

Em Direção a uma Ferramenta de Suporte a Revisão Sistemática de Literatura

César Augusto Verçosa Mariano¹, Ivan do Carmo Machado², Luanna Lopes Lobato¹

¹Departamento de Ciência da Computação – Universidade Federal de Goiás
CEP: 75704-020 – Catalão – GO – Brasil

²Departamento de Ciência da Computação – Universidade Federal da Bahia
CEP: 40170-110 – Salvador – BA – Brasil

cesar.m@catalao.ufg.br, ivanmachado@dcc.ufba.br, lll@catalao.ufg.br

Abstract. *This article introduces the concept and the characteristics of Systematic Literature Review, explaining its advantages and difficulties for the researcher. Based on the understanding of Systematic Review, we propose a tool that aims to assist in this process. We illustrate the tool workflow with a running example, and discuss the expected effectiveness and efficiency gains regarding time to search for primary studies, and quality outcomes.*

Resumo. *Este artigo apresenta o conceito e as principais características da Revisão Sistemática de Literatura, explicitando suas vantagens e dificuldades para o pesquisador. A partir do entendimento de Revisão Sistemática, propomos uma ferramenta que busca auxiliar neste processo. Ilustramos o workflow da ferramenta com um exemplo prático, e discutimos os ganhos de eficiência e eficácia esperados para a ferramenta, em relação ao tempo de busca por estudos primários e a qualidade dos resultados.*

1. Introdução

O processo de revisão sistemática (RS) de artigos disponibilizados na literatura – do termo original Systematic Literature Review – teve origem em pesquisas da área médica, em que traçou-se como objetivo a elaboração de uma seleção completa e imparcial sobre tópicos de pesquisa, seguindo processos conhecidos e bem definidos, que propiciasse a sua replicabilidade (Pai, 2004). Tal processo tem sido adaptado para a área de Ciência da Computação, e nos últimos anos, é visível a sua utilização abrangente, especialmente na pesquisa em Engenharia de Software (Kitchenham, 2007).

Kitchenham (2007) define revisão sistemática de literatura como a identificação, validação e interpretação de toda pesquisa relevante (estudos primários) disponível para uma questão de pesquisa particular, ou um tópico de uma área, ou ainda um fenômeno de interesse. A principal razão para a realização da revisão sistemática é que esta possui uma estratégia de busca bem definida, o que contribui para a integridade das informações resultantes da pesquisa. Por sua natureza de completude e imparcialidade, pode-se considerar que a revisão sistemática agrega maior valor científico quando comparado a revisões executadas de forma *ad-hoc*. As principais vantagens do uso da revisão sistemática estão relacionadas à segurança no processo seguido, a possibilidade de replicabilidade do estudo, bem como, a confiabilidade do processo, uma vez que este é executado de modo planejado e sistemático.

No entanto, apesar das vantagens apresentadas, o processo manual de revisão sistemática é trabalhoso e propenso a falhas, caso não seja cuidadosamente seguido. Neste contexto, a presente investigação propõe o desenvolvimento de uma ferramenta para automatizar o processo de seleção e extração de dados de estudos primários, objetos de avaliação na revisão sistemática.

A ferramenta, intitulada como SysReTool (acrônimo para **S**ystematic **R**eview **T**ool), tem o objetivo de auxiliar o processo de extração dos dados, de modo a facilitar o gerenciamento das informações que são coletadas de artigos e, desta forma, otimizar o tempo de extração e análise, evitando a perda de informações relevantes.

O artigo está estruturado como segue. O processo de revisão sistemática é descrito na seção seguinte. A seção 3 apresenta a proposta da ferramenta. A seção 4 apresenta o cenário de execução da ferramenta. A seção 5 apresenta as considerações finais desta investigação.

2. Revisão Sistemática

Ao conduzir pesquisas, é necessário entender a área sob investigação, de forma a propor avanços que tenham alguma significância. Assim, informação oriunda de estudos primários deve ser coletada. Entretanto, à medida que a quantidade de estudos a serem analisados escala, a extração de dados tende a ser cansativa, o que leva a uma possível perda de dados relevantes.

Walia e Carver (2009) descrevem RS como *“um processo formalizado e repetitivo no qual os pesquisadores procuram sistematicamente um corpo literário para documentar o estado do conhecimento sobre um determinado assunto”*. Dentre os benefícios de se realizar uma RS, podemos considerar que ela fornece ao pesquisador maior confiança na revisão de informações relevantes, uma vez que a seleção dos estudos é feita de forma criteriosa em relação ao escopo da pesquisa.

Segundo Sampaio (2007), a RS é metódica, explícita e passível de reprodução. Uma RS requer perguntas claras em relação ao que se procura pesquisar, buscando ter respostas precisas. Como resultado, espera-se a identificação de dados relevantes ao domínio de pesquisa, extraídos através das respostas às perguntas definidas.

A RS tem como princípios gerais a exaustão na busca dos estudos analisados, a seleção justificada dos estudos por critérios de inclusão e exclusão explícitos e a avaliação da qualidade metodológica, bem como a quantificação do efeito dos tratamentos por meio de técnicas estatísticas (Lima, 2000).

De acordo com Kitchenham (2007), a execução de uma Revisão Sistemática engloba três fases principais: (i) *planejamento*, (ii) *realização*, e (iii) *relato* da revisão.

Os estágios associados com o planejamento da revisão são:

- **Identificação da necessidade de uma revisão:** os pesquisadores devem identificar e analisar revisões sistemáticas já existentes com base no fenômeno de interesse em relação aos critérios de avaliação adequados. A RS só deve ser feita se realmente houver necessidade.
- **Especificação das questões de pesquisa:** elas são o elemento principal de condução da metodologia de RS. O processo de pesquisa deve identificar estudos primários que tratam das questões de pesquisa. O processo de extração de dados deve extrair os itens de dados necessários para responder às

perguntas. O processo de análise de dados deve sintetizar os dados, de forma que as perguntas possam ser respondidas.

- **Desenvolvimento de um protocolo de revisão:** documento que especifica os métodos que serão utilizados para realizar a revisão, a fim de reduzir a possibilidade de parcialidade gerada pelas expectativas do investigador.
- **Avaliação do protocolo de revisão:** A consistência interna do protocolo pode ser verificada para confirmar que cada processo da pesquisa é derivado das questões de pesquisa; Os dados a serem extraídos irão resolver adequadamente as questões de pesquisa; O procedimento de análise dos dados é apropriado para responder às questões de pesquisa.

Após a definição do protocolo, os seguintes passos devem ser executados:

- **Seleção de estudos primários:** os critérios de seleção identificam os estudos primários que forneçam evidência direta sobre as questões de pesquisa. Uma vez que os estudos potencialmente relevantes primários foram obtidos, eles precisam ser avaliados quanto a sua relevância atual.
- **Avaliação da qualidade do estudo:** a avaliação da qualidade dos estudos primários proporciona critérios de inclusão e exclusão mais detalhados.
- **Extração e monitoramento de dados:** o objetivo desta etapa é criar formas de extração de dados para registrar com precisão as informações que os pesquisadores obtiveram dos estudos primários.
- **Síntese dos dados:** informações extraídas sobre os estudos (ex: objetivo, proposta, contexto de investigação, avaliação da qualidade do estudo, tamanhos de amostra, resultados, etc.) devem ser tabuladas de forma consistente com a questão de revisão. As tabelas devem ser estruturadas para destacar semelhanças e diferenças entre os resultados do estudo.

A fase final da revisão está associada com o relato da revisão:

- **Especificação de mecanismos de disseminação:** geralmente os resultados de uma revisão sistemática podem ser divulgados em revistas acadêmicas e/ou conferências. No entanto, pode haver outras formas de difusão do conhecimento resultante da investigação.
- **Avaliação do relatório:** especialistas podem avaliar o relatório final gerado.

É possível perceber que o processo de RS é dispendioso e propenso a falhas, principalmente quando executado manualmente e por pessoas não especialistas. Por isso, neste trabalho é proposto o desenvolvimento de uma ferramenta para auxiliar este processo, como descrito na próxima seção.

3. Ferramenta proposta: SysReTool

Baseado nos estudos analisados (Kitchenham, 2007) (Walia, 2009) (Lobato, 2012), pôde-se observar alguns passos que devem ser executados para que haja uma eficiente e eficaz análise sistemática da literatura. Neste sentido, esses passos devem ser refletidos como funcionalidades da ferramenta proposta. Lobato (2012) apresenta um conjunto de etapas fundamentais para a condução de uma RS, são eles:

A 1ª Etapa consiste na identificação das questões de pesquisa. Ela engloba a definição das palavras-chave a serem utilizadas na busca de estudos primários. Este é o ponto de partida do processo, uma vez que é necessário que estabelecer um propósito da

pesquisa bem definido, de forma que se consiga definir questões que resultem em dados relevantes para a pesquisa. A 2ª Etapa seleciona as fontes de pesquisa que serão utilizadas para busca dos estudos primários.

A 3ª Etapa envolve a identificação dos estudos primários. É importante definir os critérios de inclusão e exclusão dos estudos, possibilitando assim que os estudos mais relevantes ao tema de pesquisa sejam incluídos. A seleção consiste na análise criteriosa de diferentes seções do artigo, de forma a elencar os estudos primários candidatos, como segue:

- **Selecionar os estudos pelo título:** os estudos obtidos em anais de conferências, periódicos, e nos motores de busca¹ são selecionados de acordo com o título (ou palavras chave). Se o título do estudo não condiz com o título, ou palavras chave fornecidas pelo usuário, não será incluído na lista de estudos armazenados pela ferramenta.
- **Eliminar trabalhos duplicados:** os estudos já armazenados pela ferramenta são verificados. Se há estudos com o mesmo nome, ou nomes bastante parecidos, um deles é excluído. Se dois trabalhos diferentes possuem a mesma informação o mais completo é considerado.
- **Selecionar estudos por critérios definidos:** o usuário pode filtrar os resultados obtidos na etapa anterior. Os estudos podem ser visualizados pela fonte em que foram identificados (motores de busca, conferências, periódicos) pelo ano (e/ou intervalo) de publicação, ou pelo nome do autor.
- **Selecionar trabalhos por resumo:** o usuário lê o resumo dos estudos restantes após as etapas anteriores e exclui, manualmente, os que ele julgar não serem relevantes para seu estudo.

A 4ª Etapa consiste na leitura e análise do artigo. O revisor faz a leitura de todo o texto e exclui, manualmente, os que ele julgar não serem relevantes para seu estudo. Cada estudo é avaliado de acordo com os critérios de qualidade. Os critérios de qualidade são necessários para pontuar os estudos selecionados, de modo a definir um ranking de relevância entre os estudos. Os critérios devem ser definidos com base nas questões de pesquisa estabelecidas na 1ª Etapa. Os estudos são pontuados de acordo com uma escala de relevância: *baixa*, *média* e *alta*. Para facilitar o gerenciamento de acordo com a pontuação, pode-se atribuir valores, e.g., *0(zero)* para baixa relevância, *0,5* para média e *1* para alta (Lobato, 2012).

A 5ª Etapa inclui a extração dos dados provenientes dos estudos selecionados. Assim, as informações relevantes são recolhidas através da análise dos trabalhos, a fim de responder às questões de pesquisa. As respostas às questões devem ser descritivas o suficiente para que, ao lê-las, o usuário consiga elaborar estratégias de solução e identifique falhas em relação ao campo de estudo.

A 6ª Etapa apresenta a síntese da pesquisa, que envolve a identificação dos resultados que podem ser utilizados como análise, sendo ressaltadas as vantagens e desvantagens em relação a área sob investigação, analisadas durante a revisão.

¹Os motores de busca comumente utilizados para a recuperação de publicações em Engenharia de Software são: ACM, IEEEEXPLORE, SpringerLink, ScienceDirect, e Scopus. Eventualmente, considera-se Google Scholar como um motor relevante, embora ele indexe os artigos listados pelos motores acima, o que pode resultar em redundância.

A ferramenta SysReTool apresenta dois estágios de execução: automático e manual. A execução automatizada tem por objetivo tornar a pesquisa mais otimizada, evitando que o usuário execute todas as tarefas; na manual, o usuário deve verificar cada parte da análise sem o intermédio de atividades automatizadas, apenas tendo a ferramenta como suporte ao gerenciamento das tarefas.

Inicialmente a ferramenta SysReTool executará uma busca automática nos motores de busca, e nos anais de conferências e periódicos– conforme definidos na 2ª Etapa. De modo adicional, o usuário poderá digitar links de sites onde deseja fazer a busca. Tais funcionalidades são ilustradas na Figura 1, que apresenta uma das telas do protótipo da ferramenta.

The screenshot shows the SysReTool interface with the following elements:

- Máquinas de busca sugeridas:** A list of search engines with checkboxes. 'IEEE' and 'ACM' are checked.
- Sites de busca adicionais:** A section for adding more search sites, featuring a text input field for 'URL' and an 'Adicionar...' button.
- Título:** A text input field for entering the search title.
- Palavras-chave:** A text input field for entering keywords.
- Iniciar busca:** A button at the bottom right to start the search process.

Figura 1. Tela do protótipo da ferramenta – funcionalidade que demonstra como o usuário inicia a pesquisa.

Os estudos, encontrados a partir das fontes sugeridas e das adicionais, serão filtrados por títulos de acordo com as palavras chave que foram informados pelo usuário na 1ª Etapa. Após este primeiro passo, é apresentada uma lista com os estudos selecionados de forma automatizada, mostrando o título, o ano e o autor. Ao clicar no título é possível visualizar o artigo por completo. Assim, o usuário tem a opção de remover, de forma manual, os artigos selecionados que não são interessantes para compor o conjunto de estudos que serão utilizados na pesquisa.

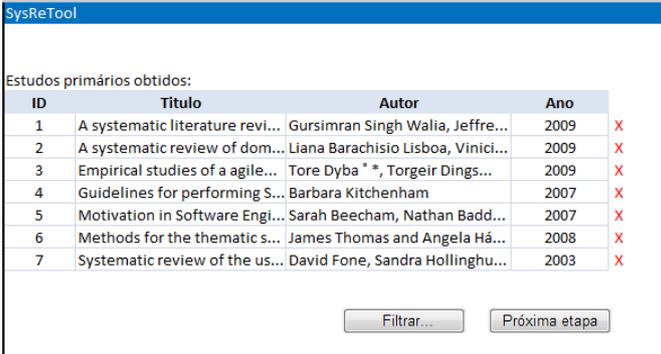
É importante ressaltar que durante a identificação dos estudos de forma automatizada, o trabalho do usuário foi reduzido e o tempo que o usuário gastaria lendo títulos foi poupado, o que poderia ocasionar cansaço, resultando na possível exclusão indevida de estudos que poderiam ser relevantes.

Após a busca automática dos estudos, a ferramenta SysReTool apresentará todos os estudos selecionados, conforme ilustrado na Figura 2, fornecendo a opção de que os estudos repetidos sejam removidos do conjunto de estudos candidatos. Esta etapa é automatizada, auxiliada pela ferramenta. O revisor pode ainda visualizar os estudos e, caso algum estudo não o interesse, é possível executar a exclusão de modo manual, clicando no X, ao lado das informações sobre o estudo.

Adicionalmente, a SysReTool possibilita que o usuário exclua artigos de acordo com o número de páginas. Essa funcionalidade, que está disponível ao clicar no botão *Filtrar*, é importante já que é possível desconsiderar estudos com número reduzido de páginas, por não estarem completos o suficiente para responder às questões de pesquisa.

Este tipo de seleção, que possibilita a exclusão dos estudos que não atendem a especificação definida de ano, autor, etc., é feita em nível da solicitação do usuário. Ou

seja, os estudos que estão fora do *range* de seleção não serão completamente excluídos, uma vez que estes são relevantes ao campo de pesquisa de um modo geral. Os estudos ficam armazenados para que a ferramenta possa desfazer a ação do usuário, caso ele deseje restaurar alguma entrada que tenha sido excluída, ou deseje fazer a seleção dos estudos considerando outras opções. Assim, para cada seleção, de acordo com a opção escolhida, os estudos serão apresentados e a lista dessa seleção pode ser salva para futuras análises, ou ainda para modificar a definição das opções de seleção.



ID	Título	Autor	Ano	
1	A systematic literature revi...	Gursimran Singh Wallia, Jeffre...	2009	X
2	A systematic review of dom...	Liana Barachisio Lisboa, Vinici...	2009	X
3	Empirical studies of a agile...	Tore Dyba * *, Torgeir Dings...	2009	X
4	Guidelines for performing S...	Barbara Kitchenham	2007	X
5	Motivation in Software Engi...	Sarah Beecham, Nathan Badd...	2007	X
6	Methods for the thematic s...	James Thomas and Angela Há...	2008	X
7	Systematic review of the us...	David Fone, Sandra Hollinghu...	2003	X

Figura 2. Tela do protótipo da ferramenta – funcionalidade que apresenta uma lista de estudos primários selecionados.

Com a definição dos estudos baseado no título, os próximos passos são a seleção dos estudos por resumo e por texto completo. Estas etapas não serão feitas automaticamente pela ferramenta, uma vez que é necessário que o usuário leia o conteúdo para fazer uma análise da relevância do estudo para a pesquisa. No entanto, apesar de ser uma análise manual a ferramenta SysReTool proverá auxílio durante todo o processo da revisão. Por exemplo, o gerenciamento dos estudos selecionados é feito pela ferramenta, onde o usuário poderá clicar no título do artigo e o resumo será exibido. Após a leitura do resumo, o usuário decidirá se aquele estudo é relevante. Caso não seja, ele será excluído. O mesmo processo será feito na última etapa, onde será exibido o texto completo e o usuário decidirá a relevância do estudo para a pesquisa.

Ao término, espera-se que tenha restado no conjunto de estudos selecionados apenas aqueles que o usuário julga serem relevantes à pesquisa em questão.

4. Cenário de execução da ferramenta

De modo a retratar o cenário de execução da ferramenta, a Figura 3 apresenta os passos que serão seguidos pela SysReTool para execução da revisão. Os dados apresentados nesta figura são fictícios, apenas para retratar a execução, em que é definida a busca e a seleção dos estudos, explicitando passo a passo a quantidade de artigos obtidos após cada passo que resulta na inclusão ou exclusão de artigos.

Como apresentado na Figura 3, existem estágios que são automatizados e outros manuais. As três primeiras etapas são executadas automaticamente pela ferramenta:

- **Etapa 1:** *Selecionar estudos por título:* a busca inicial retornou 1663 estudos. Após excluir trabalhos com títulos que claramente estavam fora do escopo restaram 99 estudos que continuaram a ser analisados.

- **Etapa 2:** *Eliminar estudos repetidos:* considerando que, havendo dois trabalhos semelhantes, o mais completo foi considerado, foram deletados 12 dos 99 estudos da etapa anterior, permanecendo 87 estudos para avaliação;
- **Etapa 3:** *Eliminar estudos por critérios definidos:* os 87 estudos foram submetidos a critérios definidos pelo usuário afim de obter uma pesquisa condizente com o perfil necessário. Restaram 48 estudos relevantes.

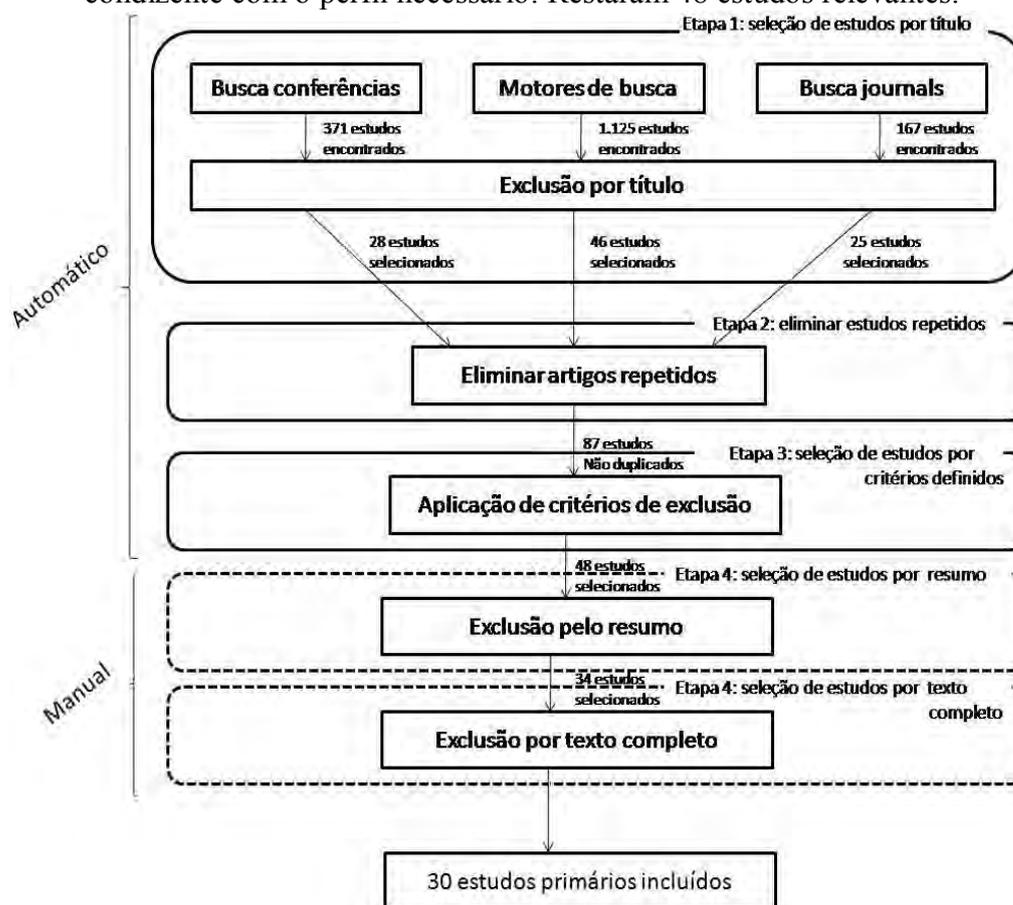


Figura 3. Passos definidos pela busca e seleção de trabalhos (adap. de Lobato (2012)).

A ferramenta armazenou uma lista com os 48 estudos relevantes obtidos após as etapas que executam o processo de eliminação e exclusão automaticamente. As próximas etapas são realizadas em modo manual, ainda com auxílio da ferramenta, que continua a armazenar a lista de estudos relevantes. As etapas manuais são:

- **Etapa 4:** *Selecionar estudos por resumo:* no exemplo, alguns estudos apresentavam informações superficiais sobre o assunto. Estes foram excluídos pelo usuário, restando 34 estudos.
- **Etapa 5:** *Selecionar estudos por texto completo:* o revisor considerou que 4, dos 34 estudos armazenados não eram relevantes e os excluiu. Logo, restaram 30 estudos ao final do processo de Revisão Sistemática.

5. Considerações Finais

Embora evidente os benefícios do uso de Revisões Sistemáticas para obtenção do estado-da-arte sobre uma determinada área de conhecimento, percebe-se que o seu processo de

elaboração, condução, e relato é trabalhoso e propenso a falhas. Desta forma, o uso de uma ferramenta computacional faz-se necessário, para que se obtenha maior qualidade dos resultados, ao passo que há redução de esforço. Uma proposta foi publicada recentemente que objetiva prover auxílio a este processo (Hernandes et al., 2012). Com base na análise dos requisitos desta ferramenta, decidimos cobrir um conjunto de aspectos que a complementasse, embora, a princípio, não haja relação entre ambas. A SysReTool inclui características de busca automáticas, que acreditamos ser elemento primordial de aumento de eficiência.

Foi publicado recentemente um estudo que verifica a viabilidade da identificação de estudos relevantes através da análise da seção de *abstract* dos artigos, e pela análise completa do texto (Torres, 2012). Baseado neste estudo nota-se que o uso da análise léxica e sintática e de expressões regulares podem auxiliar o processo de identificação de estudos primários. A classificação de relevância do estudo pode depender de alguns atributos como frequência das palavras-chave, tamanho da frase e posição das seções. De acordo com os estudos, vários métodos podem ser usados, mas ainda estão longe de chegar a um resultado ideal.

Este artigo apresenta os conceitos fundamentais de Revisões Sistemáticas de Literatura, e o workflow de uma proposta ferramenta que objetiva otimizar a realização de Revisões Sistemáticas, por meio da automatização de parte do processo., Apresentamos um exemplo prático, na tentativa de demonstrar sua viabilidade.

6. Referências

- Hernandes, E. C. M. ;Zamboni, A. B. ; Thommazo A. ; Fabbri, S. C. P. F.Using GQM and TAM to evaluate StArt a tool that supports Systematic Review. CLEI Electronic Journal, v. 15, p. 13-25, 2012.
- Kitchenham. B.; Charters, S. Guidelines for performing systematic literature reviews in software engineering (v. 2.3).Technical report, Keele University, 2007.
- Lima,M. S.; Soares, B. G. O.; Bacaltchuk, J. Psiquiatria baseada em evidências. Revista Brasileira de Psiquiatria, 2000.
- Lobato, L. L.An Approach for Risk Management in Software Product Lines. Recife: Universidade Federal de Pernambuco, 2012.
- Pai,M.; McCulloch, M.; Gorman, J. D.; Pai, N.; Enanoria, W.; Kennedy, G.; Tharyan, P.; Colford, J. M. Systematic reviews and meta-analyses: an illustrated, step-by-step guide. The National Medical Journalof India. 2004.
- Sampaio, R. F. andMancini, M. C. Estudos de Revisão Sistemática: um guia para síntese criteriosa da evidência científica. Revista brasileira de fisioterapia, São Carlos, v. 11, n.1, p. 83-89, jan/fev 2007.
- Torres, J.A.S.; Cruzes, D. S.; Salvador, L. N. Automatic Results Identification in Software Engineering Papers. Is it possible?. In: 2012 12th International Conference on Computational Science and Its Applications, Salvador-BA, p. 108-112.
- Walia, G. S. and Carver, J. C.A systematic Literature Review To Identify and Classify Software Requirement Errors. Information & Software Technology, v. 51, n. 7, pp. 1087-1109. 2009.