

SCC 5933 - METODOLOGIA DE PESQUISA CIENTÍFICA EM COMPUTAÇÃO

A COMPUTAÇÃO E A CLASSIFICAÇÃO DAS CIÊNCIAS
MÉTODOS DE PESQUISA
ESTILOS DE PESQUISAS CORRENTES EM CIÊNCIA DA COMPUTAÇÃO

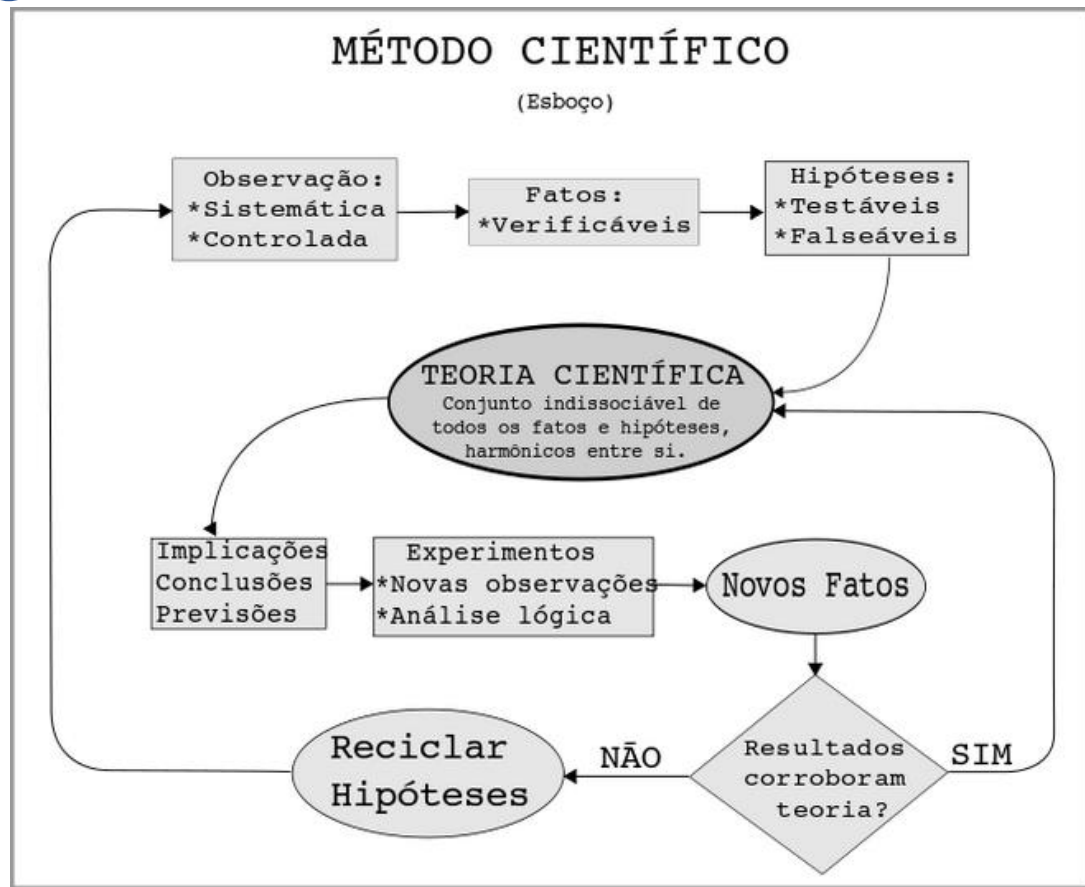
Profa. Sandra M Aluisio

Ciência

- A ciência é o esforço para descobrir e aumentar o [conhecimento humano](#) de como o [Universo](#) funciona. Refere-se tanto à (ao):
 - investigação ou estudo [racionalis](#) do [Universo](#). Tal estudo ou investigação é metódico e compulsoriamente realizado em acordo com o [método científico](#) – um processo de avaliar o [conhecimento empírico](#);
 - corpo organizado de conhecimentos adquiridos por tais estudos e [pesquisas](#). (WIKIPÉDIA, 2015).
- Face à variedade de abordagens, várias [classificações das ciências](#) foram produzidas no sentido de tentar melhor entender seus métodos e objetivos (WAZLAWICK, 2010).

Método Científico

- O **método científico** refere-se a um aglomerado de regras básicas de como deve ser o procedimento a fim de produzir conhecimento dito científico, quer seja este um novo conhecimento, quer seja este fruto de uma **totalidade, correção** (evolução) ou um **aumento da área de incidência de conhecimentos** anteriormente existentes.



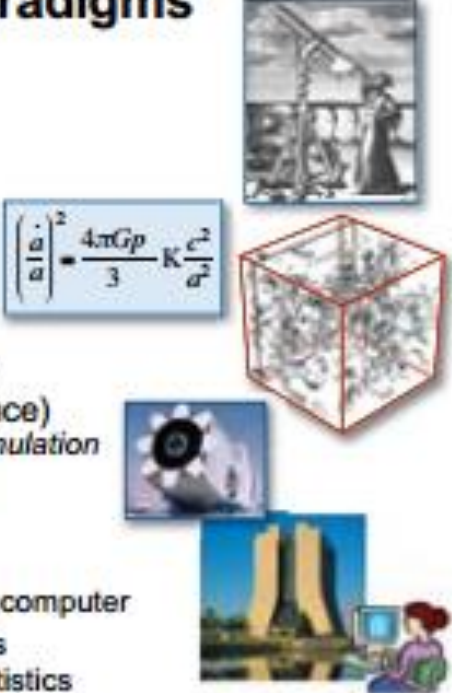
- O método é cíclico de forma a promover a contínua evolução das [teorias científicas](#).

- Popper afirmou que a ciência é um **conhecimento provisório**, que funciona através de sucessivos falseamentos.

Paradigmas da Ciência (Jim Gray on eScience: A Transformed Scientific Method)

Science Paradigms

- Thousand years ago: science was **empirical**
describing natural phenomena
- Last few hundred years: **theoretical** branch
using models, generalizations
- Last few decades: a **computational** branch
simulating complex phenomena
- Today: **data exploration** (eScience)
unify theory, experiment, and simulation
 - Data captured by instruments or generated by simulator
 - Processed by software
 - Information/knowledge stored in computer
 - Scientist analyzes database/files using data management and statistics



The collage contains four images: 1. A historical scientific experiment with a pendulum and a person. 2. A mathematical equation:
$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K \frac{c^2}{a^2}$$
 3. A 3D cube with internal structures, possibly representing a simulation or data visualization. 4. A modern scientist at a computer, representing data exploration.

We have to do better at producing tools to support the whole research cycle—from data capture and data curation to data analysis and data visualization.

Última palestra no Conselho Nacional de Pesquisa (USA) em Mountain View, CA, em Janeiro 11, 2007

The Coming Revolution in Scholarly Communication

- Internet can do **more than just make available the full text of research papers.**
 - In principle, it can **unify all the scientific data** with all the literature to create a world in which the data and the literature interoperate with each other.
- I wanted to point out that almost everything about science is changing because of the **impact of information technology.** **Experimental, theoretical, and computational science** are all being affected by the data deluge, and a fourth, “**data-intensive**” **science paradigm** is emerging.
 - The goal is to have a world in which all of the science literature is online, all of the science data is online, and they interoperate with each other. Lots of new tools are needed to make this happen.

Classificações das Ciências*

- Ciências Formais e Empíricas
- Ciências Puras e Aplicadas
- Ciências Exatas e Inexatas
- Ciências Duras e Moles
- Ciências Nomotéticas e Idiográficas

*Seção II de (WAZLAWICK, 2010).

Ciências Formais

- Estudam as ideias independentemente de sua aplicação à natureza ou ao ser humano (o que não quer dizer que não possam ser aplicadas).
- Ex.: Lógica, Matemática.

Ciências Empíricas

- Estudam os fenômenos que ocorrem no mundo real.
- Têm de fazer uso de observações para fundamentar suas descobertas.
 - **Ciências Naturais:** estudam a natureza em seus aspectos que independem da existência ou da ação do ser humano. Ex.: Astronomia, Física, Química etc.
 - **Ciências Sociais:** estudam o ser humano e suas interações. Ex.: História, Psicologia e Sociologia.

Ciências Formais x Empíricas

- **E a Computação?**
- **Ciências Formais:**
 - Teoria dos algoritmos (estruturas de dados, complexidade), teoria das linguagens formais e seus autômatos, etc.
- **Ciências Empíricas**
 - **Ciências Naturais:**
 - Eletrônica, Circuitos Lógicos
 - **Ciências Sociais:**
 - Informática na Educação, Comércio Eletrônico, Games, Interação Humano-Computador

Ciências Puras e Aplicadas

- **Ciências puras (ou básicas):** estudam os conceitos básicos do conhecimento, sem preocupação com sua imediata aplicação. Podem ser formais (p.ex., Lógica) ou empíricas (p.ex., Cosmologia).
- **Ciências aplicadas:** visam à realização de descobertas que possam ser imediatamente aplicadas a algum processo industrial ou assemelhado, visando produzir algum tipo de ganho. Ex.: Engenharias

Ciências Puras e Aplicadas

- **E a Computação?**
- **Ciência Pura:**
 - O aspecto de ciência básica da Computação é difícil de identificar visto que a maioria dos resultados em Computação possui aplicação prática... Mas há. P.ex., estudos de aprendizagem humana simulada por computador (entender processos de aprendizado via modelos que trazem teorias para explicar fenômenos)
- **Ciência Aplicada:**
 - Engenharia de Software, Informática na Educação etc.

Ciências Exatas e Inexatas

- **Ciências Exatas:** são aquelas cujos resultados são precisos. Suas leis são altamente preditivas e previsíveis.
 - Ex.: Matemática, Física, Química
- **Ciências Inexatas:** podem prever comportamentos gerais de seus fenômenos, mas nem sempre os resultados são os esperados.
 - Ex.: Meteorologia, Economia e a maioria das Ciências Sociais

Ciências Exatas e Inexatas

- **E a Computação?**
- Assim como outras ciências exatas, a Computação também tem aspectos inexatos.
 - Ex.: Algoritmos genéticos e alguns modelos de redes neurais são capazes de produzir resultados inesperados mesmo quando aplicados repetidamente a um mesmo conjunto de dados.

Ciências Duras e Moles

- **Ciências Duras:** são aquelas que usam de rigor científico em suas observações, experimentos e deduções.
- **Ciências Moles:** costumam aceitar evidências baseadas em **estudos de caso**. Isso ocorre quando é difícil ou impossível conseguir realizar experimentos totalmente controlado.

Ciências Duras

- **Ciências duras formais:** utilizam fortemente a Lógica e a Matemática como ferramentas de construção teórica.
- **Ciências duras naturais:** dependem muitas vezes de comprovação estatística para dar credibilidade a seus experimentos. Exigem grande rigor na comprovação de resultados empíricos.
 - Ex.: Medicina

Ciências Duras e Moles

- **E a Computação?**
- Normalmente entende-se a Computação como uma Ciência Dura, mas a realidade ainda, em muitos casos,
 - é que os pesquisadores têm **dificuldade em providenciar dados em quantidade suficiente** para dar suporte empírico a suas conclusões. **Precisamos de bons benchmarks!**
- Assim é comum encontrar artigos em Computação que utilizam um ou alguns poucos **estudos de caso** para tentar “validar” uma técnica, modelo ou teoria.
 - Estudo de caso é uma fonte de dados para uma pesquisa exploratória, mas **não** valida a hipótese em estudo.

Ciências Namotéticas e Idiográficas

- Ciência namotética: estudam fenômenos que se repetem e que podem levar à descoberta de leis gerais que permitam fazer previsões.
- Ciência idiográficas: analisam fenômenos únicos que não se repetem, mas que ainda assim têm validade como campo de estudo.
 - Ex: História. Estuda fatos que nunca se repetem.

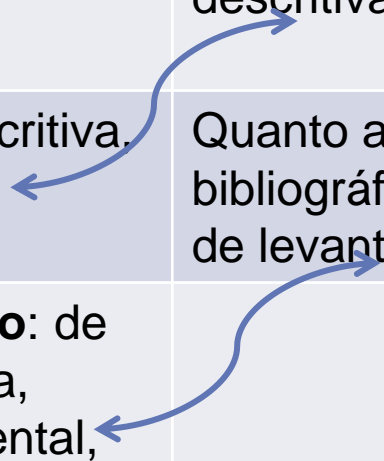
Ciências Namotéticas e Idiográficas

- E a Computação?
- Poucas áreas são idiográficas. Por exemplo, o estudo da própria história da computação e o desenvolvimento de determinadas tecnologias, como linguagens, arquiteturas.

Classificações dos Métodos de Pesquisa

- **Moresi (2003) e Waslawick (2010)** trazem diferentes **critérios** para classificar pesquisas. Tipos não são mutuamente exclusivos.
- Uma pesquisa pode ter mais de uma finalidade simultaneamente

Moresi (2003)	Waslawick (2010)
Quanto à natureza : Pesquisa Básica / Aplicada	Quanto à natureza : trabalho original e survey.
Quanto à forma de abordagem do problema : Pesquisa Quantitativa e Qualitativa	Quanto aos objetivos : exploratória, descritiva ou explicativa
Quanto aos fins : Exploratória, Descritiva, Explicativa, Metodológica, Intervencionista	Quanto aos procedimentos técnicos : bibliográfica, documental, experimental, de levantamento e pesquisa-ação
Quanto aos meios de investigação : de campo, de laboratório, telematizada, documental, bibliográfica, experimental, ex post facto, participante, pesquisa-ação, estudo de caso.	



Pesquisa Quantitativa x Pesquisa Qualitativa

Quantitativa	Qualitativa
Objetiva	Subjetiva
Testa uma teoria	Desenvolve uma teoria
Controle e precisão	Descoberta, descrição e compreensão
Mensuração	Interpretação
Possibilita análises estatísticas	Possibilita narrativas ricas
Números são os elementos básicos da análise	Ideias são os elementos básicos da análise
Testa hipóteses	Gera ideias e questões para pesquisa
Raciocínio lógico e dedutivo	Raciocínio baseado em argumentação e indutivo
Estabelece relações, causas	Descreve significados, descobertas

Adaptado de (MORESI, 2003)

Ditado Popular


- “Teoria é quando o fenômeno é bem compreendido, mas não funciona. Prática é quando funciona, mas não se sabe por quê” (WAZLAWICK, 2009).
- Em Computação: “Nada funciona e não se sabe por quê” (WAZLAWICK, 2009).

Uma área nova

- ◆ Ciência da Computação é **uma área relativamente nova** 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

Estilos da Pesquisa Corrente em Computação

- De acordo com Wazlawick (2009), os tipos de pesquisa realizados em Computação correntemente podem ser classificados nos seguintes estilos:
 1. Apresentação de **um produto**
 2. Apresentação de **algo diferente**
 3. Apresentação de **algo presumivelmente melhor**
 4. Apresentação de **algo reconhecidamente melhor**
 5. Apresentação de **uma prova**



Níveis de
Maturidade
Científica

1. Apresentação de um produto

- ◆ **Simples apresentação** de algo novo
 - ◆ Aceito apenas em **novas áreas de pesquisa**
- ◆ **Difícil de comparar** com outros que 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 **sessões especiais**:
 - ◆ Sessão de ferramentas
 - ◆ Sessão de trabalhos em andamento
 - ◆ Sessão de informática aplicada

2. Apresentação de algo diferente

- ◆ Apresentação de **uma maneira diferente** de se fazer algo
 - ◆ **Avaliação qualitativa, pouco quantitativa**
- ◆ 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 (**baixo rigor científico**)
- ◆ Requer **bons argumentos**

3. Apresentação de algo presumivelmente melhor

- ◆ Novas técnicas devem ser **comparadas** com outras da literatura
- ◆ No entanto, em muitos casos, **não há um benchmark** sobre o qual diferentes técnicas possam ser comparadas
- ◆ Em casos assim, o autor propõe o **próprio conjunto de testes** para comparação, e o aplica a todas as técnicas:
 - ◆ Mais trabalhoso
 - ◆ Possivelmente tendencioso
- ➔ É necessária uma **detalhada descrição** de como os experimentos foram realizados

3. Apresentação de algo presumivelmente 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 do tipo:
 - ◆ “O sistema é *fácil* de usar”
 - ◆ “O algoritmo trabalha com *bastante* informação”

4. Apresentação de algo reconhecidamente melhor

- ◆ Comparação da nova técnica usando-se **testes padronizados e universalmente aceitos (benchmarks)**
- ◆ **Comparação com os resultados já publicados em outros trabalhos**, não há necessidade de se fazer os 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 algo reconhecidamente melhor

- ◆ Pesquisa “**menos trabalhosa**” que **exige uma boa hipótese** → no entanto, **notadamente mais difícil**
- ◆ **Necessário:**
 - ◆ Amplo estudo
 - ◆ Reflexão
- ◆ **Problemas em aberto** são bons candidatos
 - ◆ http://en.wikipedia.org/wiki/List_of_unsolved_problems_in_computer_science
- ◆ Potencialmente útil:
 - ◆ **Conhecimento de outras áreas;** exemplo: computação bioinspirada

4. Apresentação de algo reconhecidamente melhor

- ◆ Pesquisa “**menos trabalhosa**” que **exige uma boa hipótese** → no entanto, **notadamente mais difícil**
- ◆ **Necessário:**
 - ◆ Amplo estudo
 - ◆ Reflexão
- ◆ **Problemas em aberto** são bons candidatos
 - ◆ http://en.wikipedia.org/wiki/List_of_unsolved_problems_in_computer_science
- ◆ Potencialmente útil:
 - ◆ **Conhecimento de outras áreas;** exemplo: computação bioinspirada

4. Apresentação de algo

Presumivelmente x Reconhecidamente

- ◆ Presumivelmente melhor refere-se a comparações **cujas métricas podem ser contestadas** – por exemplo, um algoritmo de processamento de imagens/textos que pode ter sido beneficiado pela escolha das imagens/textos 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 presumivelmente.

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 se **provou** que o melhor custo possível é $O(n \log n)$ *

* Algoritmos que usam comparação de chaves

5. Apresentação de uma prova

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

- ◆ Exemplo, I/O complexity bounds, $P \neq NP$

The Input/Output Complexity of Sorting and Related Problems

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

- ◆ Tese 2010: $P \neq NP$

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

<http://www.scottaaronson.com/blog/?p=458>

Exemplo: para ordenação, já se prova que o melhor custo possível é $O(n \log n)^*$

* Algoritmos que usam comparação de chaves

Tipos básicos de pesquisa

Portanto, os tipos de pesquisa apresentados podem se classificados nas seguintes categorias:

- a) Exploratória:** não há teoria formal nem medidas bem definidas: estudos de caso, e análises comparativas
 - Requer **boa argumentação**
 - De **baixa aceitação**

Tipos básicos de pesquisa

Portanto, os tipos de pesquisa apresentados podem se classificados nas seguintes categorias:

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

→ Requer **boa argumentação**

→ De **baixa aceitação**

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

→ Requer **testes estatísticos**

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

Tipos básicos de pesquisa

Portanto, os tipos de pesquisa apresentados podem se classificados nas seguintes categorias:

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

→ Requer **boa argumentação**

→ De **baixa aceitação**

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

Necessidade de uma hipótese

Necessidade de uma hipótese

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

Necessidade de uma hipótese

- ◆ **Hipótese:** uma teoria provável, mas ainda não demonstrada, ou uma suposição admissível
- ◆ A hipótese é o **coração de um trabalho científico**
 - ◆ Será **testada/colocada em prova** ao longo do trabalho
 - ◆ Ao final, deve haver **evidências de sua validade**
 - ◆ Deve-se **demonstrar** que a proposta é válida → **sucesso**

Necessidade de uma hipótese

- ◆ **Hipótese**: uma teoria provável, mas ainda não demonstrada, ou uma suposição admissível
- ◆ A hipótese é o **coração de um trabalho científico**
 - ◆ Será **testada/colocada em prova** ao longo do trabalho
 - ◆ Ao final, deve haver **evidências de sua validade**
 - ◆ Deve-se **demonstrar** que a proposta é válida → **sucesso**
- ◆ Uma maneira de se 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:

Necessidade de uma hipótese

- ◆ **Hipótese**: uma teoria provável, mas ainda não demonstrada, ou uma suposição admissível
- ◆ A hipótese é o **coração de um trabalho científico**
 - ◆ Será **testada/colocada em prova** ao longo do trabalho
 - ◆ Ao final, deve haver **evidências de sua validade**
 - ◆ Deve-se **demonstrar** que a proposta é válida → **sucesso**
- ◆ Em se tratando de técnica, uma maneira de se 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

Necessidade de uma hipótese

- ◆ Nem todo trabalho tem 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 que se reflita a respeito do trabalho de maneira mais clara
 - ◆ Direciona os experimentos
 - ◆ Indica quais trabalhos são concorrentes
 - ◆ Permite a análise dos resultados de maneira mais clara

Atividade para a próxima aula

- Checar quais tipos de pesquisa (classificação de Wazlawick (2009)) os 5 documentos possuem
- Formular uma hipótese a respeito de **seu trabalho** (não precisa ser a hipótese definitiva)

Selecionar 1 ou 2 alunos para apresentar para cada artigo encontrado, os elementos que sustentam a categorização dada àquele artigo.

Apresentação da hipótese para discussão.

Bibliografia

- WAZLAWICK, R.S. (2009). Metodologia de pesquisa para Ciência da Computação, 184p. Editora Campus/Elsevier. ISBN: 9788535235227.
- WAZLAWICK., R.S., [“Uma Reflexão sobre a Pesquisa em Ciência da Computação à Luz da Classificação das Ciências e do Método Científico”](#), *Revista de Sistemas de Informação da FSMA*, No. 6, pp. 3-10, 2010. Disponível em: http://www.fsma.edu.br/si/edicao6/FSMA_SI_2010_2_Principal_1.html