

# Uma Revisão Sistemática sobre Aplicações de Metaheurísticas Multiobjetivo ao Problema de Seleção de Atributos

Newton Spolaôr,<sup>1</sup>

Universidade Federal do ABC / Grupo Interdisciplinar em Mineração de Dados e Aplicações

Universidade Estadual do Oeste do Paraná / Laboratório de Bioinformática

Ana Carolina Lorena,<sup>2</sup>

Universidade Federal do ABC / Grupo Interdisciplinar em Mineração de Dados e Aplicações

Huei Diana Lee,<sup>3</sup>

Universidade Estadual do Oeste do Paraná / Laboratório de Bioinformática

**Resumo.** A revisão sistemática consiste em um método para pesquisa bibliográfica diferenciado, que permite dentre outras vantagens uma exploração ampla, rigorosa e reproduzível da literatura relacionada ao tema em estudo. A partir de diretrivas relacionadas à medicina e outras áreas, foram propostas recentemente diretrivas específicas para o uso da revisão sistemática em computação. Este relatório contextualiza e demonstra a prática de uma destas diretrivas por meio da descrição de uma revisão sistemática para o estudo de aplicações de metaheurísticas multiobjetivo no problema de seleção de atributos em bases de dados. O protocolo definido pode ser utilizado posteriormente em temas de pesquisa relacionados.

**Palavras-chave.** Algoritmos Genéticos, Aprendizado de Máquina, Mineração de Dados, Otimização Multiobjetivo, Revisão Sistemática.

## 1 Introdução

Um processo de pesquisa pode ser definido como uma sequência de atividades que permite reunir informações relacionadas a um tema e analisá-las, por meio de método científico, para obter conhecimento neste tema. Em computação, alguns exemplos de processos de pesquisa são a pesquisa analítica, a pesquisa empírica e a pesquisa bibliográfica [1].

---

<sup>1</sup>newtonspolaor@gmail.com

<sup>2</sup>aclorena@gmail.com

<sup>3</sup>hueidianalee@gmail.com

A pesquisa analítica envolve a prova formal de hipóteses lançadas pelos pesquisadores. Um processo de pesquisa empírico inclui estudos quantitativos e qualitativos. Enquanto um estudo quantitativo pode ser definido como a análise estatística de variáveis relacionadas às hipóteses definidas *a priori* [2], um estudo qualitativo inclui uma análise interpretativa, sem o uso de estatística, de observações usualmente não quantificáveis relacionadas ao tema estudado [3]. A pesquisa bibliográfica, caracterizada por uma coleta, análise e/ou síntese de trabalhos relevantes ao tema pesquisado, pode ser realizada de modo relativamente mais elaborado por meio de uma Revisão Sistemática (RS).

A Revisão Sistemática consiste em um método para pesquisa bibliográfica que permite a resolução de questões de pesquisa por meio de procedimentos explícitos para a identificação, seleção e avaliação de publicações [4]. Este processo é realizado com o intuito de explorar trabalhos relevantes e avaliar o tema pesquisado de modo justo, rigoroso e replicável [5]. Especificamente em relação à avaliação de trabalhos, uma RS pode incluir ou não uma metanálise, a qual corresponde à uma síntese dos resultados relatados por meio de técnicas estatísticas.

Em 1904 foi publicado a primeira metanálise na avaliação comparativa de dois estudos [6]. A primeira RS de fato surgiu em 1955 [7], em que estudava-se uma situação clínica relacionada ao efeito placebo. A popularidade deste tipo de pesquisa em áreas como a Medicina aumentou entre as décadas de 80 e 90 [4], embora em alguns temas de pesquisa seu uso ainda não tenha sido identificado. Este interesse crescente deve-se, entre outros fatores, à elaboração de diretrivas para Medicina [8], Ciências Sociais [9] e Computação [5, 10].

A RS pode ser organizada de modo geral em três etapas: (1) Planejamento, (2) Execução e (3) Apresentação dos Resultados [5]. A Etapa 1 envolve basicamente a definição das questões de pesquisa que pretende-se responder e a criação de um protocolo. Os procedimentos que integram este protocolo são efetivamente aplicados na Etapa 2, resultando em um conjunto de publicações pertinentes ao tema pesquisado. Na última etapa realiza-se a formatação e divulgação do conhecimento obtido por meio de trechos de dissertações e teses, relatórios técnicos, artigos e outros meios de publicação.

O objetivo deste relatório é introduzir diretrivas propostas em [5] e iniciativas utilizadas em revisões de Computação e na Medicina, aplicando-as na elaboração de uma RS que investiga aplicações de metaheurísticas multiobjetivo, como Algoritmos Genéticos Multiobjetivo (AGM), no problema de Seleção de Atributos (SA) em Bases de Dados (BD). Esta pesquisa constitui parte relevante de uma revisão bibliográfica realizada pelos autores em uma dissertação na Universidade Federal do ABC (UFABC) e introduzida em [11]. A descrição e uso de metanálise não pertence ao escopo deste relatório, devido ao seu uso relativamente incomum em Computação, à exceção de trabalhos como [12].

Na Seção 2 são introduzidas as atividades sugeridas pela diretiva para a realização de uma RS, enquanto que na Seção 3 se apresenta a RS elaborada pelos autores. O relatório é concluído na Seção 4 com algumas considerações finais.

## 2 Processo de Revisão Sistemática

Cada uma das etapas da RS é constituída por procedimentos que podem ser realizados conforme distintas diretrivas. A execução não necessariamente sequencial possibilita dentre outros benefícios o refinamento destes procedimentos [13]. De forma genérica, tem-se que:

### **Etapa 1.** Planejamento

- Identificação da necessidade de pesquisa;
- Financiamento da pesquisa (opcional);
- Definição das questões de pesquisa;
- Elaboração do protocolo;
- Avaliação do protocolo (opcional).

### **Etapa 2.** Execução

- Identificação de publicações;
- Seleção de publicações;
- Avaliação de publicações;
- Extração de informações das publicações;
- Síntese das informações extraídas.

### **Etapa 3.** Apresentação dos Resultados

- Definição de mecanismos de divulgação;
- Documentação da Revisão Sistemática;
- Avaliação da documentação (opcional).

### 2.1 Planejamento dos Procedimentos Iniciais

A primeira atividade a ser realizada obrigatoriamente em uma Revisão Sistemática consiste na identificação da necessidade de realizá-la [5], a qual pode incluir a sumarização do conhecimento relacionado ao tema em estudo ou a motivação para estudos adicionais. Em um trabalho acadêmico, como uma tese ou dissertação, é possível realizar uma RS para identificar o estado da arte e contextualizar a contribuição deste trabalho para o conhecimento corrente. Ainda no início do processo também é importante investigar outras RS realizadas sobre esse tema, de modo a garantir a relevância da revisão sob planejamento e possibilitar o aproveitamento de trechos de protocolos utilizados anteriormente. Estes trechos podem ser modificados para gerar uma sumarização atualizada e/ou mais ampla do estado da arte relacionado. Quando não houver RS no tema pesquisado, é possível investigar revisões realizadas em temas relativamente relacionados.

A definição das questões de pesquisa constitui a principal atividade do processo como um todo, visto que estas questões orientam distintos procedimentos realizados na sequência, como a extração e síntese de informações das publicações identificadas. A importância

destas questões é definida também pela forte influência que exercem na delimitação do escopo a ser investigado pela RS.

Diretivas na área médica recomendam a estruturação das questões conforme a população, as intervenções e as métricas de avaliação envolvidas para facilitar esta tarefa [9]. Estes conceitos se relacionam de modo que um grupo de pacientes é submetido a um conjunto de intervenções, isto é, tratamentos alternativos em investigação, os quais são comparados conforme métricas clínicas ou econômicas. Uma possível adaptação para temas de pesquisa em Computação consiste em representar população, intervenções e métricas respectivamente como um domínio de aplicação, métodos que atuam neste domínio e medidas comparativas entre estes métodos.

Apresenta-se a seguir alguns exemplos de questões de pesquisa extraídas em RS que investigam domínios relacionados às áreas de Medicina e Computação:

- Qual a precisão diagnóstica de Enema Opaco, Manometria Ano-retal e Biópsia de Sucção Retal em crianças com suspeita da doença de Hirschsprung [14]?
- Qual a incidência da Incontinência Fecal na comunidade australiana [15]?
- Qual a eficiência e segurança da graciloplastia dinâmica em relação à colostomia no tratamento da Incontinência Fecal [16]?
- Quais técnicas para seleção de testes de regressão foram avaliadas empiricamente [17]?
- O que é pesquisado em relação a testes de *softwares* baseados em busca [18, 19]?
- Quais iniciativas são adotadas no ensino de Arquitetura de *Software* [20]?
- Que temas têm sido empiricamente estudados em pesquisas relacionadas a *softwares Open Source* [21]?
- Que contribuições podem ser observadas na literatura relacionada à geração de especificações de requisitos de modelos de Engenharia de *Software* [22]?
- Em que áreas de testes não funcionais aplicam-se técnicas metaheurísticas de busca [23]?

A partir dos conceitos de população, intervenção e métricas sugerem-se formulações que podem originar questões de pesquisa. Em uma diretiva proposta para RS em Medicina descreve-se seis formulações [24] possivelmente adaptáveis para outras áreas de conhecimento:

1. Avaliação do efeito de uma intervenção;
2. Avaliação da frequência ou taxa de uma condição ou doença;
3. Determinação do desempenho de um teste diagnóstico;

População	Intervenção	Métricas
Câncer de mama	Auto-exame	Sensibilidade
Carcinoma	Mamografia	Especificidade
Neoplasia maligna	Biópsia	Sobrevida
Tumor		

Tabela 1: Listas de palavras para questão de pesquisa

4. Identificar etiologia e fatores de risco;
5. Identificar se uma condição pode ser prevista;
6. Avaliar o fator econômico de uma intervenção ou procedimento.

Algumas atividades propostas para a Etapa 1 são opcionais, mas podem ser necessárias em determinadas situações [5]. O financiamento da pesquisa ocorre quando uma instituição solicita a um grupo de pesquisadores a realização de uma RS, caracterizando uma terceirização desta tarefa. Neste caso, um relatório com informações sobre o apoio recebido deve ser anexado à documentação principal da revisão. A avaliação do protocolo é realizada de acordo com o que é definido pelo grupo de pesquisadores que realiza a RS. Mestrando e doutorando podem submeter o protocolo em desenvolvimento para seus orientadores, embora também seja possível solicitar esta tarefa a pesquisadores independentes. Durante a avaliação, poderia ser interessante testar parcialmente as estratégias utilizadas nos demais procedimentos, de modo a garantir a execução adequada destas atividades.

## 2.2 Definição de Protocolos

A elaboração prévia de um protocolo é necessária para amenizar uma possível subjetividade durante a execução de uma RS [5]. Um protocolo é constituído basicamente pelo *background* relacionado ao tema em estudo, o qual permite justificar a necessidade de realizar o processo, e pela descrição das estratégias para a realização dos procedimentos pertinentes à etapa de Execução.

A identificação de publicações é realizada por meio da aplicação de uma estratégia de busca de artigos, a qual pode ser definida baseando-se em RS anteriores e em testes iniciais. Uma abordagem sugerida para esta tarefa aproveita a estruturação das questões de pesquisa em termos de população, intervenção e métricas de avaliação para definir uma lista de palavras, termos e sinônimos para cada um destes conceitos. Posteriormente, agrupam-se os elementos de cada lista com o operador booleano OR e os elementos de listas distintas com o operador AND. Parênteses podem ser utilizados para impor ordem de prioridade a trechos destas expressões, enquanto o uso de aspas possibilita a investigação de termos constituídos de mais de uma palavra.

Para exemplificar essa abordagem considera-se a questão de pesquisa: **Quais procedimentos e técnicas podem ser utilizadas para se identificar indícios de câncer de mama?**. A partir desta questão, é possível elaborar as três listas de palavras descritas na Tabela 1.

Posteriormente, a utilização dos operadores booleanos e demais notações permite a construção da seguinte expressão de busca: ( (“câncer de mama” OR carcinoma OR “neoplasia maligna” OR tumor) AND (“auto-exame” OR mamografia OR biopsia) AND (sensibilidade OR especificidade OR sobrevida) ).

Após o estabelecimento das expressões, é necessário selecionar os locais em que a busca será realizada. Algumas opções são os motores de busca Scopus [25], Portal ACM [26], IEEE Xplore [27], ScienceDirect[28] e Pubmed [29], revistas especializadas, anais de eventos relacionados e trabalhos acadêmicos como teses e dissertações.

As publicações potencialmente relevantes que forem identificadas devem ser submetidas a um procedimento de seleção, de modo a manter apenas os trabalhos capazes de responder às questões de pesquisa. Este procedimento pode ser realizado por meio do uso de critérios de inclusão e exclusão, os quais são definidos previamente no protocolo e possivelmente refinados conforme o andamento da RS. A adoção de uma postura conservadora em relação à exclusão de publicações é importante para evitar a perda de informações relevantes. Desse modo, quando a leitura de títulos e resumos não for suficiente para se selecionar indubitavelmente uma publicação, é necessário ler outros trechos ou até mesmo o trabalho inteiro até que a seleção possa ser realizada convictamente. Apresentam-se a seguir alguns critérios utilizados em revisões nas áreas de Medicina e Computação:

**Exclusão.** Trabalhos irrelevantes e/ou duplicados [17];

**Inclusão.** Trabalhos que apresentam palavras da expressão de busca no seu título ou no seu resumo [20, 19];

**Inclusão.** Uso de instrumento válido para medir Incontinência Fecal [15];

**Inclusão.** Trabalhos com mais de 125 pacientes por gênero [15].

Quando mais de um pesquisador aplica o procedimento de seleção em um mesmo trabalho, é possível resolver eventuais divergências sobre inclusões e exclusões por meio de consenso. Quando a RS é realizada por um mestrandoo ou doutorando, as dúvidas relacionadas à seleção podem ser discutidas com o orientador ou outros pesquisadores.

A avaliação de publicações é importante devido, entre outros motivos, ao suporte que provê para a elaboração de discussões e considerações, definição de trabalhos futuros e refinamento de critérios de seleção [5]. Este procedimento é realizado por meio da aplicação de critérios de qualidade, os quais são usualmente verificados por meio de *checklists* e possibilitam identificar se as evidências destacadas em uma publicação são relevantes. Na literatura existem distintas *checklists* disponíveis, a partir da qual podem ser coletados os critérios de qualidade mais pertinentes às questões de pesquisa em investigação [30].

As *checklists* podem auxiliar na prática das seguintes abordagens [5]: (1) Definição de critérios de seleção mais detalhados e (2) Auxílio na análise e síntese das informações coletadas das publicações para responder as questões de pesquisa. Enquanto que a primeira abordagem requer um formulário para extração de informações separado do formulário principal, a outra abordagem pode ser realizada com auxílio de um formulário único. É importante também criar *checklists* separados para a avaliação de trabalhos quantitativos e qualitativos.

Ainda em relação a esse procedimento, deve-se ater à avaliação da qualidade metodológica ao invés da qualidade de escrita das publicações analisadas. Desse modo, quando um critério de qualidade não puder ser respondido, mesmo após a leitura completa do trabalho em investigação, é interessante entrar em contato com os autores da publicação para realizar a avaliação adequadamente.

Uma atividade que pode ser realizada conjuntamente à avaliação ou não, respectivamente de acordo com as abordagens 2 ou 1, contempla a extração de informações das publicações previamente selecionadas. Estas informações são registradas em formulários, os quais podem ser testados antes da aplicação do protocolo, assim como as estratégias de identificação e seleção de trabalhos. O uso de formulários como planilhas eletrônicas torna-se interessante para auxiliar em análises posteriores [5].

As informações extraídas devem ser sintetizadas para melhorar a compreensão das contribuições e identificar similaridades e diferenças relacionadas a estudos quantitativos, dentre outros motivos. A síntese pode ser quantitativa, em que se provê suporte à metanálise, ou qualitativa [5]. No último caso, as informações devem ser organizadas de modo a destacar relações de semelhança entre os trabalhos investigados. A síntese quantitativa envolve a organização das publicações de acordo com informações numéricas como tamanho das amostras, taxa de acerto e desvio padrão, as quais devem ser apresentadas de um modo comparável que considere as diferenças entre os distintos experimentos. Para realizar síntese de estudos qualitativos, sugere-se uma das abordagens descritas em [31].

Os formulários e documentos obtidos após a síntese das publicações devem ser formatados e disseminados na comunidade para a conclusão da RS. Neste contexto, o planejamento do protocolo é concluído com a definição dos meios de disseminação a serem utilizados, como conferências, revistas, posteres e *web sites*, e com a elaboração do documento principal da revisão [5]. Este documento pode ser construído na forma de artigos, os quais podem referenciar a descrição detalhada da RS em dissertações, teses ou relatórios técnicos quando houver restrição de espaço, ou ainda diretamente como trecho de trabalhos acadêmicos.

### 3 Aplicação de Revisão Sistemática

Conforme mencionado, foi realizada pelos autores uma Revisão Sistemática para investigar publicações sobre aplicações de metaheurísticas multiobjetivo, como Algoritmos Genéticos Multiobjetivo, *Ant Colony Optimization*, Otimização por Nuvem de Partículas Binárias (*Binary Particle Swarm Optimization*), Programação Genética e Têmpera Simulada, no problema de Seleção de Atributos em Bases de Dados. Uma maior contextualização sobre as metaheurísticas multiobjetivo, o problema de SA e algumas aplicações é fornecida em [11, 32, 33, 34, 35]. O protocolo proposto e as atividades realizadas em cada uma das etapas do processo são apresentados no restante desta seção.

#### 3.1 Planejamento

A elaboração desta RS foi motivada pela necessidade de atualizar o *background* e identificar o estado da arte e os métodos mais usuais, além de contextualizar a dissertação

desenvolvida pelos autores em relação a aplicações de metaheurísticas para SA [11]. Embora a dissertação tenha como enfoque apenas os AGM, torna-se interessante pesquisar outras metaheurísticas para justificar a escolha deste algoritmo. Não foram encontradas RS específicas nesse tema de pesquisa, o que motiva a definição das questões de pesquisa e do protocolo baseando-se em RS voltadas para o estudo de aplicações de metaheurísticas em problemas de Engenharia de *Software* (ES).

A elaboração das questões de pesquisa parte da definição de população, intervenções e métricas de avaliação respectivamente como o domínio de SA e algumas metaheurísticas e estratégias de avaliação destas metaheurísticas utilizadas em outros domínios. A exemplo de outras RS, são especificadas sub-questões para prover um maior detalhamento dos questionamentos. A abrangência desses conceitos permite uma ampla exploração do tema pesquisado por meio das seguintes questões:

**Q1.** Quais são as aplicações de AGM para SA em BD?

1. Quais são as medidas adotadas nestas aplicações para avaliação individual ou de subconjuntos de atributos?
2. Quais são as estratégias multiobjetivo adotadas nestas aplicações?
3. Quais são as configurações paramétricas adotadas nestas aplicações?
4. Quais são as estratégias de avaliação em cada uma destas aplicações, isto é, quais são as abordagens de avaliação do subconjunto de atributos selecionado em cada um destes trabalhos?
5. Quais são as vantagens da Otimização Multiobjetivo (OM) em relação à Otimização mono-objetivo nestas aplicações?
6. Quais são as vantagens do uso de AGM para SA nestas aplicações em relação a outras abordagens de SA em BD?
7. Quais são os métodos de SA utilizados como *baselines*<sup>4</sup> nestas aplicações?
8. Quais são os conjuntos de dados utilizados nestas aplicações?
9. São construídos conjuntos de dados artificiais nestas aplicações?
10. São realizadas customizações adicionais nos AGM nestas aplicações?
11. É adotada nestas aplicações alguma abordagem para seleção de uma única solução dentre soluções não-dominadas identificadas em algumas estratégias de OM?

**Q2.** Quais outras variações de algoritmos evolucionários multiobjetivo são aplicadas em SA em BD?

1. Quais são as vantagens de utilizar estas variações em relação aos AGM?
2. Quais são as desvantagens ou limitações em se utilizar estas variações em relação aos AGM?

---

<sup>4</sup>Técnicas utilizadas para comparação com método principal.

**Q3.** Quais outras metaheurísticas multiobjetivo são aplicadas em SA em BD?

1. Quais são as vantagens de utilizar estas variações em relação aos AGM?
2. Quais são as desvantagens de utilizar estas variações em relação aos AGM?

Constrói-se em seguida o protocolo a ser utilizado no restante do processo. O *background* foi elaborado a partir do *know-how* prévio dos autores e refinado durante o estudo das publicações identificadas no restante da RS. Dentre as estratégias relacionadas a cada procedimento, define-se primeiramente a estratégia de busca de publicações. Neste contexto, optou-se pelos motores de busca Portal ACM [26], CiteSeerX [36], IEEE Xplore [27], ScienceDirect [28], Scirus [37], Scopus [25], SpringerLink Beta [38], Wiley Interscience [39] e Web of Science [40].

Desse modo, não foram investigados trabalhos não publicados e conferências ou revistas especializadas. Esta atitude justifica-se porque acredita-se que estes tipos de trabalho são em sua maioria indexáveis ou referenciáveis por alguma publicação indexável nos distintos motores de busca selecionados. Trabalhos acadêmicos não foram considerados porque acredita-se que a maior parte deles resultaram em alguma publicação acessível pelos motores de busca.

São apresentadas na sequência as listas de termos utilizadas, as quais são definidas a partir das populações e intervenções envolvidas e de tópicos inerentes a estes conceitos. Embora não se tenha criado uma lista relacionada às métricas de avaliação, considera-se uma subquestão que trata exatamente deste conceito.

**População.** “feature selection”, “attribute selection”, “attribute reduction”, “dimensionality reduction”, “feature subset selection”, “feature ranking”;

**Intervenções. Algoritmos Genéticos.** “genetic algorithm(s)”, “genetic”;

**Otimização Multiobjetivo.** “multi-objective”, “multiobjective”, “multi objective”, “multi-criteria”, “multicriteria”, “multi criteria”, “many-objective”, “many objective”, “many-criteria”, “many criteria”, “multi-objective optimization”, “multiobjective optimization”, “multi objective optimization”, “multi-criteria optimization”, “multicriteria optimization”, “multi criteria optimization”, “multi-objective optimisation”, “multiobjective optimisation”, “multi objective optimisation”, “multi-criteria optimisation”, “multicriteria optimisation”, “multi criteria optimisation”, “many-objective optimisation”, “many-objective optimization”, “many objective optimisation”, “many-criteria optimisation”, “many criteria optimisation”, “many-objective optimisation”, “many objective optimisation”, “many-criteria optimisation”, “many criteria optimisation”

**Algoritmos Genéticos Multiobjetivo.** “multiobjective genetic algorithm (s)”, “multi-objective genetic algorithm (s)”, “multi objective genetic algorithm (s)”, “multicriteria genetic algorithm (s)”, “multi-criteria genetic algorithm (s)”, “multi criteria genetic algorithm (s)”, “many-objective genetic algorithm (s)”, “many objective genetic algorithm (s)”, “many-criteria genetic algorithm (s)”, “many criteria genetic algorithm (s)”

**Algoritmos Evolucionários.** “evolutionary”, “evolutionary algorithm (s)”, “memetic”, “memetic algorithm (s)”

**Metaheurísticas.** “metaheuristic search”, “meta-heuristic search”, “meta heuristic search”, “metaheuristic search algorithm (s)”, “meta-heuristic search algorithm (s)”, “meta heuristic search algorithm (s)”, “metaheuristic (s)”, “meta-heuristic (s)”, “meta heuristic (s)”, “ant colony”, “ant colony optimization”, “ant colony optimisation”, “artificial bee colony”, “bee colony”, “bee colony optimisation”, “bee colony optimisation”, “particle swarm”, “particle swarm optimisation”, “particle swarm optimisation”, “artificial immune”, “artificial immune system(s)”, “artificial immune optimization”, “artificial immune optimisation”, “simulated annealing”, “tabu search”, “hill climbing”, “hill-climbing”

**Algoritmos Evolucionários Multiobjetivo.** “multiobjective evolutionary algorithm (s)”, “multi-objective evolutionary algorithm (s)”, “multi objective evolutionary algorithm (s)”, “multicriteria evolutionary algorithm (s)”, “multi-criteria evolutionary algorithm (s)”, “multi criteria evolutionary algorithm (s)”, “many-objective evolutionary algorithm (s)”, “many objective evolutionary algorithm (s)”, “many-criteria evolutionary algorithm (s)”, “many criteria evolutionary algorithm (s)”

A partir destas listas, utiliza-se os operadores booleanos para a elaboração das expressões de busca. Primeiramente são definidas expressões de busca mais simples, as quais reúnem termos pertinentes a tópicos fortemente relacionados, como Algoritmos Genéticos (AG) e Algoritmos Evolucionários (AE). A integração entre estas expressões resulta nas expressões de busca definitivas apresentadas a seguir. Optou-se por formular uma única expressão de busca para cada uma das três questões de pesquisa. A última expressão é relativamente extensa, em relação às expressões utilizadas em outras RS, devido à motivação em se explorar aplicações de distintas metaheurísticas para SA. Questões de pesquisa reduzidas, com uma quantidade menor de termos relacionados a metaheurísticas, teriam uma capacidade exploratória possivelmente menor.

**AGM em SA.** (“multi-objective” OR multiobjective OR “multi objective” OR “multi-criteria” OR multicriteria OR “multi criteria” OR “many-objective” OR “many objective” OR “many-criteria” OR “many criteria”) AND (genetic OR “genetic algorithm” OR “genetic algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction” OR “feature subset selection” OR “feature ranking”)

**Algoritmos Evolucionários Multiobjetivo em SA.** (“multi-objective” OR multiobjective OR “multi objective” OR “multi-criteria” OR multicriteria OR “multi criteria” OR “many-objective” OR “many objective” OR “many-criteria” OR “many criteria”) AND ((evolutionary OR “evolutionary algorithm” OR “evolutionary algorithms”) OR (memetic OR “memetic algorithm” OR “memetic algorithms”)) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction” OR “feature subset selection” OR “feature ranking”)

**Metaheurísticas Multiobjetivo em SA.** (“multi-objective” OR multiobjective OR “multi objective” OR “multi-criteria” OR multicriteria OR “multi criteria” OR “many-objective” OR “many objective” OR “many-criteria” OR “many criteria”) AND ((metaheuristic OR “meta-heuristic” OR “meta heuristic” OR “metaheuristic algorithm” OR “meta-heuristic algorithm” OR “meta heuristic algorithm” OR “metaheuristic algorithms” OR “meta-heuristic algorithms” OR “meta heuristic algorithms” OR “metaheuristic search algorithm” OR “meta-heuristic search algorithm” OR “meta heuristic search algorithm” OR “metaheuristic search algorithms” OR “meta-heuristic search algorithms” OR “meta heuristic search algorithms”) OR (“ant colony” OR “ant colony optimization” OR “ant colony optimisation”) OR (“artificial bee colony” OR “bee colony” OR “artificial bee colony optimization” OR “bee colony optimization” OR “artificial bee colony optimisation” OR “bee colony optimisation”) OR (“particle swarm” OR “particle swarm optimization” OR “particle swarm optimisation”) OR (“artificial immune” OR “artifical immune system” OR “artifical immune systems” OR “artificial immune optimization” OR “artificial immune optimisation”) OR (“simulated annealing”) OR (“tabu search”) OR (“hill climbing” OR “hill-climbing”)) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction” OR “feature subset selection” OR “feature ranking”)

A estratégia adotada para seleção de publicações contempla inicialmente a exclusão imediata das publicações com mesmo título indexadas e identificadas simultaneamente por distintos motores de busca. Posteriormente, realiza-se a leitura do título e resumo de todas as publicações em investigação. Se a seleção não puder ser realizada indubitavelmente, considera-se a leitura da introdução e da conclusão e, se necessário, do restante do trabalho até que não reste dúvidas sobre a seleção. Os critérios de exclusão utilizados são descritos a seguir. Não são utilizados critérios de inclusão por se supor que as publicações que não puderem ser excluídas mediante os demais critérios deveriam ser imediatamente incluídas para a avaliação de qualidade.

- Publicações com enfoque em seleção de regras, protótipos, classificadores, parâmetros, arquiteturas, pesos, redutos ou outros tipos de conceitos que não podem ser representados como atributos ou colunas em uma tabela atributo-valor;
- Publicações que não realizam avaliação experimental do método proposto;
- Publicações que não realizam comparação do método proposto com algum método *baseline*;
- Publicações que são duplicadas em relação aos resultados, com exceção da versão mais completa;
- Publicações com 1 página, pôsteres, apresentações e anais de eventos;
- Publicações relacionadas à OM em que apenas um dos objetivos é relacionado à SA<sup>5</sup>;

---

<sup>5</sup>Os demais objetivos poderiam ser relacionados à seleção de exemplos, por exemplo.

- Publicações que não atendem às questões de pesquisa;
- Publicações hospedados em serviços de acesso restrito não subsidiados pela UFABC;
- Teses, dissertações, relatórios técnicos, tutoriais e editoriais.

O *checklist* considerado na avaliação das publicações é elaborado a partir da adaptação para o tema de pesquisa pertinente a este relatório de critérios sugeridos em [5] ou aplicados em outras RS, conforme enfatizado em [30], e pelo uso de critérios propostos pelos autores. Os critérios listados a seguir são validados posteriormente em aproximadamente 20% das publicações identificadas para conferir sua consistência:

- O método para SA em BD utilizado é implementado por meio de um protótipo ou sistema [22]?
- As medidas de importância, isto é, os objetivos relacionados na OM que são utilizados para SA em BD são descritas completa ou parcialmente [23]?
- A estratégia de avaliação da solução identificada envolve *baselines* metaheurísticos multiobjetivo para Seleção de Atributos?
- A estratégia de avaliação da solução identificada envolve análise de significância estatística [41]?
- A abordagem de SA em BD adotada é a filtro?
- As propriedades do conjunto de dados utilizado são completa ou parcialmente descritas [41]?
- Os autores descrevem características do problema de Seleção de Atributos que justificam o uso de metaheurísticas [19]?
- Os autores descrevem características do problema de Seleção de Atributos que justificam o uso de otimização multiobjetivo?
- Os resultados obtidos no trabalho são comparados com resultados anteriores [41]?
- São justificadas as configurações paramétricas das metaheurísticas adotadas [19]?
- Para AE, são adotadas estratégias específicas para aumentar a diversidade da população de indivíduos?
- Para AE, são adotadas estratégias específicas para inicializar a população de indivíduos [19]?
- Para AG, são adotadas estratégias específicas para amenizar a variabilidade ou randomicidade<sup>6</sup> inerente aos resultados [19]?

---

<sup>6</sup>Pseudo-aleatoriedade.

- Para OM baseada em soluções não-dominadas, é adotada alguma estratégia para selecionar uma única solução não-dominada dentre as demais identificadas pela metaheurística?
- São consideradas restrições às metaheurísticas devido aos tipos de atributos (discretos ou contínuos) presentes nos conjuntos de dados?
- São descritas as vantagens identificadas na OM em relação à Otimização mono-objetivo?
- São descritas as vantagens identificadas no uso de metaheurísticas Multiobjetivo para SA em BD nestas aplicações em relação a outros métodos de SA em BD?

Após a avaliação, aplica-se uma estratégia de extração de informações proposta em conformidade com a abordagem 2, isto é, que permite a obtenção de suporte à análise e síntese das informações coletadas das publicações para responder as questões de pesquisa. Desse modo, utiliza-se um único formulário para contemplar tanto os critérios de qualidade quanto os quesitos para extração de informações. Este formulário contém então informações gerais dos trabalhos, como autores e ano de publicação, e informações mais detalhadas e critérios aplicados previamente em [23, 18] ou propostos pelos autores desta RS. Os tópicos capturados por este formulário, na forma de colunas em uma planilha eletrônica, são apresentados a seguir:

- Abordagem de SA utilizada;
- Ano de publicação;
- Autores;
- *Baselines* utilizados em relação à metaheurística;
- Comparação com resultados de trabalhos publicados anteriormente;
- Configurações paramétricas;
- Nome, quantidade de exemplos, quantidade de atributos, quantidade de classes, natureza e pré-processamento relacionados ao(s) conjunto(s) de dados investigado(s);
- Estratégia de combate à variabilidade da metaheurística;
- Estratégia de diversidade dos indivíduos;
- Estratégia de inicialização dos indivíduos;
- Para estratégias de OM baseadas em soluções não-dominadas, estratégia de seleção de uma única solução não-dominada;
- Estratégia de OM utilizada;
- Estratégia, indutor e medidas utilizada para avaliação da melhor solução;

- Fonte de publicação;
- Justificativa(s) das configurações paramétricas adotadas;
- Medidas (objetivos) de importância de atributos utilizadas;
- Metaheurística utilizada;
- Motivação para uso de metaheurística;
- Motivação para uso de OM;
- Objetivo principal do trabalho no contexto da metaheurística multiobjetivo;
- Restrições associadas à aplicação da metaheurística;
- Melhor resultado obtido para cada medida de avaliação;
- Nome e intervalo de confiança do teste estatístico utilizado;
- Observações.

Opta-se por realizar uma síntese qualitativa das informações descritas nesse formulário, a exemplo da maioria das RS para o estudo de publicações em ES [42]. Como o método utilizado na dissertação que fundamenta a RS realizada pelos autores baseia-se na abordagem filtro, sintetiza-se com maior enfoque apenas os trabalhos que implementam esta abordagem por meio de AGM por meio da descrição deles a partir de alguns dos tópicos coletados no formulário.

O meio de divulgação da RS selecionado foi a disseminação deste relatório técnico para discentes e docentes da UFABC e para colaboradores de instituições como a Universidade Estadual do Oeste do Paraná (UNIOESTE) basicamente por meio de contatos eletrônicos.

### **3.2 Execução e Apresentação dos Resultados**

O protocolo proposto foi aplicado em meados de Dezembro de 2009. O primeiro procedimento realizado foi a identificação de publicações por meio de expressões de busca. Adaptações foram necessárias em motores de busca que não suportam expressões de busca muito extensas ou que apresentam limitações como a investigação separada de títulos, resumos e *keywords* ou de campos aproximadamente equivalentes. O SpringerLink Beta, por exemplo, restringe a consulta à expressões de até 10 palavras, o que implicou na geração manual de sub-expressões de busca reduzidas que mantivessem o mesmo escopo das expressões originais. A combinação dos resultados parciais evitou que as limitações identificadas afetassem os resultados finais deste procedimento.

As expressões e sub-expressões de busca utilizadas e a quantidade de publicações identificadas em cada motor, considerando cada questão de pesquisa, são apresentadas a seguir. Os casos em que foi possível aplicar as expressões de busca definitivas são destacados com a notação *expressão original*. A investigação das expressões em campos específicos dos trabalhos, como título e resumo, foi utilizada na maioria dos motores de busca que disponibilizam este serviço ao usuário.

- Scopus

- Q1.** (TITLE-ABS-KEY("multi-objective" OR multiobjective OR "multi objective" OR "multi-criteria" OR multicriteria OR "multi criteria" OR "many-objective" OR "many objective" OR "many-criteria" OR "many criteria") AND (genetic OR "genetic algorithm" OR "genetic algorithms") AND ("feature selection" OR "attribute selection" OR "attribute reduction" OR "dimensionality reduction" OR "feature subset selection" OR "feature ranking")): 287 publicações
- Q2.** (TITLE-ABS-KEY("multi-objective" OR multiobjective OR "multi objective" OR "multi-criteria" OR multicriteria OR "multi criteria" OR "many-objective" OR "many objective" OR "many-criteria" OR "many criteria") AND ((evolutionary OR "evolutionary algorithm" OR "evolutionary algorithms") OR (memetic OR "memetic algorithm" OR "memetic algorithms")) AND ("feature selection" OR "attribute selection" OR "attribute reduction" OR "dimensionality reduction" OR "feature subset selection" OR "feature ranking")): 256 publicações
- Q3.** (TITLE-ABS-KEY("multi-objective" OR multiobjective OR "multi objective" OR "multi-criteria" OR multicriteria OR "multi criteria" OR "many-objective" OR "many objective" OR "many-criteria" OR "many criteria") AND ((metaheuristic OR "meta-heuristic" OR "meta heuristic" OR "metaheuristic algorithm" OR "meta-heuristic algorithm" OR "meta heuristic algorithm" OR "metaheuristic algorithms" OR "meta-heuristic algorithms" OR "meta heuristic algorithms" OR "metaheuristic search algorithm" OR "meta-heuristic search algorithm" OR "metaheuristic search algorithm" OR "meta heuristic search algorithm" OR "metaheuristic search algorithms" OR "meta-heuristic search algorithms" OR "meta heuristic search algorithms") OR ("ant colony" OR "ant colony optimization" OR "ant colony optimisation") OR ("artificial bee colony" OR "bee colony" OR "artificial bee colony optimisation" OR "bee colony optimisation" OR "artificial bee colony optimisation" OR "bee colony optimisation") OR ("particle swarm" OR "particle swarm optimization" OR "particle swarm optimisation" OR ("artificial immune" OR "artifical immune system" OR "artifical immune systems" OR "artificial immune optimization" OR "artificial immune optimisation") OR ("simulated annealing") OR ("tabu search") OR ("hill climbing" OR "hill-climbing")) AND ("feature selection" OR "attribute selection" OR "attribute reduction" OR "dimensionality reduction" OR "feature subset selection" OR "feature ranking")): 72 publicações

- Web of Science

- Q1.** TS<sup>7</sup>=(("multi-objective" OR multiobjective OR "multi objective" OR "multi-criteria" OR multicriteria OR "multi criteria" OR "many-objective" OR "many objective" OR "many-criteria" OR "many criteria") AND (genetic OR "genetic algorithm" OR "genetic algorithms") AND ("feature selection" OR "attribute selection" OR "attribute reduction" OR "dimensionality reduction" OR "feature subset selection" OR "feature ranking"))

---

<sup>7</sup> Title, Author and Keywords.

selection” OR “attribute reduction” OR “dimensionality reduction” OR “feature subset selection” OR “feature ranking”)): 104 publicações

**Q2.** TS=((“multi-objective” OR multiobjective OR “multi objective” OR “multi-criteria” OR multicriteria OR “multi criteria” OR “many-objective” OR “many objective” OR “many-criteria” OR “many criteria”) AND ((evolutionary OR “evolutionary algorithm” OR “evolutionary algorithms”) OR (memetic OR “memetic algorithm” OR “memetic algorithms”)) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction” OR “feature subset selection” OR “feature ranking”)): 30 publicações

**Q3.** 11 publicações

**Sub-expresão 1.** TS=((“multi-objective” OR multiobjective OR “multi objective” OR “multi-criteria” OR multicriteria OR “multi criteria” OR “many-objective” OR “many objective” OR “many-criteria” OR “many criteria”) AND ((metaheuristic OR “meta-heuristic” OR “meta heuristic” OR “metaheuristic algorithm” OR “meta-heuristic algorithm” OR “meta heuristic algorithm” OR “metaheuristic algorithms” OR “meta-heuristic algorithms” OR “meta heuristic algorithms” OR “metaheuristic search algorithm” OR “meta-heuristic search algorithm” OR “metaheuristic search algorithms” OR “meta-heuristic search algorithms” OR (“ant colony” OR “ant colony optimization” OR “ant colony optimisation”) OR (“artificial bee colony” OR “bee colony” OR “artificial bee colony optimization” OR “bee colony optimization” OR “artificial bee colony optimisation” OR “bee colony optimisation”) OR (“particle swarm” OR “particle swarm optimization” OR “particle swarm optimisation”) OR (“artificial immune” OR “artifical immune system” OR “artifical immune systems” OR “artificial immune optimization” OR “artificial immune optimisation” OR (“simulated annealing”) OR (“tabu search”) OR (“hill climbing” OR “hill-climbing”)) AND (“feature selection” OR “attribute selection” OR “attribute reduction”))

**Sub-expresão 2.** TS=((“multi-objective” OR multiobjective OR “multi objective” OR “multi-criteria” OR multicriteria OR “multi criteria” OR “many-objective” OR “many objective” OR “many-criteria” OR “many criteria”) AND ((metaheuristic OR “meta-heuristic” OR “meta heuristic” OR “metaheuristic algorithm” OR “meta-heuristic algorithm” OR “meta heuristic algorithm” OR “metaheuristic algorithms” OR “meta-heuristic algorithms” OR “meta heuristic algorithms” OR “metaheuristic search algorithm” OR “meta-heuristic search algorithm” OR “metaheuristic search algorithms” OR “meta-heuristic search algorithms” OR (“ant colony” OR “ant colony optimization” OR “ant colony optimisation”) OR (“artificial bee colony” OR “bee colony” OR “artificial bee colony optimization” OR “bee colony optimization” OR “artificial bee colony optimisation” OR “bee colony optimisation”) OR (“particle swarm” OR “particle swarm optimization” OR “particle swarm optimisation”) OR (“artificial immune” OR “artifical immune system” OR “artifical immune systems” OR “artificial immune optimization” OR “artificial immune optimisation” OR (“simulated annealing”) OR (“tabu search”) OR (“hill climbing” OR “hill-climbing”)) AND (“feature selection” OR “attribute selection” OR “attribute reduction”))

“bee colony optimisation”) OR (“particle swarm” OR “particle swarm optimization” OR “particle swarm optimisation”) OR (“artificial immune” OR “artifical immune system” OR “artifical immune systems” OR “artificial immune optimization” OR “artificial immune optimisation”) OR (“simulated annealing”) OR (“tabu search”) OR (“hill climbing” OR “hill-climbing”)) AND (“dimensionality reduction” OR “feature subset selection” OR “feature ranking”))

- Portal ACM
  - Portal ACM Guide
    - Q1.** *expressão original:* 515 publicações
    - Q2.** *expressão original:* 446 publicações
    - Q3.** *expressão original:* 229 publicações
  - Portal ACM Digital Library
    - Q1.** *expressão original:* 108 publicações
    - Q2.** *expressão original:* 119 publicações
    - Q3.** *expressão original:* 46 publicações
- IEEE Xplore
  - Q1.** *expressão original:* 29 publicações
  - Q2.** *expressão original:* 22 publicações
  - Q3.** *expressão original:* 3 publicações
- ScienceDirect
  - Q1.** *expressão original:* 444 publicações
  - Q2.** *expressão original:* 336 publicações
  - Q3.** *expressão original:* 218 publicações
- Scirus
  - Q1.** *expressão original:* 37 publicações
  - Q2.** *expressão original:* 64 publicações
  - Q3.** *expressão original:* 0 publicações
- SpringerLink Beta
  - Q1.** 17 publicações
  - Sub-expressão 1.** (multiobjective OR “multi objective”) AND (genetic OR “genetic algorithm” OR “genetic algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 2.** (multicriteria OR “multi criteria”) AND (genetic OR “genetic algorithm” OR “genetic algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 3.** (“many objective” OR “many criteria”) AND (genetic OR “genetic algorithm” OR “genetic algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 4.** (multiobjective OR “multi objective”) AND (genetic OR “genetic algorithm” OR “genetic algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 5.** (multicriteria OR “multi criteria”) AND (genetic OR “genetic algorithm” OR “genetic algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 6.** (“many objective” OR “many criteria”) AND (genetic OR “genetic algorithm” OR “genetic algorithms”) AND (“feature subset selection” OR “feature ranking”)

## Q2. 11 publicações

**Sub-expressão 1.** (multiobjective OR “multi objective”) AND (evolutionary OR “evolutionary algorithm” OR “evolutionary algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 2.** (multicriteria OR “multi criteria”) AND (evolutionary OR “evolutionary algorithm” OR “evolutionary algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 3.** (“many objective” OR “many criteria”) AND (evolutionary OR “evolutionary algorithm” OR “evolutionary algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 4.** (multiobjective OR “multi objective”) AND (memetic OR “memetic algorithm” OR “memetic algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 5.** (multicriteria OR “multi criteria”) AND (memetic OR “memetic algorithm” OR “memetic algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 6.** (“many objective” OR “many criteria”) AND (memetic OR “memetic algorithm” OR “memetic algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 7.** (multiobjective OR “multi objective”) AND (evolutionary OR “evolutionary algorithm” OR “evolutionary algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 8.** (multicriteria OR “multi criteria”) AND (evolutionary OR “evolutionary algorithm” OR “evolutionary algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 9.** (“many objective” OR “many criteria”) AND (evolutionary OR “evolutionary algorithm” OR “evolutionary algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 10.** (multiobjective OR “multi objective”) AND (memetic OR “memetic algorithm” OR “memetic algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 11.** (multicriteria OR “multi criteria”) AND (memetic OR “memetic algorithm” OR “memetic algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 12.** (“many objective” OR “many criteria”) AND (memetic OR “memetic algorithm” OR “memetic algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Q3.** 3 publicações

**Sub-expressão 1.** (multiobjective OR “multi objective”) AND (metaheuristic OR “meta heuristic” OR “metaheuristic algorithm”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 2.** (multicriteria OR “multi criteria”) AND (metaheuristic OR “meta heuristic” OR “metaheuristic algorithm”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 3.** (“many objective” OR “many criteria”) AND (metaheuristic OR “meta heuristic” OR “metaheuristic algorithm”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 4.** (multiobjective OR “multi objective”) AND (“meta heuristic algorithm” OR “metaheuristic algorithms” OR “meta heuristic algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 5.** (multicriteria OR “multi criteria”) AND (“meta heuristic algorithm” OR “metaheuristic algorithms” OR “meta heuristic algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 6.** (“many objective” OR “many criteria”) AND (“meta heuristic algorithm” OR “metaheuristic algorithms” OR “meta heuristic algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 7.** (multiobjective OR “multi objective”) AND (“metaheuristic search algorithm” OR “meta heuristic search algorithm” OR “metaheuristic search algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 8.** (multicriteria OR “multi criteria”) AND (“metaheuristic search algorithm” OR “meta heuristic search algorithm” OR “metaheuristic search algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 9.** (“many objective” OR “many criteria”) AND (“metaheuristic search algorithm” OR “meta heuristic search algorithm” OR “metaheuristic search algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 10.** (multiobjective OR “multi objective”) AND (“meta heuristic search algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 11.** (multicriteria OR “multi criteria”) AND (“meta heuristic search algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 12.** (“many objective” OR “many criteria”) AND (“meta heuristic search algorithms”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 13.** (multiobjective OR “multi objective”) AND (“ant colony” OR “ant colony optimization” OR “ant colony optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 14.** (multicriteria OR “multi criteria”) AND (“ant colony” OR “ant colony optimization” OR “ant colony optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 15.** (“many objective” OR “many criteria”) AND (“ant colony” OR “ant colony optimization” OR “ant colony optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 16.** (multiobjective OR “multi objective”) AND (“artificial bee colony” OR “bee colony” OR “artificial bee colony optimization”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 17.** (multicriteria OR “multi criteria”) AND (“artificial bee colony” OR “bee colony” OR “artificial bee colony optimization”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 18.** (“many objective” OR “many criteria”) AND (“artificial bee colony” OR “bee colony” OR “artificial bee colony optimization”) AND

(“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 19.** (multiobjective OR “multi objective”) AND (“bee colony optimization” OR “artificial bee colony optimisation” OR “bee colony optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 20.** (multicriteria OR “multi criteria”) AND (“bee colony optimization” OR “artificial bee colony optimisation” OR “bee colony optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 21.** (“many objective” OR “many criteria”) AND (“bee colony optimization” OR “artificial bee colony optimisation” OR “bee colony optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 22.** (multiobjective OR “multi objective”) AND (“particle swarm” OR “particle swarm optimization” OR “particle swarm optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 23.** (multicriteria OR “multi criteria”) AND (“particle swarm” OR “particle swarm optimization” OR “particle swarm optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 24.** (“many objective” OR “many criteria”) AND (“particle swarm” OR “particle swarm optimization” OR “particle swarm optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 25.** (multiobjective OR “multi objective”) AND (“artificial immune” OR “artifical immune system” OR “artifical immune systems”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 26.** (multicriteria OR “multi criteria”) AND (“artificial immune” OR “artifical immune system” OR “artifical immune systems”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 27.** (“many objective” OR “many criteria”) AND (“artificial immune” OR “artifical immune system” OR “artifical immune systems”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 28.** (multiobjective OR “multi objective”) AND (“artificial immune optimization” OR “artificial immune optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 29.** (multicriteria OR “multi criteria”) AND (“artificial immune optimization” OR “artificial immune optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 30.** (“many objective” OR “many criteria”) AND (“artificial immune optimization” OR “artificial immune optimisation”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 31.** (multiobjective OR “multi objective”) AND (“simulated annealing” OR “tabu search” OR “hill climbing”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 32.** (multicriteria OR “multi criteria”) AND (“simulated annealing” OR “tabu search” OR “hill climbing”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 33.** (“many objective” OR “many criteria”) AND (“simulated annealing” OR “tabu search” OR “hill climbing”) AND (“feature selection” OR “attribute selection” OR “attribute reduction” OR “dimensionality reduction”)

**Sub-expressão 34.** (multiobjective OR “multi objective”) AND (metaheuristic OR “meta heuristic” OR “metaheuristic algorithm”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 35.** (multicriteria OR “multi criteria”) AND (metaheuristic OR “meta heuristic” OR “metaheuristic algorithm”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 36.** (“many objective” OR “many criteria”) AND (metaheuristic OR “meta heuristic” OR “metaheuristic algorithm”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 37.** (multiobjective OR “multi objective”) AND (“meta heuristic algorithm” OR “metaheuristic algorithms” OR “meta heuristic algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 38.** (multicriteria OR “multi criteria”) AND (“meta heuristic algorithm” OR “metaheuristic algorithms” OR “meta heuristic algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 39.** (“many objective” OR “many criteria”) AND (“meta heuristic algorithm” OR “metaheuristic algorithms” OR “meta heuristic algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expressão 40.** (multiobjective OR “multi objective”) AND (“metaheuristic search algorithm” OR “meta heuristic search algorithm” OR “metaheuristic search algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 41.** (multicriteria OR “multi criteria”) AND (“metaheuristic search algorithm” OR “meta heuristic search algorithm” OR “metaheuristic search algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 42.** (“many objective” OR “many criteria”) AND (“metaheuristic search algorithm” OR “meta heuristic search algorithm” OR “metaheuristic search algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 43.** (multiobjective OR “multi objective”) AND (“meta heuristic search algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 44.** (multicriteria OR “multi criteria”) AND (“meta heuristic search algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 45.** (“many objective” OR “many criteria”) AND (“meta heuristic search algorithms”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 46.** (multiobjective OR “multi objective”) AND (“ant colony” OR “ant colony optimization” OR “ant colony optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 47.** (multicriteria OR “multi criteria”) AND (“ant colony” OR “ant colony optimization” OR “ant colony optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 48.** (“many objective” OR “many criteria”) AND (“ant colony” OR “ant colony optimization” OR “ant colony optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 49.** (multiobjective OR “multi objective”) AND (“artificial bee colony” OR “bee colony” OR “artificial bee colony optimization”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 50.** (multicriteria OR “multi criteria”) AND (“artificial bee colony” OR “bee colony” OR “artificial bee colony optimization”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 51.** (“many objective” OR “many criteria”) AND (“artificial bee colony” OR “bee colony” OR “artificial bee colony optimization”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 52.** (multiobjective OR “multi objective”) AND (“bee colony optimization” OR “artificial bee colony optimisation” OR “bee colony optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 53.** (multicriteria OR “multi criteria”) AND (“bee colony optimization” OR “artificial bee colony optimisation” OR “bee colony optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 54.** (“many objective” OR “many criteria”) AND (“bee colony optimization” OR “artificial bee colony optimisation” OR “bee colony optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 55.** (multiobjective OR “multi objective”) AND (“particle swarm” OR “particle swarm optimization” OR “particle swarm optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 56.** (multicriteria OR “multi criteria”) AND (“particle swarm” OR “particle swarm optimization” OR “particle swarm optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 57.** (“many objective” OR “many criteria”) AND (“particle swarm” OR “particle swarm optimization” OR “particle swarm optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 58.** (multiobjective OR “multi objective”) AND (“artificial immune” OR “artifical immune system” OR “artifical immune systems”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 59.** (multicriteria OR “multi criteria”) AND (“artificial immune” OR “artifical immune system” OR “artifical immune systems”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 60.** (“many objective” OR “many criteria”) AND (“artificial immune” OR “artifical immune system” OR “artifical immune systems”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 61.** (multiobjective OR “multi objective”) AND (“artificial immune optimization” OR “artificial immune optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 62.** (multicriteria OR “multi criteria”) AND (“artificial immune optimization” OR “artificial immune optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 63.** (“many objective” OR “many criteria”) AND (“artificial immune optimization” OR “artificial immune optimisation”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 64.** (multiobjective OR “multi objective”) AND (“simulated annealing” OR “tabu search” OR “hill climbing”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 65.** (multicriteria OR “multi criteria”) AND (“simulated annealing” OR “tabu search” OR “hill climbing”) AND (“feature subset selection” OR “feature ranking”)

**Sub-expresão 66.** (“many objective” OR “many criteria”) AND (“simulated annealing” OR “tabu search” OR “hill climbing”) AND (“feature subset selection” OR “feature ranking”)

- Wiley Interscience

Q1. *expressão original:* 12 publicações

Q2. *expressão original:* 10 publicações

Q3. *expressão original:* 5 publicações

- CiteSeerX

**Q1.** *expressão original:* 21 publicações

**Q2.** *expressão original:* 8 publicações

**Q3.** *expressão original:* 1 publicação

Ao todo, foram identificados 3464 publicações, sendo 1574, 1302 e 588 referentes a primeira, segunda e terceira questões de pesquisa, respectivamente. Após a exclusão de trabalhos duplicados com mesmo título para cada questão de pesquisa, obteve-se 2203 ( $1012 + 815 + 376$ ) publicações. Quando este critério foi aplicado simultaneamente para todas as questões de pesquisa, reduziu-se de 2203 para 1132 a quantidade de publicações identificadas ao todo.

O procedimento de seleção de trabalhos prosseguiu com a leitura do título e resumo de todas as 1132 publicações para a aplicação dos critérios de exclusão. Foram mantidos 544 trabalhos para uma leitura mais detalhada do restante do trabalho, até que a seleção pudesse ser realizada indubitavelmente. Desse modo, obteve-se 68 publicações para a avaliação de qualidade. O formulário utilizado conjuntamente para a coleta dos critérios de qualidade e a extração das informações das publicações selecionadas pode ser solicitada aos autores deste relatório.

A validação dos critérios de qualidade foi realizada em 11 publicações, representando aproximadamente 16% dos trabalhos selecionados, para constatar que não houve nenhum erro no procedimento de avaliação dos trabalhos.

Embora a síntese das publicações selecionadas tenha sido basicamente qualitativa, optou-se também por observar o ano em que cada trabalho foi publicado, em busca de alguma tendência no interesse pelo tema pesquisado. O gráfico de tendência obtido, ilustrado na Figura 1, demonstra um crescimento na quantidade de trabalhos desenvolvidos sobre aplicações de metaheurísticas multiobjetivo para SA.

O estudo das 68 publicações avaliadas na RS permitiu sintetizar algumas informações gerais:

1. Apenas cinco trabalhos utilizam algum teste estatístico para reforçar as evidências identificadas;
2. A configuração experimental mais usual em trabalhos envolvendo AGM foi população = 100 indivíduos; taxa de *crossover* = 0,8; taxa de mutação = 0,01 e condição de parada = 1000 gerações;
3. Alguns conjuntos de dados usuais são o Ionosphere, Sonar, Vehicle, Dermatology e Wine, os quais pertencem ao repositório da UCI;
4. Poucos trabalhos investigam conjuntos de dados com quantidade extensa de atributos;
5. A inicialização randômica de indivíduos é relativamente usual;
6. O algoritmo *Non-Dominated Sorting Genetic Algorithm II* (NSGA-II) é o AGM mais usual;

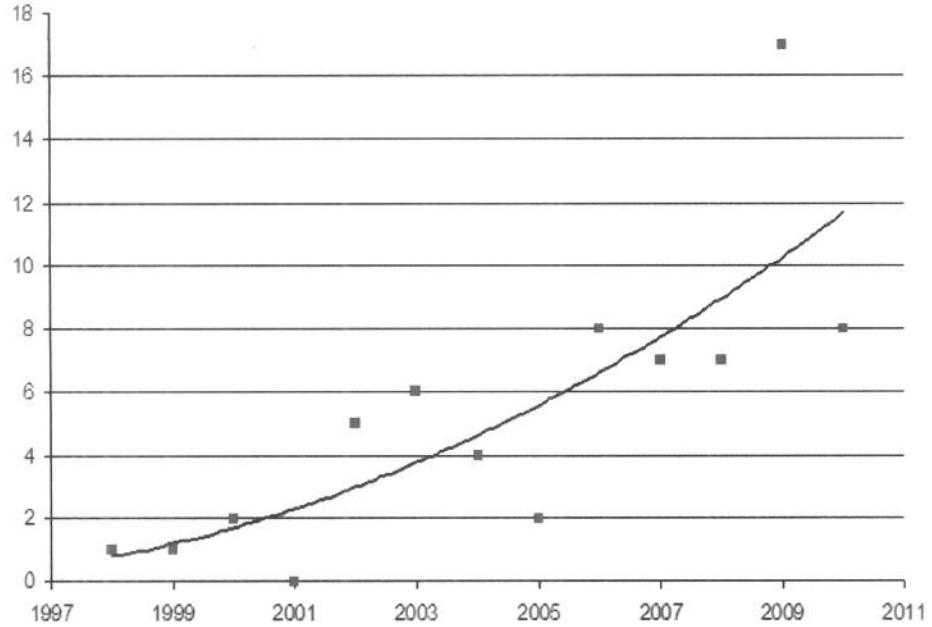


Figura 1: Tendência para quantidade de trabalhos × ano de publicação.

7. As estratégias de OM por soma ponderada e baseada em Pareto são as mais usuais.

Uma síntese mais elaborada das oito publicações selecionadas que aplicam AGM para SA conforme a abordagem filtro é apresentada a seguir. Destacou-se nesta síntese as medidas de importância de atributos utilizadas, visto que a partir de algumas delas implementou-se o método utilizado na dissertação relacionada à esta RS.

1. Com o objetivo de ranquear fatores de risco relacionados a nascimentos prematuros, aplica-se em [43] um algoritmo NSGA-II que relaciona, por meio da estratégia de Pareto, as medidas Distância Intra-classe, Distância Inter-classe e Correlação Atributo Classe. O conjunto de dados, obtido por meio de um hospital, foi submetido a procedimentos como normalização dos valores de atributos e eliminação prévia de atributos com alta redundância. O melhor resultado atingido no estudo de caso realizado foi uma taxa de acerto de 83,05%, obtido com a classificação de uma projeção construída com os quatro melhores atributos selecionados, a qual é superior às taxas atingidas com o uso de todos os atributos e com outras técnicas de SA. Projeções com quantidades diferentes de atributos também se destacaram em relação ao *baseline*.
2. Propõe-se em [44] o uso do algoritmo NSGA-II para auxiliar no problema de classificação de imagens do tipo hiperespectral. As medidas de Discriminação entre Classes e Invariância Espacial dos Atributos foram otimizadas simultaneamente por meio da estratégia de Pareto. O conjunto de imagens investigado foi processado antes da aplicação do AGM de modo a eliminar dados ruidosos obtidos com sensores descalibrados. Em [45], este trabalho é estendido para tratar tanto o contexto

supervisionado quanto o semi-supervisionado, em que apenas alguns exemplos do conjunto de dados possuem rótulos. Os resultados obtidos em ambos os trabalhos, em termos do coeficiente Kappa [46], em geral demonstram um desempenho superior dos métodos propostos em relação a AG mono-objetivo.

3. A tarefa de detecção de anomalias em pacotes TCP/IP pode apresentar melhor eficiência com a aplicação de AGM para SA, conforme realizado em [47]. A estratégia de OM por soma ponderada é aplicada para relacionar as medidas denominadas *scoring* de anormalidade e *scoring* de comunicação. Em três classificadores distintos, a aplicação proposta resulta na construção de modelos com melhor taxa de acerto e menores taxas de falso positivo e falso negativo [46] que os modelos construídos sem o uso do AGM.
4. O algoritmo NSGA-II é aplicado em [48] para auxiliar na classificação de perfis de expressões gênicas coletados por *microarray*. A natureza destes dados, que em geral possuem poucos exemplos e muitos atributos, torna a tarefa de classificação mais complexa, o que motiva o uso da SA. A Cardinalidade e a Capacidade de Discernimento entre Exemplos são as medidas de importância de atributos utilizadas no trabalho, otimizadas conjuntamente por meio da estratégia de Pareto. Os dados são previamente processados com procedimentos como a normalização e a eliminação de genes ambíguos. Experimentalmente, verificou-se em termos de taxa de acerto um desempenho superior da proposta em relação a um AG mono-objetivo e a outras técnicas para SA, incluindo um modelo com taxa de acerto de 100%.
5. As medidas de importância de atributos Inter-Correlação e Intra-Correlação, propostas em [49], são aplicadas a conjuntos de dados para análise de risco de crédito. Neste contexto, utiliza-se o algoritmo NSGA-II e, portanto, a estratégia de OM baseada na dominância de Pareto. Os dados são submetidos a um processamento inicial para normalização e outros procedimentos. Constatou-se experimentalmente, com a construção de modelos a partir de distintos classificadores analisados com diferentes critérios de avaliação, que o AGM permitiu atingir um desempenho melhor do que modelos gerados com todos os atributos e gerados com atributos selecionados por técnicas como um AG mono-objetivo e Relief [50].
6. Uma aplicação do NSGA-II para SA de dados relacionados a expressões faciais é identificada em [51]. As medidas de importância relacionadas pela estratégia de Pareto correspondem à Cardinalidade e ao Critério Fisher. A avaliação experimental realizada demonstra uma taxa de acerto de 88,18%, a qual é superior à atingida por trabalhos correlatos.
7. Em [52] é descrito um estudo sobre o uso de AGM para selecionar atributos a serem usados em combinações (*ensembles*) heterogêneas de classificadores e estimular a diversidade entre os distintos classificadores utilizados. Para atingir este objetivo, o trabalho utiliza como medidas de importância a Correlação Inter-Atributos e a Correlação Intra-Atributos dos classificadores. Ao contrário dos trabalhos anteriores, um teste estatístico é aplicado para evidenciar um desempenho significativamente

superior do método proposto em relação a muitos modelos construídos com todos os atributos.

Ainda dentre as 68 publicações investigadas, constatou-se a ocorrência de três trabalhos que utilizam outros algoritmos heurísticos, como *Binary Particle Swarm Optimization* e Programação Genética, para a OM de medidas de importância da abordagem filtro [53, 54, 55].

## 4 Considerações Finais

Neste relatório técnico foi apresentado uma Revisão Sistemática realizada para o estudo de aplicações de metaheurísticas multiobjetivo para o problema de Seleção de Atributos. Após uma contextualização sobre este método de pesquisa bibliográfica, apresentou-se o protocolo utilizado e os resultados obtidos com a aplicação de cada um dos procedimentos propostos. Embora o enfoque principal deste trabalho seja em Computação, algumas contextualizações utilizam exemplos de Revisões Sistemáticas em Medicina, visto que esta é uma das áreas em que esse método é mais tradicionalmente considerado.

A RS apresenta-se como um método de pesquisa bibliográfica interessante, dentre outros motivos, por permitir uma exploração ampla, rigorosa e reproduzível da literatura. Apesar de exigir um esforço relativamente maior que outros processos de pesquisa equivalentes, constata-se que este esforço foi recompensado pelos resultados e considerações apresentados neste relatório e pela consequente investigação relevante sobre o tema de pesquisa abordado.

Convém ressaltar como uma contribuição deste trabalho a adaptação de procedimentos e protocolos utilizados em RS e sugeridos em [5] na disciplina de Engenharia de Software para um contexto de Inteligência Artificial. O protocolo utilizado pode ser adaptado para RS em temas de pesquisa relacionados.

## Agradecimentos

À UFABC, à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, à Fundação de Amparo à Pesquisa do Estado de São Paulo e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo apoio recebido para a realização deste trabalho.

## Referências

- [1] J. Wainer, “Métodos de pesquisa quantitativa e qualitativa para a ciência da computação,” in *Atualizações em Informática*. Sociedade Brasiliense de Computacao e PUC rio, 2007, pp. 221–262.
- [2] M. Balnaves and P. Caputi, *Introduction to Quantitative Research Methods - An Investigative Approach*. SAGE Publications, 2001.

- [3] J. Ritchie and J. Lewis, *Qualitative Research Practice - A Guide for Social Science Students and Researchers.* SAGE Publications, 2003.
- [4] A. A. Castro, H. Saconato, F. Guidugli, and O. A. C. Clark, "Curso de revisão sistemática e metanálise [online]," <http://www.virtual.epm.br/cursos/metanalise>, 2002.
- [5] B. A. Kitchenham, "Guidelines for performing systematic literature reviews in software engineering," Evidence-based Software Engineering., Tech. Rep., 2007.
- [6] K. Pearson, "Report on certain enteric fever inoculation statistics," *British Medical Journal*, pp. 1243–1246, 1904.
- [7] H. K. Beecher, "The powerful placebo," *Journal of American Medical Association*, vol. 159, pp. 1602–1606, 1955.
- [8] J. P. T. Higgins and S. Green, "Cochrane handbook for systematic reviews of interventions," The Cochrane Collaboration. Acesso em: Acessa <http://www.cochrane-handbook.org/>, 2009.
- [9] M. Petticrew and H. Roberts, *Systematic Review in the Social Sciences: A Practical Guide.* Wiley-Blackwell, 2006.
- [10] J. Biolchini, P. G. Mian, A. C. C. Natali, and G. H. Travassos, "Systematic review in software engineering," COPPE/UFRJ, Tech. Rep., 2005.
- [11] N. Spolaôr, A. C. Lorena, and H. D. Lee, "Um estudo sobre a aplicação de algoritmos genéticos multiobjetivo para a seleção de atributos," in *Anais do IV Congresso da Academia Trinacional de Ciências*, 2009, pp. 1–10.
- [12] A. Jamain and D. J. Hand, "Mining supervised classification performance studies: A meta-analytic investigation," *Journal of Classification*, vol. 25, pp. 87–112, 2008.
- [13] M. Pai, M. McCulloch, J. D. Gorman, N. Pai, W. Enanoria, G. Kennedy, P. Tharyan, and J. M. J. Colford, "Systematic reviews and meta-analyses: An illustrated, step-by-step guide," *The National Medical Journal of India*, vol. 17, pp. 84–95, 2004.
- [14] F. D. Lorijn, L. C. M. Kremer, J. B. Reitsma, and M. A. Benninga, "Diagnostic tests in hirschsprung disease: A systematic review," *Journal of Pediatric Gastroenterology and Nutrition*, vol. 42, no. 5, pp. 496–505, 2006.
- [15] P. Chiarelli, W. Bower, J. Attia, D. Sibbrit, and A. Wilson, "The prevalence of faecal incontinence: a systematic review," Australian Government Department of Health and Ageing. Acesso em: <http://www.bladderbowel.gov.au/doc/FaecalIncontinenceReview.pdf>, Tech. Rep., 2002.
- [16] A. E. Chapman, P. Hewett, G. Kiroff, G. J. Maddern, B. Geerdes, T. Eyers, and J. Young, "Systematic review of dynamic graciloplasty

for the treatment of faecal incontinence," Australian Safety and Efficacy Register of New Interventional Procedures - Surgical. Acesso em: <http://www.surgeons.org/Content/ContentFolders/News/ASERNIPS/SystematicReviews/DGreview0601.pdf>, Tech. Rep., 2001.

- [17] E. Engström, P. Runeson, and M. Skoglund, "A systematic review on regression test selection techniques," *Information and Software Technology*, vol. 52, pp. 14–30, 2010.
- [18] S. Ali, L. Briand, H. Hemmati, and R. K. Panesar-Walawege, "A systematic review of the application and empirical investigation of search-based test-case generation," *IEEE Transactions on Software Engineering*, pp. 1–22, 2010.
- [19] S. Ali, L. Briand, H. Hemmati, and R. K. P. Walawege, "A systematic review of the application and empirical investigation of search-based test-case generation," Simula Research Laboratory. Acesso em: [http://simula.no/research/se/publications/Simula.SE.293/simula.pdf\\_file](http://simula.no/research/se/publications/Simula.SE.293/simula.pdf_file), Tech. Rep., 2009.
- [20] C. S. C. Rodrigues and C. M. L. Werner, "Uma revisão sistemática sobre as iniciativas realizadas no ensino de arquitetura de software," COPPE/UFRJ, Tech. Rep., 2009.
- [21] K. J. Stol, M. A. Babar, B. Russo, and B. Fitzgerald, "The use of empirical methods in open source software research - facts, trends and future directions," in *International Conference on Software Engineering*, 2009, pp. 19–24.
- [22] J. Nicolás and A. Toval, "On the generation of requirements specifications from software engineering models: A systematic literature review," *Information and Software Technology*, vol. 51, pp. 1291–1307, 2009.
- [23] W. Afzal, R. Torkar, and R. Feldt, "A systematic review of search-based testing for non-functional system properties," *Information and Software Technology*, vol. 51, pp. 957–976, 2009.
- [24] A. N. Health and M. R. Council, "How to review the evidence: systematic identification and review of the scientific literature," Australian Government. Acesso em: <http://www.nhmrc.gov.au/publications/synopses/cp65syn.htm>, Tech. Rep., 1999.
- [25] Elsevier, "Scopus," Acesso em: <http://www.scopus.com>, 2010.
- [26] A. for Computing Machinery, "The acm portal," Acesso em: <http://portal.acm.org/>, 2010.
- [27] I. of Electrical and E. Engineers, "Ieee xplore," Acesso em: <http://ieeexplore.ieee.org>, 2010.
- [28] Elsevier, "Sciencedirect," Acesso em: <http://www.sciencedirect.com>, 2010.
- [29] U. S. N. L. of Medicine, "Pubmed," Acesso em: <http://www.ncbi.nlm.nih.gov/pubmed>, 2010.

- [30] A. G. Fink, *Conducting Research Literature Reviews - From the Internet to Paper*. Sage Publication, 2004.
- [31] G. W. Noblit and R. D. Hare, *Meta-Ethnography: Synthesizing Qualitative Studies*. Sage Publications, 1988.
- [32] L. T. Bui and S. Alam, "An introduction to multiobjective optimization," in *Multi-Objective Optimization in Computational Intelligence: Theory and Practice*. Information Science Reference, 2008, pp. 1–19.
- [33] C. A. C. Coello, "Evolutionary multi-objective optimization: a historical view of the field," *Comp. Intellig. Magazine, IEEE*, pp. 28–36, 2006.
- [34] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. Chapman & Hall/CRC, 2008.
- [35] ———, *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers, 1998.
- [36] T. C. of Information Sciences and Technology, "Citeseerx," Acesso em: <http://citeseerx.ist.psu.edu>, 2010.
- [37] Elsevier, "Scirus - for scientific information," Acesso em: <http://www.scirus.com>, 2010.
- [38] Springer, "Springerlink beta," Acesso em: <http://www.springerlink.com>, 2010.
- [39] J. W. . Sons, "Wiley interscience," Acesso em: <http://www.interscience.wiley.com>, 2010.
- [40] T. Reuters, "Isi web of knowledge," Acesso em: [http://apps.isiknowledge.com/WOS\\_GeneralSearch\\_input.do?highlighted\\_tab=WOS&product=WOS&last\\_prod=WOS&SID=4AbGKcDe4f2MkGolfDO&search\\_mode=GeneralSearch](http://apps.isiknowledge.com/WOS_GeneralSearch_input.do?highlighted_tab=WOS&product=WOS&last_prod=WOS&SID=4AbGKcDe4f2MkGolfDO&search_mode=GeneralSearch), 2010.
- [41] I. K. Crombie, *The Pocket Guide to Appraisal*. British Medical Journal Books, 1996.
- [42] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, vol. 80, no. 4, pp. 571–583, 2007.
- [43] D. Zaharie, S. Holban, D. Lungceanu, and D. Navolan, "A computational intelligence approach for ranking risk factors in preterm birth," in *International Symposium on Applied Computational Intelligence and Informatics*, 2007, pp. 135–140.
- [44] L. Bruzzone and C. Persello, "A novel approach to the selection of robust and invariant features for classification of hyperspectral images," in *IEEE International Geoscience and Remote Sensing Symposium*, 2008, pp. 66–69.

- [45] ——, “A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability,” *IEEE transactions on geoscience and remote sensing*, vol. 47, pp. 3180–3191, 2009.
- [46] F. Guillet and H. J. Hamilton, *Quality Measures in Data Mining*. Springer-Verlag, 2007.
- [47] T. Shon, X. Kovah, and J. Moon, “Applying genetic algorithm for classifying anomalous tcp/ip packets,” *Neurocomputing*, vol. 69, pp. 2429–2433, 2006.
- [48] M. Banerjee, S. Mitra, and H. Banka, “Evolutionary rough feature selection in gene expression data,” *IEEE Transactions on Systems Man and Cybernetics*, vol. 37, no. 4, pp. 622–632, 2007.
- [49] C.-M. Wang and Y.-F. Huang, “Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 5900–5908, 2009.
- [50] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.
- [51] U. Tekguc, H. Soyel, and H. Demirel, “Feature selection for person-independent 3d facial expression recognition using nsga-ii.” in *International Symposium on Computer and Information Sciences*, 2009, pp. 35–38.
- [52] L. E. A. Santana, L. Silva, and A. M. P. Canuto, “Feature selection in heterogeneous structure of ensembles: a genetic algorithm approach,” in *international joint conference on Neural Networks*, 2009, pp. 1491–1498.
- [53] D. Niu, Y. Wang, and D. D. Wu, “Power load forecasting using support vector machine and ant colony optimization,” *Expert Systems with Applications*, vol. 37, no. 3, pp. 2531–2539, 2010.
- [54] K. Iswandy and A. Koenig, “Feature selection with acquisition cost for optimizing sensor system design,” *Advances in Radio Science*, vol. 4, pp. 135–141, 2006.
- [55] ——, “Feature-level fusion by multi-objective binary particle swarm based unbiased feature selection for optimized sensor system design,” in *Multisensor Fusion and Integration for Intelligent Systems, 2006 IEEE International Conference on*, 2006, pp. 365–370.