



Anderson Roque do Amaral

SISTEMA DE RECOMENDAÇÃO DE *Tags* APLICADO NA  
CATALOGAÇÃO DE RECURSOS DE APRENDIZAGEM

Sorocaba  
2014

Universidade Federal de São Carlos  
Programa de Pós-Graduação em Ciência da Computação (PPGCCS)

Anderson Roque do Amaral

SISTEMA DE RECOMENDAÇÃO DE *Tags* APLICADO NA CATALOGAÇÃO DE RECURSOS DE  
APRENDIZAGEM

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação (PPGCCS) da Universidade Federal de São Carlos como parte dos requisitos exigidos para a obtenção do título de Mestre em Ciência da Computação. Área de concentração: Engenharia de Software e Gestão do Conhecimento.

Orientadora: Prof<sup>a</sup> Dr<sup>a</sup> Luciana A. M. Zaina

Este exemplar corresponde à versão final da dissertação defendida pelo aluno, e orientada pela Prof<sup>a</sup> Dr<sup>a</sup> Luciana Aparecida Martinez Zaina

---

Sorocaba  
2014

DEDICO ESTE TRABALHO AOS  
MEUS PAIS ANTONIO PAULO DO  
AMARAL E LUCIA MARIA DO  
AMARAL QUE DESDE SEMPRE E  
ANTES DE MIM, ACREDITARAM QUE  
EU PODERIA COMEÇAR E CONTI-  
NUAR CRESCENDO COMO PESSOA,  
PROFISSIONAL E FILHO.

# Agradecimentos

Agradeço,

em primeiro lugar, a Deus por me conceder essa grande oportunidade de continuar aprendendo e estudando na área que um dia escolhi. Obrigado Senhor! Aliás nessa experiência, aprendi algo que sabia apenas por palavras. Que o Senhor escolhe para mim o que eu mesmo escolheria se visse tudo que Tu vêes. Sou feliz por todas Tuas escolhas em minha vida e por me cercar de pessoas especiais que não cessarei de agradecer.

...aos meus pais Antonio Paulo e Lucia , pelo apoio e incentivo incondicionais que me dedicaram tanto afetivamente como efetivamente.

...a minha amiga e esposa Maria Rita, por cuidar de tudo e por me colocar de volta ao foco de meu trabalho sempre que isso foi necessário.

...à minha filha Tamiris, pelo carinho...ao meu filho Kelvin, por me ajudar a descomplicar o entendimento da Complexidade de Algoritmos. Aprendi também com você. Obrigado filho.

...à minha orientadora Profa. Dra. Luciana Zaina, pela amizade, sinceridade, competência, profissionalismo, organização, incentivo, apoio e cuidado. Suas orientações foram essenciais e estiveram presentes nessa jornada do primeiro ao último passo. Agora, já não dá mais pra calcular o quanto aprendi com você. Muito obrigado.

...ao aluno de graduação Renan Lordello que em seu projeto de iniciação científica desenvolveu a ferramenta TagLink utilizada na realização dos experimentos. Muito obrigado Renan. ...ao Prof. Dr. José Fernando Rodrigues Jr., ICMC-USP, pelas contribuições durante o processo de validação da proposta.

...aos professores da UFSCar: Prof. Dr. Alexandre Alvaro, Profa Dra. Katti Faceli e Profa. Dra. Tiemi Christine Sakata, pelas disciplinas ministradas.

...aos membros da banca examinadora da qualificação, Prof. Dr. Tiago Almeida e Profa. Dra. Sahudy Montenegro, pelos comentários, sugestões e contribuições, que ajudaram a melhorar a qualidade e a redação final do texto.

...aos meus colegas de trabalho, Neusa, Luiz Carlos, Luiz Flávio, Túlio e Alex, pela amizade e permanente torcida por mim;

...à ETEC Fernando Prestes na pessoa do diretor Prof. Paulo Sérgio Germano pela amizade, por todo apoio e por disponibilizar a escola para a realização dos experimentos que validaram este trabalho; à Profa. Acidália Campos Moretti, coordenadora pedagógica da ETEC, que me ajudou muito com dicas de escrita e normas ABNT.

...aos meus queridos alunos, que participaram ativamente de todas as etapas de meu projeto, colaborando com todo o processo de validação da proposta deste trabalho.

Enfim, a todos que direta ou indiretamente me ajudaram a concluir este trabalho. Meus sinceros agradecimentos. Deus os abençoe.

Sem sonhos, a vida não tem brilho. Sem metas, os sonhos não têm alicerces. Sem prioridades, os sonhos não se tornam reais. Sonhe, trace metas, estabeleça prioridades e corra riscos para executar seus sonhos. Melhor é errar por tentar do que errar por omitir!

Augusto Cury

# Resumo

Aspectos da marcação social têm sido estudados nos últimos anos com o objetivo de abordar conceitos, aplicabilidade e técnicas relacionadas à geração e gestão de *tags* como elementos-chave para a pesquisa de recursos e classificação. A marcação social tem sido reconhecida como uma importante alternativa para a descrição dos recursos disponíveis na *Web*. No que diz respeito sobre o rápido crescimento dos recursos e suas informações na *Web*, tornou-se cada vez mais difícil de encontrar e organizar essa grande massa de dados por métodos tradicionais com base na classificação em diretórios. O processo de marcação social tem sido investigado para melhorar a catalogação de recursos, sendo que estratégias como algoritmos de recomendação de *tags*, técnicas de visualização e geração automática de *tags*, entre outras foram exploradas para esta finalidade. Dentro do contexto de *e-learning*, a marcação social pode ajudar na construção de metadados de objetos de aprendizagem. Assim, todas as vantagens associadas a este modelo de organização de recursos, surgiu como solução para sistemas de *e-learning*. A principal contribuição de um ambiente de ensino-aprendizagem é a criação, armazenamento e manutenção do catálogo de recursos de aprendizagem que pode ser verificado e atualizado continuamente. Este trabalho propõe um sistema de recomendação de *tags*, melhorando o processo de catalogação de recursos de aprendizagem na *Web*. Três experimentos foram realizados com os alunos de um curso técnico de informática para validar a proposta: o primeiro examinou a atividade de marcação social sem usar o sistema de recomendação de *tags*; o segundo foi realizado para avaliar a atividade de marcação, mas incluindo o sistema de recomendação de *tags*; e o terceiro verificou a navegação dos estudantes a partir dos recursos marcados e criados a partir do primeiro e do segundo experimentos, a fim de verificar que o apoio da recomendação foi eficaz no vocabulário descritor dos recursos.

Palavras-chave: folksonomia, marcação social, marcação colaborativa, objetos de aprendizagem

# Abstract

Aspects of social tagging have been studied in the last years aiming at addressing concepts, applicability and techniques related to the generation and management of tags as key elements for the resource research and classification. Social tagging has been recognized as an important alternative for the description of resources available on the Web. Concerning on the rapid growth of resources and their information on the Web, it has become increasingly difficult to find and organize such massive data by traditional methods as based on directories classification. The process of social tagging has been investigated to improve the cataloging of resources, being recommendation algorithms of tags, visualization techniques and automatic generation of tags, among others have been explored for this purpose. Within the e-learning context, the social tagging may assist on building of metadata of learning objects. Thus, all the advantages linked to this model of resources' organization, has emerged as solutions for e-learning systems. The main contribution of a teaching-learning environment is the creation, storing and maintenance of the catalog of learning resources that can be checked and updated continuously. This work proposes a tags recommendation system, improving the process of cataloguing of learning resources on the web. Three experiments were carried out with technician students of a informatics course to validate the proposal: the first one examined the social tagging activity without using the tag recommendation system; the second one was conducted to evaluate the tagging activity, but including the tags recommendation system; and the third one verified the students navigation from the tagged resources created from the first and the second experiment, in order to check that the support of the recommendation was effective in the descriptor vocabulary of the resource.

Key-words: folksonomy, social tagging, collaborative tagging, learning objects

# Lista de Figuras

1.1	Visão Geral do Trabalho desenvolvido . . . . .	4
2.1	Bases pesquisadas . . . . .	8
2.2	Técnica Elastic Tag Maps (fonte:M.A. Thesis) . . . . .	12
2.3	Técnica Nuvem de Tags( <i>Tag cloud</i> fonte:Dalton State) . . . . .	13
2.4	Técnica de visualização 6PLi (fonte:International Journal Design) . . . . .	14
2.5	Jaccard nas formas simétrica e assimétrica para medir co-ocorrência . . . . .	15
3.1	Ciclo de Recomendação e-TagReS . . . . .	24
3.2	Página de Pesquisa do TagLink . . . . .	35
3.3	Página com lista de recursos encontrados pelo TagLink . . . . .	36
3.4	Página para atribuição de <i>tags</i> do TagLink . . . . .	37
3.5	Página para cadastro do projeto do professor pelo TagLink . . . . .	37
4.1	Os elementos que constituem o tripé da marcação social concebida para o Experimento 1: Usuários (estudantes), recursos de aprendizagem e <i>tags</i> . . . . .	41
4.2	Distribuição do número de alunos por número de <i>tags</i> criadas. . . . .	44
4.3	Distribuição do número de alunos por número de <i>tags</i> reutilizadas. . . . .	45
4.4	Distribuição do número de novas etiquetas criadas ao longo do tempo. . . . .	47
4.5	Distribuição do número de vezes que cada <i>tag</i> foi usada (50 mais usadas) . . . . .	48
4.6	Distribuição do número de alunos por número de <i>tags</i> criadas. . . . .	52
4.7	Distribuição do número de alunos por número de <i>tags</i> reutilizadas. . . . .	53
4.8	Distribuição do número de vezes que cada <i>tag</i> foi utilizada . . . . .	54
4.9	TagLink - Globo com nuvem de <i>tags</i> . . . . .	56
4.10	TagLink - Página de Navegação: Nuvem de <i>tags</i> e Recursos relacionados . . . . .	57
4.11	Quantidade de Recursos encontrados em cada nível da nuvem de <i>tags</i> . . . . .	59



# Lista de Tabelas

2.1	Classificação dos Trabalhos Selecionados . . . . .	9
2.2	Algoritmo de Recomendação de <i>tags</i> (XU et al., 2006) . . . . .	20
3.1	Variação 1 - Aplicar ou não Jaccard na pontuação inicial . . . . .	27
3.2	Variação 2 - Usar ou não filtragem das <i>tags</i> co-ocorrentes . . . . .	27
3.3	Variação 3 - Utilizar filtragem de <i>tags</i> (Variação 1) e aplicar Jaccard ou não . . . . .	28
3.4	Variação 4 - Determinar mínimo de objetos compartilhados usando ou não a Variação 3 . . . . .	29
3.5	Variação 5 - Usar a segunda filtragem e Jaccard ou não . . . . .	29
3.6	Variação 6 - Usar ou não Jaccard levando em conta a nova definição de $T_{CO}$ . . . . .	30
3.7	Variação 7 - Usar a filtragem de <i>tags</i> co-ocorrentes, nova definição de $T_{CO}$ . . . . .	31
3.8	Variação 8 - Usar ou não a filtragem nas <i>tags</i> co-ocorrentes e Jaccard na pontu- ação inicial, nova definição de $T_{CO}$ . . . . .	31
3.9	Variação 9 - usar ou não a segunda versão da filtragem, nova definição de $T_{CO}$ . . . . .	32
3.10	Variação 10 - Usar ou não a segunda versão da filtragem e Jaccard na pontuação inicial, nova definição de $T_{CO}$ . . . . .	32
3.11	Algoritmo Adotado pelo Módulo de Recomendação . . . . .	34
4.1	Ficha Técnica do Experimento 1 . . . . .	40
4.2	Dúvidas geradas durante a condução do experimento . . . . .	43
4.3	Sumarização da Coleta de Tags . . . . .	45
4.4	<i>Tags</i> reutilizadas por cada grupo em cada fase do experimento . . . . .	46
4.5	Ficha Técnica do Experimento 2 . . . . .	48
4.6	Resumo do Experimento 2 . . . . .	51
4.7	Resumo do Experimento 2 por Grupo . . . . .	51
4.8	Deslocamento de posicionamento das <i>Tags</i> entre os experimentos . . . . .	55
4.9	Recomendações aceitas por <i>Ranking</i> . . . . .	55
4.10	Ficha Técnica do Experimento 3 . . . . .	56
4.11	Nível da Nuvem de tags por recursos encontrados . . . . .	59

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	2
1.2	Objetivo . . . . .	3
1.3	Metodologia . . . . .	3
<b>2</b>	<b>Fundamentação Teórica</b>	<b>6</b>
2.1	Introdução . . . . .	6
2.2	Metodologia da pesquisa bibliográfica . . . . .	7
2.2.1	Trabalhos Seleccionados . . . . .	9
2.3	Estado da Arte . . . . .	9
2.3.1	Visualização de <i>tags</i> . . . . .	9
2.3.2	Recomendação de <i>tags</i> . . . . .	12
2.3.3	Objetos de aprendizagem . . . . .	16
2.3.4	Geração de <i>tags</i> . . . . .	17
2.4	Algoritmos de Recomendação por Co-ocorrência . . . . .	19
2.5	Considerações Finais . . . . .	21
<b>3</b>	<b>Sistema de Recomendação de <i>Tags</i></b>	<b>23</b>
3.1	e-TagReS: Sistema de Recomendação de <i>Tags</i> Proposto . . . . .	23
3.2	Estudo do Algoritmo de Recomendação e adaptações . . . . .	25
3.2.1	Testes das variações das métricas . . . . .	25
3.2.2	Modificações para objetos pouco marcados pelo usuário . . . . .	29
3.2.3	Algoritmo de Recomendação Adotado . . . . .	33
3.2.4	TagLink: Protótipo Funcional . . . . .	35
3.3	Considerações Finais . . . . .	38
<b>4</b>	<b>Experimentação Preliminar e Validação da Proposta</b>	<b>39</b>
4.1	Experimento 1: Catalogação de Recursos sem Recomendação . . . . .	40
4.1.1	Planejamento . . . . .	41
4.1.2	Execução . . . . .	42

---

4.1.3	Análise dos resultados . . . . .	43
4.2	Experimento 2: Catalogação de Recursos com Recomendação de Tags . . . . .	48
4.2.1	Planejamento . . . . .	49
4.2.2	Execução . . . . .	50
4.2.3	Análise dos resultados . . . . .	50
4.3	Experimento 3: Navegação e Pesquisa . . . . .	55
4.3.1	Planejamento . . . . .	57
4.3.2	Execução . . . . .	58
4.3.3	Análise dos Resultados . . . . .	58
4.4	Considerações Finais . . . . .	60
<b>5</b>	<b>Conclusões e Futuros Trabalhos</b>	<b>61</b>
5.1	Contribuições . . . . .	62
5.2	Futuros Trabalhos . . . . .	63
	<b>Bibliografia</b>	<b>65</b>

## Introdução

O crescimento no número de aplicações voltadas para a Web Social tem sido exponencial nos últimos anos (GLAVINIC; GRANIC, 2008). Essas aplicações promovem a inter-relação de pessoas e de conhecimento através de interfaces ricas que permitem uma maior interação dos usuários. Os sistemas de *e-learning* acadêmicos e comerciais têm buscado se adaptar às características dessas aplicações, tornando-se mais atraentes e próximos das aplicações usadas pelos alunos no seu dia a dia (GLAVINIC; GRANIC, 2008).

A Web Social tem como característica marcante a possibilidade de marcação de conteúdo *on-line* através do uso de *tags*, o que permite aos usuários criar vocabulários que categorizam os recursos ou objetos de aprendizagem (JASCHKE et al., 2012). Formalmente, objetos de aprendizagem referem-se a entidades utilizadas no processo de ensino-aprendizagem, tais como vídeos, imagens, *software* simulador, textos, entre outras possibilidades. No domínio de *e-learning*, é desejável que os objetos de aprendizagem sejam reutilizáveis para diferentes objetivos de aprendizagem podendo ser combinados para criar objetos mais complexos (MCGREAL, 2004) Para este fim, os objetos devem ser especificados a partir de metadados que contextualizam e descrevem a sua utilização de forma padrão (VAZQUEZ; OSTROVSKAYA, 2006).

A liberdade dos usuários, na atividade de marcação de conteúdos na *Web*, deu origem ao termo "folksonomia" (folks, de pessoas; onomia, de taxonomia - classificação) (WAL, 2005). Esse termo passou a referir-se a sistemas de marcação social de auto-evolução que emergem da interação sem regras a priori para o que é criado e marcado (DAHL; VOSSEN, 2008a). Na prática da marcação social, os usuários de forma colaborativa usam marcas (*tags*) para anotar e dar sentido ao conteúdo, sendo esta atividade uma valiosa fonte de informação com o potencial de organização (indexação e classificação ou catalogação) para grandes volumes de dados.

Sistemas de marcação social supõem que os usuários irão expressar suas impressões por meio de marcas (*tags*) que eles usam para classificar o conteúdo (MILICEVIC; NANOPOULOS; IVANOVIC, 2010). Em tais sistemas, os usuários podem se comportar seguindo padrões distintos: categorizadores - que produzem *tags* que favorecem a navegação por entre os vários recursos; ou descritores - cujas *tags* são mais voltadas para a descrição dos recursos (KORNER, 2009a). O tipo de comportamento dos usuários impacta o processo de marcação de maneiras diferentes.

A marcação social provê mecanismos para classificação de objetos, atividade esta que seria dificultada devido ao grande volume de informações disponível na *Web*, mesmo quando realizada

por especialistas de um domínio de conhecimento específico (GUPTA et al., 2010).

As principais características da folksonomia (marcação social) são: flexibilidade, já que os usuários utilizam seu vocabulário de maneira dinâmica e cotidiana; identificação personalizada, pois os usuários escolhem espontaneamente as palavras que melhor descrevem o conteúdo; e colaboração, sendo prevista para aplicações *Web Sociais*.

No contexto do *e-learning*, a marcação social pode suportar a catalogação de objetos de aprendizagem com base em *tags* fornecidas quer por alunos como por professores. No caso dos estudantes, o processo de marcação diz respeito a uma experiência de reflexão na qual os alunos marcam os objetos com base em sua própria experiência (BATEMAN; BROOKS; MCCALLA, 2007). Em seguida, os repositórios de objetos de aprendizagem podem ser pesquisados através das *tags* do próprio estudante ou por outras pessoas envolvidas no processo de aprendizagem. No entanto, para uma boa catalogação ocorrer, é necessário dispor de um processo de marcação adequado que esteja de acordo com os objetivos de aprendizagem definidos.

Embora sistemas de marcação sejam impactados de maneiras diferentes, inclusive pelo perfil e comportamento do usuário (KORNER, 2009a), o vocabulário de *tags* deve ser adequadamente heterogêneo, a fim de descrever abrangentemente os objetos. Para isso, técnicas foram desenvolvidas a fim de auxiliar o processo de marcação. Dentre elas, destacam-se as técnicas de recomendação de *tags*, cuja principal função é sugerir ao usuário *tags* mais adequadas para o conteúdo ou recurso que está sendo analisado. Essas sugestões são geradas através de algoritmos que encontram *tags* mais relevantes para o objeto e para o usuário no momento da marcação. Sistemas de recomendação surgiram com o propósito de recomendar produtos, filmes, pessoas, etc, com base nos interesses do usuário, verificados a partir da interação dos mesmos com o sistema. A partir das necessidades dos sistemas de marcação social em atribuir *tags* para recursos, os sistemas de recomendação passaram a prover soluções voltadas à sugestão de *tags*.

## 1.1 Motivação

Estudos mostram que na dinâmica dos sistemas de marcação social, com o passar do tempo, o conjunto de *tags* que, descrevem um recurso ou objeto podem assumir características diversas, ora convergindo, ora divergindo, quanto a sua semântica (TRANT, 2009).

Na catalogação de recursos de aprendizagem, ao se aplicar marcação social, um dos desafios é alcançar a estabilização do conjunto de *tags* que melhor descreva um recurso específico. Essa estabilização favorece a pesquisa e conseqüentemente o processo de aprendizagem.

A recomendação de *tags* que sejam relevantes a um determinado recurso, pode apoiar a atividade de marcação e catalogação de recursos de aprendizagem por parte de alunos e professores, e melhorar a participação colaborativa em ambientes de *e-learning*.

Este trabalho está inserido no LERIS, Laboratório de Estudos em Redes, Inovação e Software, dentro do Departamento de Ciência da Computação da UFSCAR, campus Sorocaba. O LERIS trabalha para aproximar a pesquisa acadêmica a problemas práticos.

## 1.2 Objetivo

Este trabalho tem por objetivo propor um sistema de recomendação de *tags* para auxiliar a catalogação de recursos de aprendizagem que permite a participação de professores e alunos dentro de ambientes de *e-learning* com domínio de conhecimento específico.

Aspetos relevantes sobre a marcação social são abordados nesse trabalho. Mais especificamente, este trabalho buscou responder às seguintes questões de investigação dentro de ambientes de *e-learning*:

1. Um mecanismo de recomendação de *tags*, auxilia a catalogação de recursos de aprendizagem?
2. Com a recomendação de *tags*, qual perfil de marcação é favorecido: descritivo ou o categorizador?
3. O vocabulário em torno dos recursos de aprendizagem se torna mais específico com a utilização de um mecanismo de recomendação de *tags*?
4. O algoritmo e o modelo de ciclo de recomendação proposto nesse trabalho podem auxiliar o processo de marcação?

Não é objetivo deste trabalho explorar e propor técnicas que envolvam a recomendação de conteúdos ou mesmo a especificação de perfis de usuários.

## 1.3 Metodologia

A fim de responder às questões anteriores, este trabalho foi organizado em revisão bibliográfica e experimentos. Três experimentos foram realizados com estudantes dos cursos técnicos da ETEC Fernando Prestes que participaram do processo de catalogação de recursos de aprendizagem, produzindo uma base de recursos catalogados, o que permitiu avaliações em diversos aspectos. A análise dos resultados levantam conclusões sobre como o processo de marcação social pode ser realizado em um ambiente de ensino-aprendizagem eletrônica. Em cada experimento buscou-se observar o comportamento dos alunos quanto ao uso e reuso de recomendação de *tags* e quanto à navegação a partir das *tags*.

Assim, um modelo de recomendação de *tags* é proposto e validado em cada experimento.

A Figura 1.1 apresenta uma visão geral da metodologia usada nesta dissertação desde a concepção das ideias até os resultados alcançados.

Na Etapa 1:

- (a) O tema "Folksonomia e *e-Learning*" norteou o levantamento bibliográfico (estado da arte);
- (b) O levantamento do estado da arte foi conduzido através da metodologia de revisão sistemática;
- (c) Em seguida, deu-se início ao planejamento de uma ferramenta denominada TagLink, que nessa etapa teve a função de possibilitar a atividade de marcação e coleta de *tags*, sem o uso

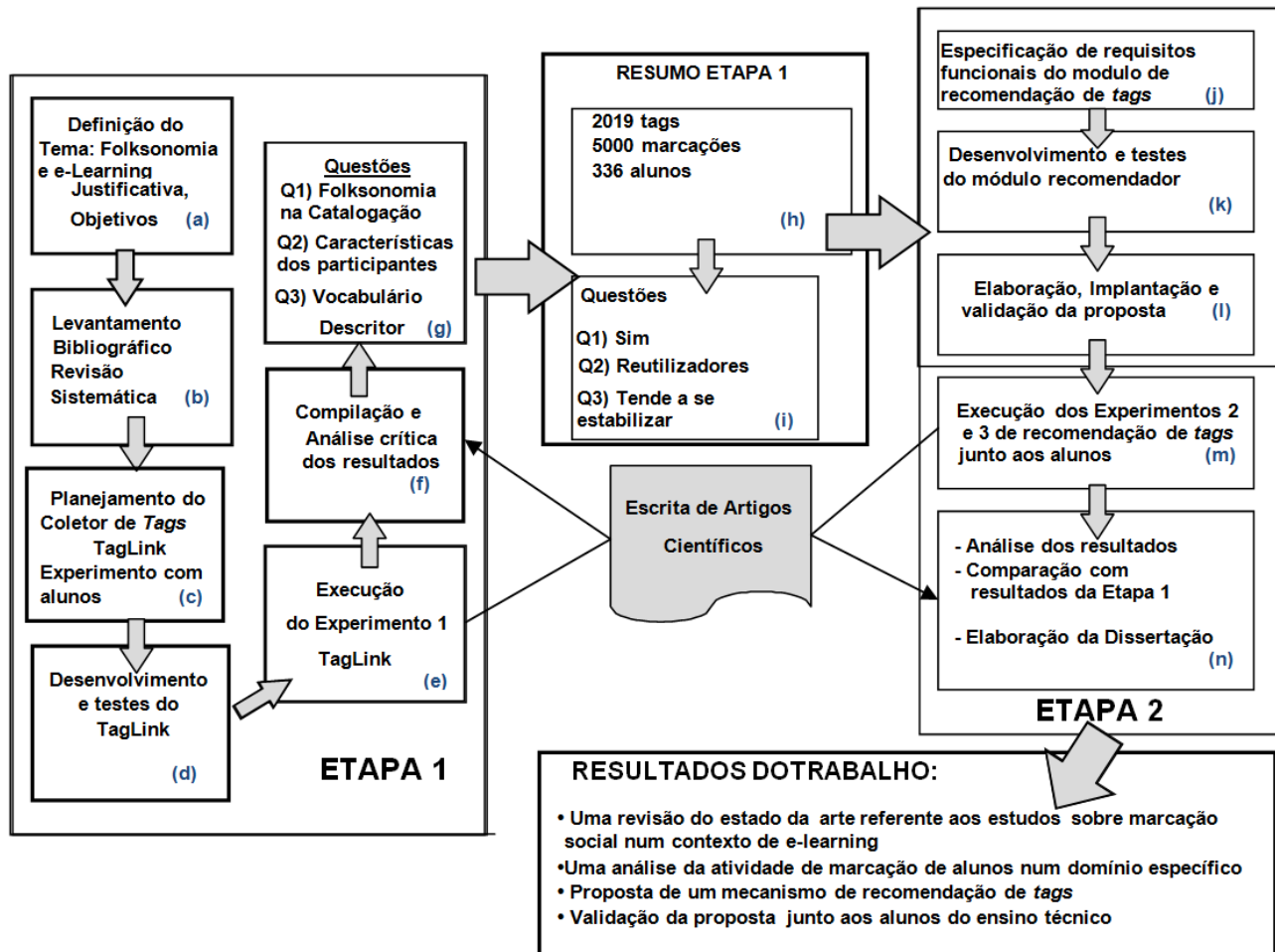


Figura 1.1: Visão Geral do Trabalho desenvolvido

de qualquer algoritmo de recomendação de *tags*; (d) Testes foram realizados com a ferramenta TagLink;

(e) O Experimento 1 foi conduzido;

(f) Os resultados foram analisados e compilados;

(g) Após a análise dos resultados, pode-se responder as seguintes questões:

- (Q1) A folksonomia pode ser aplicada na catalogação de recursos de aprendizagem? Verificou-se que a marcação social pode ser aplicada na catalogação de recursos de aprendizagem (i) pois os resultados demonstram que as *tags* utilizadas pelos alunos têm o caráter de classificar, categorizar e descrever os recursos;
- (Q2) Qual é a principal característica dos alunos participantes na atividade de marcação? Após o Experimento 1, constatou-se que os alunos participantes tem o perfil predominante de reutilizadores; (i), pois nas duas fases do primeiro experimento, foi grande o índice de reutilizações de *tags*, o que apontou para a viabilidade do desenvolvimento de um módulo de recomendação de *tags*;
- (Q3) Vocabulário descritor gerado em torno dos recursos, tende à uma estabilização (i), ou

seja, as *tags* utilizadas para descrever os recursos, com a evolução do processo, passaram a ter pouca variação por consequência do aumento das reutilizações de *tags* por parte dos alunos participantes.

**(h)** Como resultado dessa etapa, foram coletadas 2019 *tags* a partir de 5000 marcações feitas por 336 alunos .

Na Etapa 2:

**(j)** Foram definidos os requisitos funcionais do módulo de recomendação de *tags*, utilizado nos experimentos seguintes.

**(k)** O Módulo de recomendação de *tags* foi desenvolvido, intencionalmente testado e em seguida agregado ao TagLink .

**(l)** A proposta final do trabalho foi elaborada e validada junto aos alunos **(m)** A validação ocorreu através dos Experimentos 2 e 3.

**(n)** Os resultados dos experimentos conduzidos na Etapa 2, foram comparados com os resultados da Etapa 1. Finalmente, deu-se a elaboração da dissertação.

Como resultado das duas etapas, obteve-se uma revisão do estado da arte, uma análise da atividade de marcação dos alunos e a proposta do sistema de recomendação que foi validada junto aos alunos de ensino técnico.

O restante da dissertação está organizada da seguinte forma: no Capítulo 1 são apresentados os fundamentos deste trabalho e metodologia utilizada na revisão sistemática. A proposta do Sistema de Recomendação é apresentada no Capítulo 2. O Capítulo 3 detalha os estudos preliminares sem a utilização da proposta de recomendação e os experimentos realizados para a validação da proposta. Por fim, o Capítulo 4 discorre sobre as conclusões e trabalhos futuros são delineados.



## Fundamentação Teórica

### 2.1 Introdução

*e-Learning* corresponde a um modelo de ensino não presencial suportado por tecnologia pelo qual o aluno aprende através de conteúdos colocados no computador ou na *Web* e em que o professor, se existir, está à distância, utilizando a Internet como meio de comunicação (LEAL; AMARAL, 2004). Os conteúdos, normalmente, são armazenados em repositórios, isto é, coleções de informações educacionais estruturadas em um banco de dados, com o objetivo de permitir a pesquisa e o aprendizado por parte de um público específico, sejam estudantes ou não. Essas informações educacionais passaram a ser concebidas como "Objetos de Aprendizagem", termo utilizado pela primeira vez por volta de 1992 por Wayne Hodgins, que através de uma analogia com peças de "lego", observou que objetos individuais, representam um conhecimento que pode ser reutilizado, compartilhado e combinado com outros objetos e em cada nova combinação um novo conhecimento é criado (HODGINS, 2002). A partir dessa observação, Wayne propõe um novo modelo de repositório para *e-Learning*. Repositórios de objetos de aprendizagem, passaram a ser considerados como uma tecnologia poderosa e promissora, o que ocasionou um rápido crescimento de sua utilização, bem como a proposta de novos modelos de repositórios. Esse crescimento gerou a necessidade de classificação e indexação desses recursos para melhorar o processo de pesquisa das informações, cada vez mais numerosas.

A utilização de *tags* tornou-se cada vez mais frequente devido ao fato da marcação social passar a ser considerada uma excelente forma de se gerar metadados. Metadados são dados que descrevem outros dados, isto é, um item de um metadado, pode dizer do que se trata aquele dado. Dessa forma a folksonomia se apresentou com vantagens sobre a taxonomia fixa em sistemas sociais. Muitas vezes para se criar uma taxonomia fixa e hierárquica, se contratam especialistas que de certa forma conhecem o domínio em questão. Mas, mesmo assim, muitos itens não se encaixam em apenas uma categoria e desta forma, um sistema hierárquico fixo tende a se tornar rígido e inflexível não levando em conta as necessidades de classificação de cada usuário e nem seus pontos de vista diferenciados. Folksonomias fornecem um caminho mais simples, mais barato e mais natural para organizar objetos na *Web*. Sendo a folksonomia o processo de classificação feito pelas pessoas, sem regras pré-definidas e sem hierarquia, reflete diretamente o vocabulário do usuário, superando até mesmo as barreiras culturais. Por essas

razões, folksonomias são melhores que a classificação tradicional que é mais custosa e mais difícil de ser aprendida pelos usuários (GUPTA et al., 2010).

Assim, a "folksonomia" (WAL, 2005), processo de marcação social onde os usuários escolhem livremente palavras-chave para identificar recursos, se apresentou como uma solução para a deficiência encontrada no processo tradicional de classificação e organização de recursos digitais.

Técnicas de recomendação de conteúdos e *tags* foram desenvolvidas para auxiliar os usuários no processo de marcação social. Os sistemas de recomendação se utilizam das opiniões de um usuário individual ou de uma comunidade de usuários, para sugerir conteúdos e *tags*, de acordo com os interesses de pesquisa. (MILICEVIC; NANOPOULOS; IVANOVIC, 2010). Para a recomendação de *tags*, um conceito muito explorado é o da co-ocorrência. *Tags* co-ocorrentes são *tags* que foram utilizadas para a marcação de um mesmo objeto. Sendo assim, os sistemas recomendadores se utilizam de índices de co-ocorrência para sugerir *tags* que auxiliam a atividade de marcação (XU et al., 2006).

## 2.2 Metodologia da pesquisa bibliográfica

Uma Revisão Sistemática tem como objetivo produzir uma síntese completa de trabalhos publicados sobre uma questão de pesquisa específica, utilizando um processo aberto e bem definido para guiar a busca e a análise de trabalhos (KITCHENHAM, 2004). A Revisão Sistemática garante um valor científico ao trabalho. Procedimentos de Revisão Sistemática foram utilizados para a condução do levantamento bibliográfico. Para organizar e executar o estudo e identificar o estado da arte sobre o tema "marcação social" com foco em objetos de aprendizagem, foram realizadas três etapas: o planejamento; a execução e extração de artigos; e a sumarização, onde são analisados os trabalhos extraídos.

O planejamento da revisão foi motivado pelo fato de que novas perspectivas sobre as possibilidades de catalogação, indexação e busca de conteúdos para *e-Learning* na Web 2.0 tem emergido de diversas formas. Considerando este contexto de estudo delineou-se como objetivo da pesquisa, investigar soluções existentes na área *e-Learning* que envolvam marcação social e objetos de aprendizagem, identificando inclusive potenciais áreas a serem exploradas. Para realizar a investigação foram selecionadas bases científicas: IEEEXplore<sup>1</sup>, Google Scholar<sup>2</sup>, Scopus<sup>3</sup>, ACM Digital Library<sup>4</sup>, Interscience<sup>5</sup>, Citeseerx<sup>6</sup>, Periódicos Capes<sup>7</sup>.

A fim de executar a busca nas fontes de pesquisa citadas, foram definidas como ponto central as palavras chave "*Folksonomy*", "*e-learning*" e "*Learning Objects*", sendo estas procuradas nos títulos, resumos e palavras chaves dos trabalhos. Refinamentos foram realizados utilizando a palavra-chave secundária "social tagging". O termo "Folksonomy" foi escolhido por ser abrangente o suficiente permitindo que fossem encontrados um maior número possível de artigos

---

<sup>1</sup><http://ieeexplore.ieee.org>

<sup>2</sup><http://scholar.google.com>

<sup>3</sup><http://www.scopus.com/home.url>

<sup>4</sup><http://dl.acm.org>

<sup>5</sup><http://interscience.metapress.com>

<sup>6</sup><http://citeseerx.ist.psu.edu>

<sup>7</sup><http://www.periodicos.capes.gov.br>

científico dentro do tema. O termo *e-learning* foi acrescentado para buscar trabalhos que usassem marcação social sem ter o foco em objetos de aprendizagem. Isto porque, muitas vezes são apresentados experimentos com a marcação social sem abordar os recursos marcados como objetos de aprendizagem. As palavras chave também foram especificadas na língua portuguesa para a busca em fontes nacionais. Inicialmente, foram buscadas revisões sistemáticas sobre o tema, utilizando como palavras-chave as expressões "Review Framework Folksonomy" e "Review Folksonomy". Foram encontradas 6 revisões sistemáticas que foram incluídas nesse trabalho e serão abordadas adiante.

Os trabalhos que seriam incluídos no estudo deveriam: estar em inglês ou português; abordar aspectos da marcação social, folksonomia e objetos de aprendizagem; e tratar sobre técnicas e tecnologias aplicadas no contexto de *e-Learning*. Seriam excluídos artigos que não tratassem de temas relevantes à pesquisa e artigos que não possuem informações suficientes para analisar o seu contexto.

Com base no planejamento durante a execução 126 artigos foram encontrados. Dos 126 trabalhos encontrados após o refinamento da pesquisa, 91 foram excluídos por tratarem de temas não relevantes à pesquisa, como por exemplo semântica para fragmentação de vídeos, folksonomia para museus e bibliotecas, ontologias com base em folksonomia, estruturas narrativas, modelagem de conceitos médicos, etc. Além dos critérios de exclusão definidos, também foi motivo para a não seleção, artigos que estavam sem acesso devido a problemas técnicos e artigos duplicados, ou seja que são encontrados em mais de uma base. Ao final do processo, foram selecionados 35 artigos, apresentados na Figura 2.1 reportando a base em que foram encontrados.

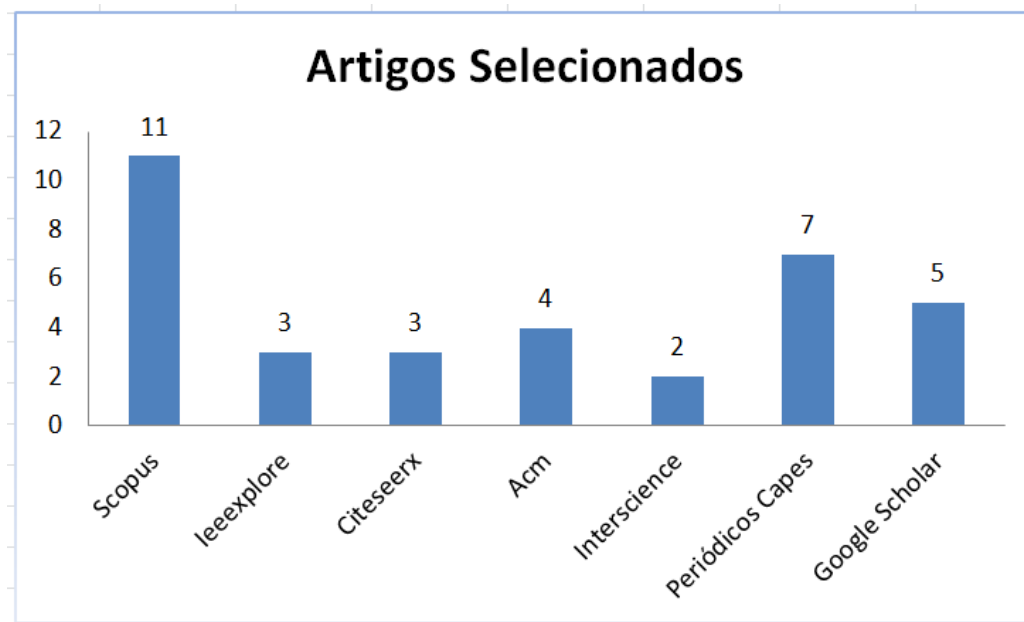


Figura 2.1: Bases pesquisadas

## 2.2.1 Trabalhos Selecionados

Os 35 trabalhos extraídos foram analisados considerando as características: visualização de *tags*, recomendação de conteúdos ou *tags*, manipulação de objetos de aprendizagem e geração de *tags*. Para isto, os trabalhos foram agrupados dentro destes tópicos, sendo que cada um pode enfatizar um ou mais de cada aspecto. A Tabela 2.1 apresenta os artigos selecionados em ordem cronológica e classificados em 4 perspectivas descritas nas colunas: (A) Visualização de *tags*, (B) Recomendação de conteúdos e/ou *tags*, (C) Objetos de aprendizagem e (D) Geração de *tags*. Este agrupamento foi realizado por refletir os temas de interesse deste trabalho. As citações que estão em **negrito** referem-se a revisões sistemáticas.

## 2.3 Estado da Arte

As subseções seguintes descrevem o estado da arte da marcação social analisando os trabalhos conforme as quatro perspectivas consideradas.

Tabela 2.1: Classificação dos Trabalhos Selecionados

Artigos	A	B	C	D	Artigos	A	B	C	D
Xu et al. (2006)		X			Sigurbjournsson e Zwol (2008)		X		
Hsieh, Lai e Chou (2006)	X	X		X	Benz et al. (2010)		X		X
<b>Bateman, Brooks e McCalla (2007)</b>			X	X	Conole e Culver (2010)	X			
Peñalvo et al. (2007)			X	X	<b>Gupta et al. (2010)</b>	X	X		X
<b>Dahl e Vossen (2008a)</b>	X		X	X	<b>Milicevic, Nanopoulos e Ivanovic (2010)</b>	X	X		X
Dahl e Vossen (2008b)			X	X	Cho et al. (2011)		X	X	
Cernea, Moral e Gayo (2008)			X	X	Wu e Bo (2011)	X	X		
Sierra e Valmayor (2008)	X	X		X	Teixeira, Sá e Fernandes (2011)	X			
Monge, Ovelar e Azpeitia (2008)		X	X		Shih e Tseng (2012)	X			X
Shih e Tseng (2008)		X	X	X	Garcia-Silva et al. (2012)		X		X
<b>Zamora e Nistal (2009)</b>		X			López, Prieta e Ogihara (2012)			X	X
Coelho (2009)	X	X	X		Strohmaier, Korner e Kern (2012)		X		
Hsieh et al. (2009)		X		X	Assis e Moura (2013)				X
<b>Trant (2009)</b>	X			X	Ribeiro, Fonseca e Freitas (2013)	X	X	X	
Golder e Huberman (2009)		X		X	Jelassi e Yahia (2013)		X		
Lops et al. (2009)		X		X	Zervas e Sampson (2014)			X	
Sinclair e Cardew-Hall (2009)	X	X			Anand e Mampilli (2014)		X		
Rigo et al. (2009)	X		X		Total	14	20	13	18

### 2.3.1 Visualização de *tags*

Num sistema de marcação social ou folksonomia, a visualização das *tags* é fundamental para apoiar a pesquisa, navegação e descoberta de informações por parte dos usuários. Estudos tem

sido desenvolvidos para aprimorar as técnicas que permitem uma visualização do vocabulário e das relações entre as *tags*. Hsieh, Lai e Chou (2006) propõe um modelo de sistema hierárquico para compartilhamento de recursos através de *tags*. Consideram a popularidade das *tags* dentro de uma *tag cloud* (nuvens de *tags*) onde as *tags* são apresentadas na ordem alfabética e escritas com fontes maiores conforme a popularidade das mesmas. Executam um experimento para coletar *tags* através de um aplicativo. Medem a precisão do modelo proposto e apresentam resultados positivos quanto ao benefício que o modelo apresentado por eles melhora a pesquisa de conteúdos.

É essencial permitir aos usuários explorar e navegar sobre a estrutura de conhecimento do domínio, tanto durante a catalogação quanto na recuperação dos recursos de informação. Rigo et al. (2009) apresentam uma análise das técnicas e ferramentas de visualização de conhecimento para amparar os usuários na escolha de termos apropriados na descrição de recursos de informação. Consideram a visualização adequada o primeiro passo para a catalogação, busca e reuso de recursos. Analisam o grau de visualização e navegação de algumas técnicas: Mapa Conceitual, Visualização Hiperbólica, Diagramas Hierárquicos, Herança Múltipla, etc. A partir deste estudo, desenvolvem um método de gerenciamento de conteúdos de um repositório.

Gupta et al. (2010) apresentam uma revisão de trabalhos que exploram o conceito de visualização. Dentre as técnicas encontradas na literatura, destaca-se a nuvem de *tags*, onde as *tags* são apresentadas como um conjunto de palavras com tamanhos, cores ou fontes diferentes conforme o peso de cada uma, medido normalmente pela frequência de uso (DAHL; VOSSEN, 2008a). Outra técnica é a visualização da evolução das *tags* onde as *tags* são apresentadas numa linha do tempo através de uma animação que permite ao usuário interagir com a apresentação e selecionar *tags* relacionadas a um ponto do tempo a partir de um controle que permite detalhar ou resumir os resultados. Também são apresentados algoritmos e ferramentas que extraem as palavras mais significativas de um documento apresentando os termos como *tags* de forma hierárquica. Assim através de um mecanismo de navegação, o usuário pode explorar a dependência entre os termos.

Além da visualização de *tags*, sistemas como o GroupMe!<sup>8</sup>, possibilitam visualização de objetos pelo tipo de mídia associado por grupos formados através de interesses comuns. Ou seja, fotos são apresentadas em miniatura, vídeos e áudios podem ser tocados por determinado grupo de usuários com mesmos interesses. Os grupos são formados a partir de inscrições feitas pelo próprio usuário, sendo que estes passam a ser notificados sobre modificações que ocorrem no grupo (MILICEVIC; NANOPOULOS; IVANOVIC, 2010). Os autores definem dois tipos de folksonomia. A folksonomia estreita (*narrow folksonomy*), onde o primeiro a atribuir uma *tag* ao objeto é o próprio criador do objeto e que impõe alguns limites, como por exemplo: um objeto não pode receber a mesma *tag* mais de uma vez, e a folksonomia larga (*broad folksonomy*), onde existe total liberdade de marcação para cada objeto. Sendo assim, a quantidade de *tags* para um mesmo objeto, na folksonomia estreita é bem menor. Nota-se que na folksonomia estreita as *tags* tem mais relação com os objetos de aprendizagem, enquanto que na folksonomia larga, as *tags* tem mais relação com os usuários (MILICEVIC; NANOPOULOS; IVANOVIC, 2010). Sendo assim, a técnica de visualização para uma folksonomia precisa levar em conta o tipo a ser

---

<sup>8</sup>[https:// groupme.com/](https://groupme.com/)

representado.

Alguns estudos que analisam a utilidade das nuvens de *tags* (*tag clouds*) na representação de folksonomias. Sinclair e Cardew-Hall (2009) afirmam que na essência, uma nuvem de *tags* traduz o vocabulário emergente de uma folksonomia em uma ferramenta de navegação social. Para explorar os benefícios e as limitações das nuvens de *tags*, conduzem um experimento. A partir de questões apresentadas aos alunos sob um determinado tema, estes deveriam encontrar as respostas a partir de uma nuvem de *tags*. Ao avaliarem os resultados, os autores concluíram que uma nuvem de *tags* é útil para a busca e descoberta de informação, pois os termos associados ao elemento provêm um resumo do conteúdo armazenado permitindo que os estudantes tenham um menor esforço cognitivo de entedimento sobre o termo.

Uma revisão das técnicas de visualização de *tags* é apresentada por Zamora e Nistal (2009). Destacam as razões pelas quais a marcação social tem potencial para ser aplicada em sistemas de *e-Learning*. Dentre elas, frisam que a marcação colaborativa pode enriquecer ainda mais as interações entre professores e alunos em torno dos objetos de aprendizagem. Por ser a marcação uma atividade de natureza reflexiva, ela pode dar aos alunos oportunidade de conceber novas ideias através da visualização das *tags* e sugestões de outros alunos. Afirmam também que a Engenharia e Tecnologia da Educação podem se beneficiar pelo fato da Web conter valiosos materiais educativos de várias disciplinas. O aspecto visual é importante condutor do aluno na atividade de marcação de objetos, pois permite que o aluno possa identificar outros objetos ou mesmo outras *tags*, facilitando a marcação dos objetos. Foram encontradas na literatura as técnicas de visualização: nuvem de *tags*; grupos de nuvem de *tags*<sup>9</sup>; *Elastic Tags Maps*<sup>10</sup> e *6Pli Interface*<sup>11</sup>. As Figuras 2.2, 2.3 e 2.4 ilustram três das técnicas de visualização citadas. (ZAMORA; NISTAL, 2009) também propõem um algoritmo de agrupamento de *tags* com o objetivo de criar uma rede de relacionamentos em uma folksonomia através da qual os usuários podem navegar por grupos de *tags* co-ocorrentes. e desenvolvem uma interface flexível, onde o usuário pode livremente mover e rearranjar as *tags* dentro das nuvens.

Por trás de cada técnica estão conceitos e algoritmos que possibilitam sua implementação. A técnica *Elastic Tag Maps* (Figura 2.2), visualiza os relacionamentos emergentes entre *tags*. O algoritmo coloca as *tags* que frequentemente são utilizadas juntas num mesmo plano 2D. Enquanto o ponteiro do mouse se desloca sobre a imagem, *tags* que tendem a co-ocorrer com a *tag* selecionada, são trazidas para frente. Clicando em uma *tag*, é possível verificar o contexto semântico da mesma.

6PLi é uma ferramenta de visualização desenvolvida para ser usada com o *del.icio.us*<sup>12</sup>. Os usuários podem navegar através de suas próprias *tags* numa rede interativa que emprega diferentes métodos de imagem 2D, 3D e círculos. É possível escolher o tipo de relação entre *tags* e os recursos são listados na parte direita da *interface*.

<sup>9</sup><http://libguides.daltonstate.edu/c.php?g=24583&p=148455>

<sup>10</sup><http://archive.stefaner.eu/write-talk/ma-thesis-visual-tools/>

<sup>11</sup><http://www.ijdesign.org/ojs/index.php/IJDesign/article/view/622/284>

<sup>12</sup><https://delicious.com/>



Figura 2.2: Técnica Elastic Tag Maps (fonte:M.A. Thesis)

### 2.3.2 Recomendação de *tags*

A marcação social permite que os usuários contribuam com *tags* de sua escolha e segundo os próprios interesses, sendo estes meios válidos para apoiar sistemas de recomendação. Com esse argumento, Lops et al. (2009), apresentam um sistema de recomendação baseado em folksonomia. O FIRSt (*Folksonomy-based Item Recommender System*) que utiliza uma estratégia integrada de técnicas e algoritmos de aprendizado de máquina para inferir interesses a partir das *tags* informadas pelo próprio usuário.

Xu et al. (2006) definem um conjunto de critérios que um sistema de marcação deve ter: cobertura de múltiplas facetas para garantir uma boa recuperação de informações; menos esforço na atividade de marcação; redução do custo envolvido na navegação e alta popularidade para garantir a qualidade da *tag*. A partir destes critérios, propõem um algoritmo colaborativo para sugestão de *tags*, aplicando esses critérios. O algoritmo utiliza também o conceito de recompensa e punição, para combater *spams*. Este conceito, consiste em que a cada passo do algoritmo, após uma *tag* ser selecionada para recomendação, *tags* redundantes são "penalizadas", ou seja, perdem pontuação, enquanto que as *tags* co-ocorrentes com a *tag* selecionada são "recompensadas" ganhando pontuação para a recomendação. A eficácia do algoritmo é comprovada através de um experimento com base no *My Web 2.0*<sup>13</sup>.

A filtragem colaborativa é a técnica que tem sido mais adotada em sistemas de recomendação de *tags*. Baseia-se em filtrar informação baseada não só considerando o conteúdo da informação, mas também na avaliação das pessoas. Além dessa técnica, outros algoritmos tem sido aplicados na recomendação de *tags* (MILICEVIC; NANOPOULOS; IVANOVIC, 2010). O PLSA é um algoritmo de recomendação baseado em probabilidade proposto por Hofmann (1999) e que recebeu várias extensões com o decorrer do tempo. Pode ser aplicado para o processamento de linguagem natural e aprendizado de máquina, tendo como principal característica a análise semântica da co-ocorrências de dados. O FolkRank, outro algoritmo, adota um modelo de grafos

<sup>13</sup><http://www.pgeorge.net/>





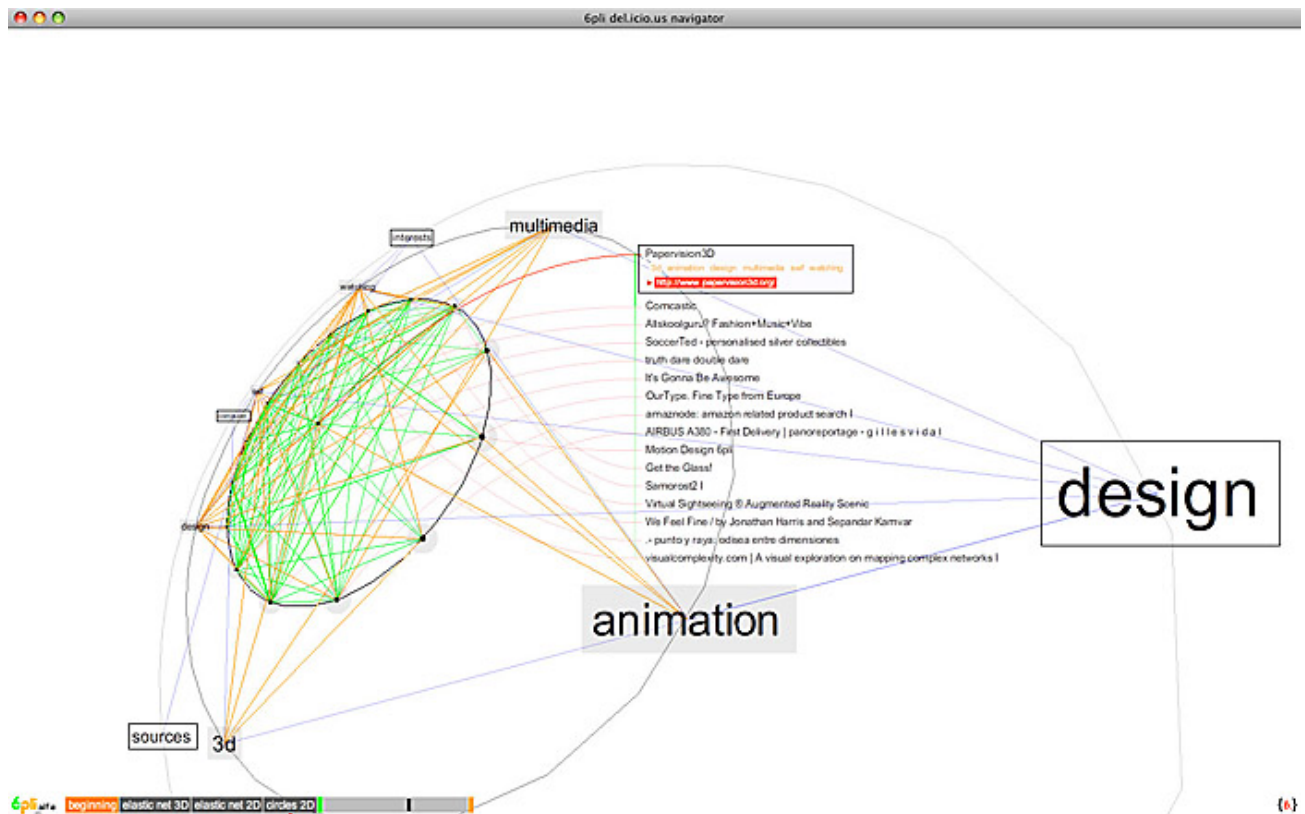


Figura 2.4: Técnica de visualização 6PLi (fonte:International Journal Design)

conteúdos para que a partir daí possam ser recomendados, e um externo, onde a recomendação passa a ser feita a partir da utilização, marcação e comentários dos próprios usuários.

Ribeiro, Fonseca e Freitas (2013), apresentam um sistema para recomendação de Objetos de Aprendizagem (OA's) aos alunos do Ambiente Virtual de Aprendizagem (AVA) *Moodle*, baseado no conjunto das *hashtags* atribuídas nas postagens dos fóruns, fazendo uso do esquema da frequência de ocorrência do termo no documento e do inverso da frequência do termo entre documentos da coleção. Nessa proposta os autores utilizam a abordagem de Classificar as *hashtags* calculando o peso e a pontuação de cada uma para a partir daí filtrarem OA's que contenham algum texto para a indexação. Desenvolvem uma extensão para o *Moodle* que tem código aberto. Concluem após um experimento que a recomendação de OA's dentro de um AVA, contribui com o processo de ensino-aprendizagem, melhorando a pesquisa e diminuindo a dispersão dos alunos na busca de conteúdos relevantes a um curso EaD (Ensino a Distância) específico.

Nos últimos anos, houve um crescimento no número de estudos que enfatizam a preocupação com o perfil do usuário para recomendações de conteúdos, *tags* e sugestão de amigos mais adequadas. Jelassi e Yahia (2013), propõe um algoritmo de recomendação que a partir do perfil do usuário numa folksonomia, procura sugerir conteúdos, *tags* e amigos mais personalizados. Para isso utilizam o conceito quadrático, que representa a personomia como um conjunto de quatro dimensões:  $QC = (U, T, R, P)$ , onde  $U$  é o conjunto de usuários,  $T$  o conjunto de *tags*,  $R$  o conjunto de recursos e  $P$  o conjunto de perfis calculado com técnicas de agrupamento e

alguns dados coletados previamente. Personomia é o conjunto de todas as *tags* utilizadas pelo usuário. Após testar o algoritmo na base de dados do site de compartilhamento de filmes MovieLens<sup>14</sup>, concluem que o conceito quadrático proposto, é um caminho eficaz para que sistemas de recomendação baseados em folksonomia ofereçam recomendações personalizadas de recursos *tags* e usuários em qualquer contexto.

A mesma abordagem de recomendação personalizada é seguida por Anand e Mampilli (2014) que propõe um método de recomendação de filmes baseada na teoria dos conjuntos *fuzzy* a partir da folksonomia gerada pela atividade de marcação colaborativa dos usuários. Assim a técnica proposta para fazer a recomendação, combina o conjunto de preferências dos usuários, informações do conteúdo e o conjunto de *tags* num único perfil para melhorar a recomendação. Aplicando o algoritmo proposto no escopo de filmes, utilizam as *tags* dos usuário para gerar a classificação de um filme específico em vários gêneros e não apenas ao gênero indicado por um especialista. O método *fuzzy* de recomendação se apresentou eficaz, comparado com outras abordagens de sistemas de recomendação.

Sigurbjournsson e Zwol (2008) analisaram o comportamento de atribuição de *tags* no *Flickr*. Concluíram que a maioria dos usuários atribuem poucas *tags* a suas fotos e em geral estas são informações de onde/que/quem e quando a foto foi tirada. Sua pesquisa teve uma contribuição substancial para a compreensão da distribuição das *tags*. Este trabalho usa a co-ocorrência para definir um *ranking* preliminar antes de refinar a recomendação. Os autores utilizaram duas estratégias para obter a co-ocorrência entre as *tags*: a primeira abordagem utiliza a similaridade baseada na fórmula de Jaccard. A Figura 2.5 apresenta respectivamente as formas simétrica e assimétrica para calcular os coeficientes de co-ocorrência, onde  $t_i$  e  $t_j$  são as *tags* para verificação da co-ocorrência e da propabilidade de um objeto ser marcado com a *tag*  $t_j$  dado que foi marcado com a *tag*.

$$J(t_i, t_j) := \frac{|t_i \cap t_j|}{|t_i \cup t_j|} \qquad P(t_j|t_i) := \frac{|t_i \cap t_j|}{|t_i|}$$

Figura 2.5: Jaccard nas formas simétrica e assimétrica para medir co-ocorrência

Além disto, os autores desenvolveram dois métodos de agregação, sendo um baseado em votos e outro em somatório. O método de agregação baseado em votos não leva em consideração a co-ocorrência das *tags* definidas pelo usuário enquanto a estratégia de somatório utiliza a co-ocorrência para a recomendação final de *tags*. Para entrega das *tags* mais relevantes, os autores desenvolveram uma função de promoção utilizando estratégias chamadas *stability*, *descriptive* e *rank*. A estabilidade (*stability*) considera que as *tags* atribuídas pelos usuários com baixa frequência são menos desejáveis que *tags* com alta frequência. Já a medida de descrição (*descriptive*) considera que *tags* com frequências muito altas provavelmente tendem a ser muito genéricas para fotos individuais. A última medida desenvolvida chamada de *rank* é a promoção do *ranking* que não utiliza a co-ocorrência, mas sim a posição da *tag* candidata.

<sup>14</sup><http://movielens.umn.edu/>

### 2.3.3 Objetos de aprendizagem

Um modelo colaborativo de construção de repositório de objetos de aprendizagem é proposto por Monge, Ovelar e Azpeitia (2008). Os autores defendem a idéia de que cada vez mais conteúdos educacionais, precisam estar disponíveis de forma aberta e multidisciplinar. Consideram a marcação social uma técnica que acrescenta uma rica semântica aos conteúdos e ajudam a dinâmica social dos repositórios de aprendizagem. Argumentam que dessa forma, professores, alunos e institutos de pesquisa, podem estruturar uma grande rede de conhecimento.

Sierra e Valmayor (2008) descrevem uma abordagem de marcação de objetos de aprendizagem para a criação e extensão de metadados, propondo que no momento da criação do objeto de aprendizagem, o criador já defina o esquema do metadado, especificando um identificador da categoria, nome, modificadores e valores. Dahl e Vossen (2008b) dissertam sobre os caminhos para a criação de metadados para objetos de aprendizagem e argumentam que a geração de metadados pode ser automática, a partir de características dos próprios objetos, seu contexto de uso, tornando a marcação dos objetos semi-automática e humana através de atividades de marcação social.

Shih e Tseng (2008) propõe um algoritmo de criação de índices baseado em folksonomia para a recuperação de objetos criados no padrão SCORM. O SCORM (*Sharable Content Object Reference Model*) é um conjunto de padrões definidos para estruturar objetos de aprendizagem para *e-Learning* publicado pela ADL<sup>15</sup>(1999). O algoritmo proposto utiliza uma função de similaridade para agrupar e classificar as *tags*.

VITAE<sup>16</sup> é uma ferramenta para recuperação de objetos de aprendizagem, que adota características de marcação social através do uso de *tags* proposta por Coelho (2009). A ferramenta implementa a recuperação de objetos de aprendizagem através de filtros realizados a partir das *tags* vinculadas aos objetos.

Cho et al. (2011) realizam um experimento e propõe uma plataforma de marcação social que auxilia os usuários a organizar, gerenciar e recuperar objetos de aprendizagem. Implementam um sistema de marcação sobre um repositório de 4000 objetos de aprendizagem e verificam como os objetos podem ser melhor descritos e pesquisados dessa abordagem.

Teixeira, Sá e Fernandes (2011) apresentam um estudo que busca prover um processo de indexação e organização de objetos de aprendizagem (OA's) para uso posterior em cenários pedagógicos virtuais a partir da integração das tecnologias de folksonomia e taxonomia.

Shih e Tseng (2012) propõe dois algoritmos para construção e manutenção de índice para objetos de aprendizagem baseado em folksonomia. O objetivo principal é auxiliar tutores voluntários que não são profissionais da educação a orientarem alunos, localizando de forma adequada recursos alternativos para serem recomendados a fim de favorecer a aprendizagem. Na abordagem apresentada por eles, folksonomias são mescladas conforme o nível de similaridade entre as *tags*. O mecanismo de manutenção do índice é disparado automaticamente sempre que o vocabulário da folksonomia principal se torna muito grande, limite este que é configurado por parâmetros. Após a condução de dois experimentos com tutoriamento em lições de matemática, analisam junto a tutores e alunos a impressão que tiveram ao utilizar o sistema. Concluem que

<sup>15</sup>Advanced Distributed Learning Initiative

<sup>16</sup><http://bookmark.aspectos.org>

índices baseados em folksonomia, propiciam maior precisão semântica na busca de recursos e podem ser aplicados para auxiliar tutores a encontrar mais rapidamente alternativas de ensino dentro de um repositório de recursos e orientar alunos que possuem dificuldade de aprendizado em determinado assunto.

Um algoritmo de Classificação *Multi-Label*(CML) é proposto por López, Prieta e Ogihara (2012). O objetivo dessa abordagem é mapear OA's dentro do padrão LOM (*Learning Object Metadata*) em tipos de consultas, para melhorar a primeira escolha do aluno ou professor. O algoritmo de classificação *multi-label* foi testado utilizando apenas os dados dos OA's e também combinando os metadados com as *tags* fornecidas pelos próprios usuários. Esse algoritmo pode ser aplicado para a organização OA's de diferentes repositórios. Após experimentos, é constatado a partir dos resultados que com a classificação *multi-label*, é possível criar classes e sub-classes de *tags*, o que permite uma organização hierárquica dos OA's. Comparado com outros métodos de classificação a CML, mostrou-se superior quanto a precisão, tanto na classificação de recursos de texto como também na classificação de OA's, estruturados no padrão *LOM*.

Zervas e Sampson (2014) desenvolvem um trabalho, no qual investigam a idéia de que os metadados de recursos educacionais (OA's) fornecidos pelos próprios criadores desses recursos, podem ser ampliados através da marcação social. Consideram que isso é diretamente influenciado pela motivação que os usuários tem para a marcação. Assim propõe uma metodologia para avaliar a motivação dos usuários e comparar a qualidade de marcação de cada grupo. Para seus experimentos, utilizam um repositório de objetos de aprendizagem existente chamado OpenScienceResources Repository<sup>17</sup>. A análise para determinar a característica dos usuários é baseada na classificação de Strohmaier, Korner e Kern (2012). Nesta classificação os usuários são identificados como catalizadores ou descritores quanto a motivação. Assim após os experimentos e efetuar comparações, concluem que usuários descritores, contribuem mais para a ampliação do vocabulário formal dos metadados, por produzirem *tags* mais voltadas a descrição dos recursos. Enquanto que a característica dos categorizadores é fornecer *tags* mais apropriadas para favorecer a navegação entre vários objetos de aprendizagem.

### 2.3.4 Geração de *tags*

A fim de descrever, compreender e analisar *tags* e sistemas de marcação, foram propostos vários modelos de geração de *tags*. Estes modelos estudam vários fatores que influenciam a geração de uma *tag*, tais como as marcações anteriores sugeridas por outros, o conhecimento dos usuários, o conteúdo dos recursos e a influência da comunidade. Dentre os modelos de geração de *tags* encontrados na literatura, pode-se citar o processo de Urna Polia. Esse modelo, utiliza uma técnica de simulação para capturar *tags* atribuídas anteriormente que são mais suscetíveis a serem novamente selecionadas (GUPTA et al., 2010). A idéia básica da simulação é colocar *tags* similares juntas num mesmo local de armazenamento (urna). Em cada passo da simulação, as *tags* selecionadas são reorganizadas, até ocorrer uma estabilização do vocabulário.

O modelo de Yule-Simon, é uma extensão do modelo Urna Polia e tem como característica, "inventar" novas *tags* para adicioná-las numa corrente de baixa probabilidade. Assim, em cada passo da simulação, verifica-se qual das *tags* já existentes na "Urna" tem maior probabilidade de

---

<sup>17</sup><http://www.osrportal.eu/>

vir a ser atribuída a um objeto que nunca foi marcado por ela. Após isso, cada *tag* vai recebendo um grau maior de probabilidade, conforme a frequência de ocorrência da mesma (GUPTA et al., 2010). Outro modelo é o baseado no valor da informação que seleciona *tags* de acordo com o valor da informação que está atribuída a um recurso. O sistema atribui o valor 1 para uma *tag* se ela identifica diretamente um recurso e valor 0 (zero) se conduz a qualquer ou a muitos recursos. Esse modelo é uma extensão do modelo "Urna Polia", e aplica probabilidade para prever marcações (GUPTA et al., 2010).

Muitos artigos utilizam como exemplo o sistema *Bibsonomy*<sup>18</sup> que incorpora técnicas chamadas *scraping* para automaticamente extrair informações a partir da página *web* que o usuário está visitando para facilitar ainda mais o processo de atribuição de *tags*. Para implementar essa técnica, o *Bibsonomy* utiliza agentes capazes de coletar informações de páginas *web*. Essas informações coletadas ficam catalogadas dentro do sistema. Assim, é possível a criação de uma biblioteca de referências que apontam para os conteúdos anotados pelos usuários (BENZ et al., 2010).

Uma das maneiras de recomendar *tags* é através da identificação de padrões de comportamento na atividade de marcação proposta por Golder e Huberman (2009). Nesse trabalho, são analisadas várias atividades de usuários dentro de um sistema de marcação social, e o que se pode observar é que após um período de atividade de marcação, padrões estáveis emergem, e podem ser aplicados para descrever e organizar recursos pessoais e compartilhados.

Cernea, Moral e Gayo (2008) propõe um o sistema chamado SOAF (Semântica dos Objetos de Aprendizagem Baseada em Folksonomia). Esse sistema tem uma arquitetura que combina técnicas automáticas de extração de informações com tecnologias de marcação colaborativa.

Wu e Bo (2011) sugerem diferentes algoritmos para formar uma rede de folksonomias. Com essa proposta, acreditam melhorar significativamente a semântica entre as *tags* com a possibilidade de aplicar até mesmo processamento de linguagens naturais e conferindo maior flexibilidade aos sistemas baseados em *tags*.

Além dos modelos de geração automática de *tags*, muitos autores se dedicaram a estudar os fatores que influenciam cada usuário no momento da marcação. A intenção de recuperação futura de objetos, a contribuição e compartilhamento e expressão de opinião pessoal, são exemplos das motivações dos usuários durante a atividade de marcação (DAHL; VOSSSEN, 2008a).

Garcia-Silva et al. (2012) apresentam uma revisão sobre as técnicas mais relevantes que objetivam descobrir e associar semântica (significado) às *tags* em folksonomias. Verificam na literatura científica, três tipos de abordagem utilizadas pelos vários autores: técnicas que utilizam clusterização; técnicas que utilizam ontologias; e técnicas híbridas que combinam as duas abordagens anteriores. Para poder comparar e avaliar as vantagens e desvantagens de cada abordagem estudada, os autores propõe um processo unificado de análise composto pelos conceitos de: seleção e limpeza dos dados; identificação do contexto; desambiguação e identificação da semântica. Listam então vários trabalhos, analisando cada um desses pontos em cada abordagem, detalhando e classificando cada trabalho quanto aos tipos de técnicas de co-ocorrências utilizados. Após essas discussões, os autores apresentam um conjunto de recomendações para melhorar cada abordagem de acordo com o objetivo da aplicação dessas técnicas, propondo

---

<sup>18</sup>[www.bibsonomy.org](http://www.bibsonomy.org)

também a combinação dos pontos fortes de cada trabalho.

Assis e Moura (2013) estudam as folksonomias enquanto redes de significados compartilhados por redes sociais. Optam por uma análise qualitativa baseada em semiótica onde consideram as *tags* como signos que são estabelecidos a partir de acordos interpretativos por parte dos usuários de uma comunidade virtual mediante a descrição e validação dos conteúdos. Enfatizam então os laços que surgem desta interatividade considerando como folksonomias podem enriquecer as linguagens formais e vice-versa, ou seja, como as linguagens formais representadas por metadatos, tesouros e ontologias podem contribuir para equilibrar o caos aparente das folksonomias. Ressaltam também a flexibilidade das folksonomias, cujo vocabulário em torno de um conteúdo pode se estabilizar devido a alta frequência de utilização de algumas *tags* por parte de grupos de usuários, e onde a renovação da linguagem pode ser atualizada com o passar do tempo pelo surgimento de novos termos e conceitos inseridos por outros usuários. Após analisarem as ferramentas *Delicious*, *Diigo* e *Stumble Upon*, e investigarem os usuários participantes dessas redes sociais, concluem que a observação dessas dinâmicas de compartilhamento pode contribuir para a ampliação das metodologias de organização intelectual da informação, pois geram extensões da memória e da cultura dos usuários.

## 2.4 Algoritmos de Recomendação por Co-ocorrência

A partir dos resultados da revisão sistemática pode-se destacar alguns algoritmos de co-ocorrência utilizados para recomendação de *tags*. Entre os algoritmos de recomendação de *tags* estudados, destacaram-se os algoritmos de co-ocorrência propostos por Sigurbjournsson e Zwol (2008) e o proposto por Xu et al. (2006). Por ser mais flexível quanto a suas métricas, isto é, em relação aos cálculos de pontuação das *tags* que são recomendadas, o algoritmo de Xu et al. (2006) foi escolhido para ser explorado. O algoritmo de Xu et al. (2006) tem como base a ideia que, dado um usuário  $U$  e um objeto  $O$ , deve-se encontrar todas as *tags* co-ocorrentes ( $T_{CO}$ ), ou seja, *tags* que compartilham pelo menos um objeto, com as *tags* que o usuário  $U$  atribuiu ao objeto  $O(T_{O;U})$ . Calcula-se a pontuação inicial dessas *tags* ( $S(t, O)$ ), e por  $n$  vezes escolhe-se a *tag* com maior pontuação e reajusta-se a pontuação das outras *tags*, a partir da *tag* escolhida. Ao final desse processo, obtém-se uma lista de  $n$  *tags* para serem recomendadas. O pseudocódigo a seguir apresenta o detalhamento do algoritmo.

Considerando:

$n$  = número de *tags* que se deseja recomendar, que deve ser maior ou igual a um.

$U$  = usuário que está solicitando a recomendação.

$O$  = O objeto em questão.

$T$  = Todas as *Tags* do repositório.

$T_{O;U}$  = *tags* dadas ao objeto  $O$  pelo usuário  $U$ .

$Q(t)$  = N° de objetos com a *tag*  $t$ .

$U(t)$  = N° de usuários que usaram a *tag*  $t$ .

$Q(t; O)$  = N° de vezes que o objeto  $O$  recebeu a *tag*  $t$ .

$CO(t, t')$  = N° de objetos que tem as *tags*  $t$  e  $t'$ .

$CO(t, t'; U) = N^{\circ}$  de objetos que o usuário  $U$  colocou as *tags*  $t$  e  $t'$ .

$CO(t, t'; O) = N^{\circ}$  de vezes que o objeto  $O$  recebeu  $t$  e  $t'$  de um mesmo usuário.

$P(t|t'; U) = \frac{CO(t, t'; U)}{Q(t')}$  = Indica o quanto para um usuário duas *tags* ,  $t$  e  $t'$ , tem significados parecidos.

$P(t|t'; O) = \frac{CO(t, t'; O)}{Q(t')}$  = Indica o quanto para um objeto  $O$ , as *tags*  $t$  e  $t'$ , tem significados complementares, uma vez que essa métrica indica o quanto essas *tags* aparecem juntas, no objeto

O algoritmo para recomendar as *tags* é apresentado na Tabela 2.2.

Tabela 2.2: Algoritmo de Recomendação de *tags* (XU et al., 2006)

01.  $T_{CO} = \{t : t \in T \wedge t \notin T_{O;U} \wedge \exists t' \in T_{O;U} \rightarrow CO(t, t') > 0\}$
02.  $\forall t \in T_{CO}, S(t, O) = \frac{U(t)}{Q(t)}$ ,  $S(t, O)$  calcula quão específica a tag é para o usuário  $U$
03. Ordene na decrescente as *tags* de  $T_{CO}$  de acordo com  $S$ .
04. Assim  $T_{CO}[0]$  é a tag  $t$  com maior valor de  $S(t, O)$  05.  $t_e = T_{CO}[0]$
06.  $Tags\ Escolhidas = \{t_e\}$
07.  $T_{CO} = T_{CO} - t_e$
08.  $i = 1$
09. Enquanto  $i < n \wedge T_{CO} \neq \emptyset$  faça:
10.  $\forall t \in T_{CO}, S(t, O) = S(t, O) - P(t|t_e; U) * S(t_e, O) + P(t|t_e; O) * S(t_e, O)$
11. Ordene  $T_{CO}$  pelo valor  $S$ .
12.  $t_e = T_{CO}[0]$
13.  $Tags\ Escolhidas = Tags\ Escolhidas \cup t_e$
14.  $T_{CO} = T_{CO} - t_e$
15.  $i = i + 1$
16. Fim Enquanto
17. Retornar *tags Escolhidas*

Observando-se o algoritmo, verifica-se que:

**linha 01:** é definido o conjunto de *tags* co-ocorrentes ( $T_{CO}$ ) às *tags* do usuário ( $U$ ) , considerando todas as *tags* do repositório ( $T$ ).

**linha 02:** é calculada a pontuação de cada *tag* do conjunto  $T_{CO}$ , dada por  $\frac{U(t)}{Q(t)}$ , onde quanto mais próximo do valor 1, mais específica é a *tag*  $t$  para o usuário  $U$ .

**Linha 03:** o conjunto  $T_{CO}$ , é ordenado na ordem decrescente

**linha 05:** obtem-se a *tag* de maior pontuação ( $T_{CO}[0]$ )

**linha 06:** a *tag* de maior pontuação é escolhida ( $t_e$ ) e associada ao conjunto "*Tags Escolhidas*".

**linha 07:** a *tag*  $t_e$  é eliminada do conjunto  $T_{CO}$ . **linha 08:** a variável  $i$  é inicializada com 1, para que o processo seja repetido por  $n$  vezes ou até que o conjunto  $T_{CO}$  esteja vazio (**linha 09**).

**linha 10:** a pontuação é recalculada para todas as *tags* restantes no conjunto  $T_{CO}$ , recompensando *tags* mais específicas para o usuário  $U$ , através da expressão  $(-S(t, O) - P(t|t_e; U))$  e

penalizando *tags* redundantes com a expressão  $(+P(t|t_e; O) * S(t_e, O))$ .

Dessa forma, as *tags* de maior pontuação são selecionadas e atribuídas ao conjunto "*Tags Escolhidas*" que ao final é retornado (**linha 17**).

## 2.5 Considerações Finais

A partir do estudo realizado observou-se que embora o potencial da marcação social seja ressaltado não foram encontrados trabalhos que abordem todas as perspectivas apresentadas na seção anterior em conjunto. Através da Tabela 2.1 observa-se que todos os estudos abordam ao menos um dos aspectos. Porém, dos 35 estudos, apenas 9 fazem referência a 3 dos aspectos analisados e nenhum chega abordar os 4 aspectos conjuntamente.

Das revisões sistemáticas encontradas, a apresentada por Gupta et al. (2010) não aborda o aspecto da visualização de *tags*. Acredita-se que tal aspecto é importante condutor do aluno na atividade de marcação de objetos de aprendizagem. O aspecto visual permite que o aluno possa identificar outros objetos ou mesmo outras *tags*, facilitando a marcação dos objetos. A visualização das *tags* é abordada por Conole e Culver (2010) através da proposta do *site* Cloudworks<sup>19</sup>. Já a revisão de Milicevic, Nanopoulos e Ivanovic (2010), não faz referência ao uso de objetos de aprendizagem. Outra revisão, de Dahl e Vossen (2008a) que destacam a evolução das folksonomias em ambientes de aprendizagem, não abordam a questão da recomendação de objetos de aprendizagem e de *tags*.

Dos trabalhos que abordam o conceito da atividade de marcação, muitos exploram a geração automática de *tags* a partir dos dados descritores do objeto de aprendizagem catalogado. Uma proposta nessa direção é apresentada por Sierra e Valmayor (2008), que apontam a criação automática e exploração de metadados para objetos de aprendizagem como um caminho para melhorar a precisão e a qualidade do acesso aos conteúdos. Cernea, Moral e Gayo (2008), também propõem a utilização de metadados, gerados a partir de uma indexação semântica de objetos de aprendizagem baseada em folksonomia, combinando técnicas automáticas de extração de informações com tecnologias de marcação colaborativa. Com isso, acreditam garantir a reusabilidade dos objetos de aprendizagem em outros contextos de aprendizagem. Pode-se observar nesse estudo, a grande tendência em se buscar técnicas e algoritmos que favoreçam a recomendação personalizada de conteúdos e *tags* além de usuários para formação de grupos, em qualquer contexto.

O crescimento da *Web Social*, caracterizada principalmente pelo armazenamento e uso de conteúdos elaborados pelo próprio usuário, tem motivado o desenvolvimento de ambientes que incorporem os princípios que permitam ao usuários fazer parte do processo construtor de informações. Dentro da área de *e-learning* esta demanda não é diferente. A marcação social (*social tagging*) é um dos princípios que permite que o aluno tenha liberdade na criação de vocabulários que classificarão e descreverão um determinado recurso de aprendizagem.

Conforme os trabalhos relacionados na revisão sistemática, observou-se que os estudos sobre marcação social tem sido explorados por muitos autores nos últimos anos, apresentando abordagens diversificadas, e colocando em destaque a importância dessa forma de organizar e indexar

<sup>19</sup><http://cloudworks.ac.uk/>



a informação. Observou-se através do estudo realizado nessa revisão que as perspectivas de visualização, geração e recomendação de *tags* são pouco exploradas conjuntamente.

Combinar as diversas tecnologias de marcação social relacionadas nesse trabalho com o objetivo de favorecer a catalogação, indexação, navegação e recomendação dos diversos tipos de objetos de aprendizagem de forma a facilitar todo o processo de ensino-aprendizagem, certamente é uma área que pode ser mais explorada. Assim acredita-se que com esse estudo obteve-se uma ampla visão sobre o estado da arte a respeito dos conceitos e técnicas envolvidos na marcação social e que são fundamentais para a aplicação adequada na área de *e-learning*.

Após o levantamento e análise do estado da arte sobre a marcação social vista sobre as 4 perspectivas abordadas, verificou-se a falta de trabalhos que proponham a catalogação de recursos de aprendizagem através da folksonomia, utilizando o recurso de recomendação de *tags*. Faltam propostas que considerem recursos de aprendizagem que não estejam necessariamente catalogados em um repositório específico. Os trabalhos que se aproximam do conceito de folksonomia em ambientes de aprendizagem (DAHL; VOSSEN, 2008a), não tratam sobre os conceitos de recomendação.

Assim, a proposta desse trabalho visa suprir esta lacuna, destacando a catalogação de recursos de aprendizagem disponíveis na *web* através da marcação social, com o auxílio de um sistema de recomendação de *tags*.

## Sistema de Recomendação de *Tags*

A aprendizagem em qualquer domínio do conhecimento passa pela dinâmica de encontrar um conteúdo, estudá-lo e compreendê-lo. Para isso, professores e tutores exercem o papel de facilitadores e direcionadores dessa aprendizagem que precisa cada vez mais ser personalizada, devido a heterogeneidade dos alunos que possuem diferentes características individuais para aprender. A área de *e-Learning* surge para contribuir com esse processo de ensino-aprendizagem, de forma a prover mecanismos que potencializem a aprendizagem. Através de recursos computacionais e da *Web*, grandes avanços ocorreram nessa direção.

A *Web* é uma fonte de informações colaborativa, onde podem-se encontrar conteúdos para prover conhecimento em diversas áreas. Contudo, é primordial prover meios para a catalogação desses conteúdos multimídias com o objetivo de apoiar a recuperação dessas informações, quando necessário. Sistemas de marcação social, através do emprego do conceito de folksonomias, oferecem a possibilidade dessa classificação personalizada, onde cada aluno pode encontrar e catalogar conteúdos, de acordo com seus interesses de aprendizagem.

Torna-se fundamental trabalhar com aspectos que possam otimizar o processo de catalogação. A existência de um mecanismo que dê suporte a essa catalogação, através de recomendação de *tags* relevantes para o recurso de aprendizagem encontrado e selecionado, pode auxiliar este processo. Isso garante uma recuperação mais precisa dos recursos e uma visão mais abrangente devido as relações entre conteúdos e alunos que surgem nesse ambiente.

Considerando essa necessidade, a proposta de um sistema de recomendação de *tags* é apresentada na Seção 3.1 e os estudos realizados para se definir o algoritmo adotado para o modelo de recomendação proposto são apresentados na Seção 3.2.

### 3.1 e-TagReS: Sistema de Recomendação de *Tags* Proposto

Um importante problema a ser considerado na marcação social é a identificação de *tags* mais adequadas e a eliminação de *spams*, uma vez que a marcação é feita de forma livre e individual por usuários finais. Num ambiente de marcação social, *spams* são *tags* consideradas fora do contexto, que são utilizadas raras vezes ou por apenas um usuário e que podem influenciar

negativamente o funcionamento de algoritmos de recomendação.

O objetivo deste trabalho é propor um sistema de recomendação de *tags* num contexto de *e-learning* que auxilie a catalogação de recursos de aprendizagem. A partir do trabalho apresentado por Xu et al. (2006), um algoritmo de recomendação foi elaborado em função do ciclo de recomendação idealizado. A Figura 3.1 apresenta o sistema proposto denominado **e-TagReS** (**Tag Recommender System - Sistema Recomendador de Tags**).

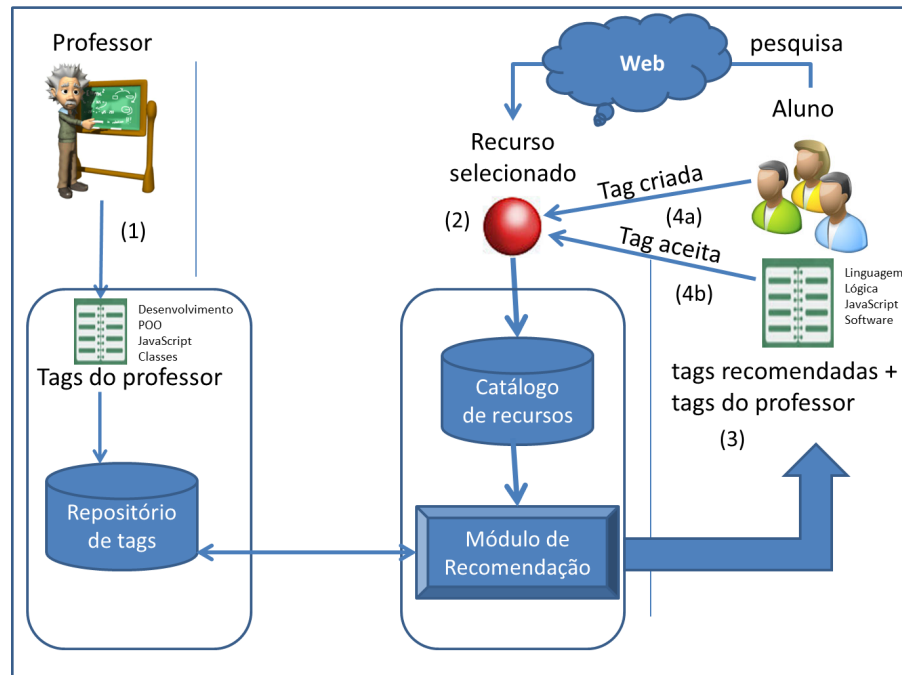


Figura 3.1: Ciclo de Recomendação e-TagReS

O objetivo do **e-TagReS** é auxiliar alunos na catalogação de recursos de aprendizagem disponíveis na *Web*. O seu principal foco é propiciar mecanismos que dêem suporte à formação de um vocabulário descritor em torno dos recursos selecionados, que favoreça a navegação e pesquisa futuras. O termo Recursos de Aprendizagem (RA's), refere-se tanto a objetos de aprendizagem previamente configurados segundo um padrão de metadados, como a qualquer conteúdo disponível na *Web* tais como, vídeos, *wikis*, textos, fóruns, que favoreçam o aprendizado de um tema específico. Foi adotado o termo (RA) em vez de (OA) para tornar o conjunto de possibilidades de conteúdos mais abrangente, não se limitando, apenas a objetos de aprendizagem pré-configurados conforme um padrão de metadados.

A partir do modelo proposto define-se que o processo de catalogação é composto de 5 etapas:

**Etapa 1:** preparação do professor que define temas que direcionarão a catalogação.

**Etapa 2:** execução de buscas pelos recursos, seguindo os temas definidos pelo professor.

**Etapa 3:** seleção das *k* tags mais relevantes que serão recomendadas.

**Etapa 4:** atribuição das tags que poderá ser:

**Etapa 4a** atribuição com tags criadas pelo aluno ou

**Etapa 4b** Atribuição com tags recomendadas e aceitas pelo aluno.

O processo de catalogação é preparado a partir de um projeto definido pelo **professor** de acordo com os objetivos de aprendizagem desejados. Após definir um tema, o professor especifica

termos (***tags do professor***)(1) que direcionarão a pesquisa dos alunos. Esses termos também serão sugeridos durante o processo de catalogação para todos os recursos pesquisados. Com isto, é contornado o problema de "partida a frio" da marcação colaborativa, ou seja, situações onde tanto recursos e/ou usuários ainda não participaram de nenhuma marcação, não possuindo assim *tags* associadas a eles.

Conforme a Figura 3.1, o processo de marcação se inicia com o **aluno** pesquisando sobre os temas que fazem parte do escopo definido pelo professor. Ao **selecionar um recurso**(2), é apresentada ao aluno uma lista das  $k$  *tags* mais relevantes provenientes do módulo de recomendação e dos termos definidos pelo professor(3). Ao atribuir *tags* a um recurso, o aluno pode optar por aceitar ou não a sugestão do algoritmo. Caso nenhuma *tag* sugerida seja aceita pelo aluno (4b), ele terá a liberdade de criar uma nova *tag* para o recurso(4a). Dessa forma, a tarefa de catalogar recursos de aprendizagem é realizada com o auxílio do mecanismo de recomendação de *tags*.

Um elemento importante no ciclo de recomendação é o algoritmo utilizado. Para decidir o melhor algoritmo foram realizadas experimentações, conforme descrito da Seção 3.2.

## 3.2 Estudo do Algoritmo de Recomendação e adaptações

Antes de se definir o algoritmo de recomendação que comporia o módulo de recomendação, um estudo preliminar foi realizado analisando-se algoritmos de co-ocorrência existentes na literatura (Seção 2.4). O algoritmo de Xu et al. (2006) (Tabela 2.2) foi estudado e adaptado à proposta deste trabalho. Observando-se o pseudo-código da Tabela 2.2, percebe-se que ao se filtrar as *tags* de  $T_{CO}$  ou modificando a forma que a pontuação inicial é calculada, e, alterando como as pontuações são reajustadas, é possível, ajustar o algoritmo para um contexto específico, pois essas modificações mudam parcialmente a forma com que o algoritmo calcula a especificidade da *tag* e sua influência em outras *tags*. A partir disto foram planejados alguns testes com o algoritmo, considerando uma base de recursos de aprendizagem que já possuíam *tags* atribuídas por alunos.

### 3.2.1 Testes das variações das métricas

Procurando ajustar o algoritmo escolhido, foi realizado uma série de testes onde filtros foram colocados nas *tags* co-ocorrentes e/ou modificações foram feitas no cálculo da pontuação inicial. Os testes foram divididos em duas etapas. Na primeira, considerou-se as *tags* que um determinado usuário atribuiu a um recurso e, na segunda parte, foram consideradas todas as *tags* que o recurso recebeu independente de quem as atribuiu a ele. Os testes tomaram como entrada um usuário com uma quantidade grande de marcações em comparação com outros usuários. Considerando um repositório já existente, os objetos analisados foram:

- *jQuery*, que tinha um número de marcações que representa objetos que foram muito intensamente marcados;
- *Android*, com um pouco menos de marcações que o *jQuery* e representando objetos que foram extensamente marcados;

- *Adobe Flash Player*, que tinha uma número médio de marcações e representa objetos que foram marcados;
- *Flash*, que tinha uma número pequeno de marcações e representa objetos que foram pouco marcados;
- *Wikipédia* sobre Windows NT, que tinha uma número escasso de marcações e representava objetos que praticamente não foram marcados;

Os testes resultaram em algumas variações e comparações, que mostram quais *tags* foram recomendadas e quais alterações foram feitas no algoritmo em cada um dos testes.

O processo de encontrar as *tags* co-ocorrentes pode-se deparar com o problema de encontrar *tags* cuja semântica não esteja tão próxima das *tags* que o usuário já atribuiu ao objeto. Para evitar que essas *tags* sejam escolhidas, optou-se por se multiplicar a pontuação inicial das *tags* co-ocorrentes pela média do valor da fórmula de Jaccard, a qual mede a similaridade entre dois conjuntos finitos. Para cada *tag* co-ocorrente e para as *tags* que o usuário usou para marcar o recurso em questão, foi aplicado a seguinte equação:

$$\forall t \in T_{CO} S'(t, O) = \frac{U(t)}{Q(t)} X \frac{1}{|T_{O,U}|} X \sum_{t' \in T_{O,U}} J(O_t, O_{t'}) \quad (1)$$

Onde  $O_t$  representa o conjunto com todos os objetos marcados com a *tag*  $t$  e  $J(O_t, O_{t'})$  é definido como:

$$J(O_t, O_{t'}) = \left| \frac{O_t \cap O_{t'}}{O_t \cup O_{t'}} \right|, \text{ para } O_t \neq \emptyset \text{ ou } O_{t'} \neq \emptyset$$

$$J(O_t, O_{t'}) = 1, \text{ para } O_t = \emptyset \text{ e } O_{t'} = \emptyset \quad (2)$$

Como  $J$  retorna valores no intervalo  $[0, 1]$ , *tags* com uma semelhança baixa tem sua pontuação inicial reduzida o que diminui a chance de que ela seja escolhida, enquanto que *tags* com uma alta similaridade mantém sua pontuação inicial próxima ao que era antes, tendo assim a sua chance de ser escolhida quase inalterada.

Essa foi a **variação 1** cujo resultado é evidenciado na Tabela 3.1, apresentando os resultados obtidos com e sem a aplicação do fator de Jaccard.

Na **variação 2**, ainda buscando resolver o problema de *tags* com pouca semelhança semântica (que não se aproximam do mesmo significado) com as *tags* usadas pelo usuário, outra modificação foi testada: filtrar o conjunto de *tags* co-ocorrentes para que contenha apenas *tags* de objetos que sejam co-ocorrentes com  $n$  ou mais *tags* que o usuário usou para marcar o objeto, ou seja:

$$T'_{CO} = \{t : t \in T_{CO} \wedge \exists t' \in T_{O,U} \rightarrow CO(t, t') > n\} \quad (3)$$

Dessa forma, estabelece-se uma quantidade mínima de objetos compartilhados para que a *tag* fosse considerada. Isso levou a exclusão de *tags* muito específicas desse conjunto ou de

Tabela 3.1: Variação 1 - Aplicar ou não Jaccard na pontuação inicial

Objeto	Sem Alterações	Usando $S'(t,O)$
<b>jQuery</b>	show php adobe google play php tutorial javascript library	javascript library organização exemplificações jquery utilidade
<b>Android</b>	google play apps downloads developer for android xbox 360	developer for android google play apps oficial tecnologia
<b>Adobe</b>	show php adobe rapido google play download	rapido adobe install aaae bom
<b>FLASH</b>	-	-
<b>Wikipédia</b>	-	-

*tags* usadas em poucos objetos, o que melhora a qualidade das *tags* devido a esta eliminação. Os resultados obtidos aplicando-se essa modificação, com  $n$  variando de 0 a 3 e os resultados obtidos juntando-se a modificação na pontuação inicial com a filtragem no conjunto de *tags* co-ocorrentes, são mostrados nas Tabelas 3.2 e 3.3 respectivamente.

Tabela 3.2: Variação 2 - Usar ou não filtragem das *tags* co-ocorrentes

Objeto	Sem alterações	Usando $T'_{CO}(n=0)$	Usando $T'_{CO}(n=1)$	Usando $T'_{CO}(n=2)$	Usando $T'_{CO}(n=3)$
<b>jQuery</b>	show php adobe google play php tutorial javascript library	show php adobe google play microsoft adobe flash	show php adobe google play php tutorial rapido	adobe google play microsoft downloads javascript	adobe google play microsoft jquery dados
<b>Android</b>	google play apps downloads developer for android xbox 360	google play apps downloads developer for android jogos para android	google play apps downloads xbox 360 microsoft	google play downloads microsoft google javascript	xbox 360 microsoft jquery games javascript
<b>Adobe</b>	show php adobe rapido google play download	show php adobe rapido google play download	show php adobe google play javascript library php tutorial	adobe google play google downloads apps	adobe google play google downloads jquery
<b>FLASH</b>	-	-	-	-	-
<b>Wikipédia</b>	-	-	-	-	-

Tabela 3.3: Variação 3 - Utilizar filtragem de *tags* (Variação 1) e aplicar Jaccard ou não

Objeto	Sem alterações	Usando $T'_{CO}(n=0)$ $e$ $S'(t, O)$	Usando $T'_{CO}(n=1)$ $e$ $S'(t, O)$	Usando $T'_{CO}(n=2)$ $e$ $S'(t, O)$	Usando $T'_{CO}(n=3)$ $e$ $S'(t, O)$
	show php	javascript library	jquery	jquery	jquery
<b>jQuery</b>	adobe google play php tutorial javascript library	organização exemplificações jquery utilidade	downloads javascript library ajax	downloads javascript library facilidade	javascript dados informação library
<b>Android</b>	google play apps downloads developer for android xbox 360	developer for android google play apps oficial tecnologia	apps google play google jelly bean developers	google play google downloads tecnologia developers	tecnologia javascript html desenvolvimento mobile
<b>Adobe</b>	show php adobe rapido google play download	rapido adobe install aaae bom	adobe videos youtube plug-in download facilidade	adobe plug-in download facilidade web	adobe web jogos efeitos bom
<b>FLASH</b>	-	-	-	-	-
<b>Wikipédia</b>	-	-	-	-	-

Na **variação 4**, seguindo os mesmos princípios uma nova versão, mais restritiva, do filtro foi testada. Nessa versão, inseriu-se a imposição de mínimo de objetos compartilhados para com a *tag* co-ocorrente em relação a todas as *tags* que usuário usou. O seguinte filtro foi gerado:

$$T''_{CO} = \{t : t \in T_{CO} \wedge \forall t' \in T_{O;U} \rightarrow CO(t, t') > n\} \quad (4)$$

Os resultados obtidos aplicando-se essa modificação, com **n** variando de 0 a 3 e os resultados obtidos juntando-se a modificação na pontuação inicial com a filtragem no conjunto de tags co-ocorrentes, são apresentados nas Tabelas 3.4 e 3.5 respectivamente.

Tabela 3.4: Variação 4 - Determinar mínimo de objetos compartilhados usando ou não a Variação 3

Objeto	Sem alterações	Usando $T''_{CO}(n=0)$	Usando $T''_{CO}(n=1)$	Usando $T''_{CO}(n=2)$	Usando $T''_{CO}(n=3)$
<b>jQuery</b>	show php adobe google play php tutorial javascript library	show php adobe google play php tutorial rapido	show php adobe google play php tutorial rapido	adobe google play microsoft downloads javascript	adobe google play microsoft downloads javascript
<b>Android</b>	google play apps downloads developer for android xbox 360	google play apps downloads developer xbox 360	google play apps downloads google javascript	google play downloads xbox 360 microsoft google	google play downloads google javascript tecnologia
<b>Adobe</b>	show php adobe rapido google play download	show php adobe rapido google play download	show php adobe google play google javascript library	adobe google play google downloads jquery	adobe google play microsoft downloads apps
<b>FLASH</b>	-	-	-	-	-
<b>Wikipédia</b>	-	-	-	-	-

Tabela 3.5: Variação 5 - Usar a segunda filtragem e Jaccard ou não

Objeto	Sem alterações	Usando $T''_{CO}(n=0)$ e $S'(t, O)$	Usando $T''_{CO}(n=1)$ e $S'(t, O)$	Usando $T''_{CO}(n=2)$ e $S'(t, O)$	Usando $T''_{CO}(n=3)$ e $S'(t, O)$
<b>jQuery</b>	show php adobe google play php tutorial javascript library	javascript library organização exemplificações jquery utilidade	jquery downloads javascript library ajax	jquery downloads javascript library facilidade	jquery javascript dados informação library
<b>Android</b>	google play apps downloads developer for android xbox 360	developer for android google play apps oficial tecnologia	apps google play google jelly bean developers	google play goole downloads tecnologia developers	tecnologia javascript html desenvolvimento mobile
<b>Adobe</b>	show php adobe rapido google play download	rapido adobe install aaae bom	adobe videos youtube plug-in download facilidade	adobe plug-in download facilidade web	adobe web jogos efeitos bom
<b>FLASH</b>	-	-	-	-	-
<b>Wikipédia</b>	-	-	-	-	-

### 3.2.2 Modificações para objetos pouco marcados pelo usuário

Ao analisar os resultados das filtragens descritas anteriormente, fica evidente que o algoritmo não consegue fazer recomendações para objetos que foram pouco marcados, ou seja, não foi possível encontrar *tags* co-ocorrentes usando os critérios originais como observado na Tabela 3.4. Para contornar essa situação, alterou-se a definição de  $T_{CO}$  para que contenha todas as *tags* com que o objeto foi marcado ( $T_O$ ), independente do usuário que a usou, menos as *tags* que o



usuário, que receberá a recomendação, usuou. Para complementar  $T_O$  quando contiver poucos elementos ( $|T_O| \leq 2$ ) foi acrescentado ao conjunto, as *tags* que co-ocorrem com  $T_O$ , assim:

$$T_O = T_O \cup \{t : t \in T \wedge \exists t' \in T_O \rightarrow CO(t, t') > 0\}, se |T_O| \leq 2 \quad (5) T_{CO} = T_O - T_{O;U}$$

Com isso, minimizou-se o problema, de não serem encontradas *tags* co-ocorrentes, uma vez que se  $|T_O| > |T_{O;U}|$ , obtem-se *tags* a serem recomendadas. Todos os testes foram repetidos com essa nova definição para  $T_{CO}$  e os resultados são apresentados nas Tabelas 3.6, 3.7, 3.8, 3.9 e 3.10.

Tabela 3.6: Variação 6 - Usar ou não Jaccard levando em conta a nova definição de  $T_{CO}$

<b>Objeto</b>	<b>Sem Alterações</b>	<b>Usando S'(t, O)</b>
<b>jQuery</b>	javascript library jquery javascript downloads utilidade	javascript library organização exemplificações jquery utilidade
<b>Android</b>	google play apps downloads developer for android google	developer for android google play apps downloads oficial
<b>Adobe</b>	adobe rapido download facilidade plug-in	rapido adobe aaae bom
<b>FLASH</b>	adobe flash programa daora livros	adobe flash programa daora backup
<b>Wikipédia</b>	win	win

Tabela 3.7: Variação 7 - Usar a filtragem de *tags* co-ocorrentes, nova definição de  $T_{CO}$ 

Objeto	Sem Alterações	Usando $T'_{CO}(n=0)$	Usando $T'_{CO}(n=1)$	Usando $T'_{CO}(n=2)$	Usando $T'_{CO}(n=3)$
<b>jQuery</b>	javascript library jquery javascript downloads utilidade	javascript library jquery javascript downloads utilidade	downloads javascript jquery dados informação	downloads javascript jquery dados informação	jquery dados informação javascrip library
<b>Android</b>	google play apps downloads developer for android google adobe	google play apps downloads developer for android google adobe	google play apps downloads google javascript adobe	google play downloads google javascript tecnologia adobe	javascript html sistema operacional tecnologia linux adobe
<b>Adobe</b>	rapido download facilidade plug-in	rapido download facilidade plug-in	videos youtube plug-in download facilidade	plug-in download facilidade jogos	jogos web download facilidade
<b>FLASH</b>	adobe flash programa daora livros	adobe flash programa daora livros	adobe flash programa daora livros	adobe flash programa daora adobe flash	adobe flash programa - -
<b>Wikipédia</b>	win	win	win	win	-

Tabela 3.8: Variação 8 - Usar ou não a filtragem nas *tags* co-ocorrentes e Jaccard na pontuação inicial, nova definição de  $T_{CO}$ 

Objeto	Sem Alterações	Usando $T'_{CO}(n=0)$ e $S'(t, O)$	Usando $T'_{CO}(n=1)$ e $S'(t, O)$	Usando $T'_{CO}(n=2)$ e $S'(t, O)$	Usando $T'_{CO}(n=3)$ e $S'(t, O)$
<b>jQuery</b>	javascript library jquery javascript downloads utilidade	javascript library organização exemplificações jquery utilidade	downloads javascript jquery library facilidade	downloads javascript jquery library facilidade	javascrip jquery php biblioteca código aberto
<b>Android</b>	google play apps downloads developer for android google	developer for android google play apps downloads oficial	google play apps google tecnologia celular	google play google tecnologia javascript downloads	javascript html sistema operacional tecnologia linux
<b>Adobe</b>	adobe rapido download facilidade plug-in	rapido adobe install aaae bom	adobe videos youtube plug-in download facilidade	adobe plug-in download facilidade web	adobe jogos web download facilidade
<b>FLASH</b>	adobe flash programa daora livros	adobe flash programa daora backup	adobe flash programa daora adobe flash	adobe flash programa daora adobe flash	adobe flash programa - -
<b>Wikipédia</b>	win	win	win	win	-

Tabela 3.9: Variação 9 - usar ou não a segunda versão da filtragem, nova definição de  $T_{CO}$ 

Objeto	Sem Alterações	Usando $T''_{CO}(n=0)$ e $S'(t, O)$	Usando $T''_{CO}(n=1)$ e $S'(t, O)$	Usando $T''_{CO}(n=2)$ e $S'(t, O)$	Usando $T''_{CO}(n=3)$ e $S'(t, O)$
<b>jQuery</b>	javascript library jquery javascript downloads utilidade	javascript library jquery javascript downloads utilidade	downloads javascript jquery dados informação	downloads javascript jquery dados informação	jquery dados informação javascript library
<b>Android</b>	google play apps downloads developer for android google	google play apps downloads developer for android google	google play apps downloads google javascript	google play downloads google javascript tecnologia	javascript html sistema operacional tecnologia linux
<b>Adobe</b>	adobe rapido download facilidade plug-in	rapido adobe install aaae bom	adobe videos youtube jogos funcional trabalhoso	adobe plug-in download facilidade jogos	adobe jogos web download facilidade
<b>FLASH</b>	adobe flash programa daora livros	adobe flash programa daora backup	adobe flash programa daora livros	adobe flash programa daora adobe flash	adobe flash programa - -
<b>Wikipédia</b>	win	win	win	win	-

Tabela 3.10: Variação 10 - Usar ou não a segunda versão da filtragem e Jaccard na pontuação inicial, nova definição de  $T_{CO}$ 

Objeto	Sem Alterações	Usando $T''_{CO}(n=0)$ e $S'(t, O)$	Usando $T''_{CO}(n=1)$ e $S'(t, O)$	Usando $T''_{CO}(n=2)$ e $S'(t, O)$	Usando $T''_{CO}(n=3)$ e $S'(t, O)$
<b>jQuery</b>	javascript library jquery javascript downloads utilidade	javascript library organização exemplificações jquery utilidade	downloads javascript jquery library facilidade	downloads javascript jquery library facilidade	javascipt jquery php biblioteca código aberto
<b>Android</b>	google play apps downloads developer for android google	developer for android google play apps downloads oficial	google play apps google tecnologia celular	google play google tecnologia javascript downloads	javascript html sistema operacional tecnologia linux
<b>Adobe</b>	adobe rapido download facilidade plug-in	rapido adobe install aaae bom	adobe videos youtube plug-in download facilidade	adobe plug-in download facilidade web	adobe jogos web download facilidade
<b>FLASH</b>	adobe flash programa daora livros	adobe flash programa daora earthquake	adobe flash programa daora adobe flash	adobe flash programa daora adobe flash	adobe flash programa - -
<b>Wikipédia</b>	win	win	win	win	-

### 3.2.3 Algoritmo de Recomendação Adotado

Analisando-se os resultados de todas as variações apresentados na tabelas, verifica-se que as melhores recomendações são feitas quando se usa  $S(t, O)$  e  $T_{CO}(n = 2)$ . Contudo quando foi redefinido  $T_{CO}$  e usada essa modificação, obtem-se uma melhora nas recomendações de objetos com poucas marcações. Porém, percebe-se também uma queda nas recomendações dos outros objetos. *Tags* que antes estavam na primeira posição devido a boa pontuação, passaram a aparecer mais abaixo sendo substituída por uma *tag* mais genérica. Considerando tal problema, o funcionamento do algoritmo foi dividido em duas partes: (1) Num primeiro momento, encontra-se as *tags* co-ocorrentes da forma original e se segue normalmente o restante do algoritmo, usando-se  $S'(t, O)$  e  $T''CO(n = 2)$ ; (2) Caso essa busca falhe, então faz-se outra busca por *tags* co-ocorrentes, usando-se a outra definição de  $T''CO$  conforme a Tabela 3.10 (**Varição 10**). O algoritmo final adotado pelo módulo de recomendação é descrito na Tabela 3.11.

Tabela 3.11: Algoritmo Adotado pelo Módulo de Recomendação

01.  $T_{CO} = \{t : t \in T \wedge t \notin T_{O;U} \wedge \exists t' \in T_{O;U} \rightarrow CO(t, t') > 0\}$
02. Se  $|T_{CO}| = \emptyset$  Então
03.     Se  $|T_O| \leq 2$  Então
04.          $T_O = T_O \cup \{t : t \in T \wedge \exists t' \in T_O \rightarrow CO(t, t') > 0\}$
05.     Fim Se
06.      $T_{CO} = T_O - T_{O;U}$
07. Fim Se
08.  $T''_{CO} = \{t : t \in T_{CO} \wedge \forall t' \in T_{O;U} \rightarrow CO(t, t') > 2\}$
09.  $\forall t \in T''_{CO}, S'(t, O) = \frac{U(t)}{Q(t)} X \frac{1}{|T_{O;U}|} X \sum_{t' \in T_{O;U}} J(O_t, O_{t'})$
10. Ordene decrescentemente as tags de  $T''_{CO}$  de acordo com  $S$ .
11. Assim  $T_{CO}[0]$  é a tag com maior valor de  $S'(t, O)$
12.  $t_e = T''_{CO}[0]$
13. Tags Escolhidas =  $\{t_e\}$
14.  $T_{CO} = T_{CO} - t_e$
15.  $i = 1$
16. Enquanto  $i < n \wedge T_{CO} \neq \emptyset$  faça:
17.      $\forall t \in T''_{CO}, S'(t, O) = P(t|t_e; U) * S(t_e, O) + P(t|t_e; O) * S'(t_e, O)$
18.     Ordene decrescentemente as tags de  $T''_{CO}$  de acordo com  $S$ .
19.     Assim  $T''_{CO}[0]$  é a tag com maior valor de  $S'(t, O)$
20.      $t_e = T''_{CO}[0]$
21.     Tags Escolhidas = Tags Escolhidas  $\cup t_e$
22.      $T_{CO} = T_{CO} - t_e$
23.      $i = i + 1$
24. Fim Enquanto
25. Retornar Tags Escolhidas

**linha 01:** é definido o conjunto de tags co-ocorrentes ( $T_{CO}$ ) às tags do usuário ( $U$ ), considerando todas as tags do repositório ( $T$ ).

**linha 02:** caso o conjunto  $T_{CO}$  for vazio, será verificado na **linha 03** se a quantidade de tags do objeto é  $\leq 2$ . **linha 04:** Se sim, acrescenta-se ao conjunto  $T_O$  tags que sejam co-ocorrentes com ao menos uma tag de  $T_O$ .

**linha 06:** são eliminadas do conjunto  $T_O$ , as tags que o usuário  $U$  utilizou para o objeto  $O$ .

**linha 08:** o conjunto  $T''_{CO}$  é gerado a partir das tags de  $T_O$  que tenham co-ocorrência em mais de 2 objetos.

**linha 9:** é calculada a pontuação de cada tag do conjunto  $T''_{CO}$ , dada por  $\frac{U(t)}{Q(t)}$  e aplicando-se a somatória do coeficiente de Jaccard ( $X \sum_{t' \in T_{O;U}} J(O_t, O_{t'})$ ), onde quanto mais próximo do valor 1, mais específica é a tag  $t$  para o usuário  $U$ .

**linha 10:** o conjunto  $T''_{CO}$ , é ordenado na ordem decrescente

**linha 12:** obtém-se a tag de maior pontuação ( $T''_{CO}[0]$ )

**linha 13:** a tag de maior pontuação é escolhida ( $t_e$ ) e associada ao conjunto "Tagsescolhidas".

**linha 14:** a tag  $t_e$  é eliminada do conjunto  $T_{CO}$ .

**linha 15:** a variável  $i$  é inicializada com 1, para que o processo seja repetido por  $n$  vezes ou até que o conjunto  $T''_{CO}$  esteja vazio (**linha 16**).

**linha 17:** a pontuação é recalculada para todas as tags restantes no conjunto  $T''_{CO}$ , recompensando tags mais específicas para o usuário  $U$ , através da expressão  $P(t|t_e; U) * S'(t_e, O)$  e penalizando tags redundantes com a expressão  $(+P(t|t_e; O) * S'(t_e, O))$ .

Dessa forma, as tags de maior pontuação são selecionadas e atribuídas ao conjunto "Tagsescolhidas"

**linha 25:** ao final, o conjunto "Tagsescolhidas" é retornado.

### 3.2.4 TagLink: Protótipo Funcional

Para tornar possível a realização dos experimentos necessários para a validação da proposta deste trabalho, foi desenvolvido um aplicativo *Web* chamado TagLink<sup>1</sup>. Essa ferramenta foi criada pelo próprio grupo de pesquisa pertencente ao LERIS da UFSCAR. O TagLink permite buscar recursos de aprendizagem na *Web* utilizando a API da Google Custom Search<sup>2</sup> e criar ou associar *tags* aos recursos. O usuário informa o termo de busca de um recurso e a ferramenta lista os objetos e seus respectivos *links*. Ao acessar um *link* é possível visualizar o respectivo *site*. Cada *link* é tratado como um recurso de aprendizagem e pode receber marcações atribuídas pelos alunos.

A Figura 3.2 apresenta a página de pesquisa do TagLink. Nesta página, o aluno pode informar o termo a ser pesquisado, além de poder iniciar uma pesquisa através dos globos que apresentam nuvens de *tags* mais relevantes.

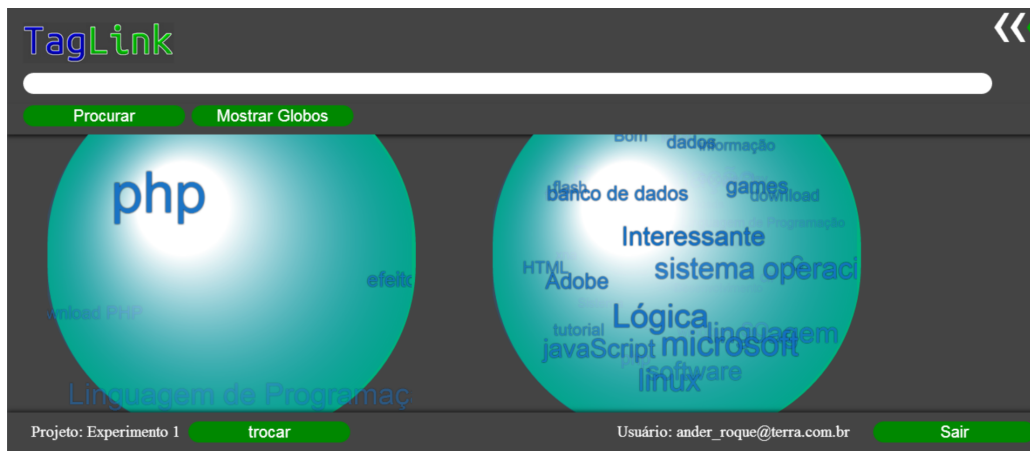


Figura 3.2: Página de Pesquisa do TagLink

Um ponto importante da API de busca utilizada é que ela permite priorizar quais são os *sites* de busca para um termo. Isto permitiu que fosse cadastrada a partir da API um conjunto de *sites* considerados relevantes ao público que iria participar do experimento. Durante o projeto foram usadas duas versões do TagLink. Na primeira, o TagLink foi configurado para permitir que os usuários pudessem criar *tags* e reusar *tags* já criadas pelo próprio usuário e associadas a outros objetos, e criadas por outros usuários e associadas ao objeto selecionado. A Figura 3.3 apresenta a página que lista as descrições dos recursos encontrados a partir do termo "Lógica de Programação".

<sup>1</sup><http://200.133.238.124/tag/>

<sup>2</sup><https://developers.google.com/custom-search/?hl=en>



Figura 3.3: Página com lista de recursos encontrados pelo TagLink

O TagLink possui um controle de usuários por *login* e senha e registra o grupo a que o aluno pertence. A ferramenta permite realizar verificações diversas relacionadas aos objetos, usuários e *tags* utilizadas: objetos que foram marcados e com quais *tags*, quantas vezes um objeto foi marcado com a mesma *tag*, registro da data de criação de uma *tag*, quantas vezes cada *tag* foi utilizada, etc.

A Figura 3.4 apresenta a página de criação e associação de *tags*. Através desta interface, o aluno pode atribuir uma *tag* recomendada, listada no controle *dropdown*, ou digitar uma nova *tag* para o recurso selecionado e analisado.

Figura 3.4: Página para atribuição de *tags* do TagLink

A segunda versão do TagLink foi concebida para apoiar os Experimentos 2 e 3. Nesta versão, passou-se a considerar a participação do professor que ao cadastrar um projeto de catalogação, define palavras-chave que direcionarão o processo de marcação. A Figura 3.5 apresenta a página de entrada para o projeto de catalogação. Além disso, o módulo de recomendação de *tags* foi incorporado ao TagLink com a implementação do algoritmo definido na Tabela 3.2.3.

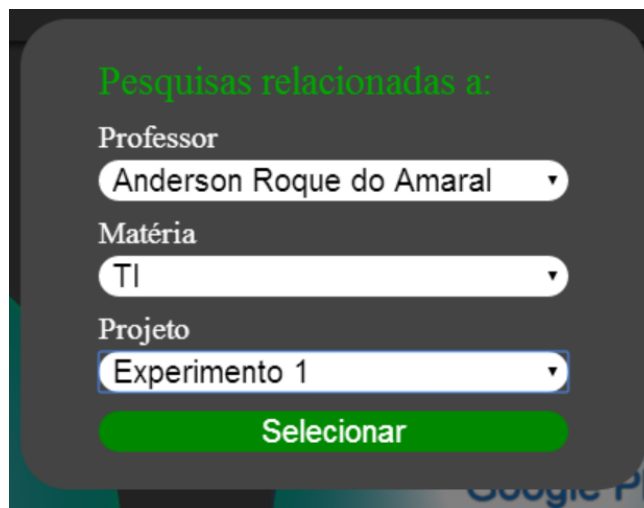


Figura 3.5: Página para cadastro do projeto do professor pelo TagLink

O principal objetivo em se criar 2 versões foi para poder realizar experimentos distintos, com e sem o módulo de recomendação. Tais experimentos são descritos no Capítulo 4.



### 3.3 Considerações Finais

Diante da necessidade de uma catalogação, classificação e indexação de recursos disponíveis na *Web* de maneira eficaz, principalmente para apoiar o processo de ensino-aprendizagem num contexto de *e-Learning*, foi proposto um modelo de recomendação descrito na Seção 3.1 chamado e-TagReS, que considera não apenas o algoritmo de recomendação (Seção 3.2.3), mas também o ciclo de recomendação necessário para apoiar a marcação social em ambiente educacional. Para a escolha do algoritmo a ser adotado, testes foram realizados para se chegar na melhor formulação conforme descrito na Seção 3.2.1. O TagLink foi desenvolvido em duas versões para que fosse possível a condução dos experimentos (Seção 4). A principal diferença entre as versões foi o módulo de recomendação, para permitir a realização de comparações entre os experimentos. Assim, o sistema de recomendação de *tags* proposto ficou estruturado para apoiar a catalogação de recursos de aprendizagem, onde o professor participa, colaborando com temas definidos por ele, que podem direcionar a atividade de marcação por parte dos alunos (2ª versão).

## Experimentação Preliminar e Validação da Proposta

A marcação social depende essencialmente da participação de usuários que interagem com o sistema, contribuindo com suas opiniões, manifestadas por meio de *tags* e atribuídas a recursos. Esse princípio, permanece válido quando a marcação colaborativa é aplicada num contexto de aprendizagem. Assim, toda a experimentação para se validar a proposta desse trabalho, foi planejada e conduzida de forma a permitir o acompanhamento do comportamento dos alunos participantes na atividade de marcação, a análise das *tags* por eles fornecidas, verificar o vocabulário gerado em torno dos recursos e observar o quanto o modelo de recomendação de *tags* proposto pode resultar em benefícios para processos de catalogação que tenham o *e-learning* como meio para favorecer a aprendizagem. Para isso ser possível, três experimentos foram organizados: (1) Marcação social sem a utilização de recomendação de *tags*, que possibilitou a análise do comportamento dos alunos no processo; (2) Marcação social utilizando o modelo de recomendação de *tags* proposto; (3) Navegação e busca de recursos, que possibilitou comparar resultados entre os experimentos (1) e (2). Os experimentos serão descritos nas próximas seções. No início de cada seção é apresentada uma ficha técnica do experimento, contendo informações relevantes sobre o mesmo.

## 4.1 Experimento 1: Catalogação de Recursos sem Recomendação

Tabela 4.1: Ficha Técnica do Experimento 1

Objetivo:	Catalogação de Recursos sem recomendação
Metodologia:	<ul style="list-style-type: none"> <li>- Os alunos foram divididos em dois grupos: Grupo A e Grupo B</li> <li>- Cada grupo de alunos recebeu 5 termos para pesquisa</li> <li>- Condução em duas Fases com inversão dos termos entre os grupos</li> <li>- Em cada fase os alunos catalogaram recursos criando e reutilizando <i>Tags</i></li> <li>- O TagLink utilizado como coletor de <i>tags</i></li> </ul>
Local:	ETEC Fernando Prestes
Período:	Fevereiro/2013 à Março/2013
Nº de alunos	336

Esse primeiro experimento teve como objetivo observar o comportamento dos estudantes em relação à associação de *tags* a recursos de aprendizagem. Assim, esse experimento foi norteado por três questões:

1. A marcação social pode catalogar com sucesso recursos de *e-learning*? Com o objetivo de verificar se esta técnica de catalogação é útil para organizar recursos de aprendizagem.
2. Como os alunos se comportam de acordo com a classificação de Korner: Categorizadores ou descritores?
3. A marcação social converge para um vocabulário descritivo e bem definido pelas *tags*?

Foi realizado um grande experimento com estudantes de cursos técnicos que marcaram recursos de aprendizagem, gerando cerca de 4.985 marcações. Os resultados desse experimento mostram que a marcação social é uma prática promissora para *e-learning*. No entanto, algumas questões devem ser abordadas para evitar um grande número de estudantes categorizadores e, também, uma convergência prematura do vocabulário de *tags*. As conclusões são específicas para a definição do experimento, mas foram generalizadas, tanto quanto possível, para se sugerir diretrizes de como usar a marcação social em *e-learning*. Nesse primeiro experimento, concebeu-se o sistema de marcação como um tripé feito de estudantes, *tags* e recursos de aprendizagem que, em conjunto, interagem para formar um vocabulário descritivo. A Figura 4.1 apresenta esse tripé, que começa a partir da pesquisa e recuperação de recursos de aprendizagem (1); procede-se à criação e reutilização de *tags* (2); e evolui através do refinamento alcançado pela associação de *tags* para os recursos(3). Estes três passos do ciclo ocorrem por tempo indeterminado, durante o qual um repositório de *tags* é construído. O repositório está organizado em dois conjuntos: "minhas *tags*", que são as *tags* criadas por um usuário específico, registrado no sistema; "*tags* globais", o universo de todas as *tags* distintas criadas por todos os usuários para os diversos recursos. Ao longo do tempo, algumas *tags* são recorrentemente usadas para descrever os recursos de aprendizagem. A convergência das *tags* a um conjunto descritivo estável define um vocabulário (4).

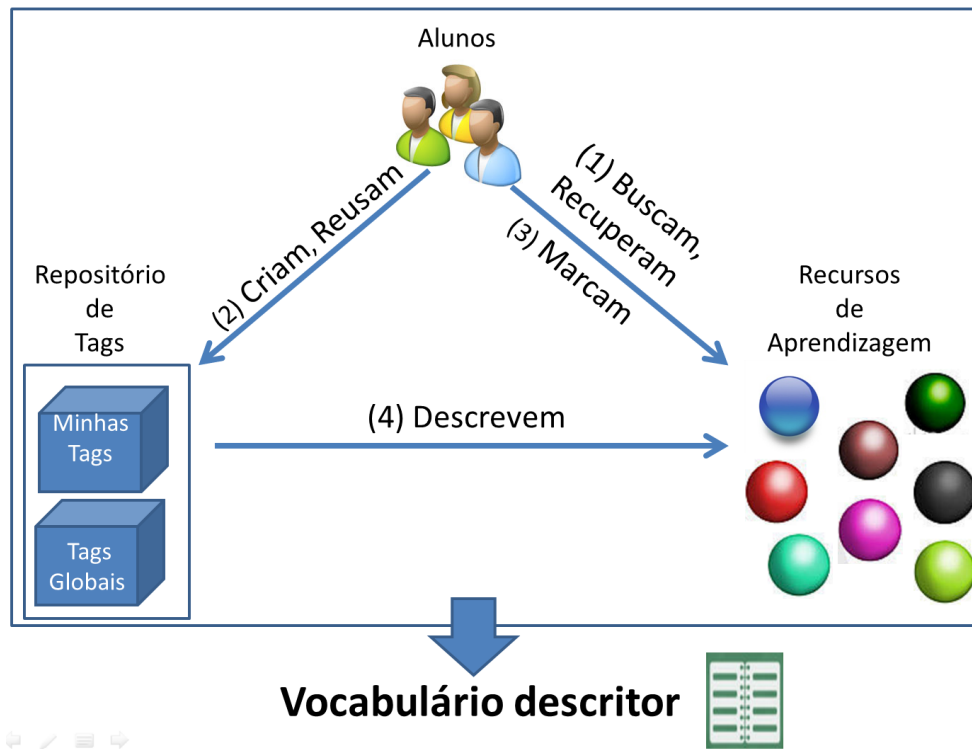


Figura 4.1: Os elementos que constituem o tripé da marcação social concebida para o Experimento 1: Usuários (estudantes), recursos de aprendizagem e *tags*

Percebeu-se que os trabalhos anteriores encontrados e descritos na Seção 2.2.1, não oferecem um ambiente adequado (*open source* e acessível) para a experimentação de acordo com a metodologia proposta e para responder as perguntas desse experimento. Por esta razão, foi necessário projetar e desenvolver o sistema TagLink, que é capaz de (1) recuperar recursos de aprendizagem, (2) criar/reutilizar *tags*, (3) apoiar a marcação de recursos, e (4) gerir um vocabulário descritivo de *tags*. O TagLink é detalhado na seção 3.2.4.

#### 4.1.1 Planejamento

Uma pesquisa preliminar revelou que a maioria dos estudantes que participariam do experimento era usuário regular de redes sociais, que estavam familiarizados com a marcação, mas nunca tinham usado essa funcionalidade para fins educacionais. Os resultados desta pesquisa preliminar levou a necessidade de preparar os estudantes antes de participarem do processo. Assim, para evitar problemas com o uso do *TagLink*, elaborou-se um tutorial detalhando os passos para se cadastrar na ferramenta, efetuar a pesquisa dos termos e atribuir *tags* aos recursos. Esse tutorial foi enviado a todos os alunos participantes.

Os estudantes foram separados em dois grupos: Grupo A, composto pelos alunos mais adultos, e Grupo B, composto pelos alunos adolescentes e mais jovens. Essa separação por faixa etária, teve como objetivo validar o quanto a idade e o conhecimento prévio dos alunos e sua relação com marcação social iria influenciar no processo, uma vez que os alunos do Grupo B tinham mais familiaridade com a atividade de marcação. Foram definidos dois conjuntos de

conceitos utilizados como ponto inicial para pesquisa. **Ta** e **Tb**, ambos com 5 termos, com o objetivo de definir um ponto inicial para a pesquisa de cada grupo, o que se faz necessário em sistemas de marcação para se evitar o problema de "partida a frio" (ZIESEMER, 2012). O primeiro conjunto, **Ta**, continha termos genéricos para a tecnologia da informação, **Ta**={lógica de programação, C#, Bancos de Dados, Windows, Linux}. O segundo conjunto, **Tb**, continha termos de programação *web*, **Tb**={jQuery, PHP, XNA, Android, Flash}. O experimento foi planejado para ser executado em duas fases: na Fase 1, o Grupo A usou o conjunto **Ta** e Grupo B usou o conjunto **Tb**; na Fase 2, o Grupo A usou o conjunto **Tb** e Grupo B usou o conjunto **Ta**.

Elaborou-se a orientação aos alunos sobre como pesquisar na *web*, usando o sistema TagLink, buscando cada termo do conjunto a eles atribuído. Na Fase 1, os alunos tiveram a missão de selecionar pelo menos um recurso de aprendizagem de cada resultado da pesquisa, e criar três ou mais *tags* para este recurso. Na Fase 2, os alunos foram orientados quanto a possibilidade de reutilizar *tags* existentes, seja própria ou de outros estudantes, além de criarem mais *tags* se quisessem. O objetivo era que os alunos construíssem um vocabulário de *tags*, criando *tags* na Fase 1, e podendo reutilizar ou criá-las na Fase 2.

#### 4.1.2 Execução

Realizou-se um experimento com estudantes do nível do ensino profissional (técnico em tecnologia da informação) da ETEC Fernando Prestes em Sorocaba. Os alunos foram instruídos a procurar e recuperar recursos de aprendizagem relacionados com os seus cursos atuais e para marcar esses recursos usando as técnicas de marcação social. A atividade dos alunos foi gravada para análise, como será relatado nas próximas seções.

Para envolver os alunos no processo e conscientizá-los a respeito de marcação social, foi feito um trabalho de divulgação do experimento, abordando os conceitos relacionados com a marcação colaborativa e recursos de aprendizagem. Conforme o planejamento, foram distribuídos panfletos aos alunos, descrevendo os objetivos, a forma que seria a condução da coleta de *tags* e os períodos referente a Fase 1 do experimento. Depois, cada grupo de alunos recebeu os 5 termos que seriam pesquisados na Fase 1: o **Grupo A** recebeu o **Ta** e o **Grupo B** o **Tb**.

Iniciou-se então a condução para coleta de *tags* (Fase 1), conforme o planejamento realizado. Os alunos estavam cientes que deveriam obrigatoriamente criar pelo menos três *tags* para um determinado recurso. Além das *tags* criadas, eles sabiam também que poderiam associar outras *tags* criadas por outros alunos. Durante a condução houve dúvidas por parte dos alunos. Dentre as dúvidas podem ser destacadas as apresentadas na Tabela 4.2 incluindo as respostas dadas aos alunos.

Tabela 4.2: Dúvidas geradas durante a condução do experimento

Perguntas dos alunos	Resposta do condutor do experimento
Na Fase 1 é mandatória a criação de <i>tags</i> inéditas?	Sim. Mas também pode haver reuso de <i>tags</i> .
Deve-se usar como <i>tags</i> apenas palavras que fazem parte do conteúdo do recurso selecionado durante a pesquisa	Não necessariamente, mas poderiam usar.
Existe a liberdade de criar <i>tags</i> segundo minha opinião particular sobre o conteúdo?	Sim.
É possível usar palavras compostas para as <i>tags</i> ?	Sim.

Deve-se observar que a liberdade de vinculação de *tags* foi dada aos alunos para haver subsídios para responder às questões Q2 e Q3, e para verificar a criação e estabilização do vocabulário utilizado no processo de marcação.

Na terceira semana teve início a Fase 2 do experimento onde houve a troca de termos para pesquisa. O **Grupo A** recebeu os **Tb** e o **Grupo B** o **Ta**. Nesta fase, a orientação dada aos alunos era que eles tinham a liberdade de atribuir uma *tag*, escolher *tags* já atribuídas ao recurso por outros alunos, escolher uma *tag* criada por ele mesmo ao marcar outro recurso na fase anterior, ou criar uma nova *tag* para o recurso. O objetivo era dar um maior grau de liberdade para verificar se o comportamento dos alunos eram de criadores ou classificadores. E também, certificar-se sobre a evolução do vocabulário usado para os recursos.

O experimento teve a duração de 2 meses. No primeiro mês, foi realizada a Fase 1; e no segundo mês, a Fase 2. Ao final do experimento, obteve-se um total de **2019** *tags* para **218** recursos selecionados por **336** alunos cadastrados no aplicativo TagLink que geraram 4.985 marcações.

### 4.1.3 Análise dos resultados

Os dados coletados durante o experimento foram analisados e serão reportados a partir das questões de pesquisa.

#### (1) A marcação social pode catalogar com sucesso recursos de *e-learning*?

Para esta questão, assumiu-se que quanto maior fosse o número de vezes e quanto mais alunos participassem do experimento, seria obtido um conjunto mais adequado de *tags* para uma análise. Analisaram-se dois aspectos: o número de vezes e de estudantes que criaram novas *tags* - Figura 4.2; e o número de vezes e de estudantes que reutilizaram *tags* já existentes - Figura 4.3.

Verificou-se que 1.279 das 2.019 *tags* (63,34%) foram utilizadas apenas uma vez, ao serem criadas; das outras, 740 *tags* (36,65%) foram reutilizadas. A Tabela 4.3 resume como os alunos utilizaram o sistema em relação a criação e a reutilização de *tags*. O conceito de distribuição gaussiana foi adotado para se analisar os gráficos das marcações feitas pelos alunos. A função gaussiana é amplamente utilizada para se analisar estatísticas e probabilidades. A principal característica das curvas de Gauss (ou curva normal) é que o "pico" da curva sempre é formado na média da distribuição de forma que os desvios à esquerda e à direita da média, mostram variações que subestimam ou superestimam a média movendo-se a um "vale". Dessa forma

considera-se que qualquer ocorrência casual de informação, quando observada por um longo período de tempo, geram resultados que tendem a se aproximar da média.

Na Figura 4.2, podemos verificar uma primeira concentração nos dados definindo uma distribuição Gaussiana-Pico em torno de 2 *tags* por aluno; e uma segunda concentração principal a definição de um centro de distribuição Gaussiana-Suave em torno de 12 *tags* por aluno. A primeira concentração era esperada, uma vez que a participação no experimento era opcional. Mesmo assim, uma fração significativa dos alunos criaram mais de 3 *tags*. A segunda concentração principal revelou que 4 vezes mais estudantes criaram a partir de 4 até 18 *tags* cada, o que mostra um bom envolvimento dos alunos no processo. Os participantes da segunda concentração principal criaram 2.420 *tags*, cerca de 11 criações de *tags* por recurso de aprendizagem.

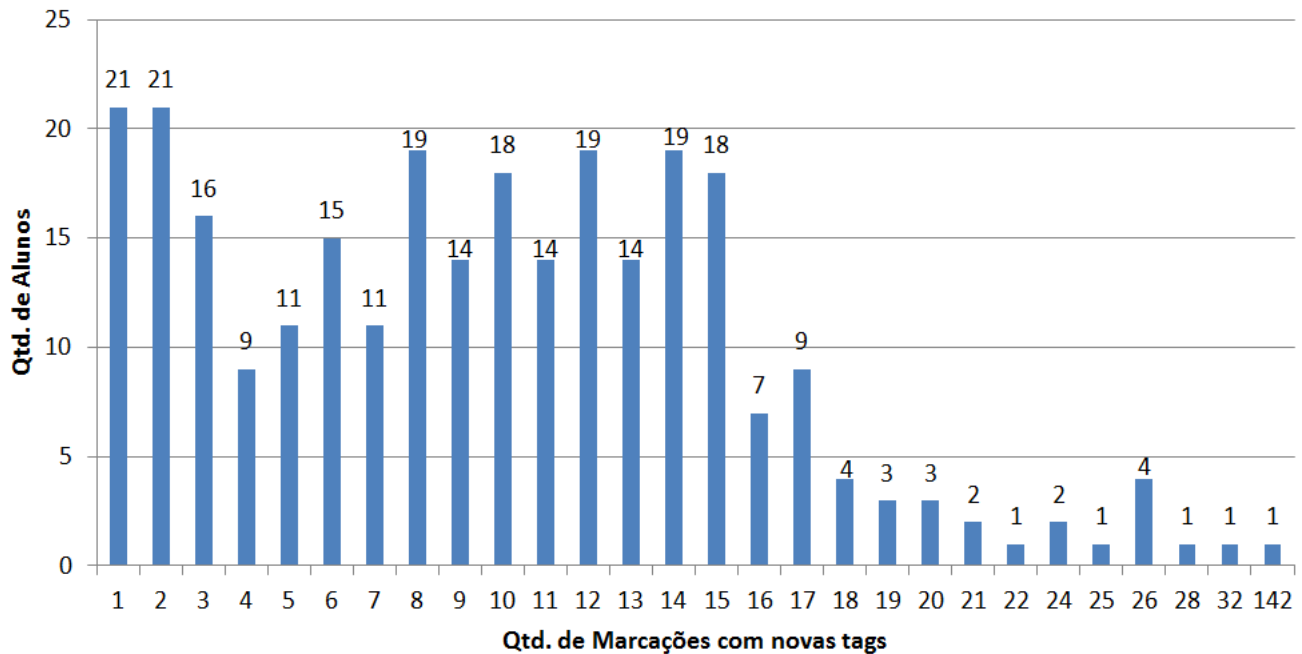


Figura 4.2: Distribuição do número de alunos por número de *tags* criadas.

Na Figura 4.3, verifica-se dois picos, um em torno de 5 *tags* reutilizadas por aluno, e outra cerca de 15 *tags* reutilizadas por aluno. A distribuição agora é deslocada em relação à distribuição de novas *tags* por aluno (Figura 4.2). No lugar em que houve um pico, agora houve um vale, indicando um aumento na participação dos alunos que não criaram as *tags* antes - mais à esquerda na distribuição. Além disso, um pequeno conjunto de estudantes, apresentou um aumento de marcações por reutilização em comparação com a criação, mais a direita na distribuição. A partir dos gráficos observa-se três comportamentos dos alunos, derivados da Tabela 4.3: um conjunto de alunos que apenas criaram novas *tags* (11, 3%), um conjunto com os alunos que apenas reutilizaram *tags* (8, 9%), e um conjunto de estudantes que fizeram as duas coisas (79, 7%). Na tabela, as colunas **C** referem-se a quantidade de criações e as colunas **R** referem-se a quantidade de reutilizações em cada fase.

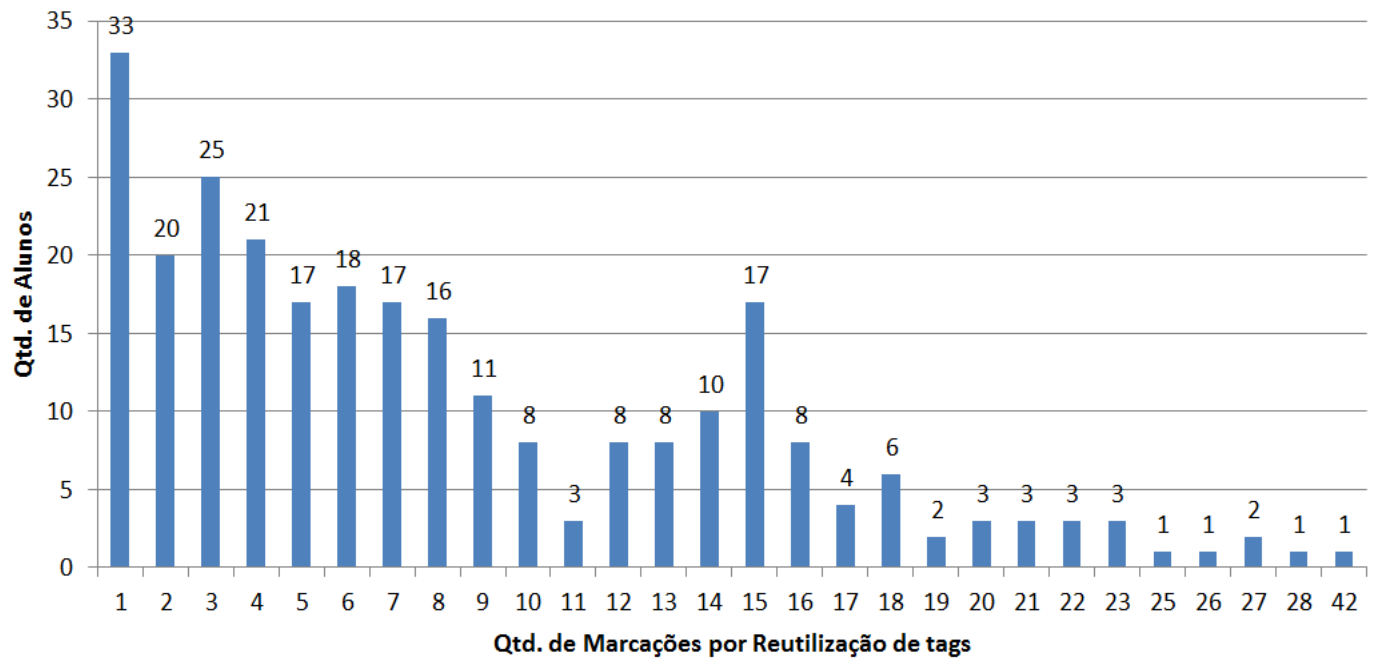


Figura 4.3: Distribuição do número de alunos por número de tags reutilizadas.

Ao considerar os gráficos das Figuras 4.2 e 4.3, pode-se afirmar que, dentro dos limites do experimento, os alunos participaram de forma satisfatória no processo de marcação, definindo *tags* suficientes para a descrição e recuperação final de recursos. Sugere-se, com evidência significativa, que a marcação social pode ser empregada com sucesso na marcação de recursos de aprendizagem.

Tabela 4.3: Sumarização da Coleta de Tags

Grupo	Fase 1		Fase 2		Totais	
	Criações	Reutilizações	Criações	Reutilizações	Criações	Reutilizações
A	1126	648	144	303	1270	951
B	1329	933	139	363	1468	1296
Totais	2455	1581	283	666	2738	2247

## (2) Como os alunos se comportam de acordo com a classificação de Körner: categorizadores ou descritores?

Para esta questão, analisou-se as *tags* mais reutilizadas pelos alunos. A Tabela 4.4 apresenta estas *tags*. É possível perceber que, com exceção da *tag* "interessante", as *tags* são de significado geral do senso comum, ou seja, mais voltadas para os conteúdos dos recursos do que para um parecer particular, uma vez que o que é interessante para alguns, não é necessariamente para todos. Este fato indica que os alunos tiveram um perfil descritor, ao invés de um perfil categorizador. Korner (2009b) define que usuários com perfil categorizador fornecem *tags* mais subjetivas que favorecem a navegação por vários objetos, enquanto que usuários com perfil descritor fornecem *tags* mais objetivas que descrevem os recursos, favorecendo assim a recuperação dos mesmos. Esta é uma evidência importante porque reforça que a marcação social em *e-learning* tende



a produzir catálogos para uso geral, em vez de catálogos de informação personalizada. Isto é importante para os seguintes aspectos:

- os catálogos são propensos a serem reutilizáveis por estudantes em cursos posteriores, que irão utilizar termos gerais para pesquisar documentos;
- os catálogos podem ser indexados de forma mais eficaz, porque a recorrência de *tags* pode gerar índices guiados por relevância;
- os recursos de aprendizagem podem ser agrupados (*cluster*) utilizando a análise de termo-frequência, descartando *tags* menos significativas e específicas;
- com *tags gerais* do senso comum, é possível combinar o sistema de busca de *e-learning* com os motores de busca (comerciais educacionais e de uso geral), melhorando a recuperação de conteúdo com recursos complementares.

Conclui-se que, embora os alunos tendam a ser categorizadores na sua atividade de marcação social, marcação que eles vêm com percepções e sentimentos pessoais, eles são mais propensos a se comportar como descritores quando a marcação social ocorre no domínio educativo. Por isso, pode-se concluir, que os alunos se comportam de acordo com o perfil descritor definido por Körner.

Tabela 4.4: *Tags* reutilizadas por cada grupo em cada fase do experimento

Tag x Grupo x Fase					
tag	Fase 1		Fase 2		Total
	Grupo A	Grupo B	Grupo A	Grupo B	
Programação	30	7	6	11	54
microsoft	20	17	6	8	51
linux	23	13	2	2	40
Lógica	27	0	1	5	33
software	13	10	7	10	40
linguagem	16	19	3	9	47
Sistema Operacional	14	10	7	10	41
javaScript	0	28	9	0	37
SO	10	3	2	9	24
Interessante	1	19	3	0	23
Total	154	126	46	64	390

### (3) A marcação social converge para um vocabulário descritivo e bem definido *detags*?

Para esta questão, foi necessário analisar o número de novas *tags* criadas no sistema ao longo do tempo do experimento (dois meses). Na Figura 4.4, podemos ver que o número de novas *tags* se comporta de acordo com uma distribuição normal com um pico próximo ao meio do período

- no 28º dia. A distribuição normal sugere que após o pico, os alunos passam a criar apenas algumas novas *tags*; um número que tende a zero à medida que se distância do pico.

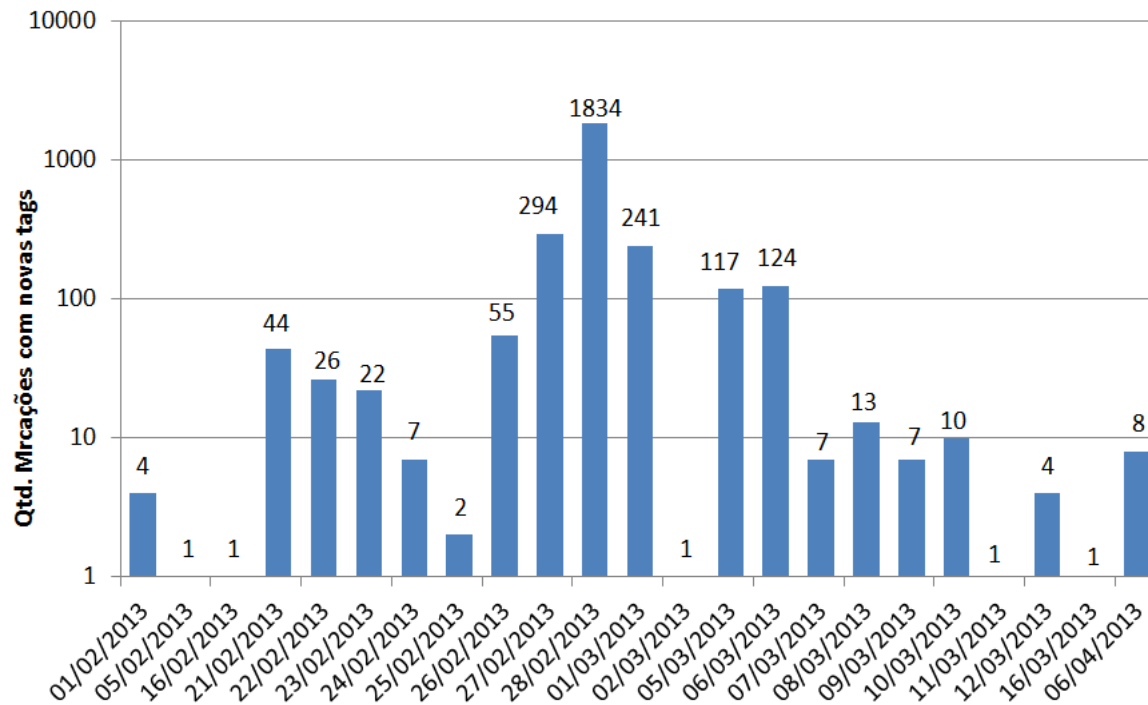


Figura 4.4: Distribuição do número de novas etiquetas criadas ao longo do tempo.

A distribuição Normal, enquanto válida, não é tão forte como seria uma distribuição de lei de potência descendente onde o pico ocorreria no início e uma "cauda longa" se estenderia para a direita; no entanto, este fato também é importante. Por quê o processo não se comportou como uma lei de potência em que a maioria das novas *tags* seriam criadas no início do período? A resposta é bastante simples quando se considera que o experimento utiliza uma interface homem-computador. Nesses sistemas, o usuário passa por uma curva de aprendizado com três fases: um início com pouca interação, um momento de interação mais efetiva e finalmente uma fase de estabilização. Verificou-se, portanto, que o lado esquerdo da distribuição normal, até quase 20 dias, foi afetado pelo período de aprendizagem e que, só depois, os usuários foram capazes de interagir plenamente com o TagLink e demonstrar seus perfis de marcação. Enquanto a Figura 4.4 mostra que as *tags* tendem a se estabilizar após o 28º dia, na Figura 4.5 pode-se observar que um pequeno conjunto de *tags* se destaca no uso do sistema. Mais precisamente, foram 2.019 *tags* diferentes, entre elas, 1.972 foram utilizadas menos que 17 vezes, não apresentadas na figura e apenas 50 foram utilizadas 17 vezes ou mais, apresentado na figura. Estas características descrevem uma forte distribuição de "cauda longa" indicando um forte desequilíbrio em como os alunos se concentram em um pequeno subconjunto de *tags*. Ao comparar este conjunto dominante com o conjunto de termos originais fornecidos no início do experimento (Seção 4.1.1), é possível observar um grande cruzamento. Este fato indica que os termos iniciais influenciaram fortemente o vocabulário e que, possivelmente, isso acelerou o processo, tal como sugerido em outros trabalhos (ZERVAS; SAMPSON, 2014) (STROHMAIER; KORNER; KERN, 2012). Portanto, com base nas evidências das Figuras 4.4 e 4.5, pode-se afirmar que a marcação social

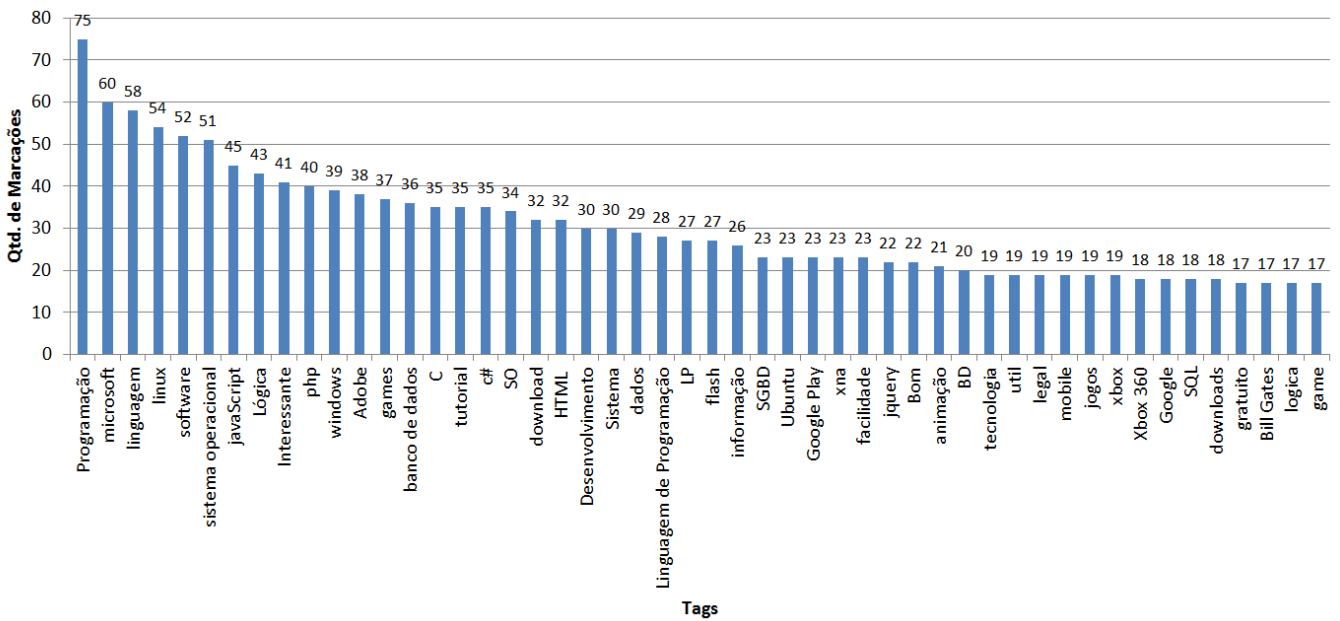


Figura 4.5: Distribuição do número de vezes que cada *tag* foi usada (50 mais usadas)

pode convergir para um conjunto bem definido de *tags*. Também, é possível argumentar que um conjunto adequado de termos de origem podem fornecer algum controle sobre esse processo, influenciando a definição do conjunto da maioria das *tags* mais frequentes e, conseqüentemente, influenciando o quanto serão descritivas. Este conjunto de termos iniciais é importante quando considera-se um cenário de aprendizagem. É uma direção que o professor provê para que os alunos possam explorar positivamente as marcações de recursos relevantes ao contexto do estudo.

## 4.2 Experimento 2: Catalogação de Recursos com Recomendação de Tags

Tabela 4.5: Ficha Técnica do Experimento 2

Objetivo:	Catalogação de Recursos com recomendação de <i>tags</i>
Metodologia:	<ul style="list-style-type: none"> <li>- Os alunos foram divididos em dois grupos: Grupo A e Grupo B</li> <li>- Cada grupo de alunos recebeu 5 termos para pesquisa</li> <li>- O professor incluiu um projeto com 5 <i>tags</i></li> <li>- Os alunos recebem recomendação de <i>tags</i> podendo aceitar ou não</li> <li>- O TagLink foi utilizado com a recomendação de <i>tags</i></li> </ul>
Local:	ETEC Fernando Prestes
Período:	Abril/2014
Nº de alunos	212

A técnica de marcação social conforme análise do Experimento 1, mostrou-se adequada para a descrição de recursos de aprendizagem de forma colaborativa pelos alunos. A partir do Experimento 1, percebeu-se que a utilização de mecanismos de recomendação de *tags* poderia auxiliar

os participantes do processo na marcação. Isso foi observado principalmente ao verificar-se que o índice de reutilização de *tags* do próprio aluno e de outros alunos, foi significativo. Dessa forma, planejou-se um novo experimento que seria apoiado por um mecanismo de recomendação de *tags* incorporado a ferramenta TagLink. Para avaliar este experimento, usaram-se as questões de pesquisa definidas no Experimento 1, pois um dos objetivos era justamente verificar se o sistema de recomendação proposto na Seção 3.1 é adequado para apoiar a catalogação de recursos de aprendizagem. Entretanto buscou-se responder outras questões:

Um mecanismo de recomendação de *tags*,

1. Auxilia a catalogação de recursos de aprendizagem?
2. Favorece mais a descrição ou a categorização dos recursos de aprendizagem por parte dos alunos?
3. Propicia um vocabulário em torno dos recursos de aprendizagem mais específico?
4. O algoritmo de recomendação proposto é eficaz?
5. O modelo de ciclo de recomendação proposto (Seção 3.1) pode ser considerado válido?

A seguir é apresentado o planejamento, a execução e análise dos resultados do Experimento 2.

### 4.2.1 Planejamento

Os alunos que participaram do Experimento 2, não foram os mesmos participantes do Experimento 1, porém possuíam o mesmo perfil quanto a faixa etária e quanto a experiência dos mesmos na utilização de sistemas de marcação social na área educacional. Considerou-se interessante, que os participantes do novo experimento tivessem acesso a recomendações baseadas em marcações geradas em outro tempo e por de outros estudantes, no Experimento 1. Sendo assim, também foi necessário fornecer orientações aos alunos sobre os conceitos e os reais objetivos do experimento. O módulo de recomendação de *tags* incorporado ao TagLink, teve como base um algoritmo que utiliza medidas de co-ocorrência e medidas de recompensa-punição que evitam *tags spam* e ajusta de forma interativa a recomendação (Tabela 3.2.3).

O Experimento 2 teria uma etapa de catalogação utilizando-se do sistema de recomendação implementado no TagLink. Essa catalogação seria avaliada por outra etapa do experimento onde se enfatizaria a navegação e pesquisa dos recursos catalogados. Como planejamento do experimento, foram seguidos os seguintes passos:

1. Semelhante ao Experimento 1, definiu-se que o professor incluiria um projeto fornecendo algumas palavras-chave(ou termos de origem para pesquisa). Essas palavras-chave seriam sempre oferecidas como recomendação durante a catalogação dos recursos de aprendizagem, para suprir as primeiras recomendações e marcações. Os alunos também foram divididos em dois grupos: Grupo A e Grupo B, sendo que cada grupo foi formado por 6

turmas. Determinou-se como nome do projeto "Experimento2: TagLink com Recomendação de *Tags*". As *tags* fornecidas pelo professor foram divididas em dois conjuntos de 5 *tags*,  $\mathbf{Ta}=\{\text{lógica, C\#, Bancos de Dados, Windows, Linux}\}$  e  $\mathbf{Tb}=\{\text{JQuery, PHP, XNA, Android, Flash}\}$ . O Grupo A foi composto por alunos das turmas do Curso Técnico em Informática e o Grupo B, por alunos das turmas do Curso Técnico de Informática para Internet e por alunos do Curso Técnico de Informática Integrado ao Ensino Médio.

2. O sistema deveria utilizar as *tags* coletadas no experimento anterior, isto é, o algoritmo de recomendação usaria o vocabulário descritor coletado no Experimento 1. Os alunos ficariam livres para a catalogação aceitando ou não as *tags* recomendadas.

## 4.2.2 Execução

A condução do experimento foi realizada em duas semanas. Na primeira semana, para envolver os alunos no processo e conscientizá-los a respeito de marcação social com recomendação, foi feito um trabalho de divulgação do experimento, abordando os conceitos relacionados e descrevendo os objetivos, a forma que seria a condução da coleta de *tags* e os períodos do experimento.

Para evitar problemas com o uso do *TagLink*, foi enviado aos alunos um tutorial detalhando os passos para se cadastrar no aplicativo, efetuar a pesquisa dos termos e atribuir *tags* aos recursos aceitando ou não a recomendação de *tags*. Depois cada grupo de alunos, Grupo A e Grupo B, recebeu os 5 termos,  $\mathbf{Ta}$  e  $\mathbf{Tb}$ , para serem pesquisados.

Na segunda semana deu-se o início da condução para coleta de *tags*, conforme o planejamento realizado. Os alunos deveriam, obrigatoriamente, atribuir pelo menos três *tags* para um determinado recurso, aceitando ou não a recomendação do sistema. Além das marcações, os alunos forneceram outras informações sobre a utilidade da recomendação de *tags* durante o processo de catalogação. Ao final de cada atribuição de *tags* através da recomendação ou não, os alunos responderam as seguintes perguntas: - As *tags* recomendadas estavam dentro do contexto da página? - Para você esse site pode ser classificado como? - O site permitiu sua interação através de algum exercício, *game*, ou enigma? - O site apresentou um vídeo? - O Conteúdo do site foi apresentado de forma Global ou Sequencial? Ao final do experimento, obteve-se um total de **2.180** marcações feitas por **212** alunos. Não houve dúvidas dos alunos durante o experimento.

## 4.2.3 Análise dos resultados

Na Tabela 4.6 estão relacionados os dados que foram levantados durante o experimento. A tabela está dividida em três blocos que apresentam: (A) os percentuais referentes a origem das marcações; (B) os percentuais quanto a avaliação dos alunos em relação à utilidade da recomendação e (C) os percentuais quanto a reutilização das *tags*. Os valores são calculados em relação ao total de alunos participantes e total de marcações realizadas pelos alunos independente do grupo a o qual o aluno pertence. Todos os percentuais de cada grupo foram calculados em função do total de marcações e participantes.

A Tabela 4.7 apresenta dados sumarizados por grupo. Isso permite uma análise que considere o perfil dos alunos participantes em relação a faixa etária e ao curso no qual estudam, pois os

Tabela 4.6: Resumo do Experimento 2

Dados		%
A	Marcações a partir das recomendações	45,76%
	Marcações com tags do professor	8,34%
	Marcações que não vieram nem de professor nem de recomendação	45,90%
B	Qtd de Alunos que acharam útil a recomendação	94,96%
	Qtd de alunos que não acharam útil a recomendação	5,04%
C	Marcações com novas tags	16,19%
	Marcações com tags já existentes	83,81%

alunos do Grupo A pertenciam a uma faixa etária maior e ao curso de Informática enquanto que os alunos do Grupo B eram mais jovens e pertenciam ao curso de Informática para Internet.

Tabela 4.7: Resumos do Experimento 2 por Grupo

Dados		Grupo A	%GA	Grupo B	%GB	Totais
A	Quantidade total de Alunos	114	53,77%	98	46,23%	212
	Quantidade total de Marcações	1187	54,42%	994	45,58%	2181
B	Marcações a partir das Recomendações	480	48,10%	518	51,90%	998
	Marcações com tags do professor	66	36,26%	116	63,74%	182
	Marcações que não vieram nem de professor nem de recomendação	641	64,04%	360	35,96%	1001
C	Qtd. de Alunos que acharam útil a recomendação	1101	53,16%	970	46,84%	2071
	Qtd. de alunos que não acharam útil a recomendação	86	78,18%	24	21,82%	110
D	Marcações com novas tags	213	60,34%	140	39,66%	353
	Marcações com tags já existentes	853	46,66%	975	53,34%	1828

### (1) Um mecanismo de recomendação de *tags* auxilia a catalogação de recursos de aprendizagem?

Observando-se a linhas A-1 da Tabela 4.6, pode-se observar que 45,76% dos alunos realizaram suas marcações aceitando a recomendação do sistema. Esse percentual considera apenas as *tags* recomendadas a partir da função de recomendação do sistema que em cada etapa, apresenta as 5 *tags* mais relevantes para o recurso selecionado. A linha A-2 da Tabela 4.6 apresenta que 8,34% dos alunos em suas marcações, escolheram *tags* que foram inicialmente definidas pelo professor como termo de origem. Por outro lado, na linha A-3 da Tabela 4.6 observa-se que 45,90% dos participantes utilizaram *tags* que não foram sugeridas nem pela função de recomendação, nem pelo professor. Esse equilíbrio comprova a liberdade que os alunos tiveram no processo de marcação, entendendo que em nenhum momento eram obrigados a aceitar a sugestão do sistema a não ser que eles próprios considerassem útil. Um detalhe interessante que pode ser notado é que ao serem questionados se a recomendação foi útil para eles ou não, 94,96% dos alunos responderam afirmativamente. Isso possibilita afirmar que mesmo os alunos que não aceitaram a recomendação, a acharam útil pois as *tags* sugeridas serviram para orientá-los e direcioná-los na escolha da própria *tag* para o recurso. Considerando que os alunos que optaram por utilizar *tags* do professor de certa forma aceitaram uma recomendação provida pelo modelo, pode-se afirmar que 54,10% dos participantes aceitaram a recomendação de *tags*, ou seja 1.180 alunos

conforme as linhas A-1 e A-2 da Tabela 4.7. Dentro do escopo do experimento é possível afirmar que um mecanismo de recomendação de *tags* auxilia a catalogação de recursos de aprendizagem, tanto quantitativamente como qualitativamente, ao proporcionar aos alunos um direcionamento sobre os termos.

## (2) Um mecanismo de recomendação de *tags* favorece mais a descrição ou a categorização dos recursos de aprendizagem por parte dos alunos?

No Experimento 1, constatou-se que os alunos participantes do processo de marcação social num ambiente de *e-learning* são propensos a se comportar mais como descritores que como categorizadores Strohmaier, Korner e Kern (2012). Para possibilitar uma comparação entre os resultados do Experimento 1 e 2, foram gerados gráficos semelhantes utilizados a seguir. A Figura 4.6 apresenta a distribuição do número de alunos por número de *tags* criadas, ou seja, inéditas no repositório de *tags* de dados. Comparando os gráficos, nota-se que com o mecanismo de recomendação, os alunos que criaram novas *tags* foi menor. Além disso, percebe-se uma maior regularidade na distribuição das criações, onde a maioria dos alunos participantes, utilizaram novas *tags* entre 1 a 4 marcações.

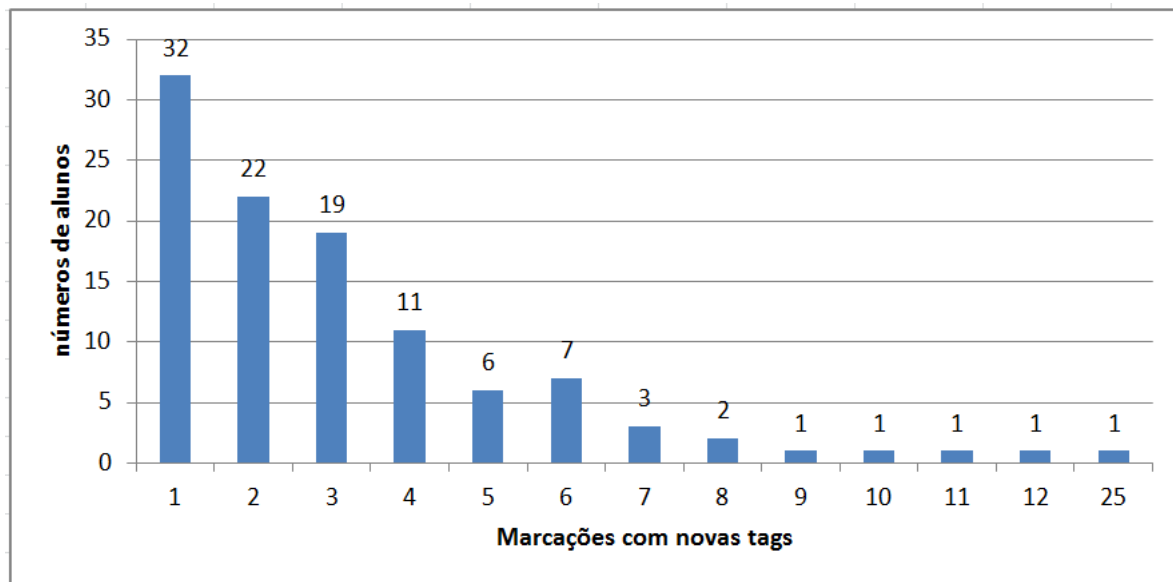


Figura 4.6: Distribuição do número de alunos por número de *tags* criadas.

Pode-se verificar na Figura 4.7 que a distribuição das *tags* reutilizadas por alunos é mais equilibrada em relação à Figura 4.3 entre os picos e vales do gráfico. Logo, pode-se afirmar que com a recomendação de *tags*, o índice de reutilizações de *tags* foi sensivelmente elevado 83,81% conforme a linha C-2 da Tabela 4.6. O fato de mais alunos reutilizarem *tags*, reforça a predominância de *tags* com significado geral e do senso comum dos participantes. Isto indica o caráter descritor das marcações com o mecanismo de recomendação. A Figura 4.8 apresenta o gráfico da distribuição das *tags* mais utilizadas, onde pode-se comprovar que entre estas, predominam termos com características de descrições mais específicas em relação ao experimento

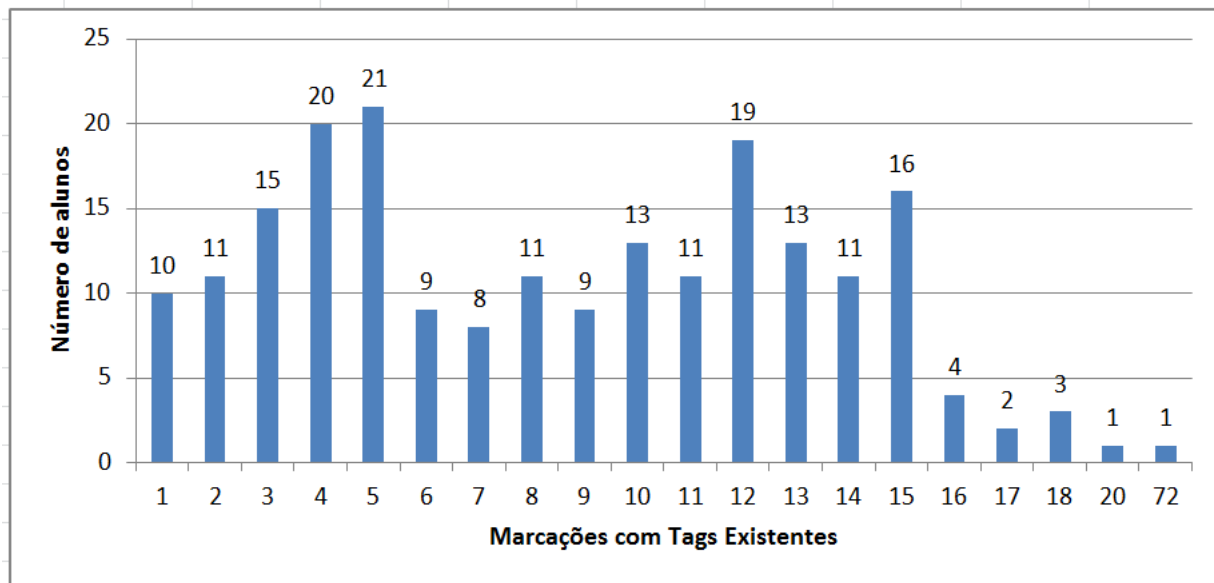


Figura 4.7: Distribuição do número de alunos por número de *tags* reutilizadas.

anterior apresentado na Figura 4.5. Dessa forma, conclui-se que mesmo usando um sistema de recomendação de *tags*, os alunos continuam tendo o perfil de descritores.

### (3) O vocabulário em torno dos recursos de aprendizagem se torna mais específico ao se utilizar um sistema de recomendação de *tags*?

Para responder esta questão, serão comparados os gráficos apresentados na Figura 4.5 do Experimento 1 com o gráfico apresentado na Figura 4.8 do Experimento 2. Eles se referem as 50 *tags* mais utilizadas durante o processo de catalogação. A curva de utilização sofreu pouca alteração em termos de quantidades, porém é fundamental analisar quais *tags* foram mais utilizadas. Mais especificamente, será observado o deslocamento ou a mudança de posição que algumas *tags* sofreram entre os dois experimentos. Observando a Figura 4.5, verifica-se que as três *tags* mais utilizadas respectivamente foram "Programação", "microsoft" e "linguagem". Na Figura 4.8, podemos perceber fatores muito interessantes. A *tag* "Programação" não aparece entre as 50 mais utilizadas, mesmo estando disponível no repositório de *tags*. A *tag* "microsoft" passou da posição 2 para a posição 7 e a *tag* "linguagem" foi deslocada para a posição 8. As três *tags* mais utilizadas no Experimento 2 foram "php", "c#" e "Android". A *tag* "Interessante" deslocou-se da posição 9 para a posição 38. Assim, observa-se que os termos mais genéricos, ou foram descartados, ou foram deslocados para baixo. Das 50 *tags* mais utilizadas, a menos utilizada foi a *tag* "Curso" que tem caráter de categorização por ser mais genérica. Semelhantemente, algumas *tags* que não foram utilizadas no Experimento 1, encontraram posição relevante no processo com recomendação como por exemplo a *tag* "Android".

A Tabela 4.8 apresenta o deslocamento de posição das *tags* mais utilizadas nos dois experimentos. São listadas apenas 20 *tags*, mas já é possível observar que *tags* mais específicas tiveram deslocamento positivo (para cima), enquanto que *tags* mais genéricas tiveram deslocamento negativo (para baixo) ou nem foram utilizadas no Experimento 2.



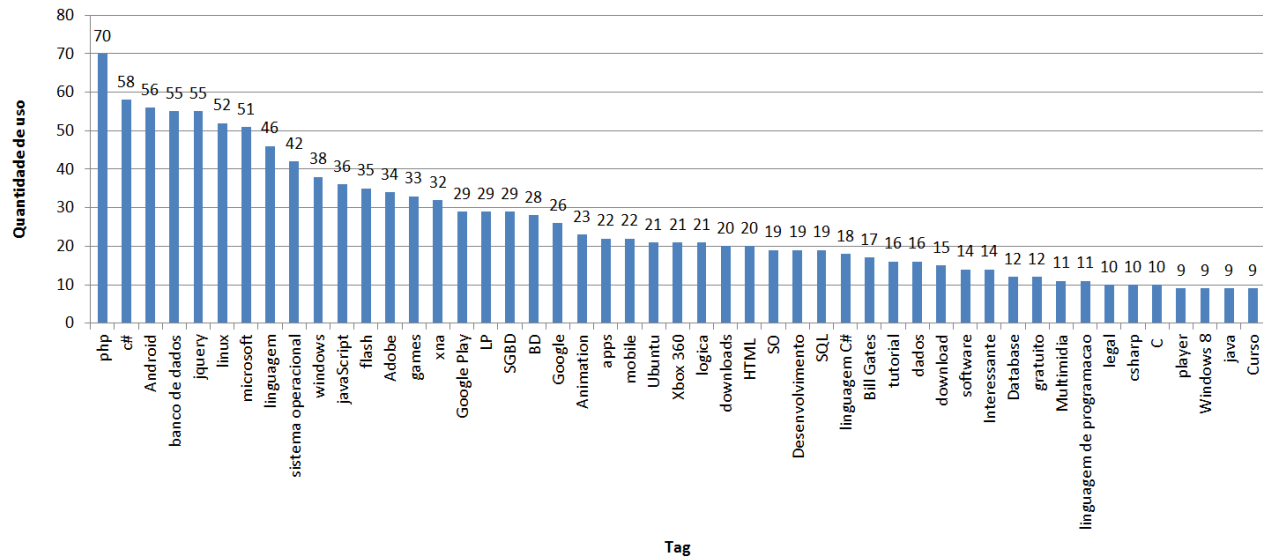


Figura 4.8: Distribuição do número de vezes que cada *tag* foi utilizada

Dessa forma, dentro do escopo do experimento, comprovou-se que o vocabulário em torno dos recursos de aprendizagem, tornou-se mais específico com o auxílio do sistema de recomendação proposto.

#### (4) O algoritmo de recomendação proposto é eficaz?

Para poder acompanhar e responder esta questão, foi implementado no TagLink, através a pontuação das *tags*, um *ranking* das 5 *tags* mais relevantes, ou seja, melhores pontuadas conforme o algoritmo proposto. São essas *tags* que são recomendadas aos alunos durante toda a catalogação. Para responder a questão desta seção, é importante verificar em que posição do *ranking* de recomendação a *tag* se encontrava ao ser escolhida para ser associada a um recurso de aprendizagem. Quanto mais alta a posição da *tag*, pode-se afirmar que maior foi a eficácia da função de recomendação proposta. Na Tabela 4.9, é apresentada a distribuição das recomendações em relação ao *ranking* das *tags* apresentadas ao aluno no momento da marcação. No topo da lista, se posicionam as *tags* com maior pontuação em relação ao recurso selecionado. Isto é **Posição 1** aponta para a *tag* mais relevante para o recurso segundo a recomendação proposta. Assim, nota-se que das 998 marcações feitas a partir da recomendação, em 246 (24,65%) delas os alunos escolheram a primeira *tag* da lista, ou seja, a considerada de maior relevância pelo sistema. Considerando que 226(22,65%) das recomendações foram aceitas utilizando *tags* da posição 2 e 183(18,34%) utilizando *tags* da posição 3.

Conclui-se que o algoritmo de recomendação proposto é eficaz, pois sugere *tags* que os alunos consideraram relevantes e escolheram para marcar o recurso.

#### (5) O modelo de ciclo de recomendação proposto (Seção 3.1) pode ser considerado válido?

Como observado na Tabela 4.6 na linha A-2, 8,34%, 182 marcações ocorreram utilizando-se as *tags* definidas como sugestão inicial pelo professor. O modelo proposto nesse trabalho e

Tabela 4.8: Deslocamento de posicionamento das *Tags* entre os experimentos

Tag	Exp1	Exp2	Desloc.
Programação	1	—	—
microsoft	2	7	-5
linguagem	3	8	-5
linux	4	6	-2
software	5	37	-32
sistema operacional	6	9	-3
javaScript	7	11	-4
Lógica	8		8
Interessante	9	38	-29
php	10	1	9
windows	11	10	1
Adobe	12	13	-1
games	13	14	-1
banco de dados	14	4	10
tutorial	15	34	-19
c#	16	2	14
C	17	45	-28
SO	18	—	—
HTML	19	28	-9
download	20	—	—

Tabela 4.9: Recomendações aceitas por *Ranking*

Posição	Recomendações	%
1	246	24,65%
2	226	22,65%
3	183	18,34%
4	184	18,44%
5	159	15,93%

apresentado na Figura 3.1, inclui a contribuição do professor como direcionador da pesquisa para a catalogação dos recursos de aprendizagem. Sem a contribuição do professor nesse processo, a estabilização do vocabulário descritor seria mais lenta e o *ranking* das *tags* recomendadas, teria menos qualidade no início das marcações, se houvesse escassez de *tags* para um recurso. Assim, mesmo recursos de aprendizagem novos, ou seja que ainda não receberam marcação alguma, terão sugestões de *tags* que serão relevantes, por estarem dentro do contexto de pesquisa. Esta característica do modelo proposto combinada com o algoritmo de recomendação que se mostrou eficaz, garantem que os objetivos da catalogação de recursos de aprendizagem sejam alcançados de forma apropriada, como demonstrado nos experimentos. Assim conclui-se que o modelo de ciclo de recomendação proposto pode ser considerado válido.

### 4.3 Experimento 3: Navegação e Pesquisa

Após o processo de catalogação, um dos benefícios esperados é a melhoria na busca e pesquisa de recursos e a facilidade de navegação por suas *tags*. A visualização das *tags* relacionadas a um recurso e de outras *tags* relacionadas entre si, possibilita que o aluno tenha visões das *tags*

Tabela 4.10: Ficha Técnica do Experimento 3

Objetivo:	Navegação e Pesquisa de recursos já catalogados
Metodologia:	- Os alunos foram divididos em dois grupos: Grupo A e Grupo B - O professor definiu apenas um assunto por classe para pesquisa - Os alunos navegaram pelas nuvens de <i>tags</i> até encontrarem o recurso pesquisado - O TagLink foi utilizado apenas para navegação e pesquisa
Local:	ETEC Fernando Prestes
Período:	Maió/2014
Nº de alunos	207

associadas a diferentes recursos. Dessa forma, o aluno pode navegar entre as *tags*, buscando novos recursos, enriquecendo assim o processo de descoberta de recursos de aprendizagem. A Figura 4.9 apresenta a interface de navegação do TagLink que possui um globo com uma nuvem de *tags*. Esta nuvem apresenta as 30 *tags* mais relevantes do sistema para os recursos catalogados. As com peso maior são mostradas com fontes maiores. As *tags* da primeira nuvem apresentada são consideradas de nível 0. Quando o aluno clicar em uma das *tags* apresentadas no globo, todos os recursos vinculados a *tag* serão listados e uma nova nuvem de *tags* co-ocorrentes será apresentada no globo de navegação.

Para esse experimento, as questões de pesquisa foram:

1. A catalogação feita nos experimentos anteriores foi de qualidade?
2. Os conceitos de visualização e navegação, implementados no TagLink, facilitaram a descoberta de recursos de aprendizagem por parte dos alunos?

Figura 4.9: TagLink - Globo com nuvem de *tags*



Figura 4.10: TagLink - Página de Navegação: Nuvem de *tags* e Recursos relacionados

### 4.3.1 Planejamento

A etapa de Navegação e Pesquisa foi organizada como se segue:

- Os alunos também foram divididos em dois grupos. Os alunos do Grupo A, pesquisariam a base de dados catalogada pelos alunos do Grupo B e vice-versa.
- O objetivo dessa vez não foi a catalogação, mas sim a busca por recursos já catalogados na etapa anterior.
- O professor definiu apenas um assunto a ser estudado. Para evitar seleções óbvias por parte dos alunos, ao invés de usar termos de pesquisa, foram definidos temas para cada turma de cada grupo.

Exemplo:

- GRUPO A: Desenvolvimento Web; Efeitos visuais em páginas *Web*, Dispositivos móveis Desenvolvimento de jogos, Formatação de páginas *Web*;
- GRUPO B: Desenvolvimento de Software, Armazenamento de Informações, Gerenciamento de recursos de máquina, Software livre (*open source*), Técnicas de desenvolvimento.

Para análise dos resultado dessa etapa, sempre quando o aluno encontrou um recurso de aprendizagem de acordo com o tema que o mesmo estava pesquisando, uma resposta foi solicitada:

- Você achou fácil navegar pelo sistema e pesquisar o tema solicitado?

Os alunos devem navegar e pesquisar os temas sobre o catálogo do grupo a que não pertence e a pesquisa ocorrerá nas duas bases: a formada sem recomendação (Experimento 1) e a formada com a recomendação de *tags* (Experimento 2).

Assim, o Experimento 3 foi planejado para verificar a qualidade da catalogação feita nos experimentos anteriores.

### 4.3.2 Execução

Para este experimento conforme o planejamento, em uma única semana, foi solicitado aos alunos que pesquisassem apenas um tema na base de recursos já catalogados. O aplicativo TagLink foi adaptado para apenas permitir a navegação e pesquisa somente através do globo de nuvem de *tags*. A cada seleção de uma *tag* dentro do globo, uma lista de recursos relacionados à *tag* escolhida era apresentada e o globo de nuvem de *tags* era atualizado com as  $k$  *tags* mais co-ocorrentes com a *tag* selecionada. Assim o aluno pôde continuar selecionando *tags* da nuvem até encontrar o recurso que atendesse sua pesquisa. Ao final cada aluno forneceu sua opinião sobre sua experiência em relação a navegação e pesquisa durante o processo. Como planejado, a navegação foi feita de forma cruzada, ou seja, no primeiro dia, alunos do GRUPO A navegaram sobre o catálogo resultante da atividade de marcação dos alunos do GRUPO B e vice-versa. Isso também ocorreu sobre o repositório de *tags* formado sem recomendação (Experimento 1) e sobre o repositório de *tags* formado com recomendação (Experimento 2). Ao todo, os alunos realizaram 431 pesquisas sendo 108 sobre a base de dados do Experimento 1 e 323 buscas sobre a base do Experimento 2.

### 4.3.3 Análise dos Resultados

Os dados coletados durante o experimento foram analisados e serão reportados a partir das questões de pesquisa.

#### (1) A catalogação feita nos experimentos anteriores foi de qualidade?

É possível inferir a qualidade da catalogação, pela facilidade ou não que os alunos tiveram para localizar recursos de aprendizagem de acordo com um assunto específico dentro do repositório construído a partir dos experimentos 1 e 2. Assim, das 431 pesquisas realizadas, em 83,06% (358), os alunos reportaram a busca e a navegação como "fácil" e em 16,94%(73) os alunos reportaram como "difícil". Esses dados apontam que a qualidade da catalogação realizada nos experimentos anteriores foi relevante para a recuperação dos recursos.

#### (2) Os conceitos de visualização e navegação implementados no TagLink facilitaram a descoberta de recursos de aprendizagem por parte dos alunos?

O caminho percorrido pelo aluno na navegação através das *tags* até encontrar o recurso procurado, indica a facilidade ou a dificuldade que o aluno teve para concluir sua busca. Cada nuvem de *tags* apresentada no globo de pesquisa, possui um nível, a partir de zero. Ou seja, o nível 0(zero) corresponde ao primeiro globo que é apresentado ao aluno. Conforme o aluno vai buscando o conteúdo, selecionando alguma *tag* do globo, um passo vai sendo adicionado e considerado, um nível vai sendo acrescentado e registrado até que o recurso seja encontrado. Convencionou-se nesse trabalho, denominar o caminho trilhado pelo aluno através do acesso ao globo de "rastros". Os rastros indicam de onde o aluno clicou no início de sua navegação, e que *tags* foi escolhendo até encontrar o recurso desejado. Dessa forma é possível medir a profundidade da busca. A Tabela 4.11 apresenta em que níveis os recursos foram encontrados nas pesquisas. O mesmo pode ser observado na Figura 4.11. Analisando estes dados, percebe-se que a maioria dos

recursos foram encontrados nos primeiros níveis das nuvens de *tags* apresentadas. Isso evidencia a qualidade da catalogação gerada anteriormente, pois o objetivo da busca é encontrar o recurso pesquisado que esteja de acordo com o tema abordado. Observa-se que isso aconteceu nos primeiros níveis.

Por outro lado, ao se observar que algumas pesquisas chegaram a níveis altos, como 12, 20 e 23, pode-se considerar dois fatores: (1) Alguns recursos não foram catalogados com qualidade, o que levou alguns alunos a terem mais dificuldade para encontrar o que procuravam; (2) A interface do TagLink com as nuvens de *tags* dispostas no globo dinâmico causou efeito tão positivo quanto a experiência de navegação e a visualização das *tags*, que alguns alunos ficaram mais tempo interagindo com os globos, antes de realmente partirem para a pesquisa dos temas. Entretanto, como a quantidade de recursos encontrados em níveis mais altos é bem pequena, prevalece o argumento que os conceitos de visualização e navegação implementados no TagLink, facilitaram a descoberta de recursos de aprendizagem por parte dos alunos.

Tabela 4.11: Nível da Nuvem de tags por recursos encontrados

Nível	Qtd. Recursos	%
0	65	49,62%
1	22	16,79%
2	12	9,16%
3	13	9,92%
4	7	5,34%
5	4	3,05%
6	2	1,53%
7	1	0,76%
12	3	2,29%
20	1	0,76%
23	1	0,76%

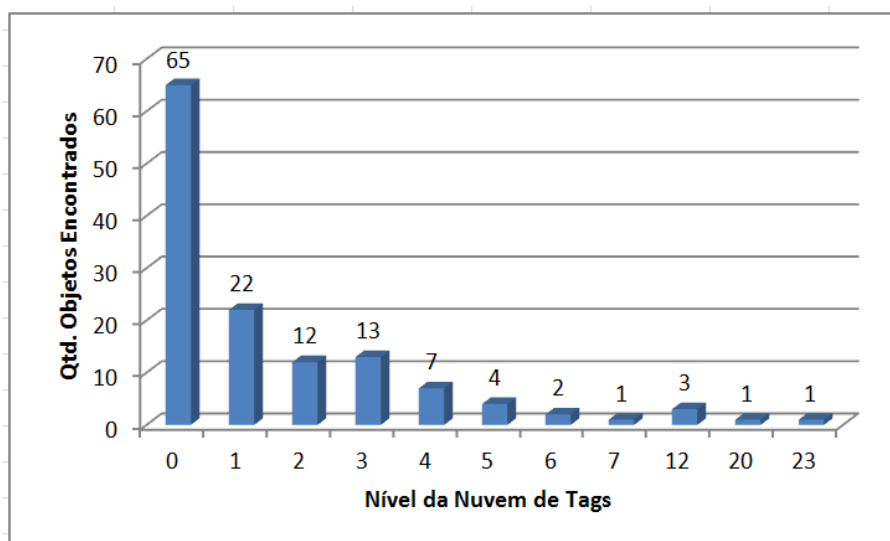


Figura 4.11: Quantidade de Recursos encontrados em cada nível da nuvem de *tags*

## 4.4 Considerações Finais

Para a validação do ciclo de recomendação, três experimentos foram realizados. O Experimento 1 foi conduzido utilizando a atividade de marcação social sem recomendação de *tags*. Nesse experimento pode-se analisar e tornar evidente a necessidade de um mecanismo de recomendação de *tags*, pois observou-se o grande número de reutilizações de *tags* durante o processo de marcação por parte dos alunos.

O Experimento 2 foi realizado utilizando-se do módulo (algoritmo e ciclo) de recomendação proposto, para analisar-se os resultados em relação ao conjunto de *tags* gerados em torno dos recursos de aprendizagem e verificar os impactos desta implementação no processo de catalogação.

Para se verificar a navegabilidade por entre as nuvens de *tags* para se encontrar os recursos catalogados, planejou-se e realizou-se o Experimento 3. Nesse experimento, as atividades de navegação e pesquisa foram realizadas tanto sobre o catálogo resultante do Experimento 1 (sem recomendação) como sobre o catálogo resultante do Experimento 2 (com recomendação). Dessa forma, foi possível analisar os vocabulários de *tags* em relação a especificidade e abrangência dos mesmos nos dois experimentos.

O **e-TagReS** apresentou-se válido e eficaz dentro do escopo deste trabalho, pois permitiu uma catalogação de recursos de aprendizagem com *tags* mais específicas, o que favorece a pesquisa e recuperação dos recursos. Outro aspecto positivo é que a maioria dos alunos envolvidos, declarou que teve facilidade na utilização do TagLink durante a atividade de marcação, devido à simplicidade do processo.

## Conclusões e Futuros Trabalhos

A necessidade da catalogação, classificação e indexação de conteúdos disponíveis na *Web*, encontrou na marcação social novas possibilidades. Com a participação mais ativa e colaborativa dos usuários através da atividade de marcação de recursos por meio de *tags*, diversas propostas surgiram como forma de motivar tal atividade. Técnicas de recomendação e visualização de *tags* passaram a ser foco de muitos estudos visando contribuir para a melhoria da atividade de criação e gerenciamento de repositórios de recursos quanto a pesquisa e navegação.

Nos ambientes de aprendizagem eletrônica, o conceito de objetos de aprendizagem passou a ganhar relevância devido à definição de repositórios de aprendizagem desenvolvidos e indexados com base em metadados; priorizando principalmente a reutilização do conhecimento em diferentes cenários de aprendizagem. Logo, percebeu-se o potencial das folksonomias para se conceber um eficiente meio de classificação de recursos, devido a sua flexibilidade e abrangência semântica. Diante disso, neste trabalho, foi proposto e desenvolvido um sistema de recomendação de *tags* para auxiliar na catalogação de recursos de aprendizagem, através da marcação social.

Após uma revisão sistemática (Capítulo 2), foi conduzido o Experimento 1 Seção(4.1) com o objetivo de se investigar o processo de marcação social sem o auxílio de um algoritmo de recomendação de *tags*. O Experimento 1, apresentou em seus resultados que os alunos utilizaram amplamente a opção de reutilizar *tags* próprias ou de outros para criarem novas marcações. Esse resultado sinalizou para o quanto seria importante um mecanismo de recomendação de *tags* que através de um algoritmo de recomendação, oferecesse *tags* mais relevantes aos alunos no momento da marcação. Desta maneira seria possível potencializar as marcações enriquecendo o vocabulário dos objetos.

A fim de se definir um algoritmo de recomendação de *tags* que fosse adequado, estudos preliminares foram efetuados, como descrito na seção 3.2, e após testes decidiu-se pelo algoritmo que se mostrou mais eficiente em relação aos objetivos da proposta. Em seguida, mais dois experimentos foram elaborados.

O Experimento 2 foi conduzido para validar o sistema de recomendação proposto. Nesse experimento, os alunos foram auxiliados pela recomendação de *tags* durante a catalogação. Com os resultados desse experimento, pode-se constatar a eficácia do algoritmo de recomendação proposto (Tabela 3.11), pois os resultados demonstraram que a maioria das recomendações aceitas pelos alunos, foram referentes a *tags* que possuíam alta pontuação de acordo com o algo-



ritmo. As *tags* sugeridas pelo professor direcionaram o processo de marcação colaborando para uma estabilização mais rápida do vocabulário em torno dos recursos. Através do experimento constatou-se que o sistema de recomendação proposto era válido e atingia seus objetivos.

Finalmente, o Experimento 3 foi conduzido com o objetivo de verificar se a catalogação resultante dos experimentos anteriores foram adequadas. Esse experimento priorizou a pesquisa de recursos de aprendizagem catalogados e a navegação pelas nuvens de *tags*. Os resultados deste experimento demonstraram que o catálogo resultante foi gerado adequadamente, pois a partir de temas gerais passados aos alunos, estes encontraram o assunto procurado já nos primeiros níveis das nuvens de *tags*, isto é, o esforço para localizar os recursos de aprendizagem procurados, foi pequeno.

Todas as questões de investigação elencadas no Capítulo 1 foram respondidas, onde pode-se mostrar que:

- Um mecanismo de recomendação de *tags* auxilia de fato a catalogação de recursos de aprendizagem.
- Em um ambiente de marcação social voltado para a catalogação de recursos de aprendizagem, e que se utiliza de um mecanismo de recomendação de *tags*, o perfil predominante dos alunos é o de "descritor" de recursos e não "categorizador".
- O vocabulário em torno dos recursos de aprendizagem se torna mais específico com a utilização de um mecanismo de recomendação de *tags*.
- O algoritmo e o modelo de ciclo de recomendação proposto nesse trabalho, podem auxiliar o processo de marcação de recursos de aprendizagem.

Dessa forma, foi mostrado como a marcação social pode ser aplicada e potencializada através de um mecanismo de recomendação de *tags* na tarefa de catalogar recursos de aprendizagem dentro de um domínio de conhecimento específico.

## 5.1 Contribuições

Este trabalho trouxe contribuições significativas das quais podem-se destacar:

- Sistema de Recomendação de *Tags* Proposto, onde a participação do professor direciona a catalogação dos recursos de aprendizagem por parte dos alunos;
- Algoritmo de recomendação adotado, intensamente testado, chegando-se a uma definição que pode ser aplicado em outros contextos;
- Ampla revisão sistemática, onde foram levantados estudos relevantes sobre a marcação social, classificados em 4 aspectos fundamentais para marcação social na área de e-learning: Visualização, Recomendação, Objetos de Aprendizagem e Geração Automática de *tags*;

- Protótipo funcional TagLink que independente do domínio de conhecimento, pode ser utilizado para a catalogação de recursos de aprendizagem, navegação e pesquisa de recursos catalogados.
- Resultados dos três experimentos que trazem conclusões significativas sobre o uso da recomendação para a marcação social com o envolvimento de 755 alunos participantes. Os experimentos podem ser reproduzidos e aplicados a outros contextos, pois o trabalho faz a descrição sistematizada de sua condução.

## 5.2 Futuros Trabalhos

Vários outros aspectos ainda podem ser analisados para a aplicação da marcação social na catalogação de recursos de aprendizagem. Durante a catalogação do Experimento 2, também foi solicitado ao aluno participante do processo que ele classificasse cada recurso marcado por ele dentro de uma "função de aprendizagem" (questionário, exercício, pesquisa, jogo, etc) e o tipo do recurso(texto, vídeo, imagem, wiki etc), segundo sua própria opinião. A classificação foi baseada nas dimensões de aprendizagem de Felder e Silverman (1998). Para isso, foi apresentado ao aluno um questionário *on-line*, com perguntas pertinentes para a classificação dos recursos.

Considerando essas informações obtidas pelo questionário, o mecanismo de recomendação poderá ser ajustado para sugerir *tags* de acordo com o perfil de um aluno, ou de um grupo de alunos, ou até mesmo de acordo com as características dos próprios recursos.

Outro trabalho futuro que pode ser feito é referente ao Experimento 3. Nesse experimento, foi registrado o caminho ("rastros") que o aluno trilhou para encontrar o recurso de aprendizagem procurado. Dessa forma, pode-se analisar e identificar, os caminhos mais comuns que interligam as *tags* e os recursos de aprendizagem e enriquecer o mecanismo de recomendação, através da sugestão de caminhos de busca mais relevantes. A partir dessas informações, também será possível analisar interesses e preferências em comum por parte dos alunos, passando para outro nível de recomendação, ou seja, a recomendação de recursos de aprendizagem a partir do perfil e dos interesses dos alunos.

Outra recomendação, que o sistema poderá fazer seria a sugestão de formação de grupos de estudos. Isto, certamente implica, no desenvolvimento de novos algoritmos de recomendação e novos conceitos de visualização que possam suportar a implantação desses novos conceitos de recomendação.

Assim, constata-se que há ainda um longo caminho a ser trilhado na direção de pesquisas que contribuam para a exploração de todo potencial que a marcação social tem para sua utilização efetiva em ambientes de ensino-aprendizagem e *e-learning*.

## Publicações

A partir da pesquisa apresentada nesta dissertação, foram publicados, em colaboração com diferentes pesquisadores, os seguintes artigos:

**AMARAL, A. R.; ZAINA, A. M.; RODRIGUES JR, J. F.** Cataloguing of learning objects using social tagging. *XL Conferencia Latinoamericana en Informática (CLEI 2014)*.

**AMARAL, A. R.; ZAINA, A. M.; RODRIGUES JR, J. F.** The use of social tagging to support the cataloguing of learning objects. *The 44th Annual Frontiers in Education (FIE 2014) Conference*.

## Submissões

O trabalho completo será submetido ao *Computers & Education Journal - Elsevier*.

# Bibliografia

ANAND, D.; MAMPILLI, B. S. Folksonomy-based fuzzy user profiling for improved recommendations. *Expert Systems with Applications*, v. 41, n. 5, p. 2424-2436, 2014.

ASSIS, J.; MOURA, M. A. Folksonomia: a linguagem das tags folksonomy: the language of the tags. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 18, n. 18, p. 85-106, 2013.

BATEMAN, S.; BROOKS, C.; MCCALLA, G. Applying collaborative tagging to e-learning. In: *In Proc. of ACM WWW*. [S.l.: s.n.], 2007.

BENZ, D. et al. The social bookmark and publication management system bibsonomy. *VLDB Journal* v. 19, Issue 6, 2010, Pages 849-875, 2010.

CERNEA, D. A.; MORAL, E. D.; GAYO, J. E. L. Soaf: semantic indexing system based on collaborative tagging. *Interdisciplinary Journal of E-Learning and Learning Objects, Annual, 2008*, 2008.

CHO, C. W. et al. A social tagging system for online learning objects. *Advanced Science Letters* v. 4, Issue 11-12, November 2011, Pages 3362-3365, 2011.

COELHO, G. O. Recuperação de objetos de aprendizagem baseada na web 2.0. *ETD - Educação Temática Digital* v. 14 n. 2 2012, 2009.

CONOLE, G.; CULVER, J. The design of cloudworks: Applying social networking practice to foster the exchange of learning and teaching ideas and designs. *Computers and Education*, v. 54 3 |, pp. 679-692., 2010.

DAHL, D.; VOSSEN, G. Evolution of learning folksonomies: social tagging in e-learning repositories. *International Journal of Technology Enhanced Learning*, v. 1, p. 35-46, 2008.

DAHL, D.; VOSSEN, G. Learning object metadata generation in the web 2.0 era. *International Journal of Information and Communication Technology Education*, July-Sept, 2008, Vol.43, p.110, 2008.

FELDER, R. M.; SILVERMAN, K. L. Learning and teaching styles in engineering education. *Journal of Engineering Education*, v. 78, p. 674-681, 1998.

GARCIA-SILVA, A. et al. Review of the state of the art: discovering and associating semantics to tags in folksonomies. *The Knowledge Engineering Review*, v. 27, n. 1, p. 57-85, 2012.

- GLAVINIC, V.; GRANIC, A. Hci research for e-learning: Adaptability and adaptivity to support better user interaction. In: *In: HCI and Usability for Education and Work. LNCS 5298*. [S.l.]: Springer Berlin Heidelberg, 2008. p. 359–376.
- GOLDER, S. A.; HUBERMAN, B. A. Usage patterns of collaborative tagging systems. *Journal of Information Science v. 32, n. 2 P. 198-208*, 2009.
- GUPTA, M. et al. Survey on social tagging techniques. *Newsletter ACM SIGKDD Explorations v. 12 Issue 1, June 2010 p. 58-72*, 2010.
- HODGINS, H. W. The future of learning objects. In: *Proceedings of the 2002 eTEE Conference 11-16 August 2002 Davos, Switzerland*. [S.l.: s.n.], 2002. p. 76–82.
- HOFMANN, T. Probabilistic latent semantic analysis. In: *Proceedings of the Fifteenth Conference Annual Conference on Uncertainty in Artificial Intelligence UAI-99*. San Francisco, CA: Morgan Kaufmann, 1999. p. 289–296.
- HSIEH, W.-T.; LAI, W.-S.; CHOU, S. T. Current developments in technology-assisted education. a collaborative tagging system for learning resources sharing. *Technological Science Education, Collaborative Learning, Knowledge Management pp. 733-1504*, 2006.
- HSIEH, W.-T. et al. A collaborative desktop tagging system for group knowledge management based on concept space. *Expert Systems with Applications v. 36, Issue 5, July 2009, p. 9513-9523*, 2009.
- JASCHKE, R. et al. Recommender systems for the social web. *Challenges in Tag Recommendations for Collaborative Tagging Systems*, p. 65–87, 2012.
- JELASSI, M. N.; YAHIA, E. M. N. S. B. A personalized recommender system based on users' information in folksonomies. In: *Proceedings of the 22nd international conference on World Wide Web companion*. [S.l.: s.n.], 2013. p. 1215–1224.
- KITCHENHAM, B. Procedures for performing systematic reviews. *Joint Technical Report Software Engineering Group. Department of Computer Science Keele University. United King and Empirical Software Engineering. National ICT Australia Ltd.*, 2004.
- KORNER, C. Understanding the motivation behind tagging. *ACM Student Research Competition*, v. 9, 2009.
- KORNER, C. Understanding the motivation behind tagging. In: *In: 20th ACM conference on hypertext and hypermedia*. [S.l.: s.n.], 2009.
- LEAL, D.; AMARAL, L. *Do Ensino em Sala ao e-Learning*. [S.l.: s.n.], 2004.
- LOPS, P. et al. A semantic content-based recommender system integrating folksonomies for personalized access. *Studies in Computational Intelligence v. 229, 2009, p. 27-47*, 2009.
- LÓPEZ, V. F.; PRIETA, F. la; OGIHARA, D. D. W. M. A model for multi-label classification and ranking of learning objects. *Expert Systems with Applications*, v. 39, n. 10, p. 8878–8884, 2012.

- MCGREAL, R. Learning objects: A practical definition. *International Journal of Instructional Technology and Distance Learning*, v. 1, n. 9, 2004.
- MILICEVIC, A.; NANOPOULOS, A.; IVANOVIC, M. Social tagging in recommender systems: A survey of the state-of-the-art and possible extensions. *Artificial Intelligence Review*, 2010.
- MONGE, S.; OVELAR, R.; AZPEITIA, I. Repository 2.0: Social dynamics to support community building in learning object repositories. In: INSTITUTE, I. S. (Ed.). *Interdisciplinary Journal of E-Learning and Learning Objects v. 4*. [S.l.]: Informing Science Institute, 2008. p. 191.
- PEÑALVO, G. et al. Learning objects for e-activities in social web. *WSEAS Transactions on Systems*, 2007.
- RIBEIRO, F. A. A.; FONSECA, L. C. C.; FREITAS, M. S. Recomendando objetos de aprendizagem a partir dashashtags postadas no moodle. In: *XXIV Simpósio Brasileiro de Informática na Educação SBIE 2013*. [S.l.: s.n.], 2013. p. 82–91.
- RIGO, W. et al. Interfaces web baseadas em conhecimento para anotação de recursos de informação e gerenciamento de repositórios. *Anais do Simpósio Brasileiro de Informática na Educação 2010, João Pessoa-PB*, 2009.
- SHIH, W.; TSENG, S. Folksonomy-based indexing for retrieving tutoring resources. In: *2012 Seventh IEEE International Conference on Wireless, Mobile and Ubiquitous Technology in Education*. [S.l.: s.n.], 2012. p. 97–101.
- SHIH, W.-C.; TSENG, S.-S. Folksonomy-based indexing for location-aware retrieval of learning contents. *5th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education, WMUTE 2008, Beijing, China*, 2008.
- SIERRA, J. L.; VALMAYOR, A. F. Tagging learning objects with evolving metadata schemas. *Eighth IEEE International Conference on Advanced Learning Technologies, ICALT 2008; Santander, Spain*, 2008.
- SIGURBJOURNSSON, B.; ZWOL, R. van. Flickr tag recommendation based on collective knowledge. In: *In WWW 08: Proceeding of the 17th International Conference on World Wide Web*. [S.l.: s.n.], 2008. p. 327–336.
- SINCLAIR, J.; CARDEW-HALL, M. The folksonomy tag cloud: When is it useful? *Journal of Information Science February 2008 v. 34 n. 1, p. 15-29*, 2009.
- STROHMAIER, M.; KORNER, C.; KERN, R. Understanding why users tag - a survey of tagging motivation literature and results. *Web Semantics: Science, Services and Agents on the World Wide Web*, v. 17, p. 1–11, 2012.
- TEIXEIRA, J. S. F.; SÁ, E. J. V.; FERNANDES, C. T. Uma proposta para indexação semântica de objetos de aprendizagem. *Paper - laclo 2011*, 2011.
- TRANT, J. Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information Vol.101*, 2009.

- VAZQUEZ, A. R.; OSTROVSKAYA, Y. A. Analysis of open technological standards for learning objects. In: *In Proceedings of 4th LA-Web, Puebla Cholula, Mexico*. [S.l.: s.n.], 2006. p. 105–108.
- WAL, V. Explaining and showing broad and narrow folksonomies. *Off the Top [blog]*, 2005.
- WU, C.; BO, Z. Tags are related: Measurement of semantic relatedness based on folksonomy network. *Computing and Informatics v. 30, Issue 1, 2011, p. 165-188*, 2011.
- XU, Z. et al. Towards the semantic web: Collaborative tag suggestions. In: *Proceedings of the Collaborative Web Tagging Workshop at the WWW 2006 - Edinburgh - Scotland*. [S.l.: s.n.], 2006.
- ZAMORA, F. S.; NISTAL, M. L. Visualizing tags as a network of relatedness. In: *Frontiers in Education Conference, 2009. FIE '09. 39th IEEE*. [S.l.: s.n.], 2009. p. 1–6.
- ZERVAS, P.; SAMPSON, D. G. The effect of users' tagging motivation on the enlargement of digital educational resources metadata. *Computers in Human Behavior*, v. 32, p. 292–300, 2014.
- ZIESEMER, A. C. Recomendação de tags para mídia social colaborativa : da generalização à personalização. 2012.