Universidade Federal da Paraíba
Centro de Informática
Programa de Pós-Graduação em Informática

# Gerenciamento de Faltas em Computação em Nuvem: Um Mapeamento Sistemático de Literatura

Clodoaldo Brasilino Leite Neto

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Informática da Universidade Federal da Paraíba como parte dos requisitos necessários para obtenção do grau de Mestre em Informática.

Área de Concentração: Informática
Linha de Pesquisa: Computação Distribuída

Alexandre Nóbrega Duarte

João Pessoa, Paraíba, Brasil

# Resumo

Fundamentação: Com o grande crescimento da popularidade da computação em nuvens, observa-se que um desafio dessa área é gerenciar falhas que possam ocorrer nas grandes infra-estruturas computacionais construídas para dar suporte à computação como serviço. Por serem extensas, possuem maior ocorrência de faltas, falhas e erros. Um trabalho que mapeie as soluções já criadas por pesquisadores de maneira simples e eficiente pode ajudar a visualizar oportunidades e saturações nesta área de pesquisa.

Objetivos: Este trabalho visa mapear de forma sistemática todo o esforço de pesquisa aplicado ao gerênciamento de faltas em computação em nuvem de forma a facilitar a identificação de áreas pouco exploradas e que pode eventualmente representar novas oportunidades de pesquisa.

Metodologia: Para este trabalho utiliza-se a metodologia de pesquisa baseada em evidências, através do método de mapeamento sistemático, sendo a pesquisa construída em três etapas de seleção de estudos.

Método: Conduzimos um mapeamento sistemático para coletar, filtrar e classificar trabalhos científicos na área. Foram inicialmente coletados 4535 artigos científicos nos grandes engenhos de busca que, após três etapas de filtragem, acabaram sendo reduzidos a 166 artigos. Estes artigos restantes foram classificados de acordo com a taxonomia definida neste trabalho.

Resultados: Observa-se que IaaS é a área mais explorada nos estudos selecionados. As funções de gerência de falhas mais exploradas são Tolerância e Remoção, e os atributos são Confiabilidade e Disponibilidade. A maioria dos trabalhos foram classificados como tipo de pesquisa de Proposta de Solução.

Conclusão: Este trabalho sumariza e classifica o esforço de pesquisa conduzido em Gerenciamento de Faltas em Computação em Nuvens, provendo um ponto de partida para pesquisas futuras nesta área.

**Palavras-chave:** computação em nuvens, gerenciamento de faltas, mapeamento sistemático, pesquisa baseada em evidências, estudo secundário.

# Abstract

Background: With the growing popularity of cloud computing, an challenge seen in this discipline is the management of faults that may occur in such big infrastructures that, because of its size, has a greater chance of ocurring faults, errors and failures. A work that maps the solutions already created by researchers efficiently should help visualizing gaps over those research fields.

Aims: This work aims to find research gaps in the cloud computing and fault management domains, aside from building a social network of researchers in the area.

Method: We conducted a systematic mapping study to collect, filter and classify scientific works in this area. The 4535 scientific papers found on major search engines were filtered and the remaining 166 papers were classified according to a taxonomy described in this work.

Results: We found that IaaS is most explored in the selected studies. The main dependability functions explored were Tolerance and Removal, and the attributes were Reliability and Availability. Most papers had been classified by research type as Solution Proposal.

Conclusion: This work summarizes and classifies the research effort conducted on fault management in cloud computing, providing a good starting point for further research in this area.

**Keywords:** Cloud Computing, Fault Management, Dependability, Mapping Study, Evidence-Based Research, Secondary Study.

# Agradecimentos

O primeiro agradecimento é a Deus, inteligência suprema e causa primeira de todas as coisas.

Agradeço ao meu orientador Alexandre Nóbrega Duarte pela ajuda fornecida. Sem a sua orientação e dedicação, não seria capaz de abrir minha mente ao ponto de construir este trabalho.

Agradeço a minha esposa, Cláudia Virgínia Albuquerque Prazim da Silva. Sua ternura, carinho e apoio emocional foram essenciais para que eu me mantivesse no caminho.

Agradeço a meu pai e meus irmãos pelo incentivo ao término desta jornada e especialmente minha mãe que, além de dar apoio, ainda serviu de exemplo pela vida acadêmica que levou e me inspirou a seguir o mesmo.

Agradeço aos colegas de mestrado pela companhia durante a jornada e no auxílio durante os estudos. Especialmente agradeço a Pedro Batista de Carvalho Filho, que foi parte integrante desta pesquisa e que além de bom colega, é um bom amigo.

Agradeço ao CNPq pelo apoio financeiro quando me foi necessário.

Agradeço aos meus colegas de trabalho na DATAPREV e no Tribunal de Justiça. A compreensão de que a construção deste trabalho cansa e os conselhos dados ficaram marcados nesta trajetória.

Agradeço a todos os amigos e familiares que compreenderam que o mestrado é bastante exigente do nosso tempo e souberam ter paciência por minha ausência.

# Conteúdo

# Lista de Símbolos

**TI** : *Tecnologia da Informação*

**PC** : *Computador Pessoal (Personal Computer)*

**NIST** : *National Institute of Standards and Technology*

**SaaS** : *Software como Serviço(Software as a Service)*

**PaaS** : *Plataforma como Serviço (Platform as a Service)*

**IaaS** : *Infraestrutura como Serviço (Infrastructure as a Service)*

**SLA** : *Acordo de Nível de Serviço (Service-Level Agreement)*

**SLR** : *Revisão Sistemática de Literatura (Sistematic Literature Review)*

**MS** : *Mapeamento Sistemático (Mapping Study)*

**CSV** : *Valores Separados por Vírgula (Comma-Separated Values)*

# Lista de Figuras

# Lista de Tabelas

# Capítulo 1

# Introdução

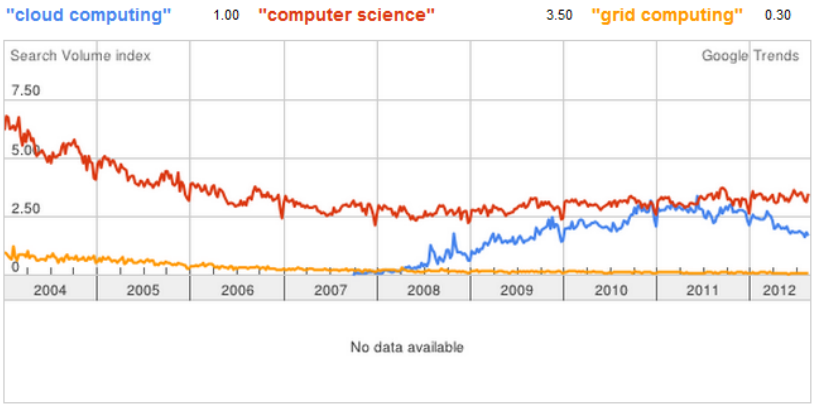
Este capítulo apresenta uma introdução ao mapeamento sistemático em computação em nuvem e gerenciamento de faltas, e está dividido nas seguintes seções: Seção 1.1 apresenta uma problematização e motivação para construção desse estudo; Seção 1.2 apresenta os objetivos gerais e específicos do trabalho; e Seção 1.3 apresenta como este trabalho está estruturado.

## 1.1 Motivação

Em 1961, o pesquisador John McCarthy trouxe à tona em um discurso no MIT Centennial uma proposta de modelo de distribuição de recursos computacionais semelhante ao utilizado para distribuir energia ou gás [Garfinkel 1999]. Este modelo, chamado Utility Computing, leva em consideração o uso de recursos computacionais sob demanda. Apesar de parecer promissor, já que nessa época tínhamos a computação realizada de maneira centralizada, acabou por perder sua relevância após o surgimento e popularização dos PC's da década de 80. [Reimer 2005] mostra o crescimento das unidades de PC's comercializadas da década de 70 até os dias atuais.

Mas à partir de 2006, de acordo com a figura 1.1, observa-se uma tendência crescente na computação distribuída em relação a esse modelo de computação sob demanda, mas desta vez com uma nova nomenclatura: Computação em nuvem. A figura 1.1 apresenta três linhas de índice de volume de buscas no engenho de busca Google. O termo "Cloud Computing" é tão buscado quanto o termo "Computer Science" em 2010 e 2011, e isso pode representar uma grande demanda ou curiosidade por serviços de computação em nuvem. Como com-

Figura 1.1: Tendências de busca do termo 'Cloud Computing'. Fonte: Google Trends




paração, podemos observar que o termo "Grid Computing", que traz conceitos de alguma similaridade à computação em nuvem, nunca teve um destaque em buscas e pesquisas neste engenho desde 2004.

A computação em nuvem, segundo [NIST 2011], traz uma abordagem de alocação de recursos de maneira rápida e sob demanda e uma cobrança apenas por recursos que foram utilizados, de maneira sistemática e mensurável. Os benefícios desse modelo de computação são diretos tanto para a academia quanto para indústria, pois representam uma redução de custos operacionais, um menor Time to Market, a capacidade de utilizar e controlar grandes infraestruturas de maneira mais rápida e abstrata com auxílio da virtualização e facilidade para as partes contratante e contratada do serviço no estabelecimento de um SLA, definindo de maneira explícita aspectos da computação que antes eram meio nebulosos ou difíceis de controlar através de contratos.

Um aspecto importante que precisa ser bem tratado em contratos da computação em nuvem é a tolerância a falhas. Com uma grande infraestrutura, aumenta-se a ocorrência de faltas, erros e falhas, e com isso corre-se o risco de perder o controle do sistema dependendo da gravidade e da frequência de ocorrência dessas falhas. [Gong et al. 2010] corrobora com esse aspecto, informando que a tolerância a falhas é intrínseca e está implícita na computação em nuvem e que paradigmas anteriores como computação paralela (*Parallel Computing*) e grades computacionais (*Grid Computing*) colaboram para isso. Assim, mecanismos de prevenção, previsão, tolerância e correção de falhas que ocorrem em sistemas são necessários, de forma a garantir vários aspectos importantes como disponibilidade, segurança, confiabi-

lidade, confidencialidade, integridade e manutenibilidade de sistemas e serviços.

Muitos trabalhos têm abordado diferentes aspectos relacionados ao uso e operação de infraestruturas de computação em nuvem. De fato, esta é uma das "áreas quentes"de pesquisa em ciência da computação e isto tem contribuído para uma alta produção de resultados científicos materializados sob a forma de artigos em jornais e conferencias. [Zhang et al. 2010]

Essa afluência de resultados e artigos é muito positiva pois indica claramente que progressos estão sendo realizados na área mas acaba impondo aos pesquisadores um novo desafio na hora de identificar lacunas e potenciais problemas a serem pesquisados.

Com o amadurecimento da área, alguns estudos secundários (como os de [Barreiros 2011], [Carvalho 2012], [Elberzhager, Münch e Nha 2012], [Jacinto 2010], entre outros) têm surgido com o objetivo de auxiliar os pesquisadores a identificarem áreas de pesquisa bem atendidas, ou mesmo já saturadas, e lacunas que podem representar novas oportunidades de pesquisa ou potenciais limitações em uma determinada área.

Ao analisar a literatura específica não identificamos nenhum trabalho que apresente um panorama geral dos resultados científicos obtidos focados na gerência de faltas em infraestruturas de computação em nuvem.

## 1.2 Objetivos

### 1.2.1 Objetivo Geral

O objetivo geral deste trabalho é mapear, de forma sistemática, os resultados científicos produzidos com foco na gerência de faltas em infraestruturas de computação em nuvem de forma a identificar problemas bem resolvidos e lacunas que possam guiar novos esforços de pesquisa.

### 1.2.2 Objetivos Específicos

Dado o objetivo geral, podemos detalhá-lo nos seguintes tópicos:

- Obter um panorama dos tópicos explorados na interseção das áreas de pesquisa de computação em nuvem e gerenciamento de faltas, identificando pesquisadores, locais

de publicação, métodos e representatividade da academia e da indústria;

- Identificar lacunas e saturações em áreas de pesquisa da gerência de faltas em computação em nuvem;

- Prover uma visão de como se relacionam pesquisadores e grupos de pesquisa, identificando suas interações na autoria de artigos.

## 1.3 Estrutura da Dissertação

Esta dissertação está organizada na seguinte forma. No capítulo 2 temos uma breve fundamentação de termos e conhecimentos relevantes à pesquisa. No capítulo 3, temos a apresentação de trabalhos relacionados ao tipo de pesquisa realizada nessa dissertação. No capítulo 4, é explicitada a metodologia de pesquisa, com a apresentação do protocolo do mapeamento a ser realizado. O capítulo 5 apresenta dados coletados da pesquisa de forma a apresentar subsídios para argumentações. O capítulo 6 apresenta respostas às questões de pesquisa propostas, de forma a promover uma discussão sobre os resultados. O capítulo 7 traz uma breve conclusão do trabalho.

# Capítulo 2

# Fundamentação Teórica

## 2.1 Pesquisa baseada em evidências

No final da década de 80 e início da de 90 houve uma mudança dramática nas práticas de pesquisa médica, pelo menos para estudos clínicos, com a adoção de um paradigma baseado em evidências. Apesar da origem na medicina clínica, o paradigma baseado em evidências foi adotado subsequentemente por um número de domínios que envolvem atividades centradas no homem, como educação, áreas não-clínicas da saúde e biblioteconomia. [Budgen et al. 2006].

O objetivo principal da medicina baseada em evidências, segundo [Sackett et al. 2000] é ”a integração da melhor evidência de pesquisa com a perícia clínica e os valores dos pacientes”.

Após a consolidação desse paradigma de pesquisa na área médica, temos o início de estudos para a adoção da pesquisa baseada em evidências pela informática no início dos anos 2000, através da engenharia de software. As pesquisas nesse paradigma se intensificaram na engenharia de software com o surgimento do projeto Evidence-Based Software Engineering em abril de 2005 na Inglaterra. Segundo [Budgen et al. 2006], o objetivo principal deste projeto é "investigar a viabilidade da adoção do paradigma baseado em evidências para a engenharia de software".

Já segundo [Kitchenham, Dyba e Jorgensen 2004], o objetivo da pesquisa baseada em evidências é prover os meios pelos quais as melhores evidências atuais de pesquisa possam ser integradas com a experiência prática e valores humanos no processo decisório de acordo

com o desenvolvimento e manutenção do software. Para [Dyba, Kitchenham e Jorgensen 2005], o objetivo da pesquisa baseada em evidências, no caso da engenharia de software, é de aperfeiçoar a tomada de decisão relacionada ao desenvolvimento e manutenção de software.

A adoção dessa técnica de pesquisa tem crescido em um ritmo intenso. Estudos realizados por [Kitchenham et al. 2010] mostram que, de Janeiro de 2004 à Junho de 2007, uma pesquisa abrangente apresentou 35 novas revisões sistemáticas de literatura (RSL) correspondendo a 33 estudos únicos. [Silva et al. 2011] informa que dois estudos terciários sobre o assunto identificaram e analisaram 54 RSLs no mesmo período, e apresenta que de julho de 2008 à dezembro de 2009 mais novos 67 RSLs foram encontrados. Esses resultados apresentam um crescimento de 124% da quantidade de publicações no período de 1,5 anos, mostrando que o crescimento das pesquisas nessa área é intenso.

Esse crescimento intenso traz alguns benefícios para a área: atrai a atenção de mais pesquisadores para o paradigma baseado em evidências, fazendo com que mais publicações apareçam e com isso também se melhore a crítica e a qualidade em relação a esses trabalhos.

Em [Dyba, Kitchenham e Jorgensen 2005], é informado que são necessários 5 passos para a construção de pesquisa baseada em evidências:

- Converter o problema relevante ou a necessidade de informação em uma questão que possa ser respondida;
- Pesquisar a literatura para a melhor evidência disponível para responder a questão;
- Avaliar criticamente a evidência para sua validade, impacto e aplicabilidade;
- Integrar a evidência avaliada com a experiência prática e os valores e circunstâncias dos clientes para fazer decisões sobre a prática;
- Avaliar desempenho do trabalho e procurar maneiras de melhorá-lo.

Apesar do crescimento intenso do paradigma baseado em evidências na informática, esse crescimento é oriundo apenas da engenharia de software. A difusão de conhecimento empírico que é evidente se restringe à engenharia de software, trazendo a impressão de que outras áreas não publicam ou pouco publicam trabalhos de pesquisa baseada em evidências.

O que acontece na realidade é que a pesquisa baseada em evidências pode trazer alguns benefícios tanto para pesquisadores quanto para profissionais do mercado.[Dyba, Kitchenham e Jorgensen 2005] e [Kitchenham, Brereton e Budgen 2010] apresentam, entre elas:

- Trazer meios valiosos de análise de viabilidade de pesquisa para candidatos a doutorado ou pesquisadores;
- Prover a estudantes de graduação e pós-graduação habilidades de pesquisa que podem ser transferíveis através da realização de trabalhos de pesquisa baseada em evidências;
- Auxiliar na tomada de decisão ao apresentar melhores práticas, análises comparativas e oportunidades no tema pesquisado, servindo para o fomento de profissionais da área.

### 2.1.1 Mapeamentos sistemáticos e Revisões sistemáticas de literatura

Quando tratamos da pesquisa baseada em evidência, observamos o destaque de duas metodologias de pesquisa que ganharam grande visibilidade pela academia: os mapeamentos sistemáticos e as revisões sistemáticas de literatura.

Os dois tipos de trabalhos possuem grandes semelhanças. De acordo com [Kitchenham, Budgen e Brereton 2011], eles possuem a mesma metodologia básica. Mas apesar de parecidas, as duas metodologias possuem diferenças que devem ser esclarecidas de forma que não deixe dúvidas sobre a metodologia que esta dissertação trata.

Uma revisão sistemática de literatura padrão é dirigida por uma questão de pesquisa muito específica que pode ser respondida por pesquisa empírica. Esta questão de pesquisa dirige a identificação de estudos primários apropriados, informam os processos de extração de dados aplicados para cada estudo primário, e determina a agregação dos dados extraídos [Kitchenham, Budgen e Brereton 2011].

Em contraste, um mapeamento sistemático revisa um tópico mais abrangente e classifica estudos primários em um domínio específico. As questões de pesquisa para tal estudo são de alto nível e incluem pontos nos quais sub-tópicos podem ser atribuídos, que métodos empíricos foram usados, e que sub-tópicos tiveram suficientes estudos empíricos para apoiar uma revisão sistemática mais detalhada [Budgen et al. 2008].

De acordo com [Kitchenham, Budgen e Brereton 2011], ”Uma diferença importante é que uma Revisão Sistemática de Literatura (do inglês *Systematic Literature Review*, SLR)

convencional faz uma tentativa de agregar estudos primários nos termos das consequências da pesquisa e investiga se os resultados da pesquisa são consistentes ou contraditórios. Em contraste, um Mapeamento Sistemático (do inglês *Mapping Study*, MS) geralmente se foca em apenas classificar a literatura relevante e agregar estudos à respeito de categorias definidas.".

Outro aspecto importante é a diferença de escopo dos estudos: enquanto mapeamentos sistemáticos focam uma área abrangente, revisões sistemáticas de literatura tratam de tópicos detalhados de uma certa área [Kitchenham, Budgen e Brereton 2011].

## 2.2 Computação em Nuvem

A computação está sendo transformada em um modelo consistente de serviços oferecidos como commodities e disponibilizados de maneira similar a serviços tradicionais como água, eletricidade, gás e telefonia. Em tal modelo, usuários acessam serviços baseados em seus requisitos sem se preocupar onde esses serviços são hospedados ou como eles são entregues. Vários paradigmas de computação prometeram entregar essa visão de computação como utilitário e dentre eles se incluem computação em clusters, computação em grades e mais recentemente computação em nuvem [Buyya, Yeo e Venugopal 2008].

Segundo [Veras e Tozer 2012], "a ideia inicial da computação em nuvem foi processar as aplicações e armazenar os dados fora do ambiente corporativo, dentro da grande rede (Internet), em estruturas conhecidas como datacenters, otimizando o uso dos recursos".

Em [Veras e Tozer 2012], eles completam ainda a afirmação anterior informando que este é o conceito atual de nuvem pública. Eles afirmam também que adotar "a arquitetura de computação em nuvem significa mudar fundamentalmente a forma de operar a TI, saindo de um modelo baseado em aquisição de equipamentos para um modelo baseado em aquisição de serviços".

A nova arquitetura introduzida pela computação em nuvem permite que as organizações escolham o modelo adequado para a arquitetura dos seus aplicativos e onde armazenar os seus dados. Isto inclui aplicativos que rodam internamente, serviços públicos de nuvem e/ou serviços privados de nuvem [Veras e Tozer 2012].

A definição para o termo computação em nuvem , desde o surgimento do termo, não tem

sido coerente entre as várias que aparecem nas publicações. O NIST (*National Institute of Standards and Technology*), visando diminuir a discrepância entre as várias definições do termo, definiu em uma publicação especial termos e definições comuns para a computação em nuvem .

A computação em nuvem é um modelo para permitir acesso sob demanda, conveniente e ubíquo via rede a um conjunto compartilhado de recursos computacionais configuráveis (servidores, redes, armazenamento, aplicações e serviços) que podem ser rapidamente providos e liberados com o mínimo de esforço de gerenciamento ou de interação com o provedor do serviço. Este modelo de é composto de cinco características essenciais, três modelos de serviço e quatro modelos de implantação [NIST 2011].

Segundo [Veras e Tozer 2012], a proposta da computação em nuvem é criar a ilusão de que o recurso computacional é infinito e ao mesmo tempo permitir a eliminação do comprometimento antecipado da capacidade. Além disso, a ideia é permitir o pagamento pelo uso real dos recursos.

Como foi adotado o modelo do NIST de referência para definir a computação em nuvem , descreveremos suas características essenciais, modelos de serviço e modelos de implantação. Vale ressaltar que apesar do conflito de definições para o termo, os modelos e características apresentadas pelo modelo são bastante difundidos e aceitos pela literatura da área.

### 2.2.1 Características Essenciais

Conforme a definição apresentada anteriormente, o [NIST 2011] define cinco características principais para que se caracterize a presença de computação em nuvem :

- Autoatendimento sob demanda: um consumidor pode unilateralmente suprir capacidades computacionais - como tempo de servidor e armazenamento em rede - o quanto necessitar automaticamente sem requerer interação humana com cada provedor de serviço.

- Amplo acesso aos serviços de rede: recursos são disponibilizados pela rede e acessados por mecanismos padrão que promovem o uso por plataformas de clientes magros (*thin clients*) e/ou estações heterogêneas (smartphones, tablets, laptops, estações de trabalho, etc).

- Agrupamento de recursos: Os recursos do provedor são unificados para servir múltiplos consumidores usando um modelo de múltiplos inquilinos, com recursos virtuais e físicos atribuídos e reatribuídos dinamicamente de acordo com a demanda do consumidor. Existe um senso de independência de localização em que o consumidor geralmente não possui controle ou conhecimento da localização exata do recurso provido mas pode especificar a localização sob um nivel de abstração mais alto (como país, estado ou datacenter de execução, por exemplo). Exemplos dos recursos incluem armazenamento, processamento, memória e banda passante de rede.

- Rápida elasticidade: recursos podem ser elasticamente providos e liberados, em alguns casos automáticamente, para escalonar rapidamente para mais ou para menos proporcionalmente à demanda. Para o consumidor, os recursos disponíveis para serem providos as vezes aparentam ser ilimitados e podem ser apropriados por eles em qualquer quantidade a qualquer tempo.

- Serviço mensurável: sistemas em nuvem automaticamente controlam e otimizam o uso de recursos alavancando e medindo a capacidade sob algum nível de abstração de mensuração apropriada ao tipo do serviço. O uso do recurso pode ser monitorado, controlado e reportado, provendo transparência tanto ao provedor quanto ao consumidor do serviço utilizado.

Apesar de não ser listada diretamente pelo NIST em suas características principais, pode-se destacar uma grande tendência da arquitetura de a se aproximar de uma arquitetura orientada a serviços, envolvendo SLA's com as características apresentadas anteriormente, facilitando a geração de contratos e regulamentação de tais características. Não devemos aqui confundir a arquitetura orientada a serviços com SOA (Service-Oriented Architecture), que é um padrão de projeto de desenvolvimento de software. A definição de serviço aqui utilizada está relacionada ao termo da economia, em que o foco deixa de ser no bem (o equipamento de infraestrutura, o software) e passa a ser no serviço, que é o equivalente intangível de um bem.

### 2.2.2 Modelos de Serviço

Os modelos de serviço, de acordo com o [NIST 2011], são os seguintes:

- Software como Serviço (SaaS): O recurso provido ao consumidor são aplicações do provedor rodando sobre uma infraestrutura em nuvem. As aplicações são acessíveis por diversos dispositivos clientes através de uma interface de cliente magro, como um browser (por exemplo, webmail) ou uma interface de um programa. O consumidor não gerencia ou controla a infraestrutura de nuvem por baixo, incluindo rede, servidores, sistemas operacionais, armazenamento ou até recursos individuais de aplicações, com a possível exceção de configurações de aplicação específicas de usuários.

- Plataforma como Serviço (PaaS): O recurso provido ao consumidor permitem implantar em uma infraestrutura de nuvem aplicações compradas ou criadas pelo consumidor usando linguagens de programação, bibliotecas, serviços e ferramentas apoiadas pelo provedor. O consumidor não gerencia ou controla a infraestrutura de nuvem por baixo incluindo rede, servidores, sistemas operacionais ou armazenamento, mas tem total controle sobre as aplicações implantadas e as possíveis configurações para o ambiente de implantação da aplicação.

- Infraestrutura como Serviço (IaaS): Os recursos providos ao consumidor incluem a provisão de processamento, armazenamento, rede e outros recursos computacionais fundamentais onde o consumidor é capaz de implantar e rodar softwares arbitrários, os quais podem incluir sistemas operacionais e aplicações. O consumidor não gerencia ou controla a infraestrutura de nuvem fornecida mas tem total controle sobre os sistemas operacionais, armazenamento e aplicações implantadas, e possivelmente controle limitado dos componentes de rede selecionados (por exemplo, firewalls).

Apesar dos modelos de serviço apresentados acima abrangerem de maneira geral os recursos computacionais contratáveis, alguns modelos alternativos vão além dessas abstrações e levam em consideração abstrações de nível mais alto como suporte como serviço, por exemplo. Muitas dessas abstrações não são relevantes, não são de aceitação a nivel dos padrões apresentados e também não possuem definição sólida, e não serão descritas ou detalhadas no trabalho.

### 2.2.3 Modelos de Implantação

De acordo com o [NIST 2011], tem-se os seguintes modelos de implantação:

- Nuvem Privada: A infraestrutura de nuvem é provida para uso exclusivo por uma única organização compreendendo múltiplos consumidores. Ela pode ser de posse de, gerenciada e operada pela organização, por terceiro, ou por alguma combinação dos dois, e pode existir dentro ou fora das instalações da organização.

- Nuvem comunitária: A infraestrutura de nuvem é provida para uso exclusivo de uma comunidade específica de consumidores de organizações que possuem interesses em comum. Ela pode ser de posse de, gerenciada e operada por uma ou mais das organizações na comunidade, por um terceiro, ou por uma combinação dos dois, e pode existir dentro ou fora das instalações das organizações.

- Nuvem pública: A infraestrutura de nuvem é provida para uso aberto ao público geral. Ela pode ser de posse de, gerenciada e operada por uma organização acadêmica, de negócio ou governamental, ou alguma combinação delas. Ela existe nas dependências do provedor de nuvem.

- Nuvem híbrida: A infraestrutura de nuvem é uma composição de duas ou mais infraestruturas de nuvem distintas (privada, comunitária ou pública) de entidades únicas, mas que são unidas por uma tecnologia padronizada ou proprietária que permite portabilidade de aplicações e dados entre .

## 2.3 Gerenciamento de Faltas

A sociedade tornou-se cada vez mais dependente de sistemas computacionais e essa dependência é especialmente sentida sobre a ocorrência de falhas. As consequências de tais eventos estão relacionadas primariamente com a economia. No entanto, algumas interrupções podem levar a causar perigo a vidas humanas como efeito secundário, ou até diretamente [Laprie 1995].

Observa-se que [Laprie 1995] levanta o seguinte questionamento: ”Até onde podemos confiar em computadores?”. Essa pergunta leva a pensar em quais são os fatores de risco ao se confiar na capacidade dos computadores. Quais são os possíveis pontos de falha? Como e quando computadores podem falhar? Que tipos e magnitudes de prejuízo um computador pode ocasionar?

Com base no que foi dito, criou-se a definição de dependabilidade. Segundo [Laprie 1995], pode-se definir a dependabilidade como ”uma propriedade de um sistema computacional tal que a confiança pode ser justificadamente colocada sobre o serviço disponibilizado”

Segundo [Laprie 1995] e [Avizienis et al. 2004], a dependabilidade é composta por 6 atributos, 4 meios e 3 prejuízos. Os atributos são:

- Confiabilidade: está relacionada à continuidade do serviço;
- Disponibilidade: está relacionada à prontidão de uso;
- Segurança: está relacionada à não-ocorrência de catástrofes no ambiente;
- Confidencialidade: está relacionada à não ocorrência de divulgação não autorizada de dados/informações;
- Integridade: está relacionada à não ocorrência de alterações impróprias às informações;
- Manutenibilidade: está relacionada à habilidade se submeter-se a reparos e evolução.

Os meios para tratar tais atributos, segundo [Laprie 1995] e [Avizienis et al. 2004], podem ser classificads como:

- Prevenção de falhas: Como prevenir a ocorrência de uma falta;
- Tolerância à falhas: Como garantir o funcionamento do sistema na presença de faltas;
- Remoção de falhas: Como reduzir a presença (em quantidade ou gravidade) de faltas;
- Previsão de falhas: Como estimar a quantidade presente, a incidência futura e a consequência de faltas.

Já os tipos de prejuízos que esses meios tratam, podem ser definidos, segundo [Laprie 1995] e [Avizienis et al. 2004], como:

- Falha: ocorre quando um desvio do comportamento esperado de um serviço disponibilizado acontece;

- Erro: é um estado do sistema que pode levar a uma subsequente falha;
- Falta: é a causa, julgada ou hipotética, de um erro.

Para simplificar o processo de entendimento desses termos relacionados a prejuízos, [Lung 2012] exemplifica da seguinte maneira: ”Pense num atacante de futebol. A especificação dele é fazer gol. Se ele deixa de fazer gol, ele falha na sua especificação. Caso surja um zagueiro e cometa uma falta nele, essa falta vai fazer com que o atacante cometa um erro em sua trajetória em direção ao gol e, como consequência, falhe no seu objetivo (sua especificação) que é fazer o gol.”

## 2.4 Considerações Finais

Neste capítulo foram apresentados conceitos e definições relevantes sobre pesquisa baseada em evidências, computação em nuvem e gerenciamento de faltas. No próximo capítulo serão apresentados trabalhos relacionados à pesquisa baseada em evidências, mostrando casos de sucesso dessa modalidade de pesquisa.

# Capítulo 3

# A metodologia de mapeamento sistemático

O mapeamento sistemático é um método que fornece uma visão geral de uma área de pesquisa e permite identificar, quantificar e analisar os tipos de pesquisas e os resultados disponíveis [Petersen et al. 2008]. Segundo [Kitchenham et al. 2007], o mapeamento sistemático é uma revisão ampla dos estudos primários de uma determinada área que visa identificar as evidências disponíveis no tópico investigado. Portanto, um mapeamento sistemático é classificado como estudo secundário, já que, depende dos estudos primários utilizados para revelar evidências e construir conhecimento.

Algumas das razões que justificam a realização de um estudo de mapeamento são examinar a extensão, alcance e natureza dos fenômenos de investigação. É uma maneira útil para mapeamento de áreas de estudo, onde é difícil visualizar a gama de materiais que possam estar disponíveis; resumir e divulgar resultados de pesquisa. Esse tipo de estudo de escopo pode descrever com mais detalhes os resultados e o alcance da investigação em determinadas áreas de estudo, proporcionando assim um mecanismo de síntese e divulgação dos resultados da investigação; e identificar as lacunas de pesquisa na literatura existente. Este tipo de estudo de escopo leva a disseminação de conclusões, a partir da literatura existente, sobre o estado global da área investigada. [Petersen et al. 2008]

Da mesma forma que um estudo de mapeamento sistemático, outro estudo secundário existente é a revisão sistemática de literatura, conhecida como um dos principais métodos da pesquisa baseada em evidências na computação, além de ser inspirada na pesquisa médica

[Petersen e Ali 2011]. Conforme foi explicado no capítulo 2, diferente de uma revisão comum da literatura presente em qualquer projeto de pesquisa, na SLR existe uma estratégia de pesquisa explícita e critérios de aceitação, permitindo que as evidências pertinentes sejam consideradas de forma sistemática.

Em [Travassos 2007], confirma-se e acrescenta-se novas informações sobre o que deve ser contemplado na fase de planejamento através de alguns passos:

- Objetivos de pesquisa devem ser listados;
- Questões de pesquisa formuladas;
- Cadeias de busca criadas de acordo com as questões de pesquisa;
- Métodos que serão utilizados para a execução da busca e análise dos dados obtidos;
- Planejamento das fontes e seleções de estudos;
- Um protocolo deve ser definido, formalizado e publicado documentalmente;

Assim, este capítulo apresenta o protocolo de uma pesquisa de mapeamento sistemático, cujo objetivo principal é fornecer uma visão da distribuição de publicações nos mais renomados meios de publicação relacionados à computação em nuvem e o gerenciamento de faltas, possibilitando, à comunidade científica, novas estratégias de pesquisa sobre recursos de gerenciamento de faltas para ou com o uso da computação em nuvem, de acordo com a identificação das questões centrais e possíveis lacunas existentes.

As seções posteriores apresentarão informações relevantes na definição do protocolo do mapeamento sistemático a ser realizado.

## 3.1 Equipe

A equipe que construirá o mapeamento é composta por:

- Clodoaldo Brasilino Leite Neto

  Papel: Autor

  Contato: clodbrasilino@di.ufpb.br

  Afiliação: Discente do Programa de Pós-Graduação em Informática - UFPB

- Pedro Batista de Carvalho Filho

  Papel: Revisor

  Contato: pedro.filho.jp@gmail.com

  Afiliação: Discente do Programa de Pós-Graduação em Informática - UFPB

- Alexandre Nóbrega Duarte

  Papel: Árbitro

  Contato: alexandre@ci.ufpb.br

  Afiliação: Docente do Programa de Pós-Graduação em Informática - UFPB

## 3.2 Questões de Pesquisa

Com o objetivo de investigar quais são as saturações e lacunas das publicações da área de gerenciamento de faltas relacionada à computação em nuvem, como as publicações em gerenciamento de faltas relacionadas à computação em nuvem estão distribuídas e que grupos de pesquisa podem ser identificados, a pesquisa parte para seis questões de investigação mais específicas que auxiliam a responder essas perguntas:

QP1 Como as publicações sobre gerência de faltas em computação em nuvem estão distribuídas ao longo dos anos?

QP2 Quais são os principais meios de publicação (congressos, revistas, etc.)?

QP3 Qual é a representatividade da academia e da indústria neste campo de pesquisa?

QP4 Quais são os tópicos mais explorados na interseção entre computação em nuvem e gerência de faltas?

QP5 Que tipos de pesquisa são realizados nesse contexto?

QP6 Existem ligações entre pesquisadores/instituições através dos estudos? Existe uma rede de pesquisadores/instituições?

## 3.3 Estratégia de Busca

Segundo [Kitchenham et al. 2007], uma estratégia deve ser usada para a pesquisa dos estudos primários, com a definição das palavras chaves, bibliotecas digitais, jornais e conferências. A estratégia usada nessa pesquisa é apresentada nas próximas subseções.

### 3.3.1 Termos-chave da pesquisa

A construção dos termos de busca foi realizada seguindo a estratégia composta pelos seguintes passos:

- A partir dos domínios de pesquisa identificados, os principais termos (palavras-chave) são identificados;
- É realizada a tradução desses termos para o inglês por ser a língua utilizada nas bases de dados eletrônicas pesquisadas;
- A cadeia de busca é gerada a partir da combinação dessas palavras-chave e sinônimos. São usados os operadores OR (ou) entre os sinônimos identificados e variáveis de mesmo domínio e AND (e) entre as palavras-chave de domínios distintos.

Dois domínios principais são abordados neste trabalho. Para computação em nuvem, utilizamos termos que tratam de modelo de serviço, enquanto para a área de gerenciamento de falhas utilizamos termos referentes ao tipo de prejuízo. Os termos e sinônimos identificados destes domínios são apresentados abaixo:

- PC1 - Computação em nuvem: cloud computing, infrastructure as a service, IaaS, software as a service, SaaS, platform as a service, PaaS;
- PC2 - Gerenciamento de Faltas: error, failure, fault, fail;

Os itens de PC1 e PC2 serão unidos através do operador OR, enquanto a ligação dos domínios será feita através do operador AND. Dessa forma, teremos a cadeia de busca final no formato: (PC1 AND PC2)

A tabela 3.1 mostra como a cadeia de busca será construída para o trabalho baseada nas palavras selecionadas acima. Variações podem ocorrer para adaptação à sintaxe dos engenhos de busca utilizados.

Tabela 3.1: Cadeia de busca para as questões de pesquisa

| Cadeia de Busca |
|---|
| ("error"OR "failure"OR "fail"OR "fault") |
| AND |
| ("cloud computing"OR |
| "infrastructure as a service"OR "IaaS"OR |
| "platform as a service"OR "PaaS"OR |
| "software as a service"OR "SaaS") |

### 3.3.2 Fontes de busca

Segundo [Kitchenham et al. 2007], as pesquisas iniciais dos estudos primários podem ser realizadas em bibliotecas digitais. Por isso, foram pesquisadas fontes de busca relevantes para a área de informática para obtenção de um bom acervo inicial de trabalhos para a pesquisa.

Os critérios para a seleção das fontes foram: (1) disponibilidade de consultar os artigos na web através da rede da universidade; (2) presença de mecanismos de busca usando palavras-chave; e, (3) importância e relevância das fontes. As fontes de pesquisa utilizadas para a busca dos estudos primários são listadas abaixo:

FB1 IEEEXplore Digital Library (http://ieeexplore.ieee.org.ez15.periodicos.capes.gov.br/);

FB2 ACM Digital Library (http://dl.acm.org.ez15.periodicos.capes.gov.br/);

FB3 ScienceDirect (http://www.sciencedirect.com.ez15.periodicos.capes.gov.br/).

FB4 SpringerLink (http://www.springerlink.com.ez15.periodicos.capes.gov.br/);

Uma vez que potenciais estudos primários tenham sido obtidos, eles precisam ser analisados para que a sua relevância seja confirmada e trabalhos com pouca relevância sejam descartados. Tendo em vista isto, critérios de exclusão são definidos para ajudar na análise desses trabalhos nas próximas seções.

## 3.4 Seleção dos estudos

Os estudos que podem fazer parte dessa pesquisa são artigos em jornais, revistas, conferências, congressos e periódicos. Uma vez que estudos potencialmente candidatos a se tornarem estudos primários tenham sido obtidos, eles precisam ser analisados para que a sua relevância seja confirmada e trabalhos com pouca relevância sejam descartados. Segundo [Travassos 2007] os critérios de exclusão devem ser baseados nas questões de pesquisa. Logo, alguns critérios de exclusão devem ser definidos na próxima subseção, baseados nos trabalhos de [Kitchenham et al. 2007] e [Travassos 2007].

### 3.4.1 Critérios de exclusão

A partir da análise do titulo, palavra-chave, resumo e conclusão, serão excluídos os estudos que se enquadrem em alguns dos casos abaixo:

CE1 Trabalhos de conteúdo irrelevante em relação aos domínios de pesquisa (cada trabalho precisa envolver os dois domínios da pesquisa);

CE2 Trabalhos repetidos dentre as diferentes fontes de busca. Constatada sua igualdade, será adotado o trabalho da primeira pesquisa, sendo os próximos descartados;

CE3 Trabalhos que não possuem resumo, introdução, conclusão e pelo menos um capítulo intermediário serão considerados incompletos e serão excluídos;

CE4 Trabalhos não escritos na língua inglesa ou portuguesa;

CE5 Trabalhos que não sejam estudos primários;

## 3.5 Processo de seleção dos estudos primários

Segundo [Kitchenham et al. 2007], Após a definição das questões de pesquisa, da estratégia usada para a busca dos estudos primários e dos critérios de inclusão e exclusão, o processo de seleção dos estudos primários é descrito abaixo como etapas de seleção:

SE1 Nesta primeira etapa, o autor fará o processo de limpeza da grande quantidade de trabalhos coletados que, através de uma rápida visualização, se encaixam nos critérios

de exclusão CE1, CE2, CE3 e CE4, de maneira fácil e objetiva. Nesta etapa serão lidos e analisados título, palavras-chave e resumo. O objetivo nesta etapa é excluir trabalhos que são facilmente selecionados por qualquer busca no portal mas não possuem nenhuma relação com o conteúdo a ser analisado, trabalhos que são incompletos (introduções de workshops, anúncios de mesas de discussão, etc.) e que não são escritos nas línguas propostas, os quais são ininteligíveis aos autores, e ainda os trabalhos repetidos. Em casos de trabalhos selecionados explicitamente expostos como estudos secundários ou terciários, estes serão excluídos de acordo com o CE5 e catalogados. Os estudos rejeitados serão quantificados de forma a fornecer dados para análise no trabalho. Quando possível, os trabalhos excluídos serão classificados em 3 categorias: trabalhos da dimensão de computação em nuvem; trabalhos da dimensão de gerenciamento de faltas e trabalhos não-relacionados aos domínios propostos. Os estudos aceitos ou duvidosos ao autor serão incluídos na lista de candidatos LC1, lista essa que serve como entrada para a segunda etapa.

SE2 Com a lista de candidatos LC1 construída, dá-se início à segunda etapa de seleção de estudos, momento este que o revisor também atua nesta seleção. A partir da leitura do título, das palavras-chave, do resumo, da introdução e da conclusão dos trabalhos da LC1, o autor constrói a lista de candidatos do autor LC1A e o revisor constrói a lista de candidatos do revisor LC1R. Através de uma filtragem mais refinada e uma análise mais profunda, acredita-se na redução de trabalhos candidatos a serem selecionados. Neste momento, serão levados os critérios de exclusão CE1, CE3 e CE5. Os estudos rejeitados nessa fase serão catalogados na lista de trabalhos excluídos LE1, os trabalhos duvidosos serão cadastrados na lista de trabalhos duvidosos LD e os trabalhos incluídos serão inseridos na lista de trabalhos candidatos incluídos LC2.

SE3 Esta etapa consiste em catalogar os trabalhos da LC2 de acordo com as taxonomias e facetas definidas de forma a construir uma lista final de trabalhos devidamente classificados de acordo com os conhecimentos abordados, e com isso permitir o início da análise dos dados à partir das informações coletadas. Neste momento serão analisadas as mesmas seções que foram analisadas na etapa de seleção da SE2 (título, palavras-chave, resumo, introdução e conclusão), desta vez com o intuito de se definir como se

classificará o artigo de acordo com as facetas definidas no tópico ”Classificação dos Dados” a seguir. Ainda neste momento artigos podem ser rejeitados ou duvidosos, e estes irão fazer parte respectivamente das listas LE2 e LD2. Os trabalhos que forem incluídos farão parte da lista de trabalhos incluídos LI e estes estão garantidos de participarem dos quantitativos da pesquisa.

As etapas SE2 e SE3 podem sofrer intervenção do árbitro de forma a resolver impasses durante a pesquisa. Existem dois casos principais em que a intervenção do árbitro é necessária: uma quando há um impasse entre o autor e o revisor, ou seja, um deles decide incluir e o outro excluir o trabalho; outra quando o autor e o revisor forem duvidosos sobre um mesmo trabalho. O árbitro tem a responsabilidade de definir o voto de minerva, dando um veredicto sobre a inclusão, exclusão e classificação do trabalho.

Assumindo a igualdade de poder entre autor e revisor, segue uma relação dos casos de aceitação/rejeição:

- Caso ambos aceitem, o trabalho será incluído;
- Caso um aceite e o outro tenha dúvida, o trabalho será incluído;
- Caso um aceite e o outro rejeite, o trabalho será julgado pelo árbitro;
- Caso ambos tenham dúvida, o trabalho será julgado pelo árbitro;
- Caso um tenha dúvida e o outro rejeite, o trabalho será rejeitado;
- Caso ambos rejeitem, o trabalho será rejeitado.

## 3.6 Estratégia de extração dos dados

Para [Kitchenham et al. 2007], o objetivo desta etapa é criar formas de extração dos dados para registrar com precisão as informações obtidas a partir dos estudos primários. Esta deve ser projetada para coletar as informações necessárias às questões. Um formulário eletrônico será utilizado para facilitar a análise posterior. Logo, para apoiar a extração e registro dos dados e posterior análise, será utilizado um software de planilhas eletrônicas.

Tem-se neste trabalho 4 tabelas de apoio à pesquisa:

- Lista de candidatos (LC): é a lista onde aparecerão os trabalhos candidatos à inclusão na pesquisa. Destes trabalhos extrai-se o título, autores, ano, país de origem, instituição de origem, local de publicação, engenho de busca de origem e um comentário caso seja necessário.

- Lista de duvidosos (LD): é a lista onde aparecerão trabalhos que são prováveis candidatos, mas o autor/revisor não está certo de sua inclusão na pesquisa. Destes trabalhos extrai-se o título, a avaliação do autor, do revisor, do árbitro e o resultado do julgamento do trabalho, quando necessários.

- Lista de excluídos (LE): é a lista de trabalhos que foram excluídos nas etapas SE2 e SE3. Serão adicionados na tabela os seguintes dados: título, autores, ano, país de origem, instituição de origem, engenho de busca de origem e critério de exclusão utilizado, podendo ser adicionado também alguma anotação sobre a exclusão.

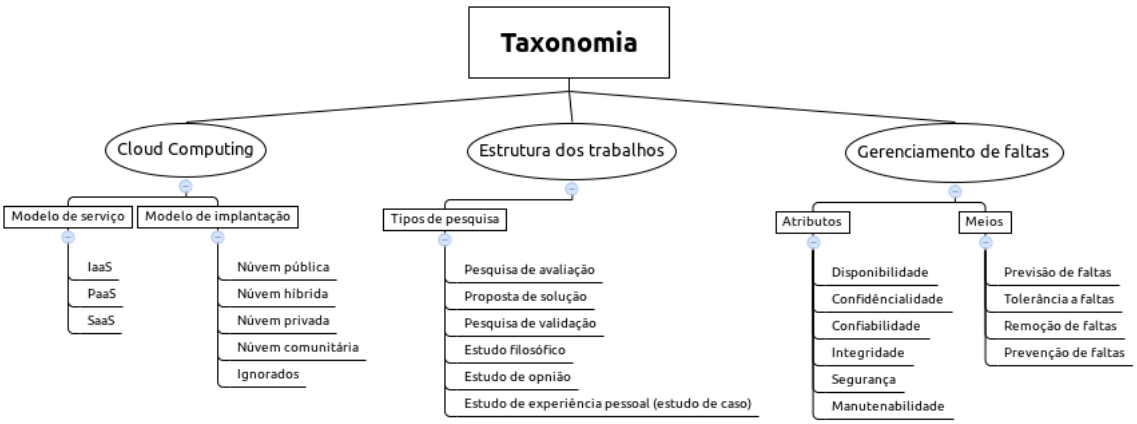
- Lista de Inclusos (LI): é a lista onde aparecerão os trabalhos que foram selecionados definitivamente para participarem da pesquisa. Esta lista de trabalhos é semelhante à lista de candidatos, mas esta inclui campos onde podem ser definidas as facetas das taxonomias utilizadas durante a classificação dos trabalhos.

## 3.7 Classificação dos dados

Após a extração dos dados, necessita-se da classificação dos dados para que seja possível realizar análises quantitativas sobre os trabalhos selecionados. Dessa forma, foram escolhidos para este trabalho três domínios para a classificação dos trabalhos: O domínio da computação em nuvem, o domínio do gerenciamento de faltas, e o domínio do tipo de pesquisa realizado. A figura 3.1 ilustra as facetas (categorias) e domínios da pesquisa. Vale ressaltar que as diferentes facetas não são mutualmente exclusivas e, excluída a de Tipo de Pesquisa, todas as outras podem ser multivaloradas a cada registro (trabalho).

Para o domínio da computação em nuvem, aplicaremos a taxonomia proposta pelo [NIST 2011], com duas facetas: a faceta de modelos de implantação (nuvem privada, pública, híbrida e comunitária) e a de modelos de serviço (software, plataforma ou infraestrutura como serviço).

Figura 3.1: Domínios e taxonomias utilizadas.




Para o domínio de gerenciamento de faltas, classificaremos os trabalhos nas facetas de atributos (confiabilidade, disponibilidade, segurança, confidencialidade, integridade e manutenabilidade) e meios de tratamento de falha (prevenção, tolerância, remoção e previsão de faltas) da dependabilidade definidos por [Avizienis et al. 2004].

Em relação ao domínio de estrutura do trabalho, utilizaremos a taxonomia definida por [Wieringa et al. 2005] para a classificação de tipos de trabalhos, de acordo com o encaixe dos trabalhos nos questionamentos propostos para cada uma das classes definidas. As classes definidas são pesquisa de avaliação, proposta de solução, pesquisa de validação, trabalhos filosóficos, trabalhos de opinião e trabalhos de experiência pessoal.

## 3.8 Síntese dos dados coletados

Com os dados extraídos e analisados, realiza-se uma síntese mais detalhada dos mesmos para a criação de mapas que representem o conhecimento gerado pela pesquisa. Essa síntese apresenta os tópicos de pesquisa que investigam a relação na pesquisa entre gerenciamento de faltas e computação em nuvem, os métodos de pesquisa usados por tal, a distribuição dos trabalhos e o envolvimento da academia e da indústria na área.

A síntese dos resultados foca em apresentar as frequências das publicações em cada categoria para tornar possível a visualização de quais dessas categorias têm sido enfatizadas em pesquisas passadas e assim, identificar lacunas e possibilidades para pesquisas futuras. Duas formas de mapa serão utilizadas para apresentação e síntese dos resultados: na forma

de tabelas de frequência que relacionam a evidência ou a categoria com a origem, através da referência ao estudo primário: na forma de gráficos como os de barras, pizza, os de bolhas, entre outros, que intersectam dados de duas ou mais categorias formando uma visão com diferentes facetas de combinação.

## 3.9 Apresentação do mapeamento

A última etapa consiste nos mapas de evidências que categorizam a área de pesquisa estudada e a avaliação dos mesmos. Os mapeamentos são apresentados de forma consistente com as questões de pesquisa, utilizando recursos como gráficos e tabelas, destacando similaridades e diferenças entre os resultados, isto é, ressalta as possíveis análises e combinação de dados.

# Capítulo 4

# Execução do Mapeamento e Resultados

Neste capítulo será apresentada a execução do protocolo de mapeamento proposto no capítulo 3, com o intuito de fornecer dados para responder as questões de pesquisa propostas. Cada seção deste capítulo apresentará as fases executadas do protocolo e seus resultados.

## 4.1 Busca e filtragem de estudos

De acordo com o protocolo definido no capítulo 3, foram realizadas as etapas de busca de trabalhos e filtragem de acordo com o processo de seleção de estudos definido. A seguir serão apresentados detalhes de cada etapa.

### 4.1.1 Resultados da busca

Para cada portal de busca definido (FB1, FB2, FB3 e FB4), foram realizadas consultas em busca de trabalhos de acordo com a cadeia de busca definida no capítulo 3.

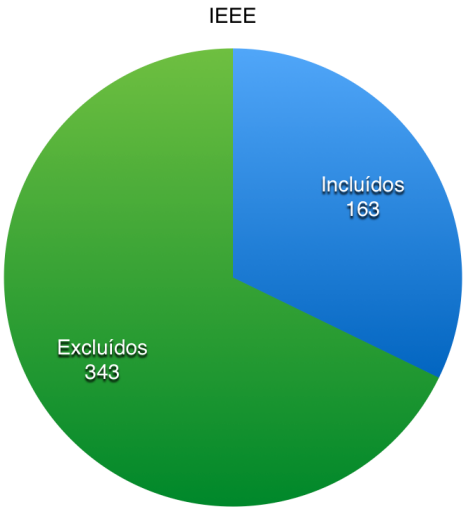
Ao executar a busca na FB1, obtivemos uma lista de 506 resultados. Já na FB2, a execução nos retornou 2405 resultados. Na FB3 a busca retornou 34 resultados e na FB4 apresentou 1590 resultados, totalizando 4535 resultados incluídos na pesquisa.

### 4.1.2 SE1 - Seleção de Estudos nº1

Na etapa de seleção SE1 4535 resultados foram analisados. A seguir apresenta-se um detalhamento para cada uma das fontes de busca definidas. Conforme a SE1 definida no capítulo

3, será preenchida neste momento a Lista de Candidatos LC1, e os trabalhos que não se encaixam nela serão descartados. Os trabalhos excluídos foram apenas contabilizados, em 3 categorias: Trabalhos Relacionados à Computação em Nuvem (TRCN), Trabalhos Relacionados à Gerência de Falhas (TRGF) e Trabalhos Não-Relacionados aos Domínios (TNR). As duas primeiras categorias se remetem a trabalhos que estão em apenas um dos domínios, enquanto a terceira se remete a textos que não apresentam contribuição a nenhum dos dois domínios.

Figura 4.1: Resultado FB1 SE1




Na FB1, conforme dito na subseção anterior, foram obtidos 506 resultados. Dessa relação, tivemos 163 trabalhos incluídos na LC1 e 343 excluídos, dentre estes 157 TRCN, 33 TRGF e 153 TNR. a figura 4.1 apresenta de forma gráfica a proporção incluídos/excluídos e a figura 4.5 apresenta o detalhamento comparativo com os outros engenhos de busca.

Na FB2 foram obtidos 2405 resultados. Dessa relação, tivemos 160 trabalhos incluídos na LC1 e 2245 excluídos, dentre estes 581 TRCN, 95 TRGF e 1569 TNR. a figura 4.2 apresenta de forma gráfica a proporção incluídos/excluídos e a figura 4.5 apresenta o detalhamento comparativo com os outros engenhos de busca.

Na FB3 foram obtidos 34 resultados. Dessa relação, tivemos 5 trabalhos incluídos na LC1 e 29 excluídos, dentre estes 14 TRCN, 2 TRGF e 13 TNR. a figura 4.3 apresenta de forma gráfica a proporção incluídos/excluídos e a figura 4.5 apresenta o detalhamento comparativo com os outros engenhos de busca.

Na FB4 foram obtidos 1590 resultados. Dessa relação, tivemos 51 trabalhos incluídos na

Figura 4.2: Resultado FB2 SE1

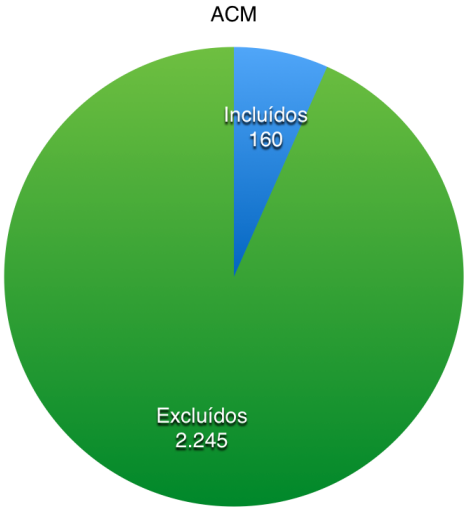



Figura 4.3: Resultado FB3 SE1

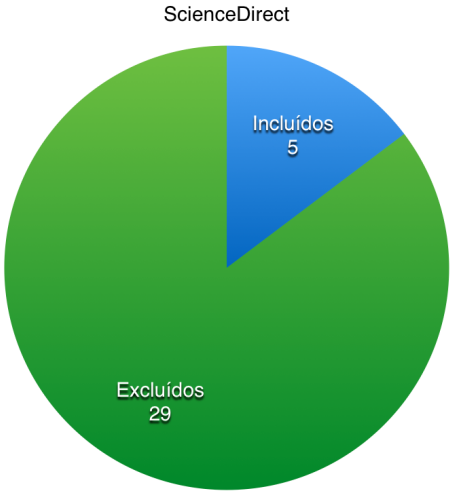
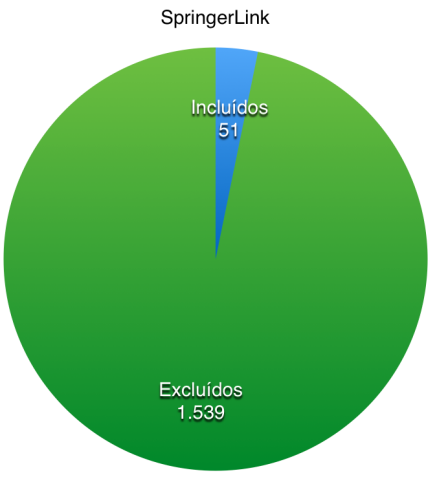
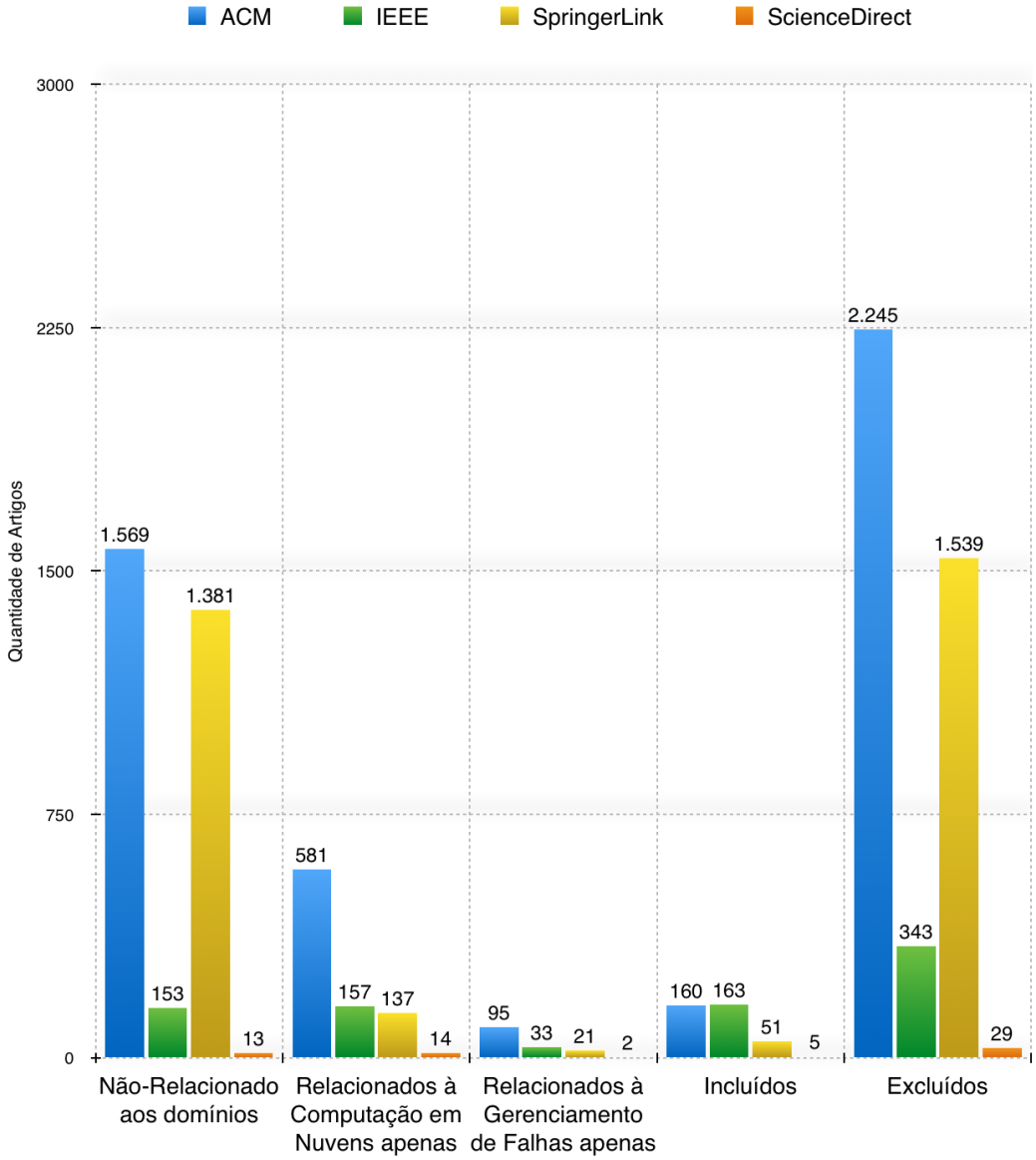



Figura 4.4: Resultado FB4 SE1



LC1 e 1539 excluídos, dentre estes 137 TRCN, 21 TRGF e 1381 TNR. a figura 4.4 apresenta de forma gráfica a proporção incluídos/excluídos e a figura 4.5 apresenta o detalhamento comparativo com os outros engenhos de busca.

Figura 4.5: Panorama Geral da SE1

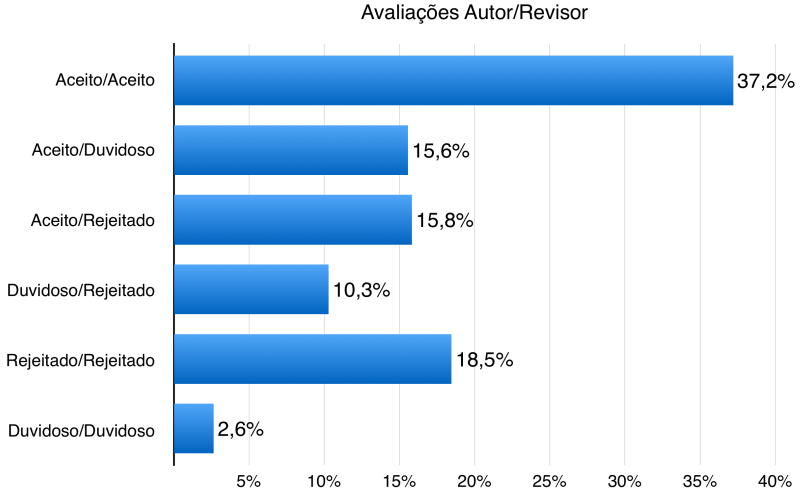



Observando a figura 4.5, podemos observar que ao final da execução da etapa SE1, dos 4535 trabalhos de entrada, tivemos um total de 379 trabalhos incluídos na LC1 e 4156 trabalhos excluídos da pesquisa, nos dando um percentual de 8,35% dos trabalhos de entrada sendo processados na segunda etapa.

### 4.1.3 SE2 - Seleção de Estudos nº2

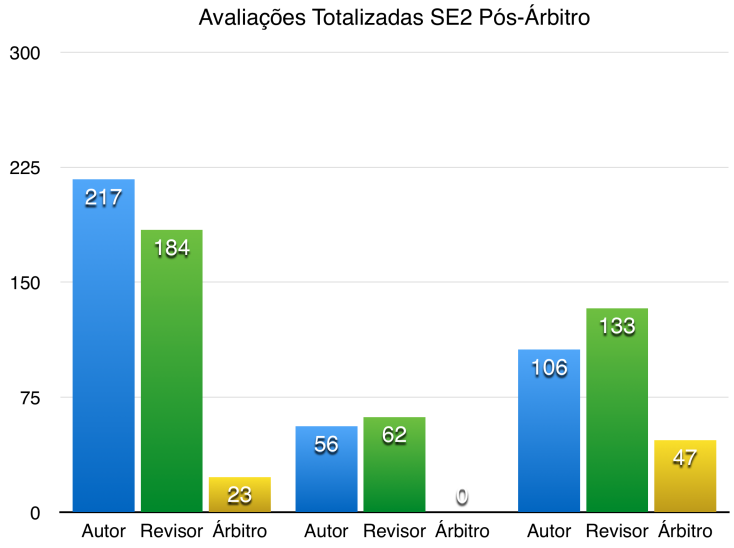
Na etapa de seleção SE2, foram avaliados os trabalhos da LC1 para a geração das seguintes listagens: Lista de Candidatos LC2, Lista de Duvidosos LD e Lista de Excluídos LE1. Como entrada desta etapa 379 trabalhos foram avaliados neste momento.

Figura 4.6: Avaliações da SE2 sem intervenção do árbitro




De acordo com as avaliações do autor, 106 trabalhos devem ser excluídos, 56 causaram dúvida em sua avaliação e 217 devem ser incluídos na pesquisa. Já o revisor avaliou os trabalhos da seguinte maneira: 133 trabalhos deveriam ser excluídos, 62 causaram dúvida e 184 deveriam ser incluídos na pesquisa. A figura 4.7 apresenta visualmetente estas totalizações.

Figura 4.7: Avaliações da SE2 com intervenção do árbitro




Com isso, podemos observar através da figura 4.7 ainda que:

- Foram considerados aceitos pelos dois autores 37,2% dos trabalhos, sendo um total de 141 trabalhos;
- 15,6% dos trabalhos foram duvidosos a um dos dois avaliadores e aceitos pelo outro, sendo aceitos para a próxima etapa da pesquisa (SE3), totalizando 59 trabalhos;
- 2,6% dos trabalhos foram duvidosos tanto ao autor quanto ao revisor, sendo neste momento responsabilidade do árbitro aceitá-los ou rejeitá-los, totalizando 10 artigos;
- 15,8% dos trabalhos foram rejeitados por um dos dois avaliadores e aceitos pelo outro, sendo neste momento responsabilidade do árbitro aceitá-los ou rejeitá-los, totalizando 60 trabalhos;
- 10,3% dos trabalhos foram duvidosos a um dos dois avaliadores e rejeitados pelo outro, sendo excluídos, totalizando 39 trabalhos;
- 18,5% trabalhos foram rejeitados pelo autor e pelo revisor, sendo excluídos, totalizando 70 trabalhos.

Com os dados sumarizados da execução da SE2 antes da intervenção do árbitro, podemos observar na figura 4.8 que dos 379 trabalhos de entrada da etapa, temos preliminarmente 200 trabalhos incluídos na LC2, 70 trabalhos a serem decididos pelo Árbitro na LD e 109 trabalhos excluídos aparecendo na LE1.

Com a LD agora preenchida, cabe ao árbitro realizar o desempate dos artigos, definindo quem será incluído na LE ou na LC2 dos artigos presentes na LD. Dos 70 trabalhos avaliados pelo árbitro, temos que 23 foram aceitos e 47 foram rejeitados, de acordo com a figura 4.7

Com os dados da avaliação do árbitro, observamos de acordo com a figura 4.9 que dos 379 artigos de entrada da SE2, temos como saída uma lista de 156 trabalhos excluídos da pesquisa e 223 trabalhos candidatos, que servirão de entrada para a etapa SE3.

### 4.1.4 SE3 - Seleção de Estudos nº3

Na etapa de seleção de estudos SE3, teremos a devida classificação dos artigos e seu descarte caso o trabalho não possa se encaixar nas seguintes premissas definidas no capítulo 3:

Figura 4.8: Resultado SE2 Pré-Árbitro

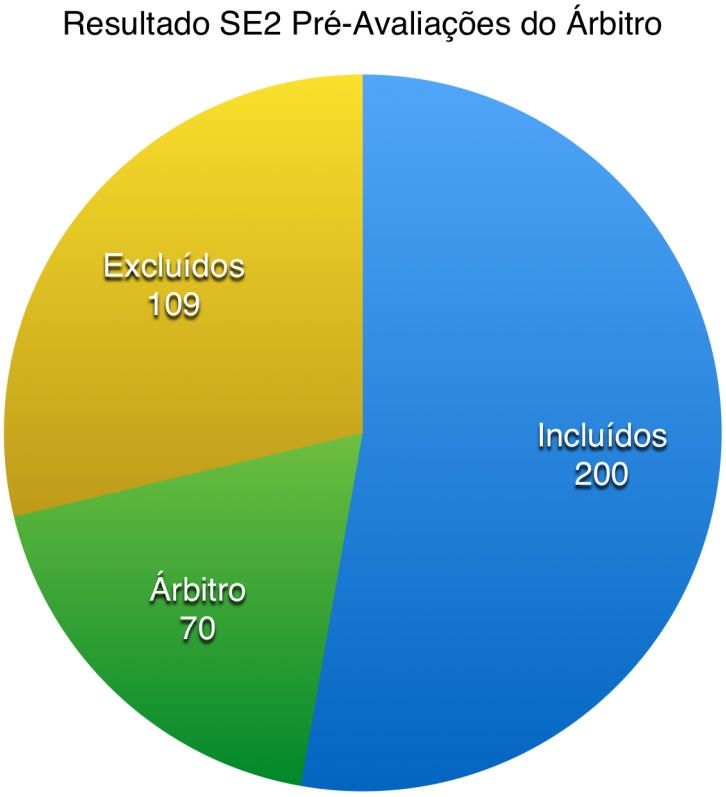
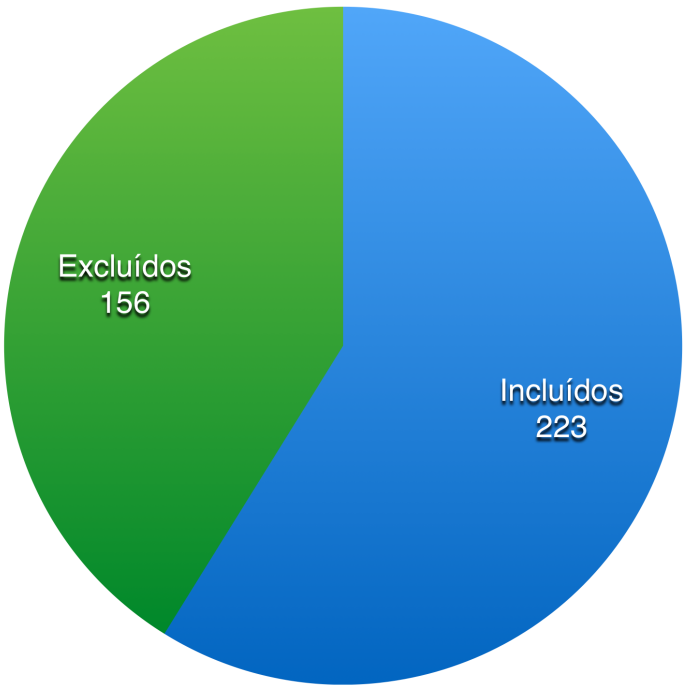



Figura 4.9: Resultado SE2

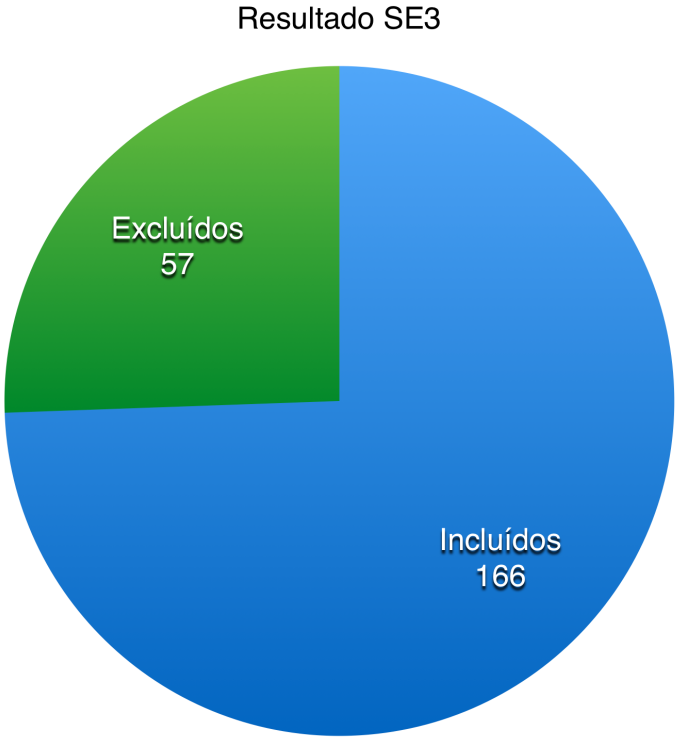


- O trabalho deve ser classificado em no mínimo uma categoria do domínio de computação em nuvem;
- O trabalho deve ter classificação válida na categoria de tipo de pesquisa;
- O trabalho deve ser classificado em no mínimo uma categoria do domínio de gerenciamento de falhas.

Com isso, podemos observar que dos 223 trabalhos incluídos provenientes da LC2, serão formadas duas listas: a lista de trabalhos inclusos LI e a lista de excluídos LE2.

A figura 4.10 mostra a saída gerada pela etapa SE3. Observamos que 166 trabalhos foram incluídos para análise e resposta das questões de pesquisa e 57 foram excluídos desta etapa.

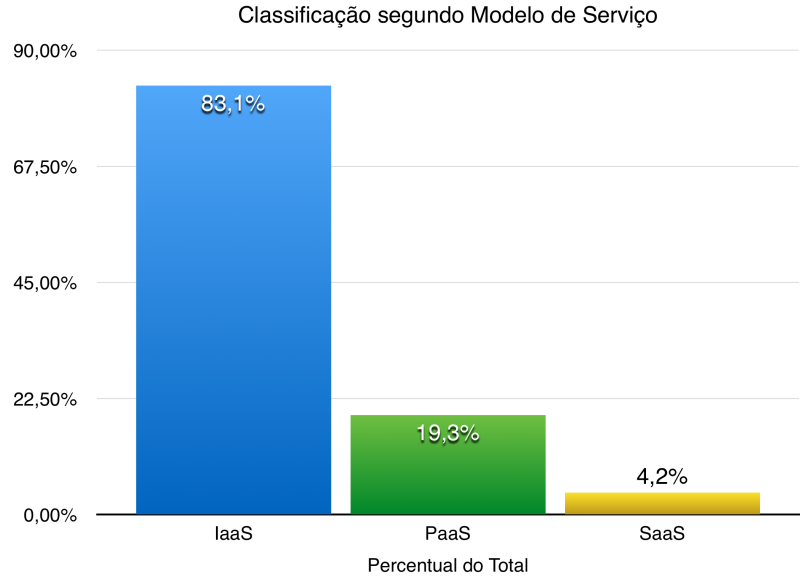
Figura 4.10: Resultado SE3 - Inclusão/Exclusão




Como nesta etapa ocorreu também a classificação dos artigos de acordo com as facetas, podemos observar os resultados específicos de cada classificação.

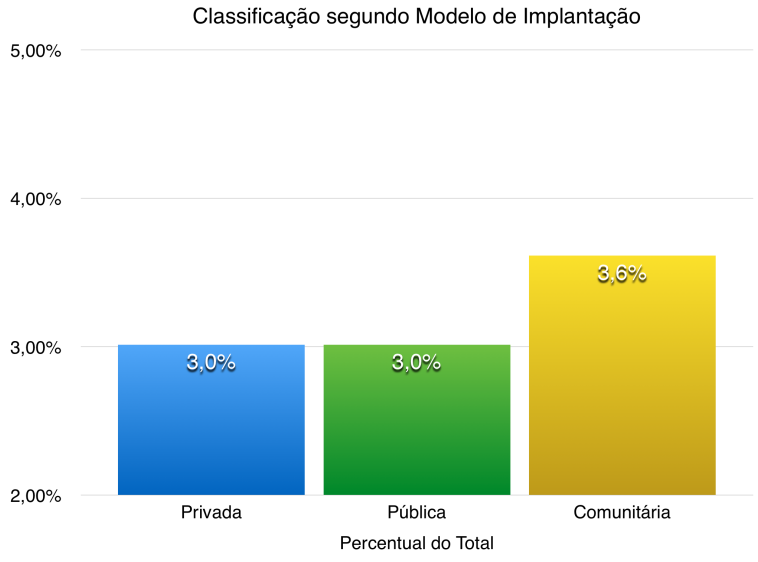
De acordo com a faceta de Modelo de Serviço, os artigos foram classificados da seguinte forma: 138 trabalhos foram classificados como IaaS, 32 como PaaS e 7 como SaaS. A figura 4.11 apresenta estes resultados como percentuais do total de artigos incluídos na pesquisa.

Figura 4.11: Resultado SE3 - Por modelo de serviço




Já de acordo com a faceta de Modelo de Implantação, os artigos foram classificados da seguinte forma: 5 trabalhos foram classificados como Nuvem Privada, 5 como Nuvem Pública e 6 como Nuvem Comunitária. A figura 4.12 apresenta estes resultados como percentuais do total de artigos incluídos na pesquisa.

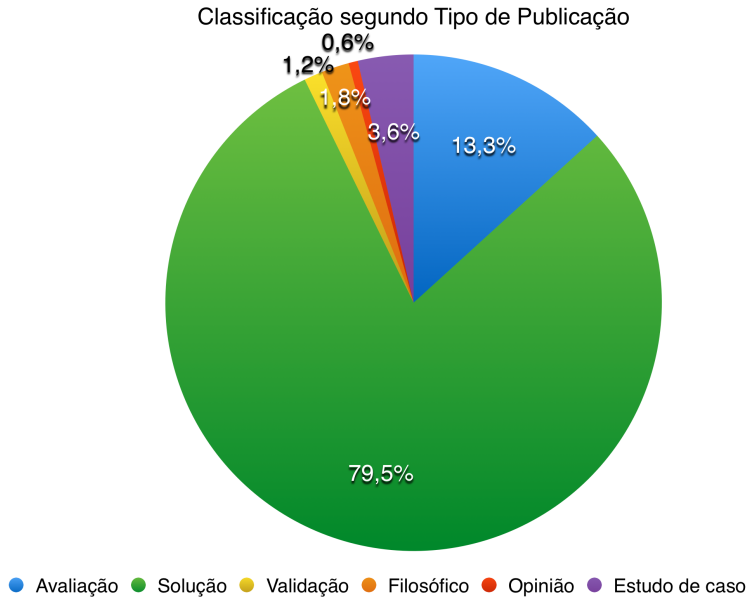
Figura 4.12: Resultado SE3 - Por modelo de implantação




Na faceta de Tipo de Pesquisa, os artigos foram classificados da seguinte forma: 22 trabalhos foram classificados como Avaliação, 132 como Solução, 2 como Validação, 3 como Filosófico, 1 como Opinião e 6 como Estudo de Caso. A figura 4.13 apresenta estes resulta-

dos como percentuais do total de artigos incluídos na pesquisa.

Figura 4.13: Resultado SE3 - Por tipo de publicação



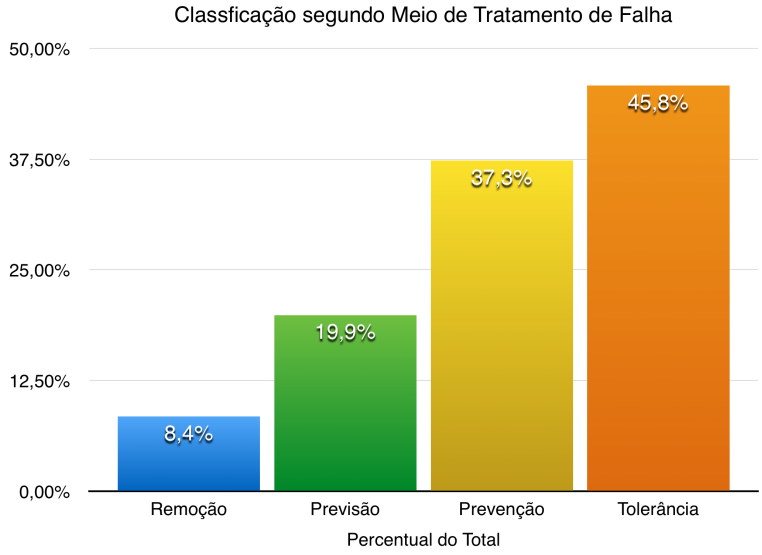

Na faceta de Atributo de Dependabilidade, os artigos foram classificados da seguinte forma: 75 trabalhos foram classificados como Disponibilidade, 87 como Confiabilidade, 7 como Confidencialidade, 24 como Manutenibilidade, 8 como Segurança e 23 como Integridade. A figura 4.14 apresenta estes resultados como percentuais do total de artigos incluídos na pesquisa.

Figura 4.14: Resultado SE3 - Por atributo de dependabilidade


Na faceta de Meio de Tratamento de Falha, os artigos foram classificados da seguinte forma: 14 trabalhos foram classificados como Remoção de Falhas, 33 como Previsão de

Falhas, 62 como Prevenção de Falhas e 76 como Tolerância a Falhas. A figura 4.15 apresenta estes resultados como percentuais do total de artigos incluídos na pesquisa.

Figura 4.15: Resultado SE3 - Por meio de tratamento de falha




## 4.2 Considerações Finais

Neste capítulo foram apresentados a aplicação prática da metodologia de pesquisa adotada, revelando detalhes da execução do mapeamento sistemático e os dados coletados nas etapas SE1, SE2 e SE3. O próximo capítulo apresenta respostas às questões de pesquisa com os dados coletados dos artigos.

# Capítulo 5

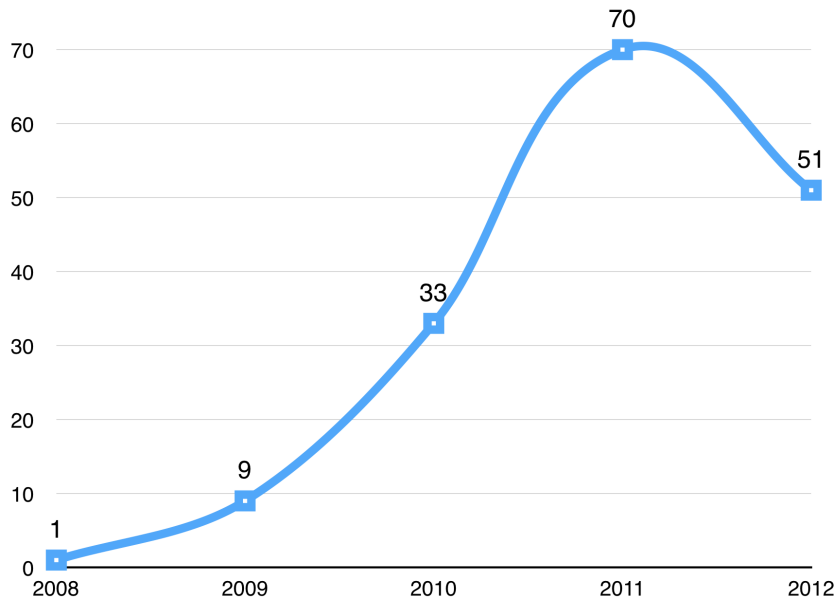
# Discussão dos resultados

Este capítulo apresentará respostas às questões de pesquisa propostas pelo protocolo definidas no capítulo 3, trazendo a discussão de dados obtidos.

## 5.1 QP1 - Como as publicações sobre gerência de faltas em computação em nuvens estão distribuídas ao longo dos anos?

Os primeiros estudos publicados sobre Gerenciamento de Faltas em Computação em Nuvem, segundo nossa pesquisa, surgiram em meados de 2008, com um crescimento acentuado até 2011. Em 2012 houve um decrescimento na quantidade de trabalhos publicados na área. Como a fase de busca dos estudos foi conduzida em Fevereiro de 2013, decidimos não incluir os resultados de 2013. Ainda é muito cedo para afirmar que existe uma migração dos esforços de pesquisa para outras áreas mas nós investigaremos este tópico no futuro de forma a identificar que isso seja uma tendência ou apenas um evento isolado. A figura 5.1 ilustra a distribuição das publicações de Gerenciamento de Faltas em Computação em Nuvens através dos anos.

Figura 5.1: Distribuição dos estudos por ano




## 5.2 QP2 - Quais são os principais meios de publicação (congressos, revistas, etc.)?

Foram identificados resultados incluídos na pesquisa em 114 conferências e jornais diferentes. A tabela 5.1 apresenta os 11 maiores meios de publicação identificados. O meio de publicação mais popular é o *ACM Symposium on Cloud Computing*, com 5 publicações. Observa-se que não há nenhum meio de publicação em destaque, provavelmente porque o assunto ainda é relativamente novo. A lista na tabela 5.1 representa apenas 24,1% do total de estudos.

A figura 5.2 apresenta a distribuição de artigos pelos engenhos de busca. *IEEEXplore* indexou 85 (51%) dos estudos selecionados, possuindo a maior retenção. O *IEEEXplore* também atingiu a maior precisão através das etapas de busca, incluindo cerca de 16,8% dos estudos obtudos desde a Busca de Estudos até a etapa SE3, seguido por *ScienceDirect* com 8,82%, *ACM* com 2,74% e *SpringerLink* com 0,75%. As figuras 5.3 e 5.4 apresentam esses dados graficamente, um com valores absolutos de frequência e o outro com percentuais do total obtido na fase de busca.

Tabela 5.1: Os 11 maiores meios de publicação

| **Meios de Publicação** | **Trabalhos** |
|---|---|
| ACM symposium on Cloud computing | 5 |
| International Conference on Autonomic Computing | 4 |
| International Conference on Cloud Computing | 4 |
| International Conference on Cloud Computing Technology and Science | 4 |
| International Symposium on Cluster, Cloud and Grid Computing | 4 |
| Middleware Doctoral Symposium | 4 |
| International Conference on Cloud Computing Technology and Science | 3 |
| International Symposium on High-Performance Parallel and Distributed Computing | 3 |
| International Workshop on Middleware for Grids, Clouds and e-Science | 3 |
| Science China Information Sciences | 3 |
| The Journal of Supercomputing | 3 |

Figura 5.2: Estudos por Fonte de Busca

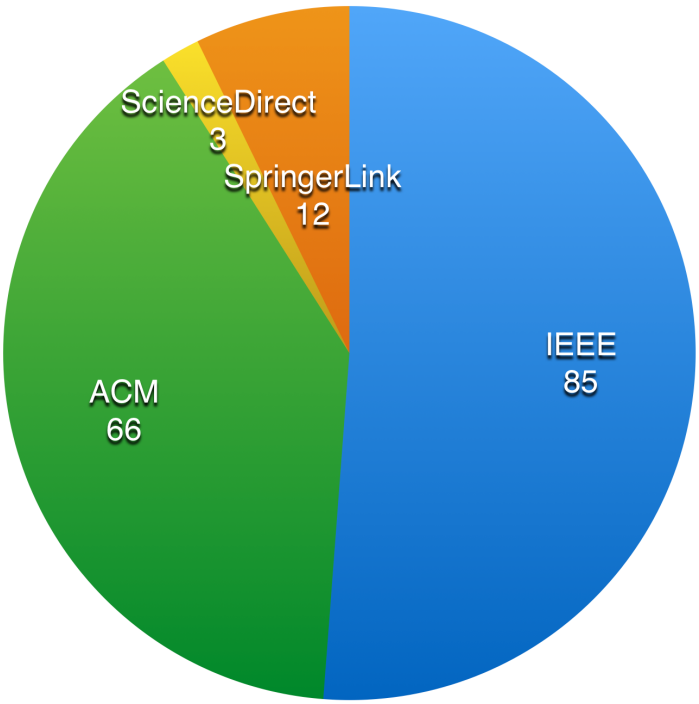


Figura 5.3: Estudos por Fonte de Busca - Evolução em valores absolutos

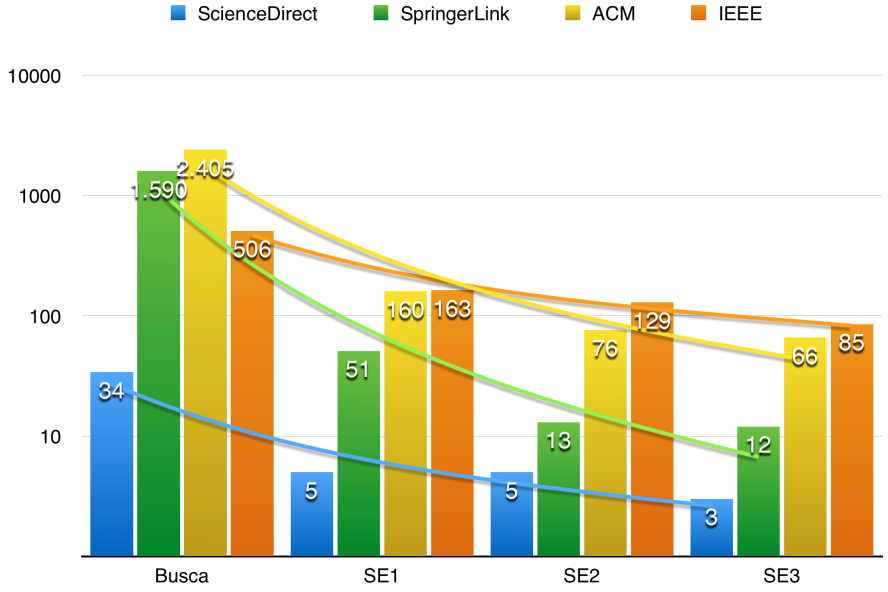
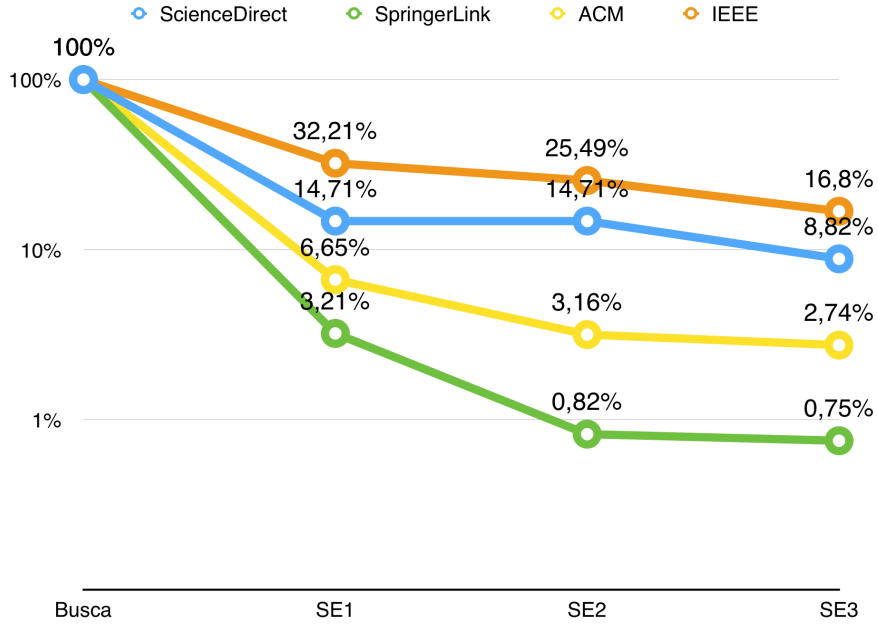



Figura 5.4: Estudos por Fonte de Busca - Evolução em valores proporcionais

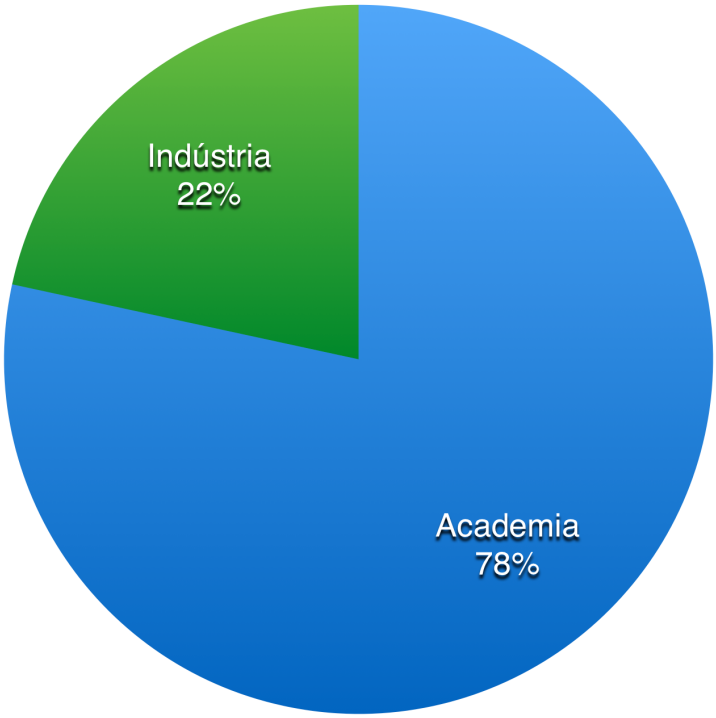


## 5.3 QP3 - Qual é a representatividade da academia e da indústria neste campo de pesquisa?

Para responder esta questão, foi armazenada a afiliação dos autores. Nós classificamos como academia estudos originados de instituições e universidades, e como indústria as empresas e indústrias, e também as outras instituições que não puderam ser consideradas como academia.

De acordo com a figura 5.5, a academia mostrou maior engajamento na publicação de resultados sobre Gerenciamento de Falhas em Computação em Nuvem, com uma fatia de 148 artigos, contra a presença de apenas 60 na indústria.

Figura 5.5: Representação da Academia e Indústria




Um aprofundamento na filiação dos autores apresenta 178 instituições aparecendo nas publicações. A University of California foi quem obteve o maior número de autores. A tabela 5.2 apresenta as 9 maiores instituições em termos de número de autores. Uma visão mais ampla da afiliação dos autores pode ser observada através da distribuição através dos países presentes na tabela 5.3. Os Estados Unidos lideram com 213 autores de um total de 523, obtendo 40,7% do total de autores. 26 países diferentes foram identificados nos

Tabela 5.2: As 9 instituições com mais autores

| **Instituição** | **Autores** |
|---|---|
| University of California | 21 |
| IBM | 20 |
| Universidade Federal de Pernambuco | 15 |
| INRIA | 14 |
| University of Illinois | 13 |
| Microsoft | 11 |
| Universidade de Lisboa | 11 |
| Georgia Institute of Technology | 10 |
| North Carolina State University | 10 |

resultados.

## 5.4 QP4 - Quais são os tópicos mais explorados na interseção entre computação em nuvens e gerência de faltas?

Para ajudar a expor a relação entre as facetas de Gerência de Faltas e Computação em Nuvem, desenhou-se um gráfico de bolhas (ou gráfico de matriz), onde no eixo X apresentamos a faceta de Modelos de Serviço, no eixo Y os atributos de dependabilidade e em cada bolha é apresentado um gráfico de pizza com a faceta de meio de tratamento de falha.

A figura 5.6 apresenta a interseção entre as facetas de Nível de Serviço de Computação em Nuvem, Atributo de Dependabilidade e Meio de Tratamento de Falha. Podemos observar que os tópicos mais explorados são: atributos de confiabilidade e disponibilidade (*reliability* e *availability*), através de prevenção e tolerância à falhas no nível de serviço IaaS.

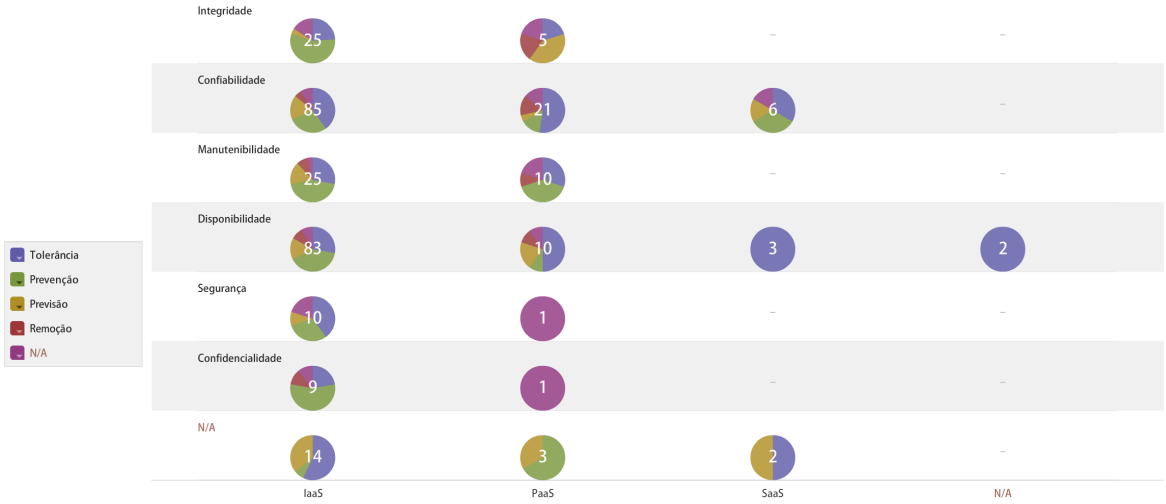
Algumas lacunas podem ser identificadas na figura 5.6. Não foram encontrados registros sobre manutenibilidade ou integridade no nível de serviço SaaS. Também não foram encontrados trabalhos sobre remoção de falhas em SaaS. Segurança e confidencialidade em PaaS não tiveram meios de tratamento de falhas evidenciados na nossa investigação.

Já a figura 5.7 apresenta a interseção entre as facetas de Tipo de Pesquisa, Modelo de Serviço e Meio de Tratamento de Falha. Observamos através da figura que Soluções para

Tabela 5.3: Os 10 países com mais autores

| **País** | **Autores** |
|---|---|
| Estados Unidos | 213 |
| China | 60 |
| Brasil | 26 |
| França | 22 |
| Portugal | 20 |
| Índia | 20 |
| Alemanha | 19 |
| Austrália | 17 |
| Coréia do Sul | 15 |
| Taiwan | 15 |

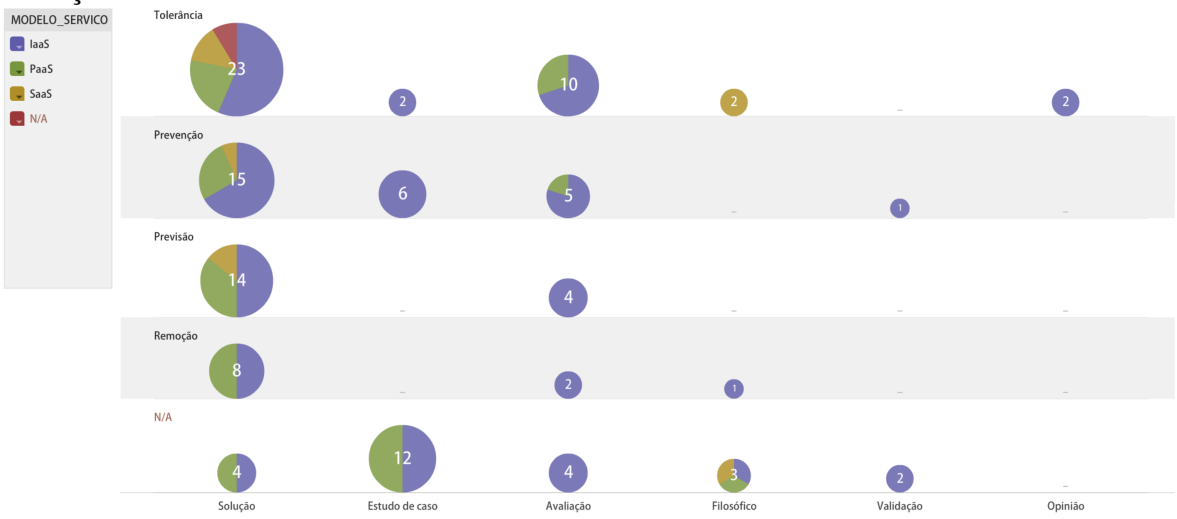
Figura 5.6: Relação entre Modelo de Serviço, Atributo de Dependabilidade e Meio de Tratamento de Falha



Tolerância é o tópico mais explorado, com 23 trabalhos. Podemos identificar ainda que os estudos de caso de prevenção e tolerância, as avaliações de previsão e remoção de falhas e ainda trabalhos de validação e opinião falaram apenas sobre IaaS. Apenas 4 dos 25 trabalhos de avaliação foram sobre PaaS, sendo os outros 21 sobre IaaS. Ainda observamos que 12 estudos de caso não puderam ser classificados pelo meio de tratamento de falha.

Pode-se identificar algumas lacunas na figura 5.7. Não há estudos de caso de Remoção e Previsão de falhas. A quantidade de trabalhos filosóficos e de validação são bem pequenas. Apesar de também serem poucos trabalhos de opinião, isso se explica devido ao caráter científico dos portais pesquisados, já que evita-se publicar trabalhos de opinião em meios acadêmicos.

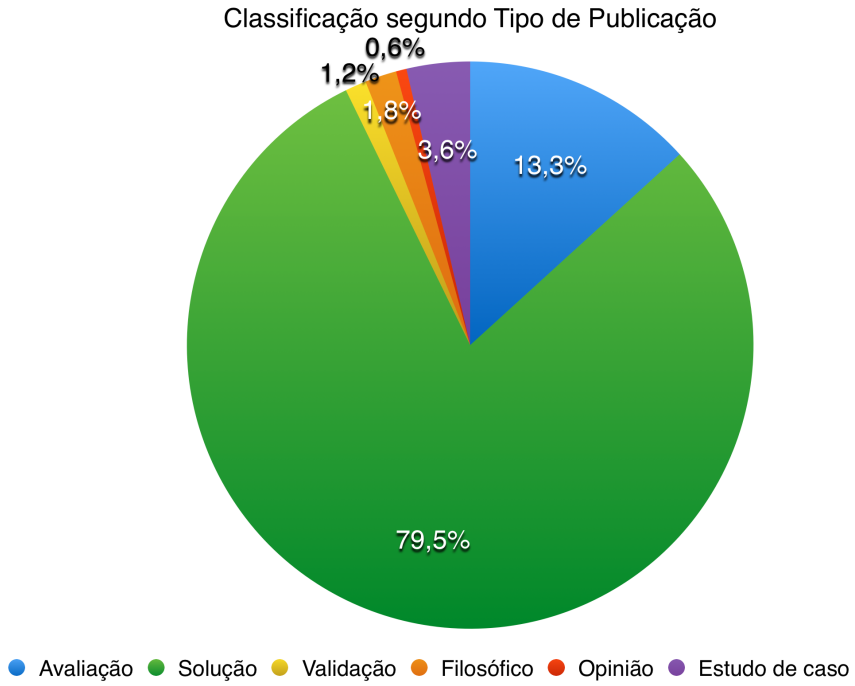
Figura 5.7: Relação entre Tipo de Pesquisa, Meio de Tratamento de Falha e Modelo de Serviço




## 5.5 QP5 - Que tipos de pesquisa são realizados nesse contexto?

Observa-se uma grande tendência na publicação de propostas de solução. A figura 5.8 mostra que 79% dos estudos foram classificados como propostas de solução. Poucos trabalhos de opinião, filosóficos e de validação foram encontrados.

Figura 5.8: Quantidade de estudos por tipo de pesquisa



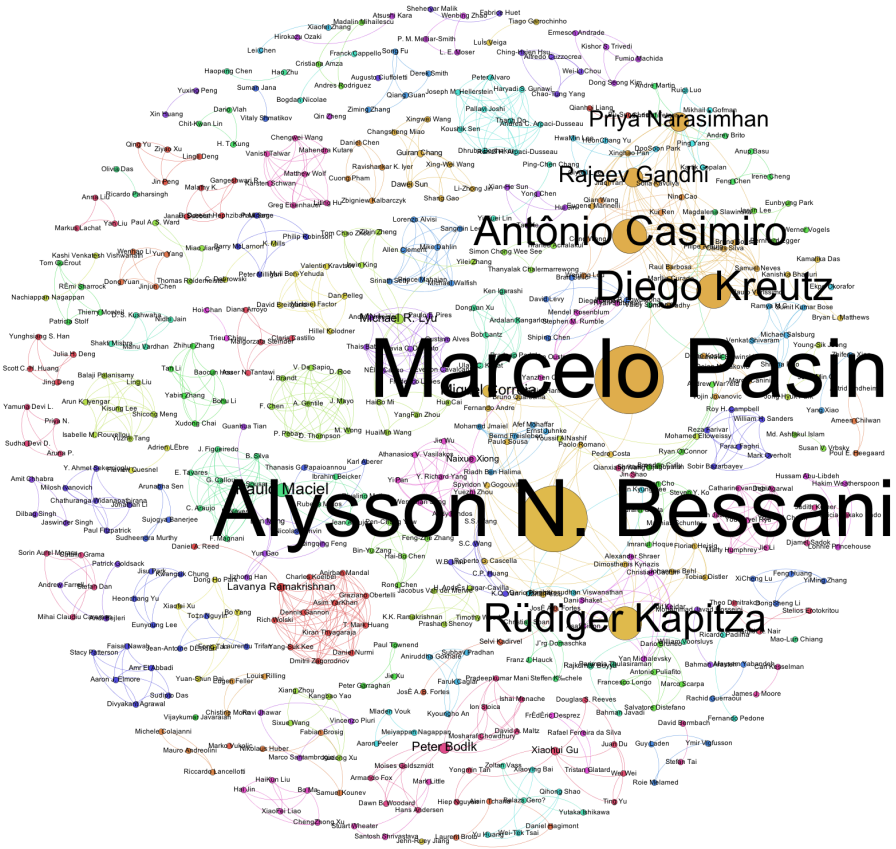

## 5.6 QP6 - Existem ligações entre pesquisadores/instituições através dos estudos? Existe uma rede de pesquisadores/instituições?

Para responder esta pergunta utilizou-se dados dos autores encontrados nos artigos e suas afiliações. A figura 5.9 apresenta um grafo de relações de autoria onde os nós representam cada autor, sendo o tamanho do nó a representação da quantidade de autores diferentes relacionados àquele, e cada aresta apresenta a ligação de co-autoria em um trabalho. O grafo foi colorido de forma a identificar conjuntos de autores de forma clara e que sejam disjuntos.

Pode-se observar que muitos conjuntos disjuntos aparecem no mapa. Isso mostra que muitos autores publicam trabalhos com seus contatos mais próximos, e não há muitas interações entre grupos de pesquisa.

Um conjunto maior em tom de amarelo possui alguns nós de maior tamanho e pode-se concluir que uma possível interação entre grupos de pesquisa pode ser encontrada. Autores como Bessani, Pasin, Kreutz, Casimiro, Gandhi, Kapitza e Narasimhan são relacionados a muitos outros e aparentam se comportar como pontes entre grupos de pesquisa.

Figura 5.9: Relação entre autores através dos estudos



## 5.7 Considerações Finais

Neste capítulo foram apresentadas respostas às questões de pesquisa propostas para o mapeamento, sendo apresentadas argumentações baseadas nos dados coletados da pesquisa. O próximo capítulo apresenta uma breve conclusão sobre o apanhado do trabalho e sugestões para trabalhos futuros.

# Capítulo 6

# Trabalhos Relacionados

## 6.1 Mapeamentos sistemáticos realizados

Nesta seção serão abordados alguns trabalhos publicados (dissertações de mestrado e artigos) que exploram a metodologia de mapeamento sistemático. Serão discutidos os trabalhos de [Barreiros 2011], [Carvalho 2012], [Elberzhager, Münch e Nha 2012] e [Jacinto 2010].

Nas subseções a seguir serão apresentados os trabalhos informados anteriormente. A cada trabalho será apresentado uma breve revisão e serão discutidos pontos chaves dos trabalhos. Ao final do capítulo, será apresentado um breve quadro comparativo.

### 6.1.1 A systematic mapping study on software engineering testbeds

O estudo de [Barreiros 2011] trata de um mapeamento sistemático sobre testbeds em engenharia de software. Ele diz que "A motivação do trabalho é produzir melhores softwares, por permitir tecnologias novas e promissoras serem adotadas pelo público em geral mais cedo".

O estudo foca em duas questões principais do problema, sendo depois detalhada em quatro questões de pesquisa para serem respondidas pelo mapeamento sistemático a ser realizado.

O trabalho não utiliza exclusivamente a metodologia do mapeamento sistemático para alcançar seu objetivo, mas ela parece ser empregada para trazer, de maneira sistemática, estudos que o ajudem a discorrer sobre o assunto. Tal observação é importante, já que a sistematização agrega valor acadêmico ao trabalho e também torna o processo de pesquisa

repetitível e passível de avaliação.

Este estudo utiliza a taxonomia de [Cooper 1988] para classificar os seus artigos. Esta taxonomia, apesar de ”proposta originalmente como um sistema de classificação para revisões sistemáticas de literatura em educação e psicologia, ela é razoavelmente aplicável a revisões sistemáticas de literatura em geral” [Barreiros 2011].

Observa-se no estudo que ”inicialmente a intenção foi de se construir uma SLR, mas baseado nas questões de pesquisa, um MS foi a melhor escolha” [Barreiros 2011]. Tal fato se constata pelas questões de pesquisa propostas pelo trabalho serem abrangentes.

O protocolo do estudo é construído com os seguintes itens: processo de busca, critérios de seleção de estudos, avaliação de relevância e estratégia de extração de dados.

Após a elaboração do protocolo, este foi executado. Ao observar o processo de seleção, podemos acompanhar a quantidade de trabalhos manuseados. Inicialmente, com a busca automática através dos engenhos de pesquisa, foram retornados 4239 estudos. Desses, apenas 106 foram considerados estudos potencialmente relevantes. Desses 106, apenas 13 foram considerados estudos primários para o trabalho, totalizando 0,3% do total retornado pelas buscas. Podemos observar nesta busca a grande quantidade de trabalhos irrelevantes para a pesquisa. Após a aplicação do protocolo, é feita a aplicação do método indutivo para avaliar qualitativamente os trabalhos primários de forma a responder suas questões de pesquisa.

Pode-se observar neste trabalho que o autor, apesar de ter a intenção de usar a metodologia de mapeamento sistemático, na verdade realiza uma revisão sistemática de literatura, e pode-se caracterizar isso pela utilização de uma abordagem qualitativa após a filtragem e seleção de artigos. De acordo com [Kitchenham, Budgen e Brereton 2011], podemos observar que ”mapeamentos sistemáticos usam a mesma metodologia básica das SLRs mas visam identificar e classificar todas as pesquisas relacionadas a um tópico abrangente em vez de responder questões sobre os méritos relativos de tecnologias que competem que SLRs convencionais visam”, e com essa afirmação podemos confirmar a presença de características de SLR no trabalho.

### 6.1.2 Um Mapeamento Sistemático de Estudos em Cloud Computing

Recentemente, o trabalho de [Carvalho 2012] foi publicado visando realizar um levantamento para investigar e apoiar o desenvolvimento e adoção de soluções em computação em

nuvem.

O trabalho é um mapeamento sistemático que foge a regra dos outros aqui relatados, pois este apresenta uma primeira tentativa de mapeamento sistemático na área de TI que foge do escopo da engenharia de software.

O trabalho é baseado na busca por respostas a oito questões de pesquisa, sendo uma delas (”QP7 - Como é feito o monitoramento do uso de recursos em Cloud Computing?” [Carvalho 2012]) relacionada a este trabalho.

Existe uma outra pergunta também no trabalho que se relaciona de maneira bem indireta com este aqui proposto: a QP8 - ”Quais os principais problemas quanto a segurança em Cloud Computing?”, já que problemas de segurança também podem ser consideradas falhas.

Apresentada a introdução do trabalho, foca-se neste momento nas informações relevantes à aplicação das técnicas de mapeamento sistemático. [Carvalho 2012] informa que a intenção inicial de seu trabalho era a construção de uma revisão sistemática de literatura, mas optou-se por um mapeamento sistemático já que o estudo utiliza questões de pesquisa mais abrangentes.

A equipe de pesquisa é composta por 3 pessoas e o trabalho apresenta questões de visão geral, sem perguntas muito específicas a um certo assunto, portais de pesquisa semelhantes aos que são usados neste trabalho e critérios de seleção sob mesmos fundamentos ao aqui propostos.

Podemos observar algumas diferenças importantes em dois pontos: a cadeia de busca definida e a classificação dos estudos.

Na cadeia de busca, procurou-se neste trabalho utilizar união para representar os domínios de pesquisa propostos para depois utilizar a interseção para ligar tais domínios de pesquisa. Em [Carvalho 2012] observamos a utilização de uma série de termos que envolvem a área de computação em nuvem unida ao termo ”cloud computing”, de uma maneira que aparentemente tenta-se extrair o máximo de conteúdo possível que use ou referencie cloud computing. Uma crítica que pode ser feita ao método usado para a construção da cadeia de busca é que, de primeira vista, ela aparenta ser uma massa de palavras sem semântica forte relacionada ao tema que tenta agregar ao máximo estudos em computação em nuvem, ou seja, não se vê uma estruturação da busca de forma a encontrar artigos que especifiquem uma certa área. De certa forma, este ponto pode ser considerado positivo, mas a forma apre-

sentada parece um pouco descuidada, onde não observamos uma estruturação da informação a ser pesquisada. O autor parece não querer especificar a cadeia de busca de forma a incluir mais trabalhos em sua pesquisa.

A pesquisa de [Carvalho 2012] foi realizada em quatro etapas. Na etapa de busca, foram obtidos 2977 artigos. Na segunda etapa, houve uma redução para 827 artigos. Na terceira, a frequência caiu para 496, e na última etapa para 301 artigos, totalizando 10,1% do total de artigos buscados.

Já em relação à classificação de estudos, observamos diferenças maiores. O autor utiliza a taxonomia de [Easterbrook et al. 2011], enquanto a proposta por este trabalho é a do [Wieringa et al. 2005].

### 6.1.3 A systematic mapping study on the combination of static and dynamic quality assurance techniques

O trabalho de [Elberzhager, Münch e Nha 2012] apresenta o intuito de, através do uso da metodologia de mapeamento sistemático, apresentar um trabalho sobre a combinação de técnicas de garantia de qualidade dinâmicas e estáticas.

Uma primeira impressão positiva que temos ao analisar este artigo é na abordagem utilizada para a construção do seu resumo. O artigo utiliza o conceito exposto por [Budgen et al. 2008] de resumos estruturados (do inglês, structured abstracts). Segundo [Budgen et al. 2008], ”Um resumo estruturado emprega um conjunto padrão de cabeçalhos através da qual autores resumem aspectos principais de um estudo, tal como seu contexto, objetivo, método, resultados e conclusões”. O uso de tal recurso traz benefícios, pois, de acordo com [Budgen et al. 2008], ”consultar o artigo completo não apenas involve tempo adicional e esforço, como também pode haver custos e demoras involvidas na obtenção destes, já que nem todos os artigos podem estar disponíveis online. Então a falta de informação em resumos pode significativamente aumentar tanto o custo quanto o tempo necessário” para realizar um mapeamento sistemático.

O trabalho utiliza de seis questões de pesquisa para nortear seus resultados. Também são utilizados quatro engenhos de busca como fontes de busca de trabalhos primários. Foram utilizados os engenhos Compendex e Inspec para o processo de seleção, enquanto o IEEE

Xplore e o ACM Digital Library foram usados apenas para validar os conjuntos de seleção.

Após a definição da cadeia de busca e executada a pesquisa através dos portais, encontrou-se 2012 artigos. Desses, foram removidos 309 duplicados, restando 1703 artigos. Aplicado um segundo filtro baseado em título e ano de publicação, foram excluídos 1597 artigos, todos considerados irrelevantes, restando apenas 106. Um segundo filtro aplicado baseado no resumo restringiu a quantidade de artigos para 60, e ao aplicar uma exclusão baseada no texto completo foram excluídos mais 12 artigos, totalizando em 48 estudos, sendo 47 deles primários e 1 secundário. De forma a não perder nenhum artigo relevante no mapeamento, os artigos considerados estudo secundário incluído foram analisados e foram incluídos no mapeamento, totalizando 51 artigos selecionados no final, 2,5% do total.

Após realizar análise dos dados, [Budgen et al. 2008] foi capaz de responder as seis questões de pesquisa de maneira satisfatória segundo eles, e pode completar afirmando que ”o maior resultado deste mapeamento sistemático é a identificação e classificação de abordagens existentes que combinam técnicas de garantia de qualidade estáticas e dinâmicas”. [Budgen et al. 2008] também diz que ”Isso pode apoiar profissionais na escolha de combinações de técnicas que compensam e também serve como uma base para pesquisadores nos termos de direções de pesquisa futuras promissoras”.

Podemos analisar este trabalho como um mapeamento sistemático de sucesso do ponto de vista da clareza do protocolo e dados expostos da execução do mapeamento, conforme foi definido pelo grupo do EBSE (Evidence-based Software Engineering) da Universidade de Durham, que são os pioneiros dessa metodologia. Apesar de tal ponto positivo, podemos ainda resalvar que o mapeamento contemplou 51 estudos apenas, e em alguns casos podemos ver que esta quantidade de estudos pode não trazer informações quantitativas suficientes.

### 6.1.4 Um mapeamento sistemático da pesquisa sobre a influência da personalidade na Engenharia de Software

O trabalho de [Jacinto 2010] visa identificar trabalhos que falem sobre a influência da personalidade na engenharia de software. A motivação para o desenvolvimento do trabalho é a necessidade, na prática da engenharia de software, de se produzir um melhor entendimento das influências do fator humano personalidade no desenvolvimento de software e de se tradu-

zir este entendimento em um corpo de conhecimento que possa ser utilizado por praticantes e pesquisadores.

O trabalho utiliza o método do mapeamento sistemático aliado ao método de comparações constantes. Analisaremos no trabalho de [Jacinto 2010] apenas o que se refere ao método do mapeamento sistemático, pois é o que convém para este trabalho.

Enquanto o método de comparações constantes possui natureza qualitativa, o mapeamento sistemático tem papel fundamental neste trabalho para a identificação de estudos primários relevantes de maneira quantitativa, para que posteriormente os trabalhos selecionados pelo mapeamento possam ser analisados qualitativamente.

O mapeamento sistemático foi guiado por 5 questões de pesquisa. A estratégia de busca utilizada foi apenas a da busca automática. Nas definições desta busca, podemos observar alguns pontos importantes, como a interseção de dois domínios de pesquisa na cadeia de busca, sendo um relacionado à engenharia de software e o outro relacionado à personalidade.

Um detalhe observado na construção dessa cadeia é que, enquanto várias palavras-chave foram adicionadas no domínio da engenharia de software, apenas uma palavra-chave foi adicionada ao domínio da personalidade, o que pode levar a conclusão de que a autora procura tendenciar seu trabalho à engenharia de software, tentando não expor muitos trabalhos relacionados à personalidade.

Foram utilizados 5 portais de busca (IEEE Xplore, ACM Digital Library, ScienceDirect, El Compendex e Scopus), e os critérios de inclusão/exclusão parecem ser de uma acentuada subjetividade.

”A partir da cadeia e fontes definidas, as buscas primárias retornaram um total de 3177 trabalhos” [Jacinto 2010]. Desses trabalhos, apenas 123 estudos foram considerados potencialmente relevantes após análise das palavras-chave e título do trabalho, ou seja, apenas 3,9%. Após a leitura da conclusão e resumo, reduziu-se o valor de potenciais relevantes para 38 trabalhos, 1,2% do total.

Com esses 38 trabalhos foi realizada a análise qualitativa dos trabalhos, de forma a identificar pontos chave e responder as questões de pesquisa elaboradas.

O que pode-se concluir, de maneira breve, é que o mapeamento sistemático, apesar de ser uma ferramenta importante para a coleta de trabalhos, não parece ser o instrumento principal de pesquisa do trabalho. [Jacinto 2010] se utiliza de uma série de gráficos de forma

a analisar dados quantitativos da pesquisa, mas no próprio texto observa-se grande parte da argumentação focada em extração de textos dos trabalhos e uso de ferramentas qualitativas para resposta às questões formuladas.

## 6.2 Considerações Finais

Nesta seção foram analisados os trabalhos de [Barreiros 2011], [Carvalho 2012], [Elberzhager, Münch e Nha 2012] e [Jacinto 2010]. Pode-se observar nestes trabalhos uma drástica redução entre a quantidade de artigos buscados automaticamente e a quantidade de artigos selecionados. Podemos notar uma média de redução para 3,5% do total nas seleções dos artigos. O próximo capítulo visa apresentar a metodologia de mapeamento sistemático utilizada nessa pesquisa.

# Capítulo 7

# Conclusão e Trabalhos Futuros

Esta dissertação apresentou um mapeamento sistemático provendo uma visão geral de pesquisas existentes em gerenciamento de falhas em computação em nuvem. Nós selecionamos 166 trabalhos científicos entre 2008 e 2012 de 4535 encontrados pela nosso método.

A classificação desses trabalhos científicos ajudou a identificar como a produção científica é distribuída através dos anos e a reconhecer quais são os principais meios de publicação utilizados. Ainda foi possível apontar tópicos de pesquisa bem explorados e lacunas que podem representar oportunidades para pesquisa futura em gerenciamento de falhas em computação em nuvem.

Pode-se concluir que a maioria dos trabalhos foca-se no estudo da IaaS para alcance da dependabilidade em computação em nuvem. Confiabilidade e disponibilidade foram atributos frequentemente discutidos em IaaS, assim como tolerância e prevenção de falhas foram tópicos frequentemente explorados em IaaS.

Alguns trabalhos futuros podem ser sugeridos de acordo com o que foi construído neste trabalho. Sugere-se investigar as lacunas e saturações de tópicos de pesquisa encontrados para possivelmente entender o motivo da existência de tal saturação/lacuna, assim como utilizar este trabalho como fonte de dados de investigação para justificativa de outros trabalhos nessas áreas.

Todos os dados utilizados na pesquisa deste trabalho foram registrados em todas as etapas e estão disponibilizados no endereço eletrônico http://goo.gl/6IigqP.

Este trabalho resultou em um artigo apresentado no congresso internacional realizado em Taiwan, o *14'th International Conference on Parallel and Distributed Computing, Applica-*

*tions and Technologies (PDCAT'13)*. O artigo pode ser acessado em sua versão eletrônica no endereço http://pdcat13.csie.ntust.edu.tw/download/papers/P10072.pdf.

Pode-se identificar pontos críticos e lições aprendidas sobre o trabalho. Primeiramente, o trabalho é bastante suscetível a mudanças em seu percurso, já que a calibragem da cadeia de busca e do tema pesquisado pode variar durante o tempo de execução do trabalho. Uma experiência feliz durante a pesquisa foi, no ato da obtenção dos dados da fonte de busca, como a obtenção dos artigos foi manual, foi identificado um possível problema de consistência nos engenhos de busca. Se a busca fosse realizada a cada dia, poderíamos ter resultados da busca diferentes, fazendo com que artigos que apareceram na primeira busca não apareçam novamente ou que artigos que não tinham surgido até agora surjam. Para contornar esse desafio, salvamos em nossos documentos as páginas de busca dos portais, garantindo assim que o estado da busca realizada no primeiro dia permanecesse por todo o trabalho. Vale ressaltar que o trabalho dos processos de seleção de estudos é bastante árduo, muito minucioso e suscetível a falhas devido ao cansaço físico gerado pela leitura de tantos trabalhos.

# Bibliografia

[Avizienis et al. 2004]AVIZIENIS, A. et al. Basic concepts and taxonomy of dependable and secure computing. *Dependable and Secure Computing, IEEE Transactions on*, v. 1, n. 1, p. 11 – 33, jan.-march 2004. ISSN 1545-5971.

[Barreiros 2011]BARREIROS, E. F. S. *A systematic mapping study on software engineering testbeds*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2011.

[Budgen et al. 2006]BUDGEN, D. et al. Investigating the applicability of the evidence-based paradigm to software engineering. In: *Proceedings of the 2006 international workshop on Workshop on interdisciplinary software engineering research*. New York, NY, USA: ACM, 2006. (WISER '06), p. 7–14. ISBN 1-59593-409-X. Disponível em: <http://doi.acm.org/10.1145/1137661.1137665>.

[Budgen et al. 2008]BUDGEN, D. et al. Presenting software engineering results using structured abstracts: a randomised experiment. *Empirical Software Engineering*, Springer Netherlands, v. 13, p. 435–468, 2008. ISSN 1382-3256. 10.1007/s10664-008-9075-7. Disponível em: <http://dx.doi.org/10.1007/s10664-008-9075-7>.

[Budgen et al. 2008]BUDGEN, D. et al. Using Mapping Studies in Software Engineering. In: *Proceedings of PPIG 2008*. [S.l.]: Lancaster University, 2008. p. 195–204. ISBN 978-1-86220-215-3.

[Buyya, Yeo e Venugopal 2008]BUYYA, R.; YEO, C. S.; VENUGOPAL, S. Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. In: *High Performance Computing and Communications, 2008. HPCC '08. 10th IEEE International Conference on*. [S.l.: s.n.], 2008. p. 5 –13.

[Carvalho 2012]CARVALHO, J. F. S. d. *Um mapeamento Sistematico de Estudos em Cloud Computing*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2012.

[Cooper 1988]COOPER, H. Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge, Technology Policy*, Springer Netherlands, v. 1, p. 104–126, 1988. ISSN 0897-1986. 10.1007/BF03177550. Disponível em: <http://dx.doi.org/10.1007/BF03177550>.

[Dyba, Kitchenham e Jorgensen 2005]DYBA, T.; KITCHENHAM, B. A.; JORGENSEN, M. Evidence-based software engineering for practitioners. *IEEE Softw.*, IEEE Computer Society Press, Los Alamitos, CA, USA, v. 22, n. 1, p. 58–65, jan. 2005. ISSN 0740-7459. Disponível em: <http://dx.doi.org/10.1109/MS.2005.6>.

[Easterbrook et al. 2011]EASTERBROOK, S. et al. Selecting empirical methods for software engineering research. In: _____. [S.l.]: N, 2011. cap. 0.

[Elberzhager, Münch e Nha 2012]ELBERZHAGER, F.; MüNCH, J.; NHA, V. T. N. A systematic mapping study on the combination of static and dynamic quality assurance techniques. *Inf. Softw. Technol.*, Butterworth-Heinemann, Newton, MA, USA, v. 54, n. 1, p. 1–15, jan. 2012. ISSN 0950-5849. Disponível em: <http://dx.doi.org/10.1016/j.infsof.2011.06.003>.

[Garfinkel 1999]GARFINKEL, S. L. *Architects of the Information Society: Thirty-Five Years of the Laboratory for Computer Science at MIT*. [S.l.]: The MIT Press, 1999.

[Gong et al. 2010]GONG, C. et al. The characteristics of cloud computing. In: *Parallel Processing Workshops (ICPPW), 2010 39th International Conference on*. [S.l.: s.n.], 2010. p. 275–279. ISSN 1530-2016.

[Jacinto 2010]JACINTO, S. d. S. *Um mapeamento sistematico da pesquisa sobre a influencia da personalidade na Engenharia de Software*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2010.

[Kitchenham, Brereton e Budgen 2010]KITCHENHAM, B.; BRERETON, P.; BUDGEN, D. The educational value of mapping studies of software engineering literature. In: *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - Volume*

*1*. New York, NY, USA: ACM, 2010. (ICSE '10), p. 589–598. ISBN 978-1-60558-719-6. Disponível em: <http://doi.acm.org/10.1145/1806799.1806887>.

[Kitchenham et al. 2007]KITCHENHAM, B. et al. 2nd internationalworkshop on realising evidence-based software engineering (rebse-2): Overview and introduction. In: *Proceedings of the Second International Workshop on Realising Evidence-Based Software Engineering*. Washington, DC, USA: IEEE Computer Society, 2007. (REBSE '07), p. 1–. ISBN 0-7695-2962-3. Disponível em: <http://dx.doi.org/10.1109/REBSE.2007.1>.

[Kitchenham et al. 2010]KITCHENHAM, B. et al. Systematic literature reviews in software engineering - a tertiary study. *Inf. Softw. Technol.*, Butterworth-Heinemann, Newton, MA, USA, v. 52, n. 8, p. 792–805, ago. 2010. ISSN 0950-5849. Disponível em: <http://dx.doi.org/10.1016/j.infsof.2010.03.006>.

[Kitchenham, Budgen e Brereton 2011]KITCHENHAM, B. A.; BUDGEN, D.; BRERETON, O. P. Using mapping studies as the basis for further research - a participant-observer case study. *Inf. Softw. Technol.*, Butterworth-Heinemann, Newton, MA, USA, v. 53, n. 6, p. 638–651, jun. 2011. ISSN 0950-5849. Disponível em: <http://dx.doi.org/10.1016/j.infsof.2010.12.011>.

[Kitchenham, Dyba e Jorgensen 2004]KITCHENHAM, B. A.; DYBA, T.; JORGENSEN, M. Evidence-based software engineering. In: *Proceedings of the 26th International Conference on Software Engineering*. Washington, DC, USA: IEEE Computer Society, 2004. (ICSE '04), p. 273–281. ISBN 0-7695-2163-0. Disponível em: <http://dl.acm.org/citation.cfm?id=998675.999432>.

[Laprie 1995]LAPRIE, J.-C. Dependable computing: concepts, limits, challenges. In: *Proceedings of the Twenty-Fifth international conference on Fault-tolerant computing*. Washington, DC, USA: IEEE Computer Society, 1995. (FTCS'95), p. 42–54. ISBN 0-8186-7146-7. Disponível em: <http://dl.acm.org/citation.cfm?id=1899254.1899261>.

[Lung 2012]LUNG, L. C. *Por que tolerância a faltas e não falha? Por que falta, erro e falha?* ago. 2012. Disponível em: <http://www.inf.ufsc.br/ lau.lung/port/definicaoTF.htm>.

[NIST 2011]NIST. *The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology*. [S.l.], 2011.

[Petersen e Ali 2011]PETERSEN, K.; ALI, N. B. Identifying strategies for study selection in systematic reviews and maps. In: *Proceedings of the 2011 International Symposium on Empirical Software Engineering and Measurement*. Washington, DC, USA: IEEE Computer Society, 2011. (ESEM '11), p. 351–354. ISBN 978-0-7695-4604-9. Disponível em: <http://dx.doi.org/10.1109/ESEM.2011.46>.

[Petersen et al. 2008]PETERSEN, K. et al. Systematic mapping studies in software engineering. *12th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 2008. Disponível em: <http://www.bcs.org/upload/pdf/ewic$_e a08_p aper 8.pdf$>.

[Reimer 2005]REIMER, J. *Total share: 30 years of personal computer market share figures*. 2005. Disponível em: <http://arstechnica.com/features/2005/12/total-share/10/>.

[Sackett et al. 2000]SACKETT, D. L. et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2nd. ed. [S.l.]: Churchill Livingstone, 2000. ISBN 0443062404.

[Silva et al. 2011]SILVA, F. Q. B. da et al. Six years of systematic literature reviews in software engineering: An updated tertiary study. *Inf. Softw. Technol.*, Butterworth-Heinemann, Newton, MA, USA, v. 53, n. 9, p. 899–913, set. 2011. ISSN 0950-5849. Disponível em: <http://dx.doi.org/10.1016/j.infsof.2011.04.004>.

[Travassos 2007]TRAVASSOS, G. H. From silver bullets to philosophers' stones: who wants to be just an empiricist? In: *Proceedings of the 2006 international conference on Empirical software engineering issues: critical assessment and future directions*. Berlin, Heidelberg: Springer-Verlag, 2007. p. 39–39. ISBN 978-3-540-71300-5. Disponível em: <http://dl.acm.org/citation.cfm?id=1767399.1767416>.

[Veras e Tozer 2012]VERAS, M.; TOZER, R. *Cloud Computing: Nova arquitetura de TI*. BRASPORT, 2012. ISBN 9788574524894. Disponível em: <http://books.google.com.br/books?id=yiggoX2aoC8C>.

[Wieringa et al. 2005]WIERINGA, R. et al. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requir. Eng.*, Springer-Verlag New

York, Inc., Secaucus, NJ, USA, v. 11, n. 1, p. 102–107, dez. 2005. ISSN 0947-3602. Disponível em: <http://dx.doi.org/10.1007/s00766-005-0021-6>.

[Zhang et al. 2010]ZHANG, S. et al. Cloud computing research and development trend. In: *Future Networks, 2010. ICFN '10. Second International Conference on*. [S.l.: s.n.], 2010. p. 93–97.