Metodologia de Pesquisa para Ciência da Computação

fonte: livro Prof. Raul Sidnei Wazlawick, 2009

Profa. M. Cristina
ICMC-USP

SCC5921 – Metodologia de Pesquisa em Visualização e Imagens

Pesquisa

- Uma pesquisa é um processo de construção do conhecimento que tem como metas principais gerar novos conhecimentos e/ou consolidar, expandir ou refutar algum conhecimento pré-existente.
- A pesquisa é basicamente um processo de aprendizagem tanto do indivíduo que a realiza, quanto da sociedade na qual esta se desenvolve.
- A pesquisa como atividade regular também pode ser definida como o conjunto de atividades orientadas e planejados pela busca de um conhecimento.

https://www.dicionarioinformal.com.br/significado/pesquisa/3055/

Método Científico

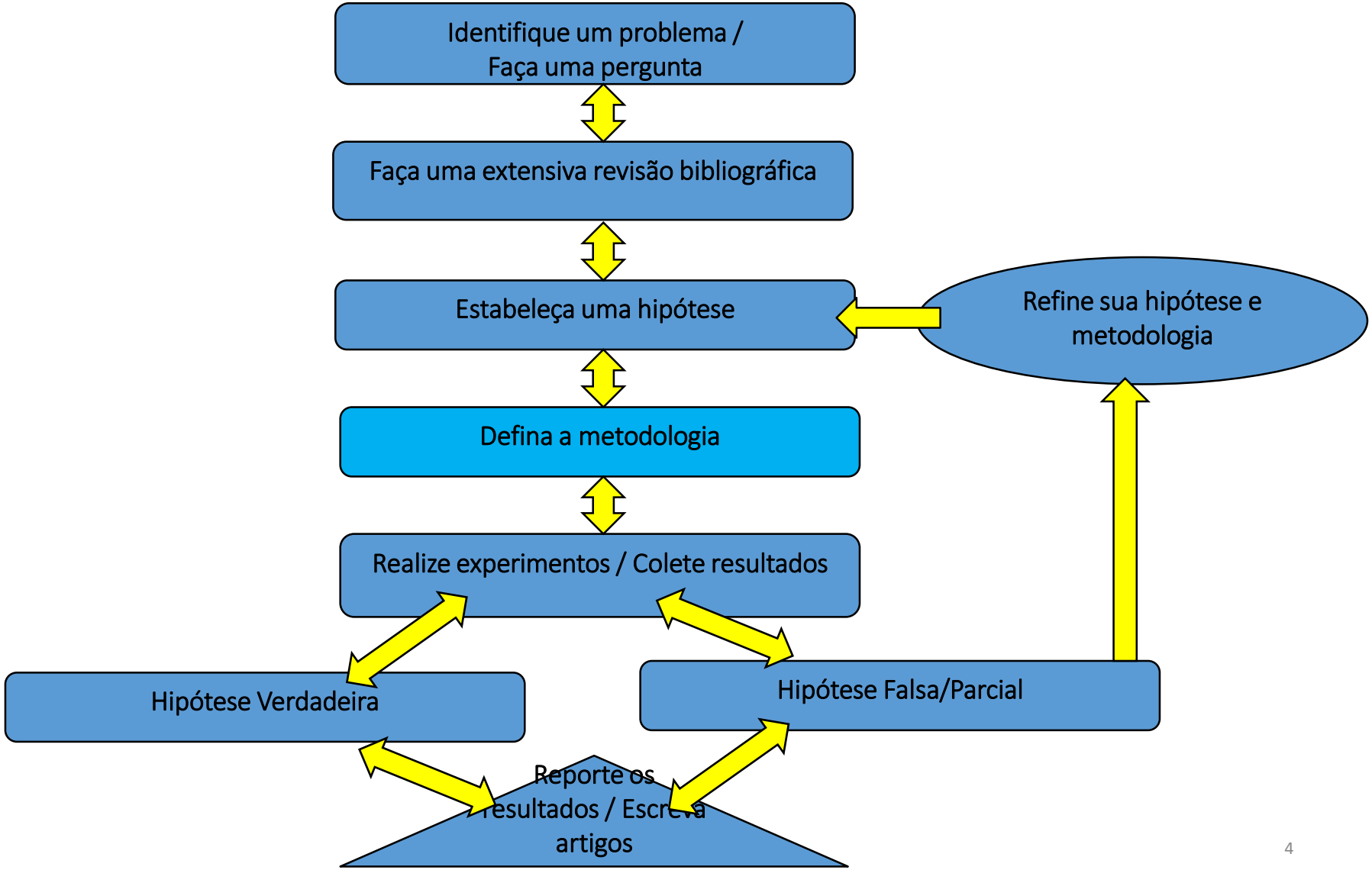
"O **método científico** é um conjunto de **regras básicas** para um cientista desenvolver um estudo experimental a fim de produzir novo conhecimento, bem como corrigir e integrar conhecimentos preexistentes. Na maioria das disciplinas científicas consiste em juntar evidências observáveis, empíricas (i.e., baseadas apenas no observado no estudo experimental) e mensuráveis, e analisá-las com o uso da lógica. "

(Wikipédia)

Foco no método experimental, ou pesquisa empírica.

Outras abordagens?

Método Científico



Tipos de pesquisa

- **Básica v. Aplicada**
- **Exploratória v. Confirmatória**
- **Quantitativa v. Qualitativa**

Tipos de pesquisa

- **Pesquisa básica:** busca oferecer um melhor entendimento de um fenômeno, tema ou método em seus fundamentos.
  - pode vir a ter aplicação
  - Ex. estudos sobre o funcionamento do cérebro

- **Pesquisa aplicada:** destinada a solucionar um problema específico em uma circunstância particular.

Tipos de pesquisa

- **Pesquisa exploratória**: visa à melhor definição ou *design* de um fenômeno ainda desconhecido. Importante em estágios iniciais do conhecimento, quando se busca familiaridade com algum objeto/disciplina

- **Pesquisa confirmatória**: em cenários em que já existe uma ou mais teorias (hipóteses), e o objetivo é descobrir se os fatos confirmam a teoria/hipótese

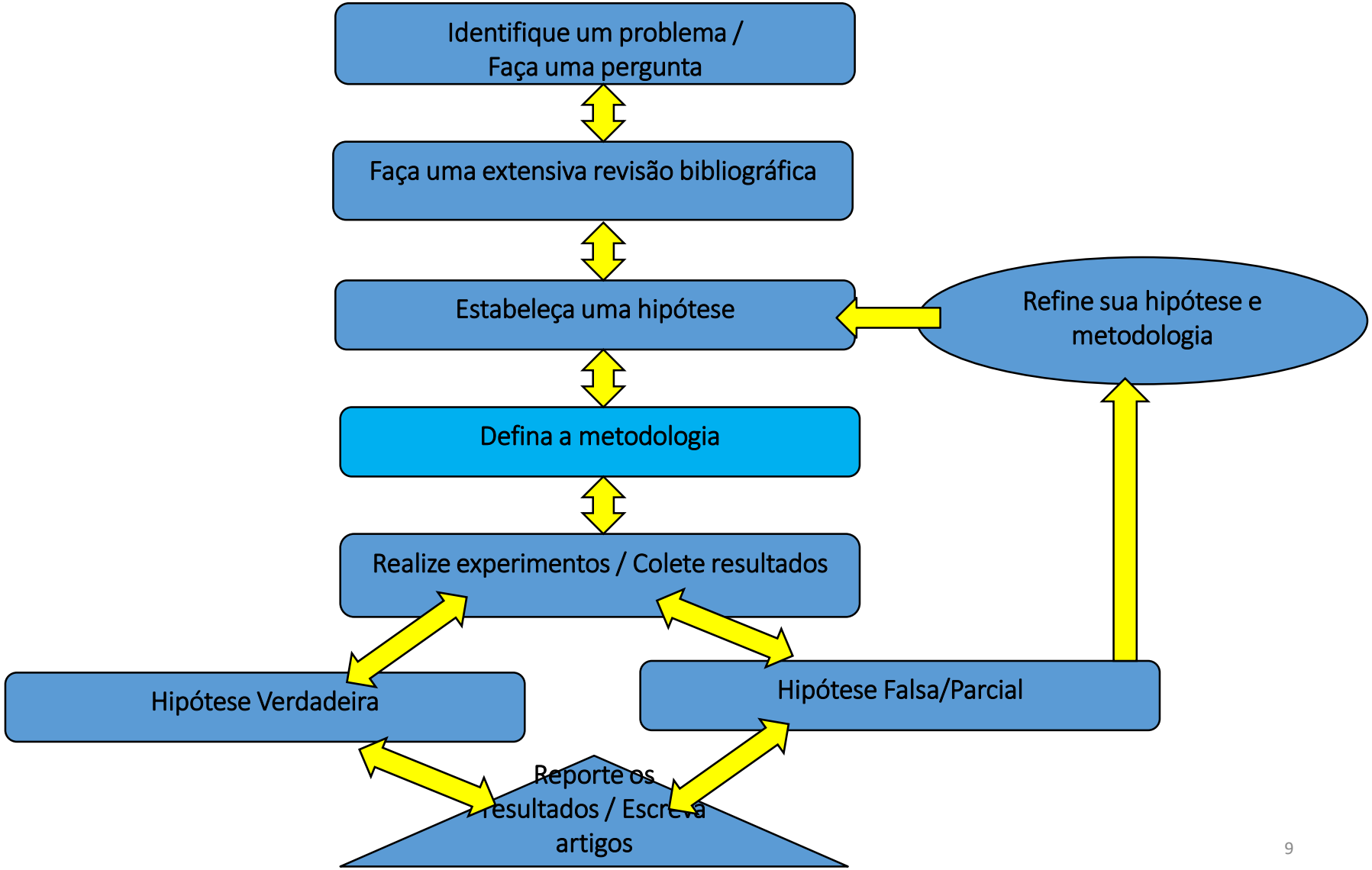
Tipos de pesquisa

- **Pesquisa quantitativa**: tem por base a coleta de dados numéricos a fim de explicar determinado campo de interesse. Normalmente, funda-se em coleta de dados, estatísticas e conclusões com base em projeções estatísticas.

- **Pesquisa qualitativa**: baseada em observação direta do comportamento ou em dados narrativos, de modo que tem por característica sínteses e análises verbais.

- Em geral, pode-se dizer que estudos confirmatórios tendem a ser quantitativos ao passo que estudos exploratórios tendem a ser qualitativos.

Método Científico



O QUE É UMA HIPÓTESE?

- É uma afirmação que pode ser desafiada (i.e., sujeita à negação, negável)

- É uma proposição que pode ser colocada à prova para determinar sua validade; pode parecer contrária, ou de acordo com o senso comum, pode ainda ser correta ou errada

- Uma hipótese conduz a uma verificação empírica

Exemplos

- **Hipótese sociológica:** ao interferir em uma cultura inserindo técnicas de produção de alimentos em larga escala, a sociedade em questão estará apta à autonomia alimentícia.
- **Hipótese biológica:** a exposição de fetos a raios X até o sexto mês de gestação altera a sua formação genética.
- **Hipótese física:** quaisquer objetos soltos a partir da mesma altura acima da superfície da terra irão atingir o solo ao mesmo tempo, sempre que a resistência do ar não constitua um fator interferente.

Exemplo

Em uma pesquisa que se propunha estudar a viabilidade e as condições de execução de cursos a distância, via Internet, foram formuladas as seguintes hipóteses:

- A infra-estrutura da Internet no Brasil possibilita o desenvolvimento de cursos a distância.
- Existe demanda por cursos a distância via Internet para profissionais da área de Comunicação.
- As ferramentas da Internet possibilitam a realização de cursos a distância interativos.

Pesquisa em Ciência da Computação

- “Frequentemente envolve a construção de um **programa**, de um **modelo**, de um **algoritmo** ou de um **sistema** *novo*.”

Fonte: J. Wainer, Métodos de pesquisa... para a CC

Retomando discussão...

1. A computação é uma ciência?

2. O que se procura avaliar ao recomendar um projeto de pesquisa?
3. O que se procura avaliar ao recomendar um artigo científico?

- http://www.fapesp.br/577

Capítulo 2

Estilos de Pesquisa Correntes em Computação

Uma área nova

- Ciência da Computação é **uma área ainda relativamente nova**, ainda em franco desenvolvimento → há necessidade de **embasamento metodológico adequado**
- Após 2000, definição clara das **carreiras**:
  - Bacharelado em Ciência da Computação
  - Bacharelado em Sistemas de Informação
  - Engenharia de Computação
  - Licenciatura em Informática

- **Tipos de pesquisa** (sem amplo consenso)
  1. Apresentação de **um produto**
  2. Apresentação de **uma solução diferente**
  3. Apresentação de **uma solução supostamente melhor**
  4. Apresentação de **uma solução reconhecidamente melhor**
  5. Apresentação de **uma prova**

Uma área nova

- **Tipos de pesquisa** (sem amplo consenso)
  1. Apresentação de **um produto** novo
  2. Apresentação de **uma solução diferente** (outra maneira de resolver um problema)
  3. Apresentação de **uma solução supostamente melhor**
  4. Apresentação de **uma solução reconhecidamente melhor**
  5. Apresentação de **uma prova** (que resolve um problema em aberto)

1. Apresentação de um produto

- **Simples apresentação** de um produto novo
  - Aceito apenas em **áreas de pesquisa recentes**
  - Estratégia exploratória

- **Difícil de comparar** (outros ainda não existem)
  - No entanto, **SEMPRE se deve comparar com alguma coisa**

- Deve-se evitar, pois é um tipo de **pesquisa de baixa aceitação**
- Em eventos científicos, encaixam-se apenas em **categorias especiais:**
  - Sessão de ferramentas
  - Sessão de trabalhos em andamento
  - Sessão de informática aplicada

1. Apresentação de um produto

- **Simples apresentação** de algo novo
  - Aceito apenas em **novas áreas de pesquisa**

Exemplo, Xgobi, Xmdv:

https://www.tandfonline.com/doi/abs/10.1080/10618600.1998.10474764

http://davis.wpi.edu/xmdv/docs/vis94.pdf

  - Sessão de ferramentas
  - Sessão de trabalhos em andamento
  - Sessão de informática aplicada

2. Apresentação de uma solução diferente

- Apresentação de **uma maneira diferente** de fazer algo ou resolver um problema
  - **Avaliação mais qualitativa**
  - Baixo rigor científico
- Exemplo:
  - Uma nova técnica de engenharia de software que não pode ser claramente comparada com outras, pois **não há métricas**
  - Apresentação de estudos de caso
- Requer **bons argumentos, exemplos convincentes…**

2. Apresentação de uma solução diferente

Exemplo, coordenadas paralelas:

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5290770

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=877389

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=809866

2. Apresentação de uma solução diferente

- Outros exemplos (artigos)

3. Apresentação de uma solução supostamente melhor

- Novas técnicas devem ser **comparadas** com outras da literatura
- No entanto, em muitos casos, **não existe um benchmark** para comparar diferentes técnicas
- Nesses casos, o autor propõe o **próprio conjunto de testes** para comparação, e o aplica a todas as técnicas:
  - Estudo experimental
  - Mais trabalhoso
  - Possivelmente tendencioso
  - ➔ É necessário apresentar uma **descrição detalhada** de como os experimentos foram realizados (confiabilidade e reproducibilidade)

3. Apresentação de uma solução supostamente melhor

- Deve-se **comparar** a técnica nova com técnicas que sejam do estado-da-arte; com os trabalhos **mais recentes**

- Fazer uso de **métricas**

- **Evitar** expressões subjetivas, como:
  - "O sistema é *fácil* de usar"
  - "O algoritmo trabalha com *bastante* informação"

3. Apresentação de uma solução supostamente melhor

Exemplo, pagerank and a priori algorithms:

http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf

http://rakesh.agrawal-family.com/papers/vldb94apriori.pdf

4. Apresentação de uma solução reconhecidamente melhor

- Comparação da nova técnica usando-se **testes padronizados e universalmente aceitos**

- **Comparação** com os **resultados já publicados em outros trabalhos**, não há necessidade de fazer experimentos com todas as técnicas alternativas, apenas com os mais recentes

- Usar a **mesma métrica e os mesmos dados**

  → o experimento pode ser facilmente **reproduzido**

- **Com sucesso → avanço do estado da arte**

4. Apresentação de uma solução reconhecidamente melhor

Exemplo, fastmap algorithm, m-tree (slim-tree):

https://dl.acm.org/citation.cfm?id=223812

http://www.cs.cmu.edu/~christos/PUBLICATIONS/EDBT_SlimTree.pdf

4. Apresentação de uma solução

Supostamente x Reconhecidamente

- Supostamente melhor refere-se a comparações **cujas métricas ou argumentos podem ser contestados** – por exemplo, um algoritmo de processamento de imagens pode ter sido beneficiado pela escolha das imagens e pela escolha do pré-processamento

- Reconhecidamente melhor refere-se a **comparações incontestáveis** – por exemplo, um algoritmo de compactação de arquivos que representa a mesma informação com menos bits e que foi testado sobre um *benchmark* aceito pela comunidade

Reconhecidamente é uma validação mais estrita do que supostamente.

5. Apresentação de uma prova

- Até aqui: uso de **evidências empíricas** e de **resultados que sugerem provas**

- Existem ainda, **pesquisas que exigem provas matemáticas de acordo com regras de lógica**
  - Exemplo: métodos formais e compiladores

- Deve-se desenvolver teoria **cuja aplicação leva logicamente a determinados resultados**, com uma demonstração formal
  - Exemplo: para ordenação, já foi **provado** que o melhor custo possível é O(n log n)*

* Algoritmos que usam comparação em domínios sem restrição, e com relação de ordem definida

5. Apresentação de uma prova

- Até aqui: uso de **evidências empíricas** e de **resultados que sugerem provas**

- Existem ainda, **pesquisas que exigem provas matemáticas de acordo com r**

- D
r

Exemplo, I/O complexity bounds, P !=NP

http://www.cc.gatech.edu/~bader/COURSES/UNM/ece637-Fall2003/papers/AV88.pdf

http://www.win.tue.nl/~gwoegi/P-versus-NP/Deolalikar.pdf

* Algo

Tipos básicos de pesquisa

a) **Exploratória**: não há teoria formal nem medidas de avaliaçãobem definidas: estudos de caso, e análises comparativas

→ Requer **boa argumentação**

➔ tem **baixa aceitação** (muito suscetível à viés...)

b) **Empírica**: baseia-se em testes bem aceitos

→ Requer **testes estatísticos**

➔Pode ser questionada, pois estatística não explica causas – pede uma **teoria complementar**

c) **Formal**: teoria demonstrável

→ Requer **lógica**

➔ Difícil de refutar

Outra visão: como obter conhecimento em CC

- Pesquisa empírica
  - Quantitativa
  - Qualitativa

- Pesquisa analítica
  - Prova matemática, a partir de certas premissas

- Pesquisa bibliográfica
  - Revisão sistemática
  - Meta-análises

fonte: J. Wainer, Pesquisa...

Outra visão: como obter conhecimento em CC

- Pesquisa Quantitativa
  - Medida de algumas variáveis objetivas, ênfase em comparação de resultados, técnicas estatísticas
  - Ex. benchmarks, simulações, competições, estudos experimentais, questionários, ...
- Pesquisa Qualitativa
  - Estudos observacionais, ...
- Foco nos processos de validação, e não no desenvolvimento de artefatos

fonte: J. Wainer, Pesquisa...

- Na área de visualização de informação avaliação e validação representam um desafio

Exemplo, avaliação de técnicas de visualização de informação:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.188.3308&rep=rep1&type=pdf

http://triton.cc.gatech.edu/hci-seminar/uploads/1/The%20Challenge%20of%20Information%20Visualization%20Evaluation.pdf

https://dl.acm.org/citation.cfm?id=2669579

https://dl.acm.org/citation.cfm?id=1168158

Toda pesquisa tem uma hipótese?

- Em geral, pesquisas confirmatórias partem de uma **hipótese de trabalho**

- A pesquisa consiste, essencialmente, em definir uma metodologia que permita verificar a validade da hipótese

Toda pesquisa tem uma hipótese?

- Existem pesquisas que são apenas exploratórias. Sua finalidade é abrir campo para que se possa construir hipóteses, como em geral ocorre em fase inicial de pesquisas de doutoramento. Portanto, o objetivo é apresentar novas evidências, ideias, esclarecer algo, e até mesmo concluir que algo não existe.

- Em dissertações é comum optar por pesquisas exploratórias a fim de apresentar uma síntese crítica de determinada teoria, ou sistematizar certa prática. Técnicas típicas da pesquisa exploratória incluem estudos de caso, observação e análise histórica de informações.

Toda pesquisa tem uma hipótese?

- **Hipótese**: uma teoria provável, mas ainda não demonstrada, ou uma suposição admissível

- A hipótese é o **cerne de um trabalho científico**
  - Será **testada/colocada em prova** ao longo do trabalho
  - Ao final, deve haver **evidências de sua validade** (ou não!)
  - Deve-se **demonstrar** que a tese é válida ➔ **sucesso**

O QUE É UMA BOA HIPÓTESE?

- Em resumo, uma hipótese de trabalho deve ser: uma **afirmação**; **simples**; **sujeita à negação**.

  - **Afirmação:** uma hipótese não é uma pergunta, uma hipótese é uma afirmação sobre algo.
  - **Simples:** uma boa hipótese é escrita em linguagem simples de maneira a expressar exatamente o que está em jogo.
  - **Sujeita à negação:** uma hipótese deve poder ser negada. Caso seja impossível estabelecer a sua negação dificilmente será considerada uma hipótese.

http://observatorio.fiocruz.br/em-foco/noticias/como-construir-uma-hipotese-de-trabalho-e-apresentar-bem-sua-pesquisa

O que se entende por negação

- Hipóteses são afirmações provisórias, e por isso, elas podem ser rejeitadas. O "como" rejeitar uma hipótese depende do seu objeto. Basicamente, uma hipótese pode ser negada de dois modos: empírica ou teoricamente.

- **Negação empírica:** é a que possibilita afastar uma hipótese empírica, i.e., baseada em fenômenos observados e medidos.
- **Negação teórica:** é a que possibilita sustentar uma afirmação, por não contradição de seus argumentos. Uma hipótese teórica fixa-se fundamentalmente em documentação, interpretação de dados ou desenvolvimento de métodos.

Necessidade de uma hipótese

- Nem todo trabalho parte de uma hipótese definida
- De fato, muitos trabalhos são desenvolvidos e defendidos (publicados) sem uma hipótese
- No entanto, a definição de uma hipótese traz muitos benefícios:
  - Norteia de maneira eficaz
  - Permite reflitir sobre o trabalho de maneira mais clara
  - Direciona os estudos experimentais
  - Indica quais trabalhos são “concorrentes”
  - Permite a análise dos resultados de maneira mais clara

Necessidade de uma hipótese

- Em se tratando de técnica, uma maneira de definir uma hipótese de trabalho é por meio da **identificação das propriedades** de cada uma das **abordagens existentes**, propondo-se uma nova que tenha, além das propriedades existentes, **propriedades adicionais** ➔ montar uma tabela, como exemplo:

	Característica 1	Característica 2	Característica 3	Característica 4
Abordagem 1	X	X		
Abordagem 2	X			
Abordagem 3		X	X	
Nova Abordagem	**X**	**X**	**X**	**X**

Conclusões da pesquisa

- **Conclusão descritiva:** a conclusão seleciona os principais aspectos encontrados ao longo da pesquisa. Uma forma eficiente de estabelecer uma conclusão descritiva é destacar as impressões dominantes do trabalho.

- **Conclusão avaliativa:** enfatiza os aspectos positivos e negativos encontrados na pesquisa. Uma forma eficiente de concluir de maneira avaliativa é estabelecendo comparações dos resultados.

- **Conclusão prescritiva:** recomenda medidas com base nos resultados obtidos. Uma forma eficiente de escrever uma conclusão prescritiva é relacionando as medidas recomendadas a cada aspecto do fenômeno estudado.

Algumas perguntas

- **Simples apresentação** de um produto novo
  - Aceito apenas em **áreas de pesquisa recentes**
  - Estratégia exploratória
- Deve-se evitar, pois é um tipo de **pesquisa de baixa aceitação**

- Pergunta: há exceções?
- Você consegue identificar trabalhos sobre “produtos” que tenham tido ampla aceitação e reconhecimento?

Retomando discussão...

1. A computação é uma ciência?

Read Paper Peter J. Denning, CACM – The Science in Computer Science

- http://denninginstitute.com/pjd/GP/GP-site/welcome.html

Atividades para próxima aula

- O que é um benchmark?
- Dê exemplos de benchmarks, e fale sobre as áreas/problemas da computação em que eles são utilizados.

Atividades para próxima aula:

- Pesquisar por um artigo que ilustre de cada um dos 5 tipos de pesquisa discutidos (na disciplina Computação).
- Formular a hipótese de trabalho para o seu projeto de pesquisa (não precisa ser a hipótese definitiva)

Apresentação de cada artigo encontrado, mostrando-se os elementos que sustentam a categorização dada àquele artigo.

Apresentação da hipótese para discussão.

Para formular sua hipótese (tentativa)

- Quero estudar/desenvolver (1. resuma o seu tópico) para descobrir/resolver (2. identifique o que você quer descobrir/resolver) de modo a ajudar o meu usuário a (3. informe a relevância de sua pesquisa).

Bibliografia

- Fonte: cap. 2 livro Prof. Wazlawick

- Wainer, J. (2007). Métodos de pesquisa quantitativa e qualitativa para a ciência de computação.

Layout e bibliografia extraídos de http://goo.gl/m9i0m

O conteúdo da apresentação não é uma reprodução exata do conteúdo do livro, não sendo, portanto, de responsabilidade do autor Prof. Raul Sidnei Wazlawick