

Systematic Literature Review Supported by Information Retrieval Techniques: A Case Study

Ramon Abilio

Diretoria de Gestão de Tecnologia da Informação,
Universidade Federal de Lavras, Lavras, MG, Brasil
ramon.abilio@dgti.ufla.br

Claudiane Oliveira

Departamento de Ciência da Computação, Universidade
Federal de Lavras, Lavras, MG, Brasil
claudiane@posgrad.ufla.br

Gustavo Vale

Departamento de Ciência da Computação, Universidade
Federal de Minas Gerais, Belo Horizonte, MG, Brasil
gustavo@dcc.ufmg.br

Flávio Morais

Diretoria de Gestão de Tecnologia da Informação,
Universidade Federal de Lavras, Lavras, MG, Brasil
flavio@dgti.ufla.br

Denilson Pereira

Departamento de Ciência da Computação, Universidade
Federal de Lavras, Lavras, MG, Brasil
denilsonpereira@dcc.ufla.br

Heitor Costa

Departamento de Ciência da Computação, Universidade
Federal de Lavras, Lavras, MG, Brasil
heitor@dcc.ufla.br

Abstract— Systematic Literature Review (SLR) is a means to synthesize relevant and high quality studies related to a specific topic or research questions. In general, a SLR has three phases, and the first one is the Primary Selection in which the selection of studies is usually performed manually reading title, abstract and keywords of each study. The number of published scientific studies has grown, increasing effort in carrying out this review. In this paper, we proposed two strategies to rank studies in decreasing order of importance, for a SLR, regarding the terms in the search string. These strategies are based on the Information Retrieval technique Vector Model. We implemented those strategies and conducted a case study to evaluate their applicability. As results, the second strategy presents 50% of precision on a recall of 80%. Among the contributions of this study, two strategies to rank relevant documents in a SLR, regarding the search string, were proposed and analyzed.

Keywords— Systematic Literature Review; Information Retrieval; Vector Model; Primary Selection

I. INTRODUÇÃO

Utilizando-se a técnica Revisão Sistemática de Literatura (RSL), pode-se avaliar e interpretar trabalhos relevantes e disponíveis na literatura para uma questão particular de pesquisa, uma área temática ou um fenômeno de interesse [1][2][3][4][8]. Essa técnica está baseada na execução de três fases [3]: i) Planejamento; ii) Execução; e iii) Análise dos Resultados. Essas fases estão subdivididas em estágios, por exemplo, a fase de Planejamento possui o estágio de Desenvolvimento do Protocolo a ser utilizado na RSL, e a fase Execução é composta pelos estágios Seleção Primária e Seleção Secundária de Estudos.

Na fase de Planejamento de uma RSL, são definidos termos (palavras e/ou expressões) relacionados ao tópico ou questão de pesquisa, juntamente com seus sinônimos, que serão utilizados na busca por trabalhos relevantes à RSL. Um termo é relacionado a seus sinônimos com o operador lógico OR, formando um grupo de termos de busca. Cada grupo é conectado pelo operador lógico AND formando uma string de busca. Essa string é utilizada em máquinas de busca online com o objetivo de recuperar trabalhos que contenham pelo menos um termo de cada grupo.

Na Seleção Primária, a string de busca é utilizada em máquinas de busca, os resultados são exportados e organizados de forma a permitir a exclusão de duplicatas e materiais que não são artigos, como: índices e resumos de anais. Após isso, os estudos são selecionados manualmente mediante a leitura de partes ou seções do texto, tais como, título, resumo e palavras-chave. Na Seleção Secundária, os estudos selecionados na Seleção Primária são lidos na íntegra. Em ambas as seleções, os critérios de inclusão e de exclusão, definidos na fase Planejamento, são considerados [3]. No resultado da Seleção Secundária, estão os estudos considerados relevantes no contexto da pesquisa. Um problema encontrado na Seleção Primária é o esforço necessário para realizar a atividade, pois a quantidade de estudos tem crescido [4] e o objetivo de reduzir o viés na seleção desses estudos requer que, pelo menos, dois pesquisadores façam a leitura do título, do resumo, das palavras-chave e, quando necessário, do texto completo [3].

Com o intuito de minimizar esse esforço e auxiliar os pesquisadores na Seleção Primária, foram realizadas propostas para ordenar documentos de acordo com sua relevância,

seleção automatizada e validação dos estudos selecionados [4][6][7][12][13]. Neste trabalho, é proposta a utilização de técnicas de Recuperação de Informação para ordenar documentos de acordo com sua relevância, considerando a string de busca utilizada na RSL. Na Recuperação de Informação, o principal foco é fornecer aos usuários fácil acesso à informação de seu interesse. Modelos de Recuperação de Informação são criados com o objetivo de produzir uma função de ranqueamento que atribui pontos aos documentos com relação a uma dada consulta [11]. Um dos modelos clássicos da Recuperação de Informação é o Modelo Vetorial, um modelo algébrico que representa documentos e consultas como vetores em um espaço t-dimensional [11].

Neste trabalho, os documentos contendo estudos científicos foram ranqueados mediante a análise do título, do resumo e das palavras-chave, com base nos termos que compõem a string de busca. Foram implementadas e avaliadas duas estratégias para ranqueamento. O estudo de caso para avaliação da aplicabilidade da técnica, no contexto de revisões sistemáticas, foi realizado em referências de uma Revisão Sistemática de Literatura cujos resultados foram publicados em 2012 [5]. As principais contribuições deste trabalho são:

- Proposta de duas estratégias, utilizando o Modelo Vetorial, para ranqueamento das referências utilizadas em uma Revisão Sistemática de Literatura de acordo com sua relevância em relação aos termos utilizados na string de busca;
- Possibilidade de utilizar as estratégias propostas para avaliação da qualidade da string de busca;
- Identificação de comportamento não esperado em uma máquina de busca em relação à consulta booleana, pois publicações podem ser retornadas sem obedecer a união dos termos de busca com o conector AND.

O restante deste artigo está organizado da seguinte forma. Alguns trabalhos relacionados são resumidamente apresentados na Seção 2. Uma contextualização sobre Revisão Sistemática da Literatura e Recuperação de Informação é apresentada na Seção 3. As estratégias propostas para o ranqueamento de estudos científicos são descritas na Seção 4. Um estudo de caso é apresentado na Seção 5. Os resultados são discutidos na Seção 6. Conclusões e sugestões de trabalhos futuros são discutidas na Seção 7.

II. TRABALHOS RELACIONADOS

Com o objetivo de reduzir o esforço na realização de uma Revisão Sistemática de Literatura no contexto de Engenharia de Software baseada em Evidências [1], uma ferramenta foi proposta para atuar como interface para realização de buscas em bases de dados, tais como, IEEEExplore¹, ACM Digital Library², Scopus³ e Science Direct⁴. Para obter estudos relevantes, foi proposta a utilização de técnicas de Recuperação de Informação: extração automática de textos baseada em

técnicas de processamento de linguagem natural e agrupamento de estudos similares utilizando as técnicas de agrupamento k-means ou k-NN. Para avaliar as técnicas propostas, foi sugerida a utilização das medidas revocação (recall) e precisão (precision). A ferramenta não foi desenvolvida, a sua implementação foi mencionada como trabalhos futuros e a avaliação dos resultados foi realizada em uma coleção de estudos de teste e usuários da área de Engenharia de Software[4].

Em outro estudo [6], as técnicas Linked Data, Text Mining e Naive Bayes foram empregadas em um processo iterativo e supervisionado utilizado na Seleção Secundária para reduzir o trabalho manual e o viés de subjetividade. A proposta consiste em estender e enriquecer o processo básico da Seleção Primária utilizando tecnologias existentes na área de Web Semântica e Text Mining no contexto da técnica Linked Data, que explora a Web como plataforma de integração de dados e informações. Para avaliar a proposta, um estudo de caso foi realizado utilizando informações de uma Revisão Sistemática de Literatura publicada em 2007. Esse estudo de caso consistiu em simular uma parte do trabalho manual realizado na revisão com a busca e a Seleção Primária. Os estudos resultantes da Seleção Primária foram utilizados para iniciar o protótipo e executar a técnica proposta. Os resultados indicaram que a técnica utilizada pode reduzir o esforço em 20% com uma revocação de 100%.

Para auxiliar na verificação dos resultados da Seleção Secundária, foi utilizada a técnica Visual Text Mining (VTM) [7]. Essa técnica auxilia em tarefas que envolvam interpretação de grande quantidade de dados textuais e integração entre as técnicas Text Mining e Information Visualization. A proposta desse estudo é alterar a fase Execução, acrescentando duas técnicas de VTM para auxiliar na revisão dos estudos selecionados (content map e citation map). Estudos de caso foram realizados para verificar a aplicabilidade da técnica e resultados indicaram que VTM pode auxiliar na Seleção Primária acelerando o processo e aumentando a confiabilidade dos resultados.

Em outro estudo [12], uma triagem semiautomatizada foi realizada na seleção de artigos. Nessa triagem, foram utilizadas Support Vector Machines (SVMs), podendo ser construídas em diferentes espaços, por exemplo, resumo e título ou texto completo. Semiautomatizar o processo de triagem é difícil por causa do desequilíbrio de artigos irrelevantes sobre os relevantes. Por isso, uma estratégia de aprendizagem ativa foi desenvolvida para conjuntos de dados desbalanceados. A triagem foi realizada em três Revisões Sistemáticas de Literatura. Em duas delas, o algoritmo foi capaz de reduzir a quantidade de referências selecionadas manualmente em 50% e, no terceiro estudo, em cerca de 40%. Nesses estudos, não houve comprometimento da qualidade e da abrangência.

Com objetivo de reduzir esforço na leitura de referências em uma Revisão Sistemática de Literatura na área da saúde, foi realizado um ranqueamento de referências relevantes dividido em duas etapas [13]: i) recuperação de texto; e ii) re-ranqueamento via classificação de texto. Na primeira etapa, restrições são adicionadas utilizando vários controles, por exemplo, o controle .mp procura determinado termo no título,

¹ <http://ieeexplore.ieee.org/>

² <http://dl.acm.org/>

³ <http://www.scopus.com/>

⁴ <http://www.sciencedirect.com/>

no resumo e no cabeçalho MeSH. Na segunda etapa, são obtidas informações do passo anterior e os artigos são ordenados de maneira hierárquica usando Support Vector Regression (SVR). Nesse trabalho, foi utilizada a ferramenta de aprendizado de máquina Weka⁵ para aplicar a técnica proposta. Essa técnica foi replicada em 17 Revisões Sistemáticas de Literatura e, como resultado, os artigos considerados relevantes estavam ranqueados entre os 30% primeiros artigos do montante de cada revisão. Esses resultados comprovaram a eficiência da técnica e, em alguns casos, os resultados atingiram índices melhores.

O objetivo, neste estudo, é aplicar técnicas de Recuperação de Informação no estágio Seleção Primária, assim como em outros três estudos [4][12][13], enquanto, em outros dois estudos, o foco foi auxiliar no estágio Seleção Secundária [6] e validar os resultados da Seleção Secundária [7]. Em relação aos estudos com foco na Seleção Primária, em um deles [4], foi proposta uma ferramenta na qual estariam presentes técnicas de Recuperação de Informação, mas não foi desenvolvida e avaliada. Neste trabalho, foi desenvolvido um protótipo com o qual pôde-se avaliar a aplicabilidade da proposta. Em outros dois estudos, as técnicas SVM [12] e SVR [13] foram utilizadas, enquanto, neste trabalho, foi utilizado o Modelo Vetorial. Neste trabalho, foram utilizadas duas estratégias baseadas na similaridade do cosseno do Modelo Vetorial. Como entrada, essas estratégias utilizam os dados usados na elaboração do protocolo da Revisão Sistemática de Literatura e a coleção de referências bibliográficas geradas pela execução da string de busca, ou seja, a sobrecarga para os pesquisadores em relação à aplicação das técnicas é mínima.

III. CONTEXTUALIZAÇÃO

A técnica de Revisão Sistemática de Literatura, conceitos de Recuperação de Informação e de Modelo Vetorial são brevemente apresentados nesta seção.

A. Revisão Sistemática da Literatura

Revisão Sistemática de Literatura (RSL) é um componente das práticas baseadas em evidências [8][9] e faz parte do que se chama de Estudos Secundários em uma pesquisa sistemática. Ao empregar RSL, a pesquisa, a seleção, a análise e a organização de documentos são facilitadas por causa dos passos e dos critérios pré-definidos [4]. Com isso, consegue-se organizar a aplicação da técnica, possibilitando encontrar conteúdo relevante sobre um tópico pesquisado em um repositório com grande quantidade de informações. O ponto de partida para aplicação da RSL consiste em identificar, formular e responder questões de pesquisa, após a obtenção e a análise dos dados [3]. A sua condução está baseada em 3 fases [4][9] (Fig. 1):

- Planejamento. O motivo da realização da Revisão Sistemática de Literatura é estabelecido. Seus tópicos são: i) Descrição da Pesquisa: motivações e objetivos para a pesquisa; ii) Definição de Questões de Pesquisa: questões relacionadas às motivações da pesquisa; iii) Desenvolvimento do Protocolo: protocolo a ser aplicado

às buscas; e iv) Avaliação do Protocolo: avaliação da aplicação da Revisão Sistemática de Literatura;

- Execução. A investigação em fontes definidas na fase anterior é realizada. O estudo e a classificação dos trabalhos encontrados podem ser feitos, guiados pelos critérios de inclusão e de exclusão. Seus tópicos são: i) Obtenção das Pesquisas: busca nas fontes definidas, organizando os resultados por algum critério para facilitar as próximas etapas; ii) Seleção Primária: primeira seleção dos resultados obtidos (título, palavras-chave e resumo dos trabalhos são lidos para verificar atendimento dos critérios); iii) Seleção Secundária: segunda seleção realizada, cujo objetivo é eliminar resultados não apropriados. Os artigos são lidos e verificados quanto ao atendimento aos critérios; e iv) Organização dos Resultados: artigos selecionados são organizados para serem analisados;
- Análise dos Resultados. A coleta e a organização dos dados extraídos dos artigos selecionados. O resultado é analisado de maneira global, gerando melhor planejamento, caso necessário. Seus tópicos são: i) Coleta e Organização dos Dados: resultados obtidos na Seleção Secundária são lidos e interpretados. Os dados são extraídos e organizados em forma de um relatório, cujo conteúdo responde as questões de pesquisa definidas no protocolo; e ii) Avaliação dos Resultados: o relatório é revisado para buscar a publicação dos resultados.

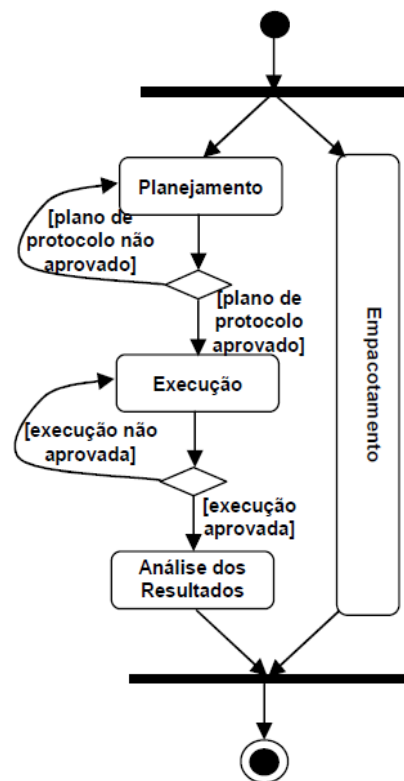


Fig. 1. Processo de Construção da Revisão Sistemática (Adaptada de [10])

⁵ <http://www.cs.waikato.ac.nz/ml/weka>

B. Recuperação de Informação

Na Recuperação de Informação, são tratados a representação, o armazenamento, a organização e o acesso a itens de informação, tais como, documentos, páginas Web, catálogos on-line, registros estruturados e semiestruturados e objetos multimídia [11]. No contexto deste trabalho, o objetivo é indexar e pesquisar por documentos contendo estudos científicos em uma coleção. Modelagem em Recuperação de Informação é um processo complexo que produz uma função de ranqueamento para atribuir pontos aos documentos com relação a uma dada consulta [11]. Um dos modelos clássicos da Recuperação de Informação é o Modelo Vetorial, um modelo algébrico que representa documentos e consultas como vetores em um espaço t-dimensional, sendo t a quantidade de termos (palavras ou expressões) distintos na coleção [11].

Seja $D = \{d_1, d_2, \dots, d_n\}$ uma coleção de N documentos. Seja $K = \{k_1, k_2, \dots, k_t\}$ o conjunto das t palavras (ou expressões) distintas que aparecem em documentos de D. Para cada par (k_i, d_j) , $k_i \in K$ e $d_j \in D$, é associado um peso $w_{i,j} \geq 0$. Um documento d_j é representado como um vetor de pesos $\vec{d}_j = \{w_{1,j}, w_{2,j}, \dots, w_{t,j}\}$. Cada $w_{i,j}$ representa a importância da palavra (ou expressão) k_i para descrever o conteúdo semântico do documento d_j . As consultas são representadas como vetores t-dimensionais. Um esquema utilizado para cálculo dos pesos, conhecido como TF-IDF (term-frequency x inverse document frequency), considera a frequência do termo no documento (TF) e a raridade do termo na coleção (IDF). O grau de similaridade entre uma consulta q e um documento d é avaliado como a correlação entre os vetores \vec{d}_j e \vec{q} . Essa correlação pode ser quantificada pelo cosseno do ângulo entre esses dois vetores (1). Essa medida de similaridade é conhecida como Similaridade do Cosseno [11].

$$\text{sim}(d, q) = \frac{\sum_{i=1}^t w_{i,d} \cdot w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,d}^2} \cdot \sqrt{\sum_{i=1}^t w_{i,q}^2}} \quad (1)$$

Após obter o resultado pela função de ranqueamento, é importante avaliar a qualidade do resultado que consiste em associar medidas quantitativas ao resultado obtido pelo sistema como resposta a um conjunto de consultas do usuário. Normalmente, essa avaliação é obtida comparando resultados do sistema de Recuperação de Informação com resultados propostos por especialistas que avaliaram a coleção de referências utilizada [11]. Duas medidas clássicas em avaliação de sistema de Recuperação de Informação são a precisão e a revocação [11]. Formalmente, sejam: q uma consulta proveniente de uma coleção de referência, R o conjunto de documentos julgados como relevantes para q e |R| a quantidade de documentos em R. Ao submeter q, obtém-se um conjunto de resultados A (|A| documentos). Considerando $|R \cap A|$ a quantidade de documentos na interseção entre R e A, pode-se definir a precisão (2), fração dos documentos recuperados relevantes, e a revocação (3), fração dos documentos relevantes recuperados.

$$\text{Precisão} = \frac{|R \cap A|}{|A|} \quad (2)$$

$$\text{Revocação} = \frac{|R \cap A|}{|R|} \quad (3)$$

IV. ESTRATÉGIAS DE RANQUEAMENTO

Durante a Seleção Primária, os resultados das buscas são exportados, reunidos em um único repositório e duplicatas e não artigos são removidos. Portanto, perde-se a ordenação por relevância realizada por máquinas de busca, como a Science Direct. Com o objetivo de ordenar novamente as referências de acordo com sua relevância considerando-se a string de busca, foram propostas duas estratégias de ranqueamento

Em ambas as estratégias, foi utilizado o Modelo Vetorial e os pesos dos termos nos documentos foram calculados considerando-se apenas informações presentes no documento de entrada, ou seja, não foram consideradas informações globais, como é o caso do IDF, que exige dados de uma coleção de documentos para ser calculado. Desse modo, a informação considerada para definir o peso de um termo em um documento é a sua frequência no documento. Portanto, quanto mais vezes os termos da consulta aparecerem em um documento, mais relevante esse documento é para a consulta. Para obter essa relevância, calcula-se o peso $w_{i,j}$ de um termo i no documento j (4), sendo $f_{i,j}$ a frequência do termo i no documento j, e $\max f_{l,j}$ a frequência do termo l, em que l é o termo mais frequente no documento j. Em ambas as estratégias, foram considerados os textos dos documentos contidos no título, no resumo e nas palavras-chave da referência.

$$w_{i,j} = \frac{f_{i,j}}{\max f_{l,j}} \quad (4)$$

A. Estratégia 1

Na Estratégia 1, foi utilizada a forma tradicional do Modelo Vetorial, que considera a contribuição parcial de cada termo da consulta. Os termos da consulta, presentes na string de busca, são tratados da mesma forma, independentemente se são ou não sinônimos, e desconhecendo termos duplicados e os conectores OR e AND. Para verificar o quão similar um documento é em relação à consulta, calcula-se o grau de similaridade (1).

B. Estratégia 2

A ideia na Estratégia 2 é definir uma função de ranqueamento que simula a expressão booleana da string de busca. Os termos de busca foram organizados em grupos, que agrupam um termo e seus sinônimos utilizando o conector OR. Os grupos são, então, conectados pelo conector AND. Para verificar a similaridade de um documento em relação à consulta, calcula-se a similaridade (5), considerando g_i o grupo i de termos, $\text{sim}(g_i)$ a máxima similaridade obtida no grupo (6), k_p o termo da posição p entre os n termos do grupo i, e $\text{sim}(d, k_p)$ a similaridade obtida por Similaridade do Cosseno

(1). Portanto, as referências mais relevantes devem possuir pelo menos um termo de cada grupo.

$$\text{sim}(d, q) = \sum_{i=1}^m \text{sim}(g_i) \quad (5)$$

$$\text{sim}(g_i) = \max \sum_{p=1}^n \text{sim}(d, k_p) \quad (6)$$

V. ESTUDO DE CASO

Nesta seção, é apresentado um estudo de caso realizado para avaliar as estratégias de ranqueamento propostas.

A. Configuração Geral

Um protótipo foi implementado em Java para avaliação das estratégias de ranqueamento dos documentos. Esse protótipo recebe, como entrada, um arquivo texto contendo referências bibliográficas no formato BibTex (Fig. 2) e um arquivo texto contendo os termos que compõem a string de busca (Fig. 3). As saídas desse protótipo são referências bibliográficas ranqueadas e o cálculo da precisão e da revocação.

Na Fig. 2, é apresentado um exemplo de uma referência, no formato BibTex, para um artigo (@ARTICLE). Entre os campos dessa referência, estão: author (autor), title (título), abstract (resumo) e keywords (palavras-chave). Para o estudo de caso, foram indexados os atributos título, resumo e palavras-chave de cada referência, extraídos do arquivo que contém as referências no formato BibTex. Esses atributos foram concatenados, formando uma única string, representado um documento.

```
@ARTICLE{Alderson1998,
author = {Alderson, A. and Hull, M.E.C. and Jackson, K. and Griffiths, L.E.},
title = {Method engineering for industrial real-time and embedded systems},
journal = {Information and Software Technology},
year = {1998},
volume = {40},
pages = {443--454},
number = {8},
month = aug,
abstract = {Real-time and embedded systems have proved troublesome to produce,
with all the difficulties of the other kinds of software-based systems
together with a number of specific additional problems. This paper
reports on a systematic approach to engineering methods and processes,
including technical management, for the complete lifecycle of real-time
and embedded systems in an industrial context. An information framework
was developed as the base for defining appropriate methods and processes
supported by tools. Existing methods and tools were utilised as the
development base. MetaCASE technology enabled iterative enhancement
under the guidance of case studies.},
doi = {10.1016/S0950-5849(98)00073-1},
issn = {0950-5849},
keywords = {Real-time and embedded systems, Development method, System development,
MetaCASE, CASE},
owner = {elsevier},
timestamp = {2012.05.03},
url = {http://www.sciencedirect.com/science/article/pii/S0950584998000731}
}
```

Fig. 2. Exemplo de Referência no Formato BibTex

O arquivo com os termos (palavras e expressões) que compõem a string de busca tem o formato apresentado na Fig. 3. Cada linha do arquivo contém os termos e seus sinônimos, separados por vírgula. As expressões devem estar entre aspas. Por exemplo, a palavra Metric possui como sinônimos as palavras: metrics, measure e measurement.

Software, application, system, product, "software product" Metric, metrics, measure, measurement Reference value, "reference values" Quality, "internal quality" Feature Oriented, "feature-oriented"

Fig. 3. Exemplo de Arquivo de Entrada com Termos da String de Busca

B. Experimento e Resultado

No estudo de caso, uma Revisão Sistemática de Literatura (RSL), publicada em 2012 [5], foi utilizada para avaliação das estratégias propostas. Os estudos considerados relevantes, na RSL, são referenciados pelos números de 1 a 13, e os títulos e referências completas estão apresentados no Anexo. As palavras, expressões e sinônimos utilizados foram:

- Software - application, system, program, product, "software product", "software system";
- Metric - metrics, measure, measurement;
- Reference value - "reference values";
- Quality - "internal quality";
- Feature Oriented - "feature-oriented";
- Aspect Oriented - "aspect-oriented";
- Product line - "product family", "product-family", "family based", "family-based";
- Modularization - modularity, modularisation;
- Maintenance - maintainability;
- Concern - aspect;
- Crosscutting - "cross-cutting".

Com os termos e sinônimos, foi montada e utilizada a seguinte string de busca:

(software OR application OR system OR program OR product OR "software product" OR "software system") AND (metric OR metrics OR measure OR measurement) AND (quality OR "internal quality") AND (modularization OR modularisation OR modularity) AND (maintenance OR maintainability) AND (("feature oriented" OR "feature-oriented" OR "product line" OR "product family" OR "family based" OR "product-family" OR "family-based") OR ("reference value" OR "reference values") AND ("aspect oriented" OR "aspect-oriented") AND (concern OR aspect AND (crosscutting OR "cross-cutting")))

Na RSL, essa string foi utilizada nas máquinas de busca IEEEExplore, Scopus e Science Direct. No total, obteve-se 672 referências. Essas referências foram importadas no software JabRef⁶ gerando um único arquivo com as referências no formato BibTex. Do total de referências, 36 foram excluídas por serem repetidas ou por não serem referências para artigos, restando 636, e 13 foram selecionadas manualmente na Seleção Primária (Anexo).

O procedimento do estudo de caso consistiu em: i) Obter o arquivo BibTex sem duplicatas e não artigos junto aos autores da RSL; ii) Realizar o ranqueamento das referências utilizando cada uma das estratégias; iii) Analisar o resultados verificando: revocação e precisão; posições das referências no ranking; e quantidade de termos da string encontrados em cada referência. No ranqueamento das referências, foi identificada a ausência do campo resumo em quatro referências e do campo palavras-chave em 81 referências. Isso ocorreu quando os resultados foram exportados pelas máquinas de busca. Optou-se por não indexar as quatro referências sem resumo, totalizando 632 referências, pois em uma análise manual elas também seriam excluídas.

1) Resultados da Estratégia 1

Como resultado do ranqueamento das 632 referências, tem-se que 600 foram recuperadas, pois possuíam ocorrências dos termos da string de busca. Na TABELA I, é apresentada a posição no ranking das 13 referências relevantes em relação ao total de referências. Por exemplo, a referência 12 ficou na primeira posição do ranking e a referência 5 ficou na posição 113. Além disso, são apresentados os valores da precisão interpolada nos 11 níveis de revocação padrão (TABELA II). Em cada linha, é apresentado o valor da precisão interpolada quando x% das referências relevantes são recuperados [11]. Por exemplo, para revocação de 60%, a precisão é 28,6%.

TABELA I. ESTRATÉGIA 1 (POSIÇÃO E REFERÊNCIA)

Posição	Referência
1	12
2	3
6	11
11	4
16	10
22	2
27	9
28	7
51	13
56	6
64	8
106	1
113	5

TABELA II. ESTRATÉGIA 1 (REVOCAÇÃO E PRECISÃO)

Revocação	Precisão
0%	100,0%
10%	100,0%
20%	50,0%
30%	36,4%
40%	28,6%
50%	28,6%
60%	28,6%
70%	17,9%
80%	17,2%
90%	11,5%
100%	11,5%

No arquivo de entrada, que contém os termos da string, há 44 palavras/expressões. Verificando a quantidade de termos e suas ocorrências nas referências ranqueadas (TABELA III), observou-se que, na referência da posição 1, há 13 termos totalizando 74 ocorrências e, na referência da posição 113, a última posição contendo referência relevante, há 5 termos com total de 17 ocorrências. Os termos com maior frequência na referência da posição 1 são: Software (18), maintainability (11) e crosscutting (11) e, na referência da posição 113, são: software (9), quality (3), metrics (2) e feature (2).

TABELA III. CONTAGEM DE TERMOS - ESTRATÉGIA 1

#	Termos	Referência/Posição	
		12 / 1	5 / 113
1	software	18	9
2	application	0	0
3	system	1	0
4	program	0	0
5	product	5	0
6	software product	2	0
7	software system	0	0
8	metric	0	0
9	metrics	5	2
10	measure	0	0
11	measurement	0	0
12	reference	0	0
13	value	0	0
14	values	0	0
15	reference value	0	0
16	reference values	0	0
17	quality	5	3
18	internal	0	0
19	internal quality	0	0
20	feature	0	2
21	oriented	0	0
22	feature oriented	0	0
23	feature-oriented	0	0

⁶ <http://jabref.sourceforge.net/>

TABELA III. CONTAGEM DE TERMOS - ESTRATÉGIA 1 (CONT)

#	Termos	Referência/Posição	
		12 / 1	5 / 113
24	aspect	0	0
25	aspect oriented	0	0
26	aspect-oriented	0	0
27	line	2	0
28	family	0	0
29	product line	2	0
30	product family	0	0
31	product-family	0	0
32	family based	0	0
33	family-based	0	0
34	based	0	1
35	modularization	0	0
36	modularity	2	0
37	modularisation	0	0
38	maintenance	3	0
39	maintainability	11	0
40	concern	7	0
41	crosscutting	11	0
42	cross-cutting	0	0
43	cross	0	0
44	cutting	0	0
Total		74	17

2) Resultados da Estratégia 2

Foram recuperadas 600 referências que possuíam ocorrências dos termos da string de busca. Na TABELA IV. , é apresentada a posição das 13 referências relevantes em relação ao total de referências. Por exemplo, as referências 3 e 1 ficaram nas posições 1 e 97, respectivamente. Além disso, são apresentados os valores da precisão interpolada nos 11 níveis de revocação padrão (TABELA V.). Por exemplo, para revocação de 80%, a precisão é 50%. Isso significa que metade das referências recuperadas é relevante e a outra metade não.

TABELA IV. ESTRATÉGIA 2 (POSIÇÃO E REFERÊNCIA)

Posição	Referência
1	3
2	12
5	2
8	4
12	11
13	10
15	9
16	7
61	6
64	13
67	8
93	5
97	1

TABELA V. ESTRATÉGIA 2 (REVOCAÇÃO E PRECISÃO)

Revocação	Precisão
0%	100,0%
10%	100,0%
20%	100,0%
30%	60,0%
40%	50,0%
50%	50,0%
60%	50,0%
70%	50,0%
80%	50,0%
90%	14,8%
100%	13,4%

Na TABELA VI. , estão representados os grupos de termos da consulta utilizados na Estratégia 2. Por exemplo, o grupo 4 é formado pela palavra quality e pela expressão “internal quality”. Considerando os 11 grupos, há termos de 8 grupos na referência da posição 1 e 7 na referência da posição 97 (TABELA VII.). Na referência da posição 1, há 49 ocorrências dos termos, enquanto, na referência da posição 97, há 28 ocorrências.

TABELA VI. GRUPOS DE TERMOS DA CONSULTA

#	Grupo
1	software, application, system, program, product, “software product”, “software system”
2	metric, metrics, measure, measurement
3	reference value, “reference values”
4	quality, “internal quality”
5	feature Oriented, “feature-oriented”
6	aspect Oriented, “aspect-oriented”
7	product line, “product family”, “product-family”, “family based”, “family-based”
8	modularization, modularity, modularisation
9	maintenance, maintainability
10	concern, aspect
11	crosscutting, “cross-cutting”

TABELA VII. CONTAGEM DE TERMOS - ESTRATÉGIA 2

Grupo	Termos	Referência			
		#1	Maior ocorrência	#97	Maior ocorrência
1	software	5	5	3	3
	application	0		0	
	system	0		1	
	program	0		0	
	product	0		0	
	software product	0		0	
	software system	0		0	

TABELA VII. CONTAGEM DE TERMOS - ESTRATÉGIA 2 (CONT)

Grupo	Termos	Referência			
		#1	Maior ocorrência	#97	Maior ocorrência
2	metric	0	7	0	0
	metrics	7		0	
	measure	0		0	
	measurement	0		0	
3	reference	0	1	0	0
	value	1		0	
	values	1		0	
	reference value	0		0	
	reference values	0		0	
4	quality	0	0	1	1
	internal	0		0	
	internal quality	0		0	
5	feature	0	6	0	4
	oriented	6		4	
	feature oriented	0		0	
	feature-oriented	0		0	
6	aspect	7	7	3	4
	oriented	6		4	
	aspect oriented	0		0	
	aspect-oriented	6		3	
7	product	0	0	0	3
	line	0		0	
	family	0		0	
	product line	0		0	
	product family	0		0	
	product-family	0		0	
	family based	0		0	
	family-based	0		0	
based	0	3			
8	modularization	0	1	1	2
	modularity	0		2	
	modularisation	1		0	
9	maintenance	0	0	0	0
	maintainability	0		0	
10	concern	0	7	0	3
	aspect	7		3	
11	crosscutting	2	2	0	0
	cross-cutting	0		0	
	cross	0		0	
	cutting	0		0	
	Total	49		28	

VI. DISCUSSÃO DOS RESULTADOS

Considerando a Estratégia 1, pode-se notar que as referências 1, 5 e 8 estão nas posições mais baixas do ranking: 106, 113 e 64, respectivamente. As 13 referências relevantes estão posicionadas entre as primeiras 113 referências do ranking. No estudo original, foram verificadas as 636 referências utilizando a seleção manual. Com o ranqueamento

gerado pela Estratégia 1, os pesquisadores poderiam encontrar os trabalhos relevantes nas primeiras 113 referências.

Com a Estratégia 2, os trabalhos relevantes ficaram entre os primeiros 15% das referências, pois a referência mais baixa no ranking ficou na posição 97. As referências 1, 5 e 8 ficaram nas posições mais baixas do ranking: 97, 93 e 67, respectivamente. Em relação à Estratégia 1, essas referências foram melhores classificadas, apesar de permanecerem nas últimas posições, e houve troca de posições entre as referências 1 e 5. Ao analisar a quantidade de termos comuns entre os documentos indexados e o arquivo com os termos da string de busca, pôde-se verificar que: i) na referência da posição 1 da Estratégia 1, a quantidade de termos é relativamente pequena, possuindo 30% dos termos da consulta, porém ela ficou na posição 1 por possuir o total de 74 ocorrências; ii) no máximo, 73% dos grupos possuem termos da string de busca; iii) 5% das referências não possuem qualquer termo da string de busca; e iv) em 13% das referências exportadas pelas máquinas de busca, não consta o campo palavras-chave.

A quantidade relativamente baixa de termos comuns entre a string de busca e as referências relevantes indica que 70% deles, que compõem a string de busca, podem ter interferido negativamente no resultado da busca, pois referências não relevantes podem ter aparecido nos resultados. Portanto, a utilização das estratégias propostas poderia auxiliar na escolha dos termos a serem utilizados na composição da string de busca e melhorar os resultados na busca por estudos na Seleção Primária.

A existência de referências que não possuem palavras no título, no resumo ou nas palavras-chave pode acontecer por causa da forma como as máquinas de busca realizam a consulta. A documentação das máquinas de busca Scopus e Science Direct indica que, ao informar uma expressão entre aspas, elas fazem busca por similaridade, ou seja, as máquinas retornam o mesmo resultado para as expressões “heart-attack” e “heart attack”, por exemplo. No entanto, a utilização de aspas na máquina de busca IEEEExplore indica uma busca exata, apesar de retornar o mesmo resultado para as expressões “solid-state” e “solid state”.

Essas máquinas de busca utilizam a técnica de stemming, ou seja, uma busca pelo termo metrics gera resultados com o termo metric, inclusive. A IEEEExplore realiza stemming inclusive entre inglês britânico e americano, mas a documentação dessa máquina de busca indica que alguns termos técnicos podem não estar no dicionário de stemming. Na IEEEExplore, essa técnica somente é aplicada se a busca não for exata, ou seja, se o termo da consulta não estiver entre aspas.

Outro recurso disponível nessas máquinas de busca é a possibilidade de indicar em quais campos o usuário deseja realizar a busca. Na IEEEExplore, existe uma opção de consulta chamada metadada que inclui resumo, termos de indexação, título do documento, título da publicação, autores, entre outros. Na publicação utilizada no estudo de caso, a busca na IEEEExplore foi realizada usando essa opção, o que pode ter gerado os 5% de referências que não possuem os termos da consulta nos campos título, resumo e palavras-chave.

Na Seleção Primária, é exigido esforço manual, além do esforço na exportação dos resultados da busca inicial. Após a execução da string de busca, o resultado deve ser exportado para ser analisado e algumas máquinas de busca não fornecem mecanismos de exportação adequados. Por exemplo, na ACM Digital Library, pode-se exportar dados, porém tem-se que exportar cada resultado individualmente. Na IEEEExplore, na Scopus e na Science Direct, pode-se exportar conjuntos de referências, mas, na IEEEExplore, os resultados podem ser exportados em blocos com tamanho máximo de 100 referências (no retorno de 1000 referências, a exportação deverá ser executada em 10 passos). Na Scopus e na Science Direct, o resultado pode ser exportado em um único passo, desde que o resultado tenha, no máximo, 1000 referências.

Além desses inconvenientes, algumas referências são exportadas faltando dados, por exemplo, resumo e palavras-chave. Pôde-se verificar que, ao executar a consulta, alguns resultados contêm os campos, mas a referência aparece incompleta ao ser exportada [5]. A ausência de algum campo pode afetar o ranqueamento, pois uma quantidade menor de termos é indexada, o que pode acarretar em uma posição baixa no ranking.

Na execução da Estratégia 2, foi obtido um resultado não esperado. Observou-se que, dos 11 grupos presentes no arquivo de consulta, 8 têm algum termo em comum com as referências retornadas pelas máquinas de busca. Pela forma como a string foi montada, deveria existir pelo menos um termo de cada grupo.

Na string de busca, existe o seguinte trecho:

```
(software OR application OR system OR program OR
product OR "software product" OR "software system") AND
(metric OR metrics OR measure OR measurement) AND
(quality OR "internal quality") AND (modularization OR
modularisation OR modularity)AND(maintenance OR
maintainability) AND...
```

Verifica-se, nesse trecho, que existem 5 grupos (termos unidos com o operador OR) separados pelo o operador AND, ou seja, o resultado deveria apresentar pelo menos um termo de cada grupo. Ao analisar a contagem de palavras, pôde-se perceber que as referências 3 e 1 (Estratégia 2) não atendem a essa restrição. Ambas as referências foram exportadas pela Science Direct. Fez-se uma simulação de consulta na Science Direct utilizando somente esse trecho da string e limitando a busca: i) nos campos: título, resumo e palavras-chave; ii) na área: Ciência da Computação; e iii) no período: 1990 a 2012. Essa ferramenta ordena o resultado por relevância⁷ e na primeira posição aparece a referência 12, assim como na Estratégia 1. Realizando a contagem de termos por grupo (Estratégia 2), verificou-se que as três primeiras referências não seguem a restrição da string de busca. Isso indica uma falha na busca booleana ou a utilização de uma estratégia não mencionada na documentação da máquina de busca, pois, se os

grupos foram unidos com AND, pelo menos, um termo de cada grupo deveria existir nos campos selecionados das publicações retornadas pela máquina de busca.

A geração de resultados que não atendem a string de busca provoca aumento no esforço na exportação dos resultados e na Seleção Primária. A seleção, normalmente, passa por leitura "aleatória", pois o resultado final da busca inicial poderá não estar ordenado por relevância. Um ranqueamento posterior das referências utilizando os mesmos termos da string de busca ou outros termos pode reduzir o esforço na Seleção Primária, pois a leitura será orientada pelo ranking, no qual as referências mais relevantes estarão no topo.

Uma questão, que deve ser considerada, é a subjetividade presente nos textos, pois o contexto pode indicar a relevância do estudo para uma questão de pesquisa e não somente a presença de um ou mais termos e seus sinônimos. Por exemplo, no campo resumo de uma referência, é mencionado que: i) a programação orientada a aspectos (POA) promete melhorar a qualidade de software provendo melhor modularização e separação de interesses; ii) foi realizada uma Revisão Sistemática de Literatura para encontrar estudos empíricos que apresentassem comparações entre abordagens POA e não-POA; e iii) os estudos analisados apontaram efeitos positivos ou negativos relacionados a performance, tamanho de código, modularidade e evolução. Logo, o contexto indica a utilização de medidas relacionadas à POA e utilizadas para avaliar a qualidade de software, pois foram pesquisados estudos empíricos.

A referência 1 foi selecionada pela subjetividade e foi recuperada no processo de ranqueamento da Estratégia 1 por possuir 5 termos da string de busca com total de 17 ocorrências. A referência 5 também foi selecionada pela subjetividade e ocupa a posição 97 na Estratégia 2 com 28 termos da consulta pertencentes a 7 grupos. No gráfico da Fig. 4, são apresentadas as curvas de revocação versus precisão para os 11 níveis de revocação padrão obtidas a partir dos resultados da Estratégia 1 e da Estratégia 2. A Estratégia 2 apresentou precisão maior ou igual à Estratégia 1 para os 11 níveis, sendo a mais indicada para este estudo de caso.

Nos níveis de revocação iguais a 90% e 100% os valores da precisão foram baixos para referências selecionadas como relevantes, em decorrência da seleção subjetiva. As posições que as referências relevantes ocupam em cada estratégia estão no gráfico da Fig. 5. Nessa análise, é mostrada a variação que as referências tiveram no ranking. Por exemplo, as referências 1 e 5, que se destacam em maior posição. Essas referências foram selecionadas analisando critérios subjetivos. Em contrapartida, as referências 2 e 13, com baixa representatividade, são referências no topo do ranking. Essas foram as referências com maior frequência dos termos da consulta.

⁷ A relevância, indicada na documentação da Science Direct, é calculada pela frequência relativa dos termos da consulta em cada publicação.

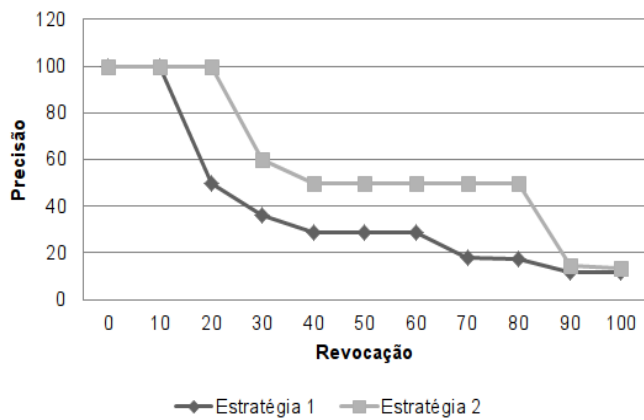


Fig. 4. Revocação x Precisão

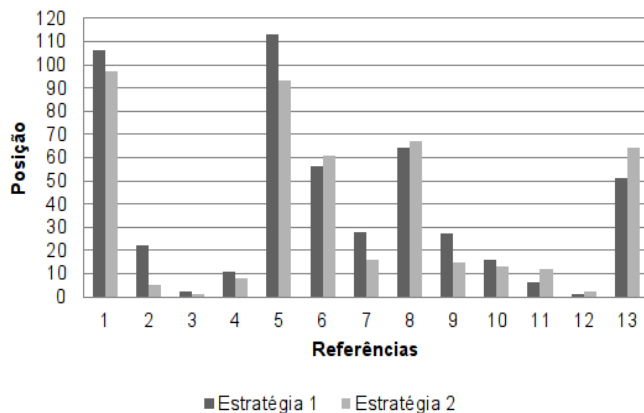


Fig. 5. Variação da posição de cada referência em cada estratégia

VII. CONCLUSÃO

A Revisão Sistemática de Literatura é uma das principais técnicas para síntese de estudos relevantes a um determinado tópico ou questão de pesquisa. Em sua primeira fase, denominada Seleção Primária, título, resumo e palavras-chave de cada referência retornada na busca inicial devem ser lidos com o objetivo de identificar, inicialmente, estudos relevantes para responder a questão de pesquisa. A Seleção Primária demanda esforço dos pesquisadores por causa do aumento da quantidade de trabalhos publicados. Com o objetivo de reduzir esse esforço, neste trabalho, foi proposta a utilização de técnicas de Recuperação de Informação para ranqueamento de documentos de acordo com sua relevância. Duas estratégias foram apresentadas: i) Estratégia 1 consiste em utilizar o Modelo Vetorial na forma tradicional; e ii) na Estratégia 2, foi proposta uma função de ranqueamento que simula a expressão booleana da string de busca.

A avaliação dessas estratégias foi realizada em um estudo de caso. Ambas as estratégias mostraram-se promissoras para redução do esforço manual, pois, no estudo de caso realizado, documentos relevantes foram posicionados entre 15% a 20% das referências do topo do ranking. As estratégias não indicam

um percentual de referências a serem lidas a priori, mas o ranqueamento possibilita a leitura de referências mais aderentes à string de busca elaborada. Algumas limitações deste trabalho incluem a necessidade de (i) avaliar outros estudos de caso para melhor confirmação dos resultados e (ii) considerar subjetividade dos textos analisados. Ainda assim, o estudo mostrou-se promissor, pois o objetivo é utilizar ferramenta proposta como auxílio no processo de seleção e não para substituir o trabalho humano.

Como trabalhos futuros, pretende-se transformar o protótipo desenvolvido em uma ferramenta disponível, via Web, para a comunidade científica. Além disso, pretende-se ampliar a funcionalidade da ferramenta para identificar artigos duplicados e para eliminar referências a materiais que não são artigos. Por fim, sugere-se avaliar o seu uso para análise do texto completo dos artigos e a adição de novas evidências para melhorar o ranqueamento.

ANEXO

Referências selecionadas como relevantes na Revisão Sistemática utilizada no estudo de caso.

1. M. S. Ali, M. Ali Babar, L. Chen, K-J Stol. "A Systematic Review of Comparative Evidence of Aspect-Oriented Programming". *Journal of Information and Software Technology*. v.52, n.9, pp. 871-887, 2010.
2. P. Arpaia, M. L. Bernardi, G. Di Lucca, V. Inglese, G. Spiezia. "An Aspect-Oriented Programming-Based Approach to Software Development for Fault Detection in Measurement Systems". *Journal of Computer Standards & Interfaces*. v.32, n.3, pp. 141-152, 2010.
3. E. K. Piveta, A. Moreira, M. S. Pimenta, J. Araújo, P. Guerreiro, R. T. Price. "An Empirical Study of Aspect-Oriented Metrics". In: *Science of Computer Programming*. 2012.
4. E. Figueiredo, C. Sant'Anna, A. Garcia, C. Lucena. "Applying and Evaluating Concern - Sensitive Design Heuristics". *Journal of Systems and Software*. v.85, n.2, pp. 227-243, 2012.
5. R. Brcina, M. Riebisch. "Architecting for evolvability by means of traceability and features". In: *23rd IEEE/ACM International Conference on Automated Software Engineering*. pp. 72-81, 2008.
6. A. L. Medeiros, E. Figueiredo, I. Galvão, A. Garcia, T. Batista, C. Sant'Anna. "Concern-Based Assessment of Architectural Stability: A Comparative Study". In: *4th Brazilian Symposium on Software Components, Architectures and Reuse*. pp. 27-29, 2010.
7. R. Burrows, A. Garcia, F. Taiani. "Coupling Metrics for Aspect-Oriented Programming: A Systematic Review of Maintainability Studies". In: *4th International Conference on Evaluation of Novel Approaches to Software Engineering*. 2009.

8. M. Couto, M. Valente, E. Figueiredo. "Extracting Software Product Lines: A Case Study Using Conditional Compilation". In: European Conference on Software Maintenance and Reengineering. pp. 1-4, 2011.
9. J. M. Conejero, J. Hernández, E. Jurado, K. van den Berg. "Mining early aspects based on syntactical and dependency analyses". *Science of Computer Programming*. v.75, n.11, pp. 1113-1141, 2010.
10. F. d'Amorim, P. Borba. "Modularity Analysis of Use Case Implementations". *Journal of System and Software*. v.85, n.4. pp. 1012-1027, 2012.
11. E. Figueiredo, C. Sant'Anna, A. Garcia, T. T. Bartolomei, W. Cazzola, A. Marchetto. "On the Maintainability of Aspect-Oriented Software: A Concern-Oriented Measurement Framework". In: 12th European Conference on Software Maintenance and Reengineering. pp. 183-192, 2008.
12. J. M. Conejero, E. Figueiredo, A. Garcia, J. Hernández, E. Jurado. "On the relationship of concern metrics and requirements maintainability". *Information and Software Technology*. v.54, n.2, pp. 212-238, 2012.
13. M. Revelle, M. Gethers, D. Poshyvanyk. "Using Structural and Textual Information to Capture Feature Coupling in Object-Oriented Software". *Journal of Empirical Software Engineering*. v.16, n.6, pp. 773-811, 2011.

AGRADECIMENTOS

Os autores agradecem a Capes, CNPq e Fapemig pelo apoio financeiro.

REFERÊNCIAS

- [1] T. Dyba, B. A. Kitchenham, M. Jorgensen. "Evidence-based software engineering for practitioners". *IEEE Software*, vol.22, n.1, pp.58-65. 2005

- [2] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman. "Systematic literature reviews in software engineering - A systematic literature review". *Information and Software Technology*, vol.51, n.1, pp. 7-15, 2009.
- [3] B. Kitchenham. "Guidelines for Performing Systematic Literature Review in Software Engineering". EBSE Technical Report, 2.3, Keele University, 2007.
- [4] H. Ramampiaro, D. Cruzes, R. Conradi, M. Mendona. "Supporting evidence-based Software Engineering with collaborative information retrieval". In: International Conference on Collaborative Computing: Networking, Applications and Worksharing, pp.1-5, 2010.
- [5] R. Abilio, P. Teles, H. Costa, E. Figueiredo. "A Systematic Review of Contemporary Metrics for Software Maintainability". In: 2012 Sixth Brazilian Symposium on Software Components Architectures and Reuse (SBCARS), pp.130-139, 2012.
- [6] F. Tomassetti, G. Rizzo, A. Vetro, L. Ardito, M. Torchiano, M. Morisio. "Linked data approach for selection process automation in systematic reviews". In: 15th Annual Conference on Evaluation and Assessment in Software Engineering, pp. 31-35, 2011.
- [7] K. R. Felizardo, G. F. Andery, F. V. Paulovich, R. Minghim, J. C. Maldonado. "A visual analysis approach to validate the selection review of primary studies in systematic reviews". *Information and Software Technology*, vol.54, n.10, pp. 1079-1091, 2012.
- [8] B. Kitchenham. "Procedures for Performing Systematic Reviews". Keele University Technical Report TR/SE-0401. Disponível: <http://www.idi.ntnu.no/emner/empse/papers/kitchenham_2004.pdf>. Acesso: maio/2011.
- [9] R. Rocha, C. Costa, R. Prikladnicki, R. R. de Azevedo, I. H. de Farias Júnior, S. Meira. "Modelos de Colaboração no Desenvolvimento Distribuído de Software: Uma Revisão Sistemática da Literatura". In: Workshop de Desenvolvimento Distribuído de Software. 2010.
- [10] G. Travassos, J. Biolchini. "Revisões Sistemáticas Aplicadas a Engenharia de Software". In: XXI Simpósio Brasileiro de Engenharia de Software. 2007.
- [11] R. Baeza-Yates, B. Ribeiro-Neto. "Modern Information Retrieval". 2nd edition, 2011.
- [12] B. C. Wallace, T. A. Trikalinos, J. Lau, C. E. Brodley and C. H. Schmid. "Semi-automated screening of biomedical citations for systematic reviews". In: *BMC Bioinformatics*, 11:55:67, 2010.
- [13] D. Martinez, S. Karimi, L. Cavedon, T. Baldwin. "Facilitating biomedical systematic reviews using ranked text retrieval and classification". In: Proceedings of the Thirteenth Australasian Document Computing Symposium (ADCS2008), pp. 53-60, 2008.