

WAZLAWICK, Raul Sidnei. Metodologia de Pesquisa para Ciência da Computação. Rio de Janeiro: Elsevier, 2008.

Capítulo 1 – Introdução

Capítulo 2 – Estilos de Pesquisa Correntes

Capítulo 3 – Preparação de um Trabalho de Pesquisa

3.1 Escolhendo o Objetivo de Pesquisa

3.2 A Revisão Bibliográfica

3.3 O Objetivo

3.4 O Método de Pesquisa

3.5 Justificativa

3.6 Resultados Esperados

3.7 Limitações do Trabalho

3.8 Discussão

Capítulo 4 – Análise Crítica da Proposta de Monografia

Capítulo 5 – Escrita da Monografia

Capítulo 6 – Escrita de Artigo Científico

Capítulo 7 – Plágio

Capítulo 8 – Níveis de Exigência do Trabalho de Conclusão

Referências

Capítulo 3

Preparação de um Trabalho de Pesquisa

A preparação de um trabalho de pesquisa é uma etapa que deve ser realizada antes que se comece a escrever sobre a pesquisa.

Parece estranho, mas muitas vezes esse é um ponto que deve ser ressaltado. Alunos, ansiosos por escrever a monografia, começam a escrever páginas e páginas sem terem feito nenhuma pesquisa. Uma ressalva: é necessário diferenciar o termo “pesquisa”, que pode ser entendido como revisão bibliográfica, da sua significação como pesquisa científica, ou seja, produção de conhecimento novo. A revisão bibliográfica, diga-se de passagem, não produz conhecimento novo, mas apenas supre ao estudante conhecimentos que lhe faltavam.

Portanto, não se recomenda que o aluno comece a escrever sua monografia ou trabalho de conclusão de curso sem ter realizado alguma pesquisa que tenha produzido conhecimento novo. Explicando melhor: não há necessidade de sair escrevendo o capítulo de revisão bibliográfica antes de saber o que efetivamente será feito em termos de produção de conhecimento.

A revisão bibliográfica de um trabalho de pesquisa em Computação, em geral, não deve ser um tratado sobre a área de pesquisa. Muitas vezes, um aluno que começa a escrever o capítulo da revisão antes de decidir o objetivo da pesquisa acabará escrevendo demais e desnecessariamente. Esse capítulo será cansativo para o leitor, e muitas vezes ele não compreenderá por que determinados assuntos são ali colocados se não são abordados na monografia em si, que vai iniciar mais adiante.

O segredo de um trabalho de pesquisa de sucesso consiste em ter um bom objetivo. Uma vez definido o objetivo do trabalho, tudo o mais gravita em redor dele. A justificativa vai dizer por que vale a pena buscar esse objetivo; o método informa como o objetivo pode ser alcançado; os resultados esperados mostram o que muda no mundo após o objetivo ser atingido; o capítulo de revisão bibliográfica vai apresentar os conceitos necessários para a compreensão do objetivo e os trabalhos relacionados ao objetivo.

3.1. Escolhendo o Objetivo de Pesquisa

A escolha de um objetivo de pesquisa é frequentemente a tarefa mais difícil em um trabalho de mestrado ou doutorado.

Muitas vezes, o objetivo é confundido com o tema da pesquisa. O tema da pesquisa pode ser a influência da batata inglesa na língua portuguesa. Mas nesse caso qual é o objetivo? Ou seja, o que se quer provar? *demonstrar?*

O objetivo normalmente comporta uma *hipótese de trabalho*. Um bom objetivo de pesquisa normalmente terá a forma “demonstrar que a hipótese x é verdadeira”.

Nem todo objetivo pode ser considerado um bom objetivo de pesquisa. Por exemplo, algo do tipo “o objetivo deste trabalho é aumentar os meus conhecimentos na área de estudo” pode até ser muito sincero, mas não convence ninguém de que algum conhecimento novo para a humanidade será produzido. Portanto, isso deve ser evitado.

Outro objetivo algumas vezes encontrado é a forma “propor...”. Alguma coisa é proposta, normalmente um método, uma abordagem, uma técnica, um algoritmo, uma comparação, ou qualquer outra coisa. A questão é: se o autor fizer a proposta, então o objetivo estará atingido? Se o aluno se propõe a propor e propôs, então está proposto! O que for proposto não é necessariamente melhor ou diferente daquilo que já existia antes. Então, o estágio da pesquisa neste caso ainda é dos mais ingênuos.

É necessário que o objetivo diga que aquilo que está sendo proposto é melhor do que alguma outra coisa ou que resolve algum problema que antes não podia ser resolvido.

Segundo Chinneck (1988), a descrição de um problema de pesquisa tem três partes:

- Um enunciado preciso da questão ou problema de que trata a monografia.
- Uma explicação por referência direta à bibliografia de que tal questão de pesquisa ainda não foi tratada.
- Uma discussão sobre por que é importante tratar essa questão de pesquisa.

O item (b) falhará se o aluno não conseguir deixar claro que a questão de pesquisa nunca foi tratada. Uma boa revisão bibliográfica é necessária para apresentar tal justificativa com suficiente autoridade. Devem ser evitadas afirmações do tipo “não encontrei nada parecido”. O aluno deve sempre mostrar o que encontrou nas fontes relevantes que examinou e comparar aquilo que foi encontrado com aquilo que ele pretende fazer. Se ele disser que não encontrou nada, a banca examinadora provavelmente pensará que ou ele não pesquisou direito ou está tratando um problema de pouco interesse. Em alguns casos pode ocorrer que efetivamente nada de muito semelhante seja encontrado, mas em todo o caso, sempre existe algum problema que possa ser considerado o mais próximo possível. Haverá situações em que serão encontradas abordagens quase idênticas, variando em poucos detalhes; e, em outros casos, a abordagem mais próxima será tão distinta que será necessária uma boa dose de explicação para que se possa entender por que ela é relevante. Leonardo da Vinci não conhecia a tecnologia utilizada pelos modernos aviões, então baseou seus estudos sobre máquinas voadoras no modelo mais próximo que estava disponível na sua época, a estrutura de voo dos pássaros.

Por outro lado, é importante que um trabalho de pesquisa se relacione principalmente com conhecimentos que sejam inerentes ao estado da arte, ou seja, que sejam recentes. Não faria muito sen-

tido hoje, com todos os conhecimentos de engenharia aeronáutica disponíveis, ignorar essas informações e tentar construir máquinas voadoras baseadas nos modelos de Leonardo.

Apesar disso, os conhecimentos antigos não devem ser de todo ignorados. Muitas vezes o conhecimento antigo aliado ao estado da arte pode produzir resultados muito interessantes. No caso da indústria aeronáutica, por exemplo, tenta-se hoje produzir aviões flexíveis que, de certa forma, imitam as estruturas de voo dos pássaros. Mas isso é feito com base em conhecimentos atuais, não se está reinventando a roda, mas aperfeiçoando conceitos já existentes.

3.1.1. O Caminho para a Escolha de um Objetivo de Pesquisa

Para que alguém seja capaz de pensar em um objetivo de pesquisa relevante, essa pessoa deve conhecer a área de pesquisa na qual está trabalhando. Portanto, o caminho lógico consiste de três passos:

- Escolher um *tema* de pesquisa, ou seja, uma área de conhecimento na qual se vai trabalhar.
- Realizar a *revisão bibliográfica*. A não ser que o autor já seja especialista na área escolhida, ele vai precisar ler muitos trabalhos já publicados nessa área para saber o que está sendo feito (estado da arte) e o que ainda precisa ser feito (problemas em aberto).
- Definir o *objetivo* de pesquisa. Uma vez feita a revisão bibliográfica, o objetivo de pesquisa possivelmente será fortemente relacionado com um dos problemas em aberto verificados no passo anterior.

Em poucos casos, a revisão bibliográfica será feita depois da definição do objetivo de pesquisa. Não é razoável conceber que alguém decida por um caminho de pesquisa sem conhecer exatamente a área onde vai atuar. Essa atitude possivelmente levará a objetivos mal definidos e que precisarão ser revistos quando o pesquisador perceber que está reinventando a roda.

Mas é possível ainda que os passos (b) e (c) devam ser repetidos algumas vezes de forma a refinar o objetivo (Figura 3.1). De

fato, ao realizar a revisão bibliográfica sobre um determinado tema, o pesquisador terá ideias sobre aspectos do tema que ainda não foram explorados, e esses aspectos darão origem a um objetivo de pesquisa. Mas antes de começar a gastar energia tentando atingir esse objetivo, o pesquisador deve ainda refinar sua pesquisa bibliográfica tentando verificar se tal objetivo já não foi perseguido em trabalhos anteriores e que tipo de resultados foi obtido.

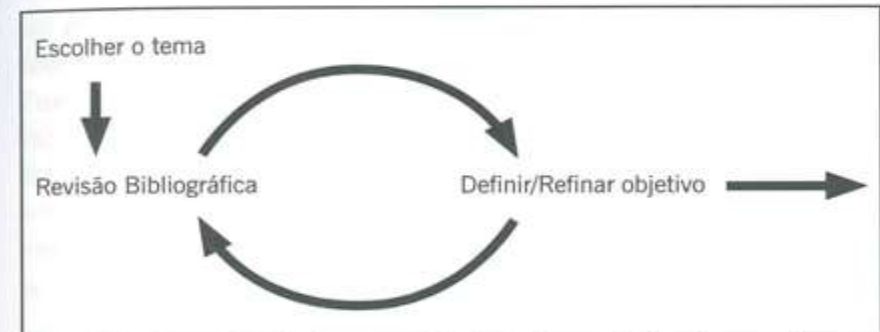


Figura 3.1: O caminho lógico para a definição de um objetivo de pesquisa.

Na história da catapulta, o tema de pesquisa foi o rio que cortava a cidade. Quando foi estabelecido o objetivo de pesquisa “encontrar uma forma segura de atravessar o rio”, o aluno deveria ter verificado na literatura os principais trabalhos já publicados sobre travessia de rios. Mas ele se limitou apenas a informações sobre a composição dos rios e, portanto, perdeu informações importantíssimas para sua pesquisa, ou seja, que já existiam métodos para atravessar rios e que ele deveria primeiro tentar melhorar esses métodos já existentes, ou pelo menos conhecê-los e seus defeitos, antes de enveredar pela pesquisa de algo totalmente novo. No extremo desse raciocínio fica a possibilidade de que alguns anos antes outros pesquisadores já tivessem tentado usar a catapulta para cruzar rios, sem saber um do outro. A cada vez que o estudo é repetido, esforços são despendidos, medidas tomadas e conclusões (possivelmente as mesmas) são obtidas, sem que nenhum avanço ocorra para a ciência.

3.1.2. O Tema

O tema da pesquisa frequentemente depende do interesse do aluno e do orientador. Não se recomenda, em hipótese alguma, uma pesquisa cujo tema não seja compatível com os conhecimentos do orientador. No caso do aluno, recomenda-se que ao passar da graduação ao mestrado e do mestrado ao doutorado procure trabalhar no mesmo tema, embora, buscando objetivos distintos. Por que isso? Porque, se o aluno se mantiver no mesmo tema, o passo de revisão bibliográfica será concluído mais rapidamente ao passar de um estágio para outro em sua formação. Bastará que ele se atualize nos últimos desenvolvimentos da área para poder decidir por um bom objetivo de pesquisa. Se o aluno mudar de tema, então terá de fazer toda a revisão bibliográfica sobre outro tema, o que lhe tomará muito tempo.

Não é impossível que uma pessoa com formação em uma área faça mestrado ou doutorado em uma área totalmente diferente, mas com certeza essa pessoa terá muito mais trabalho e levará muito mais tempo para amadurecer os conceitos da nova área do que uma pessoa que já tem a vivência na área.

O tema pode ser especializado a partir de uma grande área em subáreas cada vez mais específicas. Por exemplo:

1. Ciência da Computação.
 - 1.1. Inteligência Artificial.
 - 1.1.1. Métodos de busca.
 - 1.1.1.1. Busca heurística.
 - 1.1.1.1.1. Algoritmo A*.

Nessa lista, cada item é uma especialização do item anterior, mas cada um é apenas um tema de pesquisa, embora cada vez mais específico.

Indo para outra direção, pode-se combinar um tema de pesquisa com uma área de aplicação. O tema, possivelmente, será mais específico do que geral. Não faz muito sentido, por exemplo, falar de “aplicação da Ciência da Computação no problema da pavimentação

das estradas”. Faria mais sentido um tema de pesquisa como “aplicação de busca heurística no problema do transporte de máquinas para pavimentação de estradas”.

Mesmo assim, em alguns casos encontram-se trabalhos cujo tema é algo do tipo “aplicação de técnicas de Inteligência Artificial para resolver o problema x”. O pesquisador deve ter a noção de proporção para saber se o nível de especificidade do tema de pesquisa é adequado ou não. No exemplo anterior, ao se falar de Inteligência Artificial, abre-se um leque enorme de possibilidades (pode-se conferir como a área é extensa no livro de Russel e Norvig, 1995). Portanto, um tema tão extenso quanto este não é adequado. Ao se falar de busca heurística, porém, o leque é reduzido a alguns poucos algoritmos conhecidos, e o tema fica mais viável. Quanto mais amplo o tema, maior a quantidade de livros e artigos que terão de ser lidos. Portanto, recomenda-se buscar temas cada vez mais específicos antes de se propor um objetivo de pesquisa. Quando se escolhe um tema de pesquisa que tem aplicação em outra área, deve-se tomar cuidado. Quando se está realizando um mestrado ou doutorado em Ciência da Computação, deve-se observar que a principal contribuição do trabalho deve ser para a área da Computação. Ou seja, o problema a ser resolvido deve estar ligado à não-existência ou inadequação das ferramentas de computação existentes, e não aos aspectos ou técnicas da área de aplicação (pelo menos não predominantemente). O aluno, nesse caso, deve evitar a tentação de *contribuir* para uma área que ele não conhece ou que conhece muito pouco. Por exemplo, uma monografia sobre Informática aplicada à Medicina deve apresentar contribuições em relação às ferramentas de Computação aplicada à Medicina e às melhorias que podem ser obtidas. Dificilmente o aluno de Computação irá propor e defender uma nova técnica cirúrgica. Não quer dizer que seja impossível que isso aconteça, nem quer dizer que o aluno não seja execrado na banca por algum médico que esteja ali presente, mas é um risco a ser evitado.

3.1.3. O Problema

Uma monografia deve apresentar uma solução para um problema. Inicialmente, portanto, um problema deve ser identificado. Seria errado iniciar a monografia simplesmente resolvendo criar um novo método para isso ou aquilo.

No caso da catapulta, o aluno propôs e testou um novo método para cruzar rios. Aqui vem uma questão interessante. Segundo o aluno, o problema consistia em cruzar o rio. Mas esse problema identificado já não é mais um problema sem solução, porque existem diferentes abordagens para cruzar um rio: ponte, balsa, teleférico etc. Então, caso o aluno quisesse insistir neste tema, teria de indicar quais problemas as soluções existentes apresentam. Ou seja, quais os problemas encontrados quando se tenta cruzar um rio com uma ponte ou com uma balsa? Ele poderia descobrir, por exemplo, que pontes são muito caras e balsas são muito lentas. Se todas as soluções existentes apresentarem algum tipo de problema, então é possível que se esteja abrindo caminho para uma nova abordagem. Caso contrário, as pessoas continuarão a cruzar rios com os meios usuais e não com a *nova* abordagem.

Algumas propostas de pesquisa são apresentadas inicialmente sem ter um problema claramente identificado. Por exemplo: “este trabalho propõe usar a metáfora de formigueiro para modelar pacotes em uma rede”. Esse tema até pode vir a ser um trabalho interessante, mas qual é o problema que essa modelagem vai resolver? O que há de errado com outras formas de modelagem, sejam elas quais forem, que essa metáfora do formigueiro vai possivelmente resolver?

Segundo Griffiths (2008), se o autor não consegue estabelecer claramente qual é o problema tratado em sua monografia, será muito difícil para outras pessoas especularem sobre os possíveis usos dela. Também será difícil avaliar se ela obteve sucesso.

3.1.4. Perspectiva Profissional

O tema de pesquisa a ser escolhido, em qualquer nível de formação, além de ser do agrado do aluno, deve estar relacionado com a sua perspectiva de desenvolvimento profissional.

Não faz sentido gastar vários anos investindo em uma pesquisa na área X, para depois trabalhar o resto da vida na área Y. O ideal seria que cada pesquisa gerasse algum tipo de produto ou uma melhoria a um produto existente que pudesse, após o final do curso, gerar algum tipo de perspectiva de trabalho e de renda para o pesquisador.

Isso é mais difícil quando se trata de pesquisa básica, mas mesmo assim esse tipo de pesquisa pode gerar patentes que em algum momento poderão se tornar produtos e gerar riquezas para o país. Não é admissível que todo um esforço despendido durante o mestrado ou mesmo durante um doutorado acabe sendo depois simplesmente guardado em uma gaveta.

3.2. A Revisão Bibliográfica

A revisão bibliográfica, conforme já comentado, não produz conhecimento novo, mas apenas supre as deficiências de conhecimento que o pesquisador tem em uma determinada área. Portanto, ela deve ser muito bem planejada e conduzida.

Supondo que o pesquisador seja praticamente iniciante em uma determinada área, ele deveria iniciar a pesquisa lendo algum tipo de *survey* sobre o assunto. Não é difícil encontrar artigos ou mesmo livros que abordem toda uma área de conhecimento na forma de *survey*. Os livros, aliás, servem exatamente a esse propósito.

Pode-se iniciar a pesquisa com uma leitura de trabalhos mais abrangentes que deem uma visão do todo para depois ir se aprofundando cada vez mais em temas cada vez mais específicos.

Quando se faz uma pesquisa em que alguma técnica de computação é aplicada a alguma outra área do conhecimento, é necessário que se faça a revisão bibliográfica sobre a técnica em si, sobre a área de aplicação e, mais do que tudo, sobre as aplicações que já foram tentadas com essa técnica ou com técnicas semelhantes na mesma área ou em áreas equivalentes. Exemplificando, um aluno pretende desenvolver um sistema multiagentes para auxiliar controladores de voo. Esse aluno deve conhecer profundamente os sistemas multiagentes e deverá conhecer também os problemas que os controla-

dores de voo enfrentam para exercer sua profissão. Porém, ele não deve pensar, como algumas vezes acontece, que essa é a primeira vez que alguém vai tentar desenvolver um sistema multiagentes para esse tipo de aplicação. O aluno da catapulta, mencionado anteriormente, estudou rios e estudou catapultas, mas não procurou saber se alguém já tinha tentado atravessar um rio usando uma catapulta. Se essa pesquisa existisse e ele tivesse acesso a ela, teria visto que os resultados não eram animadores e talvez tivesse escolhido outro tema de pesquisa antes de ter dedicado a maior parte do seu tempo de mestrado a algo infrutífero.

Uma monografia sobre pesquisa aplicada que apresenta uma boa revisão bibliográfica da ferramenta de computação e da área de aplicação, mas que não menciona nenhuma tentativa anterior de aplicação dessa ferramenta na área sofre da “Síndrome da Intersecção Esquecida”. Uma monografia com esse problema estará possivelmente “reinventando a roda”.

A Síndrome da Intersecção Esquecida em geral é justificada pelos alunos com frases do tipo “mas não encontrei nada parecido com o que estou fazendo”. Esse raciocínio negativo deve ser evitado. Nunca se deve dizer que não se achou nada semelhante. Algo sempre deve ser apresentado como referência. Essa referência poderá ser mais semelhante ou menos semelhante à abordagem usada de um ponto de vista relativo. Mas sempre a abordagem mais semelhante de todas (por menos semelhante que seja) deve ser buscada.

Pode ser interessante pensar assim: “Ninguém fez algo parecido com o que estou fazendo, *mas* muitas coisas já foram feitas pelos seres humanos ao longo da sua história. Então, eu poderia classificar as coisas que já foram feitas em termos de grau de semelhança com aquilo que estou fazendo. As coisas mais parecidas com o meu trabalho serão minha referência, mesmo que a semelhança seja pequena”.

Assim, evita-se o *fundamento vazio*, ou seja, dizer que seu trabalho é original porque ninguém nunca fez nada parecido. Não se deve fundamentar todo um trabalho de pesquisa em uma negação.

Deve-se mostrar o que outros fizeram, e depois mostrar que o trabalho feito é diferente ou melhor do que essas referências.

3.2.1. Fichas de Leitura

Durante todo o processo de leitura, é fundamental que sejam feitas anotações. Conceitos-chave e ideias novas devem ser anotados sempre que forem detectados na leitura. É necessário que se saiba de onde tais ideias e conceitos saíram. Em geral, inicia-se uma ficha de leitura, seja em papel, seja no computador, escrevendo a referência bibliográfica da obra sendo consultada. Em seguida são feitas as anotações relevantes.

Essas fichas serão extremamente importantes no futuro para saber de onde as ideias saíram. Depois de ler algumas dezenas de artigos sobre um determinado assunto, será difícil lembrar-se de onde saíram determinadas ideias.

Porém, deve-se ter em mente que o conjunto das fichas de leitura não é o mesmo que o capítulo de revisão bibliográfica. As fichas são apenas um registro, com memória, de leituras feitas, que é organizado por fonte bibliográfica. Esse trabalho normalmente será desenvolvido antes da definição do objetivo de pesquisa, pois, como foi comentado, consiste exatamente na busca de informações para que esse objetivo possa ser definido. Já o capítulo de revisão bibliográfica será organizado após a execução da pesquisa. Ou seja, após definir o objetivo, definir o método de trabalho, executar os experimentos, coletar os resultados e esboçar as conclusões, é que o pesquisador vai organizar o capítulo de revisão bibliográfica onde ele vai inserir os principais conceitos e trabalhos correlatos relevantes para a compreensão da sua pesquisa.

Se as fichas de leitura são organizadas por fonte bibliográfica, o capítulo de revisão bibliográfica não deve ser organizado dessa forma, mas por conceitos. Ou seja, em vez de dizer tudo o que um autor pensa sobre vários conceitos, deve-se dizer o que vários autores pensam sobre um conceito de cada vez.

3.2.2. Tipos de Fontes Bibliográficas

Há vários tipos de fontes bibliográficas. Cada uma terá sua utilidade em determinados momentos da pesquisa. Os *livros* normalmente contêm informação mais completa, didática e bem amadurecida. O objetivo do livro é justamente apresentar uma determinada área da ciência de forma didática e bem fundamentada.

Raramente serão encontradas em livros informações sobre trabalhos futuros, que conduzem a ideias de pesquisa.

Alguns livros se dedicam especificamente à apresentação de problemas em aberto em determinadas áreas, mas estes não são maioria. Em geral, tais livros são produto de eventos científicos. A maioria dos livros didáticos procura apresentar apenas a informação que já está consolidada.

Ideias de pesquisa serão encontradas mais facilmente em artigos curtos que normalmente são publicados em *eventos* ou *periódicos*. A maioria das ciências exatas preza mais a publicação em periódicos. A área de Computação, porém, tem características distintas em relação a esse aspecto, visto que considera publicações em eventos como sendo tão importantes, e muitas vezes até mais importantes, que publicações em periódicos.

Isso significa que, na área de Computação, bons artigos poderão ser encontrados tanto em eventos quanto em periódicos. Que diferenças podem então ser esperadas?

O processo de submissão e publicação em eventos e periódicos é diferente. Por isso, pode-se esperar diferentes tipos de artigos nesses dois veículos.

Eventos normalmente têm uma data-limite para entrega de trabalhos. De um conjunto de trabalhos enviados e avaliados, os melhores são encaminhados para publicação. Em geral, são sugeridas algumas modificações, mas dificilmente uma segunda rodada de avaliação é realizada.

Já no caso de periódicos, não existem datas-limite, a não ser no caso de edições especiais. Os artigos são submetidos, avaliados, e revisões são sugeridas. Posteriormente, os artigos são enviados de

novo, avaliados e assim por diante. Esse processo de ida e volta pode acontecer várias vezes e em geral pode levar alguns anos até que o artigo seja publicado.

Então, a diferença que se pode esperar é a seguinte: os artigos em eventos terão informações mais atuais, mas poderão variar bastante em termos de qualidade. Já os artigos em periódicos terão sido arduamente revisados e lapidados ao longo de iterações entre autores e revisores, mas quando publicados talvez já não sejam mais tão atuais quanto os artigos em eventos.

Fica a ressalva de que toda regra tem exceção, porque existem periódicos que publicam artigos rapidamente e eventos cujo índice de exigência é tão alto que publicará apenas artigos do mais alto grau de excelência.

3.2.3. O que deve ser Necessariamente Lido

1) O pesquisador iniciante em uma determinada área deverá começar sua revisão bibliográfica pelos *surveys*. Livros sobre o assunto também seriam uma boa escolha nesse momento. Algumas ferramentas de pesquisa bibliográfica, como *citeseer* (<http://citeseer.nj.nec.com/impact.html>), permitem uma busca específica por artigos do tipo *survey*. Além disso, existem publicações especializadas nesse tipo de artigo.

Essas coletâneas apresentam ao pesquisador o estado da arte da área de pesquisa e sua evolução histórica, indicando diferentes desdobramentos e as principais realizações.

2) Na sequência, alguns trabalhos clássicos devem ser buscados, normalmente destacados nos *surveys*. Além disso, quando se usa uma ferramenta de pesquisa bibliográfica eletrônica, é possível, muitas vezes, solicitar que os trabalhos publicados em uma determinada área sejam ordenados a partir de sua importância. Um trabalho será mais importante na medida em que for citado por outros trabalhos. Os trabalhos clássicos são aqueles que já receberam ao longo do tempo o maior número de citações.

3) Continuando a pesquisa, deverão ser buscadas as fontes mais recentes sobre o assunto da pesquisa. Artigos muito atuais dificilmente serão clássicos, pois ainda não terá havido tempo para que sejam citados em outras publicações. Porém, é fundamental que um trabalho de pesquisa tenha como referência também os desenvolvimentos mais recentes na área.

3.2.4. Leitura Crítica

A leitura de trabalhos científicos não deve ser encarada apenas como um aprendizado. O pesquisador deve exercer, antes de tudo, o espírito crítico, para questionar a validade de todas as informações registradas nos textos que estão sendo lidos. A aceitação passiva de tudo aquilo o que é lido não gera no pesquisador o espírito de busca por novas informações.

Para que o tema de pesquisa se transforme em um objetivo, é necessário que a cada instante a leitura produza questionamentos. Sem perguntas não há respostas. Muitas perguntas que o pesquisador fizer a si mesmo ao longo da leitura de um texto possivelmente ainda não terão respostas e serão, portanto, excelentes candidatas a objetivo de pesquisa.

Algumas perguntas-chave poderão ajudar o leitor a transformar uma leitura passiva em uma leitura rica e geradora de ideias para pesquisa. Entre elas:

- a) De onde o autor parece tirar suas ideias?
- b) O que foi obtido como resultado deste trabalho?
- c) Como este trabalho se relaciona com outros na mesma área?
- d) Qual seria um próximo passo razoável para dar continuidade a essa pesquisa?
- e) Que ideias de áreas próximas poderiam ser aproveitadas neste trabalho?

As perguntas geradoras mencionadas também poderiam ser usadas para avaliar a qualidade do trabalho que está sendo lido. No caso da primeira pergunta, se não for possível descobrir de onde o autor tira suas ideias, provavelmente se estará diante de um trabalho

fraco, pois as ideias devem vir de referências bibliográficas, ou da observação de fenômenos, ou então são hipóteses criadas pelo autor, as quais serão comprovadas ao longo do trabalho.

Em geral, os autores não podem simplesmente escrever frases como “o interesse pela Internet tem crescido muito ao longo dos últimos anos”. Uma afirmação como esta, embora, à primeira vista, seja consensual, não pode deixar de ter uma base. Essa base pode ser uma referência a outro trabalho, que tenha realizado uma pesquisa sobre o assunto. A base também pode ser um levantamento estatístico realizado pelo próprio autor que demonstra a validade da afirmação. Mais adiante será visto que essa frase específica ainda comporta vários outros problemas, por exemplo:

- a) Como se define e se mede “interesse”?
- b) Como se conceitua “crescer muito”?
- c) Que período de tempo compreende os “últimos anos”?

Em relação à segunda pergunta geradora, “o que exatamente foi obtido como resultado deste trabalho em particular?”, se não houver possibilidade de resumir em poucas palavras a contribuição real do trabalho, então, possivelmente, o texto será confuso e mal organizado, não deixando clara a efetiva contribuição do artigo para a comunidade científica.

Em relação à terceira pergunta geradora, “como este trabalho se relaciona com outros na mesma área?”, espera-se, a princípio, que o próprio artigo deixe bem claro, citando adequadamente os trabalhos correlatos. Caso isso não seja feito, o leitor poderá tentar estabelecer as relações entre o trabalho lido e outras obras. Muitas vezes, aspectos importantes sobre o trabalho (falhas) são descobertos através dessas comparações.

Em relação à quarta pergunta geradora, “qual seria um próximo passo razoável para dar continuidade a essa pesquisa?”, a resposta poderá ser um excelente objetivo de pesquisa. Muitas vezes as questões de pesquisa já estão colocadas no trabalho pelos autores na esperança de que outros grupos deem continuidade à pesquisa.

A quinta pergunta, “que ideias de áreas próximas poderiam ser aproveitadas neste trabalho?”, traz ao pensamento do pesquisador possíveis melhoramentos ao trabalho sendo estudado em função de conceitos correlatos de outras áreas de pesquisa, que possivelmente não eram conhecidos dos autores do trabalho. Dessa forma, o sucesso da aplicação de algum desses conceitos correlatos no trabalho em questão poderá dar origem a uma interessante hipótese de pesquisa, que, se tiver uma justificativa plausível, poderá ser um excelente objetivo de pesquisa.

3.2.5. Exposição à Pesquisa

Além da leitura, o pesquisador, na fase de geração de ideias, deverá estar constantemente exposto a um ambiente científico. No caso de alunos de mestrado e doutorado, na fase de elaboração do objetivo de pesquisa, é fundamental que se tente, pelo menos uma vez por semana, gerar uma ideia de pesquisa para ser discutida com o orientador.

Regularmente, o pesquisador deverá ler pelo menos os resumos dos artigos publicados nos principais periódicos e eventos na sua área de pesquisa. Além dos resumos, deve-se procurar ler pelo menos um ou dois artigos de maior relevância para a área de pesquisa.

Além disso, sempre que possível, o pesquisador deverá participar de palestras e seminários nos quais poderá trocar ideias com outros pesquisadores, além de observar a forma de trabalho de outros grupos de pesquisa. No caso de alunos de mestrado e doutorado, isso implica também participar, como ouvinte, do maior número possível de defesas de teses e dissertações, mesmo que não sejam referentes à sua área de pesquisa específica.

3.2.6. A Ideia de Pesquisa

A ideia de pesquisa, ou problema, surge a partir da leitura e observação. Pode-se comparar a ideia de pesquisa a uma semente que germina no solo da revisão bibliográfica. Com uma revisão bi-

bliográfica superficial ou inexistente até se pode ter ideias, mas possivelmente serão ideias fracas, que dificilmente se desenvolverão de forma saudável. Já com uma boa revisão bibliográfica, tem-se um solo forte para germinação de ideias consistentes.

Portanto, antes de tentar ter ideias, o aluno deve ter experimentado uma boa carga de leitura relacionada ao tema.

3.2.7. Como Sistematizar a Pesquisa Bibliográfica

Um dos problemas que os alunos eventualmente encontram, conforme já mencionado, é justificar que não encontraram nada efetivamente ligado ao assunto que desejam pesquisar. Buscas desorganizadas dificilmente levarão a bons resultados e dificilmente darão segurança ao aluno quando lhe for perguntado “mas será que alguém já não fez isso?”.

Fica aqui então uma sugestão sobre como proceder à pesquisa bibliográfica de maneira sistemática. Cada aluno poderá mudar essas sugestões de acordo com suas necessidades ou disponibilidade. Sugere-se seguir estes passos:

- Listar os títulos de periódicos e eventos relevantes para o tema de pesquisa e os títulos de periódicos gerais em computação que eventualmente possam ter algum artigo na área do tema de pesquisa.
- Obter a lista de todos os artigos publicados nos últimos cinco (ou mais) anos nesses veículos.
- Selecionar dessa lista aqueles títulos que tenham relação com o tema de pesquisa.
- Ler o *abstract* desses artigos e, em função da leitura, classificá-los como relevância “alta”, “média” ou “baixa”.
- Ler os artigos de alta relevância e fazer fichas de leitura anotando os principais conceitos e ideias aprendidos. Anotar também títulos de outros artigos possivelmente mencionados na bibliografia de cada artigo (mesmo que com mais de cinco anos) e que pareçam relevantes para o trabalho de pesquisa. Incluir

esses artigos na lista dos que devem ser lidos (inicialmente o *abstract* e, se for relevante, o artigo todo).

- f) Dependendo do caso, ler também os artigos de relevância média e baixa, mas iniciando sempre pelos de alta relevância.

O conjunto de referências mencionadas no terceiro passo deve ser efetivamente produzido na forma de uma listagem que possa ser posteriormente consultada ou até indicada na bibliografia do trabalho como fonte de pesquisa. Portanto, essa fase do processo deve ser bem documentada.

Depois do último passo, o aluno poderá decidir se:

- Já tem material suficiente para elaborar uma ideia de pesquisa consistente.
- Precisa expandir a pesquisa examinando artigos mais antigos (expandindo o passo "b") ou periódicos menos relevantes (expandindo o passo "a").

A decisão sobre se já há material suficiente para finalizar a revisão bibliográfica usualmente será tomada em conjunto com o orientador.

Uma última técnica ainda para expandir a revisão bibliográfica, quando necessário, é consultar as referências bibliográficas citadas nos trabalhos mais importantes consultados.

3.2.8. Como Terminar a Revisão Bibliográfica

Em geral, existe sempre conhecimento novo sendo produzido em todas as áreas da ciência. A revisão bibliográfica, portanto, praticamente não termina nunca.

Mas o trabalho de pesquisa deve ter um término claro. E é necessário saber que, durante o seu desenvolvimento, apenas a revisão bibliográfica não produz nenhum conhecimento novo.

É necessário, portanto, passar gradualmente da fase de revisão bibliográfica para a fase de realização propriamente dita do trabalho de pesquisa, ou seja, a construção de teorias e realização de experimentos para testar hipóteses.

3.3. O Objetivo

O objetivo da pesquisa deve ser diretamente verificável ao final do trabalho. Um bom objetivo de pesquisa possivelmente irá demonstrar que alguma hipótese sendo testada é ou não verdadeira.

Portanto, o objetivo geral e os objetivos específicos do trabalho devem ser expressos na forma de uma condição não trivial cujo sucesso possa vir a ser verificado ao final do trabalho. Um objetivo bem expresso em geral terá verbos como "demonstrar", "provar", "melhorar" (de acordo com alguma métrica definida) etc.

Deve-se tomar cuidado com certos verbos que determinam objetivos cuja verificação é trivial e, portanto, inadequada. Entre eles pode-se citar "propor", "estudar", "apresentar" etc. Se o objetivo do trabalho é *propor* algo, basta que a coisa seja proposta para que o objetivo seja atingido e, portanto, essa forma é trivial e inadequada, pois a definição do objetivo não menciona a qualidade daquilo que será proposto.

Se o objetivo do trabalho é *estudar* algo, então ele terá sido alcançado se aquilo foi estudado, não importando se alguma nova informação foi aprendida ou não, sendo, portanto, inadequado como objetivo de pesquisa. Estudar, normalmente, é o objetivo do *aluno* e não do trabalho.

Se o objetivo do trabalho consiste em *apresentar* algo, novamente ele é trivial e inadequado. Uma simples apresentação não produz necessariamente conhecimento novo. Por exemplo, "o objetivo deste trabalho é apresentar os operadores da lógica booleana"; tal objetivo pode ser alcançado com um pequeno texto explicando os operadores conhecidos, mas, como não traz informação nova, não é um objetivo de pesquisa.

A proposta, o estudo e a apresentação podem ser justificáveis como objetivo de pesquisa desde que o objeto da proposta, estudo ou da apresentação seja algo original.

Segundo Chinneck (1988), uma monografia deve apresentar uma contribuição original ao conhecimento. Dessa forma, ao final do trabalho, o estudante deverá ser capaz de mostrar que identifi-

cou um problema que valia a pena ser resolvido, mas que ainda não havia sido. Além disso, o estudante deverá mostrar que ele resolveu o problema que propôs e apresentar a solução.

Em função disso, Chinneck conclui que um avaliador, ao ler o texto de uma monografia, vai procurar responder às seguintes questões:

- Qual é a questão de pesquisa que o aluno propôs?
- É uma boa questão? (Já foi respondida alguma vez? Vale a pena respondê-la?)
- O aluno conseguiu convencer que a questão foi respondida adequadamente?
- O aluno fez uma contribuição adequada ao conhecimento?

A falha em encontrar respostas para alguma dessas questões poderá colocar o aluno em apuros, sendo que a banca avaliadora provavelmente exigirá revisões extensas no trabalho ou poderá até reprovar o candidato.

3.3.1. A Extensão do Objetivo de Pesquisa

Um objetivo de pesquisa, dependendo do nível de formação pretendido (graduação, especialização, mestrado ou doutorado), não pode ser demasiadamente trivial nem demasiadamente complexo.

Um objetivo demasiadamente trivial será atingido rapidamente, mas dificilmente será defensável perante uma banca, pois a cada grau acadêmico espera-se do aluno não apenas inspiração, mas também transpiração, ou seja, não basta ter uma boa ideia, é necessário trabalhar sobre ela demonstrando todos os seus diferentes aspectos dentro do nível de complexidade exigido pelo grau almejado.

Um objetivo demasiadamente complexo dificilmente será atingido no tempo disponível para o cumprimento dos requisitos do grau almejado. Objetivos complexos podem ser colocados como projetos de longo prazo na carreira de pesquisadores já formados, os quais terão, muitas vezes, décadas à frente e equipes de pesquisa para buscar esses objetivos.

Alunos de graduação e pós-graduação devem atingir os objetivos colocados dentro do tempo regulamentar que seus cursos estabelecem, e, portanto, a complexidade desses objetivos deve ser consistente com esse tempo. Não basta, para a obtenção de um grau, apresentar uma ideia complexa que não pode ser demonstrada ou concluída no tempo disponível. Alegar para a banca que não houve tempo para concluir o trabalho dificilmente é uma boa desculpa.

Sendo assim, o aluno deverá sempre buscar apoio no seu orientador e a definição de um objetivo de pesquisa que possa ser buscado no tempo disponível, de acordo com o curso realizado. Quanto mais experiente for o orientador, ou seja, quanto mais trabalhos ele tiver orientado com sucesso em sua vida, mais se pode esperar que seja capaz de apoiar o aluno na escolha de um objetivo adequado.

3.3.2. Objetivo de Pesquisa *versus* Objetivo Técnico

É aceitável que um trabalho de graduação e mesmo de especialização tenham objetivos técnicos, ou seja, espera-se nesses graus que os alunos sejam capazes de demonstrar que aprenderam determinados conceitos e conseguem colocá-los em prática. Assim, é aceitável que um aluno de graduação, ao final de seu curso, desenvolva um sistema usando conceitos aprendidos durante o curso e que apresente o sistema como trabalho final. Porém, esse tipo de abordagem não é aceitável nos níveis de mestrado e doutorado. Espera-se que o mestre e o doutor tenham de alguma forma contribuído para o avanço do conhecimento. Assim, o desenvolvimento de um sistema, embora possa ser necessário para provar alguma hipótese previamente estabelecida, não é em si suficiente para a concessão de um título de mestre ou doutor.

Se o simples desenvolvimento de um sistema ou de um protótipo fosse suficiente para outorgar um grau acadêmico, então as universidades deveriam estar conferindo título de mestre a todos os programadores ou analistas que diariamente desenvolvem sistemas, muitas vezes complexos, nas suas empresas. Se isso não acontece é porque existe algo mais nas dissertações e teses que precisa ser

buscado. Esse algo mais é o *conhecimento*. Uma monografia é um documento que apresenta de forma organizada uma contribuição para o estado da *arte*, apresentando, portanto, informações que não eram conhecidas e que a partir do momento em que são publicadas passam a integrar o corpo de conhecimento relevante para quem for atuar em uma determinada área.

Os trabalhos *técnicos*, que são aceitáveis na graduação e na especialização, apenas usam o conhecimento já disponível.

Os trabalhos *científicos*, que devem obrigatoriamente ser desenvolvidos no mestrado e no doutorado, devem, além de usar o conhecimento já disponível, criar novos conhecimentos, associando-os dentro de uma estrutura coerente àqueles que já são conhecidos. Sendo assim, o desenvolvimento de sistemas ou protótipos servirá apenas como apoio para demonstrar a aplicabilidade desses novos conhecimentos, se assim for necessário.

3.3.3. Os Objetivos Específicos

Os objetivos específicos devem ser escolhidos da mesma forma que o objetivo geral, ou seja, devem ser não triviais e verificáveis ao final do trabalho. Normalmente, os objetivos específicos não são etapas do trabalho, mas subprodutos. Deve-se tomar cuidado para não confundir os objetivos específicos com os passos do método de pesquisa.

A implementação de um protótipo ou a coleta de dados empíricos possivelmente serão etapas dentro de um trabalho e, portanto, parte do método de trabalho. Esses passos não são, assim, objetivos específicos.

Deve-se entender, portanto, que os objetivos específicos são detalhamentos ou subprodutos do objetivo geral. Se o objetivo geral consiste em provar uma determinada hipótese, os objetivos específicos podem estabelecer a prova de uma série de condições associadas a tal hipótese.

3.4. O Método de Pesquisa

Em geral as monografias têm um capítulo ou seção designados como “metodologia”. Entretanto, *metodologia* seria o estudo dos métodos. Apesar do uso corrente, linguisticamente seria mais correto afirmar que um trabalho científico individualmente tem um *método* de pesquisa e não uma metodologia.

Segundo a Wikipédia (2009a), “A Metodologia é o estudo dos métodos. (...) Tem como finalidade captar e analisar as características dos vários métodos disponíveis, avaliar suas capacidades, potencialidades, limitações ou distorções e criticar os pressupostos ou as implicações de sua utilização.”

Difícilmente um aluno escrevendo uma monografia apresenta uma *metodologia* com essas características. O usual é que, dependendo do tipo de problema, se escolha um método de trabalho e siga utilizando-o até comprovar ou refutar as hipóteses. Assim, os termos *método* e *metodologia* serão aqui claramente diferenciados.

Ainda segundo a Wikipédia (2009b), “O método científico é um conjunto de regras básicas para desenvolver uma experiência a fim de produzir novo conhecimento, bem como corrigir e integrar conhecimentos preexistentes. Na maioria das disciplinas científicas consiste em juntar evidências observáveis, empíricas (ou seja, baseadas apenas na experiência) e mensuráveis e as analisar com o uso da lógica.”

Este livro trata de metodologia porque apresenta um estudo de métodos. Mas o livro em si teve um método para ser concebido e escrito, o qual não é aqui descrito. Porém, uma monografia deverá sempre descrever o método usado para chegar às conclusões.

O método propriamente dito de um trabalho científico só pode ser estabelecido depois que o objetivo tiver sido definido. Por esse motivo, no caso da Computação, normalmente a revisão bibliográfica não deveria nem fazer parte do método. A revisão bibliográfica consiste em um passo do trabalho no qual o aluno vai iniciar ou aprofundar seus conhecimentos em um campo do saber para que possa então propor um objetivo que seja coerente com o grau que deseja obter. Ou seja, a etapa de revisão bibliográfica não seria parte

do método, mas um *prerrequisito* para realização do trabalho de pesquisa, pois quem não estudou o assunto não tem como propor um objetivo válido.

O método consiste na sequência de passos necessários para demonstrar que o objetivo proposto foi atingido, ou seja, se os passos definidos no método forem executados, os resultados obtidos deverão ser convincentes.

O método deve então indicar se protótipos serão desenvolvidos, se modelos teóricos serão construídos, quais experimentos eventualmente serão realizados, como os dados serão organizados e comparados, e assim por diante, dependendo do tipo de trabalho.

A definição do método de pesquisa é um passo fundamental a ser executado logo após a definição do objetivo. Dado o objetivo, o método descreve o caminho para atingi-lo. Assim, deverá ser suficiente trilhar o caminho descrito pelo método para se alcançar o objetivo. Se o objetivo e o método foram coerentes, então grande parte do trabalho de pesquisa já terá sido executada, restando apenas a execução dos passos descritos no método.

Porém, descrever um conjunto de passos que constitua um método de trabalho científico aceitável exige alguns conhecimentos sobre o método científico que serão detalhados nesta seção. A não observância do método científico pode levar a conclusões erradas ou forçadas.

Propostas metodologicamente ingênuas não são totalmente incomuns em computação. Coisas como “trabalhar com dois grupos, um com a ferramenta e outro sem a ferramenta”, até poderia ser parte de um método, mas não é suficiente. Se a diferença entre as médias dos dois grupos for de 0,5 ponto percentual, pode-se concluir que um grupo foi melhor que o outro? Ou pode ter sido obra do acaso? E se a diferença for de cinco pontos percentuais? Como saber? Existem algumas informações trazidas pela estatística que devem ser do conhecimento de qualquer pessoa que se aventure a desenvolver pesquisa científica.

Além disso, existe ainda a possibilidade do mau uso da lógica. Os antigos sofistas eram bastante requisitados entre os gregos para argumentação. Entretanto, nem sempre a lógica usada por eles era exatamente a que poderia ser aceita cientificamente. Um exemplo hilário de uso tortuoso da lógica aparece no filme *Monty Python em Busca do Cálice Sagrado*, quando Bedevere propõe um método infalível para determinar a identidade de uma bruxa. Ao final de uma série de argumentações pseudológicas, ele conclui que, se a mulher pesar tanto quanto um pato, será uma bruxa. Basicamente, segundo Bedevere, bruxas queimam e madeira também queima. Madeira flutua na água assim como os patos. Logo, se a mulher pesar tanto quanto um pato é feita de madeira e, portanto, é uma bruxa.

3.4.1. Dados versus Conceitos

O método de pesquisa não consiste apenas em coletar dados para suportar a hipótese de trabalho. É necessário elaborar um discurso ponderado e esclarecedor a partir desses dados. O aspecto mais importante de uma monografia é o pensamento crítico e não apenas a coleta de informações. Trabalhos acadêmicos que se restringem à realização de pesquisas de opinião através de questionários com a consequente tabulação dos dados e apresentação de gráficos não terão validade se não trouxeram consigo alguma informação nova.

Lakatos e Marconi (2006) identificam o questionário como um instrumento de pesquisa que dispensa a presença do pesquisador. Porém, deve-se evitar a distribuição de questionários para serem respondidos se não se sabe com antecedência qual a informação que se está buscando, ou seja, qual a hipótese que se está tentando comprovar.

Além disso, em geral respostas diretas e interpretações simplistas não são as mais interessantes. Por exemplo, há algum tempo pesquisadores realizaram uma pesquisa na Inglaterra, onde entrevistaram homens e mulheres perguntando quantos parceiros sexuais haviam tido ao longo da vida. O resultado médio para as mulheres foi três e o resultado médio para os homens foi 10. Uma interpre-

tação ingênua e simplista desse fato diria que o homem em média tem mais parceiros do que as mulheres. Porém, considerando-se que existe aproximadamente o mesmo número de homens e mulheres na sociedade, isso é impossível, visto que cada vez que um homem tem uma nova parceira, uma mulher (a parceira) também tem automaticamente um novo parceiro. Portanto, as duas médias deveriam ser praticamente iguais. A conclusão da pesquisa foi, portanto, que ou os homens mentem para mais, ou as mulheres mentem para menos, ou ambos mentem quando se trata de quantificar o número de parceiros.

3.4.2. Pesquisa Experimental e Não-Experimental

Uma das caracterizações clássicas das formas de pesquisa consiste na classificação da pesquisa experimental e não-experimental.

Basicamente, a pesquisa não-experimental consiste no estudo de fenômenos sem a intervenção sistemática do pesquisador. Por exemplo, um pesquisador que estude o dia a dia de uma empresa de desenvolvimento de software para detectar ali determinadas práticas previamente catalogadas está fazendo uma pesquisa não-experimental, pois ele age apenas observando e tirando conclusões a partir de um arcabouço teórico preconcebido.

Já a pesquisa experimental implica que o pesquisador sistematicamente provocará alterações no ambiente a ser pesquisado de forma a observar se cada intervenção produz os resultados esperados.

No exemplo anterior, se o pesquisador resolver criar artificialmente situações de estresse dentro da empresa para verificar como os funcionários reagem, então ele estará realizando pesquisa experimental.

3.4.3. Objetividade

Uma característica bastante importante da pesquisa científica é a objetividade. O pensamento humano permite a tirada de conclusões que nem sempre são objetivas, como no caso das opiniões.

Por exemplo, um desenvolvedor pode considerar que a programação orientada a objetos é mais fácil de usar do que a programação estruturada. Isso é apenas uma questão de opinião. Mesmo que o desenvolvedor consiga estabelecer todo um conjunto de argumentos para justificar essa opinião, ela ainda carece de objetividade científica. Por quê? Porque a facilidade de uso não é uma conclusão à qual chegaria qualquer observador independente. Um programador experimentado em orientação a objetos pode considerar muito fácil trabalhar com objetos e mensagens, mas um iniciante ou ainda um programador experimentado em Cobol poderia ter dificuldade com tais conceitos. Esse último ainda poderia sugerir que Cobol é que é muito mais fácil de usar.

Segundo Kerlinger (1980), *“a condição principal para satisfazer o critério de objetividade é, idealmente, que quaisquer observadores com um mínimo de competência concordem com seus resultados”*.

Os experimentos e as observações no trabalho científico devem, então, ser tratados de forma objetiva. Qualquer observação ou grandeza que se queira avaliar deve ser definida de forma a que leituras possam ser feitas independentemente do observador que as toma. Por exemplo, o pesquisador poderia dizer que um sistema é “fácil de usar” se um determinado conjunto de tarefas predefinido puder ser executado por um usuário com um determinado grau de treinamento dentro de um período de tempo predeterminado. A objetividade do conceito passa a ser então a pertinência a um conjunto discreto.

Ainda é possível definir um fenômeno objetivamente, mas de forma difusa, isto é, em vez de trabalhar com apenas duas categorias como “fácil” ou “difícil”, trabalha-se com um grau de dificuldade inerente. Novamente, a definição objetiva deve considerar uma medição que possa ser realizada de forma independente por diferentes observadores e ainda assim chegar aos mesmos resultados. Elaborando sobre o exemplo anterior, então, seria possível definir o grau de dificuldade como o tempo médio que usuários com determinado grau de treinamento levam para realizar um conjunto de tarefas pre-

viamente definido usando a ferramenta. Nesse caso, diversas medições chegarão a valores objetivos e comparações poderão ser feitas entre sistemas.

Porém, ter uma definição objetiva de um fenômeno não é suficiente para se produzir um trabalho de qualidade. É necessário observar ainda a qualidade da própria definição. Ou seja, poderá haver pesquisadores que não concordem que o tempo médio para realizar um conjunto de tarefas seja uma boa definição para “facilidade de uso”. Levando o caso a um extremo para melhor compreensão, se um pesquisador definisse que um sistema é fácil de usar apenas se puder ser utilizado via Internet, essa definição dificilmente seria aceita, pois se afasta muito daquilo que em geral se considera, mesmo intuitivamente, como facilidade de uso.

Então, para trabalhar com observações de fenômenos, o pesquisador deve definir de maneira objetiva o fenômeno que vai observar e convencer os demais de que essa definição é razoavelmente intuitiva, mas em especial útil, para chegar a algum resultado.

3.4.4. Empirismo

Há um ditado que já vem se tornando clássico em computação: “Teoria é quando o fenômeno é compreendido, mas não funciona. Prática é quando funciona, mas não se sabe por quê. Na Computação coexistem a teoria e a prática: nada funciona e não se sabe por quê.”

Essa prática tem sido um dos significados adotados para o termo “empírico” por desenvolvedores na área de computação. Não importa por que, basta que funcione. Se funcionar está certo! Esse ponto de vista pragmático, porém, não corresponde ao significado de *empírico* dado pela ciência. Segundo Kerlinger (1980), “*empírico significa guiado pela evidência obtida em pesquisa científica sistemática e controlada*”.

A computação, enquanto ciência, fundamenta suas pesquisas no empirismo e não no princípio da autoridade. Em computação, na maioria das vezes, pouco importa a opinião deste ou daquele expoente, mas as conclusões objetivas obtidas empiricamente.

A falta de empirismo pode levar a conclusões erradas. Na idade média acreditava-se que o homem tinha uma costela a menos do que a mulher, já que a Bíblia relatava que a mulher havia sido criada a partir de uma costela do homem. Essa crença poderia ser refutada sem dificuldades caso alguém usasse de empirismo para simplesmente contar a quantidade de costelas em um e outro sexo.

Já Descartes (2004) acreditava que o coração batia porque era o lugar mais quente do corpo, e seu calor inflava o sangue que entrava nele. Essa conclusão obtida com alguma reflexão mas pouca evidência empírica não serve para mais nada hoje, pelo menos na área da Medicina.

O empirismo é importante para a ciência, portanto, porque é uma maneira sensata de olhar o mundo. Não basta acreditar em sua intuição ou nas palavras dos mestres. É preciso verificar objetivamente se o fenômeno descrito realmente é verdadeiro.

Antes foi mencionado o fato de que o interesse pela Internet vem crescendo muito nos últimos anos. Será verdade? Quem disse? E de onde observou esse fato? É um sentimento comum e intuitivo. Possivelmente vários livros e artigos sobre Internet têm falado isso ao longo dos anos. Mas será essa afirmação realmente verdadeira hoje? Não estará o interesse pela Internet estável ou diminuindo? O leitor provavelmente responderá “claro que não” a essa última pergunta. Mas como pode ter tanta certeza? Fez alguma observação empírica? Tem dados? Afinal o que é *interesse*? Como se mede?

Essas dúvidas são importantes para o cientista que quer realmente entender os fenômenos e descobrir novos conhecimentos. Se os médicos atualmente continuassem acreditando que o coração bate porque aquece o sangue, baseados na afirmação de Descartes, pouquíssimos avanços da Medicina teriam sido conseguidos. Duvidar das conclusões de outros cientistas e duvidar do próprio senso comum, muitas vezes, é a chave para grandes descobertas.

3.4.5. Variáveis

Foi mencionada anteriormente a existência de fenômenos para os quais interessa ao cientista realizar medições objetivas. Em geral, tais fenômenos que podem ser medidos objetivamente são identificados como *variáveis*. Uma variável é um nome que se dá a um fenômeno que pode ser medido e que varia conforme a medição. Se não variasse seria uma *constante* e não teria maior interesse para a pesquisa.

Variáveis em experimentos, assim como as variáveis em programas computacionais possuem um domínio, ou seja, um conjunto de valores dentro do qual a variável se altera. Por exemplo, uma temperatura em geral será um número racional limitado inferiormente pelo zero absoluto (o valor literal dependerá da escala que se use). Uma variável relacionada à temperatura não poderá assumir o valor “Z” ou “falso”, já que esses não pertencem ao seu domínio.

O domínio de uma variável pode ser discreto ou contínuo. Variáveis contínuas assumem valores reais, a ideia de contínuo vem do fato de que entre dois valores sempre existe um terceiro. Já as variáveis discretas assumem seus valores em conjuntos cujos elementos podem ser ordenados ou em conjuntos finitos. O número de participantes de uma videoconferência é um fenômeno representado em uma variável cujo domínio é discreto e corresponde ao conjunto dos números naturais. Os valores desse conjunto podem ser ordenados, e entre o n -ésimo valor e o $(n+1)$ -ésimo valor não existe uma terceira possibilidade para qualquer n que seja um número natural.

Algumas variáveis discretas assumem seus valores em conjuntos finitos. Tais variáveis são denominadas *categóricas*. Por exemplo, as notas que um estudante de mestrado na UFSC pode obter em uma disciplina variam no conjunto {A, B, C, E}. Em ciência, muitas vezes fenômenos observados são classificados com variáveis categóricas, pois estas são em geral mais fáceis de compreender do que valores contínuos. Caso se avaliasse alunos com valores contínuos, seria razoável concluir que um aluno com nota 6,7812 é melhor do que um aluno com nota 6,7811?

Em alguns casos é possível estabelecer regras de transformação de valores contínuos para discretos. Esse processo, chamado *discretização*, consiste em atribuir um valor discreto diferente para variados intervalos de valores contínuos. Por exemplo, considerando-se notas arredondadas para uma casa decimal após a vírgula, pode-se estabelecer uma correspondência entre o domínio contínuo e a variável categórica da seguinte maneira: notas de 0,0 a 4,9 poderiam ser consideradas como o conceito E, notas de 5,0 a 6,9, como conceito C, notas de 7,0 a 8,9, como conceito B, e notas de 9,0 a 10,0, como conceito A. Note-se, porém, que quando se aplica o conceito de arredondamento para uma casa decimal depois da vírgula o domínio da variável já foi transformado de contínuo em discreto, pois, dessa forma, as notas possíveis limitam-se a 101 valores: {0,0, 0,1, 0,2, ..., 9,8, 9,9, 10,0}.

Na pesquisa científica é importante classificar ainda as variáveis como *medidas* ou *manipuladas*. Uma variável *medida* é aquela cujo fenômeno será observado pelo pesquisador. Por exemplo, quantas vezes um usuário de uma ferramenta vai olhar no manual para obter informações para desempenhar a tarefa que lhe foi proposta. Essa variável tem como domínio o conjunto dos números naturais, e seus valores não são determinados pelo observador, mas simplesmente medidos.

Já a variável *manipulada* é aquela que o experimentador vai deliberadamente modificar para realizar seu experimento. Por esse motivo, tal variável também é chamado de variável *experimental*. Um exemplo de variável manipulada poderia ser o número de passos da tarefa repassada aos usuários. O pesquisador poderia passar a alguns usuários, por exemplo, tarefas com cinco passos, a outros com 10 passos, a outros com 15 passos etc. Assim, ao fazer a experimentação, o pesquisador *manipula* a variável referente ao número de passos da tarefa e *observa* o comportamento da variável medida que consiste em contar quantas vezes o usuário vai olhar o manual da ferramenta.

Mas por que pesquisadores manipulam uma ou mais variáveis enquanto observam outras? É porque eles querem encontrar *dependências* entre essas variáveis. No exemplo anterior, possivelmente o pesquisador estaria tentando descobrir se tarefas mais longas implicam ou não o usuário consultar o manual do sistema mais vezes.

A princípio pode-se testar a dependência entre quaisquer variáveis manipuladas e observadas. Mas nem sempre esse teste fará sentido. Antes de analisar uma dependência experimentalmente, o pesquisador em geral desenvolve uma teoria ou hipótese. No caso anterior, a hipótese poderia ser que quanto maior a tarefa, maior a consulta ao manual feita pelo usuário. Outra hipótese possível seria que não importa o tamanho da tarefa, isso não influenciará no número de vezes que o usuário consulta o manual. É uma dependência que tanto em um caso quanto no outro parecem valer a pena testar.

Mas nem sempre dependências que podem ser testadas fazem sentido enquanto hipótese ou teoria. Por exemplo, manipular a variável “número de botões na tela do aplicativo” e medir a variável “número de vezes que o usuário espirra enquanto está usando a ferramenta”, pode até ser divertido, mas dificilmente haverá uma conexão entre os dois fenômenos. Outro exemplo seria observar o dia da semana e tentar verificar se programas rodam mais rapidamente na sexta-feira. O pesquisador até poderia criar experimentos controlados para mensurar isso. Mas dificilmente encontraria uma conexão entre o dia da semana e a velocidade dos programas. Até porque a teoria corrente sobre programas não estabelece nenhuma conexão entre sua velocidade e o dia da semana. Tal hipótese então se revelaria falha.

Dai vem a importância de se trabalhar com uma boa teoria em pesquisa. Não basta realizar experimentos e encontrar relações entre variáveis. É preciso ter uma teoria que procure explicar o porquê dessas relações.

Há aqui também a caracterização, bastante comum em pesquisa, dos termos *variável dependente* e *variável independente*. A variável independente é aquela que, se supõe, influencia outra. A variável dependente é a influenciada.

Uma hipótese de pesquisa científica, então, em geral terá uma associação de implicação, ou seja, antecedente/consequente, entre uma ou mais variáveis independentes e uma ou mais variáveis dependentes. Por exemplo, o número de passos em uma tarefa implica o aumento do número de vezes que o usuário consulta o manual? Essa seria uma hipótese de pesquisa em que a variável dependente é o número de consultas ao manual, e a variável independente é o número de passos da tarefa. Em geral, o pesquisador manipula a variável independente e mede a dependente. Nesse caso, ainda, como ambas as variáveis são numéricas, pode-se afirmar que se busca encontrar uma dependência *direta*, ou seja, quanto maior o número de passos, maior o número de consultas. Em outros casos, se poderia pesquisar dependências inversas, em que o aumento no valor da variável independente causa uma diminuição na variável dependente.

Ainda existe a dependência linear e não-linear. A linear pode ser aproximada por um polinômio de primeiro grau. Por exemplo, poderia ser observado que, para cada passo na tarefa, o usuário faz mais duas consultas ao manual. Nesse caso, a relação entre o número de passos x e o número de consultas y poderia ser expressa pela função $y = 2x$.

Dependências não-lineares usualmente são representadas por polinômios de grau superior a um ou ainda por equações com exponenciais, raízes ou logaritmos. Este livro apresentará apenas os conceitos relacionados à dependência linear. Para o leitor que queira se aprofundar em outras formas de dependência recomenda-se a leitura de um bom livro de estatística, como o de Barbetta, Reis e Bornia (2008).

3.4.6. Variância e Desvio-Padrão

Na pesquisa científica, frequentemente o pesquisador defronta-se com o problema de analisar conjuntos de dados. Por exemplo, ao avaliar um determinado sistema, o pesquisador contabiliza o tempo de interação de cada pessoa dentre um conjunto previamente

definido. Em geral, a *média* é considerada uma medida importante na avaliação de conjuntos de valores. Por exemplo, se quatro pessoas foram analisadas e os tempos medidos em minutos foram 10, 12, 14, 9, então se pode dizer que o tempo médio observado foi de 11,25 minutos.

A média aritmética simples consiste em somar os valores e dividir o resultado pela quantidade de valores. Por isso, a média acima resultou do cálculo $(10+12+14+9)/4 = 11,25$.

Porém, na pesquisa científica há outra medida importante, que é a *variância* do conjunto de valores. Considerem-se os três conjuntos¹ de valores a seguir:

{10, 12, 14, 9}

{1, 20, 2, 22}

{11, 11, 11, 12}

É possível notar certa semelhança entre eles? Aparentemente são conjuntos bem diferentes. Mas todos têm a mesma média: 11,25. A média, então, passa alguma informação sobre a natureza dos conjuntos de valores, mas ela sozinha não é a única informação importante.

Qual a diferença notável entre esses conjuntos? O terceiro apresenta apenas valores muito próximos da média (ou muito próximos entre si). Já o segundo apresenta valores bastante distanciados da média. O primeiro conjunto é um caso intermediário.

Essa observação do distanciamento dos elementos em relação à média é chamada de *variância*. Então, além da média, o pesquisador deve ficar atento também à *variância* do conjunto de valores, já que esta complementa a caracterização do conjunto.

A variância de um conjunto pode ser definida de forma numérica. Basicamente deseja-se uma variância alta quando os valores se afastam muito da média, seja para cima, seja para baixo. Então, a

variância pode ser o cálculo da diferença de cada valor do conjunto em relação à média deste. No caso dos três exemplos anteriores, subtraindo 11,25 de cada valor o resultado seria:

$$\{10-11,25, 12-11,25, 14-11,25, 9-11,25\} = \{-1,25, 0,75, 2,75, -2,25\}$$

$$\{1-11,25, 20-11,25, 2-11,25, 22-11,25\} = \{-10,25, 8,75, -9,25, 10,75\}$$

$$\{11-11,25, 11-11,25, 11-11,25, 12-11,25\} = \{-0,25, -0,25, -0,25, 0,75\}$$

O resultado corresponde à intuição: o terceiro conjunto, de baixa variância, tem valores absolutos pequenos referentes à diferença entre os valores originais e a média. Já o segundo conjunto tem os valores absolutos mais altos, e o primeiro tem valores intermediários.

Agora, para obter um valor escalar para a variância se poderia somar os valores obtidos. Porém, alguns deles são negativos. Não seria intuitivo que um valor negativo pudesse anular a influência de um valor positivo no cálculo da variância. Inclusive, pela própria definição até aqui usada, essas somas invariavelmente resultariam em zero (o leitor pode conferir somando os valores dos três conjuntos anteriores). Então, a simples soma dos valores não é uma solução. É preciso somar apenas valores de mesmo sinal para que se tenha efetivamente um escalar que corresponda à medida da distância dos valores em relação à média do conjunto.

Uma opção seria trabalhar com o valor absoluto de cada diferença. Mas a definição oficial de variância eleva os valores das diferenças ao quadrado como forma de obter apenas valores positivos. Elevar esses valores ao quadrado também permite aumentar a influência dos valores mais afastados da média. Essa escolha tem origem em questões cuja explicação foge ao escopo deste livro.

Então, elevando ao quadrado cada um dos valores dos conjuntos de diferenças, obtêm-se:

$$\{(-1,25)^2, 0,75^2, 2,75^2, (-2,25)^2\} = \{1,5625, 0,5625, 7,5625, 5,0625\}$$

$$\{(-10,25)^2, 8,75^2, (-9,25)^2, 10,75^2\} = \{105,0625, 76,5625, 85,5625, 115,5625\}$$

$$\{(-0,25)^2, (-0,25)^2, (-0,25)^2, 0,75^2\} = \{0,0625, 0,0625, 0,0625, 0,5625\}$$

1 O conjunto, na matemática, não admite repetição de elementos. No entanto, aqui a palavra "conjunto" é usada em sentido mais amplo, significando "coleção" ou "multiconjunto" e, portanto, admitindo repetição de elementos.

Observa-se que os valores do terceiro conjunto são muito baixos (próximos de zero), enquanto os do segundo conjunto são altos e os do primeiro conjunto são intermediários, como se poderia esperar.

O próximo passo é somar os valores obtidos para os elementos de cada conjunto, ficando-se respectivamente com:

$$\sum \{1,5625, 0,5625, 7,5625, 5,0625\} = 14,75$$

$$\sum \{105,0625, 76,5625, 85,5625, 115,5625\} = 382,75$$

$$\sum \{0,0625, 0,0625, 0,0625, 0,5625\} = 0,75$$

Novamente como esperado, o maior valor, 382,75, fica com o segundo conjunto, o menor com o terceiro e o intermediário com o primeiro.

Porém, somando-se simplesmente as diferenças, como foi feito antes, fica-se com um valor de variância que aumenta conforme o tamanho do conjunto. Isso não é intuitivo. A variância de um conjunto seria a dispersão *média* dos elementos em relação à média do conjunto. Então ela não deveria crescer com o tamanho do conjunto. É necessário então, para eliminar esse efeito, dividir o valor escalar encontrado pelo número de elementos do conjunto.

Porém, não é intuitivo dizer que um conjunto com um único elemento tenha variância. No mínimo dois elementos são necessários para que faça sentido dizer que eles se afastam da média. Assim, a divisão não será feita pelo número de elementos no conjunto, mas pelo número de elementos menos um. Assim, no caso de um conjunto com um elemento, a variância será o resultado de uma divisão por zero e, portanto, indefinida.

No exemplo anterior, aplicando-se a divisão por três aos valores de variância obtidos, fica-se com os valores finais de:

$$14,75/3 = 4,9166...$$

$$382,75/3 = 127,5833...$$

$$0,75/3 = 0,25$$

O que se pode concluir aqui é que o segundo conjunto é o que mais varia, ou seja, seus elementos mais se afastam da média, enquanto o terceiro conjunto é o que menos varia, ou seja, ele é o que tem os elementos relativamente mais próximos da média.

A fórmula da variância pode ser então assim definida:

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Onde s_x^2 é o símbolo comumente usado para representar a variância de X , n é o número de elementos do conjunto, \bar{x} representa a média aritmética do conjunto e x_i representa cada um dos elementos do conjunto no somatório.

O *desvio-padrão* é uma medida também bastante utilizada para analisar conjuntos e é definido simplesmente como a raiz quadrada da variância, ou seja:

$$s_x = \sqrt{s_x^2}$$

onde s_x é o símbolo comumente usado para representar o desvio-padrão.

No exemplo anterior, os valores de desvio-padrão dos três conjuntos são respectiva e aproximadamente:

$$\sqrt[3]{4,9166...} = 2,217355783$$

$$\sqrt[3]{127,5833...} = 11,29527925$$

$$\sqrt[3]{0,250,75 / 3} = 0,5$$

3.4.7. Covariância

A variância é uma medida muito importante para a pesquisa científica. Basicamente pode-se dizer que não haveria muito o que fazer se os fenômenos não variassem. Se cada variável estudada tivesse variância zero, então os valores seriam previsíveis e pouco se poderia aprender sobre a natureza das coisas.

Entretanto, os fenômenos variam, e quanto maior a variância, mais interessante pode ser o fenômeno. A pesquisa em engenharia de software, por exemplo, tenta descobrir formas de estimar quanto tempo um programa levaria para ser desenvolvido. É uma medida difícil porque mesmo que se tenha uma descrição detalhada de cada função ou caso de uso a ser desenvolvido, o tempo que o desenvolvedor levaria para programar cada um desses elementos poderia variar de alguns minutos a várias semanas.

Saber o tempo esperado (ou o tempo médio) ao criar os programas necessários para realizar um caso de uso² pode até ser útil ao fazer previsões para conjuntos de casos de uso, mas médias aritméticas de um conjunto de valores só costumam ser boas estimativas quando uma quantidade significativa de valores está em jogo. Saber o tempo médio que se gasta para programar um caso de uso não permite prever quanto tempo vai-se levar para programar um dado caso de uso, tomado individualmente. Similarmente, mesmo sabendo que em 1.000 jogadas de uma moeda, aproximadamente 500 serão cara e aproximadamente 500 serão coroa, não há meios de saber qual resultado será obtido em uma jogada específica.

No caso de conjuntos que variam muito, como do tempo que se leva para programar casos de uso, será que não se trataria de considerar que não se tem um único conjunto mas sim um certo número de subconjuntos, cada um dos quais com características distintas? Cada um com sua própria média e variância?

Voltando ao conjunto de maior variância do exemplo anterior, {1, 20, 2, 22}. Observando-se esse conjunto, não seria possível concluir que talvez se trate de dois subconjuntos distintos? Ou seja, {1, 2} e {20, 22}. Nesse caso, o que se tem são dois subconjuntos com médias distintas, e cada um dos quais com variância bem menor do que a do conjunto original.

Seria necessário, então, construir uma teoria para determinar quais as causas de certos elementos apresentarem medidas associa-

das ao primeiro subconjunto ou ao segundo subconjunto, ou seja, qual a causa da existência desses dois subconjuntos.

Voltando ao exemplo dos casos de uso. Não seria mais útil imaginar que seria possível classificar esses casos de uso em, por exemplo, simples, médios e complexos, para obter conjuntos de medidas com variância menor e, portanto, mais previsíveis? E se, por exemplo, os casos de uso simples pudessem ser programados em menos de um dia, os médios entre um dia e uma semana e os complexos em mais de uma semana? Seria mais fácil fazer previsões sobre o tempo que se levaria para desenvolver um sistema. Bastaria contar a quantidade de casos de uso simples, médios e complexos e multiplicar cada quantidade pela média de tempo associada a cada subconjunto. A informação seria mais relevante, possivelmente, do que uma única média aplicada ao conjunto inteiro dos casos de uso, pois com essa abordagem seriam usadas características dos elementos do conjunto para reduzir a incerteza sobre eles.

Porém, essa subclassificação introduz outro fator de incerteza: não se sabe, *a priori*, se a forma de determinar que um caso de uso é simples, médio ou complexo *realmente* classifica os casos de uso em subconjuntos nos quais os valores de tempo de desenvolvimento tenham variância mais baixa.

Nesse ponto poderia ser aplicada a experimentação científica para validar a hipótese de que uma determinada técnica de classificação dos casos de uso efetivamente classifica a complexidade destes adequadamente, ou seja, que esse sistema de classificação colocará no conjunto “simples” os casos de uso que efetivamente sejam mais rápidos de programar, e no conjunto “complexos” os casos de uso mais difíceis, ficando os demais no conjunto “médios”.

Para testar essa hipótese é necessário comparar dois conjuntos de valores: o valor dado a um caso de uso pelo método de classificação e o valor do tempo que efetivamente se leva para programar o caso de uso. Para efetuar essa comparação é necessário usar o conceito de *covariância*, ou seja, determinar em que grau os dois conjuntos variam conjuntamente.

2 O conceito de caso de uso é descrito de forma compreensível por Wazlawick (2004).

Espera-se, então, que ao se ter as medidas dos tempos, os maiores tempos estejam no conjunto “complexos”, os menores tempos no conjunto “simples” e os demais tempos no conjunto “médios”. Para dar um tratamento totalmente numérico aos conjuntos é possível denotar os conjuntos simples, médios e complexos por números. Alguns trabalhos de referência na área de pontos de caso de uso sugere aplicar valores numéricos 1, 2 e 3 para casos de uso simples, médios e complexos, respectivamente. Nesse caso, o que se busca é uma covariância entre esses valores de pontos de casos de uso e os valores das medidas de tempo obtidas. Pode ser questionado se a escala é correta. Por exemplo, não seria mais adequado ter valores como 1, 2 e 5 em vez de 1, 2 e 3? É uma dúvida válida, mas como a literatura propõe os valores 1, 2 e 3 e não é objetivo deste exemplo questionar esse aspecto da teoria, então trabalha-se com esses valores mesmo e não com outros (se esse exemplo fosse um projeto de pesquisa, o fato de que não será considerada a hipótese de a proporção ser diferente poderia ser uma *limitação* do trabalho).

Supondo que o pesquisador queira saber se um determinado método de estimativa realmente classifica os casos de uso adequadamente, ele poderia fazer experimentos da seguinte forma: (1) tomar um conjunto de casos de uso aleatoriamente escolhidos cujo tempo de programação já seja conhecido (neste caso, trata-se de um *benchmark*; se o tempo não for conhecido, o pesquisador terá de efetuar ele mesmo os cálculos, possivelmente solicitando a um ou mais programadores que desenvolvam os casos de uso para verificar quanto tempo levam); (2) aplicar o método para classificar cada caso de uso com os valores 1, 2 ou 3 e verificar se existe covariância entre os dois conjuntos de valores obtidos. Supondo que os valores obtidos pelo método de classificação para um conjunto de 10 casos de uso sejam os que estão apresentados na Tabela 3.1, pode-se verificar intuitivamente que para os valores mais altos de tempo tem-se a classificação com 3 pontos e os valores mais baixos

de tempo estão classificados com 1 ponto, portanto, intuitivamente parece haver covariância.

Tabela 3.1: Exemplo de covariância alta e direta

Caso de Uso	Tempo conhecido (horas)	Pontos de caso de uso
UC1	1	1
UC2	18	2
UC3	4	1
UC4	67	3
UC5	22	2
UC6	12	2
UC7	2	1
UC8	7	1
UC9	18	2
UC10	55	3

A covariância é considerada alta porque efetivamente os menores valores da coluna “tempo” coincidem com os menores valores da coluna “pontos de caso de uso”, e os maiores valores da coluna “tempo” coincidem com os maiores valores da coluna “pontos de caso de uso”. A variância também é considerada direta ou positiva porque quanto maior o valor da coluna “tempo”, maior o valor da coluna “pontos de caso de uso”.

Em outra situação, poderia ser observada uma relação inversa, ou seja, quanto maior o valor numa coluna, menor na outra. Se fosse assim, então a covariância seria negativa.

Também seria possível que não houvesse nenhuma ou pouca relação entre os tempos conhecidos e o resultado do método de classificação. Um método de classificação totalmente arbitrário não poderia, a princípio, gerar alta covariância, como, por exemplo, um método que atribua aos casos de uso valores sequenciais 1, 2 e 3 em função da ordem em que estes aparecem. A Tabela 3.2 apresenta uma comparação entre os tempos conhecidos e os resultados desse método arbitrário.

Tabela 3.2: Exemplo de baixa covariância

Caso de uso	Tempo conhecido (horas)	Pontos de caso de uso
UC1	1	1
UC2	18	2
UC3	4	3
UC4	67	1
UC5	22	2
UC6	12	3
UC7	2	1
UC8	7	2
UC9	18	3
UC10	55	1

Intuitivamente é possível perceber que não existe uma relação bem definida entre os valores altos da coluna “tempo” e da coluna “pontos” nem entre os valores baixos da coluna “tempo” e os da coluna “pontos”.

A questão agora consiste em como tratar numericamente a covariância para que se possa, por exemplo, comparar diferentes métodos de classificação de casos de uso, decidindo quais são os melhores.

Numericamente, então, a covariância pode ser calculada a partir dos desvios da média em cada conjunto. Se no primeiro conjunto um valor se desvia muito para cima da sua média, espera-se que no segundo conjunto o valor correspondente também se desvie muito para cima da sua própria média. Por outro lado, se ocorre o inverso, ou seja, se o desvio no segundo conjunto ocorre para baixo da média, então há uma relação negativa.

No final, se todos os desvios para cima ou para baixo são semelhantes nos dois conjuntos, haverá covariância alta e positiva. Se os desvios forem sempre invertidos, ou seja, para cima em um conjunto e para baixo no segundo e vice-versa, então a covariância também será alta, mas negativa. Finalmente, se em alguns casos os desvios

coincidem e em outros não, têm-se valores positivos e negativos, levando a uma covariância próxima de zero e, portanto, baixa.

Assim, para calcular a covariância costuma-se usar o produto das diferenças de cada elemento em relação à média do conjunto ao qual ele pertence. Seja $X = \{x_1, x_2, \dots, x_n\}$ o primeiro conjunto e $Y = \{y_1, y_2, \dots, y_n\}$ o segundo conjunto, ambos de mesmo tamanho, pois só é possível calcular covariância se para cada elemento de X corresponde um elemento de Y , então a covariância é calculada em função dos produtos: $(x_i - \bar{x})(y_i - \bar{y})$.

Dessa forma, cada vez que os elementos correspondentes nos dois conjuntos se desviam conjuntamente para cima da média ou para baixo da média, o resultado será positivo. Se um se desvia para cima e outro para baixo, o resultado será negativo. Quanto mais os dois elementos se desviam da média, maior o valor absoluto do produto.

A covariância pode então ser calculada simplesmente como o somatório desses produtos, dividido pelo número de elementos do conjunto de valores menos um, ou seja:

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Onde s_{xy} é o símbolo usual para representar a covariância entre os conjuntos X e Y .

Agora, conhecendo a definição matemática da covariância pode-se aplicar esse conceito para identificar entre os dois métodos qual é mais adequado para fazer estimativas de tempo de desenvolvimento de casos de uso. A Tabela 3.3 apresenta o cálculo da covariância para o primeiro método. A média dos tempos conhecidos, \bar{x} , é 20,6 e a média dos valores de classificação, \bar{y} , é 1,8.

Tabela 3.3: Covariância para o primeiro método

Caso de uso	Tempo conhecido (horas): X	Pontos de caso de uso: Y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
UC1	1	1	-19,6	-0,8	15,68
UC2	18	2	-2,6	0,2	-0,52
UC3	4	1	-16,6	-0,8	13,28
UC4	67	3	46,4	1,2	55,68
UC5	22	2	1,4	0,2	0,28
UC6	12	2	-8,6	0,2	-1,72
UC7	2	1	-18,6	-0,8	14,88
UC8	7	1	-13,6	-0,8	10,88
UC9	18	2	-2,6	0,2	-0,52
UC10	55	3	34,4	1,2	41,28

Então, o valor da covariância entre os dois conjuntos consiste na somatória da última coluna da Tabela 3.3 dividido por $n-1$, ou seja: $149,2/9 = 16,57777...$. Mas o que significa esse número? A partir dele é possível concluir que os valores de um conjunto são afetados pelos valores do outro conjunto? A covariância não produz um número normalizado, como será explicado em seguida. Então, o valor da covariância serve basicamente para comparar um par de conjuntos com outro par de conjuntos, desde que as unidades de medida dos dois pares sejam idênticas.

Seguindo esse raciocínio, é de se esperar que o método arbitrário, apresentado na Tabela 3.2, designado agora como Y' apresente um valor de covariância bem mais baixo do que o primeiro método, designado na Tabela 3.3 como Y . A Tabela 3.4 apresenta o cálculo da covariância para o método arbitrário, levando em conta que a média dos valores obtidos pelo método arbitrário é 1,9.

Tabela 3.4: Covariância para o método arbitrário

Caso de uso	Tempo conhecido (horas): X	Pontos de caso de uso: Y'	$(x_i - \bar{x})$	$(y'_i - \bar{y}')$	$(x_i - \bar{x})(y'_i - \bar{y}')$
UC1	1	1	-19,6	-0,9	17,64
UC2	18	2	-2,6	0,1	-0,26
UC3	4	3	-16,6	1,1	-18,26
UC4	67	1	46,4	-0,9	-41,76
UC5	22	2	1,4	0,1	0,14
UC6	12	3	-8,6	1,1	-9,46
UC7	2	1	-18,6	-0,9	16,74
UC8	7	2	-13,6	0,1	-1,36
UC9	18	3	-2,6	1,1	-2,86
UC10	55	1	34,4	-0,9	-30,96

Assim, a covariância entre esses dois conjuntos é a somatória da última coluna da Tabela 3.4 dividido por 9, ou seja, $-70,4/9 = -7,822222...$. O sinal negativo indica que se existe eventualmente alguma covariância, esta é negativa, mas o valor absoluto obtido para o método arbitrário é menos do que a metade do valor obtido para o primeiro método. Ou seja, a covariância do primeiro método em relação aos valores de tempo é maior do que a covariância do método arbitrário.

3.4.8. Correlação

Como mencionado anteriormente, o valor absoluto da covariância não diz muita coisa sobre um conjunto estar influenciando o outro ou não. Por esse motivo usa-se mais frequentemente a *correlação*. Esta tem basicamente o mesmo significado intuitivo da covariância, mas o resultado numérico varia entre -1 e 1 , em que perto de -1 significa correlação negativa muito forte, perto de 1 significa correlação positiva muito forte e perto de 0 significa ausência de correlação.

Existem vários tipos de cálculo de correlação possíveis. O mais conhecido é o coeficiente de correlação de Pearson, obtido dividindo-se o resultado de $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, que (dividido por $n-1$) é usado para obter a covariância, pelo produto do desvio-padrão de cada um dos dois conjuntos de valores, ou seja:

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_X s_Y}$$

Retornando ao exemplo anterior, pode-se calcular a correlação do primeiro método da seguinte forma:

- Já se tem o valor de $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, que é 149,2.
- Calcula-se o desvio-padrão do conjunto dos tempos como a raiz quadrada da variância desse conjunto: $s_X = 67,94409467$.
- Calcula-se o desvio-padrão do conjunto dos resultados do primeiro método como a raiz quadrada da variância desse conjunto: $s_Y = 2,366431913$.
- Calcula-se o coeficiente de correlação de acordo com a fórmula anterior: $r_{XY} = 149,2 / (67,94409467 * 2,3661913) = 0,928041193$.

Para o método arbitrário tem-se:

- O valor de $\sum_{i=1}^n (x_i - \bar{x})(y'_i - \bar{y}')$ é -70,4.
- Desvio-padrão do conjunto dos tempos é o mesmo: $s_X = 67,94409467$.
- Desvio-padrão dos resultados do método arbitrário: $s_Y = 2,626785107$.
- Coeficiente de correlação é, portanto, $r_{XY} = -70,4 / (67,94409467 * 2,626785107) = -0,39445403$.

Pode-se tirar duas conclusões desses valores. Em primeiro lugar, o primeiro método tem um índice de correlação com os tempos

conhecidos bastante próximo a 1 e, portanto, parece ser um método de boa qualidade. Já o método arbitrário tem uma correlação negativa e mais próxima de 0 do que de 1, não podendo ser considerado, portanto, como um bom método para estimativa de tempo.

A intuição, porém, diz que a correlação do método arbitrário deveria ser ainda mais próxima de zero. O método arbitrário atribui valores sequenciais que nenhuma relação tem com os tempos dos casos de uso. Por que não é assim? É que o conjunto de valores examinado é pequeno. Pode-se observar que em alguns casos há até coincidência entre o primeiro método e o método arbitrário. Essas coincidências é que podem levar a crer que o método funcione. Levando a um extremo, se em vez de usar os 10 valores que foram usados para cada método se tivesse trabalhado apenas com os dois primeiros valores, os dois métodos dariam exatamente o mesmo resultado e seriam considerados equivalentes. Mas isso é apenas uma coincidência devido ao pequeno número de situações analisadas.

Um problema a ser considerado, portanto, quando se faz esse tipo de experimento para encontrar correlação entre dois conjuntos de valores é saber se o tamanho da amostra é suficiente para considerar a correlação como efetivamente significativa ou se o valor foi obtido apenas por coincidência. Felizmente existe um método para testar isso. Como a correlação é um índice que deve convergir para um valor à medida que se aumenta o tamanho do conjunto considerado, então não existe certeza absoluta sobre o índice, a não ser que o conjunto seja infinito. Porém, conjuntos infinitos são impossíveis de trabalhar empiricamente. Trabalhando então sempre com conjuntos finitos, é necessário estabelecer um índice de confiabilidade aceitável. Em geral, em estatística, trabalha-se com o índice de 95% de certeza, considerado suficientemente alto para a maioria das aplicações.

O teste para verificar então se um coeficiente de correlação é ou não significativo vai considerar o índice de certeza. Se esse índice for estabelecido em 95%, então, dependendo do tamanho do conjunto examinado, existirá um valor limite para que um coeficiente de cor-

relação seja considerado significativo. O cálculo desses valores foge ao escopo deste livro, mas a Tabela 3.5 apresenta alguns valores que podem ser usados como referência.

Tabela 3.5: Valores mínimos de correlação para ser considerada significativa com 95% de certeza

n	r_{xy} mínimo	n	r_{xy} mínimo	n	r_{xy} mínimo
3	.99692	13	.5529	27	.3809
4	.95000	14	.5324	32	.3494
5	.8783	15	.5139	37	.3246
6	.8114	16	.4973	42	.3044
7	.7545	17	.4821	47	.2875
8	.7067	18	.4683	52	.2732
9	.6664	19	.4555	62	.2500
10	.6319	20	.4438	72	.2319
11	.6021	21	.4329	82	.2172
12	.5760	22	.4227	92	.2050

Considerando agora que os experimentos anteriores foram realizados com 10 casos de uso, o valor mínimo de r_{xy} para que a correlação seja considerada significativa é, pela Tabela 3.5, de 0,6319 no caso de correlação positiva e máximo de -0,6319 no caso de correlação negativa. Então, o primeiro método, com correlação 0,928041193 já pode ser considerado suficientemente analisado para se concluir que efetivamente existe uma forte relação entre o número de pontos de caso de uso que ele produz e o tempo esperado para programar esses casos de uso.

Já o método arbitrário, com correlação de -0,39445403, não atingiu um valor de correlação suficiente para que possa ser considerado que possui correlação negativa com os tempos.

Portanto, o resultado é conclusivo para o primeiro método e inconclusivo para o método arbitrário. Foi demonstrado que o primeiro método atende à expectativa, mas não foi demonstrado nada a respeito do segundo método.

Caso o pesquisador estivesse tentado demonstrar que o método arbitrário funciona, ele teria de aumentar o tamanho do conjunto de valores analisados e calcular um novo índice de correlação, aplicando então outros valores mínimos de acordo com a Tabela 3.5. Porém, o bom-senso diz que, considerando-se a natureza totalmente independente entre os valores dos tempos e os resultados do método arbitrário, por maior que seja o conjunto de valores nunca se chegará a provar qualquer tipo de correlação entre os dois conjuntos. No limite haverá tantos elementos no conjunto que o índice de correlação será praticamente igual a zero.

3.4.9. A Hipótese de Pesquisa

Um aspecto que diferencia o trabalho científico do trabalho técnico é a existência de uma hipótese de pesquisa. A *hipótese* é uma afirmação da qual não se sabe a princípio se é verdadeira ou falsa. O trabalho de pesquisa consiste justamente em tentar provar a veracidade ou falsidade da hipótese.

Um objetivo sem uma boa hipótese pode ser muito arriscado. Anteriormente foi dito que o objetivo consiste em tentar produzir algum conhecimento que ainda não existe. Mas se não houver uma boa hipótese para justificar esse objetivo, corre-se o risco de realizar a pesquisa sem obter resultados. Por exemplo, ter como objetivo de pesquisa provar que $P=NP$ é perfeitamente válido, pois esse problema é relevante à sociedade e o conhecimento necessário para resolver o problema ainda não existe. Porém, com que hipótese o pesquisador vai trabalhar? Se o problema de pesquisa for colocado simplesmente como “provar que $P=NP$ ”, o pesquisador poderá ficar Tateando a esmo, e o risco de fracasso será muito grande.

É necessário, portanto, ter uma hipótese.

Segundo Comer (2008), uma tese é uma hipótese ou conjectura. O texto da tese ou monografia é um documento em que o aluno apresenta argumentos a favor de sua tese. Daí a confusão que muitas vezes se faz com o termo “tese”, que pode representar tanto o documento escrito, quanto a hipótese de pesquisa.

O método, como discutido anteriormente, deverá indicar como os testes deverão ser feitos. Ao final dos experimentos haverá evidências a favor ou contra a hipótese inicial. Nesse caso, alguém poderá perguntar, “e se não se conseguir provar que a hipótese era válida?”. A resposta para essa pergunta dependerá de quão relevante era a hipótese original. Uma hipótese qualquer escolhida a esmo, sem nenhum tipo de justificativa, se não for confirmada, não traz nenhuma informação nova para a área de pesquisa. Mas uma hipótese sólida e bem justificada, com evidências de validade, que ao final é invalidada, pode produzir a informação interessante. No pior dos casos provará que aquilo que eventualmente se poderia aceitar intuitivamente como verdadeiro não resistiu à prova. É dessa forma que muitos mitos podem ser derrubados.

Sendo assim, além do objetivo, hipótese e metodologia, é fundamental que o trabalho de pesquisa tenha como base uma boa justificativa para a escolha da hipótese. Uma hipótese bem justificada no início do trabalho aumenta as chances de sucesso. Em primeiro lugar, é mais provável que ela seja verdadeira do que uma hipótese sem justificativa. Em segundo lugar, se ela for falsa, o trabalho terá o mérito de ter derrubado algum mito.

O trabalho científico na área de Computação consiste então em formular uma hipótese e coletar evidências para comprovar a sua validade. Essas evidências podem ser obtidas basicamente de três formas:

- Construindo uma teoria, que a partir de fatos aceitos e deduções válidas prove que a hipótese é verdadeira.
- Realizando certo número de experimentos controlados, que estatisticamente comprovem a validade da hipótese. Porém, deve-se ter em mente que esse tipo de comprovação é sempre sujeito a erros. Normalmente se aceita que hipóteses sejam comprovadas com 95% de certeza.
- Realizar estudos de caso, comparativos, argumentações, colher opiniões através de questionários e outras formas que dificilmente constituem uma prova, mas que podem ser evidências da validade da hipótese.

É possível também realizar um trabalho estruturado sobre combinações de duas ou três formas anteriores. De qualquer maneira, o importante é que a partir da formulação da hipótese, o pesquisador esteja engajado no processo de busca de evidências, e que essas evidências sejam estruturadas em um discurso coerente, em que ele apresenta a sua argumentação para a comunidade científica.

Um problema de pesquisa, então, em geral vai perguntar como duas ou mais variáveis se relacionam, e se existe correlação positiva ou negativa entre os valores das variáveis. A existência dessas correlações, porém, ainda não prova causas. Uma teoria consistente que explique causa e efeito precisa também ser elaborada, além da validação empírica. Isso acontece porque algumas vezes duas variáveis até se correlacionam com alto índice, mas as causas envolvidas podem não ser tão diretas.

Cita-se como exemplo o caso de uma empresa que decidiu verificar se funcionários bem alimentados trabalhavam melhor. A empresa passou a servir um café da manhã saudável para seus funcionários em todas as filiais e, em praticamente todas, a produtividade aumentou. Mas será realmente verdade? Qual a explicação? Será que foi mesmo a alimentação que fez os funcionários trabalharem mais? Na sequência a empresa testou retirar o café da manhã saudável para verificar se os trabalhadores retornavam ao ritmo anterior. Para surpresa de todos, a produtividade cresceu ainda mais.

Como um cientista lidaria com esses dados aparentemente contraditórios? O caso aqui é que se trata de verificar o real motivo do aumento de produtividade. Este não ocorreu por conta de uma alimentação melhor, mas pelo fato de que houve uma significativa modificação no dia a dia da empresa. Funcionários estão acostumados a perceber que modificações na empresa em geral implicam demissões. Por isso, cada vez que se observa uma modificação no ambiente (introdução ou retirada do café da manhã), os funcionários tendem a trabalhar mais para serem notados e garantirem seus empregos.

3.5. Justificativa

Foi comentado anteriormente que uma hipótese de trabalho é muito arriscada se não estiver solidamente apoiada em uma boa justificativa que apresente evidências de que vale a pena investir tempo e recursos na tentativa de comprovar a hipótese. Quem em sua consciência se proporia a trabalhar dois anos para provar que o método arbitrário definido anteriormente faz boas previsões em estimativa de esforço? Uma boa hipótese precisa ser *justificável*.

Em uma monografia, pode-se justificar o tema de pesquisa, mas mais importante ainda é justificar a escolha do objetivo e da hipótese. Por exemplo, se o tema de pesquisa é “compactação de texto”, o objetivo de pesquisa é obter um algoritmo com maior grau de compactação do que os algoritmos comerciais, e a hipótese de pesquisa pode consistir em utilizar um determinado modelo de rede neural para realizar essa compactação, então a justificativa do tema deverá se concentrar em mostrar que é necessário obter algoritmos de compactação melhores. Adicionalmente, a justificativa da hipótese deverá se concentrar em apresentar evidências de que o modelo de rede neural escolhido poderá produzir resultados melhores do que os algoritmos comerciais.

Em geral, a justificativa do tema aparece na contextualização do trabalho, em que se tenta justificar ao leitor que o problema escolhido realmente é relevante (no exemplo anterior, compactação de textos). Mas na maior parte das vezes esse convencimento é pacífico. Mais difícil é justificar uma hipótese de trabalho, pois para isso será necessário apresentar alguma evidência de que uma determinada linha de pesquisa pode levar a bons resultados quando ainda não se efetuou essa pesquisa (no exemplo anterior, justificar o uso do modelo específico de redes neurais para compactar textos). Essas evidências podem ser referências a outros trabalhos que eventualmente mostraram algum tipo de resultado que aponte para a viabilidade da hipótese escolhida, ou ainda em dados colhidos preliminarmente pelo próprio autor do trabalho ou em um estudo de caso.

3.6. Resultados Esperados

Em geral, os resultados esperados são situações que o autor de um trabalho espera que ocorram, caso seus objetivos sejam atingidos. Os resultados esperados normalmente fogem ao escopo do trabalho. O autor da pesquisa não tentará obter os resultados esperados ao final da pesquisa. Eles são posteriores.

Isso diferencia os resultados esperados dos objetivos. Os objetivos serão perseguidos pelo autor, e ao final do trabalho ele dirá se foram ou não atingidos. Os resultados esperados possivelmente ocorrerão após a conclusão do trabalho.

Por exemplo, o objetivo do trabalho poderá ser definir um método de cálculo de esforço para desenvolvimento de software mais preciso do que os métodos do estado da arte. O autor da pesquisa deverá ter uma boa hipótese para fundamentar esse objetivo em primeiro lugar. Depois, ele deverá realizar um conjunto de experimentos que, juntamente com uma base teórica, demonstrarão a validade ou não da hipótese.

Esse autor poderá apresentar, inicialmente, como resultados esperados de seu trabalho, a adoção do seu método pela indústria e um melhor desempenho das empresas produtoras de software que venham a utilizar esse método.

Como se vê aqui, é praticamente impossível que o autor obtenha esses resultados esperados durante a realização de sua pesquisa. Mas eles poderão eventualmente ocorrer depois. Também é possível que não ocorram, pois, por outros motivos quaisquer, poderá acontecer que nenhuma empresa venha a adotar o seu método.

Assim, pode-se dizer que os objetivos devem ser verificáveis ao final do trabalho, inclusive os objetivos específicos. Já os resultados esperados são apenas esperanças e não podem necessariamente ser verificados ao final do trabalho.

No início do trabalho de pesquisa, uma forma de se tentar determinar quais são os resultados esperados do trabalho consiste em fazer a pergunta “o que possivelmente mudaria no mundo se eu atingisse os objetivos da minha tese/monografia?”.

3.7. Limitações do Trabalho

Ao contrário do que alunos iniciantes muitas vezes pensam, não é possível resolver todos os problemas da humanidade em dois ou três anos de trabalho (“síndrome de querer mudar o mundo”, ou “síndrome do Prêmio Nobel”).

Um trabalho de pesquisa pode começar muitas vezes com um objetivo demasiadamente amplo e, portanto, inalcançável durante o tempo disponível para a realização do curso. Sendo assim, muitas vezes é necessário realizar cortes nos objetivos, ou limitar a forma de persegui-los. Em vez de demonstrar que uma hipótese é sempre verdadeira, pode-se optar por demonstrar que ela é verdadeira apenas em determinadas condições, para as quais foi possível realizar testes convincentes. Por exemplo, o método de estimativa de esforço mencionado na seção anterior poderia ser comprovadamente mais preciso apenas para uma determinada classe de sistemas, como, por exemplo, sistemas baseados em Web. O fato de que o método não foi testado com outros tipos de sistemas impõe uma limitação ao trabalho.

As limitações são, portanto, aspectos do trabalho dos quais o autor tem consciência e reconhece a importância, mas não tem condições de abordar no tempo disponível.

É importante, em trabalhos de pesquisa, que as limitações conhecidas sejam claramente identificadas pelo autor desde o início. Isso evitará que o próprio autor muitas vezes se perca em divagações ou buscando aspectos que extrapolam os objetivos iniciais. Isso evita também que o leitor crie expectativas demasiadamente amplas sobre o trabalho, que serão depois frustradas.

Novamente, espera-se que uma boa interação com o orientador ajude o aluno a colocar as devidas limitações nos seus objetivos, para que o trabalho possa ser concluído com sucesso no tempo disponível.

3.8. Discussão

De acordo com o que foi visto neste capítulo, o trabalho de pesquisa deverá estar enquadrado em um tema que, como área de conhecimento, deverá ser plenamente conhecido pelo pesquisador. Dentro do tema, o pesquisador deverá estabelecer um objetivo a ser buscado. Esse objetivo deverá estar baseado em uma hipótese de trabalho, que deve ter uma boa justificativa para ter sido escolhida. O método vai esclarecer como a hipótese será comprovada pelo autor do trabalho, e as limitações deixarão claros quais aspectos não serão abordados.

É compreensível a dificuldade de muitos alunos que ingressam, especialmente no mestrado, em compreender essa estrutura e realizar um trabalho organizado dessa forma. Surgem assim dissertações que muitas vezes são meramente uma apresentação de um sistema, ou uma proposta testada em apenas uma ou duas situações, ou, ainda, dissertações que se concentram em coletar dados e não elaboram adequadamente os conceitos que os dados representam.

Essa dificuldade deve-se, especialmente, ao fato de que pela primeira vez, talvez, em sua vida o aluno será colocado diante de um trabalho individual extenso, em que a sua iniciativa será fundamental para o sucesso. Trabalhos escolares, mesmo na graduação e na especialização, resumem-se, muitas vezes, apenas à pesquisa bibliográfica. O aluno simplesmente coleta material de várias fontes e organiza essa informação de uma maneira pessoal. A estrutura da pesquisa científica, especialmente no mestrado e no doutorado, vai muito além da pesquisa bibliográfica, como se procurou mostrar neste capítulo.