

UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

LUCAS PUPULIN NANNI

DESiRe: uma abordagem dinâmica para recomendação de resultados
de busca em atividades de pesquisa exploratória

Maringá
2015

LUCAS PUPULIN NANNI

DESiRe: uma abordagem dinâmica para recomendação de resultados
de busca em atividades de pesquisa exploratória

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Departamento de Informática, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientadora: Prof.^a Dr.^a Valéria Delisandra Feltrim

Maringá
2015

Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca Central - UEM, Maringá, PR, Brasil)

N184d Nanni, Lucas Pupulin
DESiRe: uma abordagem dinâmica para recomendação de resultados de busca em atividades de pesquisa exploratória / Lucas Pupulin Nanni. -- Maringá, 2015.
90 f. : il. color., figs., tabs., quadro + apêndice

Orientadora: Prof.^a Dr.^a Valéria Delisandra Feltrim.

Dissertação (mestrado) - Universidade Estadual de Maringá, Centro de Tecnologia, Departamento de Informática, Programa de Pós-Graduação em Ciência da Computação, 2015.

1. Recomendação dinâmica de resultados de busca.
2. Recuperação da informação - Pesquisa Exploratória - Busca personalizada. 3. Recuperação interativa de informação. I. Feltrim, Valéria Delisandra, orient.
II. Universidade Estadual de Maringá. Centro de Tecnologia. Departamento de Informática. Programa de Pós-Graduação em Ciência da Computação. III.
Título.

CDD 21.ed. 025.4

MN-002060

FOLHA DE APROVAÇÃO

LUCAS PUPULIN NANNI

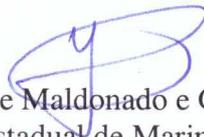
DESiRe: uma abordagem dinâmica para recomendação de resultados de busca em atividades de pesquisa exploratória

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Departamento de Informática, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Ciência da Computação pela Banca Examinadora composta pelos membros:

BANCA EXAMINADORA



Profa. Dra. Valéria Delisandra Feltrim
Universidade Estadual de Maringá – DIN/UEM



Prof. Dr. Yandre Maldonado e Gomes da Costa
Universidade Estadual de Maringá – DIN/UEM



Prof. Dr. Roberto Pereira
Universidade Estadual de Campinas – IC/UNICAMP

Aprovada em: 24 de agosto de 2015.

Local da defesa: Sala 120, Bloco C56, *campus* da Universidade Estadual de Maringá.

Ao professor Sérgio Roberto Pereira da Silva, que deixou saudades.

Agradecimentos

Agradeço e dedico este trabalho aos meus pais, Inês Pupulin Nanni e José Nanni Junior, por todo o incentivo e apoio que foi concedido. Agradeço também ao professor Dr. Sérgio Roberto Pereira da Silva (*in memoriam*) pela oportunidade de ter sido seu aluno e orientando durante a graduação e mestrado.

À professora Dr. Valéria Delisandra Feltrim, que me acolheu como orientando após a perda do professor Sérgio e se dedicou indistintamente ao desenvolvimento deste trabalho. Aos colegas do Grupo de Sistemas Interativos Inteligentes (GSII) pelas discussões, ajuda nos momentos difíceis e principalmente pela amizade, em especial: José Valderlei da Silva (Jesus), Josiane Melchiori Pinheiro e Roberto Pereira, pela dedicação e comprometimento em auxiliar o desenvolvimento do meu trabalho após o momento tão delicado em que perdemos o nosso mentor.

Aos colegas de mestrado, os quais faço questão de mencionar individualmente: Alexandre Augusto Giron, Alexandre Cassimiro Andreani, Aline de Paula Sanches, Alisson Gaspar Chiquitto, Ariel Gustavo Zuquello, Douglas Francisquini Toledo, Emanuel Felipe Duarte, Felipe Pereira Perez, Frank Willian Cardoso de Oliveira, Geazzy Brasilino Marçal Zanoni, Jaime William Dias, Leandro Lago da Silva, Luciano Fiorin Junior, Marcio Henrique Gimenes Bera, Mauricio Begnini, Paulo Roberto de Oliveira, Pietro Martins de Oliveira, Ricardo Theis Geraldi, Rodolfo Miranda Pereira, Romulo de Aguiar Beninca e Yoji Massago. Um grupo do qual me orgulho de ter feito parte por toda a dedicação e companheirismo, e também pela costumeira pausa diária para tomar um café na mercearia próxima ao prédio do departamento.

À Maria Inês Davanço, secretária do mestrado, agradeço a atenção e o apoio. Sua competência e dedicação são grandes exemplos e vão muito além de suas atribuições. Finalmente, agradeço o apoio financeiro cedido pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), ao Programa de Pós-Graduação em Ciência da Computação (PCC) e ao Departamento de Informática (DIN) da Universidade Estadual de Maringá (UEM).

“Everything should be made as simple as possible, but not simpler.”
(Albert Einstein)

DESiRe: uma abordagem dinâmica para recomendação de resultados de busca em atividades de pesquisa exploratória

Resumo

A dificuldade para recuperar informações relevantes e de qualidade na *Web* tem se agravado a medida que a complexidade e disponibilidade de recursos informacionais aumentam. Grande parte dessa dificuldade decorre da deficiência dos buscadores em identificar o contexto informacional da busca que o usuário está realizando. Esse problema se acentua ainda mais quando o usuário está engajado em atividades complexas de busca como a de Pesquisa Exploratória. Como forma de suprir tal deficiência, diversos trabalhos têm proposto ferramentas e técnicas associadas aos conceitos de recuperação adaptativa e busca personalizada para capturar as diferentes necessidades de informação dos usuários e provê-los com resultados relevantes considerando seus interesses individuais. Entretanto, os trabalhos desenvolvidos até então se restringem apenas à página de resultados para apresentar os esforços da personalização, deixando de explorar a possibilidade de recomendar resultados enquanto o usuário navega pelo espaço de busca. Dessa forma, propomos a DESiRe, uma abordagem dinâmica para a recomendação de resultados de busca capaz de apresentar resultados relevantes enquanto o usuário navega pelo espaço de busca recuperado. A abordagem se baseia em um modelo do interesse de busca do usuário para reclassificar os documentos recuperados e permitir que um conjunto de resultados relevantes seja recomendado ao usuário durante a navegação. Para tornar o processo de recomendação dinâmico, o modelo é construído conforme o usuário navega pelos documentos recuperados e informa o seu interesse por eles. Os resultados obtidos por meio de simulações mostraram que a abordagem proporcionou uma melhoria média de 88% na qualidade do *ranking* oferecido pelo Google. Além disso, o processo de recomendação foi igualmente efetivo, proporcionando recomendações de alta qualidade com um número relativamente alto de resultados inéditos. Avaliações preliminares também mostraram que a abordagem é capaz de lidar com *feedbacks* considerados suspeitos e de propagar o modelo de interesse pelas diversas atividades que podem compor uma mesma tarefa de busca.

Palavras-chave: Recuperação interativa de informação. Busca personalizada. Pesquisa Exploratória. Recomendação dinâmica de resultados de busca.

DESiRe: a dynamic approach for results recommendation in exploratory search activities

Abstract

The difficulty to retrieve relevant and quality information on the Web has worsened as the complexity and availability of information resources increase. Much of such difficulty stems from the deficiency of search engines to identify the informational context of the search performed by the user. This problem increases when the user is engaged in complex search activities such as Exploratory Search. To overcome this deficiency, several studies have proposed tools and techniques related to adaptive retrieval and personalized search concepts in order to capture the different users' information needs and provide them with relevant results considering their individual interests. However, works done so far restricted themselves only to the results page to display the customization efforts, failing to explore the possibility of recommending results while the user navigates through the search space. Thus, we propose Desire, a dynamic approach for recommending search results able to present relevant results while the user navigates through the retrieved search space. The approach is based on a model of the user's search interest to re-rank the retrieved documents, allowing to recommend a set of relevant results to the user during navigation. To make dynamic the process of recommendation, the model is built as the user navigates the retrieved documents and informs his interest about them. The results obtained by simulations have shown that our approach provided an average improvement of 88% in the quality of the ranking provided by Google. In addition, the recommendation process was equally effective, providing high quality recommendations with a relatively high number of unseen results. Preliminary evaluations have also shown that the approach is able to handle suspect feedbacks and to propagate the model of interest among the several activities that can comprise the same search task.

Keywords: Interactive information retrieval. Personalized search. Exploratory Search. Dynamic search results recommendation.

Lista de ilustrações

Figura 1 – As atividades de busca definidas por Marchionini (2006).	22
Figura 2 – Aplicação do modelo do usuário em diferentes etapas do processo de personalização	26
Figura 3 – Interação entre <i>Feedbacks</i> implícitos e explícitos para coletar informações sobre o perfil do usuário e mantê-lo atualizado durante a personalização	28
Figura 4 – Reorganização da página de resultados proposta por Shen, Tan e Zhai (2005b).	32
Figura 5 – Reorganização da página de resultados proposta por Raman, Joachims e Shivaswamy (2011).	33
Figura 6 – Representação comportamental da DESiRe.	35
Figura 7 – Exemplificação da DESiRe como uma extensão para um navegador <i>Web</i> .	36
Figura 8 – Adaptação do processo de busca às etapas do processo de recomendação previsto para o processo de recomendação.	37
Figura 9 – Estrutura de uma tarefa de busca.	38
Figura 10 – Representação do espaço vetorial definido pelos termos extraídos de dois documentos fictícios.	39
Figura 11 – Representação visual do processo de aproximação do Modelo do Interesse de Busca (MIB) aos documentos relevantes.	41
Figura 12 – Organização da coleção de teste construída.	45
Figura 13 – Distribuição do grau de relevância para cada atividade de busca.	58
Figura 14 – Distribuição sumarizada do grau de relevância dos documentos recuperados para as atividades A1, A2, A3a, A3b, A4b e A5.	59
Figura 15 – Distribuição dos documentos relevantes.	60
Figura 16 – Processo de recomendação adaptado para a simulação das atividades. .	65

Lista de quadros

Quadro 1 – Problema de busca: “Benefícios do consumo de chá”.	47
Quadro 2 – Problema de busca: “Café ou chá, qual é melhor para a saúde?”. . . .	48
Quadro 3 – Problema de busca: “Tratamento de queimaduras por água fervente”. .	49
Quadro 4 – Problema de busca: “Aprender sobre Bitcoins”.	50
Quadro 5 – Problema de busca: “Efeitos das cores sobre o humor”.	51
Quadro 6 – Problema de busca: “Localização da ilha de Java”.	52

Lista de tabelas

Tabela 1 – Grau de concordância κ para as tarefas de julgamento de relevância.	56
Tabela 2 – Classificação e recomendação resultantes para cada <i>feedback</i> capturado (f.c.) durante a simulação realizada para a atividade A4a.	67
Tabela 3 – Métrica NDCG calculada para as posições do <i>ranking</i> resultante em cada iteração da simulação realizada para a atividade A4a.	69
Tabela 4 – Métrica NDCG-média calculada para as posições do <i>ranking</i> resultante em cada iteração das simulações realizadas para todas as atividades.	70
Tabela 5 – Métrica NDCG@10 calculada para o <i>ranking</i> resultante das iterações da simulação realizada para cada atividade.	70
Tabela 6 – Métrica NDCG-média calculada para as posições do <i>ranking</i> puro resultante em cada iteração das simulações realizadas para todas as atividades.	71
Tabela 7 – Métrica NDCG calculada para as recomendações resultantes das iterações da simulação realizada para cada atividades.	72
Tabela 8 – Métrica NDCG-média calculada para as recomendações derivadas de diferentes segmentações em cada iteração das simulações realizadas para todas as atividades.	73
Tabela 9 – Número médio de documentos inéditos recomendados por diferentes segmentações em cada iteração das simulações realizadas para todas as atividades.	74
Tabela 10 – Tempo médio aferido para as etapas do processo de recomendação a partir de 100 execuções de cada simulação.	74
Tabela 11 – Métrica NDCG calculada para as posições do <i>ranking</i> original da atividade A3b (A) e do <i>ranking</i> reclassificado (B) a partir da propagação do MIB construído durante a atividade A3a.	76
Tabela 12 – Métrica NDCG calculada para as posições do <i>ranking</i> original da atividade A4b (A) e do <i>ranking</i> reclassificado (B) a partir da propagação do MIB construído durante a atividade A4a.	76
Tabela 13 – Métrica NDCG-média calculada para as reclassificações e recomendações produzidas em cada iteração das simulações nas situações em que o processo de atualização do MIB ignora e trata a confiança sobre o <i>feedback</i> capturado.	77
Tabela 14 – Classificação e recomendação resultantes para cada <i>feedback</i> capturado (f.c.) durante a simulação realizada para a atividade A1.	89

Tabela 15 – Classificação e recomendação resultantes para cada <i>feedback</i> capturado (f.c.) durante a simulação realizada para a atividade A2.	89
Tabela 16 – Classificação e recomendação resultantes para cada <i>feedback</i> capturado (f.c.) durante a simulação realizada para a atividade A3a.	89
Tabela 17 – Classificação e recomendação resultantes para cada <i>feedback</i> capturado (f.c.) durante a simulação realizada para a atividade A3b.	89
Tabela 18 – Classificação e recomendação resultantes para cada <i>feedback</i> capturado (f.c.) durante a simulação realizada para a atividade A4a.	90
Tabela 19 – Classificação e recomendação resultantes para cada <i>feedback</i> capturado (f.c.) durante a simulação realizada para a atividade A4b.	90
Tabela 20 – Classificação e recomendação resultantes para cada <i>feedback</i> capturado (f.c.) durante a simulação realizada para a atividade A5.	90
Tabela 21 – Classificação e recomendação resultantes para cada <i>feedback</i> capturado (f.c.) durante a simulação realizada para a atividade A6.	90

Lista de abreviaturas e siglas

CLEF	<i>Conference and Labs of the Evaluation Forum</i>
DCG	<i>Discounted Cumulative Gain</i>
LSI	<i>Latent Semantic Indexing</i>
MAP	<i>Mean Average Precision</i>
MEV	Modelo de Espaço Vetorial
MIB	Modelo do Interesse de Busca
MRR	<i>Mean Reciprocal Rank</i>
NDCG	<i>Normalized Discounted Cumulative Gain</i>
NLTK	<i>Natural Language Toolkit</i>
NTCIR	<i>NII Testbeds and Community for Information access Research</i>
OHSUMED	<i>Oregon Health and Science University School of Medicine</i>
RI	Recuperação de Informação
SRI	Sistema de Recuperação de Informação
TF-IDF	<i>Term Frequency–Inverse Document Frequency</i>
TREC	<i>Text Retrieval Conference</i>
UCAIR	<i>User-Centered Adaptive Information Retrieval</i>

Sumário

1	Introdução	16
2	Revisão de literatura	19
2.1	A procura por informação	19
2.1.1	As atividades de procura por informação	19
2.1.2	A Pesquisa Exploratória	21
2.2	Personalização de buscas na <i>Web</i>	24
2.2.1	Modelagem dos interesses do usuário	25
2.2.2	Recuperação Interativa e <i>Feedback</i> de Relevância	27
2.3	Considerações finais	29
3	DESiRe: uma abordagem dinâmica para recomendação de resultados de busca	30
3.1	Trabalhos relacionados	31
3.2	A abordagem DESiRe	34
3.3	O Processo de recomendação	37
3.3.1	Criação do MIB	38
3.3.2	Atualização do MIB	40
3.3.3	Reclassificação e recomendação dos resultados	42
3.4	Considerações finais	43
4	Construção da coleção de teste	44
4.1	Definição das atividades de busca	46
4.2	Recuperação dos documentos	53
4.3	Julgamento de relevância	54
4.3.1	Interpretações de relevância	55
4.3.2	Avaliação da concordância entre juízes	55
4.3.3	Determinação da relevância final dos documentos	57
4.4	Avaliação da distribuição de relevância	57
4.5	Considerações finais	61
5	Avaliação da DESiRe	62
5.1	Simulação das atividades de busca	65
5.2	Avaliação do desempenho da DESiRe	68
5.2.1	Avaliação da eficácia das reclassificações	68
5.2.2	Avaliação da eficácia das recomendações	72
5.2.3	Avaliação da eficiência da DESiRe	74
5.2.4	Avaliação da propagação do MIB	75
5.2.5	Avaliação da confiança sobre o <i>feedback</i>	76

5.3	Considerações finais	77
6	Conclusão	79
	Referências	82
	APÊNDICE A – Resultados das simulações	88

1 Introdução

Nós, seres humanos, somos exploradores por natureza. Por meio da interação com o ambiente ao nosso redor, visamos satisfazer nossas necessidades de integração e aprendizagem sobre nosso mundo. Conforme exploramos, reunimos informações a fim de desenvolver diversas habilidades intelectuais complexas como a compreensão, aplicação, análise, síntese e avaliação em um determinado domínio do conhecimento (WHITE; ROTH, 2009). Segundo Maslow *et al.* (1970), essas são habilidades fundamentais para a satisfação de nossas necessidades sociais e psicológicas, favorecendo assim a busca natural pela autorrealização.

Por décadas, cientistas da informação e psicólogos têm estudado como as pessoas manipulam, armazenam, recuperam e disseminam informações em diversos ambientes e situações, como em bibliotecas, organizações, e mais recentemente, na *Web*. Como afirmam White e Roth (2009), a *Web* tem se tornado fonte primária para aquisição de conhecimento e buscas cada vez mais complexas são realizadas nesse contexto. Seus usuários recorrem constantemente a mecanismos de busca para satisfazer uma grande variedade de necessidades e desejos, os quais podem variar desde curiosidades cotidianas a pesquisas mais complexas (HEARST, 2009). Esse fenômeno pode ser justificado tanto pela facilidade de acesso proporcionada quanto pela abundância de informação disponível.

Apesar de a abundância de informação representar um aspecto positivo da *Web*, ela também apresenta um risco para a sua utilidade, acarretando uma grande sobrecarga de informação para o usuário. Simon (1971) estabelece uma relação de consumo entre a informação e a atenção despendida pelos seus receptores. Ele afirma que “a riqueza de informação resulta diretamente em pobreza de atenção”, trazendo à tona a necessidade de preservá-la de maneira eficiente em meio ao excesso de informação que pode consumi-la.

Dessa maneira, torna-se evidente que, aliado à busca por informação, o imenso volume informacional que se faz presente na *Web* tem trazido grandes desafios para os trabalhos que pretendem tratá-lo em relação aos seus usuários. Os desafios se tornam ainda maiores quando atividades complexas de busca estão envolvidas no processo de aquisição de conhecimento pelo usuário. Em buscas que envolvem atividades de investigação e aprendizagem, como na Pesquisa Exploratória (MARCHIONINI, 2006; WHITE; ROTH, 2009), o usuário geralmente não domina o assunto procurado e, portanto, não é capaz de empregar os termos adequados à consulta.

Lidar com um grande volume de informação associado à falta de clareza dos objetivos do usuário e a sua própria falta de conhecimento no domínio da pesquisa são apenas alguns dos fatores que contribuem para tais desafios. A *Web* tem se tornado

um ambiente informacional cada vez mais complexo e a grande diversidade de usuários existente também tem dificultado a recuperação de informações relevantes. Dessa forma, diferentes condições de busca somadas à falta de clareza dos objetivos do usuário dificultam a identificação do seu interesse de busca, provocando uma baixa precisão na recuperação da informação.

Como forma de tratar tais problemas, pesquisadores começaram a explorar novos caminhos para melhorar a experiência de busca do usuário. Uma das alternativas investigadas reside no conceito de **busca personalizada** (MICARELLI *et al.*, 2007; PITKOW *et al.*, 2002), o qual, diferentemente da busca tradicional, procura caracterizar individualmente os usuários para provê-los de resultados que atendam suas necessidades sob diferentes contextos. Para tanto, os sistemas de busca personalizada armazenam informações sobre as preferências e interesses dos usuários, criando modelos de perfis individualizados que possam ser integrados ao processo de recuperação de informação.

Dentre as diferentes técnicas de modelagem, um grande esforço tem sido empregado em modelos de interesse de curto prazo (SHEN; TAN; ZHAI, 2005b; HARDTKE; WERTHEIM; CRAMER, 2009; RAMAN; JOACHIMS; SHIVASWAMY, 2011; BRANDT *et al.*, 2011; SAFIYA; MARTIN; ELIZABETH, 2013). Tal abordagem permite capturar mudanças rápidas de contexto e inferir o interesse associado a uma tarefa específica de busca. Isso se faz necessário, principalmente, em atividades que envolvem investigação e aprendizado, nas quais o assunto procurado é inédito ao usuário ou difere de suas preferências gerais. Geralmente, tais abordagens consistem em construir um modelo de interesses do usuário durante uma atividade de busca específica e empregá-lo na recuperação de documentos mais relevantes sempre que o usuário realiza uma nova busca, ou na reclassificação dos documentos já recuperados quando o usuário interage com a página de resultados.

Embora apresentem abordagens eficazes para inferir o interesse do usuário e empregá-los na recuperação ou reclassificação dos resultados, os trabalhos desenvolvidos até então se restringem apenas à estrutura da página de resultados para apresentar os esforços da personalização oferecida. O mesmo ocorre para as ferramentas mais populares¹ de busca existentes atualmente, como Google Search, Yahoo Search, Ask e Bing. Tal estratégia exige que o usuário sempre recorra à página de resultados a fim de ser beneficiado pela personalização, limitando as oportunidades em que a mesma pode ser apresentada. Por outro lado, essas oportunidades podem ser ampliadas se a personalização for apresentada durante a navegação pelos documentos recuperados, tornando o processo de busca mais dinâmico.

Dessa forma, o nosso trabalho tem como objetivo explorar uma maneira mais interativa e eficaz para a apresentação de documentos de busca relevantes. Para tanto,

¹ Disponível em: <http://www.alexa.com/topsites/category/Computers/Internet/Searching/Search_Engines>. Acessado em 01/06/2015

propomos a DESiRe, uma abordagem dinâmica para recomendação de resultados de busca, baseada no interesse do usuário durante uma tarefa específica de Pesquisa Exploratória na *Web*. A recomendação é construída a partir da reclassificação dos documentos recuperados e apresentada dinamicamente ao usuário sempre que um *feedback* de interesse é capturado durante o processo de busca. A recomendação de resultados durante a navegação tem como papel antecipar a descoberta de documentos inéditos e de interesse que foram recuperados pela busca, mas que corriam o risco de não serem visualizados pelo usuário por estarem em posições mais distantes no *ranking*.

A avaliação da nossa proposta difere das realizadas pelos trabalhos relacionados por apresentar uma análise sistemática do comportamento da abordagem a partir de simulações executadas sobre uma coleção de teste. Dessa forma, ao exercitar várias iterações do processo de recomendação podemos medir a eficiência e a eficácia de suas diferentes etapas. No contexto do nosso trabalho, a eficiência e eficácia da abordagem estão respectivamente relacionadas ao tempo computacional consumido durante sua execução e à qualidade das reclassificações e recomendações produzidas. De maneira complementar, também investigamos como a informação adquirida durante uma atividade de busca pode ser propagada para as atividades seguintes e como a noção de confiança sobre o *feedback* do usuário pode alterar a qualidade da recomendação.

Nosso trabalho está organizado em seis capítulos, sendo que neste primeiro **introduzimos** o contexto em que o trabalho se encontra, as motivações para o seu desenvolvimento, e os objetivos almejados com a sua realização. O [Capítulo 2](#) apresenta a **revisão de literatura** realizada para o nosso trabalho, detalhando o contexto em que o mesmo está inserido. Nesse capítulo, associamos o comportamento dos usuários durante atividades de busca na *Web* com a complexidade desse ambiente de busca e apresentamos como alguns desafios e dificuldades envolvidos no processo de Procura por Informação na *Web* podem ser tratados. O [Capítulo 3](#) apresenta a proposta de **uma abordagem dinâmica para recomendação de resultados de busca em atividades de Pesquisa Exploratória na Web**, descrevendo a metodologia utilizada para sua construção e detalhando as etapas essenciais do processo de recomendação proposto. O [Capítulo 4](#) apresenta a **coleção de teste** que criamos para a avaliação da nossa proposta. Nele discutimos as etapas que cumprimos para a construção da coleção e os resultados obtidos a partir da sua avaliação interna. O [Capítulo 5](#) apresenta a metodologia que adotamos para a **avaliação da abordagem**, bem como os resultados obtidos por ela. Nele avaliamos as diferentes características da abordagem em relação a sua eficiência e eficácia. Por fim, o [Capítulo 6](#) apresenta a **conclusão** deste trabalho, bem como algumas direções que a abordagem deve seguir futuramente.

2 Revisão de literatura

2.1 A procura por informação

A fim de propor qualquer abordagem que apoie os usuários na recuperação de informação, precisamos primeiramente entender o processo de Procura por Informação (*Information Seeking*), incluindo as estratégias empregadas pelos usuários quando realizam buscas na *Web*. Vários modelos de Procura por Informação foram propostos, alguns focando os aspectos temporais, outros os comportamentos que os indivíduos apresentam enquanto procuram informações. [Hearst \(2009\)](#) apresentou várias tentativas de descrever o processo de Procura por Informação e as atividades envolvidas pelo mesmo. Ela cita que o modelo padrão de Procura por Informação foi inicialmente proposto por [Marchionini \(1989\)](#) como “um caso especial de resolução de problemas” e, de maneira semelhante a outros pesquisadores ([SUTCLIFFE, 1998](#); [MARCHIONINI; WHITE, 2007](#)), define que o processo deve incluir:

1. Reconhecer e interpretar o problema de informação;
2. Estabelecer um plano de busca;
3. Realizar a busca;
4. Avaliar os resultados e, se necessário, interagir com o processo novamente.

A execução dessas etapas pelo usuário a fim de resolver um determinado problema de informação (ou problema de busca) resulta no que denominamos de **atividade de busca**. Dependendo da complexidade da atividade de busca envolvida, essas etapas nem sempre estarão claras para o usuário, exigindo que o mesmo realize diversas iterações durante a realização do processo. A fim de distinguir as possíveis iterações sobre o processo de busca, denominaremos um conjunto de atividades que visam resolver um mesmo problema de busca como uma **tarefa de busca**.

2.1.1 As atividades de procura por informação

Segundo [Marchionini \(2006\)](#), existem três tipos de atividades que podem ser realizadas na Procura por Informação. A primeira delas, a atividade de **procura** (*lookup*), é a mais

comum e tem sido o foco dos principais mecanismos atuais de busca na *Web*. Geralmente as tarefas de procura são adequadas para estratégias de busca analítica, na qual consultas bem especificadas geram resultados precisos, sem a necessidade de verificação e comparação dos itens recuperados. O paradigma predominante é o de “consulta – resposta”, no qual consultas são emitidas ao mecanismo de busca e um conjunto de itens potencialmente relevantes ao usuário é oferecido como resposta (WHITE; ROTH, 2009). Esse paradigma exige que o consumo de informação pelo usuário seja realizado de maneira independente do mecanismo de busca, ao navegar dentre um grande volume de documentos orientado apenas por sua necessidade de informação e seu instinto.

Entretanto, a *Web* tem se tornado fonte primária para aquisição de conhecimento e buscas cada vez mais complexas são realizadas nesse contexto, exigindo múltiplas iterações de consulta, navegação pós-consulta e exame detalhado dos resultados. Aliadas ao crescimento do volume de informação disponível, Marchionini (2006) define outras duas atividades de busca relacionadas à Procura por Informação, desta vez, focadas na aquisição de conhecimento pelo usuário. As atividades de **aprendizagem** e de **investigação**, como são definidas pelo autor, envolvem diversas iterações de busca, requerendo que o usuário desempenhe um grande esforço cognitivo em análises, comparações e julgamento dos documentos recuperados. Os usuários que desenvolvem essas atividades geralmente são caracterizados por almejam a aquisição de conhecimento, compreensão de conceitos, habilidades e interpretação de ideias.

Buscas relacionadas ao aprendizado são cada vez mais frequentes, ao passo que materiais primários são disponibilizados na *Web* (WHITE; ROTH, 2009). Tais buscas envolvem múltiplas iterações e retornam objetos instanciados de várias mídias (textos, gráficos, imagens, vídeos), requerendo do usuário processamento e interpretação cognitiva para a avaliação e julgamento qualitativo sobre eles. Marchionini utiliza a terminologia da taxonomia de Bloom (BLOOM, 1984) para correlacionar os objetivos das buscas que suportam a aprendizagem, como a aquisição de conhecimento, compreensão de conceitos ou habilidades, interpretação de ideias, e comparação ou agregação de dados e conceitos.

Assim como as buscas por aprendizagem, as buscas que apoiam a investigação também envolvem múltiplas iterações, porém, essas podem ocorrer durante períodos longos de tempo e podem retornar resultados que serão criticamente avaliados antes de serem integrados à base de conhecimento do usuário. Buscas investigativas visam alcançar os níveis mais superiores da taxonomia de Bloom, tais como análise, síntese, e avaliação, e requerem um conhecimento prévio substancial. Além de encontrar novas informações, buscas investigativas procuram descobrir lacunas no conhecimento de forma que novas pesquisas possam ser iniciadas e becos sem saída possam ser evitados (MARCHIONINI, 2006).

De maneira complementar a Marchionini (2006), Broder (2002) propõe que as buscas realizadas na *Web* podem ser classificadas como **navegacionais**, quando a intenção

da busca é atingir uma determinada página previamente conhecida; **transacionais**, quando a intenção é realizar alguma atividade mediada pela *Web*, como o consumo de serviços e o comércio de bens; e **informacionais**, quando a intenção é adquirir alguma informação que se assume estar presente em uma ou mais páginas *Web*.

Dessa forma, torna-se evidente a existência de uma correlação entre as atividades de busca informacionais definidas por Broder e as atividades de investigação e aprendizagem apresentados por Marchionini. No decorrer do nosso trabalho, tomaremos a liberdade de tratar as atividades de busca como “informacionais” quando seu objetivo for a aquisição de informação, seja como fim a aprendizagem ou a investigação sobre um determinado assunto.

2.1.2 A Pesquisa Exploratória

Atualmente, motores de busca como Google Search¹, Bing², Yahoo Search³, DuckDuckGo⁴, dentre outros, apoiam de forma satisfatória os usuários que possuem sua necessidade de informação bem definida, dado o grande investimento realizado por seus desenvolvedores em tecnologias de *ranking* e respostas instantâneas (WHITE; ROTH, 2009). Situações de busca como “indicados ao oscar 2015” e “comprar série livros Harry Potter” são consideradas bem definidas, já que grande parte das informações necessárias para a recuperação dos resultados se encontra nos próprios termos da consulta. Entretanto, em situações em que a necessidade de informação é vaga, ou não está completamente definida, os usuários normalmente partem de uma busca inicial mais abrangente e, de forma incremental, refinam a consulta a partir da leitura e entendimento dos documentos recuperados. Marchionini (2006), White, Kules e Bederson (2005) denominam essa atividade de busca e refinamento como **Pesquisa Exploratória**, uma atividade na qual os usuários executam buscas com objetivos abertos ou abstratos a fim de construir conhecimento sobre um determinado assunto.

De maneira geral, pesquisadores e estudantes que necessitam dar início a novos trabalhos, bem como jornalistas e professores que devem estudar sobre um assunto de interesse, mas não familiar, são alguns exemplos de usuários que desempenham atividades de Pesquisa Exploratória. Geralmente, a busca realizada nessas situações é motivada por problemas complexos, nos quais a terminologia utilizada é desconhecida ou mal compreendida, ou ainda o conhecimento da estrutura do espaço de informação é insuficiente.

Para satisfazer sua necessidade de informação, o usuário explora o espaço de busca procurando por informações que sejam úteis e de seu interesse. Entretanto, uma vez que o espaço de busca é vasto e quase desconhecido, tal exploração é geralmente ineficiente. Para usuários leigos, o problema se torna ainda mais difícil, o que pode frustrar completamente

¹ Disponível em: <<https://www.google.com>>. Acessado em: 01/06/2015

² Disponível em: <<http://www.bing.com>>. Acessado em: 01/06/2015.

³ Disponível em: <<http://search.yahoo.com>>. Acessado em: 01/06/2015.

⁴ Disponível em: <<https://duckduckgo.com>>. Acessado em: 01/06/2015.

seus esforços. Vale ressaltar que o termo leigo não se refere apenas à falta de habilidade com o mecanismo de busca, mas também ao desconhecimento do próprio assunto que está sendo buscado. Tal desconhecimento implica na incapacidade de expressar adequadamente a necessidade de informação e, conseqüentemente, em uma recuperação de baixa qualidade.

A [Figura 1](#), em uma versão adaptada e traduzida de [Marchionini \(2006\)](#), ilustra os três tipos de atividades de busca definidos pelo autor e destaca a aprendizagem e a investigação como componentes-chave da Pesquisa Exploratória. O formato de nuvens sobrepostas sugere que as três atividades podem ocorrer paralelamente e de maneira mesclada durante o processo de busca. Abaixo de cada atividade são listados os seus respectivos objetivos.



Figura 1 – As atividades de busca definidas por [Marchionini \(2006\)](#).

Segundo [White e Roth \(2009\)](#), a Pesquisa Exploratória é um conceito que está sendo continuamente modificado e moldado por uma variedade de pesquisas relacionadas. Entretanto, eles caracterizam a Pesquisa Exploratória apresentando alguns atributos que a diferenciam de outros tipos de busca por informação:

1. As atividades de Pesquisa Exploratória podem ser constituídas de múltiplas consultas e sessões que podem se estender por dias, semanas, ou meses, dependendo da natureza da tarefa de busca (uma busca por informação pertinente a compra de um *smartphone* pode tomar menos tempo do que o conjunto de buscas relacionadas à pesquisa realizada para uma dissertação de mestrado, por exemplo). O aspecto temporal de uma Pesquisa Exploratória pode ser apoiado por sistemas capazes de armazenar

consultas e históricos de longo-prazo que retenham informações de preferências do usuário sobre diversas sessões de busca.

2. A necessidade de informação que motiva uma Pesquisa Exploratória é geralmente indeterminada, persistente, e multifacetada. O indeterminismo é relacionado à incerteza sobre a informação disponível, ou à falta da informação sobre a natureza do problema de busca. Pesquisas exploratórias são baseadas em uma mescla de curiosidades gerais e específicas, e enfatizam o aprendizado e investigação.
3. O objetivo da Pesquisa Exploratória vai além de simplesmente localizar informações em função das atividades associadas com aprendizagem e entendimento. Isto é, a tarefa de busca não existe de maneira isolada do contexto das tarefas ao seu redor. Pesquisas exploratórias são geralmente conduzidas para auxiliar pessoas a fazerem decisões mais informadas ou para melhorar seu entendimento em um determinado tópico.
4. Os comportamentos de iteração observados durante uma Pesquisa Exploratória são geralmente uma combinação de navegação e busca focada, com maior ênfase na navegação. As pessoas utilizam a navegação como forma de resolver a incerteza e confusão que pode ocorrer conforme novas informações são encontradas.
5. Pesquisas exploratórias podem envolver a colaboração de várias pessoas de maneira síncrona ou assíncrona. Dada a forte relação entre a Pesquisa Exploratória, o uso da informação e o entendimento da informação, é provável que ocorra participações de outras pessoas durante o processo de busca.
6. A avaliação de sistemas que suportam a Pesquisa Exploratória requer uma metodologia que tenha como alvo a aprendizagem e intuição, bem como resultados de tarefas e utilidade do sistema. Para determinar quão bem sistemas suportam atividades de Pesquisa Exploratória, eles devem ser avaliados em termos de suas habilidades em facilitar os elementos chave da exploração (como auxiliar os usuários a obter novos *insights*, assistir o aprendizado e oferecer suporte para a realização de decisões críticas).

As características apresentadas por [White e Roth \(2009\)](#) sugerem pontos importantes que devem ser tratadas durante o desenvolvimento de ferramentas que apoiem a Pesquisa Exploratória. Embora não tenhamos a pretensão de tratar todos os pontos descritos, podemos destacar a incerteza que permeia a busca, a multiplicidade das sessões de busca com enfoque na navegação e a avaliação focada no aprimoramento do usuário como as principais características abordadas pelo nosso trabalho.

2.2 Personalização de buscas na *Web*

Conforme a *Web* se populariza, os Sistemas de Recuperação de Informação (SRI) precisam ser capazes de acomodar uma grande diversidade de usuários e suas respectivas habilidades. Em contrapartida, a maioria dos usuários não consegue lidar corretamente com o modelo de “consulta – resposta” implementado por grande parte dos sistemas atuais de busca (WHITE; ROTH, 2009). Esses mecanismos e as técnicas em que se baseiam assumem que consultas precisas e bem formadas serão fornecidas pelos usuários, o que atualmente está distante da realidade na *Web*. Em vez disso, muitas das consultas na *Web* são vagas ou ambíguas e não representam a real necessidade de informação do usuário, que por vezes falha em ser satisfeita.

Pesquisas recentes mostram que mesmo os principais mecanismos de busca ainda sofrem com o insucesso na recuperação de resultados relevantes. Em um estudo com mais de 20.000 consultas, Coyle e Smyth (2007) mostraram que em 52% dos casos, os usuários não encontravam o que esperavam dentre o conjunto de resultados retornado pelos buscadores. Em um trabalho mais recente, utilizando técnicas mais aprimoradas, Nettleton e Gonzalez-Caro (2012) observaram uma frequência de insucesso de 34%, apontando que esse problema ainda persiste de forma acentuada.

Outro fator que corrobora para tal insucesso se deve à desconsideração do contexto de busca pelos SRI atuais. O contexto de busca pode ser entendido como o conjunto de assuntos (ou tópicos) associados a uma atividade de busca desenvolvida pelo usuário e está geralmente relacionado às suas intenções e interesses específicos, os quais podem diferir de suas preferências gerais. Uma vez que a relevância do conteúdo retornado pela busca está diretamente relacionada aos interesses do usuário, ela será mal estimada se tais interesses destoarem de suas preferências gerais. Por exemplo, um usuário que possua preferência por automóveis pode, em um dado momento, estar interessado em saber a “velocidade máxima atingida por um jaguar”, em que a palavra “jaguar” refere-se ao animal, e não ao carro fabricado por essa marca.

Em resposta à falta de precisão dos resultados, pesquisadores começaram a explorar novos caminhos para melhorar a experiência de busca do usuário. Uma visão amplamente aceita é a de que as buscas na *Web* necessitam se tornar cada vez mais personalizadas para que as necessidades individuais dos usuários sejam satisfeitas (MICARELLI *et al.*, 2007). O conceito de **personalização** discute a natureza generalista das buscas, na qual usuários distintos utilizando a mesma consulta receberão aproximadamente a mesma lista de resultados, apesar de seus interesses serem diferentes. Informações adicionais sobre os usuários, como suas preferências e seu contexto de busca, podem ser usadas para fornecer resultados que correspondam melhor aos seus interesses individuais.

O principal objetivo da personalização de buscas na *Web* é adaptar os resultados de busca de forma individualizada para cada usuário, baseando-se em seus perfis de características, interesses e preferências. Para tanto, os mecanismos de busca personalizada

procuram reconhecer o verdadeiro objetivo das consultas submetidas construindo modelos capazes de representar a necessidade de informação dos usuários. Estudos recentes como os elaborados por Safiya, Martin e Elizabeth (2013), White *et al.* (2013), Veningston e Shanmugalakshmi (2012), Matthijs e Radlinski (2011), têm concentrado esforços na modelagem automática dos interesses dos usuários para provê-los de resultados de busca personalizados.

2.2.1 Modelagem dos interesses do usuário

Segundo Micarelli *et al.* (2007), o simples fato de rastrear as páginas visitadas pelo usuário e suas consultas submetidas ao mecanismo de busca representa uma técnica de modelagem de seus interesses, a partir da qual características importantes podem ser aprendidas e então utilizadas para se obter resultados mais relevantes. De maneira geral, o modelo do perfil do usuário (modelo do usuário) é o principal componente utilizado para fornecer resultados adaptados às necessidades particulares de cada usuário.

Os modelos podem consistir desde questionários simples sobre as preferências explícitas do usuário até estruturas informacionais dinâmicas capazes de determinar o conhecimento do usuário sobre determinado assunto. Por exemplo, o protótipo ifWeb (ASNICAR; TASSO, 1997) faz uso de modelos baseados em redes semânticas para criar uma representação dos tópicos de interesse do usuário e então filtrar os resultados recuperados pelo mecanismo de busca. Já o trabalho realizado por Wang *et al.* (2011) utiliza redes neurais com *backpropagation* para modelar o perfil do usuário e então otimizar o processo de personalização.

Outro exemplo é a busca personalizada implementada pelo buscador Google, o qual analisa o comportamento de navegação do usuário armazenado pela ferramenta Web History⁵ para sugerir resultados relacionados aos seus interesses. O modelo ainda é composto pelo contexto espacial do usuário, *i.e.*, a região na qual o usuário realiza a busca, e por dados associados a sua rede social Google Plus⁶.

De acordo com Micarelli *et al.* (2007), a modelagem do perfil do usuário pode afetar a Recuperação de Informação (RI) em três diferentes situações, como ilustra a Figura 2. Na situação (a), o modelo do usuário é empregado como parte do processo de recuperação. A classificação é um processo unificado no qual o perfil do usuário é utilizado para pontuar os documentos. Já na situação (b), o modelo é empregado como um agente reclassificador externo. O perfil do usuário tem sua participação em uma etapa posterior à recuperação, reclassificando o conjunto de documentos de acordo com as características representadas pelo modelo. Por fim, a situação (c) emprega o modelo na modificação da consulta do usuário. O modelo afeta diretamente a representação da necessidade de informação do usuário, a modificando ou a estendendo.

⁵ Disponível em: <<https://support.google.com/websearch/answer/54068>>. Acessado em 01/06/2015.

⁶ Disponível em: <<https://plus.google.com>>. Acessado em 01/06/2015.

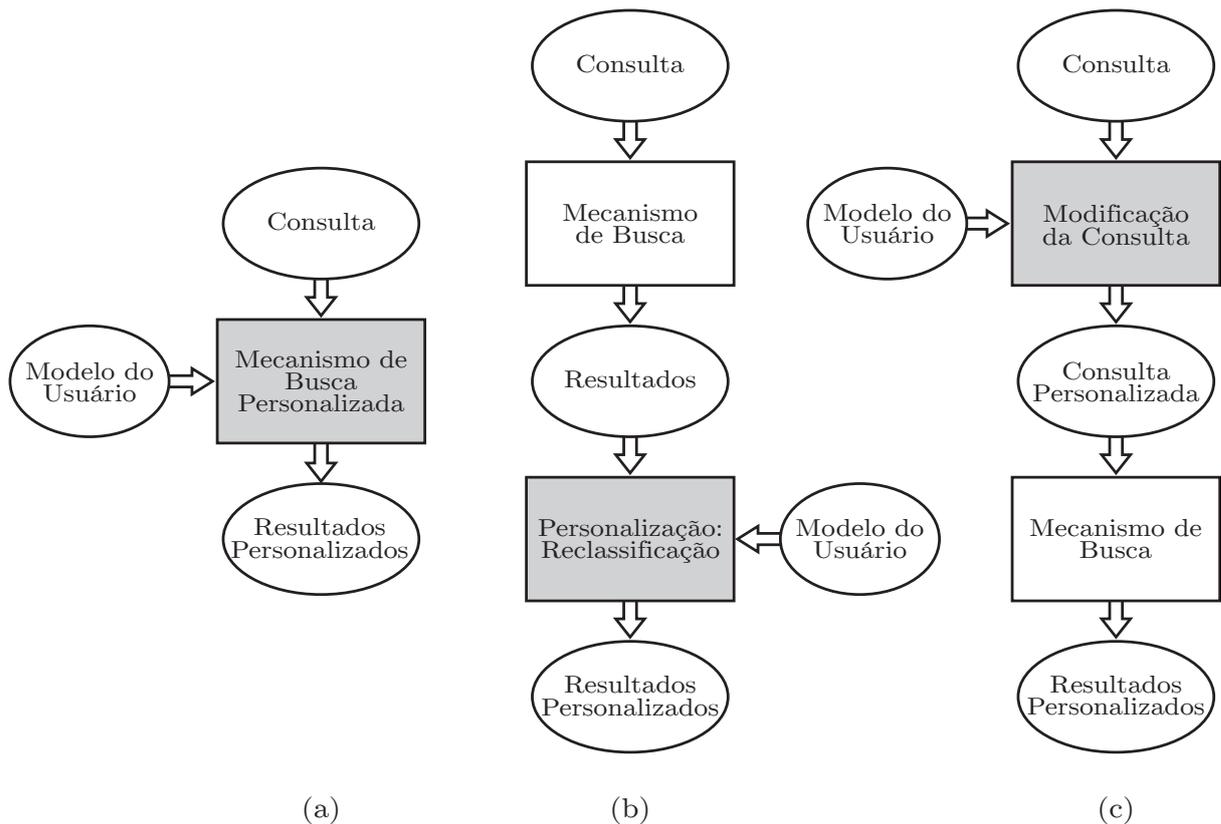


Figura 2 – Aplicação do modelo do usuário em diferentes etapas do processo de personalização. Adaptado de Micarelli *et al.* (2007).

Micarelli *et al.* (2007) sugere que a primeira técnica é mais propensa a fornecer respostas rápidas à consulta do usuário, já que o sistema tradicional de classificação pode ser adaptado diretamente para incluir a personalização, evitando processamento repetido ou supérfluo. Entretanto, devido a fortes restrições de tempo de resposta, grande parte dos mecanismos de busca simplesmente não implementam a personalização, uma vez que ela pode gerar uma sobrecarga no processo de classificação.

Por outro lado, reclassificar os documentos utilizando um agente externo ao SRI permite ao sistema manter o processo de classificação convencional e ao mesmo tempo fornecer a possibilidade da personalização seletiva por parte do usuário. Diversos sistemas empregam essa abordagem no lado cliente, (MICARELLI; SCIARRONE, 2004; PITKOW *et al.*, 2002; SPERETTA; GAUCH, 2005), no qual uma ferramenta se conecta ao mecanismo de busca e recupera um conjunto de documentos para ser analisado localmente. Para evitar o esforço de armazenar e processar cada documento que é apresentado na lista de resultados, a análise é geralmente aplicada apenas aos recursos melhor classificados, ou então apenas a descrição (*snippet*) associada a cada resultado é analisada.

Outro aspecto importante da modelagem dos interesses do usuário é a maneira com que o conhecimento sobre tais interesses é adquirido. Basicamente, podemos distinguir o conhecimento entre dados sobre o usuário (*user data*) e dados sobre o uso da ferramenta de

busca pelo usuário (*usage data*). Dessa forma, enquanto algumas abordagens analisam as características pessoais do usuário, outras empregam técnicas de mineração de dados para capturar a informação comportamental do usuário. De certa forma, as duas abordagens são complementares e trabalham para adquirir o conhecimento necessário a fim de construir um modelo de interesse de busca.

No âmbito das abordagens orientadas a *usage data*, o comportamento do usuário pode ser capturado de duas maneiras em relação à relevância dos documentos visitados. A primeira maneira, por meio de *feedbacks* de relevância explícitos, permite que o usuário informe explicitamente documentos ou termos que devem ser utilizados no aprimoramento do modelo de interesse de busca. Embora simples, esse tipo de *feedback* é raramente utilizado, uma vez que ele requer interações diretas com usuário, acarretando uma carga cognitiva adicional ao mesmo (JANSEN; SPINK; SARACEVIC, 1999).

Como forma de superar o problema gerado pelo *feedback* explícito, pesquisadores exploraram o uso de *feedbacks* de relevância implícitos. Com o *feedback* implícito, o comportamento do usuário, como cliques em *links*, a rolagem da página visitada e o tempo despendido na leitura dos documentos é discretamente monitorado e utilizado para expandir a compreensão da necessidade de informação do usuário após a consulta (WHITE; RUTHVEN; JOSE, 2002). A interação dos dois tipos de *feedbacks* para a construção de um modelo do perfil do usuário pode ser melhor entendido pelo diagrama ilustrado pela Figura 3.

2.2.2 Recuperação Interativa e *Feedback* de Relevância

Em termos gerais, a Recuperação Interativa refere-se a qualquer ambiente de recuperação no qual o sistema e seus usuários interagem durante uma tarefa de busca. Em um alto nível, as interações podem ser classificadas em duas dimensões: interações orientadas ao sistema e interações orientadas ao usuário. As interações orientadas ao sistema envolvem solicitações diretas aos usuários para que forneçam explicitamente *feedbacks* a fim de compreender melhor a necessidade de informação envolvida na busca. Exemplos incluem solicitar aos usuários que selecionem os documentos mais relevantes a partir de um conjunto apresentado (TEEVAN, 2008; TEEVAN; DUMAIS; HORVITZ, 2005), ou solicitar aos usuários que selecionem facetas (ou subtópicos) que melhor descrevam o seu interesse (WHITE; MORRIS, 2007). Embora sejam intrusivas à experiência de busca do usuário, tais abordagens podem ser eficazes quando estamos lidando com consultas difíceis (TEEVAN; DUMAIS; HORVITZ, 2005).

Já as interações orientadas ao usuário envolvem ambientes nos quais o sistema de recuperação fornece recomendações que podem ser aceitas, rejeitadas ou julgadas pelos usuários. Normalmente, o sistema de recuperação responde a uma sequência de consultas, dentro de uma sessão em particular, com recomendações de resultados adicionais ou sugestões de reformulação da consulta (LEE; LIU; CHO, 2005; AULA; MAJARANTA;

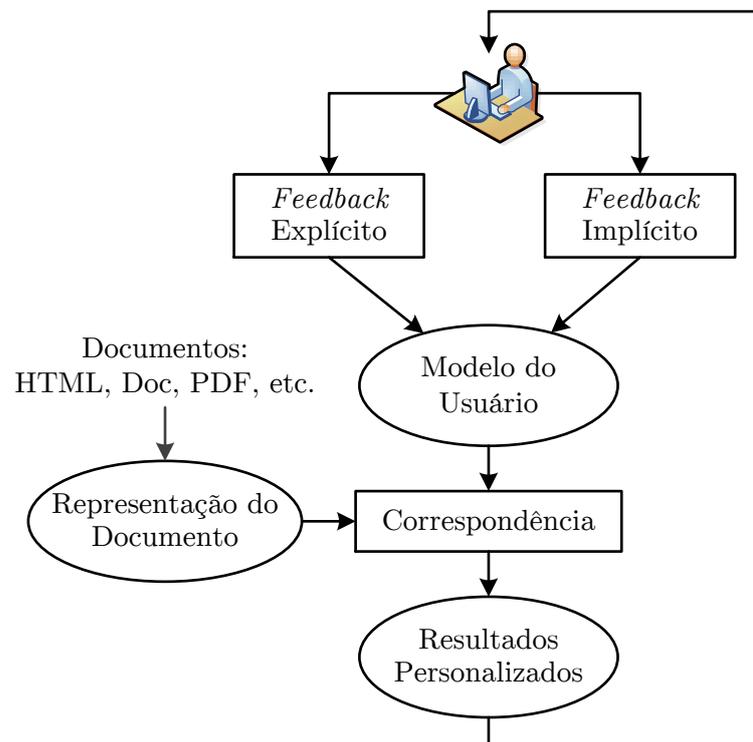


Figura 3 – Interação entre *Feedbacks* implícitos e explícitos para coletar informações sobre o perfil do usuário e mantê-lo atualizado durante a personalização. Figura adaptada de Micarelli *et al.* (2007).

RÄIHÄ, 2005). Nessa estratégia, os usuários fornecem *feedbacks* implícitos quando acessam os resultados ou aceitam a reformulação da consulta sugerida, por exemplo. Tal *feedback* pode então ser usado para fornecer recomendações mais úteis para as consultas seguintes durante a tarefa de busca (BLANCO *et al.*, 2011).

Dentre as possibilidades de interação entre o sistema e o usuário, o *Feedback* de Relevância é o método primário pós-consulta para aprimorar automaticamente a representação computacional do interesse de busca do usuário. Em sistemas que empregam esse método, *feedbacks* fornecidos pelos usuários ou juízes especialistas são utilizados para aprimorar a consulta a fim de gerar modelos mais detalhados sobre a necessidade de informação envolvida na busca (RUTHVEN; LALMAS, 2003). Uma vez que as interações do usuário, por definição, fornecem *feedbacks* para o sistema de recuperação, a técnica de *Feedback* de Relevância pode ser vista como um caso especial da Recuperação Iterativa.

Os modelos mais conhecidos de *Feedback* de Relevância incluem abordagens baseadas no Modelo de Espaço Vetorial (MEV) (ROCCHIO, 1971; SALTON; BUCKLEY, 1997), em modelos linguísticos (PONTE, 2000; RADLINSKI *et al.*, 2009), e na expansão de consulta (BUCKLEY *et al.*, 1995; CRONEN-TOWNSEND; ZHOU; CROFT, 2004; CAO *et al.*, 2008). Embora distintos, todos os métodos seguem a mesma motivação de que os interesses codificados no modelo aprimorado devem se aproximar aos documentos fornecidos pelo

Feedback de Relevância. Dessa forma, o modelo pode ser utilizado para a recuperação de documentos mais relevantes em relação ao interesse apresentado pelo usuário.

2.3 Considerações finais

Neste capítulo apresentamos os conceitos-chave presentes na literatura que fundamentam o nosso trabalho. Dentre eles destacamos o processo de Procura por Informação ao qual os usuários se submetem quando estão interessados em satisfazer alguma necessidade de informação. Ainda relacionados a esse conceito, apresentamos os tipos de atividades de busca que podem ser desempenhadas a fim de se satisfazer uma necessidade de informação. Como discutimos, daremos atenção às atividades de busca informacionais, em especial as de Pesquisa Exploratória, já que é esse tipo de atividade que fornece os maiores desafios encontrados durante a busca e que mais carece de atenção dos buscadores atuais.

Além disso, comentamos como o conceito de personalização permite às ferramentas de busca empregar informações adquiridas sobre o perfil do usuário para que resultados mais relevantes e alinhados aos seus interesses sejam recuperados. A modelagem do perfil do usuário baseada na sua interação com a ferramenta de busca permite que seus interesses sejam representados com maior precisão, proporcionando a recuperação de documentos mais relevantes a sua necessidade de informação.

No capítulo a seguir apresentamos alguns trabalhos relacionados ao nosso que utilizam o conceito de recuperação interativa para representar os interesses do usuário durante a busca e provê-lo com resultados mais relevantes. Baseado nesses trabalhos, justificamos e apresentamos a nossa proposta no mesmo capítulo.

3 *DESiRe: uma abordagem dinâmica para recomendação de resultados de busca*

Como apresentado nas seções anteriores, a dificuldade para recuperar informações relevantes e de qualidade na *Web* tem se agravado a medida que a complexidade e disponibilidade de recursos informacionais aumentam. Grande parte dessa dificuldade decorre da deficiência dos buscadores em identificar o contexto informacional da busca que o usuário está realizando. Esse problema se acentua ainda mais quando o usuário está engajado em atividades complexas de busca como a de Pesquisa Exploratória, a qual geralmente não é suportada de maneira adequada pelos mecanismos de busca atuais. Em atividades desse tipo, o usuário geralmente se mostra inseguro em relação ao espaço de busca investigado e não é capaz de expressar corretamente sua necessidade de informação, resultando em uma recuperação ineficaz.

Como forma de suprir tal deficiência, diversos trabalhos têm proposto ferramentas e técnicas associadas aos conceitos de recuperação adaptativa de informação e busca personalizada com a finalidade de apoiar os usuários que realizam esse tipo de atividade (SHEN; TAN; ZHAI, 2005b; LI *et al.*, 2007; HUTCHISON; MITCHELL, 2007; RAMAN; JOACHIMS; SHIVASWAMY, 2011; SAFIYA; MARTIN; ELIZABETH, 2013; WANG *et al.*, 2013). De maneira geral, tais propostas sugerem que modelos capazes de representar o perfil do usuário sejam empregados durante a busca, permitindo a recuperação de documentos mais relevantes e associados às preferências e interesses do usuário (GAUCH *et al.*, 2007).

Entretanto, o perfil do usuário nem sempre é suficiente para representar com precisão o seu interesse de busca. Geralmente, em atividades que envolvem investigação e aprendizado, a necessidade de informação pode estar associada a um interesse de busca específico, o qual pode não corresponder às preferências globais representadas pelo perfil do usuário. Em tais situações, os usuários procuram por assuntos inéditos que não estão presentes em seus históricos de busca e, portanto, não estão representados em seus perfis. Dessa forma, modelos capazes de representar o interesse específico do usuário durante uma busca se fazem necessários quando atividades desse tipo são desempenhadas. Diferentemente dos modelos aplicados para a construção do perfil do usuário, esses modelos devem ser construídos dinamicamente conforme a busca é realizada.

3.1 Trabalhos relacionados

Embora diversos trabalhos tenham empregado estratégias para modelar o perfil do usuário durante as atividades de busca, eles geralmente não exploram a possibilidade de aplicá-lo dinamicamente, reprimindo a possibilidade de beneficiar os usuários logo que evidências sobre seus interesses são identificadas. Como forma de suprir tal necessidade, vários trabalhos têm utilizado o conceito de **reclassificação dinâmica** para fornecer resultados mais relevantes sempre que o perfil do usuário é atualizado. O termo “reclassificação dinâmica” foi formalmente introduzido por Brandt *et al.* (2011), porém, outros trabalhos precedentes já exploravam esse mesmo conceito.

Shen, Tan e Zhai (2005a) investigaram SRIs adaptativos ao utilizarem informações do histórico do usuário bem como o seu comportamento de cliques em resultados com a finalidade de aprimorar os algoritmos de recuperação. Os autores propuseram a utilização de modelos linguísticos para representar a necessidade de informação do usuário, criando para tanto quatro modelos estatísticos capazes de incorporar informações do contexto de busca a um modelo de recuperação baseado no método de divergências de Kullback-Leibler (ZHAI; LAFFERTY, 2001). Para avaliar os modelos propostos, os autores utilizaram uma coleção de teste composta por 30 tópicos extraídos da TREC e respectivos históricos de buscas e dados de acessos aos resultados construídos sobre os tópicos escolhidos. Sobre a coleção de teste, os autores construíram uma interface de busca adaptativa, na qual três voluntários conduziram buscas sobre os tópicos definidos. Os usuários podiam consultar, reformular a consulta, examinar o sumário dos documentos e examinar os documentos. Utilizando a métrica *Mean Average Precision* (MAP), os autores verificaram que quando utilizada a informação de contexto da busca, isto é, o histórico de busca aliado aos dados de acesso, a precisão melhorou até 92% em relação a recuperação sem o contexto. Além disso, foi verificado que grande parte desse aumento foi ocasionado pelo uso da informação sobre os cliques do usuário (*usage data*), a qual sozinha proporcionou um aumento próximo de 100% para os 20 primeiros documentos recuperados.

Posteriormente, utilizando o mesmo conceito, os autores construíram o *User-Centered Adaptive Information Retrieval* (UCAIR) (SHEN; TAN; ZHAI, 2005b), uma interface experimental de busca adaptativa que utiliza o *feedback* implícito do usuário para reclassificar dinamicamente os resultados de busca. O agente de busca proposto é capaz de reformular a consulta e reclassificar ativamente os resultados se baseando em um modelo de usuário orientado ao contexto. Sempre que um novo resultado é visitado pelo usuário ou uma nova consulta é realizada, o agente recalcula seu interesse de busca e reclassifica os resultados apresentados. A Figura 4 ilustra a página de resultados original (a) e a versão reorganizada pela abordagem logo após o modelo do usuário ser recalculado (b). Os resultados que residiam em posições mais profundas no *ranking* e emergiram após a reclassificação são destacados por um indicador junto ao sumário do documento.

The figure consists of two side-by-side screenshots of a search results page for 'Jaguar'.
 (a) **Página original:** Shows a list of search results. The top result is 'Jaguar' from www.jaguar.com/. Below it is a Wikipedia entry for 'Jaguar'. A button labeled 'Visita' is next to the second Wikipedia result. A box labeled 'Resultados ainda não visualizados' points to the bottom of the list.
 (b) **Página reorganizada:** Shows the same search results but re-ranked. The top result is 'Jaguar Cars - Wikipedia, the free encyclopedia'. A box labeled 'Resultado visitado' points to the top result. A box labeled 'Resultados elevados' points to the second and third results, indicating they have moved up in the ranking.

(a) Página original.

(b) Página reorganizada.

Figura 4 – Reorganização da página de resultados proposta por Shen, Tan e Zhai (2005b).

Os autores avaliaram o agente solicitando a seis participantes que utilizassem o sistema para realizar buscas sobre diferentes tópicos e que avaliassem manualmente a relevância dos 30 resultados melhor classificados. Os resultados apresentados consistiam de uma mistura entre classificações do Google e do UCAIR, de forma que os participantes não pudessem distingui-los. As avaliações realizadas pelos os autores mostraram que as reclassificações geradas pelo UCAIR proporcionaram um aumento de 20% na precisão média do *ranking* em relação aos resultados apresentados pelo Google.

Marchi (2010) também propôs uma metodologia semelhante, porém empregando uma adaptação da técnica de *Latent Semantic Indexing* (LSI) para modelar o perfil do usuário durante a realização de suas buscas. O modelo é construído e aplicado de forma incremental a partir dos documentos selecionados pelo usuário e então é empregado na reclassificação dos resultados sempre que um novo documento relevante é acessado pelo usuário. A adaptação sugerida pela autora consiste em reduzir a dimensão do modelo aplicando um corte de 80% no conjunto de termos extraídos dos documentos recuperados, viabilizando a aplicação da técnica de LSI na construção do modelo. A avaliação realizada mostrou que a proposta apresenta uma melhora de até 23% em relação à classificação gerada pelo Google. Além disso, simulações com usuários mostraram que a precisão média obtida a partir da navegação em cinco resultados foi 13% superior à média obtida a partir da navegação sobre o primeiro documento relevante, *i.e.*, sem a aplicação do modelo para a reclassificação.

Apesar de os três trabalhos apresentarem abordagens eficazes para a seleção de resultados relevantes, a apresentação dos mesmos ainda é realizada sob o mesmo paradigma da página de resultados. Apresentar os resultados reclassificados aproveitando a mesma disposição em lista apresentada pela busca inicial pode dificultar o entendimento da

abordagem pelo usuário, uma vez que a percepção da lista de resultados é alterada a cada nova classificação (TEEVAN, 2008).

Como forma de resolver tal problema, Raman, Joachims e Shivaswamy (2011) propuseram uma abordagem de classificação dinâmica que permite adaptar a página de resultados em dois níveis, superando as restrições presentes na classificação convencional. Nessa abordagem, a interação do usuário com o primeiro nível da classificação é utilizada para inferir o seu interesse, possibilitando que um segundo nível contendo resultados mais relevantes seja inserido na classificação. Essa abordagem permite apresentar a reclassificação sugerida de maneira mais organizada e efetiva ao usuário. A Figura 5 ilustra a página de resultados reorganizada após a inferência do interesse do usuário.

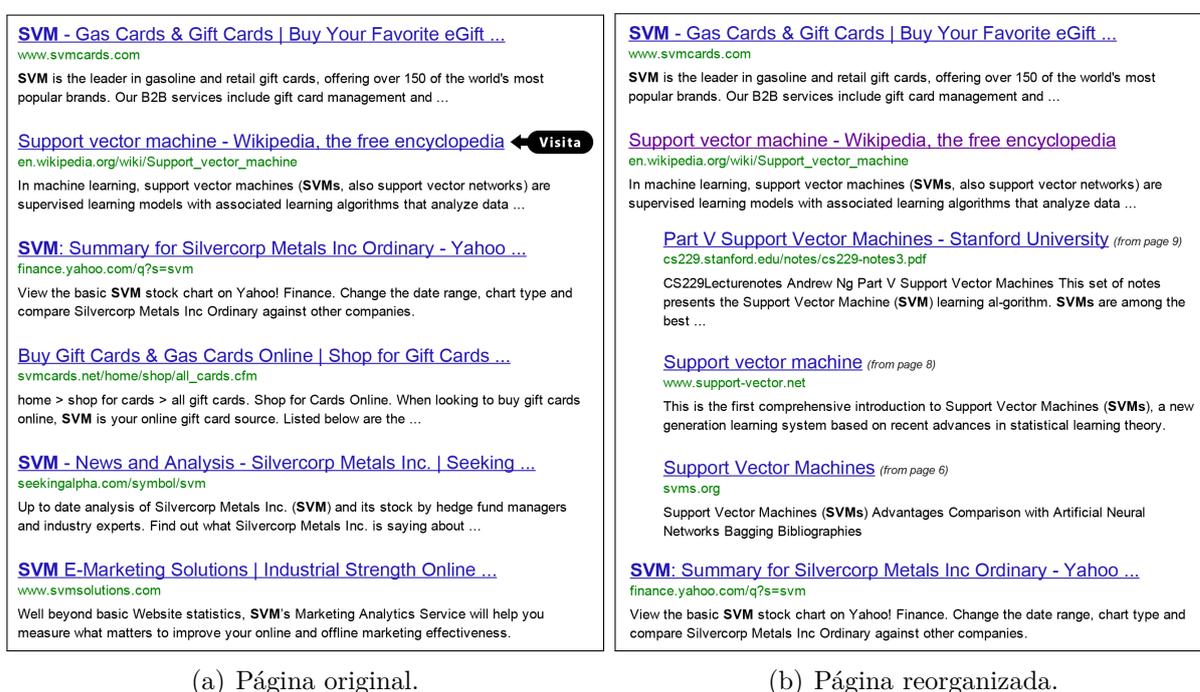


Figura 5 – Reorganização da página de resultados proposta por Raman, Joachims e Shivaswamy (2011).

Em experimentos realizados com coleções de teste, os autores verificaram que a abordagem de reclassificação de duplo nível proporcionou um aumento de até 60% na precisão da recuperação quando comparada à recuperação de apenas um nível, isto é, sem o emprego da classificação dinâmica.

Já Brandt *et al.* (2011) teorizaram o conceito de classificação dinâmica ao propor e avaliar um modelo genérico capaz de construir *rankings* dinamicamente a partir do *feedback* interativo do usuário. O modelo interpreta o *ranking* utilizando uma árvore de decisão, em vez da classificação utilizada em SRI convencionais. Junto ao modelo, um *framework* teórico também foi proposto com a finalidade de guiar o projeto e avaliação de algoritmos de classificação dinâmica. A partir do *framework* proposto, os autores derivaram um algoritmo de classificação e garantiram, por meio de provas teóricas, que

rankings dinâmicos podem fornecer ganhos substanciais no desempenho de recuperação. O algoritmo também foi avaliado empiricamente utilizando coleções de teste, apresentando uma precisão superior a 40% em relação a recuperação convencional estática.

Por fim, em um trabalho mais recente, Kim *et al.* (2013) utilizaram a ferramenta Surf Canyon¹ para estudar como os usuários interagem com resultados de busca que são modificados dinamicamente utilizando o seu *feedback* implícito. O Surf Canyon pode ser tratado como um agente reclassificador externo ao buscador capaz de prover uma página de resultados dinâmica também estruturada em dois níveis. Os documentos selecionados pelo usuário no primeiro nível são utilizados para aprimorar um modelo do usuário representado por um MEV construído a partir dos documentos recuperados pela busca. A cada documento selecionado, o modelo do usuário é atualizado e os documentos reclassificados, sendo que os documentos que assumiram posições mais elevadas no novo *ranking* são apresentados no segundo nível da estrutura de resultados (CRAMER, 2010). Ao analisar o comportamento do usuário em um conjunto de experimentos controlados, os autores verificaram que o uso de conteúdo dinâmico leva a um maior envolvimento dos usuários no processo de busca. Para garantir o controle completo das tarefas de busca, também foi realizado um estudo com 91 participantes e dois tipos de tarefas de busca, no qual foi verificado que o recurso de classificação dinâmica melhora o desempenho do usuário em tarefas de Pesquisa Exploratória de forma significativa tanto em relação ao tempo de conclusão da tarefa quanto ao número de resultados visitados.

3.2 A abordagem DESiRe

Embora os trabalhos apresentados proponham abordagens distintas e eficazes tanto para inferir o interesse do usuário quanto para aplicar o modelo aprendido e apresentar os resultados reclassificados, eles se restringem apenas à reorganização da própria página de resultados para apresentar os esforços da personalização. Apoiado nos conceitos e técnicas de modelagem desenvolvidos pelos trabalhos anteriores, nosso trabalho procura explorar uma maneira mais dinâmica e interativa para a apresentação dos resultados de busca. Dessa forma, propomos a DESiRe, uma abordagem dinâmica de recomendação de resultados de busca, baseada nos interesses do usuário durante uma tarefa de Pesquisa Exploratória na *Web*.

No contexto do nosso trabalho, o emprego do termo “dinâmica” se refere não apenas à capacidade da abordagem em capturar o interesse do usuário durante a busca, mas também à capacidade de recomendar resultados enquanto o usuário navega pelos documentos recuperados. Diferentemente dos trabalhos relacionados que apresentamos, nossa proposta não se restringe à página de resultados para oferecer as recomendações, ampliando as oportunidades para que a personalização seja aproveitada.

¹ Disponível em: <<http://www.surfcanyon.com/>>. Acessado em: 01/06/2015.

A recomendação de resultados durante a navegação tem como papel antecipar a descoberta de documentos inéditos e de interesse que foram recuperados pela busca, mas que corriam o risco de não serem visualizados pelo usuário. Quando utilizamos o termo “inéditos” estamos nos referindo aos documentos que não foram classificados na primeira página de resultados, e portanto, não podem ser acessados diretamente após a busca. Essa estratégia de antecipação é apropriada para a situação atual de busca na *Web*, na qual os buscadores convencionais não são capazes de classificar adequadamente os documentos recuperados, mesmo quando consultas bem definidas são empregadas na busca (vide avaliação realizada na [Seção 4.4](#)).

Com a finalidade de simplificar o entendimento da proposta, a DESiRe pode ser vista como um agente recomendador, cujo fluxo de execução consiste em construir dinamicamente um Modelo do Interesse de Busca (MIB) conforme o usuário navega pelos documentos recuperados e informa o seu interesse por eles. A cada *feedback* informado explicitamente ou implicitamente pelo usuário, o modelo é aprimorado e os documentos são reclassificados em relação ao modelo atualizado. A reclassificação produzida é então utilizada para derivar um conjunto de resultados que serão recomendados ao usuário durante a navegação. Uma representação do comportamento desse agente é ilustrada pela [Figura 6](#).

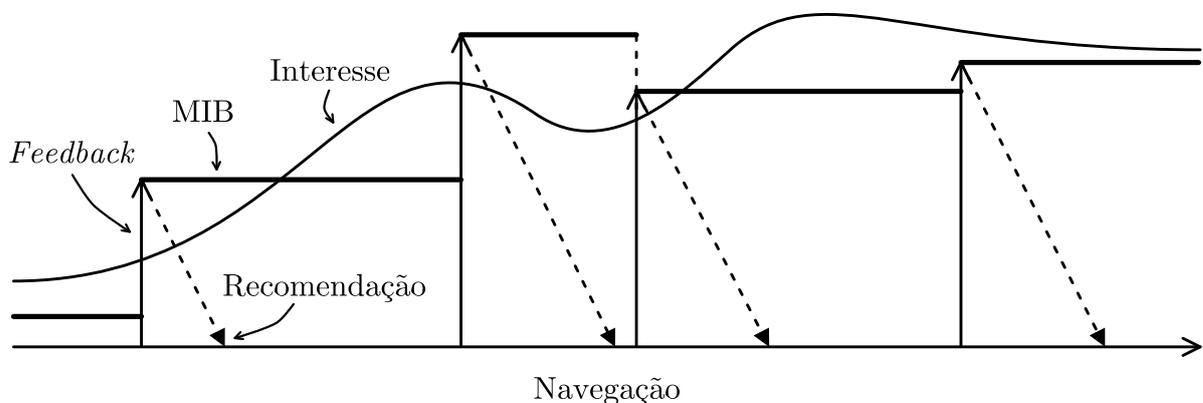


Figura 6 – Representação comportamental da DESiRe.

Como apresentado pela [Figura 6](#), as recomendações produzidas pela DESiRe (setas pontilhadas) são entregues ao usuário ao longo de sua navegação pelos documentos recuperados pela busca. Sempre que um *feedback* (setas verticais) é capturado durante a navegação a partir da interação do usuário com os documentos, o MIB (linhas horizontais) é reajustado para se aproximar do interesse do usuário (curva), permitindo que documentos mais relevantes sejam fornecidos novamente a ele.

Como podemos observar na [Figura 6](#), o interesse de busca pode sofrer oscilações conforme o usuário explora e aprende com os documentos navegados. Essas oscilações devem ser acompanhadas pelo modelo a cada *feedback* informado, procurando dessa forma se aproximar do interesse latente no usuário. É importante observar que a aproximação

do MIB não depende apenas do processo de aprimoramento do modelo, mas também da qualidade da recomendação fornecida. Se os documentos recomendados forem relevantes para o usuário, eles provavelmente serão considerados para os próximos *feedbacks* e a aproximação do MIB será mais rápida. Por outro lado, se a recomendação não for eficaz, o usuário poderá ter dificuldades em encontrar documentos do seu interesse, refletindo uma aproximação mais lenta por parte do modelo.

Outra vantagem de tratarmos a abordagem como um agente recomendador é permitir que ela também seja entendida como uma ferramenta externa aos mecanismos de busca, possibilitando que a captura do *feedback* e a apresentação dos resultados recomendados seja realizada por meio de extensões criadas para navegadores *Web*. Conforme é exemplificado na Figura 7, a extensão para o navegador é responsável por capturar o *feedback* do usuário e apresentar a ele os resultados recomendados. Por se tratar de um recurso externo à página de resultados e ao próprio buscador, uma nova recomendação pode ser apresentada sempre que um *feedback* for informado. Embora a DESiRe seja baseada no *feedback* do usuário para a reformulação do modelo, a maneira como essa informação pode ser capturada não foi tratada neste trabalho, deixando essa questão ser respondida pelo desenvolvedor da ferramenta.

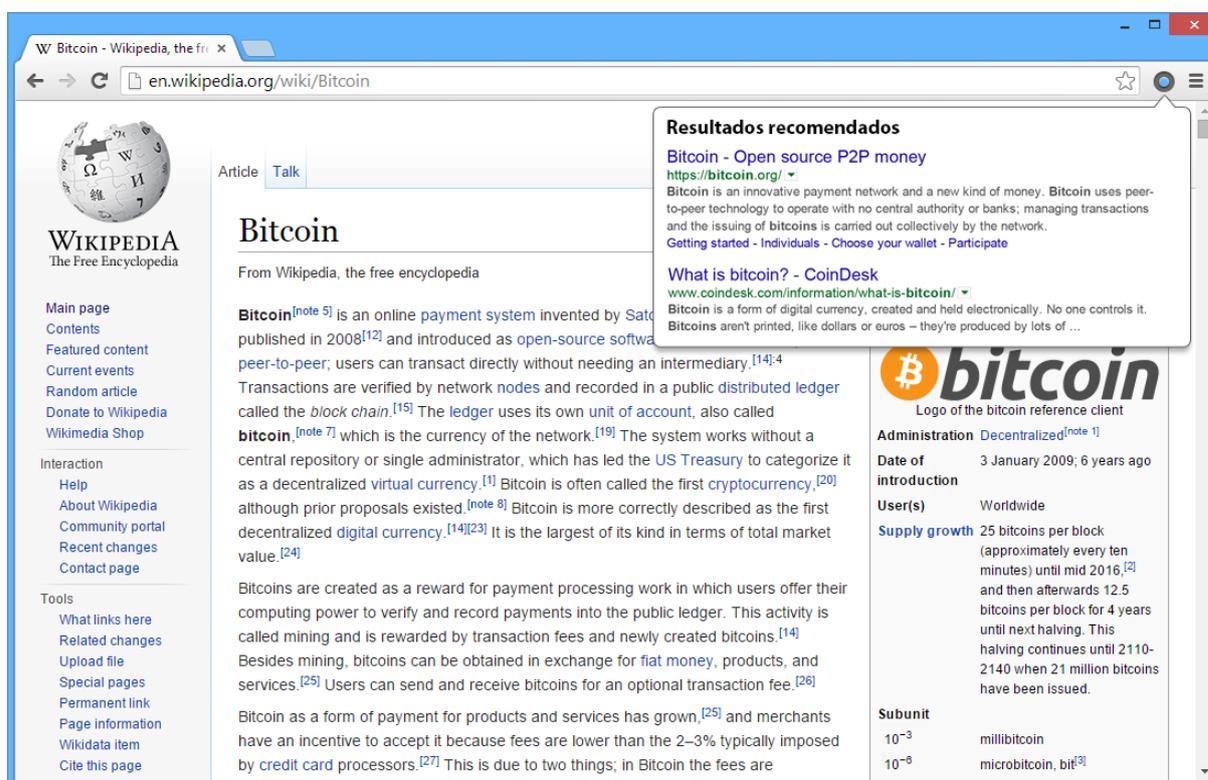


Figura 7 – Exemplificação da DESiRe como uma extensão para um navegador *Web*.

Com a nossa proposta, desejamos verificar se as reclassificações proporcionadas pela DESiRe possuem qualidade superior em relação à classificação gerada por buscadores *Web* convencionais, e que o processo de recomendação é igualmente eficaz, fornecendo

documentos relevantes e inéditos ao usuário. Uma maneira direta de identificar tais evidências é avaliando empiricamente a eficácia da DESiRe frente a uma simulação simplificada do comportamento de busca do usuário. O [Capítulo 5](#) descreve essa simulação e apresenta a avaliação que realizamos para a abordagem.

3.3 O Processo de recomendação

O processo de recomendação que propomos para a DESiRe consiste de uma sequência de etapas bem definidas que devem ser adicionadas ao processo padrão de busca desempenhado pelo usuário. Para que as etapas propostas sejam efetivamente aplicadas ao processo padrão, o fluxo de interações do mesmo deve ser adaptado conforme o diagrama ilustrado pela [Figura 8](#).

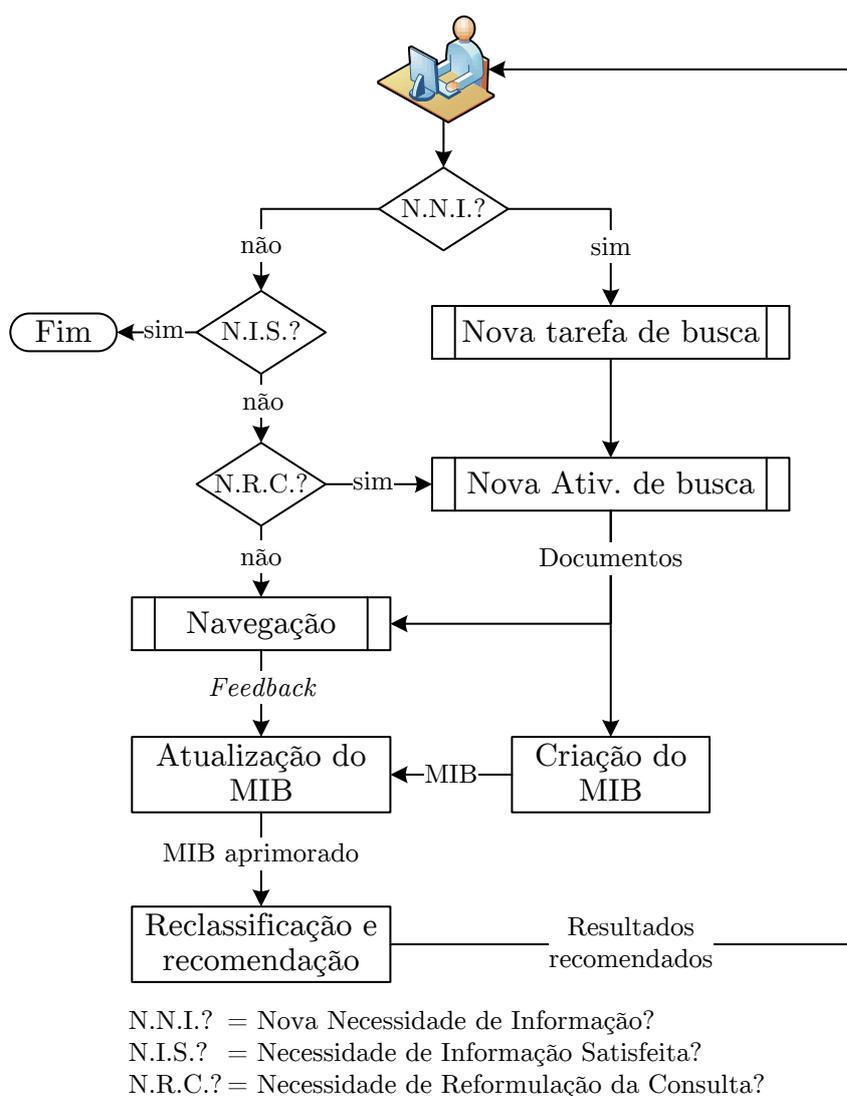


Figura 8 – Adaptação do processo de busca às etapas do processo de recomendação previsto para o processo de recomendação.

Como podemos observar, o processo de recomendação é inicializado sempre que o usuário inicia uma tarefa de busca na *Web*. Para cada atividade de busca realizada, uma etapa preliminar de criação do **MIB** é executada uma única vez antes do primeiro *feedback* ser informado. Em seguida, as etapas de atualização do **MIB**, reclassificação e recomendação dos resultados são executadas para cada *feedback* informado até que uma nova atividade de busca seja realizada, reiniciando o processo.

Conforme comentado na [Seção 2.1](#), definiremos uma tarefa de busca como um conjunto de atividades de busca informacionais realizadas para satisfazer uma necessidade de informação específica. A tarefa inicia sempre que o usuário admite uma nova necessidade de informação e é finalizada quando a mesma é satisfeita. A [Figura 9](#) ilustra a composição de uma tarefa de busca junto à evolução do interesse do usuário durante um período de tempo. A tarefa de busca somente é finalizada quando a última atividade de busca for encerrada e a necessidade de informação do usuário for satisfeita.

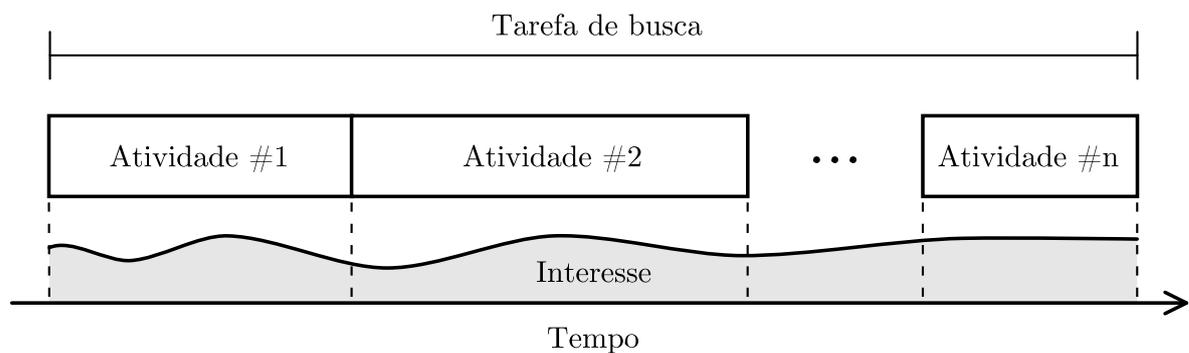


Figura 9 – Estrutura de uma tarefa de busca.

Para que o processo se estenda para as atividades de busca seguintes, algum método capaz de identificar mudanças de contexto deve ser aplicado para verificar se o **MIB** estabelecido até então é condizente com a nova intenção de busca do usuário. Embora tal método não seja tratado pelo nosso trabalho, a avaliação realizada na [Subseção 5.2.4](#) verifica a viabilidade de propagar o modelo para atividades de busca seguintes ao assumir que o contexto da busca é mantido.

As seções a seguir descrevem em maiores detalhes as etapas que compõem o processo de recomendação proposto, envolvendo a criação, evolução e aplicação do **MIB** proposto.

3.3.1 Criação do **MIB**

Um aspecto fundamental da nossa abordagem reside na maneira com que o interesse do usuário é representado e mantido durante a tarefa de busca. Essa representação é de responsabilidade do **MIB**, o qual associa as informações textuais contidas na consulta realizada e nos documentos indicados pelo usuário. Tal associação é construída com a

utilização do **MEV** (SALTON; WONG; YANG, 1975), um modelo algébrico capaz de representar os documentos recuperados e o **MIB** como vetores de características que podem ser facilmente comparados em relação à sua semelhança. Embora o **MEV** seja um modelo proposto há mais de quatro décadas, ele se baseia em uma representação algébrica simplificada quando comparado com modelos mais sofisticados como o **LSI**, permitindo que suas operações sejam realizadas com custo computacional reduzido. Essa característica é essencial quando estamos trabalhando no contexto de buscas na *Web*, o qual exige um tempo de resposta curto para as interações com o usuário.

No espaço vetorial estabelecido pelo **MEV**, as dimensões de cada vetor correspondem aos termos extraídos do conjunto de documentos considerados na modelagem. Para a nossa abordagem, o modelo é criado sobre um subconjunto dos documentos melhor classificados pela busca, uma vez que o conjunto dos documentos recuperados pode ser excessivamente grande para ser representado completamente. A **Figura 10** exemplifica o espaço vetorial definido por três termos t_i extraídos a partir de dois documentos d_j fictícios. O **MIB** e os documentos são representados como vetores pertencentes ao espaço criado.

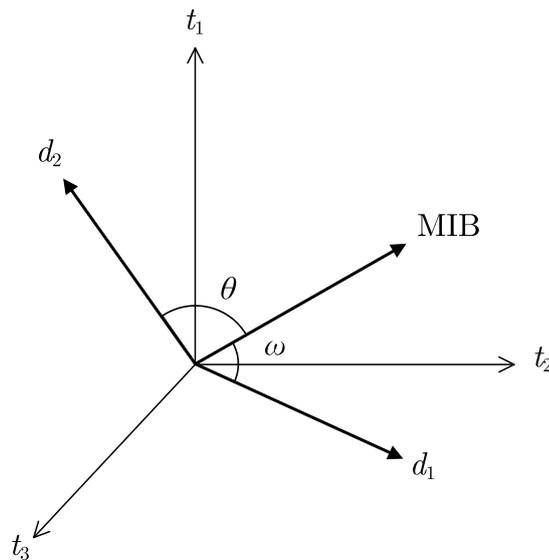


Figura 10 – Representação do espaço vetorial definido pelos termos extraídos de dois documentos fictícios.

Como podemos observar na **Figura 10**, os valores atribuídos às dimensões de cada vetor dependem do peso que cada termo exerce sobre o documento representado. Em nossa proposta, o peso atribuído às dimensões dos vetores é calculado pela estatística *Term Frequency–Inverse Document Frequency* (**TF-IDF**), uma heurística que procura determinar a importância de um termo para um documento em função da sua frequência (tf) e distinção em relação ao conjunto dos documentos (idf). Dessa forma, para que um termo seja considerado importante a um determinado documento, ele deve ocorrer com frequência nesse documento e não deve ser comum no conjunto de documentos. O peso atribuído a um termo t em d , $d \in \mathcal{D}$, em que \mathcal{D} é o conjunto de documentos representados no espaço

vetorial, pode ser calculado pela seguinte equação:

$$\text{TF-IDF}_{t,d} = tf(t, d) \times idf(t, \mathcal{D}), \quad (3.1)$$

na qual $tf(t, d)$ é a frequência do termo t em d e $idf(t, \mathcal{D}) = \log_2 \frac{|\mathcal{D}|}{|\{d \in \mathcal{D} \mid t \in d\}|}$.

Para que as técnicas descritas anteriormente possam ser aplicadas na criação do MIB, o conteúdo textual dos documentos deve ser previamente processado a fim de se extrair um conjunto de termos representantes que será utilizado na construção do MEV. A determinação desse conjunto é realizada com a aplicação de técnicas clássicas de pré-processamento linguístico, como a normalização textual, segmentação de sentenças e palavras, remoção de *stopwords*, e a *stemização* (BAEZA-YATES; RIBEIRO-NETO, 2011). Vale destacar que durante a construção de ferramentas que implementem a nossa abordagem, é importante estabelecer um projeto no qual o processamento textual não seja realizado pelo dispositivo do usuário, mas sim por um serviço que proverá a ferramenta. Encarregar o cliente de descarregar e processar os documentos recuperados pela busca, mesmo que apenas os melhor classificados, pode demandar tempo suficiente para tornar a abordagem inviável. Assim, o conjunto de termos processados deve ser previamente transmitido ao dispositivo do usuário para que o MIB seja construído localmente. A Seção 5.2 avalia a viabilidade dessa estratégia ao considerar o tempo de transmissão dos termos processados.

3.3.2 Atualização do MIB

No processo de recomendação dos resultados, a etapa de atualização do MIB ocorre sempre que o usuário informa um documento relevante dentro do espaço de busca recuperado. Essa etapa é responsável por compor ao modelo atual à informação adquirida com o *feedback*, fornecendo um modelo evoluído que represente com maior precisão os interesses do usuário. A composição consiste de uma adaptação do Algoritmo de Rocchio para *Feedback* de Relevância (ROCCHIO, 1971) e é realizada por meio da combinação linear entre o MIB e o vetor representante do documento informado pelo usuário.

Supondo que os documentos relevantes estejam correlacionados de alguma maneira, o objetivo da função de atualização é “moldar” o MIB de forma que ele se aproxime da representação vetorial desses documentos, se tornando mais semelhante a eles. Quanto maior for a semelhança entre o modelo e os documentos relevantes, maior é a probabilidade desses documentos ocuparem posições mais elevadas na reordenação do *ranking*. A Figura 11 ilustra esse conceito, destacando a posição do modelo e sua vizinhança antes e após a atualização.

A Equação 3.2 estabelece como uma versão evoluída do modelo (MIB_{n+1}) pode ser calculada em função do modelo atual (MIB_n) e do documento d associado ao *feedback*

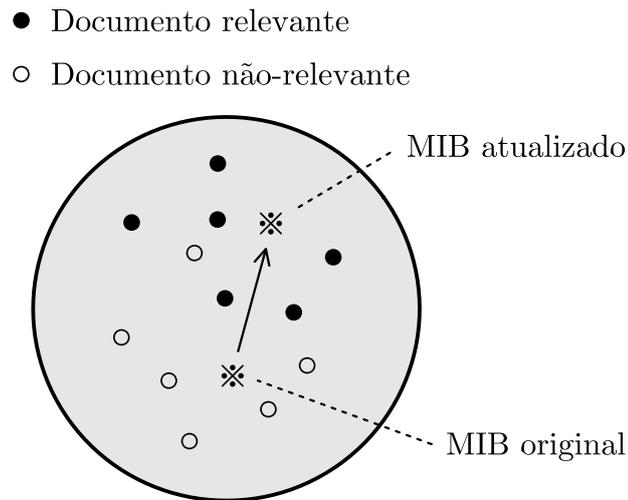


Figura 11 – Representação visual do processo de aproximação do MIB aos documentos relevantes.

informado. Na equação, α e β são parâmetros que definem o comportamento da atualização do modelo e estão diretamente relacionados à confiança sobre o *feedback* informado.

$$\overrightarrow{\text{MIB}}_{n+1} = \alpha \cdot \overrightarrow{\text{MIB}}_n + \beta \cdot \vec{d} \quad (3.2)$$

De maneira geral, queremos adotar configurações em que a informação adquirida seja privilegiada em *feedbacks* confiáveis e amenizada para *feedbacks* suspeitos. Essas configurações podem ser alcançadas ajustando os valores assumidos individualmente pelos parâmetros α e β . Valores superiores a um, por exemplo, indicam que a carga informacional (pesos das dimensões do vetor) associada ao parâmetro será potencializada durante a atualização. Já valores inferiores a um indicam que a carga informacional não será considerada integralmente para a evolução do modelo.

Dessa forma, tais observações permitem estabelecer uma correlação entre os parâmetros apresentados. Se considerarmos uma configuração em que $\alpha < \beta$, a informação adquirida pelo *feedback* será privilegiada em relação ao que já foi consolidado pelo modelo. Essa configuração é desejada quando a confiança sobre o *feedback* é alta e, portanto, queremos que o modelo atualizado adquira a informação associada de maneira potencializada. Por outro lado, se considerarmos uma configuração em que $\alpha > \beta$, a informação estabelecida pelo modelo atual será resguardada em relação à associada ao *feedback*. Essa configuração é desejada quando a confiança sobre o *feedback* não é alta e, portanto, não queremos que a informação associada tenha grande influência sobre o modelo.

A confiança sobre um determinado *feedback* pode estar associada à maneira como ele é informado. Por simplificação, podemos considerar que os *feedbacks* informados explicitamente pelo usuário são sempre confiáveis, embora tal garantia dependa da segurança do usuário em relação ao espaço de busca explorado. Por outro lado, quando o *feedback* é

informado de maneira implícita, a confiança sobre o mesmo é determinada pelo método de captura utilizado. Nesse caso, é necessário determinar uma função capaz de traduzir o grau de confiança atribuído pela técnica de *feedback* para os parâmetros de atualização do modelo. A avaliação realizada na [Subseção 5.2.5](#) demonstra a importância de considerar a confiança sobre o *feedback* ao comparar a qualidade das recomendações geradas em situações em que ela é ignorada.

3.3.3 Reclassificação e recomendação dos resultados

A reclassificação e recomendação dos resultados é uma etapa que decorre diretamente da atualização do modelo e, portanto, é executada sempre que a etapa anterior é encerrada. Como discutido anteriormente, o objetivo da reclassificação é fornecer um *ranking* que atenda aos interesses do usuário e que possibilite derivar uma recomendação eficaz dos resultados recuperados. Dessa forma, o processo de reclassificação deve reordenar os documentos em função da sua semelhança com o modelo atual, representando assim o seu grau de proximidade com os interesses do usuário.

A representação do MIB e dos documentos recuperados em um espaço vetorial comum é uma característica fundamental para que a métrica de semelhança entre seus componentes seja estabelecida. Nesse aspecto, a Similaridade de Cosseno se apresenta como uma métrica conveniente para ser utilizada em espaços vetoriais, estabelecendo a semelhança entre dois vetores a partir de um cálculo trigonométrico simples e que independe do número de dimensões consideradas. Baseada no cálculo do cosseno do ângulo formado entre os vetores, a métrica resulta em um julgamento de orientação e não de magnitude. Quão menor for o ângulo entre os dois vetores, mais próximo de um será o cosseno desse ângulo e maior será a semelhança entre eles. Dessa forma a similaridade entre o MIB e um determinado documento pode ser computada pela seguinte equação:

$$\text{sim}(\text{MIB}, d) = \cos(\theta) = \frac{\overrightarrow{\text{MIB}} \cdot \vec{d}}{\|\overrightarrow{\text{MIB}}\| \times \|\vec{d}\|} = \frac{\sum_{i=1}^n \overrightarrow{\text{MIB}}_i \times \vec{d}_i}{\sqrt{\sum_{i=1}^n (\overrightarrow{\text{MIB}}_i)^2} \times \sqrt{\sum_{i=1}^n (\vec{d}_i)^2}}, \quad (3.3)$$

na qual θ é o ângulo formado entre os vetores que representam o MIB e o documento d , como apresentado na [Figura 10](#).

Após a reordenação, o *ranking* deve ser segmentado em um ponto a fim de determinar os documentos que serão recomendados. Todos os documentos pertencentes ao primeiro segmento e que ainda não foram associados a algum *feedback* são sequencialmente selecionados para compor a recomendação. A [Subseção 5.2.2](#) avalia duas estratégias para a definição do ponto de segmentação (ou ponto de recomendação), sendo uma pré-fixada para todas as recomendações e outra autônoma. A estratégia autônoma procura determinar um ponto de recomendação ideal com base na medida de semelhança atribuída aos documentos do *ranking*.

Para a avaliação, utilizamos uma versão ótima do algoritmo de clusterização *k-means* unidimensional (WANG; SONG, 2011) para agrupar os 10 primeiros documentos não selecionados do *ranking* de forma ordenada em dois *clusters*, cuja fronteira em comum define o ponto de recomendação. A construção ordenada dos dois *clusters* permite que o primeiro *cluster* represente a sequência de documentos que devem ser recomendados. Embora a clusterização utilizada seja ótima, não é possível garantir que tal *cluster* contenha apenas documentos relevantes, uma vez que esse processo depende diretamente da organização do *ranking* que será segmentado. Por outro lado, a segmentação ótima permite que o número de documentos não-relevantes recomendados seja minimizado em função da métrica de semelhança utilizada para a classificação dos mesmos.

3.4 Considerações finais

Neste capítulo apresentamos a DESiRe, uma abordagem para a recomendação dinâmica de resultados de busca na *Web*. Diferentemente dos trabalhos relacionados que também apresentamos aqui, nossa proposta procura explorar uma maneira mais dinâmica e interativa para apresentação dos resultados de busca. Em seguida, descrevemos a DESiRe em função de um processo de recomendação que estabelecemos para a abordagem. O processo foi detalhado e discutido em suas diferentes etapas, compreendendo a criação e atualização de um MIB, bem como a reclassificação e recomendação dos resultados recuperados pela busca. Com o propósito de comparar o desempenho obtido pela DESiRe em relação ao que é fornecido pelos buscadores, devemos dispor de uma coleção de teste que reflita a recuperação realizada por esses mecanismos quando expostos a atividades de Pesquisa Exploratória. Dessa forma, o capítulo a seguir apresenta e avalia em detalhes a coleção de teste que criamos para esse fim.

4 Construção da coleção de teste

Segundo Manning, Raghavan e Schütze (2008), a RI se desenvolveu como uma área altamente empírica, requerendo que uma avaliação abrangente e cuidadosa seja empregada para demonstrar o desempenho das técnicas aplicadas pelos SRIs. Sem a avaliação adequada, não conseguimos medir quão bem um SRI está satisfazendo as necessidades dos usuários, nem comparar objetivamente o desempenho entre SRIs diferentes. Baeza-Yates e Ribeiro-Neto (2011) definem que a avaliação da eficácia da recuperação é o processo de associar sistematicamente uma métrica aos resultados produzidos pelo SRI em resposta a um conjunto de consultas realizadas pelo usuário.

Uma vez que a avaliação da eficácia da recuperação é baseada na noção de relevância dos documentos recuperados, o cálculo de tal métrica geralmente requer que os resultados produzidos pelo sistema sejam comparados com os sugeridos por humanos para um mesmo conjunto de consultas. Dessa forma, uma coleção composta por um conjunto \mathcal{D} de documentos pré-selecionados, um conjunto Q de consultas sobre esses documentos e um conjunto \mathcal{R} de julgamentos de relevância atribuídos a cada par (q_m, d_j) , $q_m \in Q$ e $d_j \in \mathcal{D}$ pode ser utilizada como base para o emprego dessa métrica. Dentre as diversas coleções de teste disponíveis, podemos citar a Reuters¹, OHSUMED (HERSH *et al.*, 1994), NewsGroup (LANG, 1995), NTCIR² e CLEF³ como algumas das principais referências utilizadas na avaliação da recuperação de informação.

Atualmente, outro grupo de coleções que tem recebido grande destaque é a TREC⁴, uma conferência de recuperação de informação que compila anualmente um grande volume de documentos sobre diversos assuntos. Dentre suas diversas trilhas, a *Web Track* é a que se dedica à avaliação de tecnologias de recuperação *Web*, oferecendo coleções com centenas de consultas e milhares de documentos avaliados. A TREC *Web* construída em 2014, por exemplo, é composta por um *pooling* de 870 milhões de páginas *Web* sobre a ClueWeb12⁵ a partir de 50 tópicos de busca distintos.

Embora a TREC *Web* esteja consolidada no âmbito da avaliação de técnicas de recuperação na *Web*, optamos por desenvolver uma coleção especificamente para o nosso

¹ Disponível em: <<http://trec.nist.gov/data/reuters/reuters.html>>. Acessado em: 01/06/2015.

² Disponível em: <<http://research.nii.ac.jp/ntcir/data/data-en.html>>. Acessado em: 01/06/2015.

³ Disponível em: <<http://www.imageclef.org/datasets>>. Acessado em: 01/06/2015.

⁴ Disponível em: <<http://trec.nist.gov/>>. Acessado em: 01/06/2015.

⁵ Disponível em: <<http://boston.lti.cs.cmu.edu/clueweb12/>>. Acessado em: 01/06/2015.

trabalho. A necessidade de construir uma coleção de teste específica, em vez de utilizar coleções já existentes, se dá pela natureza particular do problema que estamos tratando. Uma vez que a proposta aborda a reclassificação dos documentos recuperados por buscas informacionais na *Web*, é necessário dispor de uma coleção que reflita diretamente esse contexto. Diferentemente das coleções convencionais, para avaliar o efeito gerado por nossa abordagem em relação ao que é oferecido pelos buscadores atuais, os documentos pertencentes à nossa coleção devem ser recuperados a partir de buscas realizadas na *Web*. Dessa forma, além de possibilitar a avaliação da abordagem em relação a sua eficácia, a coleção também permite que as reclassificações proporcionadas pela abordagem sejam comparadas ao *ranking* originado pelo buscador utilizado.

Baseada na organização descrita anteriormente, a coleção de teste que construímos consiste de oito conjuntos de documentos \mathcal{D}_t recuperados a partir de oito atividades de busca informacional $t \in \mathcal{T}$ realizadas na *Web*. Cada atividade de busca é realizada por uma consulta distinta com a finalidade de satisfazer a necessidade de informação associada a um problema de busca $p \in \mathcal{P}$ que a originou. Ao todo, seis problemas de busca distintos foram considerados. Adicionalmente, a coleção compõe um conjunto \mathcal{R} de julgamentos de relevância sobre cada par (t, d) , tal que $d \in \mathcal{D}_t$. A [Figura 12](#) ilustra a organização da coleção.

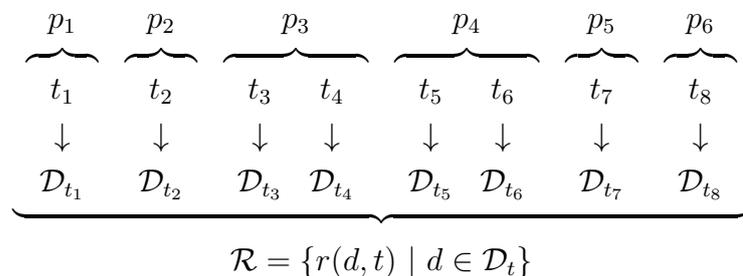


Figura 12 – Organização da coleção de teste construída.

A construção da coleção de teste seguiu um processo definido pelas três etapas descritas a seguir:

1) Definição das atividades de busca

Compreende a criação de um conjunto de problemas de busca e a especificação das suas respectivas atividades de busca.

2) Recuperação dos documentos

Compreende a emissão das consultas geradas ao buscador e a recuperação dos documentos associados aos resultados retornados por cada busca.

3) Julgamento de relevância

Compreende a atribuição de relevância aos documentos em relação à necessidade de informação que os recuperou.

As seções seguintes apresentam os detalhes da construção da coleção de teste proposta, bem como a sua avaliação em relação à distribuição de relevância dos documentos recuperados.

4.1 Definição das atividades de busca

A construção da coleção de teste foi determinada pela definição de oito atividades de busca na *Web*, cada qual associada a um problema de busca informacional. Os problemas de busca foram descritos em função de uma necessidade de informação que deve ser satisfeita por uma consulta durante a respectiva atividade de busca. Os Quadros 1 a 6 a seguir descrevem as oito atividades definidas, estabelecendo sua relação com os seis problemas de busca descritos. Embora não haja restrições quanto à língua utilizada por nossa abordagem, a língua inglesa foi empregada na especificação das consultas devido ao seu alcance global, favorecendo a utilização da coleção por outros trabalhos.

Quadro 1 – Problema de busca: “Benefícios do consumo de chá”.

p_1 : Benefícios do consumo de chá

Necessidade de Informação: Durante o intervalo de trabalho em sua empresa, você geralmente vai à cozinha e toma uma xícara de café. Certo dia, sua colega trouxe uma garrafa de chá e ofereceu para os demais colegas, alegando que, além de ser mais saudável do que o café, poderia proporcionar os mesmos benefícios. Desconfiado, você resolveu pesquisar na *Web* sobre os benefícios à saúde proporcionados pelo consumo de chá e, como forma de recolher documentos que tragam tal informação, você realizou a seguinte atividade de busca:

Atividade de busca [A1]

Consulta realizada: “*health benefits of drinking tea*”.

Características: Atividade de dificuldade moderada, envolvendo o processo de tomada de decisão a partir da investigação e análise crítica de diferentes conceitos. A consulta foi expressa de maneira adequada, empregando as palavras-chave envolvidas na necessidade de informação.

Critérios de avaliação

Relevância Alta: O documento deve trazer informações completas sobre os benefícios à saúde proporcionados pelo consumo de chá. É importante que o documento traga informações sobre os diferentes tipos de chá e seus respectivos benefícios. A existência de uma comparação entre o chá e o café, a riqueza de informações e a existência de citações a estudos científicos são fatores que contribuem diretamente para uma alta relevância.

Relevância Moderada: O documento não atende a todas as restrições de um documento de alta relevância e é superficial na explicação dos benefícios ou apenas os cita.

Relevância Baixa: O documento apresenta pouca ou nenhuma informação relacionada aos benefícios trazidos pelo consumo de chá.

Quadro 2 – Problema de busca: “Café ou chá, qual é melhor para a saúde?”.

***p*₂: Café ou chá, qual é melhor para a saúde?**

Necessidade de Informação: Na empresa em que trabalha, você foi responsável por realizar uma pesquisa para decidir qual das bebidas, café ou chá, é a melhor para a saúde. A sua pesquisa irá auxiliar na tomada de decisão de qual bebida deverá ser servida de forma contínua pela cozinha da empresa. Como forma de recolher documentos que tragam comparações entre essas duas bebidas, você realizou a seguinte atividade de busca na *Web*:

Atividade de busca [A2]

Consulta realizada: “*coffee or tea which is better for health*”

Características: Atividade de dificuldade moderada, envolvendo o processo de tomada de decisão a partir da investigação e comparação de diferentes conceitos. A consulta foi expressa de maneira adequada, ressaltando os objetos envolvidos na comparação (*coffee* e *tea*) e o objetivo da mesma (*better for health*).

Critérios de avaliação

Relevância Alta: O documento deve trazer comparações entre as duas bebidas, apresentando os prós e contras de cada uma. Não é necessário que o documento se posicione à escolha, mas se o fizer, ela deve ser justificada. A riqueza de informações e a existência de citações a estudos científicos contribuem diretamente para uma alta relevância.

Relevância Moderada: O documento não atende a todas as restrições de um documento de alta relevância. O documento é superficial nas comparações ou traz as vantagens e/ou desvantagens de apenas uma das bebidas.

Relevância Baixa: O documento apresenta pouca ou nenhuma informação relacionada aos benefícios trazidos pelo café ou pelo chá.

Quadro 3 – Problema de busca: “Tratamento de queimaduras por água fervente”.

***p*₃: Tratamento de queimaduras por água fervente**

Necessidade de Informação: Seu amigo se queimou após entrar em contato com água fervente e pediu conselhos à você sobre quais medidas ele deve tomar. Como forma ajudá-lo, você se prontificou a procurar na *Web* informações de como tratar queimaduras causadas por água fervente. Como você não sabe ao certo qual é a gravidade dos ferimentos, você deve ter preferência por documentos que tragam a maior quantidade possível de informação sobre os tipos de queimaduras, como identificá-las e qual o procedimento correto para tratá-las.

Atividade de busca [A3a]

Consulta realizada: “*how to treat burns*”

Características: Atividade de dificuldade baixa, envolvendo o processo de aquisição de conhecimento a partir da análise crítica sobre conceitos simples. A consulta foi expressa com carência de especificidade, não informando a causa da queimadura.

Atividade de busca [A3b]

Consulta realizada: “*how to treat burn from boiling water*”

Características: Atividade de dificuldade baixa, envolvendo o processo de aquisição de conhecimento a partir da análise crítica sobre conceitos simples. A consulta foi expressa de maneira adequada, especificando a causa da queimadura (*from boiling water*).

Critérios de avaliação

Relevância Alta: O documento deve trazer informações completas sobre os tipos de queimaduras causadas por água fervente e os procedimentos que devem ser adotados durante as etapas do tratamento.

Relevância Moderada: O documento não é tão completo quanto um de alta relevância, mas deve trazer ao menos os procedimentos que devem ser adotados durante as etapas do tratamento.

Relevância Baixa: O documento apresenta pouca ou nenhuma informação relacionada ao tratamento de queimaduras causadas pelo contato com água ou líquidos quentes.

Quadro 4 – Problema de busca: “Aprender sobre Bitcoins”.

***p*₄: Aprender sobre Bitcoins**

Necessidade de Informação: Você estava no bar com seus amigos e ouviu um deles comentar que começou a “minerar” Bitcoins. Ele explicou superficialmente o que era, dizendo que se tratava de uma moeda digital. Você ficou curioso sobre o assunto e resolveu procurar na *Web* explicações do que se trata tal moeda e como ela funciona. Duas atividades foram realizadas, a primeira mais abrangente e a segunda mais específica.

Atividade de busca [A4a]

Consulta realizada: “*bitcoin*”

Características: Atividade de dificuldade moderada, envolvendo o processo de aquisição de conhecimento a partir da análise crítica sobre conceitos elaborados. A consulta foi expressa com carência de informação, não indicando o propósito da busca (aprender sobre Bitcoins).

Atividade de busca [A4b]

Consulta realizada: “*what is bitcoin and how does it work*”

Características: Atividade de dificuldade moderada, envolvendo o processo de aquisição de conhecimento a partir da análise crítica sobre conceitos elaborados. A consulta foi expressa de maneira adequada, especificando o propósito da busca associado à necessidade de informação.

Critérios de avaliação

Relevância Alta: O documento deve trazer informações completas do que se trata a moeda Bitcoin, como ela funciona, como você pode obtê-la e como pode usá-la. Informações técnicas de quem a criou e a tecnologia utilizada para sua criação são complementares.

Relevância Moderada: O documento não é tão completo quanto um de alta relevância, mas deve trazer ao menos informações sobre o funcionamento da Bitcoin.

Relevância Baixa: O documento apresenta pouca ou nenhuma informação relacionada ao funcionamento da Bitcoin.

Quadro 5 – Problema de busca: “Efeitos das cores sobre o humor”.

***p*₅: Efeitos das cores sobre o humor**

Necessidade de Informação: Você decidiu pintar as paredes do seu quarto, mas não decidiu com que cor pintar. Em uma conversa com seu colega, ele comentou que existem estudos que correlacionam os estados de humor com determinadas cores e sugeriu que você pesquisasse sobre tais correlações antes de pintar o quarto. Como você não sabe ao certo quais são essas correlações e como elas ocorrem, ou se elas realmente existem, você resolveu procurar na internet alguma constatação com respaldo científico de como o humor pode ser afetado pelas cores realizando a seguinte atividade de busca:

Atividade de busca [A5]

Consulta realizada: “*how color affects mood*”

Características: Atividade de dificuldade moderada, envolvendo o processo de aquisição de conhecimento a partir da análise crítica sobre conceitos elaborados. A consulta foi expressa de maneira adequada, expressando informações-chave da necessidade de informação.

Critérios de avaliação

Relevância Alta: O documento deve trazer informações completas sobre como as cores afetam nosso humor e como elas estão relacionadas aos estados de humor. A existência de citações a estudos científicos que fortalecem a veracidade (ou não) dessas relações são fatores que contribuem para uma alta relevância.

Relevância Moderada: O documento não é tão completo quanto um de alta relevância, mas deve trazer ao menos informações sobre como as cores se relacionam aos estados de humor ou quais são essas relações.

Relevância Baixa: O documento apresenta pouca ou nenhuma informação relacionada à influência das cores sobre o humor.

Quadro 6 – Problema de busca: “Localização da ilha de Java”.

***p*₆: Localização da ilha de Java**

Necessidade de Informação: Para um trabalho na faculdade você precisa saber sobre a geografia e localização da ilha de Java. Para tanto, você executou a seguinte atividade de busca na *Web*:

Atividade de busca [A6]

Consulta realizada: “*where is java located*”

Características: Atividade de dificuldade baixa, envolvendo o processo de aquisição de conhecimento a partir da análise crítica sobre conceitos simples. A consulta foi expressa de maneira adequada, porém ambígua. O termo “Java” é polissêmico e a característica “*located*” não é capaz de desambiguá-lo: Java:[ilha] ↔ *located*: [localização geográfica], Java:[linguagem de programação] ↔ *located*: [local de instalação].

Critérios de avaliação

Relevância Alta: O documento deve trazer informações detalhadas sobre a localização e geografia da ilha de Java.

Relevância Moderada: O documento traz apenas informações superficiais sobre a localização da ilha de Java.

Relevância Baixa: O documento apresenta pouca ou nenhuma informação relacionada à localização da ilha de Java.

Como pode ser observado nos quadros anteriores, cada atividade de busca possui características particulares que descrevem o processo cognitivo envolvido na busca e a maneira com que a consulta foi expressa. A dificuldade de cada atividade foi estabelecida a partir da complexidade do processo e dos conceitos envolvidos na busca. Além disso, um conjunto de critérios de avaliação foi associado a cada problema de busca com a finalidade de auxiliar o processo de julgamento de relevância dos documentos recuperados, o qual será discutido em detalhes mais adiante.

Ainda em relação aos quadros anteriores, especificamente os que descrevem os problemas p_3 e p_4 , podemos observar a definição de atividades de busca congêneres, isto é, atividades distintas, mas que foram originadas a partir de um mesmo problema de busca. Para esses casos, a necessidade de informação associada é expressa por uma consulta inicial que, posteriormente, é redefinida por uma nova consulta mais precisa em relação à necessidade de informação envolvida. O objetivo de criar atividades de busca congêneres é possibilitar a avaliação da abordagem em relação à continuidade das buscas. Como apresentado no [Capítulo 3](#), a proposta prevê que a informação adquirida durante uma determinada atividade de busca possa ser reutilizada em atividades posteriores que preservem o mesmo contexto.

4.2 Recuperação dos documentos

Uma vez definidas as atividades de busca, cada consulta associada foi submetida ao buscador Google e o conteúdo textual dos 100 primeiros documentos recuperados foi armazenado. A decisão pela utilização da ferramenta de busca da Google foi tomada baseada em sua popularidade e reconhecida qualidade de busca. Entretanto, outros mecanismos de busca como Bing, Yahoo Search e DuckDuckGo também poderiam ter sido utilizados para compor a coleção construída.

A fim de minimizar as possíveis influências de personalização geradas pelo buscador, uma versão adaptada do Google (Google Custom Search⁶) foi utilizada. Dessa maneira foi possível especificar um buscador que não considerasse elementos externos à própria definição das consultas, reduzindo o possível viés provocado por um perfil de usuário pré-definido. Além disso, o buscador foi configurado para proporcionar apenas *rankings* orgânicos, isto é, isentos de propagandas ou resultados patrocinados.

A tarefa de recuperação dos documentos foi realizada no dia 03/06/2014 e, portanto, a classificação produzida pelo buscador e o conteúdo dos documentos recuperados são referentes a essa data. Por se tratarem de documentos hospedados na *Web*, não houve garantia de que todos estivessem acessíveis no momento em que seriam recuperados. Para que o número de documentos ausentes fosse reduzido, os endereços inacessíveis foram monitorados durante uma semana. Ao todo, cinco documentos não puderam ser recuperados

⁶ Disponível em: <<https://developers.google.com/custom-search/>>. Acessado em: 01/06/2015.

nesse período, o que resultou em um montante de 795 textos válidos. Entretanto, para simplificar as estatísticas que serão aplicadas adiante, trataremos a coleção como o conjunto completo dos 800 documentos, equalizando o número de documentos recuperados por página.

4.3 Julgamento de relevância

A etapa de Julgamento de Relevância consistiu em avaliar a relevância de cada documento recuperado $d \in \mathcal{D}_t$ em relação à necessidade de informação associada à atividade t que o recuperou. Para tanto, uma tarefa de julgamento foi criada para cada atividade de busca, contendo um conjunto de critérios que permitiam atribuir um grau de relevância ternário (*baixa*, *moderada* e *alta*) a cada documento. Os critérios de avaliação estabelecidos para cada tarefa de julgamento foram descritos individualmente nos quadros dos respectivos problemas de busca.

Para que o processo de avaliação dos documentos se torne confiável, é necessário que eles sejam julgados por mais de um juiz. Dessa forma, para cada tarefa de julgamento criada, um grupo distinto de três juízes foi selecionado para executá-la. Ao todo, 10 juízes foram convidados para a realização voluntária das tarefas. Eles se caracterizavam como usuários habituados com atividades de Pesquisa Exploratória na *Web* mas sem conhecimento prévio sobre os problemas de busca definidos.

As tarefas de julgamento foram conduzidas em uma ferramenta que desenvolvemos especialmente para esse fim. Nela, o conjunto de documentos a ser avaliado era apresentado aos juízes junto da escala de relevância que deveria ser selecionada para cada documento. Durante a tarefa, os juízes tinham a liberdade de navegar pelo conjunto de documentos fornecido, podendo saltar e rever os documentos já avaliados. Todos os juízes realizaram as tarefas de julgamento individualmente e em um prazo pré-estabelecido de duas semanas.

Embora apenas o conteúdo textual dos documentos tenha sido considerado na composição da coleção, os documentos foram avaliados em seu contexto original (como páginas *Web*), contendo possivelmente gráficos e outras mídias que também influenciaram na decisão de relevância. A finalidade de conduzir as tarefas de julgamento dessa maneira é fazer com que a coleção reflita as noções de relevância encontradas em atividades de busca reais.

A fim de aprimorar a definição dos problemas de busca e dos critérios de avaliação, os juízes realizaram uma tarefa piloto para que pudessem apontar qualquer defeito ou ambiguidade na formulação das tarefas. Além das sugestões propostas, foi verificado que os juízes aprimoravam o entendimento dos critérios estabelecidos conforme aprendiam com o próprio conteúdo do documento avaliado. Como forma de amenizar essa influência, os juízes receberam sequências randomizadas dos documentos recuperados e foram instruídos a realizar revisões no julgamento conforme houvesse necessidade.

4.3.1 Interpretações de relevância

Além de possibilitar que métricas mais aprimoradas sejam empregadas na avaliação da recuperação, o uso do julgamento ternário permite também que outras duas interpretações de relevância sejam derivadas a partir da conversão de sua graduação. Tal conversão consiste em resumir o julgamento ternário em uma classificação binária [**relevante**; **não-relevante**], permitindo que a relevância atribuída aos documentos seja interpretada de acordo com a necessidade da avaliação a ser realizada.

A primeira interpretação derivada, a de *relevância binária relaxada* (r_{rlx}^2), consiste em converter os valores do julgamento ternário r^3 sobre um documento d conforme a [Equação 4.1](#). Ela permite que documentos de relevância **moderada** e **alta** sejam considerados **relevantes**, o que a torna adequada para a avaliação de concordância dos julgamentos, já que promove um aumento na chance de correspondência entre os juízes. ([Subseção 4.3.2](#)).

$$r_{rlx}^2(d) = \begin{cases} \text{relevante,} & \text{se } r^3(d) = \text{alta} \vee \text{moderada} \\ \text{não-relevante,} & \text{se } r^3(d) = \text{baixa} \end{cases} \quad (4.1)$$

Já a segunda interpretação derivada, a de *relevância binária restrita* (r_{rst}^2), consiste na conversão dos valores de r^3 conforme a [Equação 4.2](#). Ela implica que apenas documentos de alta relevância sejam considerados **relevantes**. Isso é útil quando se quer realizar uma avaliação que exija a identificação dos documentos mais relevantes da coleção, como é feito na análise de distribuição de relevância da coleção ([Seção 4.4](#)) e na simulação da abordagem proposta ([Capítulo 5](#)).

$$r_{rst}^2(d) = \begin{cases} \text{relevante,} & \text{se } r^3(d) = \text{alta} \\ \text{não-relevante,} & \text{se } r^3(d) = \text{moderada} \vee \text{baixa} \end{cases} \quad (4.2)$$

4.3.2 Avaliação da concordância entre juízes

A fim de promover uma interpretação confiável da relevância dos documentos, devemos garantir que o processo de atribuição de relevância foi consistente, o que pode ser verificado por meio do cálculo da concordância entre os juízes. A [Tabela 1](#) relaciona, com nível de significância $\alpha = 0,05$, os valores de κ obtidos para os julgamentos considerando diferentes combinações entre os três juízes e as interpretações de relevância. Os valores apresentados para os pares de juízes se referem à métrica tradicional do κ de Cohen ([COHEN, 1960](#)). Quando considerados os três juízes simultaneamente, os valores se referem ao κ de Fleiss ([FLEISS, 1971](#)), uma versão generalizada da métrica para múltiplos avaliadores. Os valores em destaque representam o κ de maior valor em cada tarefa.

Como podemos observar na [Tabela 1](#), ao nos atentarmos apenas aos valores destacados, as tarefas de julgamento que apresentaram os menores valores de κ se referem às atividades que derivaram dos problemas de busca p_1 e p_3 . Em contrapartida, os maiores

Tabela 1 – Grau de concordância κ para as tarefas de julgamento de relevância.

	Juízes	T1	T2	T3a	T3b	T4a	T4b	T5	T6
r^3	A, B, C	0,144	0,197	0,195	0,151	0,919	0,872	0,490	0,888
	A, B	0,045	0,015	0,367	0,245	0,909	0,856	0,399	0,887
	A, C	0,187	0,120	0,144	0,094	0,939	0,936	0,645	0,946
	B, C	0,209	0,454	0,095	0,122	0,907	0,823	0,425	0,830
r_{rlx}^2	A, B, C	0,219	0,347	0,286	0,240	0,957	0,960	0,480	0,962
	A, B	0,113	0,160	0,413	0,367	0,968	0,960	0,452	0,942
	A, C	0,226	0,260	0,244	0,192	0,968	0,980	0,585	1,000
	B, C	0,330	0,659	0,201	0,162	0,935	0,940	0,419	0,942
r_{rst}^2	A, B, C	0,208	0,115	0,100	0,094	0,944	0,839	0,572	0,860
	A, B	0,146	-0,115	0,412	0,174	0,917	0,811	0,422	0,884
	A, C	0,133	-0,051	0,010	0,061	0,957	0,919	0,736	0,904
	B, C	0,318	0,471	0,016	0,069	0,957	0,787	0,532	0,790

valores de κ foram observados para as atividades advindas dos problemas p_4 e p_6 . Tal resultado pode ser associado à dificuldade encontrada pelos juízes durante a realização dessas tarefas. Um fator que influencia na dificuldade do julgamento se refere à própria dificuldade das atividades de busca, já que atividades com maior dificuldade de realização podem demandar mais capacidade do juiz durante o julgamento.

Entretanto, a dificuldade das tarefas não atua em primeiro plano, uma vez que obtivemos κ 's altos e baixos tanto para atividades de dificuldade baixa quanto para atividades de dificuldade moderada. Dessa forma, existem outros fatores além da dificuldade das atividades que contribuíram com a dificuldade das tarefas de julgamento e que não puderam ser tratados. Uma hipótese sobre esse comportamento reside na variabilidade dos documentos recuperados. Pudemos observar, de maneira subjetiva, que as tarefas que produziram κ 's baixos estão associadas às atividades que obtiveram uma baixa distinção de assunto entre os documentos recuperados, o que possivelmente dificultou o julgamento da relevância. Por outro lado, as atividades que recuperam documentos com maior distinção estão associadas às tarefas que produziram os maiores valores de κ .

Outra observação que cabe à Tabela 1 é a de que a interpretação que apresentou os valores mais elevados de κ foi a r_{rlx}^2 , permitindo que graus significativos de relevância (moderada e alta) pudessem ser melhor correspondidos. Além disso, os maiores graus de concordância foram obtidos quando considerados apenas os julgamentos dos pares de juízes. Tal comportamento indica que, mesmo com tarefas bem descritas e critérios de avaliação bem definidos, a percepção de relevância é distinta entre os usuários, principalmente quando a tarefa de busca não é trivial. Por outro lado, os melhores pares de juízes se mantiveram inalterados para as atividades congêneres, indicando que não houve disparidade no entendimento do contexto da busca quando uma nova atividade era realizada.

4.3.3 Determinação da relevância final dos documentos

O emprego de três juízes nas tarefas de julgamento é essencial para estabelecer uma classificação confiável da relevância dos documentos. Entretanto, esse método resulta na atribuição de três avaliações a cada documento, exigindo que sua relevância final seja determinada em função delas. Dessa forma, assumiremos como relevância final de um documento a relevância atribuída pelo par de juízes que melhor concorda na tarefa de julgamento envolvida. Para os julgamentos em que tal par não concorda, a relevância final é definida pela mediana entre as relevâncias atribuídas pelos três juízes. Dessa forma, para cada documento $d \in \mathcal{D}_t$, sua relevância final pode ser expressa por:

$$r_f(d) = \begin{cases} r_d^x, & \text{se } r_d^x = r_d^y \\ \widetilde{R}_d, & \text{se } r_d^x \neq r_d^y \end{cases} \quad (4.3)$$

onde r_d^j é a relevância atribuída ao documento d pelo juiz j , \widetilde{R}_d é a mediana das relevâncias atribuídas pelos três juízes ao documento d , e (x, y) é o par de juízes que melhor concorda na tarefa de julgamento t .

A partir da composição entre o cálculo da relevância final e das interpretações de relevância apresentadas, funções específicas podem ser criadas para a identificação dos documentos relevantes. Por exemplo, a composição $r_f \circ r_{rst}^2$ permite calcular a relevância final a partir da interpretação binária restrita sobre as relevâncias atribuídas a um determinado documento. O objetivo dessas composições é possibilitar a avaliação da distribuição de relevância da coleção e, posteriormente, a avaliação da proposta.

4.4 Avaliação da distribuição de relevância

Conforme discutido no início deste capítulo, a coleção de teste criada permite não apenas a avaliação da nossa abordagem, mas também avaliar a qualidade da recuperação oferecida pelo buscador utilizado, já que as características associadas à coleção são um reflexo direto do desempenho do buscador. Dessa forma, uma maneira de avaliar a qualidade de sua classificação é analisando a distribuição dos documentos relevantes recuperados pelas atividades de busca.

Ao assumir o cálculo da relevância final, podemos avaliar a distribuição de relevância a partir de duas perspectivas. A primeira delas, relacionada ao julgamento ternário, permite verificar como os graus de relevância se distribuem sobre os documentos recuperados. A [Figura 13](#) ilustra a frequência de documentos julgados com os graus de relevância **baixa**, **moderada** e **alta** em cada tarefa de julgamento.

Ao analisar o gráfico ilustrado pela [Figura 13](#), é possível verificar que as atividades A4a e A6 apresentaram uma distribuição de relevância atípica, resultando em uma grande disparidade entre a frequência de documentos de baixa e alta relevância. Tal comportamento

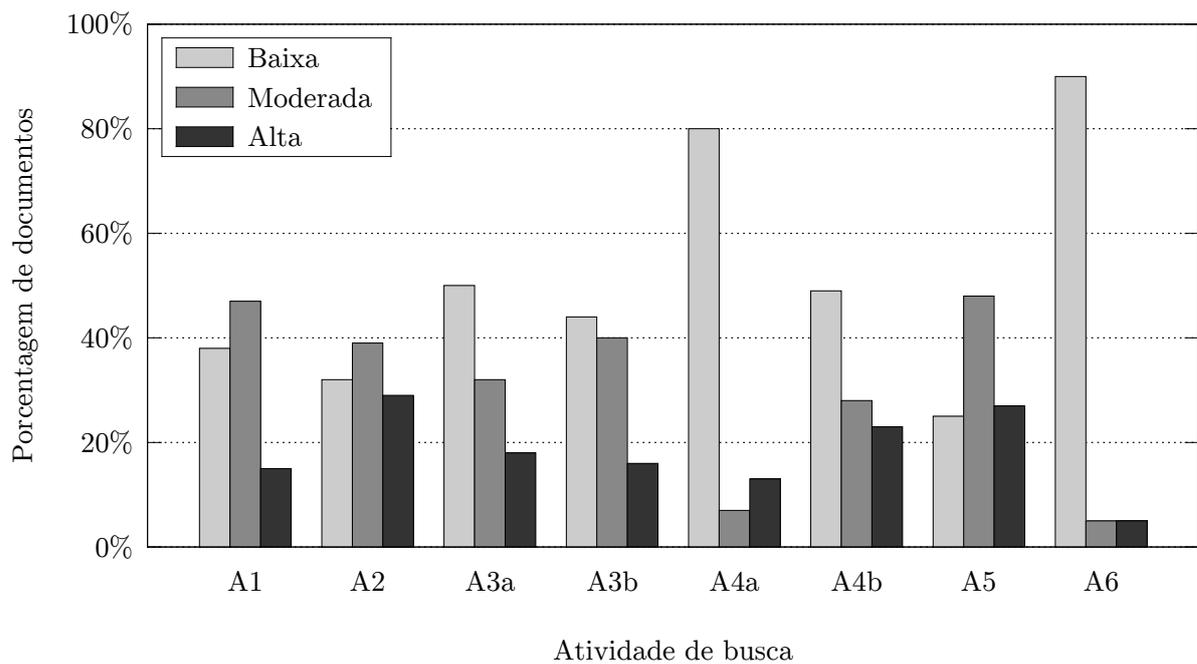


Figura 13 – Distribuição do grau de relevância para cada atividade de busca.

pode ser atribuído às características particulares de cada atividade de busca. Como apresentado na [Seção 4.1](#), as atividades A4a e A6 são caracterizadas respectivamente pela ambiguidade e carência de informação na especificação de suas consultas, o que contribuiu para uma recuperação de baixa precisão.

Em relação à atividade A6, a ambiguidade entre os conceitos [Java: ilha] e [Java: linguagem de programação] não pôde ser resolvida pelo buscador mesmo quando associado ao termo “*located*” (localizado), já que ambos conceitos possuem a propriedade de “localização”. Como consequência, resultados referentes à [Java: linguagem de programação] foram retornados em maior número, provavelmente por serem mais recorrentes do que os relacionados à [Java: ilha]. Já para a atividade A4a, a carência de informação na especificação da consulta e, novamente, a priorização de assuntos mais recorrentes pelo buscador, resultaram em poucos resultados relevantes. Nesse caso, grande parte dos documentos recuperados tratavam sobre o investimento e aplicações financeiras da moeda, e apenas uma pequena parcela dos documentos discorriam sobre a moeda propriamente dita.

Dessa forma, a fim de prover uma interpretação mais consistente dos dados obtidos, as duas atividades foram consideradas *outliers* e, portanto, suprimidas das análises seguintes⁷. A [Figura 14](#) apresenta o comportamento da distribuição sumarizada dos graus de relevância com a remoção dessas atividades *outliers*.

Como podemos observar na [Figura 14](#), os graus de relevância **baixa** e **moderada** representam a maioria dos julgamentos realizados, totalizando respectivamente 39,67% e

⁷ Embora suprimidas das análises, as atividades não comprometem a coleção criada, permitindo que sejam mantidas na avaliação da proposta.

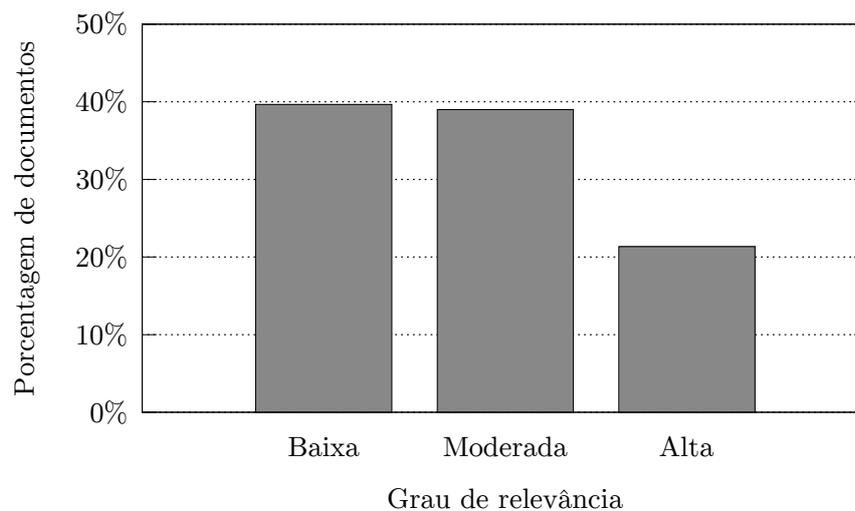


Figura 14 – Distribuição resumida do grau de relevância dos documentos recuperados para as atividades A1, A2, A3a, A3b, A4b e A5.

39% das avaliações. Já o grau de relevância *alta*, o de menor frequência, representa apenas 21,33% dos documentos avaliados. Dessa forma, para que a identificação dos documentos relevantes seja realizada da maneira mais distinta possível, é necessário que a relevância final seja calculada pela composição $r_f \circ r_{rst}^2$. Descartando as duas atividades *outliers*, a aplicação da composição anterior fornece um total de 128 documentos relevantes sobre os 600 documentos recuperados. Já quando a coleção é considerada integralmente, a mesma composição fornece 147 documentos relevantes, os quais foram utilizados como referência para a verificação do desempenho da abordagem.

A noção de relevância estabelecida anteriormente permite que uma nova avaliação seja realizada, dessa vez em relação à distribuição dos documentos relevantes sobre os *rankings* retornados pelo Google. Mantendo a análise isenta das atividades *outliers*, a [Figura 15](#) apresenta o gráfico da distribuição dos 128 documentos relevantes sobre os *rankings* retornados pelo buscador. Assim como é realizado por padrão pelo Google, os 100 documentos recuperados por cada atividade foram organizados sequencialmente em 10 páginas com 10 resultados cada. Os valores descritos sobre as barras de frequência do gráfico representam a métrica *Mean Average Precision* ([MAP](#)) obtida para cada página de resultados.

Segundo [Manning, Raghavan e Schütze \(2008\)](#), a métrica [MAP](#) é frequentemente utilizada para a avaliação da recuperação pois é capaz de fornecer uma medida única de sua qualidade dentre os diferentes níveis de cobertura, apresentando uma boa discriminação e estabilidade. Para uma única necessidade de informação, a Precisão Média (*Average Precision*) é o valor médio das precisões obtidas após cada documento relevante recuperado no *ranking* até n . Tal valor pode ser calculado por meio da [Equação 4.4](#), na qual $P(k)$ é a precisão do *ranking* até um ponto k (caso o documento k seja relevante, 0 caso contrário),

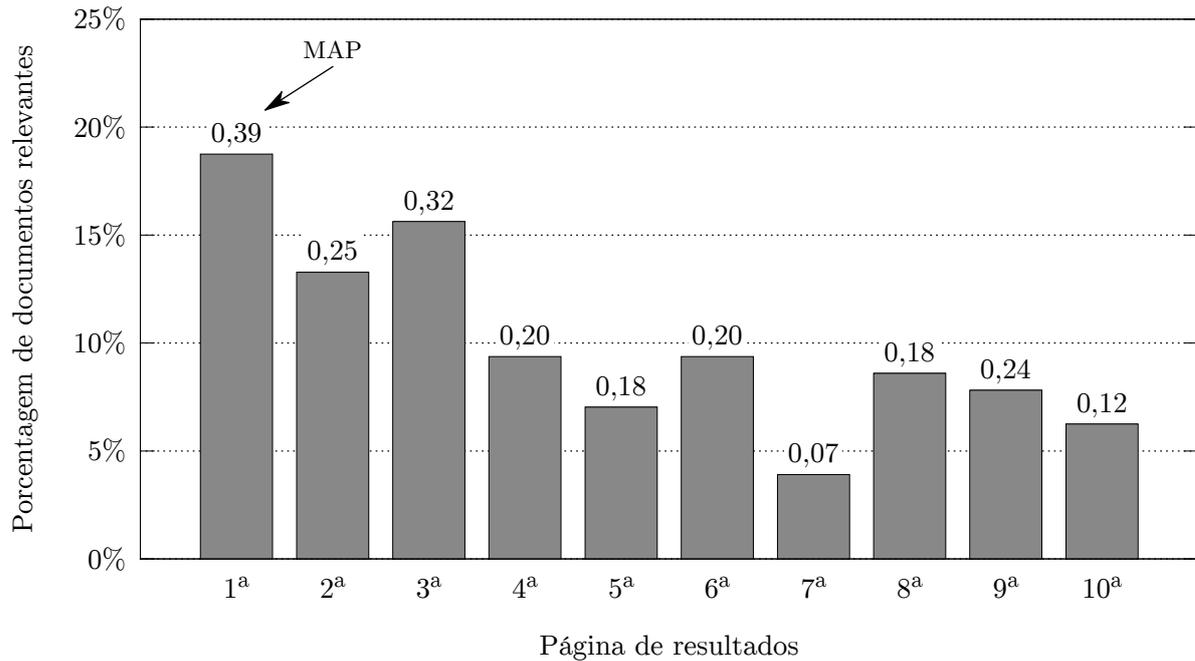


Figura 15 – Distribuição dos documentos relevantes.

e R_n o conjunto de documentos relevantes até n .

$$\text{Pm}(R) = \frac{\sum_{k=1}^n P(k)}{|R_n|} \quad (4.4)$$

Dessa forma, a métrica **MAP** é a média aritmética das Precisões Médias obtidas a partir de todas as necessidades de informação $q \in Q$. A [Equação 4.5](#) descreve o cálculo da métrica **MAP**.

$$\text{MAP}(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \text{Pm}(R_k^j) \quad (4.5)$$

Ao analisar a [Figura 15](#), é possível verificar que a frequência de documentos relevantes e o valor da **MAP** decrescem, embora de maneira não uniforme, conforme a profundidade do *ranking* aumenta. Apesar de ser possível identificar uma tendência em recuperar os documentos relevantes em posições superiores do *ranking*, o desempenho da classificação interna às páginas não foi satisfatório, resultando em uma **MAP** baixa inclusive para os 10 primeiros resultados. De maneira geral, valores baixos de **MAP** indicam que os documentos relevantes recuperados não foram classificados adequadamente pelo buscador, deixando de ocupar as primeiras posições do *ranking*.

Supondo um cenário otimista em que o usuário navega até a terceira página à procura de documentos que satisfaçam sua necessidade de informação, ele encontrará, em média, 10,17 ($\sigma=3,53$) documentos relevantes (48,41% de todos os documentos relevantes recuperados). Esse problema se intensifica conforme o usuário se restringe às páginas mais superficiais. No pior caso, quando ele acessa somente a primeira página, apenas quatro ($\sigma=1,00$) documentos relevantes, em média, terão chance de serem visualizados (ou 19,05% de todos documentos relevantes).

Situações como as descritas anteriormente indicam um comportamento inadequado do buscador em relação às atividades de busca, já que o desejado é que o maior número possível de resultados relevantes seja classificado logo nas primeiras páginas de resultados. Como foi observado, além da recuperação ter apresentado uma precisão baixa, a maioria dos resultados relevantes recuperados correriam o risco de não serem sequer encontrados pelos usuários devido a sua classificação inadequada. Tal comportamento se torna ainda mais prejudicial quando atividades de Pesquisa Exploratória estão envolvidas. Nesses casos, o usuário geralmente está incerto sobre o próprio espaço de busca recuperado e necessitará de um tempo ainda maior para identificar os documentos relevantes.

4.5 Considerações finais

Neste capítulo apresentamos a coleção de teste que guiou a avaliação da abordagem proposta. Devido à natureza particular do problema que estamos tratando, decidimos pela construção da nossa própria coleção de teste. Isso permitiu refletir determinadas características envolvidas no contexto de buscas informacionais, bem como fornecer uma referência para a análise do desempenho da recuperação do buscador utilizado.

A coleção foi construída a partir dos documentos recuperados por oito atividades de busca realizadas em uma versão adaptada do buscador Google. Cada documento teve sua relevância julgada por três juízes em relação a necessidade de informação envolvida. Embora a concordância entre os juízes não tenha se mostrado alta quando os três julgamentos foram considerados simultaneamente, graus de concordância mais elevados foram observados para determinados pares de juízes. Dessa forma, a relevância final de cada documento foi estabelecida a partir de um consenso entre os julgamentos realizados pelos pares de juízes que melhor concordaram em cada tarefa de julgamento realizada.

Por fim, a coleção foi avaliada em relação à distribuição dos documentos relevantes. Foi verificado que, embora o buscador utilizado tenha recuperado a maioria dos documentos relevantes nas primeiras páginas do *ranking*, o número de documentos relevantes recuperados foi baixo e a distribuição deles não foi adequada. Isso indica que o buscador Google não foi capaz de identificar corretamente o interesse envolvido nas atividades de busca, e conseqüentemente, não conseguiu classificar adequadamente os resultados recuperados.

No capítulo seguinte apresentamos a avaliação realizada para a abordagem proposta. Nele discutimos as métricas empregadas durante a avaliação e a maneira com que a coleção de teste foi utilizada para a verificar o desempenho de diversos aspectos da abordagem, como a reclassificação dos documentos, a atualização do modelo e os métodos de seleção dos documentos para a recomendação.

5 Avaliação da DESiRe

Para determinar o desempenho das técnicas aplicadas em nossa abordagem, devemos realizar uma avaliação que permita mensurar quão bem ela satisfaz as expectativas dos usuários em relação a sua eficiência e a sua eficácia no processo de recomendação. Quando avaliamos a eficiência do processo, estamos interessados em verificar se os seus objetivos são cumpridos de maneira viável para que a DESiRe seja implementada como uma ferramenta de busca na *Web*. O principal aspecto avaliado é o tempo de resposta do processo, compreendendo as etapas de construção, atualização e aplicação do MIB. Por outro lado, quando tratamos da eficácia do processo, procuramos verificar se os seus objetivos são cumpridos de maneira satisfatória para usuário. O principal propósito dessa análise é encontrar evidências de que as classificações proporcionadas pela DESiRe superam a qualidade das geradas por buscadores *Web* convencionais, e que o processo de recomendação é igualmente eficaz, fornecendo documentos relevantes e inéditos ao usuário.

A abordagem padrão para avaliar a eficácia de uma recuperação se baseia na noção de documentos relevantes e não relevantes, tendo como métricas frequentemente utilizadas a precisão e a cobertura. Para que possam ser aplicadas, tais métricas exigem que todos os documentos recuperados pelo SRI tenham sido examinados pelo usuário, o que na prática dificilmente acontece. Em vez disso, os documentos são ordenados de acordo com o grau de relevância e então apresentados em uma lista que geralmente se distribui por diversas páginas. O usuário então examina a lista de resultados classificados a partir do topo, fazendo com que as métricas de precisão e cobertura variem conforme ele explora a lista retornada. Dessa forma, uma avaliação apropriada requer que uma curva de precisão *vs.* cobertura seja traçada e a área resultante abaixo dessa curva seja avaliada, resultando em um valor único que represente a eficácia da recuperação. Outras métricas derivadas da precisão, como a Precisão em n ($P@n$), MAP, *Mean Reciprocal Rank* (MRR) e *F-Measure* também traduzem a eficácia da recuperação em um valor único, auxiliando a comparação de desempenho entre diferentes SRIs. Além disso, elas também reduzem problemas causados por documentos relevantes localizados nas últimas posições do *ranking* ao considerarem no cálculo a posição assumida pelo documento. Essa característica é particularmente importante quando os documentos são classificados em forma de lista, já que os usuários geralmente valorizam os documentos melhor classificados.

Embora amplamente utilizadas, a precisão e cobertura, bem como suas métricas derivadas, tratam os documentos apenas como relevantes e não relevantes. Como resultado, a distinção entre diferentes técnicas pode ficar ofuscada quando uma recupera resultados “altamente” relevantes no topo do *ranking* e outra apenas “ligeiramente” relevantes nas mesmas posições. Essa limitação pode ser superada adotando-se avaliações de relevância graduais e métricas que as combinem efetivamente, como o *Discounted Cumulative Gain* (DCG) (JÄRVELIN; KEKÄLÄINEN, 2002). Além de traduzir a eficácia da recuperação em um único valor numérico a partir da interpretação gradual da relevância dos documentos, o DCG também emprega uma função de desconto sobre a classificação, atribuindo pesos diferentes para cada posição do *ranking*. De maneira geral, o DCG se baseia na ideia de que documentos altamente relevantes devem ser melhor classificados que os documentos ligeiramente relevantes, que por sua vez, devem ser melhor classificados que os documentos menos relevantes.

Conforme é descrito pela Equação 5.1, o DCG consiste de uma soma de pesos sobre o grau de relevância *rel* dos documentos classificados até uma determinada posição *p*. O peso, ou desconto, é calculado por uma função decrescente da posição *i* do item classificado, representando o decaimento da probabilidade do usuário visualizar um documento conforme ele se aprofunda no *ranking*. Na equação, rel_i representa um valor numérico atribuído a cada nível de relevância do documento *i*, sendo 0 associado à relevância baixa, 1 à moderada e 2 à alta.

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log_2(i + 1)} \quad (5.1)$$

Embora o tenhamos definido de maneira específica, o DCG pode ser considerado uma família de métricas de classificação, uma vez que a escolha da função de desconto é flexibilizada. Para esta avaliação, optamos pelo desconto logarítmico ($\frac{1}{\log(i+1)}$) com ênfase na recuperação de documentos relevantes ($2^{rel_i} - 1$). Além de ser a abordagem mais comumente utilizada, o desconto logarítmico permite que a métrica seja empregada na diferenciação de técnicas de classificação de forma consistente quando normalizada (WANG *et al.*, 2013).

Visto que os *rankings* podem variar de tamanho dependendo da consulta utilizada e da situação em que são gerados, a comparação da eficácia de uma determinada técnica de classificação para as diversas configurações geradas não pode ser alcançada consistentemente utilizando apenas o DCG. Dessa forma, o ganho acumulado até uma posição *p* do *ranking* deve ser normalizado pelo valor do maior DCG possível até essa mesma posição. Esse valor é conhecido como DCG-ideal até *p* (IDCG_p) e pode ser calculado a partir do DCG sobre o *ranking* ordenado pela relevância dos seus documentos. Assim, o *Normalized Discounted Cumulative Gain* (NDCG) para um determinado *ranking* até a posição *p* é computado conforme a Equação 5.2.

$$NDCG_p = \frac{DCG_p}{IDCG_p} \quad (5.2)$$

Com o emprego da normalização, o valor resultante do **NDCG** é definido entre $[0; 1]$, permitindo que a métrica seja interpretada como uma medida da qualidade geral da classificação. Valores de **NDCG** próximos a 0 representam que a classificação não foi adequada e provavelmente não atende às expectativas do usuário. Já valores de **NDCG** próximos a 1 indicam que a classificação tende à ideal e, provavelmente, atende às expectativas do usuário. Dessa forma, dentre as métricas apresentadas, a **NDCG** é a mais apropriada para a avaliação da DESiRe, pois fornece uma visão da satisfação do usuário em relação à eficácia da recomendação.

Assim como o estabelecimento das métricas envolvidas na avaliação, a definição da metodologia adotada para sua realização é igualmente importante. Uma maneira de se conduzir a avaliação envolve implementar a abordagem como um protótipo de ferramenta de busca, e em laboratório, realizar experimentos com usuários de busca na *Web*. Durante a experimentação podemos, por exemplo, solicitar aos participantes que utilizem a ferramenta construída para realizar atividades de busca informacionais na *Web* e que avaliem a relevância dos documentos recuperados e recomendados. As avaliações realizadas pelos participantes podem ser utilizadas para medir diretamente a eficácia da abordagem, bem como comparadas com avaliações obtidas a partir de atividades de busca executadas sem o auxílio da ferramenta. A eficiência também pode ser medida diretamente questionando os participantes sobre o tempo de resposta da ferramenta. Embora esse tipo de experimento procure representar as condições reais de um processo de busca, sua realização é de difícil controle e pode se tornar cara devido à necessidade de recrutar usuários em laboratório. Além disso, a implementação de tal ferramenta, mesmo quando limitada ao nível de detalhamento de um protótipo, deve contemplar algum tipo de interação com o usuário, o que por hora, não está no escopo do nosso trabalho.

Uma segunda maneira, mais apropriada ao nosso trabalho, se baseia na instanciação da DESiRe como um agente recomendador que possibilite a verificação do seu desempenho por meio de simulações do comportamento do usuário durante a busca. Nessa estratégia, as atividades de busca e os respectivos documentos recuperados devem ser previamente armazenados em uma coleção de referência, a qual deve codificar as expectativas dos usuários na forma de julgamentos de relevância que permitam avaliar a eficácia da abordagem em relação ao que é oferecido por ferramentas convencionais de busca. Assim como a eficácia, a avaliação da eficiência deve ser realizada de maneira indireta, mensurando o tempo médio de resposta do agente durante as simulações e o comparando com um limiar máximo aceitável. Embora a estratégia de simulação não consiga representar em completude o comportamento de busca do usuário e o próprio contexto da busca, ela permite exercitar as diversas etapas do processo de recomendação de maneira controlada e sistematizada. A seção a seguir descreve a simulação considerada para a DESiRe bem como as avaliações decorrentes da sua execução.

5.1 Simulação das atividades de busca

A simulação das atividades de busca do usuário é o principal recurso que utilizamos para exercitar o processo de recomendação proposto para a DESiRe. Com ela, procuramos representar a captura do interesse do usuário de maneira simplificada, isolando questões sobre a interação com o usuário e a distinção entre *feedbacks* explícitos e implícitos. Embora essa simulação não possibilite representar com precisão as variáveis presentes em uma situação real de busca na *Web*, ela permite que as diferentes etapas do processo sejam sistematicamente avaliadas por meio da aplicação da coleção de teste que construímos.

Como pode ser observado na [Figura 16](#), a simulação acompanha o processo de recomendação estabelecido na proposta, porém adaptado para a utilização da coleção de teste criada. Para tanto, um simulador de *feedbacks* é empregado para selecionar e informar os documentos ao agente. Além disso, quatro iterações são realizadas sobre esse processo como forma de simular uma sequência de *feedbacks* capturados. A decisão por quatro *feedbacks* foi arbitrária, porém suficiente para avaliar o comportamento do processo durante a simulação.

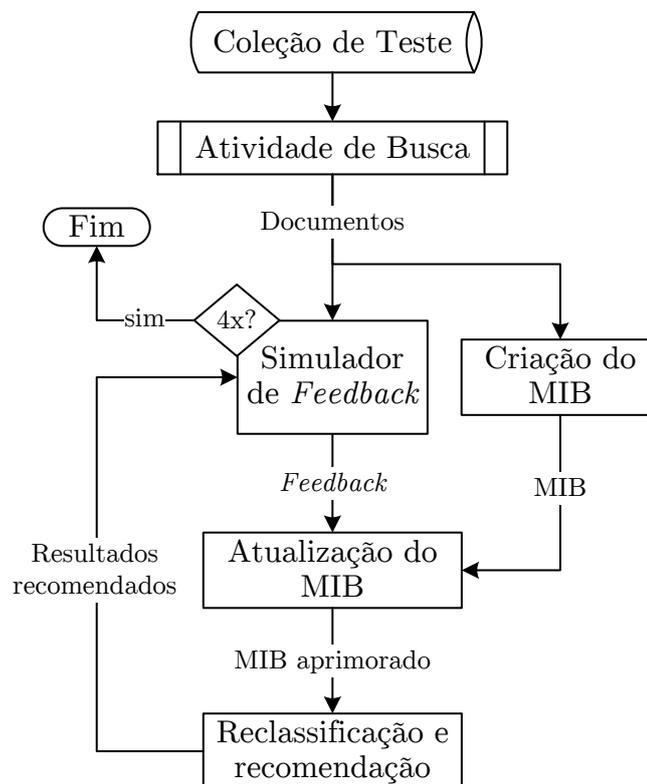


Figura 16 – Processo de recomendação adaptado para a simulação das atividades.

A simulação é inicializada com a criação do [MIB](#) a partir dos documentos recuperados pela atividade de busca presente na coleção de teste, representando uma situação que seria encontrada logo após o usuário realizar uma busca na *Web*. Após a criação do

modelo, o agente está preparado para que a primeira iteração da simulação seja executada. O *feedback* informado nessa iteração será associado ao documento relevante¹ melhor classificado pelo *ranking* original. Essa estratégia procura representar a navegação sequencial do usuário pela lista de resultados e, posteriormente, a captura do seu interesse sobre um documento relevante que tenha visitado.

A primeira iteração da simulação se encerra com a reclassificação do *ranking* a partir do modelo atualizado e com a recomendação de uma sequência de resultados derivada do *ranking* gerado. As outras três iterações seguintes são executadas de maneira análoga, mas dessa vez, utilizando a sequência de resultados recomendados para a seleção dos documentos associados aos próximos *feedbacks*. Essa decisão representa a inspeção realizada pelo usuário sobre os resultados recomendados durante a navegação. Se tais documentos forem relevantes, a necessidade do usuário recorrer à página de resultados será minimizada.

É importante observar que a simulação que definimos anteriormente considera uma situação ideal, na qual temos a garantia de que o documento informado é realmente relevante, e portanto, o *feedback* é confiável. Conforme discutimos na [Subseção 3.3.2](#), a confiança sobre o *feedback* permite que os parâmetros de atualização do modelo sejam ajustados a fim de privilegiar a informação adquirida em relação à informação que já foi consolidada pelo modelo atual. Portanto, como forma de traduzir essa situação, o processo de atualização do modelo foi configurado com $\alpha = 1$ e $\beta = 1,25$ para todas as simulações. Por outro lado, em uma situação real de busca, nem todos os *feedbacks* capturados são confiáveis, já que o usuário pode se mostrar inseguro em relação ao espaço de busca explorado. Dessa forma, com o propósito de verificar o comportamento da DESiRe quando uma situação de incerteza é considerada, apresentamos na [Subseção 5.2.5](#) uma avaliação do tratamento da confiança sobre os *feedbacks* capturados.

A fim de possibilitar a execução da simulação discutida, uma instância da DESiRe foi criada utilizando a linguagem de programação Python (versão 2.7.9)² e ferramentas consolidadas no cenário de processamento e modelagem textual. Para a etapa de pré-processamento linguístico, utilizamos a segmentação de palavras e a *stemização* implementadas pela ferramenta NLTK (versão 3.0)³ (BIRD; KLEIN; LOPER, 2009). A remoção das *stopwords*, por sua vez, foi baseada em um índice composto por palavras disponibilizadas pela NLTK e pelo banco de dados MySQL. Já a estruturação do MIB foi apoiada pela ferramenta Gensim (versão 0.10.2)⁴ (ŘEHŮŘEK; SOJKA, 2010), a qual permitiu criar o MEV de maneira eficiente e simplificada.

A [Tabela 2](#) ilustra o processo de simulação apresentado anteriormente relacionando as reclassificações e recomendações resultantes de cada iteração (*i*) da simulação realizada

¹ Conforme discutido nos capítulos anteriores, a identificação dos documentos relevantes foi realizada utilizando a composição $r_f \circ r_{rst}^2$.

² Disponível em: <<http://www.python.org/>>. Acessado em: 01/06/2015.

³ Disponível em: <<http://www.nltk.org/>>. Acessado em: 01/06/2015.

⁴ Disponível em: <<http://radimrehurek.com/gensim/>>. Acessado em: 01/06/2015.

para a atividade de busca A4a, a qual proporciona uma visão clara do funcionamento da simulação. Para auxiliar a compreensão do processo, os documentos foram identificados de acordo com a posição em que se encontravam no *ranking* original e os documentos relevantes (2, 7, 20, 26, 28, 42, 45, 51, 61, 74, 76, 80, 96) foram destacados em negrito.

Tabela 2 – Classificação e recomendação resultantes para cada *feedback* capturado (f.c.) durante a simulação realizada para a atividade A4a.

<i>i</i>	f.c.	<i>Ranking</i>										Recomendação		
•	•	1	2	3	4	5	6	7	8	9	10	•		
1	2	<u>2</u>	20	80	51	5	74	76	42	61	96	20	80	51
2	20	<u>20</u>	<u>2</u>	80	51	76	5	61	42	74	96	80	51	76
3	80	<u>20</u>	<u>2</u>	<u>80</u>	51	76	61	42	96	5	74	51	76	61
4	51	<u>20</u>	<u>2</u>	<u>80</u>	<u>51</u>	76	61	96	42	74	5	76	61	96

A primeira linha da [Tabela 2](#) apresenta a configuração das 10 primeiras posições do *ranking* original, logo antes de ocorrer a primeira iteração da simulação. As demais linhas se referem ao *ranking* e recomendação resultantes de cada iteração para o respectivo *feedback* capturado (f.c.). Como pode ser verificado, o documento 2 foi selecionado na primeira iteração por se tratar do documento relevante melhor classificado no *ranking* original. Nas iterações seguintes, o critério é mantido, porém as seleções são realizadas a partir da sequência de documentos recomendados. Os documentos selecionados até uma determinada iteração foram destacados com um grifo.

Para essa atividade em particular, o Google recuperou apenas dois documentos relevantes na primeira página de resultados, sendo um na segunda posição e outro na sétima posição. Durante a simulação, o agente instanciado reclassificou grande parte dos documentos relevantes nas primeiras posições do *ranking* logo na primeira iteração. Nas iterações seguintes, os documentos relevantes já reclassificados foram reorganizados conforme sua semelhança com o MIB construído enquanto novos documentos relevantes foram trazidos para a primeira página de resultados.

Outra observação importante é referente à posição assumida no *ranking* pelos documentos selecionados para o *feedback*. Embora esses documentos tenham sido classificados nas primeiras posições, tal comportamento se alterou de maneira particular para cada atividade de busca simulada, não permitindo que o mesmo fosse generalizado. A tabela também torna claro como a recomendação é uma sequência dos documentos melhor classificados, exceto pelos documentos demarcados que já foram associados a algum *feedback*, e portanto, não devem ser reapresentados ao usuário. Para o caso exemplificado anteriormente, a recomendação consistiu da sequência dos três documentos melhor classificados pela iteração. A avaliação de outras estratégias de segmentação do *ranking* é apresentada na [Subseção 5.2.2](#). As tabelas com os resultados das simulações realizadas para todas as atividades de busca são apresentadas no [Apêndice A](#).

Embora a discussão aqui realizada tenha mostrado evidências do desempenho da DESiRe, o mesmo não pode ser quantificado sem que uma métrica adequada seja empregada para mensurar e compará-lo em relação ao que é oferecido pelo Google. Dessa forma, a seção a seguir apresenta a avaliação do desempenho da DESiRe, contemplando métricas capazes de quantificar a eficácia e eficiência da abordagem a partir das simulações realizadas para cada uma das atividades busca presentes na coleção de teste.

5.2 Avaliação do desempenho da DESiRe

A avaliação do desempenho da DESiRe consiste em verificar a eficácia e a eficiência da abordagem por meio da avaliação das reclassificações e recomendações produzidas, bem como do tempo despendido por cada etapa do processo de recomendação. A avaliação das reclassificações é fundamental para verificarmos se o modelo foi capaz de representar os interesses de busca do usuário e recuperar os documentos relevantes de maneira adequada. Ao medir a qualidade das reclassificações realizadas durante as simulações podemos avaliar o comportamento desempenhado pelo agente e comparar sua qualidade geral em relação aos *rankings* fornecidos pelo Google.

Já a avaliação da recomendação permite verificar se os documentos apresentados são realmente relevantes ao usuário e se a sequência com que foram sugeridos também é adequada. Assim como as reclassificações, as recomendações também foram avaliadas em função da qualidade das sequências sugeridas. Conforme discutimos anteriormente, a métrica **NDCG** é capaz de traduzir a relevância e posição atribuídas a cada documento recuperado em um grau de qualidade da classificação.

Embora a avaliação das reclassificações seja importante para determinar a eficácia da DESiRe, a viabilidade de implementá-la como uma ferramenta de buscas na *Web* fica condicionada a uma avaliação de sua eficiência. Para tanto, também avaliamos o tempo despendido por cada etapa projetada para abordagem por meio das simulações realizadas com o agente instanciado. As seções a seguir detalham e apresentam os resultados obtidos pelas avaliações discutidas anteriormente.

5.2.1 Avaliação da eficácia das reclassificações

O conceito de qualidade fornecido pela métrica **NDCG** foi o principal método utilizado para avaliarmos a eficácia das reclassificações produzidas. Quando aplicada em cada posição do *ranking*, o **NDCG** permite avaliar como os documentos relevantes foram distribuídos, além de apontar a qualidade dessa classificação até a última posição considerada. Para determinar a qualidade do *ranking* decidimos considerar apenas as suas 10 primeiras posições, compreendendo a primeira página de resultados, as quais possuem maior probabilidade de serem visualizadas pelos usuários.

A [Tabela 3](#) apresenta os valores da métrica calculada em cada posição dos *rankings* gerados pelas iterações da simulação exemplificada pela [Tabela 2](#). O cálculo da **NDCG** em todas as posições do *ranking* fornece uma visão da organização dos documentos recuperados em função da qualidade acumulada até a posição considerada. Como forma de comparar a qualidade alcançada pelas reclassificações com o que foi oferecido pelo Google, os valores da **NDCG** calculados para cada posição do *ranking* original foram dispostos na primeira linha da tabela.

Tabela 3 – Métrica **NDCG** calculada para as posições do *ranking* resultante em cada iteração da simulação realizada para a atividade A4a.

<i>i</i>	Posição do <i>ranking</i>									
	@1	@2	@3	@4	@5	@6	@7	@8	@9	@10
•	0,000	0,387	0,296	0,246	0,258	0,230	0,300	0,277	0,257	0,241
1	1,000	1,000	1,000	1,000	0,913	0,922	0,929	0,935	0,939	0,943
2	1,000	1,000	1,000	1,000	1,000	0,928	0,935	0,940	0,944	0,948
3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,953	0,956
4	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,958

Como podemos observar na [Tabela 3](#), o Google não foi capaz de classificar um documento relevante para a primeira posição do *ranking*, resultando em uma **NDCG** igual a zero nessa posição. Já para a segunda posição, desta vez ocupada por um documento relevante, o valor da **NDCG** sofreu um acréscimo, mas ainda permaneceu baixo devido ao desconto gerado pela classificação inadequada ocorrida para a posição anterior. Em relação às reclassificações realizadas, observamos um aumento considerável da **NDCG** em todas as posições dos *rankings* resultantes. Logo na primeira reclassificação ($i = 1$), além do documento associado ao *feedback* ter ocupado a primeira posição, outros três documentos relevantes ocuparam as posições seguintes, atingindo qualidade máxima para essas posições. Um comportamento semelhante foi observado para as demais iterações, nas quais novos documentos relevantes foram adicionados ao *ranking* a cada reclassificação. Para a última iteração ($i = 4$), praticamente todas as posições do *ranking* foram ocupadas por documentos relevantes.

Uma maneira de verificar a qualidade geral de cada reclassificação é observando o valor da **NDCG** calculada para a última posição do *ranking* (**NDCG@10**). Logo na primeira iteração, o **NDCG@10** atingiu o valor de 0,943, representando um aumento de 292% na qualidade do *ranking* em relação ao que foi oferecido pelo Google. Já para a última iteração, o **NDCG@10** foi de 0,958, representando um aumento final de 298% na qualidade do *ranking*. Apesar de fornecer fortes indícios de que a DESiRe é capaz de reclassificar os resultados de busca de maneira eficaz, a avaliação apresentada anteriormente é particular à atividade A4a e não é suficiente para tal constatação. Como forma de obter uma visão geral do comportamento das reclassificações, o valor médio da **NDCG** foi calculado para todas as

posições dos *ranking* gerados a partir de todas as atividades de busca. A [Tabela 4](#) relaciona a **NDCG**-média calculada para as reclassificações obtidas a partir das oito simulações.

Tabela 4 – Métrica **NDCG**-média calculada para as posições do *ranking* resultante em cada iteração das simulações realizadas para todas as atividades.

<i>i</i>	Posição do <i>ranking</i>									
	@1	@2	@3	@4	@5	@6	@7	@8	@9	@10
•	0,500	0,548	0,566	0,520	0,496	0,473	0,502	0,506	0,502	0,504
1	1,000	1,000	0,980	0,956	0,929	0,927	0,926	0,916	0,916	0,907
2	1,000	1,000	1,000	1,000	0,989	0,972	0,963	0,960	0,957	0,940
3	1,000	1,000	1,000	1,000	1,000	0,991	0,984	0,966	0,956	0,948
4	1,000	1,000	1,000	1,000	1,000	0,991	0,992	0,973	0,957	0,949

Embora os resultados discutidos até então sejam particulares à atividade A4a, ao comparar os valores relacionados nas Tabelas 3 e 4 verificamos que, de maneira geral, a qualidade média das reclassificações produzidas por todas as atividades foi semelhante à atividade apresentada como exemplo. Outra maneira de se obter uma visão geral do comportamento das reclassificações, desta vez distinguindo as atividades simuladas, é avaliando a qualidade geral dos *rankings* produzidos em cada uma das simulações. A [Tabela 5](#) relaciona a **NDCG@10** calculada para os *rankings* produzidos em cada iteração da simulação realizada para cada atividade de busca. Os valores da **NDCG@10**-média refletem os apresentados na [Tabela 4](#) e estão acompanhados aos respectivos desvios padrão observados.

Tabela 5 – Métrica **NDCG@10** calculada para o *ranking* resultante das iterações da simulação realizada para cada atividade.

<i>i</i>	Atividade								$\mu \pm \sigma$
	A1	A2	A3a	A3b	A4a	A4b	A5	A6	
•	0,541	0,733	0,296	0,539	0,241	0,667	0,554	0,461	$0,504 \pm 0,157$
1	0,781	0,880	0,786	0,894	0,943	1,000	1,000	0,972	$0,907 \pm 0,082$
2	1,000	0,913	0,855	0,832	0,948	1,000	1,000	0,972	$0,940 \pm 0,063$
3	1,000	0,911	0,951	0,815	0,956	0,954	1,000	1,000	$0,948 \pm 0,058$
4	0,958	0,954	0,954	0,904	0,958	0,910	0,956	1,000	$0,949 \pm 0,028$

Em relação ao comportamento particular de algumas atividades, verificamos que o Google obteve um desempenho baixo na classificação dos documentos recuperados, especialmente para as atividades A3a e A4a, resultando em **NDCGs@10**-média iguais a 0,296 e 0,241 respectivamente. Esse comportamento pode ser atribuído principalmente à falta de especificidade na elaboração da consulta, uma característica que está associada

a ambas as atividades. Já os melhores desempenhos foram obtidos para as atividades realizadas a partir de consultas bem especificadas como a A2 e A4b.

Em relação ao comportamento geral observado para as atividades, verificamos que a qualidade média atingida pelo Google foi razoável, resultando em uma $NDCG@10$ -média de 0,504. Devido a qualidade das recuperações terem sido influenciadas pelas características particulares de cada atividade, o desvio padrão observado para a qualidade dos *rankings* originais foi alto quando comparado aos desvios obtidos pelas qualidades de cada reclassificação.

Conforme os *rankings* originais foram reclassificados, a $NDCG@10$ -média aumentou e o respectivo desvio padrão diminuiu para cada iteração, indicando que a qualidade das reclassificações convergiram para um alto valor. Logo na segunda iteração, foi possível verificar um aumento aproximado de 80% para a $NDCG@10$ -média em relação às classificações apresentadas pelo Google. Para a última iteração, a qualidade média das reclassificações atingiu um aumento próximo a 88% em relação ao *ranking* original fornecido pelo buscador.

Embora as Tabelas 4 e 5 apresentem o desempenho geral da DESiRe em relação às reclassificações proporcionadas, a avaliação da qualidade real dos *rankings* gerados em cada iteração fica prejudicada pela inclusão dos documentos envolvidos nos *feedbacks*. Uma vez que estes documentos não serão mais considerados para as recomendações, uma análise deve ser realizada sem a sua presença. A Tabela 6 relaciona os valores $NDCG$ -médios para os *rankings* gerados assim como foi realizado anteriormente na Tabela 4, porém desconsiderando os documentos selecionados em cada iteração.

Tabela 6 – Métrica $NDCG$ -média calculada para as posições do *ranking* puro resultante em cada iteração das simulações realizadas para todas as atividades.

i	Posição do <i>ranking</i>									
	@1	@2	@3	@4	@5	@6	@7	@8	@9	@10
•	0,500	0,548	0,566	0,520	0,496	0,473	0,502	0,506	0,502	0,504
1	1,000	0,968	0,936	0,905	0,906	0,908	0,897	0,899	0,890	0,897
2	1,000	1,000	0,980	0,956	0,945	0,942	0,940	0,920	0,925	0,915
3	1,000	1,000	0,980	0,970	0,941	0,929	0,921	0,914	0,908	0,890
4	1,000	0,968	0,975	0,937	0,913	0,904	0,890	0,886	0,861	0,862

Observando a Tabela 6 podemos verificar que os valores médios da $NDCG$ decaíram em relação aos apresentados na Tabela 4. Isso se deve ao comportamento da DESiRe em classificar os documentos associados ao *feedback* nas primeiras posições do *ranking*. Uma vez que esses documentos possuem relevância alta, a sua permanência no *ranking* contribuía para uma $NDCG$ elevada. Entretanto, mesmo com a remoção dos documentos selecionados, a $NDCG$ -média se mostrou alta inclusive para últimas posições da primeira página. Além disso, a $NDCG$ -média obtida para a primeira posição do *ranking* atingiu

o valor máximo em todas as iterações, indicando que para todas as iterações havia a possibilidade de pelo menos um novo documento relevante ser recomendado.

5.2.2 Avaliação da eficácia das recomendações

A avaliação da recomendação seguiu a mesma estratégia empregada na avaliação das reclassificações. Para cada iteração da simulação, a qualidade da recomendação foi medida pela métrica **NDCG** calculada para a última posição da sequência de documentos sugeridos. A [Tabela 7](#) relaciona os valores de **NDCG** obtidos para as recomendações de cada uma das atividades simuladas, juntamente da **NDCG**-média e desvio padrão calculados a partir de todas as simulações.

Tabela 7 – Métrica **NDCG** calculada para as recomendações resultantes das iterações da simulação realizada para cada atividades.

<i>i</i>	Atividade								$\mu \pm \sigma$
	A1	A2	A3a	A3b	A4a	A4b	A5	A6	
1	1,000	0,844	0,803	0,844	1,000	1,000	1,000	1,000	0,936 \pm 0,083
2	1,000	1,000	0,844	1,000	1,000	1,000	1,000	1,000	0,980 \pm 0,052
3	1,000	1,000	1,000	0,844	1,000	1,000	1,000	1,000	0,980 \pm 0,052
4	1,000	1,000	1,000	0,803	1,000	1,000	1,000	1,000	0,975 \pm 0,065

Como podemos observar na [Tabela 7](#), enquanto para a maioria das atividades as recomendações atingiram qualidade máxima em todas as iterações, outras atividades mostraram um comportamento evolutivo ou até mesmo instável em relação a qualidade das recomendações. Esse comportamento está diretamente relacionado com a maneira com que os documentos foram reclassificados e com a estratégia de segmentação utilizada para derivar a recomendação. Mesmo *rankings* de alta qualidade podem gerar recomendações com qualidade reduzida se a segmentação incluir algum documento não relevante que tenha sido classificado inadequadamente. Uma vez que a recomendação geralmente compreende menos posições que o *ranking*, o cálculo de **NDCG** para a mesma fica extremamente sensível à inclusão de documentos não relevantes.

A avaliação realizada anteriormente considerou recomendações derivadas a partir da segmentação prefixada na terceira posição de cada *ranking* gerado. Entretanto, conforme discutimos na [Subseção 3.3.3](#), outras estratégias de segmentação podem ser consideradas para a derivação das recomendações. A [Tabela 8](#) relaciona a **NDCG**-média calculada para as recomendações derivadas utilizando segmentações prefixadas na segunda (Fixa@2), terceira (Fixa@3) e quarta (Fixa@4) posições do *ranking*, bem como utilizando uma segmentação autônoma baseada na organização dos documentos reclassificados. Além disso, a qualidade média global foi calculada a partir de todas as recomendações como um todo, não fazendo distinção entre as iterações.

Tabela 8 – Métrica **NDCG**-média calculada para as recomendações derivadas de diferentes segmentações em cada iteração das simulações realizadas para todas as atividades.

i	Segmentação			
	Fixa@2	Fixa@3	Fixa@4	Autônoma
1	$0,968 \pm 0,085$	$0,936 \pm 0,083$	$0,905 \pm 0,083$	$0,943 \pm 0,098$
2	$1,000 \pm 0,000$	$0,980 \pm 0,052$	$0,956 \pm 0,057$	$0,984 \pm 0,043$
3	$1,000 \pm 0,000$	$0,980 \pm 0,052$	$0,970 \pm 0,053$	$0,941 \pm 0,065$
4	$0,968 \pm 0,085$	$0,975 \pm 0,065$	$0,937 \pm 0,065$	$0,922 \pm 0,081$
Global	$0,984 \pm 0,062$	$0,968 \pm 0,067$	$0,942 \pm 0,070$	$0,948 \pm 0,078$

Dentre as estratégias pré-fixadas avaliadas, a que divide o *ranking* puro na segunda posição (Fixa@2) obteve os maiores valores de **NDCG**-médio para todas as iterações. Esse resultado pode ser atribuído à posição utilizada para a segmentação. Quanto menor for o tamanho do segmento, menor será a probabilidade de incluir algum documento não relevante e prejudicar a qualidade da recomendação. Tal comportamento fica evidente quando observamos a qualidade média geral, a qual reduziu gradativamente a medida que segmentos maiores foram considerados. Por outro lado, todas as estratégias de segmentação atingiram uma alta qualidade, apresentando uma pequena variação entre elas.

Já a segmentação autônoma, apesar de ter obtido apenas a terceira melhor qualidade média, recomendou em média $4 \pm 3,3$ documentos, superando as estratégias Fixa@3 e Fixa@4. O desvio padrão observado para a segmentação autônoma também se mostrou superior em relação as outras segmentações, o que era esperado já que a diversificação de segmentos é um comportamento característico dessa estratégia.

Outra avaliação que podemos realizar em relação às estratégias de segmentação é sobre número de documentos inéditos recomendados. Como discutido no [Capítulo 3](#), um documento é considerado inédito quando ele não se faz presente na primeira página de resultados. A [Tabela 9](#) apresenta o número médio de documentos inéditos recomendados em cada iteração por cada estratégia de segmentação. A média global desses documentos foi calculada suprimindo a distinção entre as iterações.

Como podemos observar, a quantidade de documentos inéditos recomendados aumenta conforme segmentos maiores são considerados. Entretanto, quando avaliamos a porcentagem de documentos inéditos recomendados, essa regra não é mantida. De maneira geral, essa taxa se manteve aproximadamente a mesma para todas as estratégias, variando de 77% para as segmentações Fixa@3 e Fixa@4, até 78% e 80% para as segmentações autônoma e Fixa@2 respectivamente. Esse resultado indica que a taxa de documentos inéditos recomendados foi alta mesmo quando segmentos menores eram considerados.

Tabela 9 – Número médio de documentos inéditos recomendados por diferentes segmentações em cada iteração das simulações realizadas para todas as atividades.

i	Segmentação			
	Fixa@2	Fixa@3	Fixa@4	Autônoma
1	$1,6 \pm 0,5$	$2,3 \pm 0,4$	$3,0 \pm 0,5$	$2,8 \pm 2,6$
2	$1,5 \pm 0,5$	$2,1 \pm 0,8$	$3,0 \pm 0,7$	$2,4 \pm 2,2$
3	$1,5 \pm 0,5$	$2,4 \pm 0,7$	$3,1 \pm 0,8$	$4,0 \pm 3,7$
4	$1,8 \pm 0,7$	$2,6 \pm 0,7$	$3,3 \pm 0,8$	$3,9 \pm 2,7$
Global	$1,6 \pm 0,6$	$2,3 \pm 0,7$	$3,1 \pm 0,7$	$3,3 \pm 3,0$

5.2.3 Avaliação da eficiência da DESiRe

Como discutimos anteriormente, a aplicação da DESiRe em um ambiente real de busca na *Web* está condicionada a avaliação da sua eficiência. Ao mensurar o tempo despendido em cada etapa do processo de recomendação, podemos verificar por meio do agente instanciado evidências que indiquem a viabilidade de implementação da abordagem. Dessa forma, os tempos médios consumidos pelas etapas do processo de recomendação foram aferidos a partir de 100 execuções de cada simulação em um computador de porte pessoal⁵. A [Tabela 10](#) relaciona os tempos médios obtidos durante a criação e atualizações do MIB, bem como durante as reclassificações do *ranking* e recomendações geradas em cada iteração.

Tabela 10 – Tempo médio aferido para as etapas do processo de recomendação a partir de 100 execuções de cada simulação.

Etapa	Tempo $\mu \pm \sigma$ (s)
Criação do MIB	$0,903 \pm 0,137$
Atualizações do MIB	$0,005 \pm 0,008$
Reclassificações	$0,003 \pm 0,006$
Recomendações (autônoma)	$0,001 \pm 0,003$
Total	$0,912 \pm 0,132$

Como podemos verificar na tabela anterior, o tempo médio total despendido nas etapas do processo, *i.e.*, o tempo de resposta do agente, não ultrapassou 0,912 segundos. Esse valor é aceitável para uma aplicação *Web*, principalmente pelo fato de que a navegação do usuário não está condicionada ao término do processamento, permitindo que coexistam de maneira paralela. Adicionalmente, considerando que as simulações foram executadas sobre a configuração de um computador pessoal, tal evidência fortalece a viabilidade de implementação da DESiRe como uma ferramenta de busca que resida no dispositivo

⁵ O computador utilizado para as simulações executava o sistema operacional Linux (Ubuntu 14.10) sobre um processador Intel i3-540 em 3.06 GHz e 8 GB de memória RAM DDR3 em 1333 MHz.

do usuário, como uma extensão para o navegador, por exemplo. Nesse caso, devemos também considerar os atrasos de transmissão e descompactação do modelo criado para cada atividade de busca. Ainda utilizando as simulações realizadas anteriormente, verificamos que o tamanho médio dos modelos compactados atingiu $70,8 \pm 10,4$ KB, o que não implica em uma sobrecarga na transmissão dos dados já existente durante a busca. A descompactação do modelo também não representa um gargalo para o processamento da ferramenta, consumindo em média $0,002 \pm 0,005$ segundos.

5.2.4 Avaliação da propagação do MIB

Um dos aspectos que propomos para a DESiRe foi a possibilidade de propagar o MIB construído durante uma atividade de busca para as atividades seguintes pertencentes a uma mesma tarefa de busca. O propósito da propagação é permitir que, dentro de um mesmo contexto, a informação consolidada pelo modelo durante uma determinada atividade de busca possa ser reutilizada em outra atividade alvo a fim de antecipar a reclassificação do *ranking* fornecido pelo buscador.

Para que o MIB propagado possa ser reaproveitado pela atividade alvo, sua representação deve ser convertida em função do MEV criado para essa atividade. Dessa forma, o MIB construído durante a atividade de busca antecessora pertencerá ao mesmo espaço vetorial dos documentos recuperados pela atividade alvo, permitindo que seja empregado na reclassificação desses documentos.

Esta avaliação tem como objetivo verificar o ganho de qualidade sobre o *ranking* original de uma atividade alvo quando o modelo propagado a partir de uma atividade antecessora é empregado na reclassificação dos seus documentos recuperados. Para tanto, utilizaremos os pares de atividades de busca congêneres definidas na coleção de teste para simular a propagação do modelo entre as mesmas. Por se tratarem de atividades realizadas sobre o mesmo contexto, consideramos que a propagação do MIB entre elas é factível.

A avaliação consiste em estabelecer o MIB da atividade de busca antecessora ao realizar sua simulação utilizando o processo padrão que definimos. Em seguida, o MIB é readequado para o MEV da atividade alvo e empregado na reclassificação dos documentos fornecido pelo Google para essa atividade. Definimos os pares (A3a, A3b) e (A4a, A4b) como as relações entre as atividades antecessoras e alvo, uma vez que A3a e A4a representam versões iniciais das suas respectivas atividade sucessoras A3b e A4b.

As Tabelas 11 e 12 apresentam os valores da NDCG calculada para cada posição dos *rankings* original ($r=A$) e reclassificado ($r=B$) pela propagação dos modelos originados pela atividades antecessoras A3a e A4a para as respectivas atividades alvo A3b e A4b. Como podemos observar nas tabelas, em ambas as simulações houve um aumento da qualidade do *ranking* reclassificado pelo modelo propagado em relação ao *ranking* apresentado pelo buscador.

Tabela 11 – Métrica **NDCG** calculada para as posições do *ranking* original da atividade A3b (A) e do *ranking* reclassificado (B) a partir da propagação do **MIB** construído durante a atividade A3a.

<i>r</i>	Posição do <i>ranking</i>									
	@1	@2	@3	@4	@5	@6	@7	@8	@9	@10
A	0,333	0,333	0,490	0,576	0,544	0,521	0,565	0,546	0,508	0,539
B	1,000	0,742	0,803	0,836	0,857	0,873	0,884	0,840	0,852	0,819

Tabela 12 – Métrica **NDCG** calculada para as posições do *ranking* original da atividade A4b (A) e do *ranking* reclassificado (B) a partir da propagação do **MIB** construído durante a atividade A4a.

<i>r</i>	Posição do <i>ranking</i>									
	@1	@2	@3	@4	@5	@6	@7	@8	@9	@10
A	1,000	0,742	0,803	0,724	0,672	0,600	0,637	0,666	0,689	0,667
B	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,953	0,956

Para a propagação entre as atividades A3a e A3b, a métrica **NDCG@10** atingiu o valor de 0,819 para a reclassificação, representando um aumento de 52% em relação ao *ranking* original. Além disso, seis documentos relevantes foram reclassificados nas dez primeiras posições do *ranking*, contra apenas dois providos pelo Google. Já para a propagação entre A4a e A4b, a métrica **NDCG@10** atingiu 0,956 para a reclassificação, representando um aumento de 43% em relação ao *ranking* original. Para essa propagação, nove documentos relevantes foram classificados nas dez primeiras posições do *ranking* contra apenas dois providos pelo Google. Dessa forma, os resultados que obtivemos por esta avaliação indicam que, em uma situação controlada na qual temos a garantia de que as atividades de busca permanecem sobre o mesmo contexto, a propagação do **MIB** entre uma atividade antecessora e outra atividade alvo proporcionam, de maneira geral, reclassificações com maior qualidade em relação às apresentadas pelo buscador.

5.2.5 Avaliação da confiança sobre o *feedback*

De forma complementar às avaliações realizadas anteriormente, o objetivo desta avaliação é verificar a importância de se tratar o conceito de confiança sobre o *feedback* informado pelo usuário. Como discutimos anteriormente, o tratamento da confiança sobre o *feedback* permite ao processo de atualização do **MIB** reduzir o impacto causado por um documento não relevante que seja implicitamente informado pelo usuário. Dessa forma, a fim de possibilitar a avaliação desse impacto, modificamos o fluxo padrão das simulações para que um *feedback* suspeito associado a um documento não relevante fosse incluído na segunda

iteração. Assim, podemos definir e avaliar duas situações para o processo de atualização do modelo, uma ignorando a confiança sobre o *feedback* capturado e outra a tratando.

Para a situação em que a confiança é ignorada, mantivemos os parâmetros α e β inalterados para todas as iterações, assim como o fizemos durante a avaliação da eficácia da abordagem. Já para a situação em que a confiança é tratada pelo processo de atualização do MIB, simulamos um ajuste no valor de β de 1,25 para 0,25 na segunda iteração, e de α de 1 para 0,75 na terceira iteração. Ao reduzir o valor de β na segunda iteração, estamos amenizando o impacto de uma informação incondizente com os interesses do usuário sobre o modelo construído. Quando reduzimos o valor de α na terceira iteração, estamos reforçando uma atenuação na informação já consolidada em função da informação suspeita adquirida na iteração anterior.

A Tabela 13 relaciona a NDCG-média geral calculada para as reclassificações e recomendações geradas em cada situação considerada. A segunda iteração está destacada representando o momento em que o *feedback* suspeito foi informado. Como já era esperado, a primeira iteração não mostrou alterações para nenhuma das situações, uma vez que o comportamento da DESiRe é mantido até a captura do primeiro *feedback* suspeito. Já para a segunda iteração, momento em que o *feedback* suspeito foi capturado, podemos observar uma queda na qualidade do *ranking* e recomendações para ambas as situações. Comparando as qualidades resultantes a partir desse ponto, verificamos que quando o conceito de confiança é ignorado pela atualização do modelo, o *feedback* suspeito provoca uma queda mais acentuada na qualidade do processo em relação à observada quando a confiança é tratada. Nesse segundo caso, a qualidade é recuperada mais rapidamente e atinge um valor mais alto na última iteração, se aproximando dos valores obtidos quando o *feedback* suspeito não era fornecido.

Tabela 13 – Métrica NDCG-média calculada para as reclassificações e recomendações produzidas em cada iteração das simulações nas situações em que o processo de atualização do MIB ignora e trata a confiança sobre o *feedback* capturado.

i	Confiança ignorada		Confiança tratada	
	Reclassificação	Recomendação	Reclassificação	Recomendação
1	0,907 \pm 0,082	0,936 \pm 0,083	0,907 \pm 0,082	0,936 \pm 0,083
2	0,673 \pm 0,094	0,716 \pm 0,193	0,825 \pm 0,073	0,922 \pm 0,111
3	0,804 \pm 0,080	0,811 \pm 0,155	0,874 \pm 0,058	0,936 \pm 0,083
4	0,856 \pm 0,066	0,879 \pm 0,165	0,920 \pm 0,056	0,980 \pm 0,052

5.3 Considerações finais

Neste capítulo apresentamos a avaliação da abordagem DESiRe, a qual consistiu principalmente em verificar a eficácia e eficiência do processo de recomendação. Para tanto,

instanciamos a abordagem como um agente recomendador e adotamos a estratégia de simulação das atividades de busca como metodologia de avaliação. Dessa forma, foi possível realizar uma avaliação sistematizada da abordagem, cobrindo de maneira isolada cada etapa do processo.

A eficácia da DESiRe foi medida em função da qualidade das reclassificações e recomendações. Como observamos durante as simulações, a qualidade média dos *rankings* reclassificados apresentou um acréscimo de até 88% em relação aos *rankings* originais. Além disso, analisando os *rankings* puros produzidos pelas reclassificações, verificamos que ao menos um novo documento relevante era oferecido em cada iteração.

Já em relação às recomendações, avaliamos diferentes estratégias de segmentação do *ranking* para a sugestão dos documentos. Verificamos que tanto as segmentações pré-fixadas quanto a segmentação autônoma produziram recomendações de alta qualidade, atingido **NDCGs**-média globais acima de 0,942. Entretanto, analisando a relação entre qualidade e número de documentos recomendados, verificamos que a segmentação autônoma superou as demais segmentações. O mesmo pôde ser observado para a avaliação do número de documentos inéditos recomendados, na qual a segmentação autônoma também superou as demais estratégias.

Embora a avaliação tenha se limitado a simulações sobre uma instância da DESiRe, os resultados obtidos indicam que as reclassificações proporcionadas possuem qualidade superior em relação à classificação gerada por um buscador *Web* convencional, e que o processo de recomendação é igualmente eficaz, fornecendo documentos relevantes e inéditos ao usuário. Na avaliação da eficiência da abordagem, mostramos a viabilidade de implementá-la como uma ferramenta de busca que resida no dispositivo do usuário, atingindo um tempo de resposta inferior a 0,912 segundos.

Por fim, avaliamos a possibilidade de propagação do **MIB**, demonstrando que a reutilização da informação consolidada durante uma atividade de busca pode ser reutilizada por outra atividade alvo para reclassificar de maneira eficaz os documentos recuperados pela busca, desde que o contexto da mesma seja mantido. Além disso, durante a avaliação do conceito de confiança sobre o *feedback*, verificamos que quando a confiança é tratada pelo processo de atualização do **MIB**, o impacto causado por um documento não relevante associado a um *feedback* suspeito é inferior ao causado quando a confiança não é tratada.

O capítulo seguinte encerra esta dissertação com as conclusões obtidas a partir do desenvolvimento e avaliação da proposta. Nesse capítulo também discutimos as contribuições que acreditamos ter proporcionado com o trabalho, bem como limitações identificadas para a DESiRe. A partir dessas limitações descrevemos algumas direções para o desenvolvimento de trabalhos futuros.

6 Conclusão

Neste trabalho apresentamos a DESiRe, uma abordagem dinâmica para recomendação de resultados de busca, baseada no interesse do usuário durante uma tarefa específica de busca na *Web*. O propósito da DESiRe foi explorar uma maneira mais dinâmica e interativa para a apresentação dos resultados de busca, se diferenciando dos trabalhos desenvolvidos até então que restringiram seus esforços de personalização apenas à reorganização das páginas de resultados. Ao realizarmos a recomendação de resultados relevantes enquanto o usuário navega pelo espaço de busca, antecipamos a descoberta de documentos inéditos e de interesse que foram recuperados, mas que corriam o risco de não serem visualizados por residirem em posições profundas no *ranking*.

Como forma de representar os interesses do usuário durante uma atividade de busca, propomos a utilização do MIB, um modelo baseado no MEV capaz de agregar as informações associadas aos *feedbacks* capturados e reclassificar os documentos recuperados pela busca. Dessa forma, a partir do modelo criado, estabelecemos um processo de recomendação que pudesse ser adicionado ao processo padrão de busca. De maneira geral, tal processo consistiu em criar o MIB a partir dos documentos recuperados pela atividade de busca e, para cada *feedback* capturado, atualizar o modelo de acordo com a informação adquirida, reclassificar os documentos recuperados e recomendá-los ao usuário.

Para verificar se a DESiRe é capaz de oferecer recomendações relevantes ao usuário de maneira viável, medimos sua eficácia e eficiência a partir de simulações realizadas com uma coleção de teste que criamos especificamente para esse objetivo. Em relação a coleção criada, ela ainda possibilitou medir o desempenho de recuperação do mecanismo de busca Google em relação a um conjunto de atividades de busca que definimos. A avaliação que realizamos para a coleção apontou que o mecanismo utilizado não foi capaz de classificar adequadamente os resultados, fortalecendo a premissa de que os esforços para melhorar a experiência de busca do usuário ainda não foram exauridos.

Em relação a avaliação que realizamos para a DESiRe, verificamos que o processo de recomendação foi eficaz e eficiente, oferecendo recomendações de alta qualidade em um tempo de resposta viável para a sua aplicação no contexto de buscas na *Web*. Além disso, avaliamos de maneira preliminar a capacidade da abordagem em propagar o MIB pelas diversas atividades que podem compor uma mesma tarefa de busca. Como resultado obtivemos que a propagação do modelo contribuí com reclassificações de maior qualidade em

relação às apresentadas pelo buscador. Por fim, avaliamos também de maneira preliminar como o conceito de confiança sobre o *feedback* capturado pode alterar a qualidade da recomendação. Quando simulamos a captura de um *feedback* não relevante, verificamos que o impacto causado na qualidade final da recomendação era amenizado quando o mesmo era tratado como suspeito.

Dessa forma, as evidências observadas durante a avaliação da DESiRe nos permitiram concluir que a abordagem proposta possui um grande potencial de emprego em ferramentas que visam apoiar os usuários em atividades de Pesquisa Exploratória na *Web*. Também foi possível concluir que, mesmo ao utilizar técnicas tradicionais para representar o interesse do usuário, conseguimos atingir bons resultados para a recomendação dinâmica de documentos relevantes.

Embora os resultados obtidos por meio das simulações tenham mostrado o potencial da abordagem, a mesma necessita ser validada em contextos reais de busca na *Web*. Dessa forma, como trabalho futuro, espera-se realizar um novo conjunto de avaliações, desta vez solicitando a usuários que desempenhem tarefas de Pesquisa Exploratória na *Web* com o apoio de uma ferramenta de busca que implemente a DESiRe.

Outro trabalho futuro que vislumbramos está associado à limitação da abordagem em relação ao espaço de busca recuperado. Caso o buscador não seja capaz de classificar algum resultado relevante no espaço de busca considerado pela DESiRe, a mesma não será capaz de realizar uma recomendação eficaz. Uma maneira de amenizar essa situação é possibilitar a incorporação ao MIB de documentos externos ao espaço de busca, mas que podem ser atingidos pelos usuários por meio da navegação a partir das referências (*links*) presentes nos documentos recuperados.

Outra limitação conhecida é em relação à identificação do contexto de busca. Como discutimos durante o trabalho, identificar a mudança desse contexto permite delimitar o espaço de uma tarefa de busca de tal forma que o MIB construído possa ser propagado de maneira segura e efetiva pela sequência de atividades que compõe essa tarefa. Dessa forma, como extensão imediata do nosso trabalho, vislumbramos o desenvolvimento de um método capaz de identificar a mudança de contexto durante uma tarefa de busca.

Como direção para um outro trabalho futuro, estabelecemos a investigação de técnicas de *feedback* implícito e o estudo de como elas podem ser incorporadas ao processo de atualização do modelo. De maneira simplificada, essa direção consiste em desenvolver um método capaz traduzir o comportamento do usuário para os valores de α e β de forma que eles representem a confiança do sistema sobre o *feedback* capturado.

Também consideramos como uma direção para trabalho futuro o estudo da viabilidade de se empregar modelos mais aprimorados que o MEV para a representação do MIB. Como alternativas temos a técnica de LSI e até mesmo de modelos que excedam o espaço vetorial, como estruturas baseadas em ontologias e em redes de conceitos.

Finalmente, em um aspecto mais amplo da nossa abordagem, devemos estudar como as recomendações influenciam as interações do usuário durante o processo de busca. Embora tenhamos verificado que as recomendações são qualitativamente eficazes, também devemos verificar sua eficácia em participar, e conseqüentemente alterar, o processo de busca desempenhado pelo usuário.

Referências

ASNICAR, F.; TASSO, C. IfWeb: a prototype of user model-based intelligent agent for document filtering and navigation in the World Wide Web. *Sixth International Conference on User Modeling*, Chia, 1997. Citado na página 25.

AULA, A.; MAJARANTA, P.; RÄIHÄ, K.-J. Eye-tracking reveals the personal styles for search result evaluation. In: *Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction*. Berlin, Heidelberg: Springer-Verlag, 2005. (INTERACT'05), p. 1058–1061. ISBN 3-540-28943-7, 978-3-540-28943-2. Citado 2 vezes nas páginas 27 e 28.

BAEZA-YATES, R.; RIBEIRO-NETO, B. *Modern Information Retrieval: The Concepts and Technology behind Search*. Harlow, Essex, England: Addison-Wesley Professional, 2011. v. 82. 944 p. Citado 2 vezes nas páginas 40 e 44.

BIRD, S.; KLEIN, E.; LOPER, E. *Natural Language Processing with Python*. 1^a. ed. Boston, USA: O'Reilly Media, Inc., 2009. Citado na página 66.

BLANCO, R.; HALPIN, H.; HERZIG, D. M.; MIKA, P.; POUND, J.; THOMPSON, H. S.; DUC, T. T. Repeatable and reliable search system evaluation using crowdsourcing. In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, 2011. (SIGIR '11), p. 923–932. ISBN 978-1-4503-0757-4. Citado na página 28.

BLOOM, B. S. *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*. New York, USA: Addison Wesley Publishing Company, 1984. 207 p. ISBN 0582280109. Citado na página 20.

BRANDT, C.; JOACHIMS, T.; YUE, Y.; BANK, J. Dynamic ranked retrieval. *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*, ACM Press, New York, NY, USA, p. 247, 2011. Citado 3 vezes nas páginas 17, 31 e 33.

BRODER, A. A taxonomy of web search. *SIGIR Forum*, ACM, New York, NY, USA, v. 36, n. 2, p. 3–10, set. 2002. ISSN 0163-5840. Citado na página 20.

BUCKLEY, C.; SALTON, G.; ALLEN, J.; SINGHAL, A. Automatic query expansion using SMART: Trec 3. In: *Proceedings of the Third Text REtrieval Conference (TREC3)*. Gaithersburg, Maryland, USA: NIST Special Publication, 1995. p. 500–225. Citado na página 28.

CAO, G.; NIE, J.-Y.; GAO, J.; ROBERTSON, S. Selecting good expansion terms for pseudo-relevance feedback. In: *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, 2008. (SIGIR '08), p. 243–250. ISBN 978-1-60558-164-4. Citado na página 28.

COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, v. 20, p. 37–46, 1960. ISSN 0013-1644. Citado na página 55.

- COYLE, M.; SMYTH, B. Information Recovery and Discovery in Collaborative Web Search. In: *Proceedings of the 29th European Conference on IR Research*. Berlin, Heidelberg: Springer-Verlag, 2007. (ECIR'07), p. 356–367. ISBN 978-3-540-71494-1. Citado na página 24.
- CRAMER, M. *Dynamic search engine results employing user behavior*. San Francisco, USA: United States Patent and Trademark Office, 2010. US8095582 B2. Citado na página 34.
- CRONEN-TOWNSEND, S.; ZHOU, Y.; CROFT, W. B. A framework for selective query expansion. In: *Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2004. (CIKM '04), p. 236–237. ISBN 1-58113-874-1. Citado na página 28.
- FLEISS, J. L. Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin*, v. 76, n. 5, p. 378–382, 1971. ISSN 0033-2909. Citado na página 55.
- GAUCH, S.; SPERETTA, M.; CHANDRAMOULI, A.; MICARELLI, A. User profiles for personalized information access. In: BRUSILOVSKY, P.; KOBSA, A.; NEJDL, W. (Ed.). *The Adaptive Web*. Berlin: Springer Berlin Heidelberg, 2007, (Lecture Notes in Computer Science, v. 4321). p. 54–89. ISBN 978-3-540-72078-2. Citado na página 30.
- HARDTKE, D.; WERTHEIM, M.; CRAMER, M. Demonstration of Improved Search Result Relevancy Using Real-Time Implicit Relevance Feedback. In: *Understanding the user - workshop in conjunction with SIGIR'09*. [S.l.]: CEUR-WS, 2009. Citado na página 17.
- HEARST, M. A. *Search User Interfaces*. New York, New York, USA: Cambridge University Press, 2009. 404 p. ISBN 0521113792. Citado 2 vezes nas páginas 16 e 19.
- HERSH, W.; BUCKLEY, C.; LEONE, T. J.; HICKAM, D. Ohsumed: An interactive retrieval evaluation and new large test collection for research. In: *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: Springer-Verlag New York, Inc., 1994. (SIGIR '94), p. 192–201. ISBN 0-387-19889-X. Citado na página 44.
- HUTCHISON, D.; MITCHELL, J. C. *The Adaptive Web*. Berlin, Germany: Springer Berlin Heidelberg, 2007. v. 4321. ISBN 978-3-540-72078-2. Citado na página 30.
- JANSEN, B. J.; SPINK, A.; SARACEVIC, T. The use of relevance feedback on the web: Implications for web ir system design. In: *Proceedings of WebNet World Conference on the WWW and Internet 1999*. Honolulu, Hawaii: Association for the Advancement of Computing in Education (AACE), 1999. p. 550–555. Disponível em: <<http://www.editlib.org/p/7478>>. Citado na página 27.
- JÄRVELIN, K.; KEKÄLÄINEN, J. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, v. 20, n. 4, p. 422–446, out. 2002. ISSN 10468188. Citado na página 63.
- KIM, J. Y.; CRAMER, M.; TEEVAN, J.; LAGUN, D. Understanding how people interact with web search results that change in real-time using implicit feedback. In: *Proceedings of the 22nd ACM international conference on Conference on information & knowledge*

management - CIKM '13. New York, NY, USA: ACM Press, 2013. p. 2321–2326. ISBN 9781450322638. Citado na página 34.

LANG, K. Newsweeder: Learning to filter netnews. In: RUSSELL, A. P. (Ed.). *Machine Learning Proceedings 1995*. San Francisco, USA: Morgan Kaufmann, 1995. p. 331 – 339. ISBN 978-1-55860-377-6. Citado na página 44.

LEE, U.; LIU, Z.; CHO, J. Automatic identification of user goals in web search. In: *Proceedings of the 14th International Conference on World Wide Web*. New York, NY, USA: ACM, 2005. (WWW '05), p. 391–400. ISBN 1-59593-046-9. Citado 2 vezes nas páginas 27 e 28.

LI, L.; YANG, Z.; WANG, B.; KITSUREGAWA, M. Dynamic adaptation strategies for long-term and short-term user profile to personalize search. In: DONG, G.; LIN, X.; WANG, W.; YANG, Y.; YU, J. (Ed.). *Advances in Data and Web Management*. Berlin, Germany: Springer Berlin Heidelberg, 2007, (Lecture Notes in Computer Science, v. 4505). p. 228–240. ISBN 978-3-540-72483-4. Citado na página 30.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to Information Retrieval*. New York, New York, USA: Cambridge University Press, 2008. v. 1. 496 p. ISSN 08912017. ISBN 0521865719. Citado 2 vezes nas páginas 44 e 59.

MARCHI, K. R. C. *Uma Abordagem para Personalização de Resultados de Busca na Web*. Dissertação (Mestrado) — Universidade Estadual de Maringá, 2010. Citado na página 32.

MARCHIONINI, G. Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science*, v. 40, n. 1, p. 54–66, jan. 1989. ISSN 0002-8231. Citado na página 19.

MARCHIONINI, G. Exploratory search: from finding to understanding. *Communications of the ACM*, v. 49, n. 4, p. 41, abr. 2006. ISSN 00010782. Citado 6 vezes nas páginas 9, 16, 19, 20, 21 e 22.

MARCHIONINI, G.; WHITE, R. W. Find What You Need, Understand What You Find. *International Journal of Human-Computer Interaction*, v. 23, n. 3, p. 205–237, dez. 2007. ISSN 1044-7318. Citado na página 19.

MASLOW, A. H.; FRAGER, R.; FADIMAN, J.; MCREYNOLDS, C.; COX, R. *Motivation and personality*. New York, NY, USA: Harper & Row, 1970. v. 2. Citado na página 16.

MATTHIJS, N.; RADLINSKI, F. Personalizing web search using long term browsing history. In: *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*. New York, New York, USA: ACM Press, 2011. p. 25. ISBN 9781450304931. Citado na página 25.

MICARELLI, A.; GASPARETTI, F.; SCIARRONE, F.; GAUCH, S. Personalized Search on the World Wide Web. p. 195–230, 2007. Citado 5 vezes nas páginas 17, 24, 25, 26 e 28.

MICARELLI, A.; SCIARRONE, F. Anatomy and Empirical Evaluation of an Adaptive Web-Based Information Filtering System. *User Modeling and User-Adapted Interaction*, v. 14, n. 2/3, p. 159–200, jun. 2004. ISSN 0924-1868. Citado na página 26.

- NETTLETON, D. F.; GONZALEZ-CARO, C. Analysis of User Behavior for Web Search Success Using Eye Tracker Data. *2012 Eighth Latin American Web Congress*, IEEE, p. 57–63, out. 2012. Citado na página 24.
- PITKOW, J.; SCHÜTZE, H.; CASS, T.; COOLEY, R.; TURNBULL, D.; EDMONDS, A.; ADAR, E.; BREUEL, T. Personalized search. *Communications of the ACM*, v. 45, n. 9, set. 2002. ISSN 00010782. Citado 2 vezes nas páginas 17 e 26.
- PONTE, J. Language models for relevance feedback. In: CROFT, W. (Ed.). *Advances in Information Retrieval*. Philadelphia, NY, USA: Springer US, 2000, (The Information Retrieval Series, v. 7). p. 73–95. ISBN 978-0-7923-7812-9. Citado na página 28.
- RADLINSKI, F.; BENNETT, P. N.; CARTERETTE, B.; JOACHIMS, T. Redundancy, diversity and interdependent document relevance. *SIGIR Forum*, ACM, New York, NY, USA, v. 43, n. 2, p. 46–52, dez. 2009. ISSN 0163-5840. Citado na página 28.
- RAMAN, K.; JOACHIMS, T.; SHIVASWAMY, P. Structured learning of two-level dynamic rankings. *Proceedings of the 20th ACM international conference on Information and knowledge management - CIKM '11*, ACM Press, New York, NY, USA, p. 291, 2011. Citado 4 vezes nas páginas 9, 17, 30 e 33.
- ŘEHŮŘEK, R.; SOJKA, P. Software Framework for Topic Modelling with Large Corpora. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Valletta, Malta: ELRA, 2010. p. 45–50. Citado na página 66.
- ROCCHIO, J. J. Relevance feedback in information retrieval. In: G, S. (Ed.). *The SMART Retrieval System*. Englewood Cliffs, NJ, USA: Prentice Hall, 1971. p. 313–23. Citado 2 vezes nas páginas 28 e 40.
- RUTHVEN, I.; LALMAS, M. A survey on the use of relevance feedback for information access systems. *Knowl. Eng. Rev.*, Cambridge University Press, New York, NY, USA, v. 18, n. 2, p. 95–145, jun. 2003. ISSN 0269-8889. Citado na página 28.
- SAFIYA, A. S.; MARTIN, B.; ELIZABETH, U. Enhancing the degree of personalization through Vector Space Model and Profile Ontology. In: *The 2013 RIVF International Conference on Computing & Communication Technologies - Research, Innovation, and Vision for Future (RIVF)*. Hanoi, Vietnam: IEEE, 2013. p. 248–252. ISBN 978-1-4799-1350-3. Citado 3 vezes nas páginas 17, 25 e 30.
- SALTON, G.; BUCKLEY, C. Readings in information retrieval. In: JONES, K. S.; WILLETT, P. (Ed.). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1997. cap. Improving Retrieval Performance by Relevance Feedback, p. 355–364. ISBN 1-55860-454-5. Citado na página 28.
- SALTON, G.; WONG, A.; YANG, C. S. A vector space model for automatic indexing. *Communications of the ACM*, ACM, New York, NY, USA, v. 18, n. 11, p. 613–620, nov 1975. ISSN 0001-0782. Citado na página 39.
- SHEN, X.; TAN, B.; ZHAI, C. Context-sensitive information retrieval using implicit feedback. In: *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, 2005. (SIGIR '05), p. 43–50. ISBN 1-59593-034-5. Citado na página 31.

- SHEN, X.; TAN, B.; ZHAI, C. Implicit user modeling for personalized search. In: *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2005. (CIKM '05), p. 824–831. ISBN 1-59593-140-6. Citado 5 vezes nas páginas [9](#), [17](#), [30](#), [31](#) e [32](#).
- SIMON, H. A. Designing organizations for an information rich world. In: GREENBERGER, M. (Ed.). *Computers, communications, and the public interest*. Baltimore: Johns Hopkins Press, 1971. p. 37–72. ISBN 0-8018-1135-X. Citado na página [16](#).
- SPERETTA, M.; GAUCH, S. Personalized Search Based on User Search Histories. In: *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*. Compiegne, France: IEEE, 2005. p. 622–628. ISBN 0-7695-2415-X. Citado na página [26](#).
- SUTCLIFFE, A. Towards a cognitive theory of information retrieval. *Interacting with Computers*, Elsevier, v. 10, n. 3, p. 321–351, jun 1998. ISSN 09535438. Citado na página [19](#).
- TEEVAN, J. How people recall, recognize, and reuse search results. *ACM Transactions on Information Systems (TOIS)*, ACM, New York, NY, USA, v. 26, n. 4, p. 19:1–19:27, out. 2008. ISSN 1046-8188. Citado 2 vezes nas páginas [27](#) e [33](#).
- TEEVAN, J.; DUMAIS, S. T.; HORVITZ, E. Personalizing search via automated analysis of interests and activities. In: *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, 2005. (SIGIR '05), p. 449–456. ISBN 1-59593-034-5. Citado na página [27](#).
- VENINGSTON, K.; SHANMUGALAKSHMI, R. Enhancing personalized web search re-ranking algorithm by incorporating user profile. In: *2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12)*. Coimbatore, Tamilnadu, India: IEEE, 2012. p. 1–6. Citado na página [25](#).
- WANG, H.; HE, X.; CHANG, M.-W.; SONG, Y.; WHITE, R. W.; CHU, W. Personalized ranking model adaptation for web search. *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '13*, ACM Press, New York, NY, USA, p. 323, 2013. Citado na página [30](#).
- WANG, H.; SONG, M. Ckmeans.1d.dp: Optimal k-means clustering in one dimension by dynamic programming. *The R journal*, v. 3/2, p. 29–33, December 2011. ISSN 2073-4859. Citado na página [43](#).
- WANG, S.; XU, K.; ZHANG, Y.; LI, F. Search Engine Optimization Based on Algorithm of BP Neural Networks. *2011 Seventh International Conference on Computational Intelligence and Security*, IEEE, p. 390–394, dez. 2011. Citado na página [25](#).
- WANG, Y.; WANG, L.; LI, Y.; HE, D.; LIU, T.; CHEN, W. A theoretical analysis of NDCG type ranking measures. *CoRR*, abs/1304.6480, 2013. Citado na página [63](#).
- WHITE, R. W.; CHU, W.; HASSAN, A.; HE, X.; SONG, Y.; WANG, H. Enhancing personalized search by mining and modeling task behavior. In: *Proceedings of the 22Nd International Conference on World Wide Web*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2013. (WWW '13), p. 1411–1420. ISBN 978-1-4503-2035-1. Citado na página [25](#).

- WHITE, R. W.; KULES, B.; BEDERSON, B. Exploratory Search Interfaces: Categorization, Clustering and Beyond. *SIGIR Forum*, Citeseer, v. 39, n. 2, p. 52–56, 2005. Citado na página 21.
- WHITE, R. W.; MORRIS, D. Investigating the querying and browsing behavior of advanced search engine users. In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, 2007. (SIGIR '07), p. 255–262. ISBN 978-1-59593-597-7. Citado na página 27.
- WHITE, R. W.; ROTH, R. A. Exploratory Search: Beyond the Query-Response Paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, v. 1, n. 1, p. 1–98, jan. 2009. ISSN 1947-945X. Citado 6 vezes nas páginas 16, 20, 21, 22, 23 e 24.
- WHITE, R. W.; RUTHVEN, I.; JOSE, J. M. Finding relevant documents using top ranking sentences: An evaluation of two alternative schemes. In: *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, 2002. (SIGIR '02), p. 57–64. ISBN 1-58113-561-0. Citado na página 27.
- ZHAI, C.; LAFFERTY, J. Model-based feedback in the language modeling approach to information retrieval. In: *Proceedings of the Tenth International Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2001. (CIKM '01), p. 403–410. ISBN 1-58113-436-3. Citado na página 31.

APÊNDICE **A**

Resultados das simulações

As Tabelas 14 a 21 relacionam as reclassificações e recomendações resultantes de cada iteração (i) da simulação realizada para as atividades de busca presentes na coleção de teste. Para auxiliar a compreensão das tabelas, os documentos foram identificados de acordo com a posição em que se encontravam no *ranking* original e os documentos relevantes foram destacados em negrito. A primeira linha de cada tabela apresenta a configuração das 10 primeiras posições do respectivo *ranking* original, logo antes de ocorrer a primeira iteração da simulação. As demais linhas se referem ao *ranking* resultante de cada iteração para o respectivo *feedback* capturado (f.c.). Os itens grifados representam os documentos já informados em algum *feedback* até a iteração considerada e os itens cercados por retângulos representam os documentos recomendados em uma determinada iteração. Apenas as recomendações obtidas pela segmentação autônoma do *ranking* foram representadas, não impedindo que as demais estratégias também possam ser observadas.

Tabela 14 – Classificação e recomendação resultantes para cada *feedback* capturado (f.c.) durante a simulação realizada para a atividade A1.

<i>i</i>	f.c.	<i>Ranking</i>									
•	•	1	2	3	4	5	6	7	8	9	10
1	2	<u>2</u>	46	53	8	34	3	95	57	17	80
2	46	46	53	<u>2</u>	8	22	67	80	26	85	7
3	53	46	53	<u>2</u>	8	22	67	80	85	26	7
4	8	46	53	<u>8</u>	<u>2</u>	22	67	80	85	26	47

Tabela 15 – Classificação e recomendação resultantes para cada *feedback* capturado (f.c.) durante a simulação realizada para a atividade A2.

<i>i</i>	f.c.	<i>Ranking</i>									
•	•	1	2	3	4	5	6	7	8	9	10
1	1	<u>1</u>	15	27	26	77	33	55	29	60	12
2	15	<u>1</u>	15	27	33	29	12	60	47	26	77
3	27	<u>1</u>	15	27	29	60	12	47	26	33	77
4	29	<u>1</u>	27	29	<u>15</u>	60	47	12	26	37	62

Tabela 16 – Classificação e recomendação resultantes para cada *feedback* capturado (f.c.) durante a simulação realizada para a atividade A3a.

<i>i</i>	f.c.	<i>Ranking</i>									
•	•	1	2	3	4	5	6	7	8	9	10
1	8	<u>8</u>	33	<u>2</u>	15	<u>9</u>	27	100	11	26	28
2	33	<u>8</u>	33	<u>9</u>	26	<u>2</u>	15	100	27	77	11
3	9	33	<u>9</u>	<u>8</u>	76	26	15	<u>2</u>	77	69	100
4	76	<u>9</u>	76	33	<u>8</u>	26	15	77	2	69	100

Tabela 17 – Classificação e recomendação resultantes para cada *feedback* capturado (f.c.) durante a simulação realizada para a atividade A3b.

<i>i</i>	f.c.	<i>Ranking</i>									
•	•	1	2	3	4	5	6	7	8	9	10
1	3	<u>3</u>	4	57	36	93	10	56	34	54	94
2	4	<u>4</u>	<u>3</u>	57	93	10	36	49	34	12	72
3	57	<u>4</u>	57	<u>3</u>	10	93	36	34	45	1	81
4	10	<u>4</u>	57	10	<u>3</u>	93	36	34	27	45	12

Tabela 18 – Classificação e recomendação resultantes para cada *feedback* capturado (f.c.) durante a simulação realizada para a atividade A4a.

<i>i</i>	f.c.	<i>Ranking</i>									
•	•	1	2	3	4	5	6	7	8	9	10
1	2	<u>2</u>	20	80	51	5	74	76	42	61	96
2	20	20	<u>2</u>	80	51	76	5	61	42	74	96
3	80	20	<u>2</u>	80	51	76	61	42	96	5	74
4	51	20	<u>2</u>	80	51	76	61	96	42	74	5

Tabela 19 – Classificação e recomendação resultantes para cada *feedback* capturado (f.c.) durante a simulação realizada para a atividade A4b.

<i>i</i>	f.c.	<i>Ranking</i>									
•	•	1	2	3	4	5	6	7	8	9	10
1	1	<u>1</u>	8	20	14	27	54	82	72	7	9
2	8	<u>1</u>	8	18	82	72	20	7	17	14	27
3	18	8	18	<u>1</u>	72	82	17	20	10	43	7
4	72	8	18	72	<u>1</u>	43	17	82	10	22	20

Tabela 20 – Classificação e recomendação resultantes para cada *feedback* capturado (f.c.) durante a simulação realizada para a atividade A5.

<i>i</i>	f.c.	<i>Ranking</i>									
•	•	1	2	3	4	5	6	7	8	9	10
1	2	<u>2</u>	43	97	5	17	6	19	18	82	59
2	43	43	<u>2</u>	97	6	5	17	10	19	24	82
3	97	<u>2</u>	43	97	17	6	5	24	19	10	81
4	17	<u>2</u>	97	17	43	6	5	81	19	94	24

Tabela 21 – Classificação e recomendação resultantes para cada *feedback* capturado (f.c.) durante a simulação realizada para a atividade A6.

<i>i</i>	f.c.	<i>Ranking</i>									
•	•	1	2	3	4	5	6	7	8	9	10
1	1	<u>1</u>	75	37	7	32	17	3	8	11	41
2	75	<u>1</u>	75	37	7	32	17	3	8	11	41
3	37	<u>1</u>	75	37	7	32	17	3	8	11	86
4	7	<u>1</u>	37	7	75	32	17	3	8	11	86