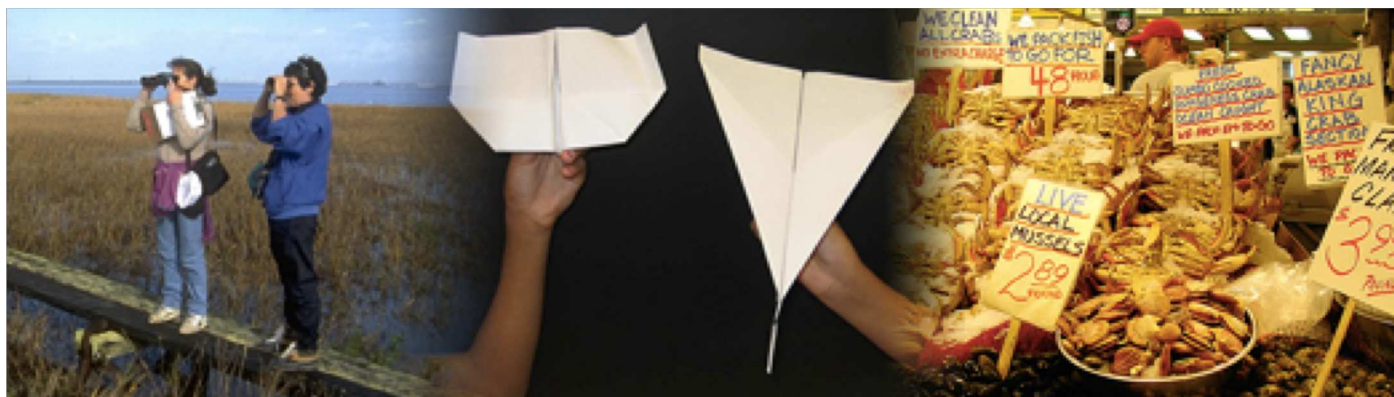


Método Científico e Escrita de Artigos



What comes to mind when you
hear the word Science?



Source: http://undsci.berkeley.edu/article/intro_01



These images reflect only some aspects of Science, since it has several facets



Source: http://undsci.berkeley.edu/article/intro_01

What is Science?

- Science is both a **body of knowledge** and a **process**
 - It is not only isolated and static facts listed in a textbook
 - It is also a process of discovery that allows us to link isolated facts into coherent and comprehensive understandings of the natural world

Science is **useful**

- The knowledge generated by science is powerful and reliable
- It can be used to develop new technologies, treat diseases, and deal with many other sorts of problems

The Brick Wall of Knowledge

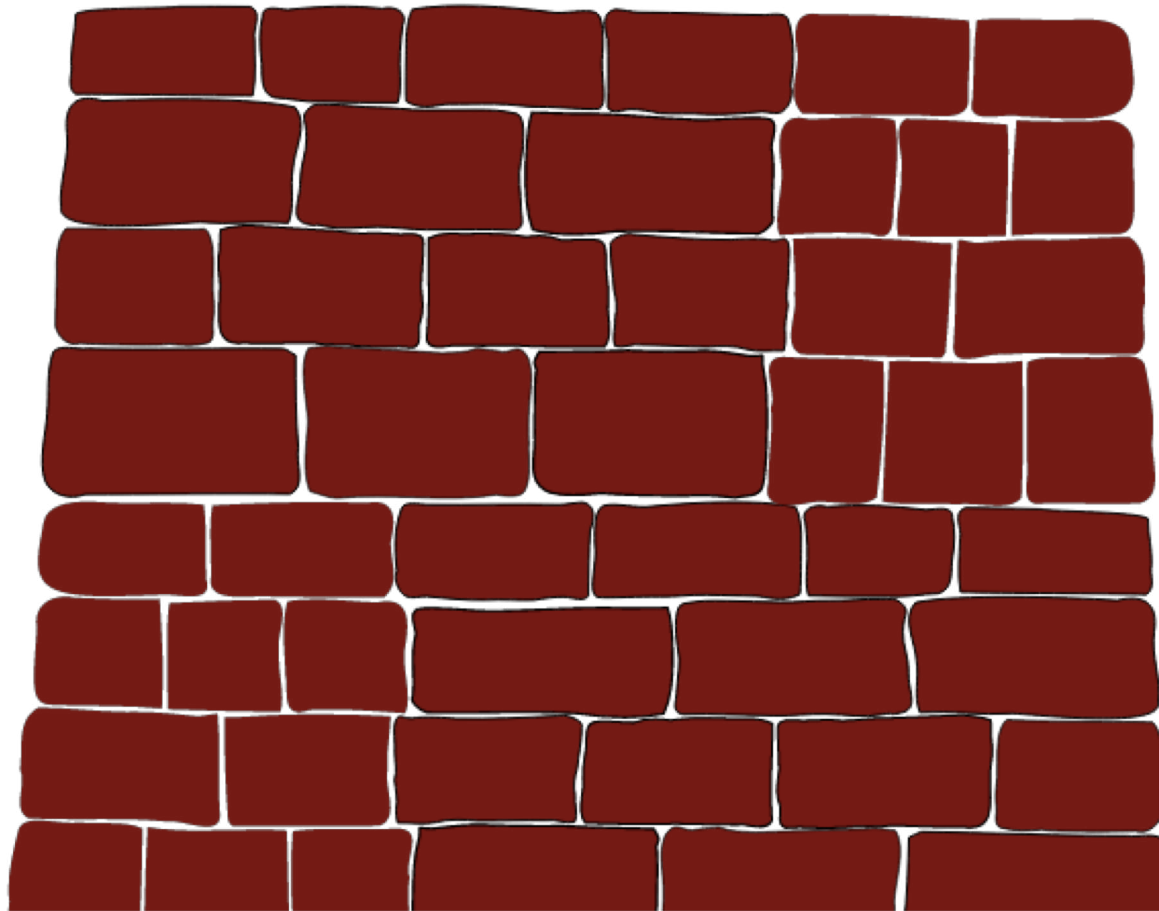


Image Source: <https://clipartfest.com/categories/view/cc90dca431547f9c9d22265d24142adced0b8cf4/brick-wall-clip-art.html>

Science is **ongoing**

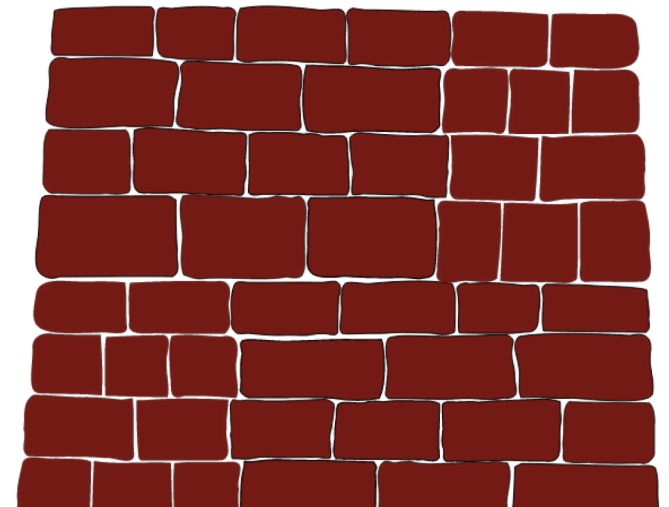
- Science is continually refining and expanding our knowledge of the universe, and as it does, it leads to new questions for future investigation
- Science **will never be "finished"** (the wall to be built is infinite)

Science is a **global human endeavor**

- People all over the world participate in the process of science
- And you can too!

How can I participate?

How can I add a
brick to the Wall of
Knowledge?



Every baby knows the
scientific method!



1 Make an observation.



2 Form a hypothesis.



4 Analyze the data.



3 Perform the experiment.



6 Invite others to reproduce the results.



5 Report your findings.

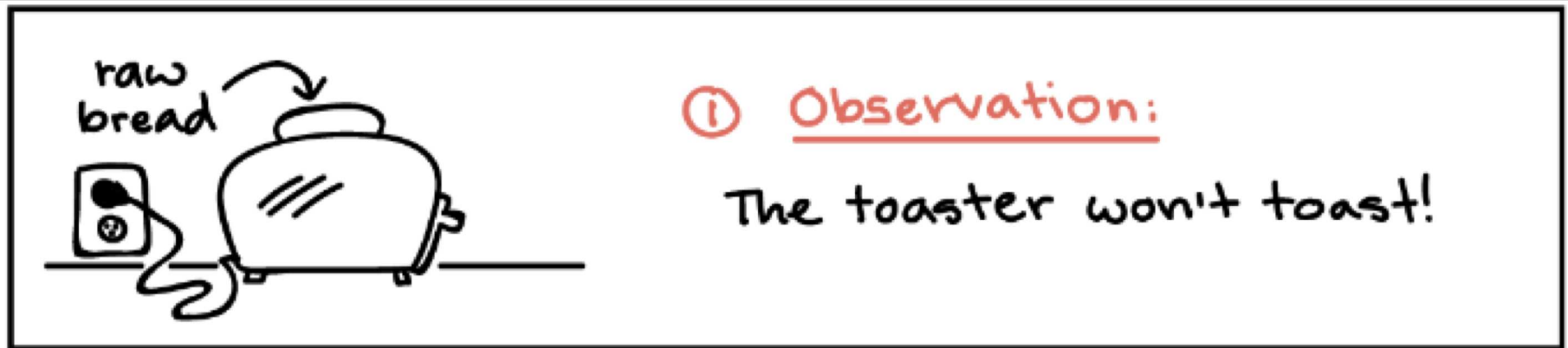
Scientific Method

1. *Make an Observation*
2. *Ask a Question*
3. *Formulate an Hypothesis*
4. *Make a prediction based on the hypothesis*
5. *Test the prediction*
6. *Iterate: use the results to make new hypotheses or predictions*

Source: <https://www.khanacademy.org/science/biology/intro-to-biology/science-of-biology/a/the-science-of-biology>

Example: Fail to Toast

1. Make an observation

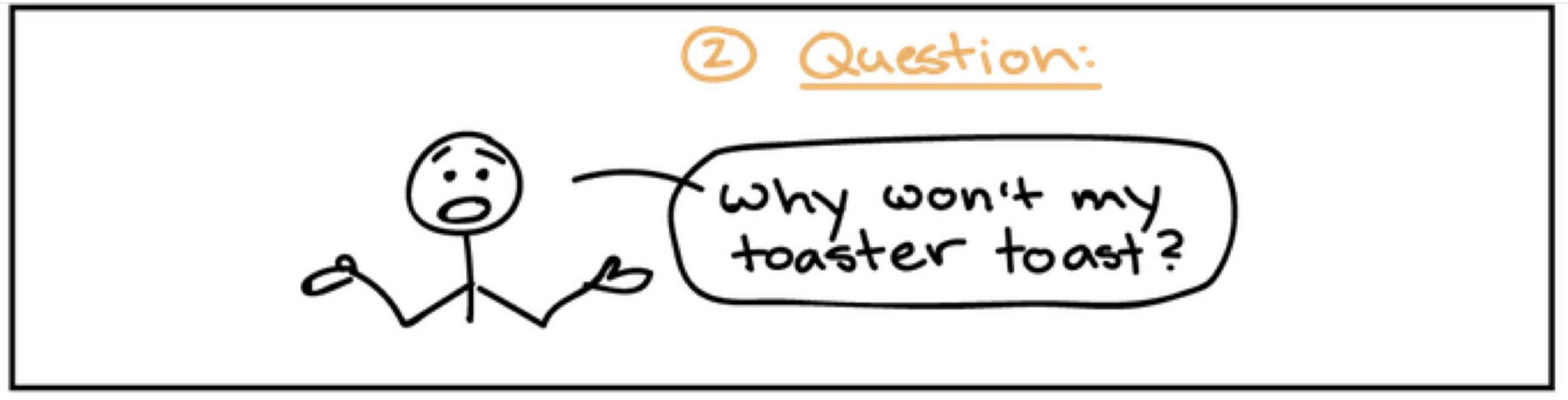


Source of the example: <https://www.khanacademy.org/science/biology/intro-to-biology/science-of-biology/a/the-science-of-biology>

Example: Fail to Toast

2. Ask a question

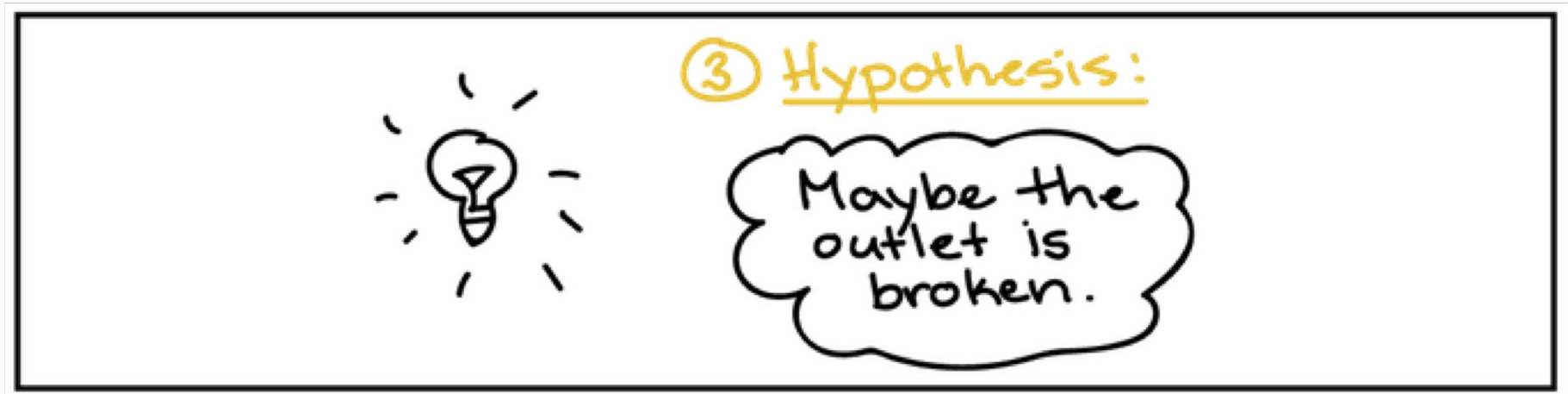
Why didn't my bread get toasted?



Example: Fail to Toast

3. Propose a hypothesis

A hypothesis is a potential answer to the question, one that can somehow be tested.



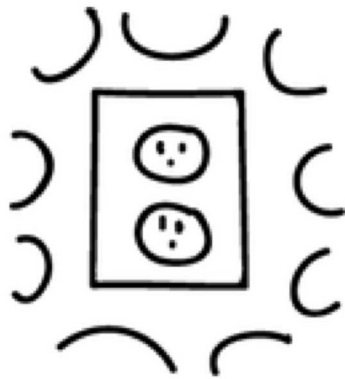
Hypothesis

- A hypothesis **is not necessarily the right explanation**
- Instead, **it's a possible explanation that we can test** to see if it is likely correct, or if we need to make a new hypothesis

Example: Fail to Toast

4. Make a prediction

A prediction **is an outcome we'd expect to see if the hypothesis is correct.**



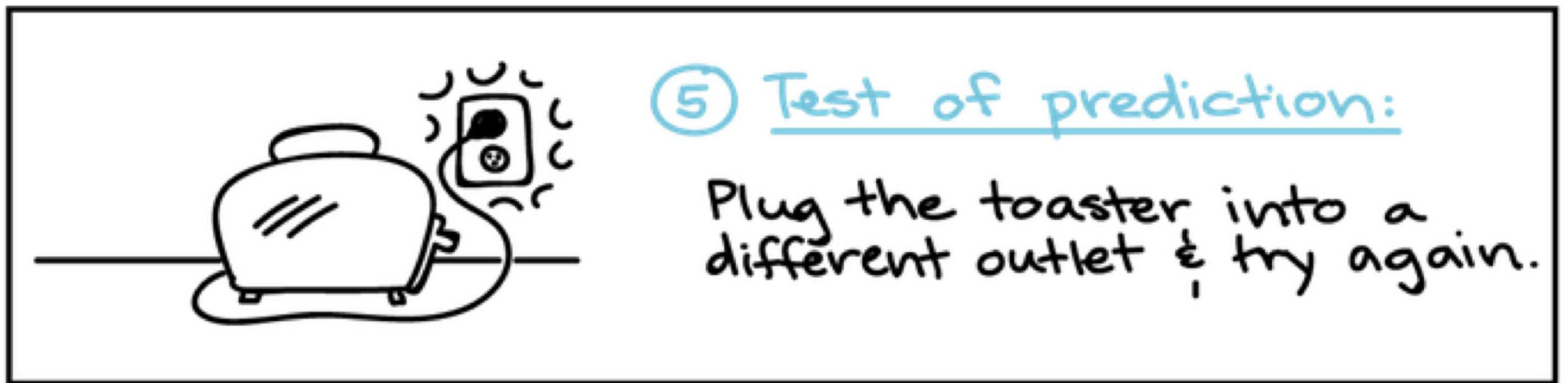
④ Prediction:

If I plug the toaster into a different outlet, then it will toast the bread.

Example: Fail to Toast

5. Test the prediction

To test the hypothesis, we need to **make an observation or perform an experiment** associated with the prediction.



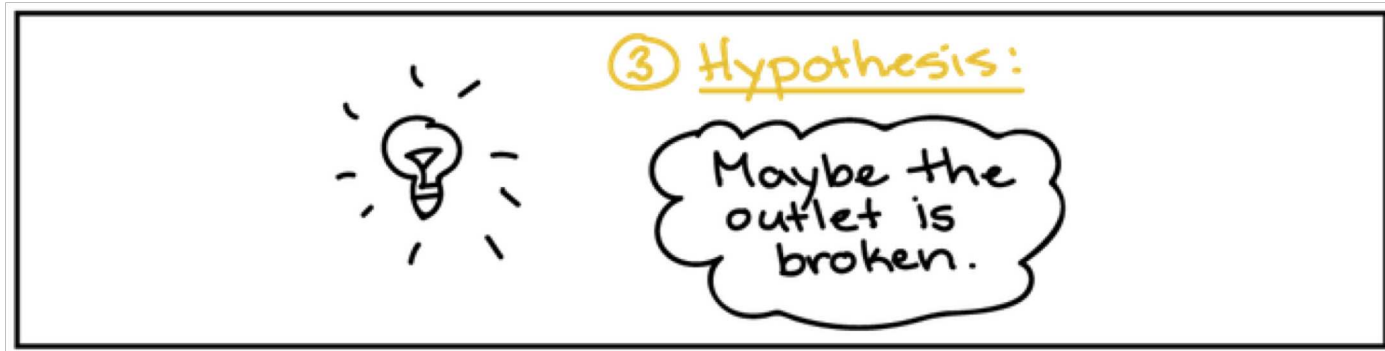
Test the Prediction Outcomes

- If the toaster does toast, then the hypothesis is supported—likely correct
- If the toaster doesn't toast, then the hypothesis is not supported—likely wrong

Supporting or Contradicting a Hypothesis

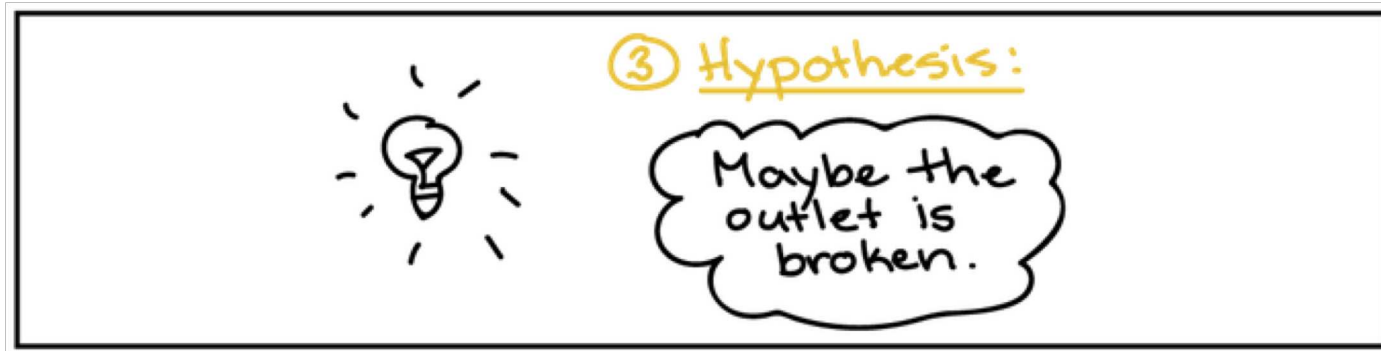
- **Results that support** a hypothesis can't conclusively prove that it's correct, but **they do mean it's likely to be correct**
- If **results contradict a hypothesis**, that hypothesis is probably not correct. Unless there was a flaw in the test—a possibility we should always consider—a contradictory result means that **we can discard the hypothesis and look for a new one**

What can we say about the Hypothesis...



	First outlet working?	Second outlet working?
Toaster does work		
Toaster doesn't work		

What can we say about the Hypothesis...



	First outlet working?	Second outlet working?
Toaster does work	Probably no	Probably yes
Toaster doesn't work	Cannot tell	Cannot tell

Example: Fail to Toast

6. Iterate

The last step of the scientific method is to reflect on our results and use them to guide our next steps.

And the result is...



My bread toasts!

Hypothesis is
supported.



My bread still
won't toast.

Hypothesis is
not supported.

⑥ Iteration time!

But what is
actually wrong
with that outlet?

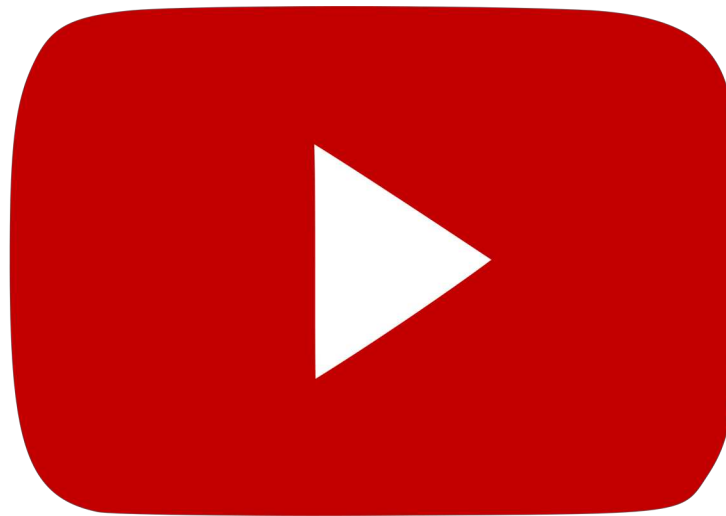
Hmm... maybe
there is a broken
wire in the toaster.

If toaster still doesn't work

- New hypotheses:
 - Both outlets are working and the toaster is broken
 - Both outlets are defective and the toaster is working
- New prediction:
 - If I plug another device in the outlets, then it will turn on.

And so on...

Video



Link to the video: <https://www.youtube.com/watch?v=N6IAzlugWw0>

Making Assumptions

- All scientific tests involve making assumptions

Example of a Hypothesis Test

- Suppose we want to test the hypothesis that **substance A stops bacterial growth**
- To test the hypothesis, we spread a mixture of substance A and some bacterial growth medium in a Petri dish, and spread a mixture of the inert substance B and some bacterial growth medium in another Petri dish
- We then wait 1 day to check the results

Source: http://undsci.berkeley.edu/article/0_0_0/howscienceworks_01



Even a fairly straightforward experiment will rely on some assumptions:

We assume that substance B does not affect growth.



We assume that bacteria will grow on the growth medium.



We assume that the pen used to mark the dishes does not affect growth.



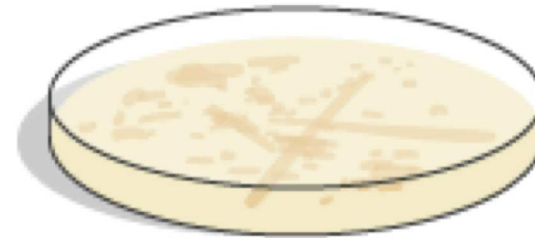
We assume that one day is long enough for colonies to grow.

Making Assumptions

- The previous example contained perfectly reasonable assumptions that **can be tested**
- All the assumptions **need to be justified** with
 - New experiments
 - Past tests performed by other scientists
- Some assumptions may remain untested because our knowledge in the field suggests that assumption is safe

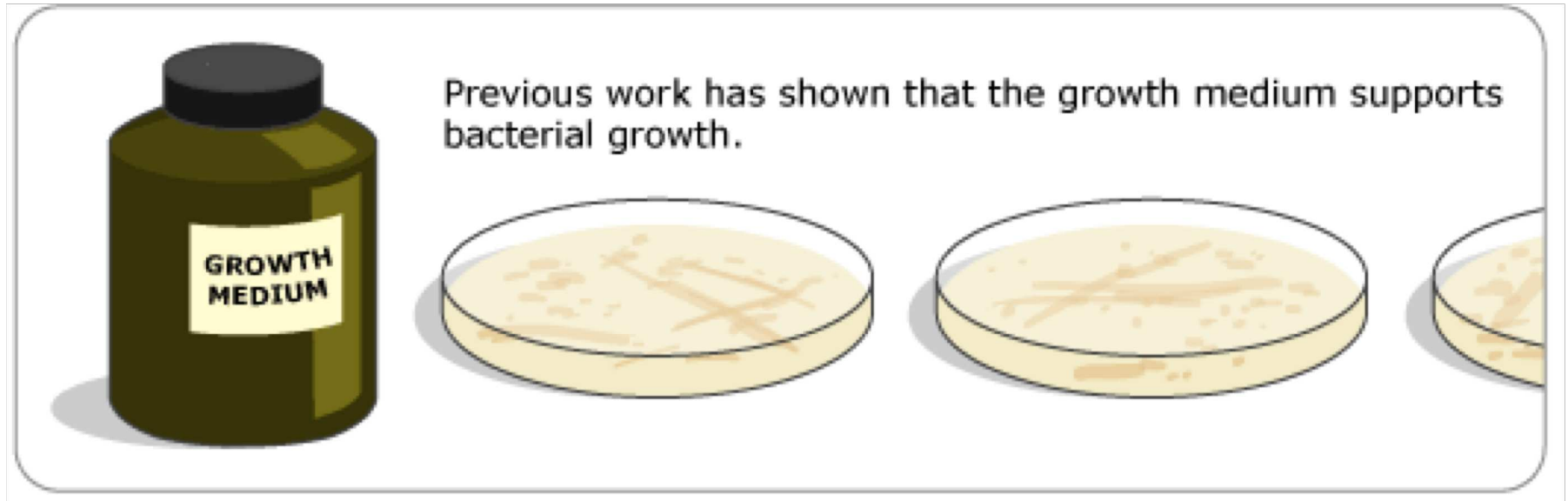
Check Assumptions

A separate experiment verifies that substance B is inert.



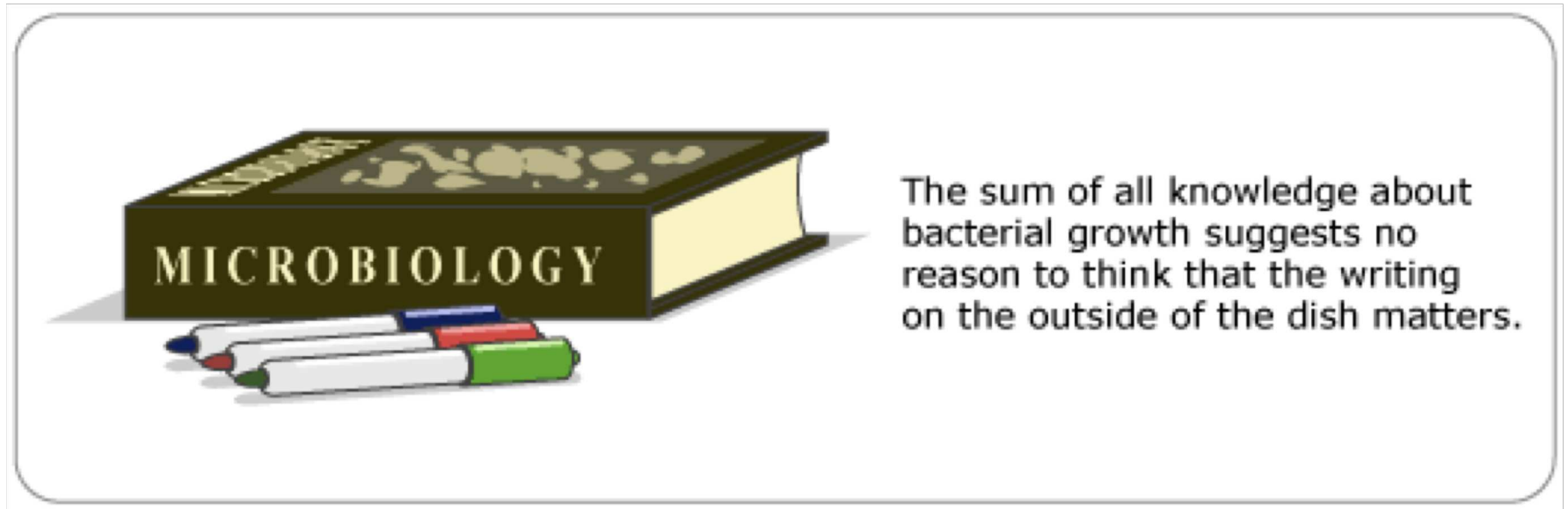
Source: http://undsci.berkeley.edu/article/0_0_0/howscienceworks_01

Check Assumptions



Source: http://undsci.berkeley.edu/article/0_0_0/howscienceworks_01

Check Assumptions



Source: http://undsci.berkeley.edu/article/0_0_0/howscienceworks_01

Observation beyond our eyes

- Making an observation does not mean we have to “see it with our own eyes” – in some cases, this may be impossible
- Several tools can be used to make observations: telescopes, microscopes, all kinds of sensors, radar, thermometers, etc.
- For Computer Science, this is almost always de case
- **Precision is a key** when making observations

Source: http://undsci.berkeley.edu/article/0_0_0/howscienceworks_01

Experiments

- An experiment is a test that involves **manipulating some factor** in a system in order to see how that affects the outcome
- They can be as simple as rolling a ball at different angles to check how that affects the speed of the ball, or as complex as removing a species from a controlled habitat to see how that affects other species

Exercise

- You have developed a new sort algorithm in your thesis
- Which hypothesis would you use?
- Which experiments would you design to test each of your hypothesis?

Hypothesis x Research Question

- Research questions are always written as questions
- Hypothesis are written as statements
- Example of research question: "What is the effect of heat on the effectiveness of bleach?"
- Example of hypothesis: "I predict heat will diminish the effectiveness of bleach."

Source: <https://sciencing.com/the-difference-between-research-questions-hypothesis-12749682.html>

Hypothesis x Research Question

- When using a research question, the researcher would think about **how to phrase the question** to ensure its scope is **not too broad, too narrow** or **impossible to answer**

Source: <https://sciencing.com/the-difference-between-research-questions-hypothesis-12749682.html>

Hypothesis x Research Question

- Both are used in Computer Science Research
- Both require experiments to prove/refute the hypothesis or answer the research questions

Source: <https://sciencing.com/the-difference-between-research-questions-hypothesis-12749682.html>

Types of Experiments

- ***in vivo*** (from the Latin *within the living*)
 - An experiment that is conducted on living organisms, usually in their own habitats
- ***in vitro*** (from the Latin *within the glass*)
 - An experiment that is conducted in a controlled environment, such as a laboratory
- ***in virtuo***
 - An experiment that is conducted in the form of a computer simulation, but allow agents to change the course of the simulation
- ***in silico***
 - An experiment where everything is simulated

Amigoni and Schiaffonati "Multiagent-Based Simulation in Biology". In: Model-Based Reasoning in Science, Technology and Medicine, 2007.

In the beginning



In-vivo



In-vitro

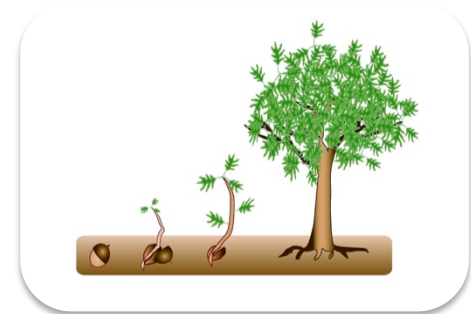
However in-vivo and in-vitro experiments are...



Costly



Risky



Slow

From in-vivo to in-silico



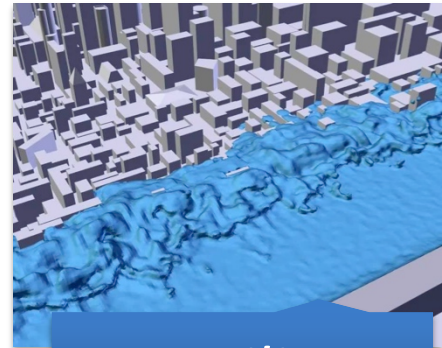
In-vivo



In-vitro



In-virtuo



In-silico



Travassos and Barros "Contributions of in virtuo and in silico experiments for the future of empirical studies in software engineering." WSESE 2003

Dealing with produced data



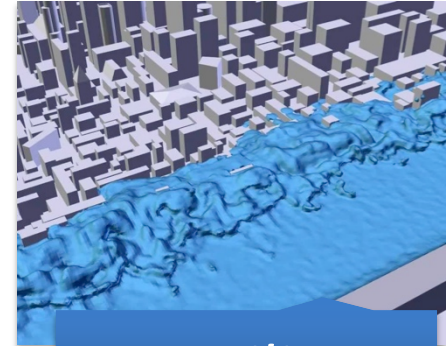
In-vivo



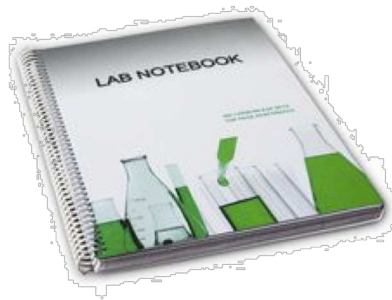
In-vitro



In-virtuo

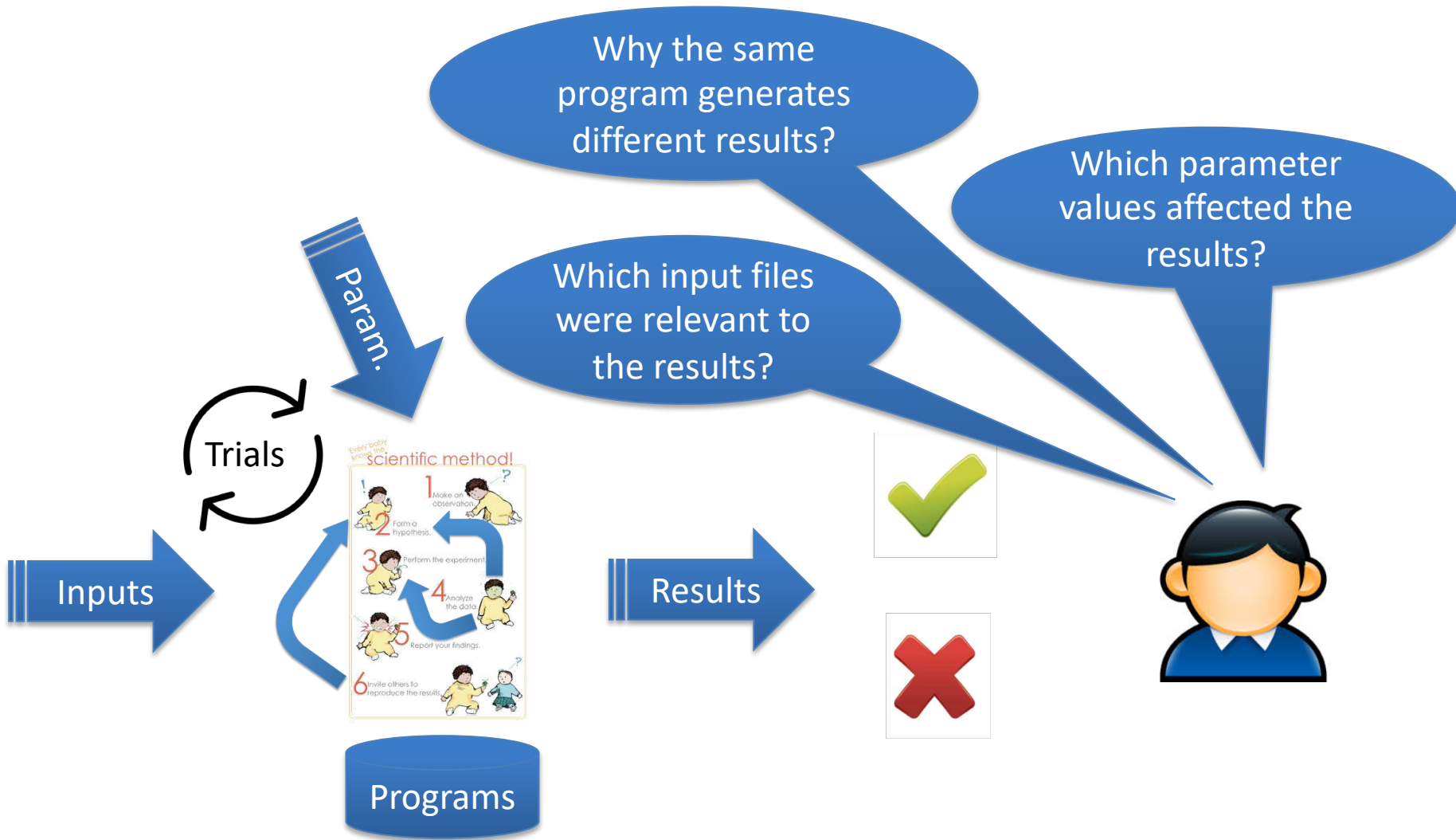


In-silico



Travassos and Barros "Contributions of in virtuo and in silico experiments for the future of empirical studies in software engineering." WSESE 2003

How to reason from the data?



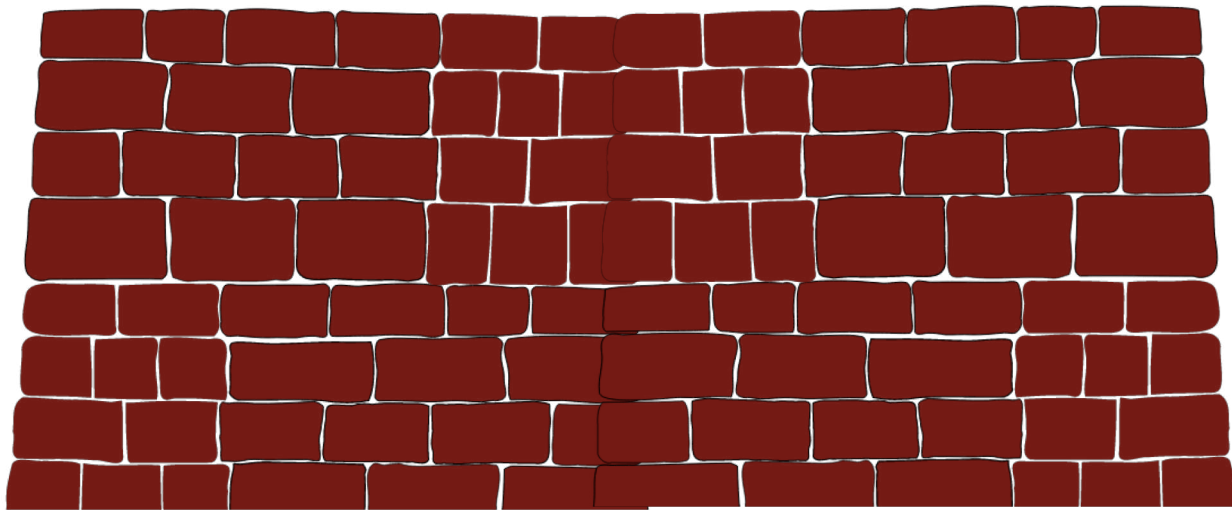
How to reason from the data?

Disciplina e-Science na pós

Programs

Building a Strong Wall of Knowledge

- Accepted scientific ideas are reliable because they have been **subjected to rigorous testing**, but as new evidence is acquired and new perspectives emerge these ideas can be revised



Publishing

- Publishing results is a way of **making your findings available** so that other researchers can build upon them
- Publications describe a study and report any details that one might need to evaluate that study (background information, data, statistical results, graphs, maps, explanations of how the study was performed and how the researchers drew their conclusions)

Publishing

- Publishing is also “quality control” of your research
- Publications go through a rigorous process of peer-review (journal or conference)

Publishing

- There is also the “non noble” reason for publishing...



No money



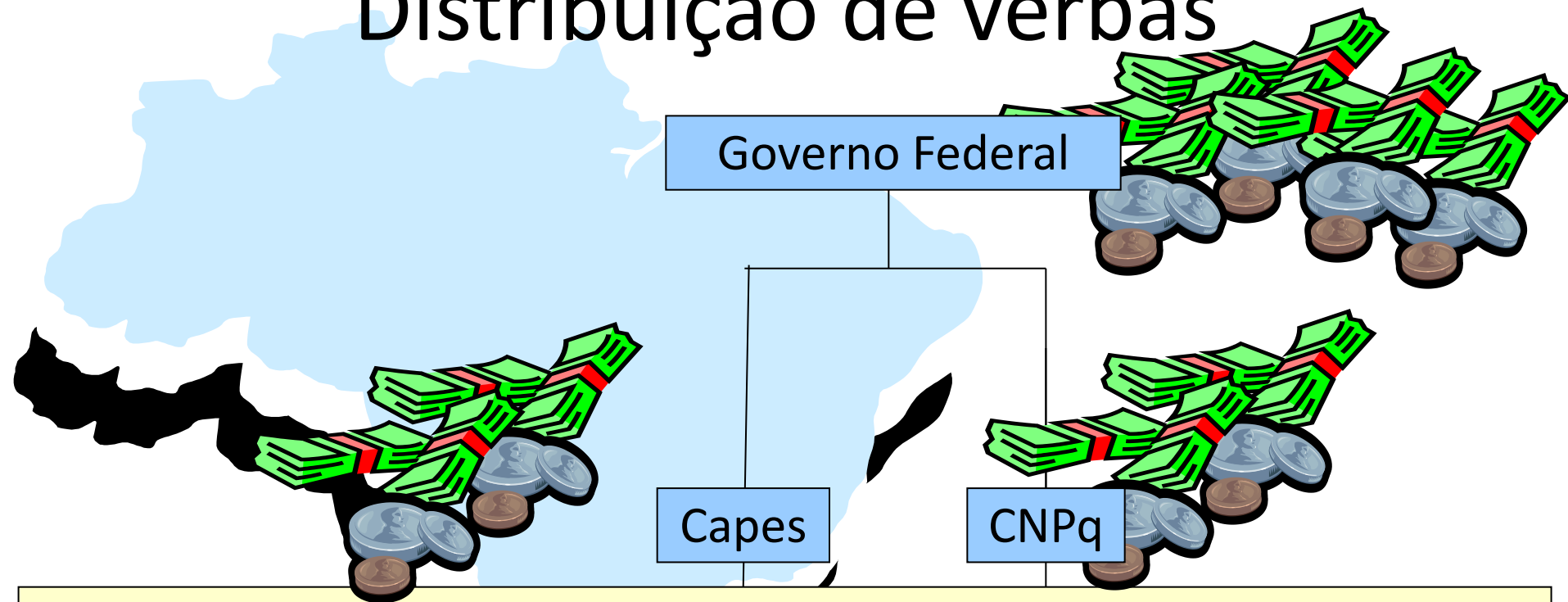
No
research

Dinheiro pra quê?

- Bolsas
- Professores
- Papel para a impressora
- Equipamentos para o laboratório
- Limpeza das salas
- Sabonete para o banheiro
- Livros para biblioteca
- etc...



Distribuição de verbas



Universidades Públicas e Privadas



Distribuição de verbas

Universidades Públicas e Privadas



- Não é uniforme...
 - Cada curso de Pós-Graduação recebe recursos de acordo com sua **nota** na **avaliação** da Capes

Avaliação da Capes

- CAPES avalia todos os cursos de Pós-Graduação do país
- Avaliação é feita a cada 4 anos
- Nota de 3 a 7

Pós-Graduação em Computação

MESTRADO E DOUTORADO

EXCELÊNCIA EM PESQUISA



CONCEITO 6 - 2017



Instituto de
computação

#POSGRAD-IC-UFF

#OrgulhoDeSerUFF

posgrad.ic.uff.br

Critérios de Avaliação



- Definidos em um documento chamado **Documento de Área**
- Documento de Área da **Computação** define **como** as **notas** para os cursos de Computação **são** **calculadas**



ACESSO À
INFORMAÇÃO

NOSSAS AÇÕES

Avaliação

Bolsas /
Estudantes

Educação a
Distância

Formação de
Professores da
Educação Básica

Tecnologia da
Informação

Prêmio Capes de
Tese

Prêmios

SOBRE A CAPES

CENTRAL DE
CONTEÚDOS



Imagens



Vídeos



Áudios



Publicações

Ciência da Computação

Publicado: Sexta, 24 Junho 2011 16:54 | Última Atualização: Quinta, 10 Maio 2018 12:42

Coordenador

Paulo Roberto Freire Cunha (UFPE)

Coordenador Adjunto

José Carlos Maldonado (USP)

Coordenador de Mestrado Profissional

Avelino Francisco Zorzo (PUC/RS)

Contato

E-mail: 02.comp@capes.gov.br

▼ Avaliação Quadrienal 2017

Disponibilizado na WEB	Nome do documento	Formatos disponíveis
20/12/2017	Relatório de Avaliação	PDF 2,5MB

▼ Documentos de área

Disponibilizado na WEB	Nome do documento	Formatos disponíveis
16/12/2016	Documento de área 2017	PDF 1280 KB

Critérios de Avaliação

- Uma parte considerável (**40%**) da nota vem da Produção Intelectual do Programa de Pós-Graduação (**PUBLICAÇÕES!!**)
- **Comitê de Computação da Capes** mede a qualidade através de vários critérios
 - **Congressos**: h5-index calculado com o auxílio do Google Metrics
 - **Periódicos**: h5-index calculado com o auxílio do Google Metrics

* h5-index de uma conferência ou periódico é o número **h** de artigos publicados nesse fórum nos últimos 5 anos que possuem pelo menos **h** citações

Qualis

- Congressos e periódicos recebem uma classificação
- Escala não é linear

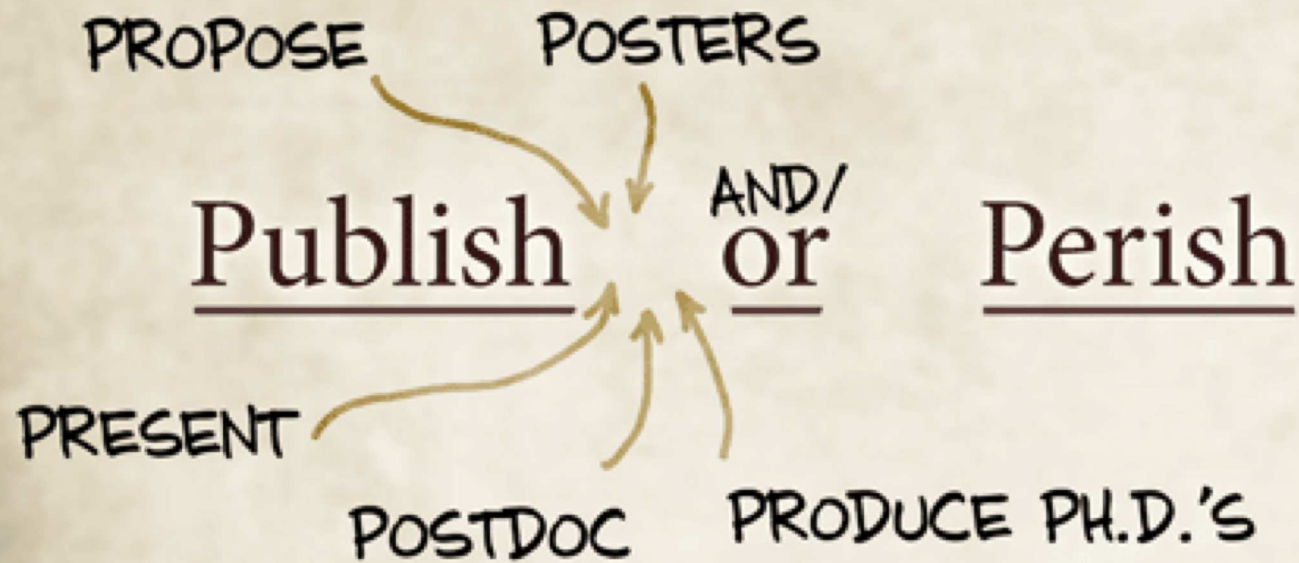
A1	A2	B1	B2	B3	B4	B5	C
1,00	0,85	0,70	0,50	0,20	0,10	0,05	0

- A Computação é a **única área** que tem **Qualis para Congressos**



Periódicos

ISSN	Título	Área de Avaliação	Classificação
0360-0300	ACM COMPUTING SURVEYS	CIÊNCIA DA COMPUTAÇÃO	A1
0730-0301	ACM TRANSACTIONS ON GRAPHICS	CIÊNCIA DA COMPUTAÇÃO	A1
0254-5330	ANNALS OF OPERATION RESEARCH	CIÊNCIA DA COMPUTAÇÃO	A1
1572-9338	ANNALS OF OPERATIONS RESEARCH (DORDRECHT. ONLINE)	CIÊNCIA DA COMPUTAÇÃO	A1
0096-3003	APPLIED MATHEMATICS AND COMPUTATION	CIÊNCIA DA COMPUTAÇÃO	A1
1568-4946	APPLIED SOFT COMPUTING (PRINT)	CIÊNCIA DA COMPUTAÇÃO	A1
0004-3702	ARTIFICIAL INTELLIGENCE (GENERAL ED.)	CIÊNCIA DA COMPUTAÇÃO	A1
0005-1098	AUTOMATICA (OXFORD)	CIÊNCIA DA COMPUTAÇÃO	A1
1367-4803	BIOINFORMATICS (OXFORD. PRINT)	CIÊNCIA DA COMPUTAÇÃO	A1
1471-2105	BMC BIOINFORMATICS	CIÊNCIA DA COMPUTAÇÃO	A1
1467-5463	BRIEFINGS IN BIOINFORMATICS	CIÊNCIA DA COMPUTAÇÃO	A1
1007-5704	COMMUNICATIONS IN NONLINEAR SCIENCE & NUMERICAL SIMULATION	CIÊNCIA DA COMPUTAÇÃO	A1
0001-0782	COMMUNICATIONS OF THE ACM	CIÊNCIA DA COMPUTAÇÃO	A1
0018-9162	COMPUTER (LONG BEACH, CALIF. PRINT)	CIÊNCIA DA COMPUTAÇÃO	A1
1467-8659	COMPUTER GRAPHICS FORUM (ONLINE)	CIÊNCIA DA COMPUTAÇÃO	A1
0167-7055	COMPUTER GRAPHICS FORUM (PRINT)	CIÊNCIA DA COMPUTAÇÃO	A1



JORGE CHAM © 2011

WWW.PHDCOMICS.COM

Qualis

- Deve ser usado para avaliar **cursos**, e **não pessoas**
- Consulte sempre seu **orientador** para saber onde publicar
 - Ele conhece a área e sabe quais são os melhores fóruns

FLUXO DE PUBLICAÇÃO: CONGRESSO X PERIÓDICO

Congresso



Congresso



**ANTES DO
DEADLINE**

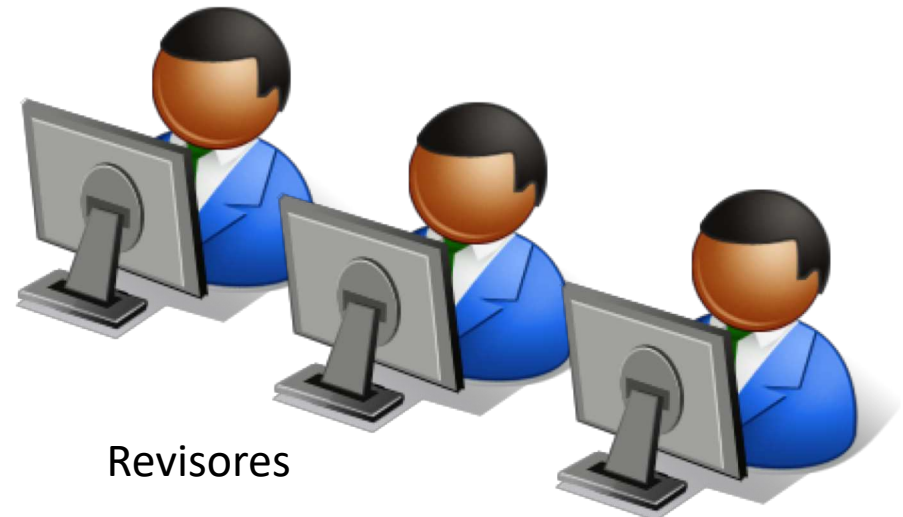
Congresso



Congresso



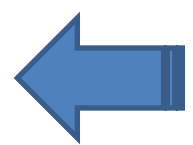
3 Avaliações



Congresso



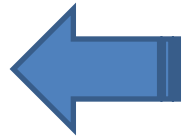
Resultado



Congresso



Resultado



ACEITO
RECUSADO
REBUTTAL*

E quais são as opções disponíveis?

- Como ficar sabendo quais congressos estão com datas de submissão abertas?
- Chamadas de Trabalhos
 - **áreas de interesse** do evento
 - **formato** que o artigo deve ter
 - **datas limite de submissão**
 - composição do **comitê de programa**
 - etc

Listas de email

- Lista da SBC
 - É uma lista geral para onde são enviadas chamadas de trabalhos de congressos de diversas áreas de pesquisa



Listas de email

- Listas internacionais específicas de cada área

DBWORLD
(Banco de Dados)



SEWORLD
(Eng. Software)



IEEE_VIS
(Visualização)



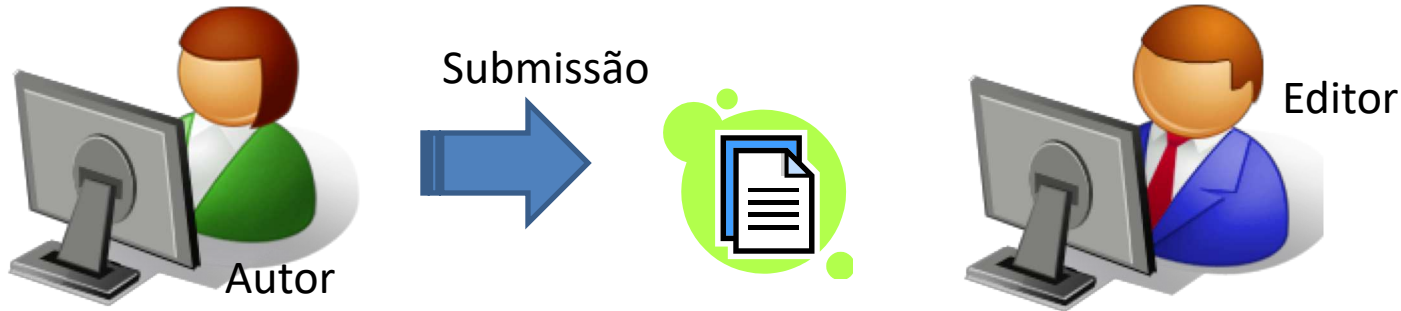
(**pergunte para seu orientador** qual é a lista específica da sua área de pesquisa)

Periódico



Corpo de Revisores

Periódico



**A QUALQUER
TEMPO
(FLUXO CONTÍNUO)**

Periódico



Periódico



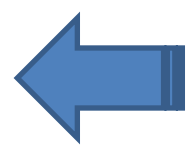
3 Avaliações



Periódico



Resultado



Periódico

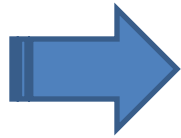


ACEITO
RECUSADO
MAJOR REVIEW
MINOR REVIEW

Periódico



Submissão Revisada +
Carta de Resposta



**DENTRO DA DATA
ESTIPULADA PELO
EDITOR**



CARTA DE RESPOSTA AOS REVISORES

Objetivo

- Responder os comentários dos revisores
 - Frisar o que foi feito para atender cada comentário

META:
**Evitar que o revisor leia
o artigo todo novamente**

Response to Reviewers

Dear editors and reviewers

We would like to thank you for your valuable suggestions. We have taken them into account to improve the paper in several ways. In the following, we explain the actions we took to address each of your comments.

- For better readability and clarity, the organization of the paper should be changed to break long, wordy sections into shorter subsections.

We have followed your suggestion and divided Section 3, which now is Section 4 into several subparagraphs that act as subsections. Section 4.1 now has two subparagraphs, and section 4.2 has four subparagraphs.

- In general, the paper lacks formalism. Important concepts and terms (e.g., an XML fragment, horizontal XML fragmentation, vertical XML fragmentation, correctness of a given XML fragmentation, etc.) should be more formally defined. Formal definitions can then be illustrated with simple examples (e.g., a running example throughout the paper would be nice).

We have created a new section (Section 3) that provides all of these formal definitions. In this section, we formally define XML documents, types, path expressions, XML fragments, horizontal fragments, vertical fragments, hybrid fragments and correctness rules. All of these definitions are illustrated by examples.

Response to Reviewers

Dear editors and reviewers

We would like to **thank you for your valuable suggestions**. We have taken them into account to improve the paper in several ways. In the following, we explain the actions we took to address each of your comments.

Introdução agradecendo

- For better readability and clarity, the organization of the paper should be changed to break long, wordy sections into shorter subsections.

We have followed your suggestion and divided Section 3 (which now is Section 4) into several subparagraphs that act as subsections. Section 4.1 now has two subparagraphs, and section 4.2 has four subparagraphs.

Repita o que o revisor disse. Explique o que foi feito para atender cada comentário

- The paper compares fragmentation in relational databases and XML databases in various parts of the paper. I have found this a bit confusing. Is it really necessary? The paper could instead focus directly on XML fragmentation. The work in relational data fragmentation is probably more mature and well established. So, maybe the contrast between the two could be brought up when the authors are discussing the open research challenges in XML fragmentation.

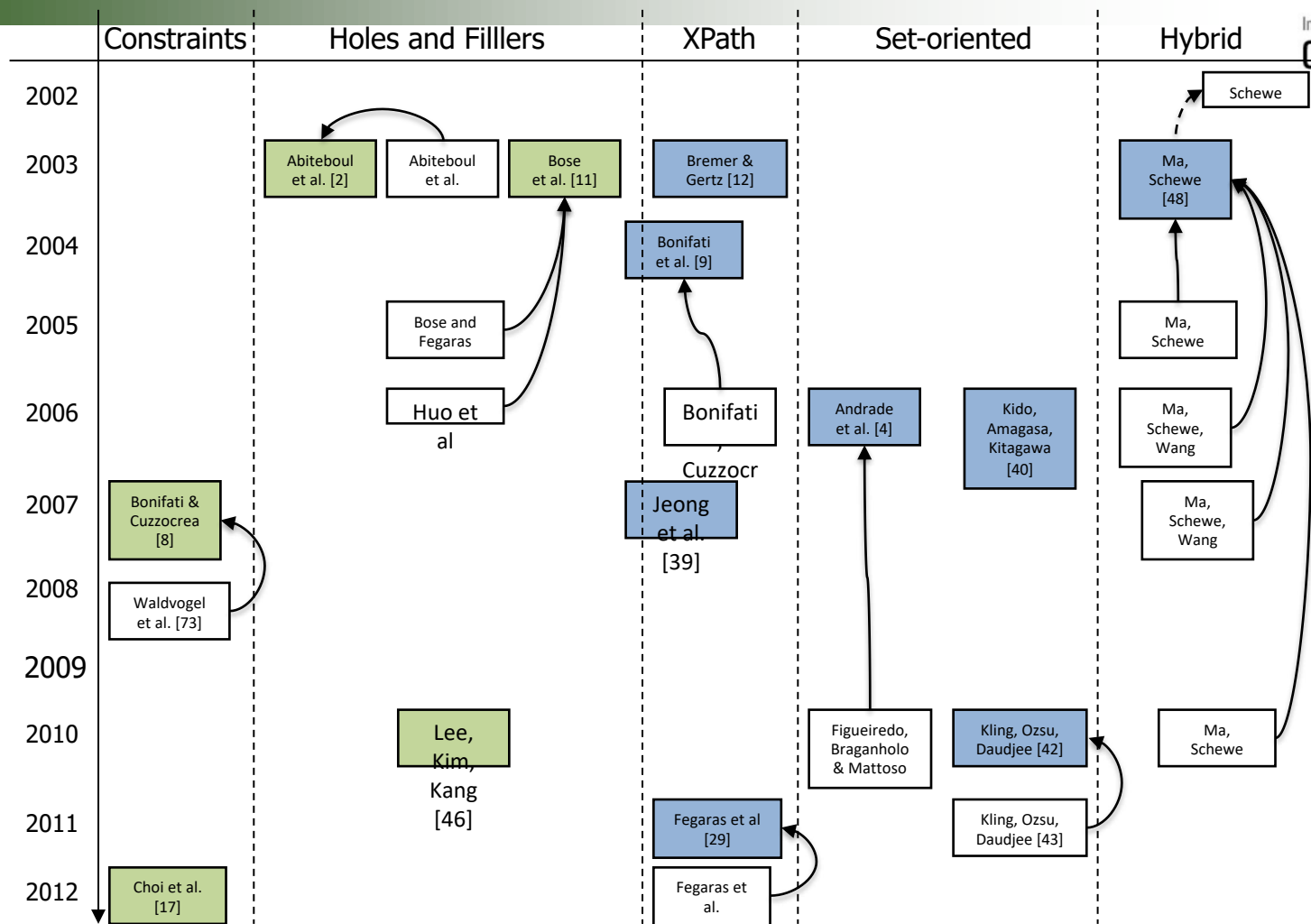
Because it is well established, **we believe that keeping the analogy with the relational model is beneficial.** We are not comparing but rather using it as a reference model. It helps the readers to establish a parallel between the concepts in the two models (from relational to XML). Because of this, we kept Section 2 as an informal grasp into XML fragmentation, and then in Section 3 we introduce formal definitions for XML fragmentation.

Discorde quando achar apropriado, mas explique os motivos

In section 5 (discussion), a timeline shows the appearance of the different approaches. However, I think that what is important here is to understand the evolution of the approaches by analyzing how each approach has improved the previous one. Therefore, instead of just mentioning the works one could mention the involved features within the timeline. I also understand that what I am suggesting is not easy to implement in a latex table.

In fact, **this is a great suggestion, we have followed it and we really think it improved the paper a lot.** We have changed the figure to include lanes that group approaches that follow the same principles. Besides, we added arrows from derived approaches to the approach that originated them. Features are discussed on Tables 1 and 2. The new figure is shown below.

Elogie as boas sugestões



Sempre que possível, coloque o texto ou figura alterados diretamente na carta de resposta aos revisores

Page 4: Please clarify what is "relate all types of generated provenance data." This reads a bit vague.

Answer: We have changed that sentence as follows:

"However, to analyze the complete provenance trail, scientists need to connect all types of generated provenance data (prospective, explicit retrospective, and implicit retrospective provenance)."

Sempre que possível, coloque o texto alterado diretamente na carta de resposta aos revisores

Response to Reviewers

Como não fazer...

The authors would like to thank the reviewers for their suggestions, which we have addressed as follows.

REVIEWER 1

C1: Summary of the article

C2: Justified in the introduction

C3: Explained in Section 3

C4: Future work

Response to Reviewers

Como não fazer...

The authors would like to thank the reviewers for their suggestions, which we have addressed as follows.

**Avaliador de mal humor em
3, 2, 1...**

REVIEWER

C1: Summary

C2: Justified in the introduction

C3: Explained in Section 3

C4: Future work

REDAÇÃO DE ARTIGOS

Formato

- Definido pelo veículo para o qual o artigo será submetido
- Ver regras no CALL FOR PAPERS
 - No site do evento
 - No email de call for papers enviado para alguma lista de emails

JIDM PAPERS

Submissions in this category should present new and exciting research results in any of the SBBD topics of interest. JIDM papers should describe complete research results. They must not have been simultaneously submitted to any other forum (conference or journal), nor should they have already been published elsewhere.

They must be written in English and formatted according to the JIDM template. The templates for submission are available at <http://bit.ly/1e77l6W>. Papers must be submitted directly to the journal editor using the SEER submission site (<https://seer.lcc.ufmg.br/index.php/jidm>). Submissions will be judged by the JIDM editorial board. For this category, submissions are made continuously throughout the year. Authors of papers that are accepted until August 7th will be invited to present their paper during SBBD 2017 and the paper will be published in the Journal of Information and Data Management (JIDM).

FULL PAPERS

Submissions in this category should present research results and discussion in any of the SBBD topics of interest. They are mainly targeted to the Brazilian DB community for research diffusion. Full papers submitted to SBBD must not have been simultaneously submitted to any other forum (conference or journal), nor should they have already been published elsewhere. The acceptance of a paper implies that at least one of its authors will register for the symposium to present it. Submitted papers will be reviewed based on originality, relevance, technical soundness and clarity of presentation.

Full papers must be written in Portuguese or English and not exceed 12 pages, according to the SBC template. (...)

Uma Questão de Estilo...

- Congressos e periódicos fornecem estilos prontos para serem usados na redação do artigo
- **Não tentem fazer tudo do zero**
- Baixem e usem o **template!**

SBC



ACM



IEEE

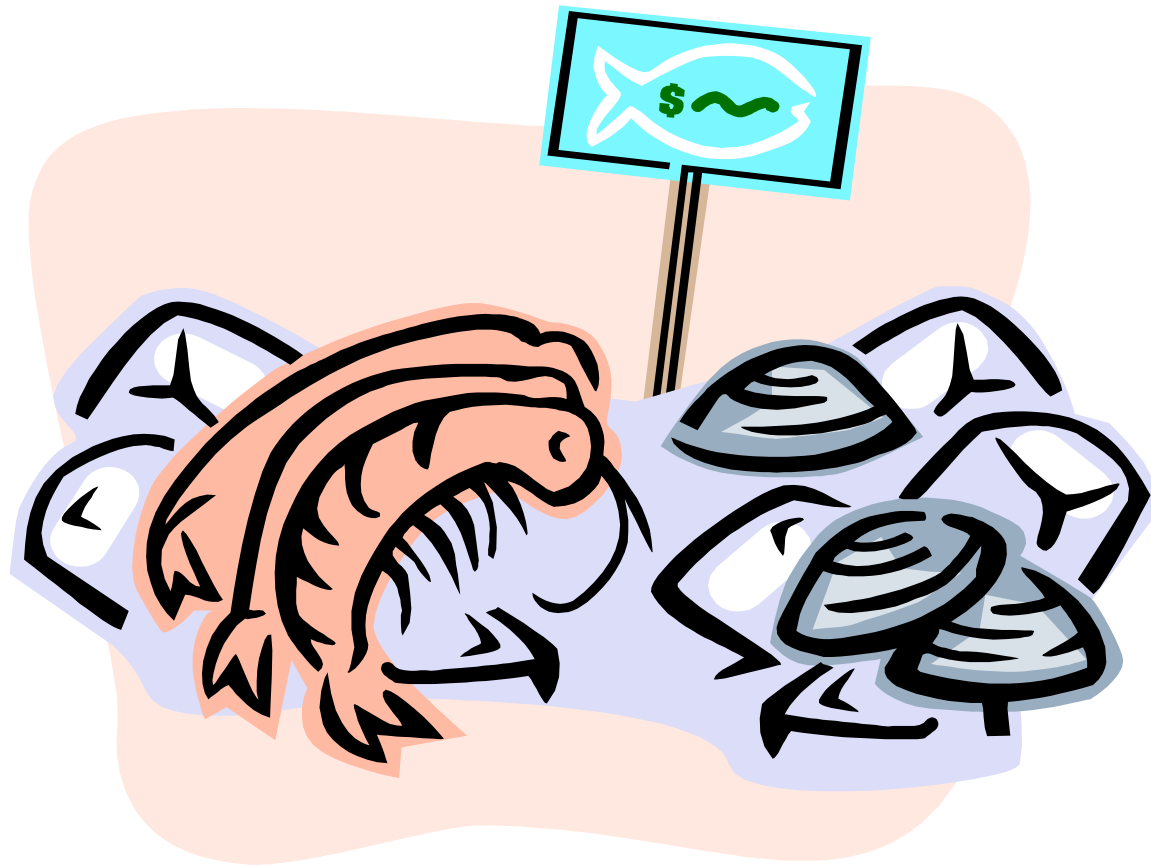


LNCS



Ok, já sei o formato...
Mas e o conteúdo do artigo?

Essencial (!!!)



Passos para a escrita de artigo

CONTEXTO

Passos para a escrita de artigo

PROBLEMA

CONTEXTO

Passos para a escrita de artigo

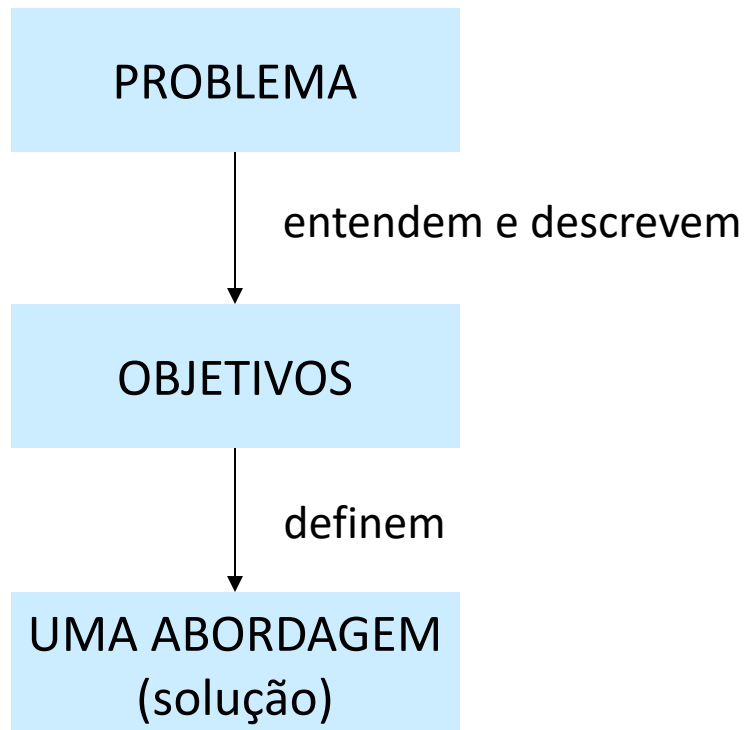
PROBLEMA

entendem e descrevem

OBJETIVOS

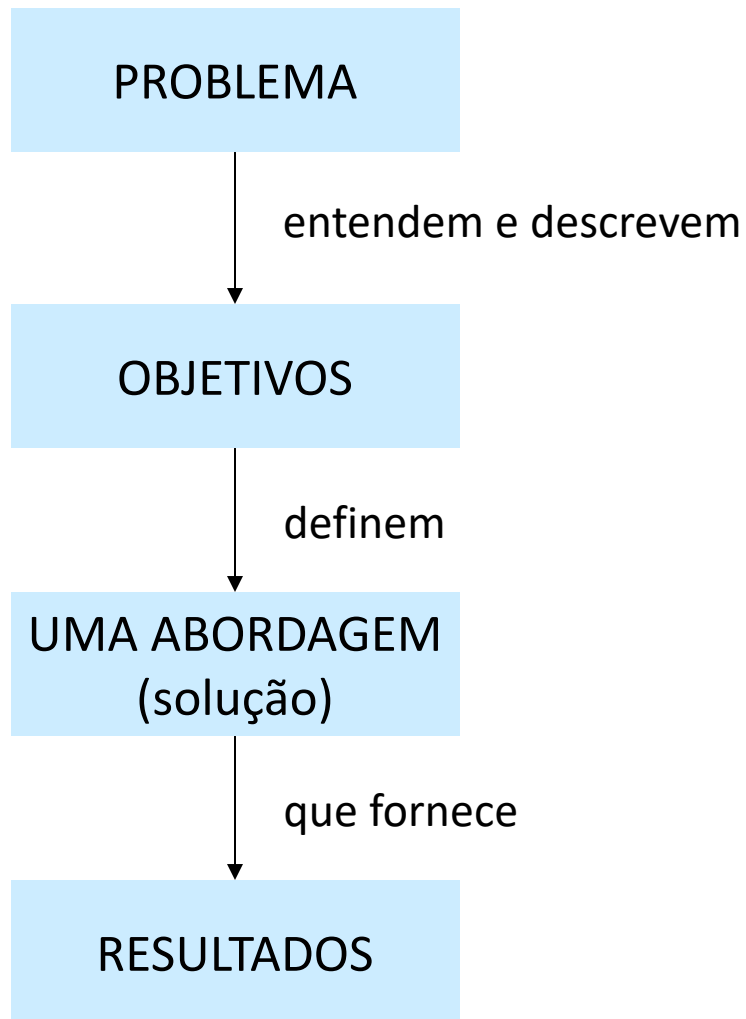
CONTEXTO

Passos para a escrita de artigo



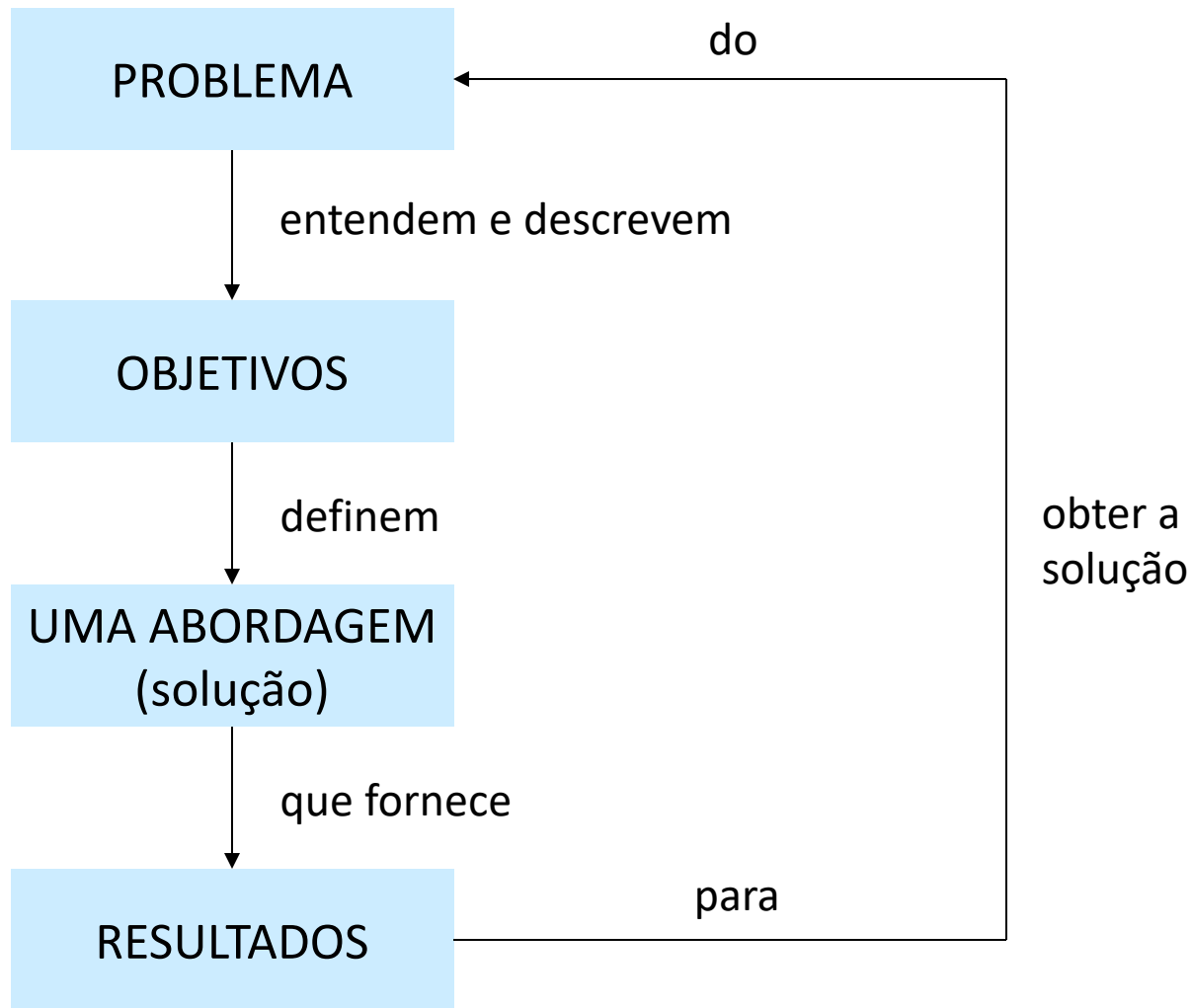
CONTEXTO

Passos para a escrita de artigo

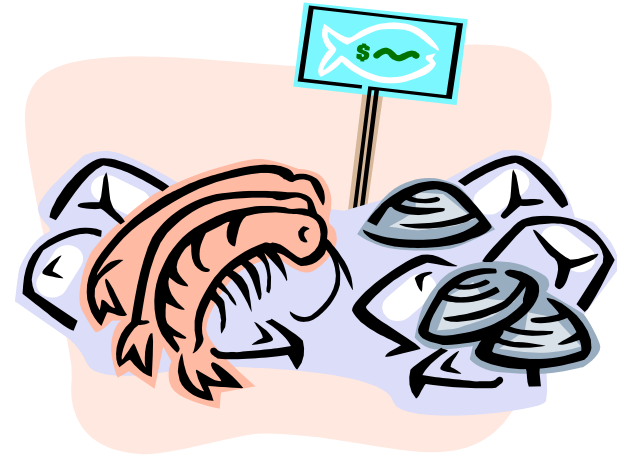
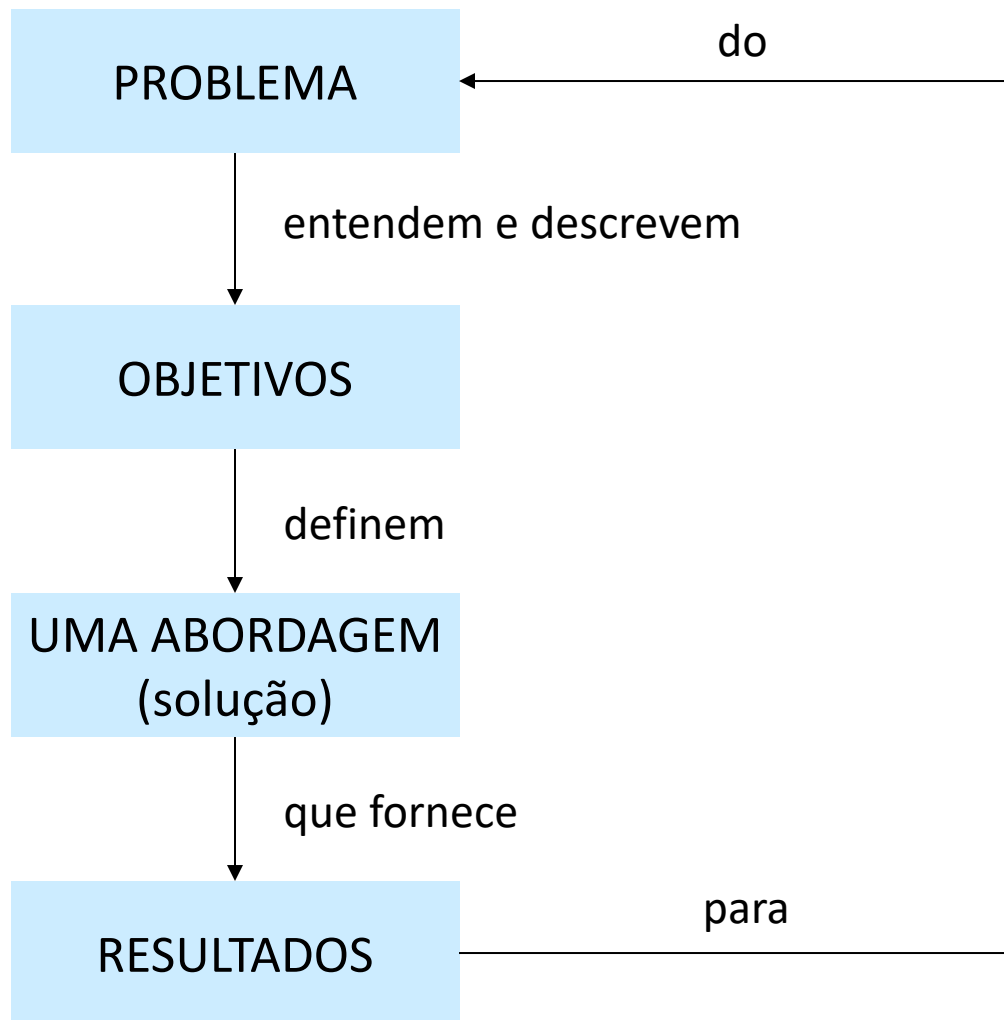


CONTEXTO

Passos para a escrita de artigo



Passos para a escrita de artigo



obter a
solução

CONTEXTO

Onde Vender o Peixe?

Título

Resumo

Introdução

Conclusão

Onde Vender o Peixe?

Título

Resumo

Introdução

Conclusão

Título

- Menor resumo de um artigo (by Leonardo Murta)
- TEM que **vender** o trabalho

BIRCH: An Efficient Data Clustering Method for Very Large Databases

Of snowstorms and Bushy trees

Zhang, Ramakrishnan, Livny. **BIRCH: An Efficient Data Clustering Method for Very Large Databases**. SIGMOD 1996: 103-114

Rafi Ahmed, Rajkumar Sen, Meikel Poess, Sunil Chakkappen. **Of Snowstorms and Bushy Trees**. PVLDB 7(13), 2014: 1452-1461

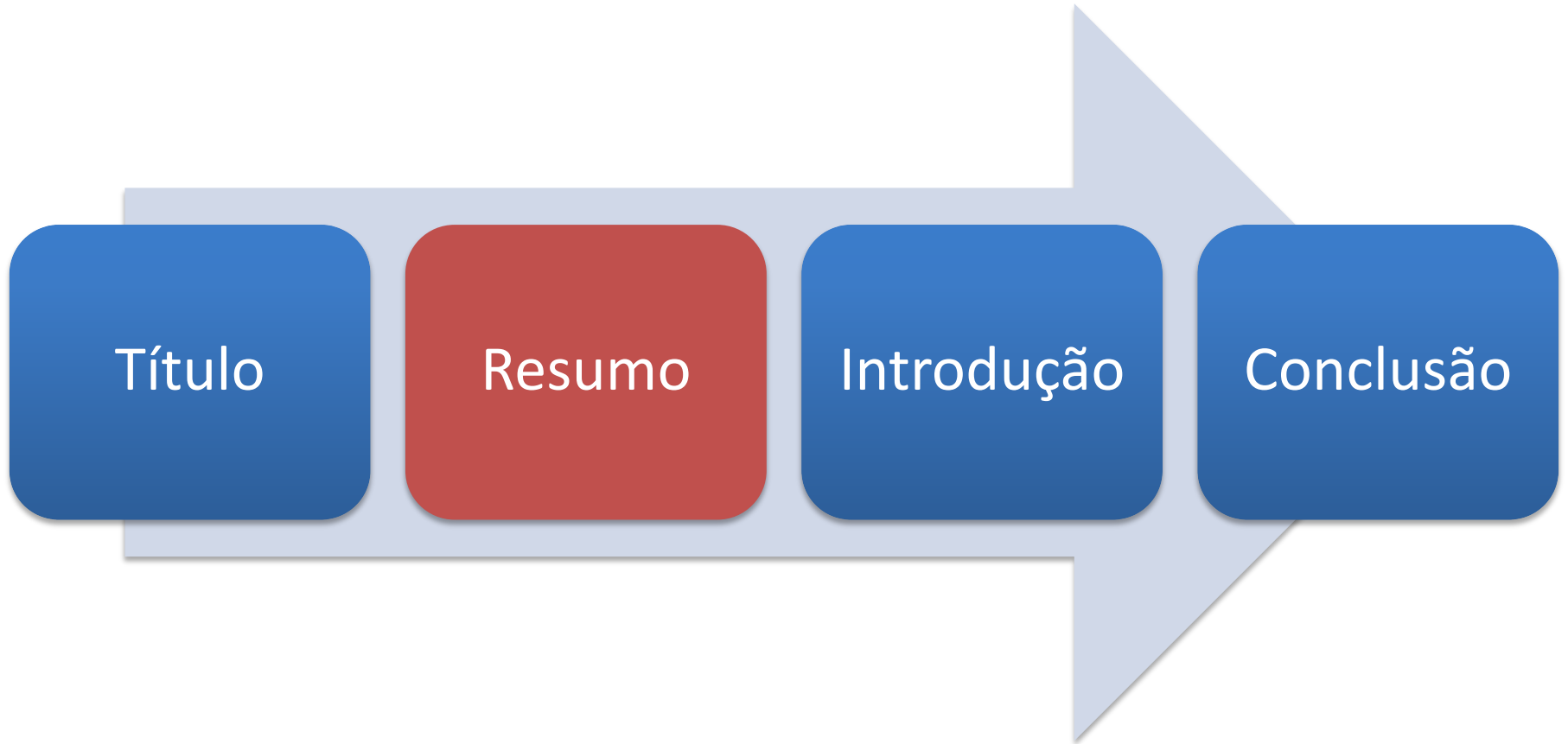
BIRCH: An Efficient Data Clustering Method for Very Large Databases

Of snowstorms and Bushy trees

Zhang, Ramakrishnan, Livny. **BIRCH: An Efficient Data Clustering Method for Very Large Databases**. SIGMOD 1996: 103-114

Rafi Ahmed, Rajkumar Sen, Meikel Poess, Sunil Chakkappen. **Of Snowstorms and Bushy Trees**. PVLDB 7(13), 2014: 1452-1461

Onde Vender o Peixe?



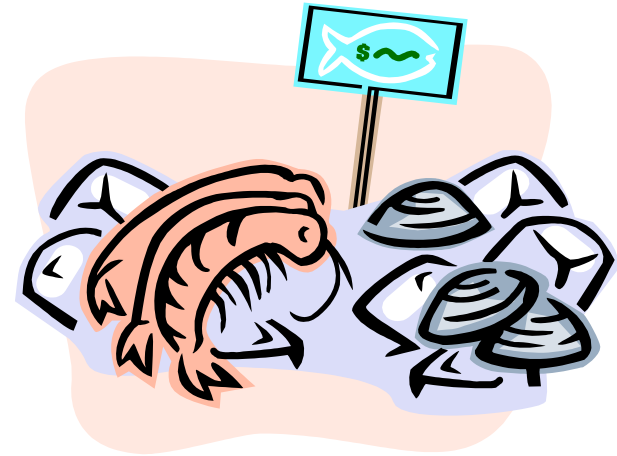
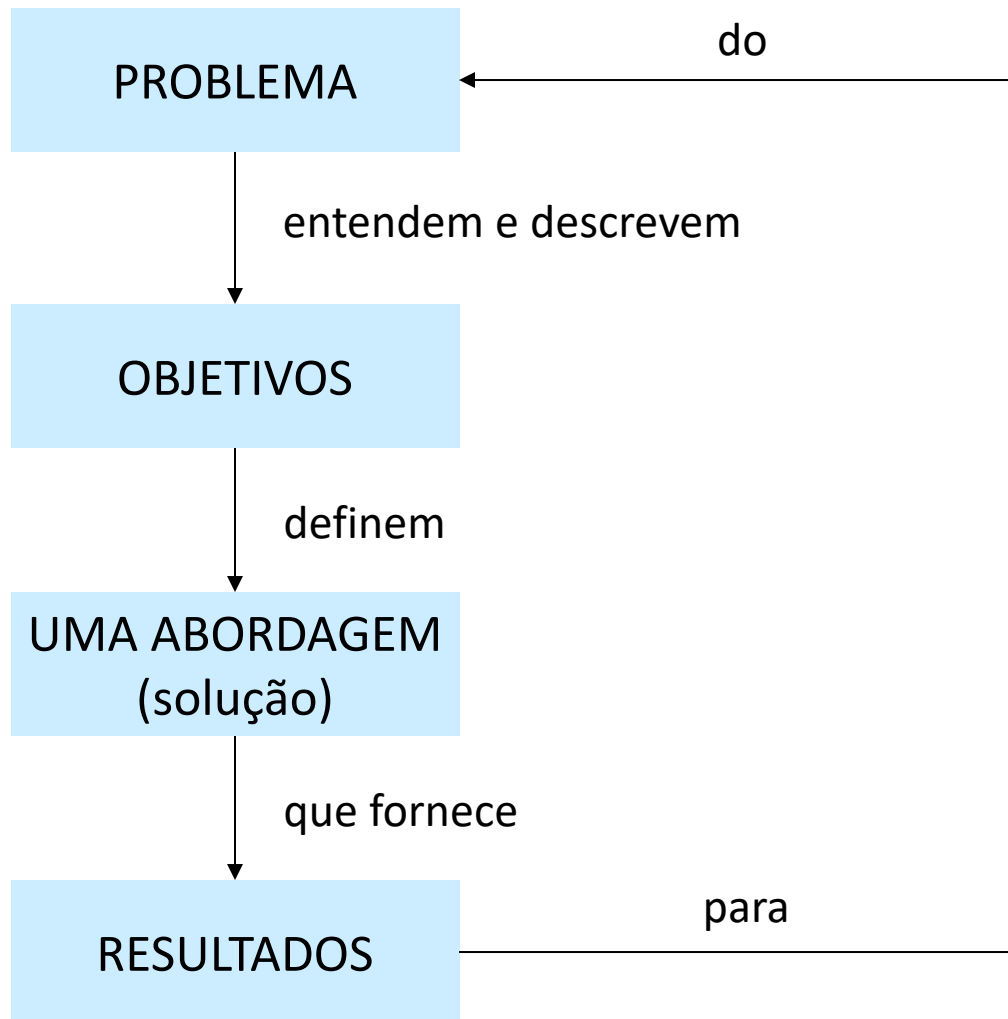
Resumo

- Um único parágrafo (50 a 200 palavras)
- Função: dizer aos leitores se vale a pena ler o restante do artigo
- Atenha-se ao essencial

Resumo NÃO deve ter...

- Fórmulas matemáticas
- Descrição da organização do texto
- Acrônimos
- Abreviaturas
- Referências (só em casos muito específicos)

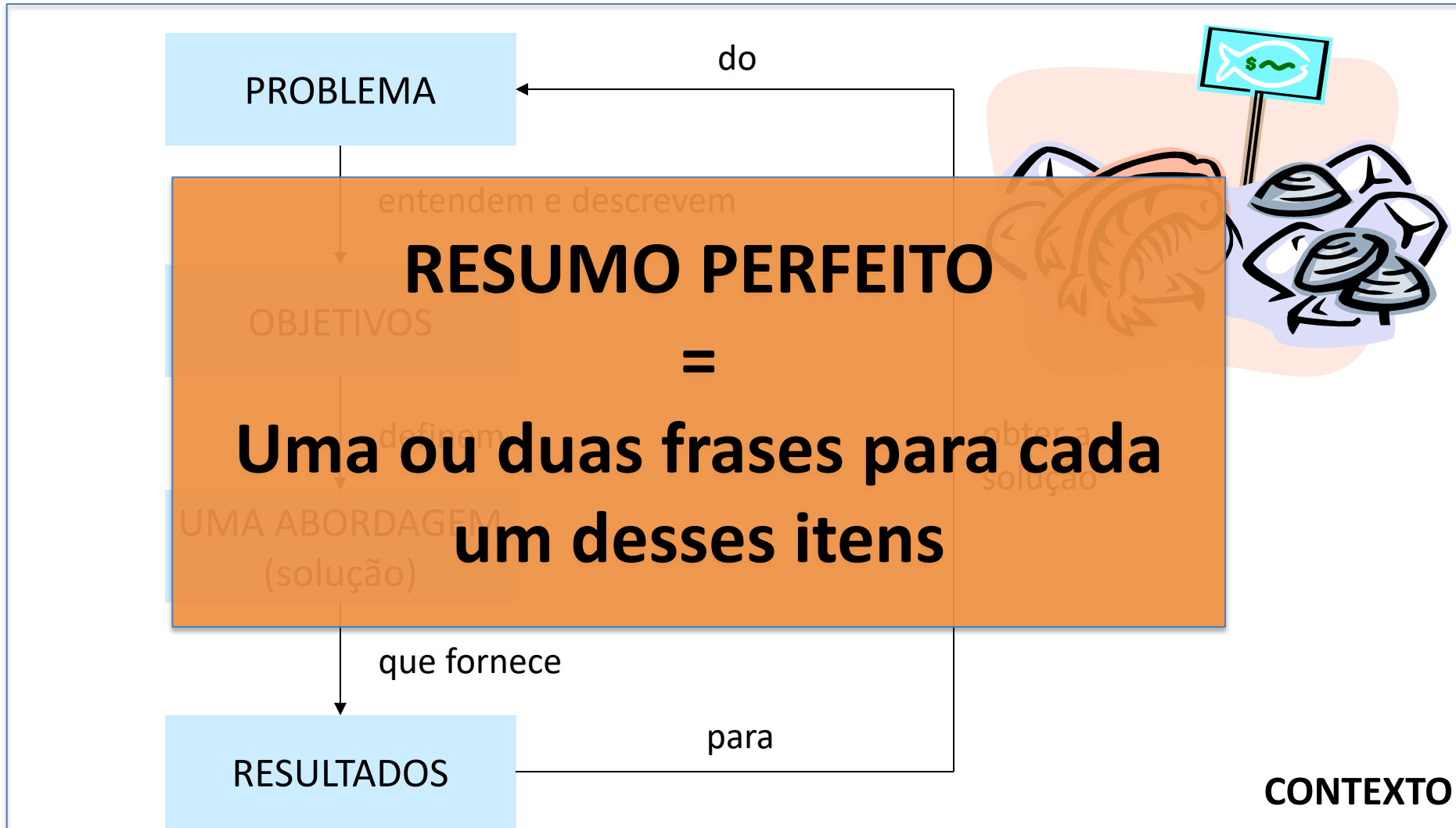
Resumo



obter a
solução

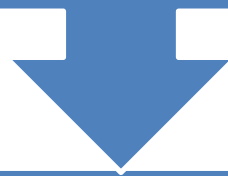
CONTEXTO

Resumo



Entregue o ouro...

Espaço necessário em memória pode ser reduzido significativamente



Espaço necessário em memória pode ser reduzido em 60%

Entregue o ouro...

Nós definimos um novo algoritmo de
inversão



Nós definimos um novo algoritmo de
inversão baseado em listas “mover
para frente”

Finding useful patterns in large datasets has attracted considerable interest recently, and one of the most widely studied problems in this area is the identification of clusters, or densely populated regions, in a multi-dimension dataset. Prior work does not adequately address the problem of large datasets and minimization of I/O costs. This paper presents a data clustering method named BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies), and demonstrates that it is especially suitable for very large databases. BIRCH incrementally and dynamically clusters incoming multi-dimensional metric data points to try to produce the best quality clustering with the available resources (i.e., available memory and time constraints). BIRCH can typically find a good clustering with a single scan of the data, and improve the quality further with a few additional scans. BIRCH is also the first clustering algorithm proposed in the database area to handle “noise” (data points that are not part of the underlying pattern) effectively. We evaluate BIRCH’s time/space efficiency, data input order sensitivity, and clustering quality through several experiments. We also present a performance comparison of BIRCH versus CLARANS, a clustering method proposed recently for large datasets, and show that BIRCH is consistently superior.

Contexto

Finding useful patterns in large datasets has attracted considerable interest recently, **and one of the most widely studied problems in this area is the identification of clusters, or densely populated regions, in a multi-dimension dataset. Prior work does not adequately address the problem of large datasets and minimization of I/O costs.** This paper presents a data clustering method named BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies), and demonstrates that it is especially suitable for very large databases. BIRCH incrementally and dynamically clusters incoming multi-dimensional metric data points to try to produce the best quality clustering with the available resources (time and space constraints). BIRCH can typically find a good clustering with a single scan of the data, and improve the quality further with a few additional scans. BIRCH is also the first clustering algorithm proposed in the database area to handle “noise” (data points that are not part of the underlying pattern) effectively. We evaluate BIRCH’s time/space efficiency, data input order sensitivity, and clustering quality through several experiments. We also present a performance comparison of BIRCH versus CLARANS, a clustering method proposed recently for large datasets, and show that BIRCH is consistently superior.

Problema (note posição em relação a trabalhos relacionados)

Finding useful patterns in large datasets has attracted considerable interest recently, and one of the most widely studied problems in this area is the identification of clusters, or densely populated regions, in a multi-dimension dataset. Prior work does not adequately address the problem of large datasets and minimization of I/O costs. **This paper presents a data clustering method named BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies), and demonstrates that it is especially suitable for very large databases.**

BIRCH incrementally and dynamically clusters incoming multi-dimensional metric data points to try to produce the best quality clustering with the available resources (i.e., available memory and time constraints). BIRCH can typically find a good clustering of the data, and improve the quality further with a few additional scans. BIRCH is also the first clustering algorithm proposed in the database area to handle “noise” (data points that are not part of the underlying pattern) effectively. We evaluate BIRCH’s time/space efficiency, data input order sensitivity, and clustering quality through several experiments. We also present a performance comparison of BIRCH versus CLARANS, a clustering method proposed recently for large datasets, and show that BIRCH is consistently superior.

Objetivo

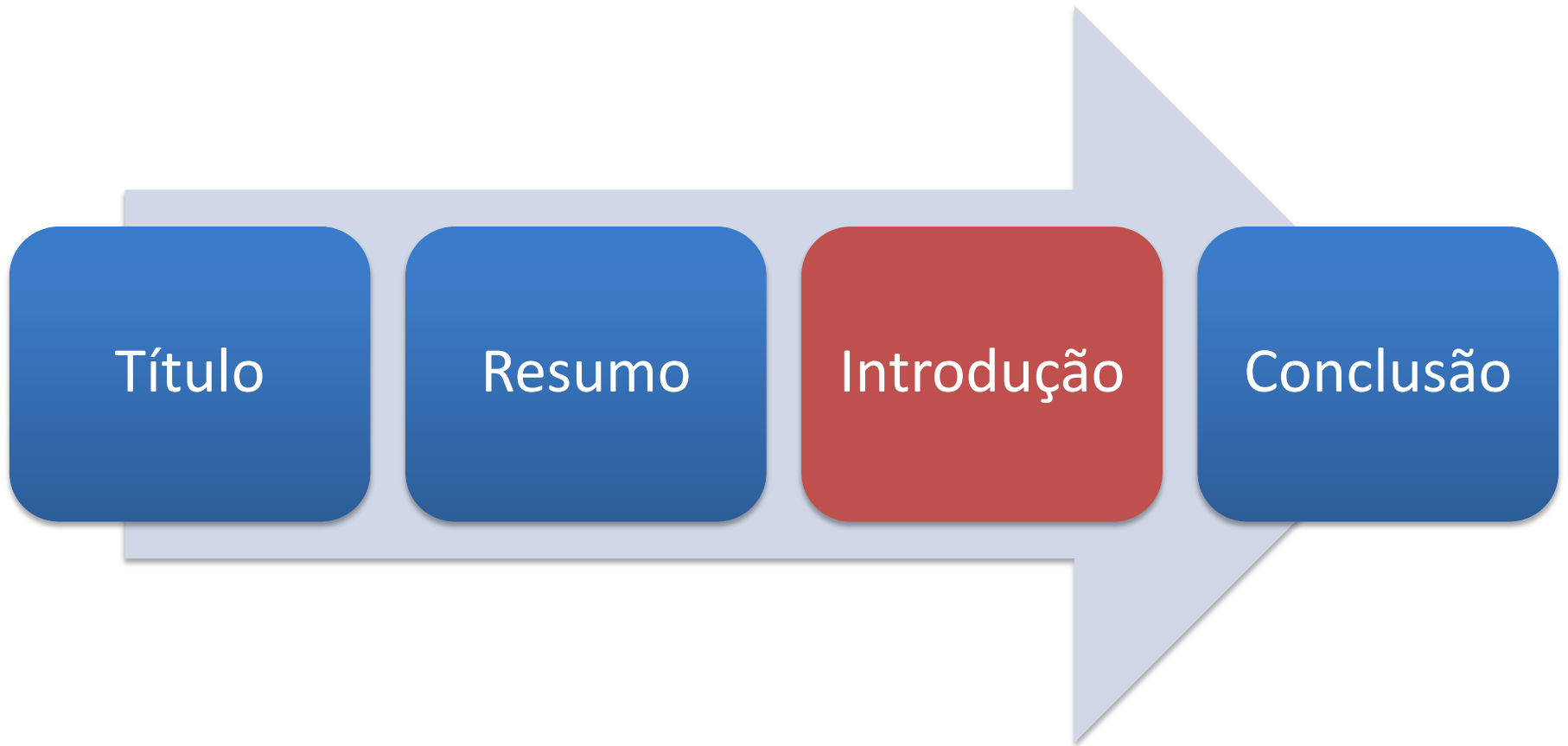
Abordagem

Finding useful patterns in large datasets has attracted considerable interest recently, and one of the most widely studied problems in this area is the identification of clusters, or densely populated regions, in a multi-dimensional dataset. Prior work does not address the problem of large datasets and minimization of I/O costs. This paper presents a data clustering method named BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies), and demonstrates that it is especially suitable for very large databases. **BIRCH incrementally and dynamically clusters incoming multi-dimensional metric data points to try to produce the best quality clustering with the available resources (i.e., available memory and time constraints). BIRCH can typically find a good clustering with a single scan of the data, and improve the quality further with a few additional scans. BIRCH is also the first clustering algorithm proposed in the database area to handle “noise” (data points that are not part of the underlying pattern) effectively.** We evaluate BIRCH’s time/space efficiency, data input order sensitivity, and clustering quality through several experiments. We also present a performance comparison of BIRCH versus CLARANS, a clustering method proposed recently for large datasets, and show that BIRCH is consistently superior.

Finding useful patterns in large datasets has attracted considerable interest recently, and one of the most widely studied problems in this area is the identification of clusters, or densely populated regions, in a multi-dimension dataset. Prior work does not adequately address the problem of large datasets and minimization of I/O costs. This paper presents a data clustering method named BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies), and demonstrates that it is especially suitable for very large databases. BIRCH incrementally and dynamically clusters incoming multi-dimensional metric data points to try to produce the best quality clustering with the available resources (i.e., available memory and time constraints). BIRCH can typically find a good clustering with a single scan of the data, and improve the quality further with a few additional scans. BIRCH is also the first clustering algorithm proposed in the database area to handle “noise” (data points that are not part of the underlying pattern) effectively. **We evaluate BIRCH’s time/space efficiency, data input order sensitivity, and clustering quality through several experiments. We also present a performance comparison of BIRCH versus CLARANS, a clustering method proposed recently for large datasets, and show that BIRCH is consistently superior.**

Resultados

Onde Vender o Peixe?



Introdução

- Mesma estrutura do resumo
- Transformar cada parte do resumo em **um parágrafo ou mais**
- Incluir um parágrafo no final com a **estrutura do artigo**

Dica

- Use um exemplo motivacional
 - Exemplos ajudam o leitor a entender o problema e a abordagem proposta pelo seu trabalho
 - Diga **explicitamente** qual seria o resultado da sua abordagem quando aplicado ao exemplo

Exemplo de uso de Figura para ilustrar resultados

(...)

Using the results of this paper, we will show that the view of figure 2(a) is updatable for all insertions, deletions and modifications. That is, there is a unique, side effect free translation from any update on this view to the underlying relations of figure 1. The view is produced by the following query:

VIEW 1

$$V_{YearLocation} = (Year, Location) (\pi (IdConf, ConfName, Year, Location) (Conferences \bowtie ConfLocation))$$

This query is an example of a class which we call *well-nested project-select-join queries*. Views of this class are always updatable.

IdConf	ConfName	YearLocation	
		Year	Location
DEXA	Conference on Database and Expert Systems Applications	2001	Munich, Germany
PODB	Symposium on Principles of Database Systems	2001	Santa Barbara, California
		2002	Madison, Wisconsin
VLDB	Conference on Very Large Data Bases	2002	Hong Kong, China
		2003	Berlin, Germany

(a)

IdConf	Details		
	Year	Location	Title
DEXA	2001	Munich, Germany	Querying the Web
VLDB	2002	Hong Kong, China	Databases and IP
	2002	Hong Kong, China	Web Survey

(b)

Figure 2: (a) View 1 (b) View 2

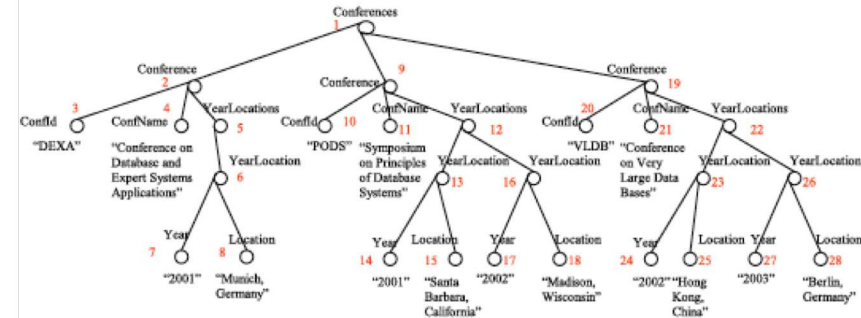


Figure 3: View 1 in XML

Introdução deve Responder

- Qual é o **problema** que o seu trabalho pretende resolver?
- Por que os **trabalhos existentes** não resolvem esse problema?
- **Como** o seu trabalho aborda o problema?
- Quais **resultados** experimentais seu trabalho obteve?
- Como seu trabalho está **organizado**?

Organização do Artigo

- Agora é a hora de planejar o restante da escrita!

Estrutura

1. Introdução

2. Background (*)

3. Trabalhos Relacionados

4. Abordagem

5. Avaliação (*)

6. Conclusão

Agradecimentos (*)

Referências

Estrutura

1. Introdução

2. Background (*)

3. Trabalhos Relacionados

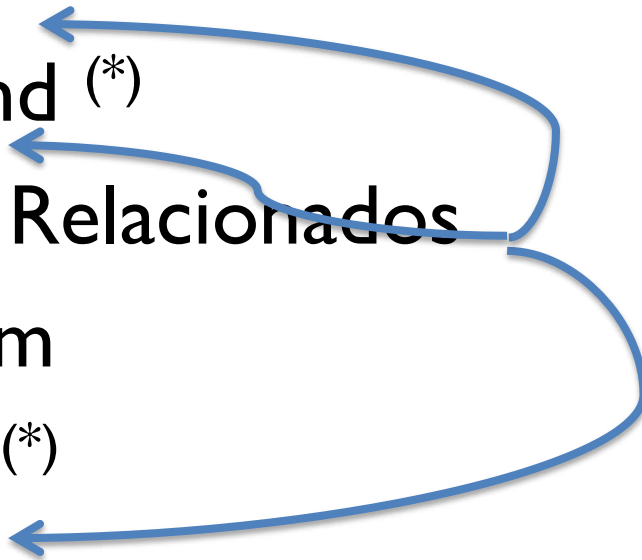
4. Abordagem

5. Avaliação (*)

6. Conclusão

Agradecimentos (*)

Referências

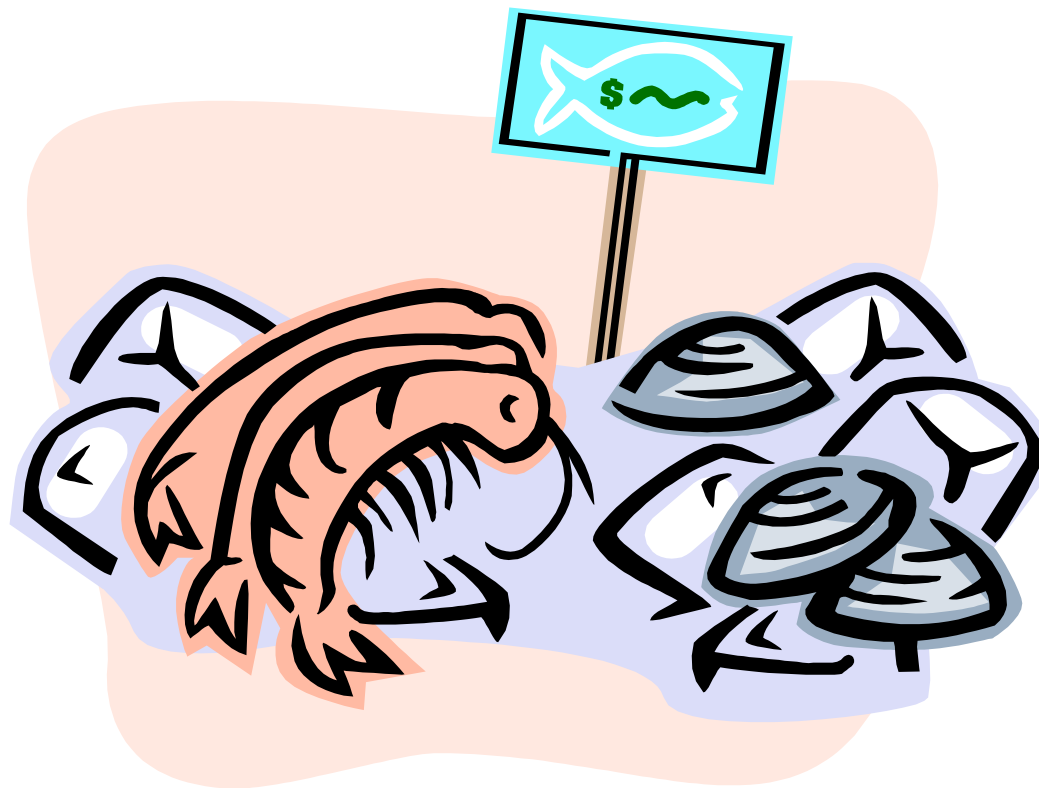


Onde Vender o Peixe?



Conclusão

- Parágrafo inicial: **Resumo** do que foi apresentado no artigo
- Frisar as **contribuições**
- Discutir **limitações**
- Discutir **trabalhos futuros**



VENDIDO!

Mas é só isso?

Reforçando a Motivação...

Trabalhos Relacionados

Trabalhos Relacionados

- Lição de casa que deve ser feita ANTES da elaboração da abordagem
 - Quem são seus concorrentes?
 - Quais problemas eles deixam em aberto?
 - Qual a relação deles entre si?

Como descobrir?

- Uso de ferramentas como bibliotecas digitais (ACM, IEEE)
- Uso de ferramentas que controlam citações

Google Acadêmico



CiteSeerX





COLLECTING

THEN



NOW



JORGE CHAM © 2011

WWW.PHDCOMICS.COM

Objetivos da Revisão da Literatura

- Evidenciar que você conhece o que existe na área
- Posicionar-se em relação ao que já existe
- Evidenciar lacunas que seu trabalho preenche

Dicas

- Não basta dizer que Fulano fez isso, Beltrano fez aquilo outro
 - É preciso evidenciar as **lacunas**, **comparar** os trabalhos entre si, **mostrar que você de fato entendeu** o que eles fazem e quais os **pontos fortes** e **fracos** de cada um
- Uma boa e profunda revisão da literatura muitas vezes pode ser **publicada** em forma de **survey**

Durante a Redação

- Garantir que todo trabalho citado aparece nas referências bibliográficas
- Garantir que todas as referências foram citadas

(Falaremos de ferramentas para auxiliar nisso na aula de quinta-feira)

Cuidado na construção das frases

Qi et al. [22] constructed a graphical model that clusters dependent sources into groups and measures the quality of each group as a whole (instead of each individual source).

**Nome do autor faz parte da
frase**

Cuidado na construção das frases

Correlation between sources are studied in two bodies of works. First, copy detection has been surveyed **in [10]** for various types of data and studied **in [3, 5, 6, 7, 16]** for structured data. [...] Second, [...]

Reescrevendo

Correlation between sources are studied in two bodies of works. First, copy detection has been surveyed for various types of data [10] and for structured data [3, 5, 6, 7, 16]. [...] Second, [...]

Uso de Software de Apoio

- Latex
 - JabRef/Bibtex
- Word
 - Zotero
 - Reference Manager
 - EndNote

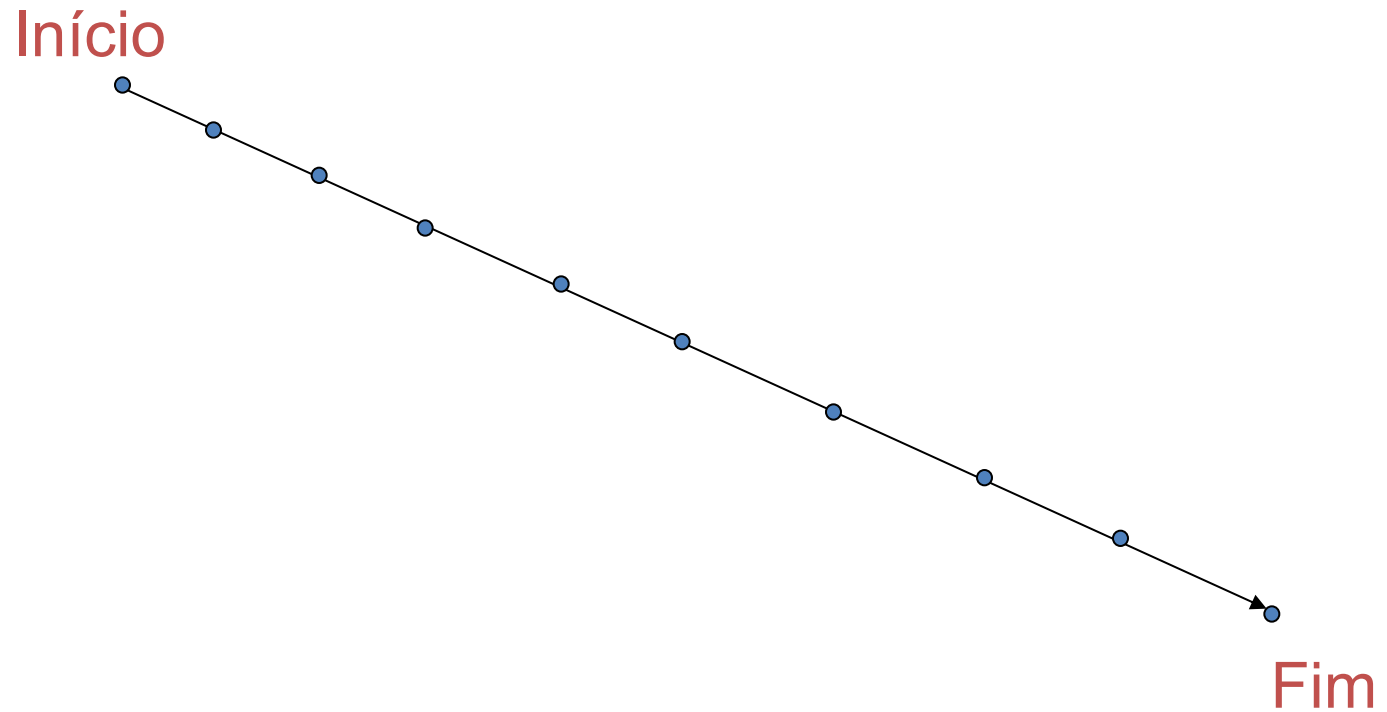
Questões Secundárias (mas não menos importantes!)

- Correção Gramatical
- Figuras
- Tabelas
- Formatação

Planejamento

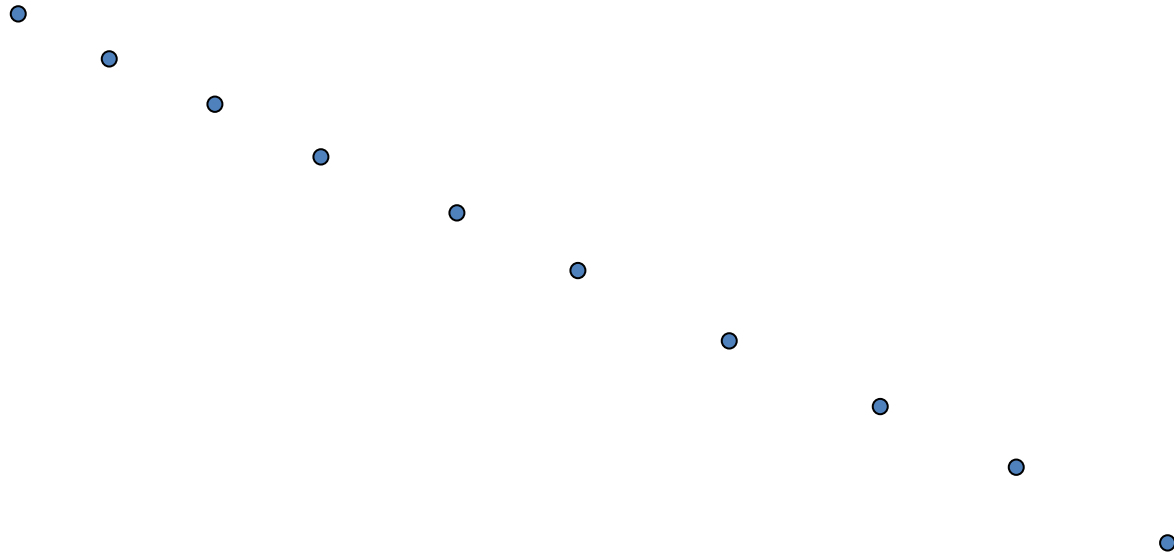
- Diversas formas de se escrever um artigo
- Escolha a que funciona melhor para você
 - Iniciar pela abordagem, que normalmente é mais fácil;
ou
 - Começar pela estrutura do artigo e ir construindo o
todo a partir daí (técnica de redação através de
scripts); ou
 - ...

Um bom artigo



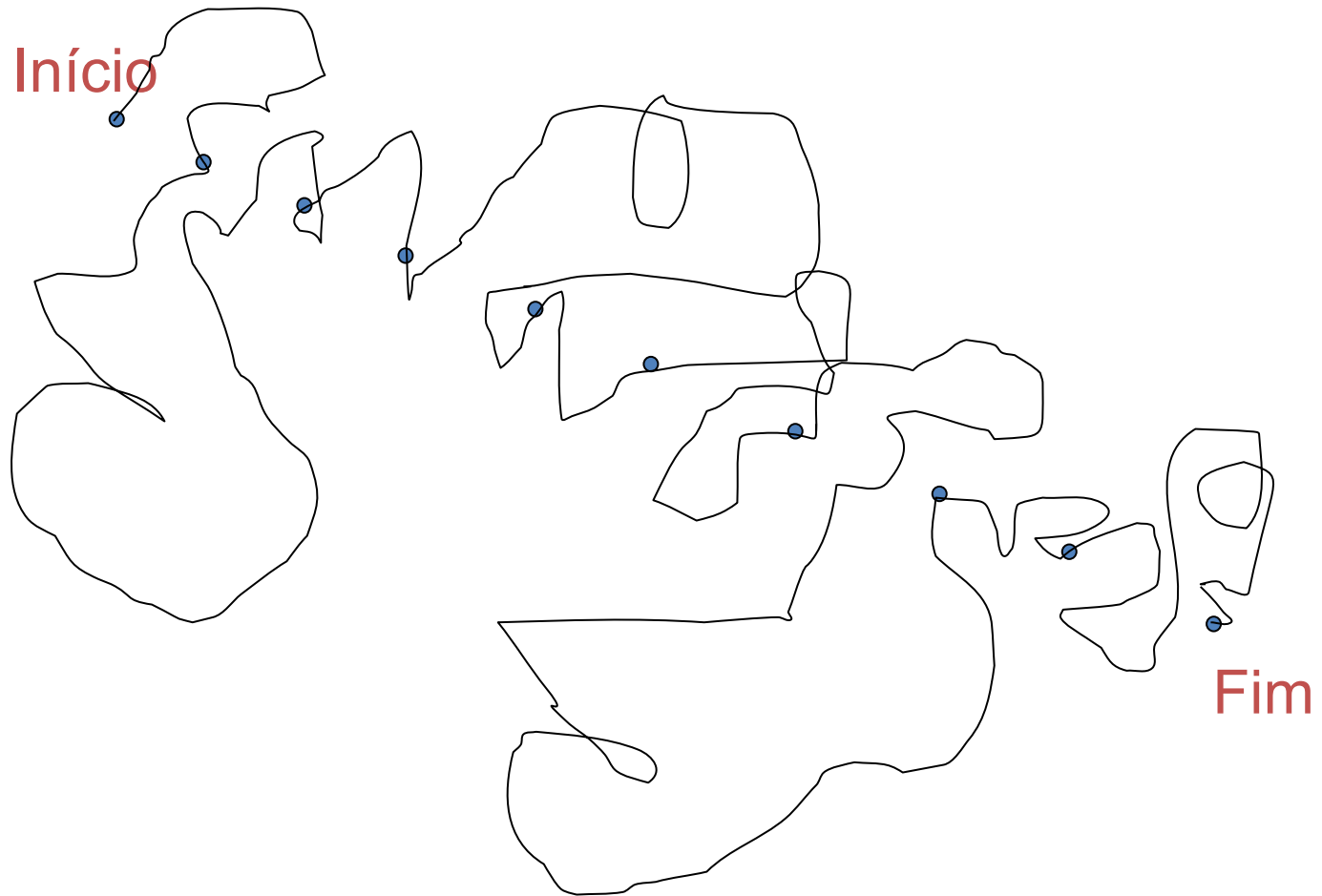
Um Artigo Ruim

Início



Fim

Um Artigo Ruim



Comentários dos revisores

- Leve-os em consideração na preparação da versão final do artigo
 - Na maioria das vezes, são comentários pertinentes, mesmo que o revisor não tenha sabido se expressar bem
 - Não adianta se revoltar com os comentários.
 - Mãos à obra!



HOLIDAY!

HOW A NORMAL PERSON CELEBRATES:

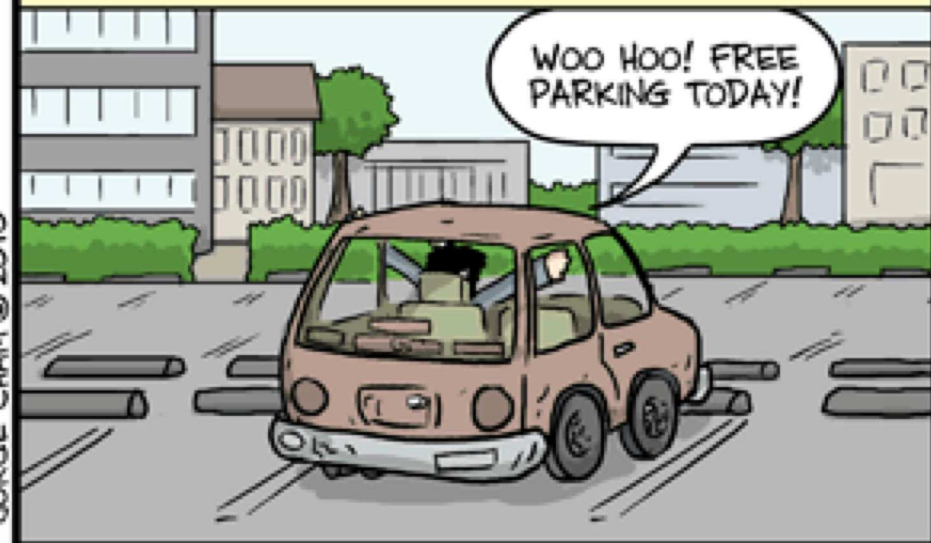
WOO HOO! NO
WORK TODAY!



JORGE CHAM © 2010

HOW A GRAD STUDENT CELEBRATES:

WOO HOO! FREE
PARKING TODAY!



WWW.PHDCOMICS.COM

Método Científico e Escrita de Artigos



h5

- h-index dos últimos 5 anos

O que é h-index?

- **H-index = N** se os **N** artigos mais citados possuem ao menos **N** citações
 - Serve para periódicos, congressos, pesquisadores

Bibliography

- [The scientific method](#), Khan Academy
- [How Science works](#), Understanding Science
- [Wikipedia](#), Scientific Method