and practice of vocabulary related to description of schedule in school
591	Spanish lesson - present tense - ar verbs: nosotros	Además de la comprensión oral desarrollar en los estudiantes la idea de la formación de la primera persona del plural en el presente de indicativo
592	Spanish lesson - present tense - ar verbs: él/ella/usted	Además de la comprensión oral presentar la formación estructural de algunos verbos en tercera persona del singular del presente de indicativo
594	Spanish lesson - present tense - ar verbs: yo form	Presentar a los estudiantes la formación del presente de indicativo en primera persona
597	Stereotypes of First Nations people	To work on listening as one of the four skills in English language and to get familiar to the culture of the countries in which English is spoken
602	Travelling	To work on listening as one of the four skills in english language and to get familiar to the culture of the countries in which english is spoken

Continua na página seguinte

Tabela 15 – *Continuação da página anterior*

Código	Título	Objetivo
603	Covent Garden	To work on listening as one of the four skills in english language and to get familiar to the culture of the countries in which english is spoken
646	Short Poetry Collection 002 - A childs nightmare	The main goal of this colection is to practice the trivial English students skills such as listening text comprehension and oral skills. The audio also helps to improve the students vocabulary and to increase the knowledge in literature
659	Los derechos del animal (2)	Además de desarrollar la capacidad de la comprensión oral obtener una competencia cultural relacionada con el uso de la lengua y ponerla como un medio de acceso a los diversos tipos de manifestaciones del lenguaje
660	Los derechos del animal (1)	Además de desarrollar la capacidad de la comprensión oral obtener una competencia cultural relacionada con el uso de la lengua y ponerla como un medio de acceso a los diversos tipos de manifestaciones del lenguaje
661	My Chinese portrait (2)	The audio aims to work on listening comprehension and practice of vocabulary related to cultural aspects. It provides a good opportunity to discuss about stereotypes and multicultural themes

ANEXO B RECOMENDAÇÕES REALIZADAS A PARTIR DA INTERAÇÃO DOS USUÁRIOS I, J E K

Tabela 16: Recomendações realizadas a partir da interação dos usuários I, J e K

Objeto selecionado	Objeto recomendado	Valor
155	356	4.059700848181420
155	439	0.010284217713809999
177	555	5.623404391201620
177	506	0.02181577381956
177	364	0.00130210145424
354	355	0.00355263866103
356	155	10.881962459841000
356	441	0.060794910859199996
356	67	0.021510696092890004
555	177	0.00215010750306
660	402	0.00315378606692
79	269	0.010697033227680001
454	497	0.30882120321477
454	455	0.00594061292163
455	454	0.2723010109470101
493	539	0.029526063187519998
539	493	0.12881065481581
539	356	0.00432487546444
354	597	0.00195978342842
439	354	0.046824165543670006
497	454	0.7381423323402302
660	659	0.023005167238