

# OFICINA

## Escrita de Artigos





Home

Publications

Courses

Students

Service

Contact

## Vanessa Braganholo

Associate Professor (Prof. Associado I) - **Instituto de Computação**

**Universidade Federal Fluminense**, Niterói, RJ, Brazil.

Co-Chair of **SIGMOD DEMO 2015**, with Juliana Freire

Chair of the Brazilian Symposium on Databases, 2015

Associate Editor of **Journal of Information and Data Management (JIDM)**

From 2006 to 2010, I worked at **Universidade Federal do Rio de Janeiro** as an Assistant Professor.

(DSc., **Instituto de Informática, UFRGS**, 2004)

(BSc., **Instituto de Informática, UFRGS**, 1998)

# Agenda

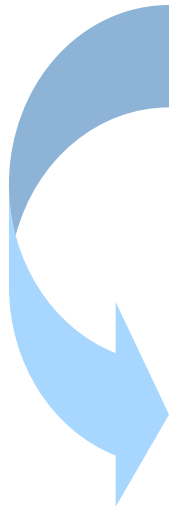
- Importância de Publicar
- Fluxo de Publicação: Congresso x Periódico
- Carta de Resposta aos Revisores
- Redação de Artigos
- Considerações Finais

# Agenda

- **Importância de Publicar**
- Fluxo de Publicação: Congresso x Periódico
- Carta de Resposta aos Revisores
- Redação de Artigos
- Considerações Finais

# Por quê escrever Artigos?

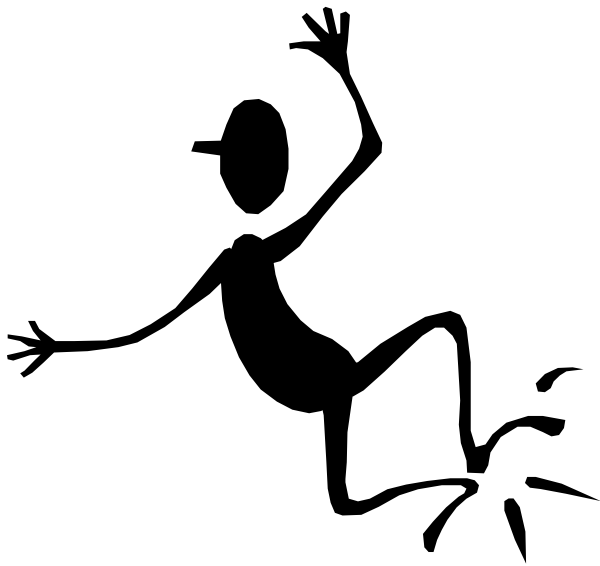
## 1) Divulgar conhecimento



**Ciência** avança através de **pesquisas**

**Pesquisas** são divulgadas através de **publicações**

# De que adianta...



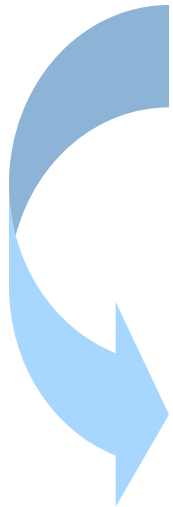
Descobrir a **cura** da AIDS

Se **ninguém** ficar  
sabendo disso?



# Por quê escrever Artigos?

## 2) Dinheiro



Sem **dinheiro**

Não há **pesquisa**

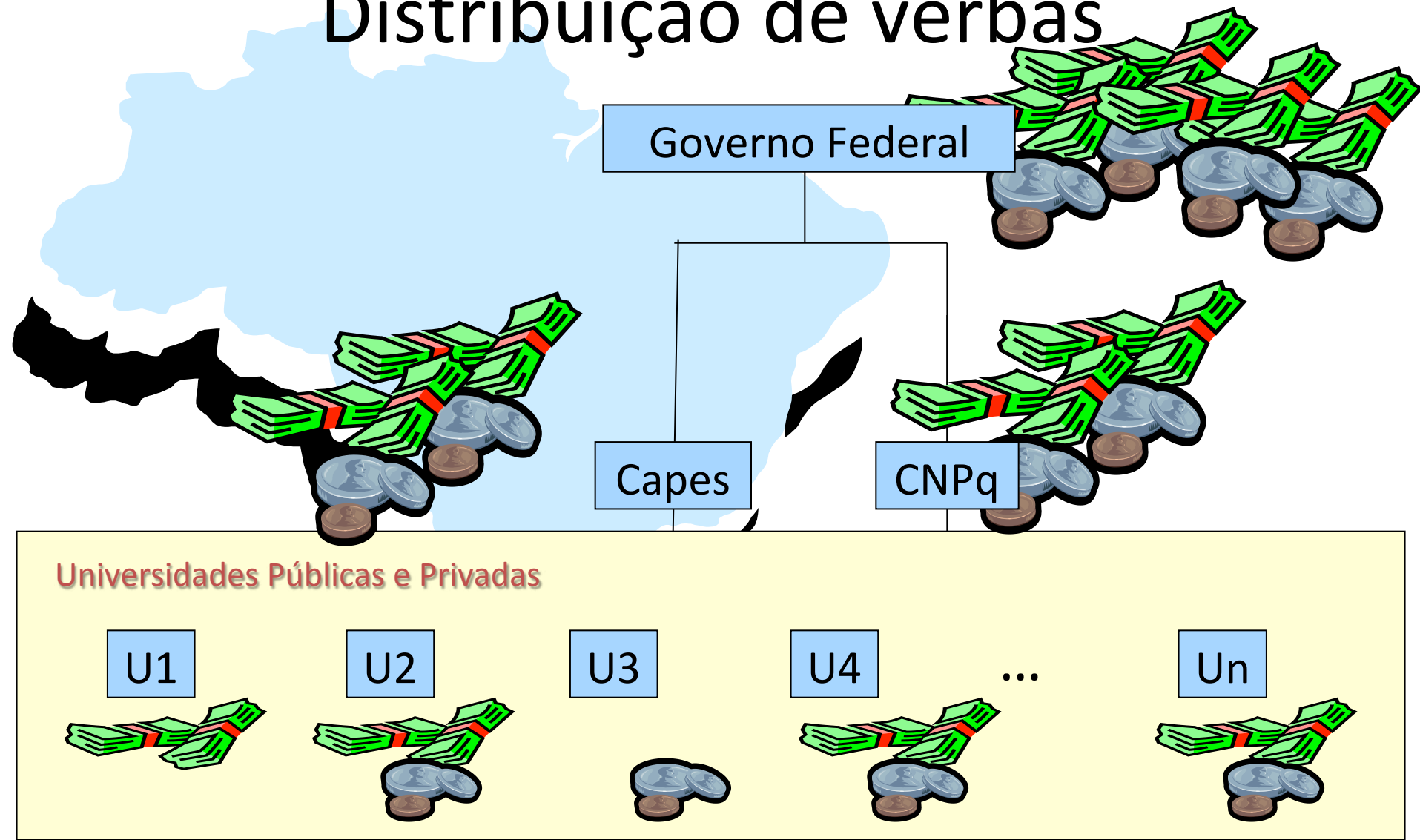
# Dinheiro pra quê?

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





# Distribuição de verbas



# 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

- Avalia todos os cursos de Pós-Graduação do país
- Feita a cada 3 anos
- Nota de 1 a 7

# Critérios de Avaliação

- Definidos em um documento chamado **Documento de Área** <http://qualis.capes.gov.br/webqualis/> menu “documentos de área”)
- Documento de Área da **Computação** define **como** as **notas** para os cursos de Computação **são calculadas**
- Uma parte considerável (35%) da nota vem da Produção Intelectual do Programa de Pós-Graduação (**PUBLICAÇÕES!!**)

# Qualidade dos veículos de publicação

- Comitê de Computação da Capes mede a qualidade através de vários critérios
  - Congressos: h-index calculado com o auxílio do Google Scholar
  - Periódicos: h-index calculado com o auxílio do Google Scholar (HG) e JCR

# O que é h-index?

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



**INÍCIO**

**SOBRE O SHINE**

**EQUIPE**

**CONTATO**

**SBC**

**UFAM**

**ICOMP**



## SBBD - Simpósio Brasileiro de Bancos de Dados

SBBD - Simpósio Brasileiro de Bancos de D

Artigos de  
2001

até  
2014

2000 2005 2010 2014

**Enviar**

H-Index = 14

SBBD tem 14 papers com 14 ou mais citações entre 2001 e 2014:

1

**Privacy preserving clustering by data transformation**

Stanley Oliveira, Osmar R. Zaiane

**Ano:** 2003. **Citações:** 126

2

**Siphoning Hidden-Web Data through Keyword-Based Interfaces**

Luciano Barbosa, Juliana Freire

**Ano:** 2004. **Citações:** 103

3

**A Geographic Knowledge Base for Semantic Web Applications**

Marcirio Silveira Chaves, Mário J. Silva, Bruno Martins

**Ano:** 2005. **Citações:** 50





## Vanessa Braganholo

Professor of Computer Science, Universidade Federal Fluminense

[Databases](#)

E-mail confirmado em ic.uff.br - [Página inicial](#)

Meu perfil é público

Editar

Seguir

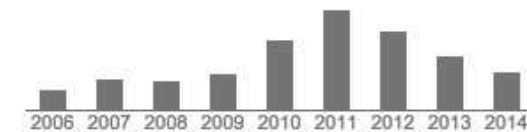
[Alterar foto](#)

Google Acadêmico

ic.uff.br



Índices de citações	Todos	Desde 2009
Citações	799	575
Índice h	15	13
Índice i10	25	16



<input type="checkbox"/>	Título	Adicionar	Mais	1-20	Citado por	Ano
<input type="checkbox"/>	From XML view updates to relational view updates: old solutions to a new problem				77	2004
	VP Braganholo, SB Davidson, CA Heuser Proceedings of the Thirtieth international conference on Very large data ...					
<input type="checkbox"/>	Integrating The Galileo Applications For Simulation of Offshore Systems Via The GXML Unified Format				55	2009
	IHF Santos, V Braganholo, M Mattoso, BP Jacob, C Albrecht XXX CILAMCE-Iberian Latin American Congress on Computational Methods in ...					
<input type="checkbox"/>	Towards supporting the life cycle of large scale scientific experiments				54	2010
	M Mattoso, C Werner, GH Travassos, V Braganholo, E Ogasawara, ... International Journal of Business Process Integration and Management 5 (1 ...					
<input type="checkbox"/>	On the updatability of XML views over relational databases.				48	2003
	VP Braganholo, SB Davidson, CA Heuser WebDB, 31-36					
<input type="checkbox"/>	Updating relational databases through XML views				45	2002
	V de Paula Braganholo Universidade Federal do Rio Grande do Sul					
<input type="checkbox"/>	Exploring many task computing in scientific workflows				42	2009
	E Ogasawara, D de Oliveira, F Chirigati, CE Barbosa, R Elias, ... Proceedings of the 2nd Workshop on Many-Task Computing on Grids and ...					

# Qualidade dos veículos de publicação

## Congressos e periódicos:

- Recebem uma nota entre A1, A2, B1, B2, B3, B4, B5 e C
- A Computação é a única área que tem Qualis para congressos

Por ISSN do Periódico

Por Título do Periódico

Por Classificação / Área de Avaliação

**Lista Completa**

Área de Avaliação: CIÊNCIA DA COMPUTAÇÃO

Estrato: A1

● Pesquisar

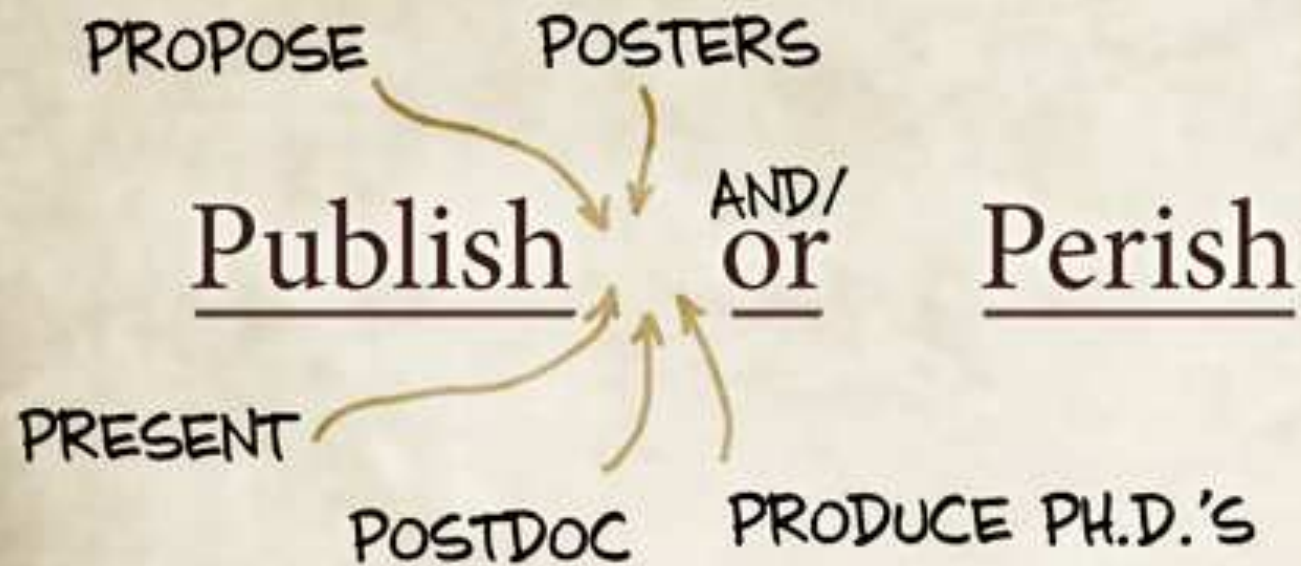
ISSN	Título	Estrato	Área de Avaliação	Classificação
0360-0300	ACM Computing Surveys	A1	CIÊNCIA DA COMPUTAÇÃO	Atualizado
0730-0301	ACM Transactions on Graphics	A1	CIÊNCIA DA COMPUTAÇÃO	Atualizado
1748-7188	Algorithms for Molecular Biology	A1	CIÊNCIA DA COMPUTAÇÃO	Atualizado
1134-3060	Archives of Computational Methods in Engineering	A1	CIÊNCIA DA COMPUTAÇÃO	Atualizado
0004-3702	Artificial Intelligence (General Ed.)	A1	CIÊNCIA DA COMPUTAÇÃO	Atualizado
0005-1098	Automatica (Oxford)	A1	CIÊNCIA DA COMPUTAÇÃO	Atualizado
0929-5593	Autonomous Robots	A1	CIÊNCIA DA COMPUTAÇÃO	Atualizado
1573-7527	Autonomous Robots (Dordrecht. Online)	A1	CIÊNCIA DA COMPUTAÇÃO	Atualizado
0140-525X	Behavioral and Brain Sciences (Print)	A1	CIÊNCIA DA COMPUTAÇÃO	Atualizado
1367-4811	Bioinformatics (Oxford. Online)	A1	CIÊNCIA DA COMPUTAÇÃO	Atualizado

« «

**1** 2 3 4 5 6 7 8 9 10

» »

Exportar PDF



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

# Mas...

- 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
  - <https://grupos.ufrgs.br/mailman/listinfo/Sbc-l/>



# Listas de email

- Listas internacionais específicas de cada área
  - DBWORLD (área de Banco de Dados)
    - <http://www.cs.wisc.edu/dbworld/>
  - SEWORLD (área de Engenharia de Software)
    - <http://serl.cs.colorado.edu/~serl/seworld/>
  - ...

# Agenda

- Importância de Publicar
- **Fluxo de Publicação: Congresso x Periódico**
- Carta de Resposta aos Revisores
- Redação de Artigos
- Considerações Finais

# Congresso



Comitê de Programa  
Revisores



# Congresso



**ANTES DO  
DEADLINE**

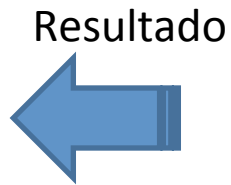
# Congresso



# Congresso



# Congresso

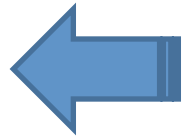




# Congresso



Resultado



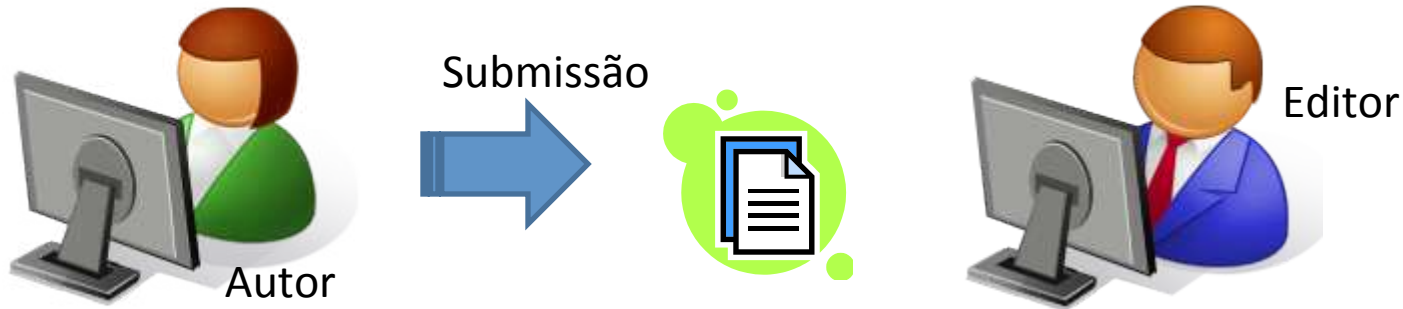
**ACEITO**  
**RECUSADO**  
**REBUTTAL\***

# Periódico



Corpo de Revisores

# Periódico



**A QUALQUER  
TEMPO  
(FLUXO CONTÍNUO)**

# Periódico



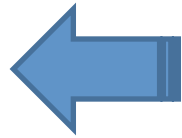
# Periódico



# Periódico



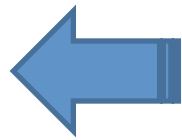
Resultado



# Periódico



Resultado



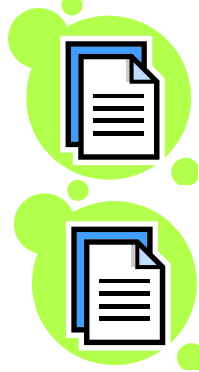
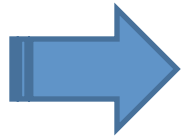
**ACEITO**  
**RECUSADO**  
**MAJOR REVIEW**  
**MINOR REVIEW**



# Periódico



Submissão Revisada +  
Carta de Resposta



**DENTRO DA DATA  
ESTIPULADA PELO  
EDITOR**



# Agenda

- Importância de Publicar
- Fluxo de Publicação: Congresso x Periódico
- **Carta de Resposta aos Revisores**
- Redação de Artigos
- Considerações Finais

# 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.

**Elogie as boas sugestões**

# Response to Reviewers

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

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

# Agenda

- Importância de Publicar
- Fluxo de Publicação: Congresso x Periódico
- Carta de Resposta aos Revisores
- **Redação de Artigos**
- Considerações Finais

# Formato

- Definido pelo veículo para o qual o artigo será submetido
- Ver regras no CALL FOR PAPERS

For the 2014 edition, SBBD will accept two kinds of submissions:

**JIDM articles.** Submissions in this category should present interesting results or novel thought-provoking ideas in any of the topics of interest. They must be written in English and not exceed 12 pages, according to the format of the Journal of Information and Data Management (JIDM). The templates for submission are available at <http://bit.ly/1eil5co>. Articles must be submitted by the deadline through EasyChair at <https://www.easychair.org/conferences/?conf=jidmvol5>

**SBBD proceedings.** Submissions in this category should present initial results or on-going work in any of the topics of interest. They must be written in Portuguese or English and not exceed 10 pages, according to the format of the SBC. The templates for submission are available at <http://bit.ly/1coeoWk>. Papers must be submitted by the deadline through EasyChair website at <https://www.easychair.org/conferences/?conf=sbbd2014>

# 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**

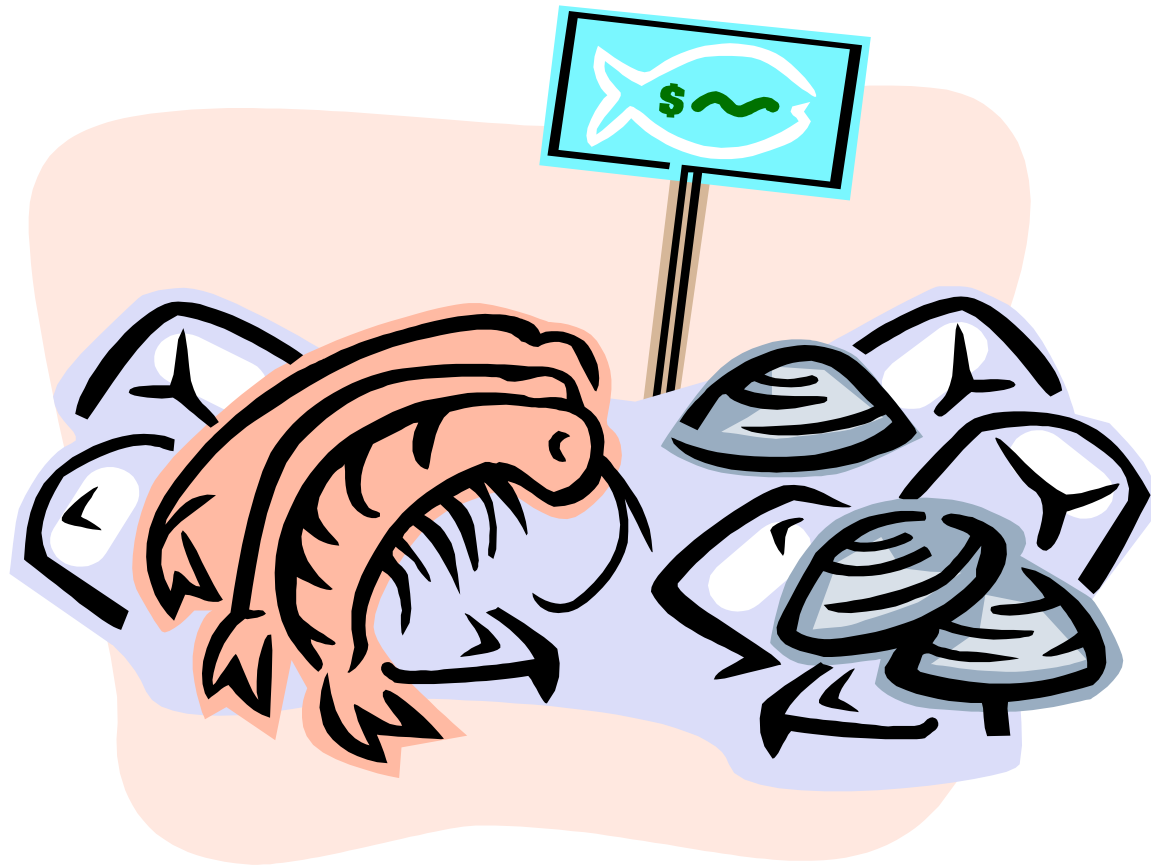


**LNCS**





# Essencial (!!!)



# Passos para a escrita de artigo

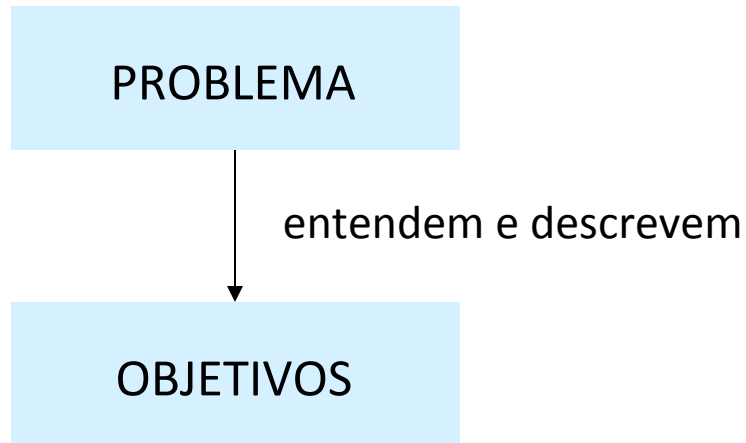
**CONTEXTO**

# Passos para a escrita de artigo

PROBLEMA

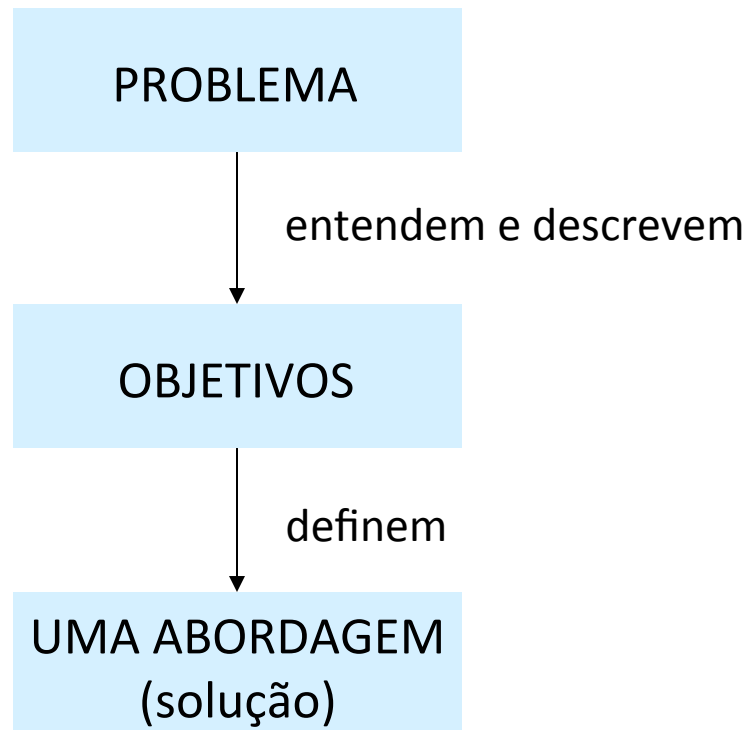
**CONTEXTO**

# Passos para a escrita de artigo



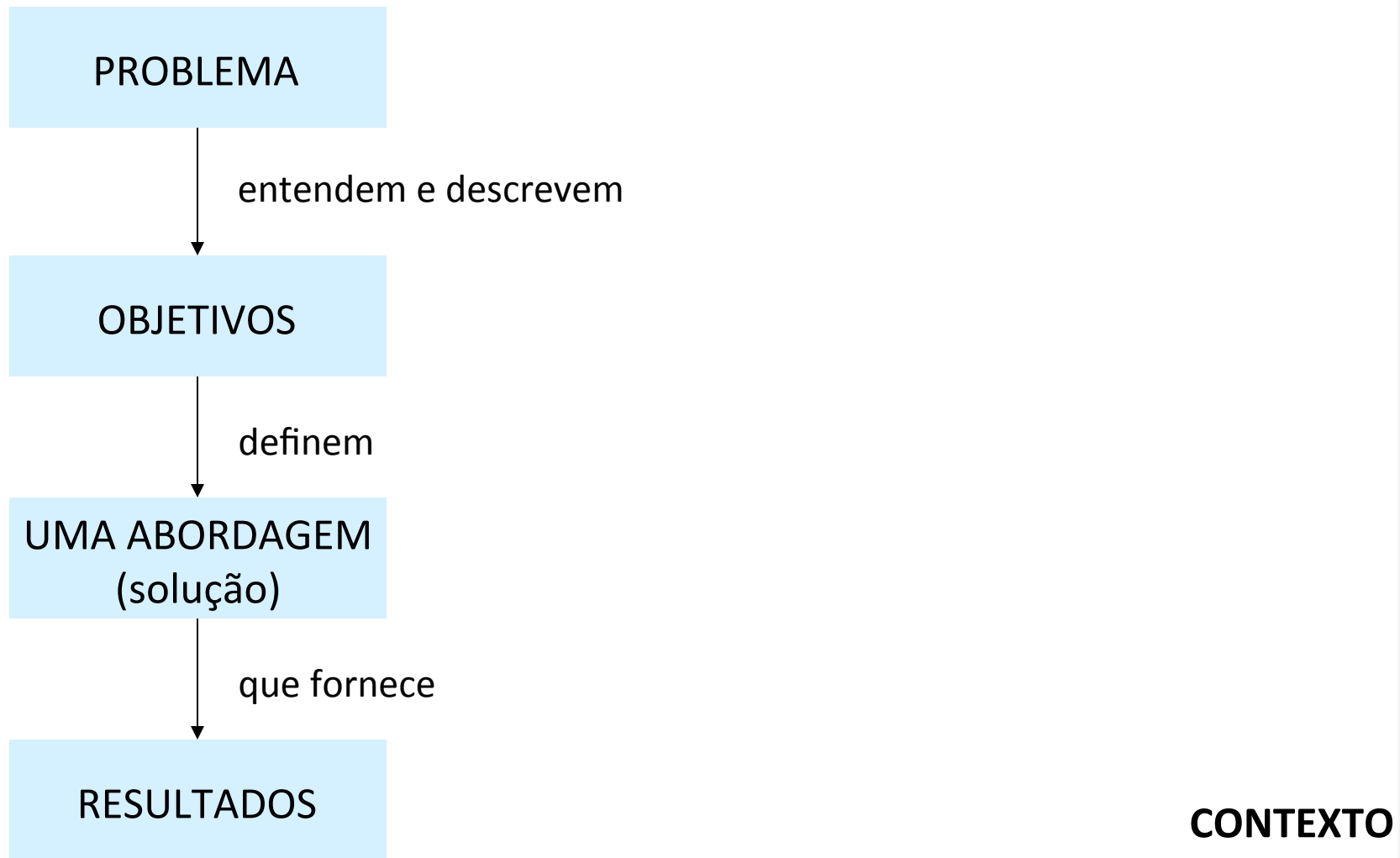
**CONTEXTO**

# Passos para a escrita de artigo

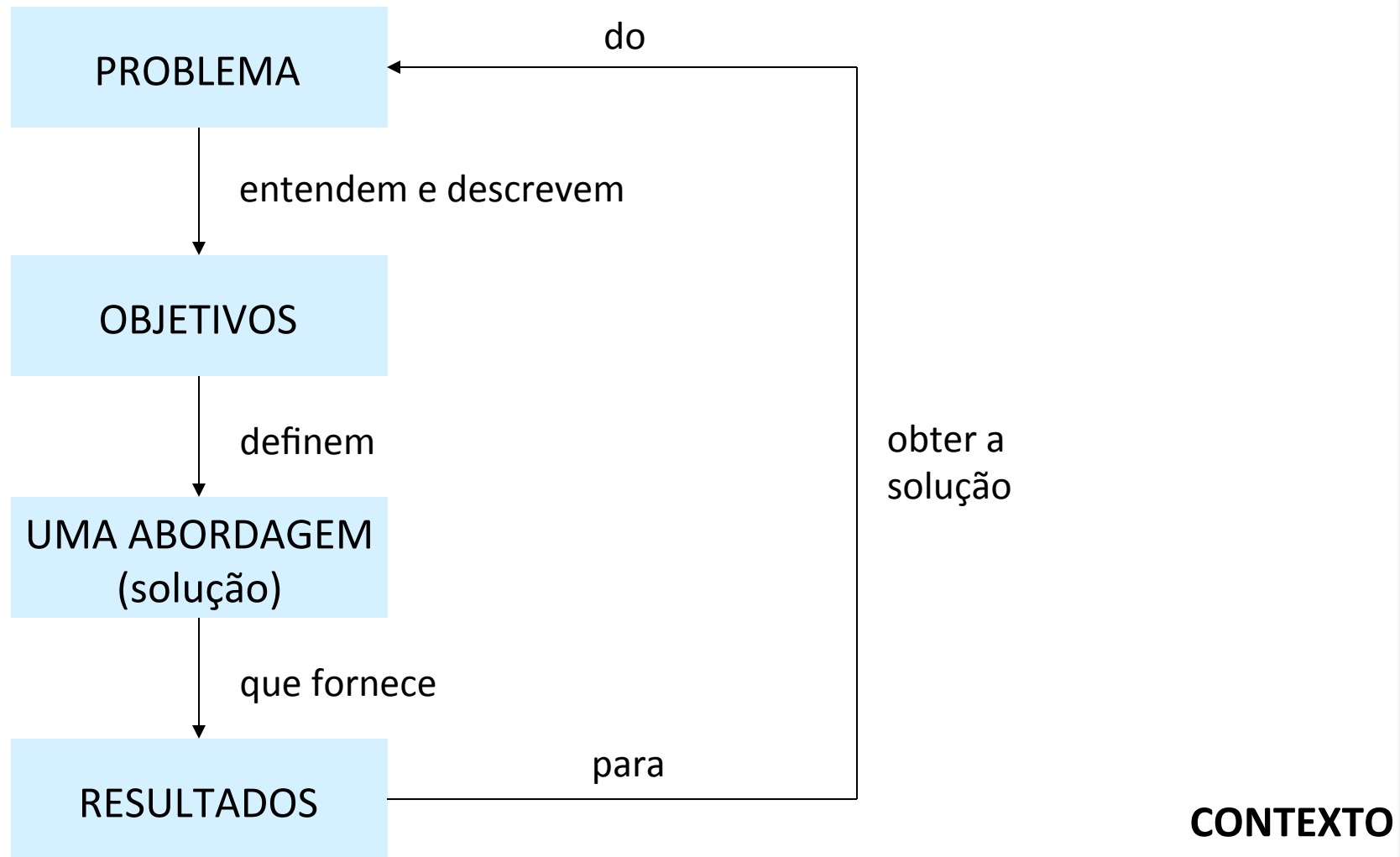


**CONTEXTO**

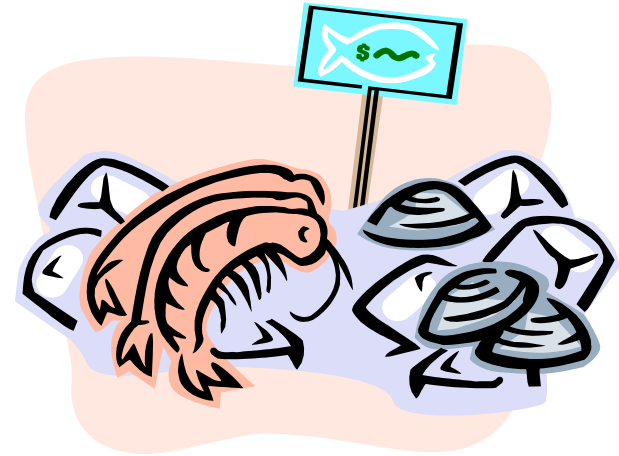
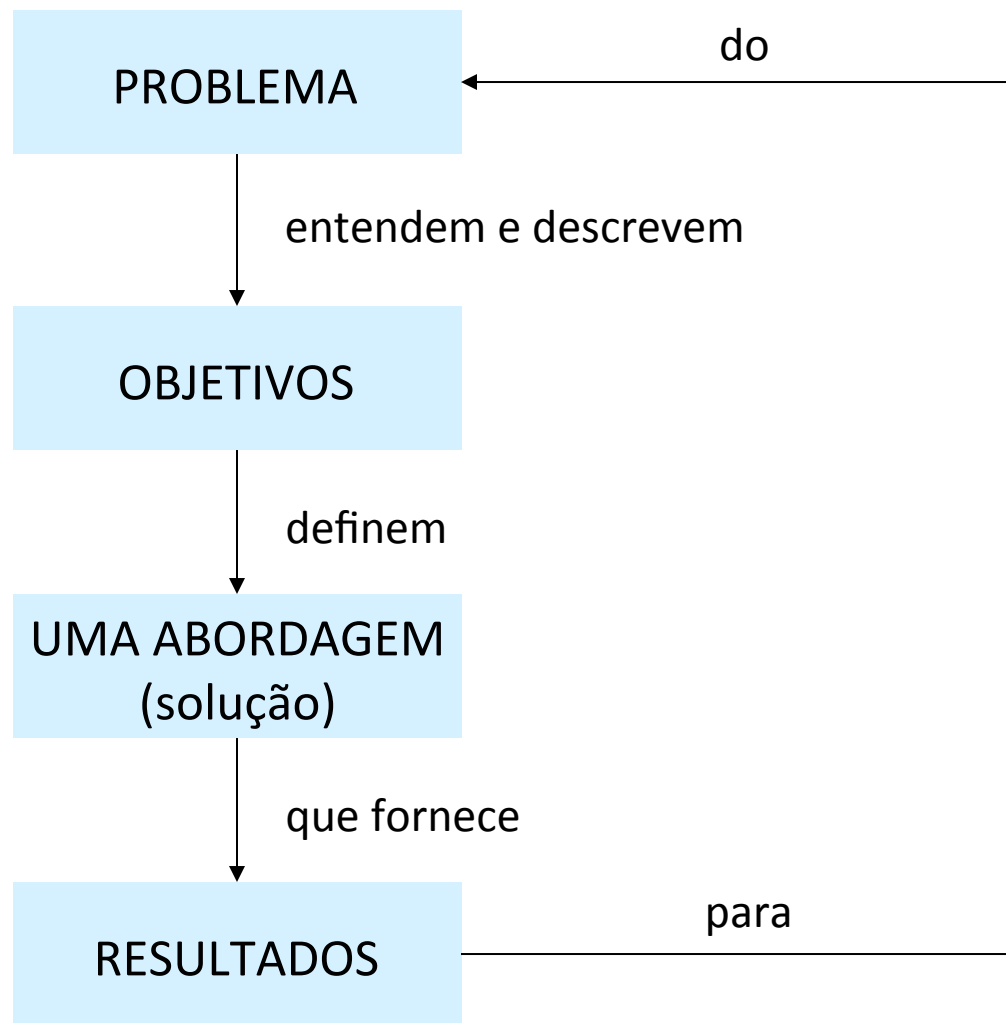
# Passos para a escrita de artigo



# Passos para a escrita de artigo



# Passos para a escrita de artigo



obter a  
solução

**CONTEXTO**



# Onde Vender o Peixe?



# 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

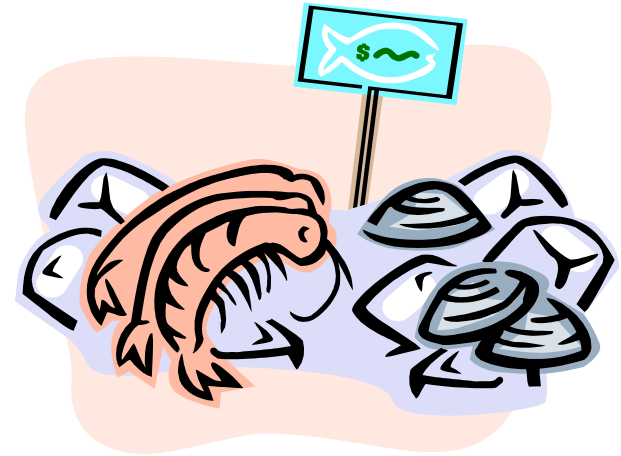
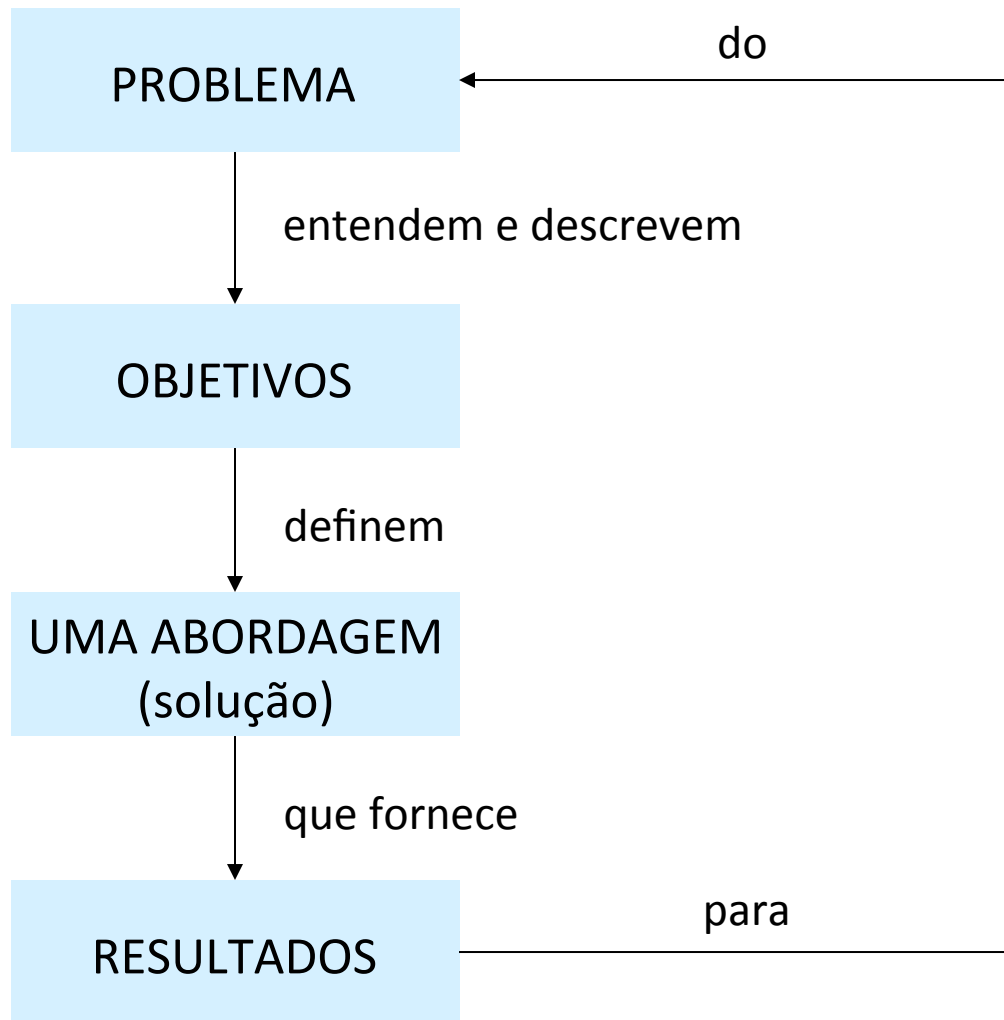
- Um único parágrafo (50 a 200 palavras)
- Função: dizer aos leitores se vale a pena ler o restante do artigo
- Atenha-se ao essencial

Zobel, 2001

# Resumo NÃO deve ter...

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

# Resumo

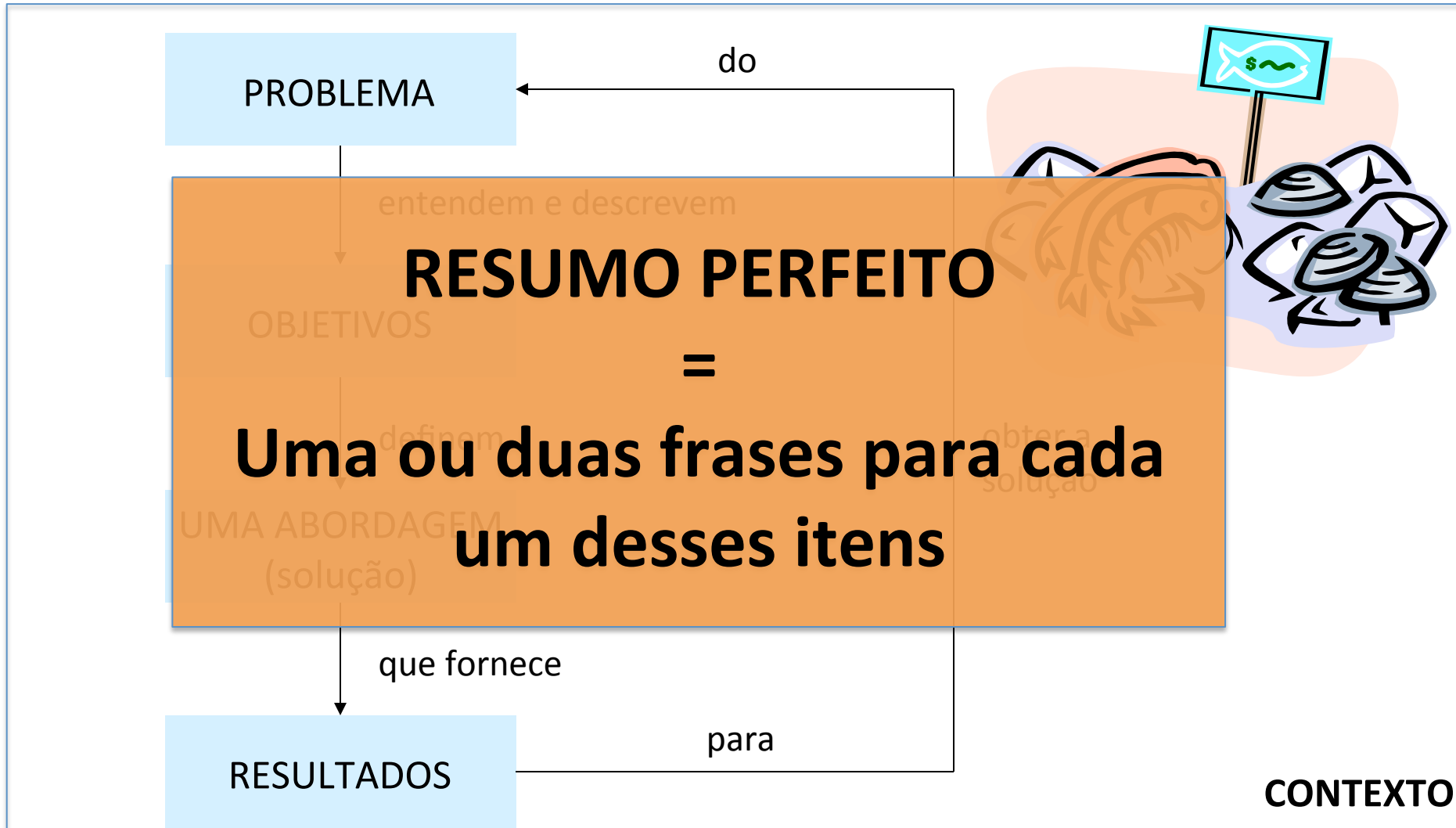


obter a  
solução

**CONTEXTO**

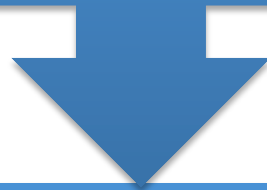


# Resumo



# Entregue o ouro...

Espaço necessário em memória pode ser reduzido significativamente

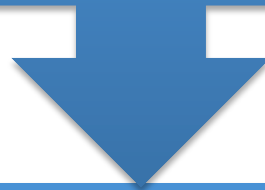


Espaço necessário em memória pode ser reduzido em 60%

Zobel, 2001

# Entregue o ouro...

Nós definimos um novo algoritmo de inversão



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

Zobel, 2001

**Many applications rely on Web data and extraction systems to accomplish knowledge-driven tasks.** Web information is not curated, so many sources provide inaccurate, or conflicting information. Moreover, extraction systems introduce additional noise to the data. We wish to automatically distinguish correct data and erroneous data for creating a cleaner set of integrated data. Previous work has shown that a naïve voting that trusts data provided by the majority or at least a certain number of sources may not work well in the presence of copying between the sources. However, correlation between sources can be much broader than copying: sources may provide data from complementary domains (negative correlation), extractors may focus on different types of information (negative correlation), and extractors may apply common rules in extraction (positive correlation, without copying). In this paper we present novel techniques modeling correlations between sources and applying it in truth finding. We provide a comprehensive evaluation of our approach on three real-world datasets with different characteristics, as well as on synthetic data, showing that our algorithms outperform the existing state-of-the-art techniques.

Many applications rely on Web data and extraction systems to accomplish knowledge-driven tasks. **Web information is not curated, so many sources provide inaccurate, or conflicting information. Moreover, extraction systems introduce additional noise to the data.** We wish to automatically distinguish correct data and erroneous data for creating a cleaner set of integrated data. Previous work has shown that a naïve voting that trusts data provided by the majority or at least a certain number of sources may not work well in the presence of copying between the sources. However, correlation between sources can be much broader than copying: sources may provide data from complementary domains (negative correlation), extractors may focus on different types of information (negative correlation), and extractors may apply common rules in extraction (positive correlation, without copying). In this paper we present novel techniques modeling correlations between sources and applying it in truth finding. We provide a comprehensive evaluation of our approach on three real-world datasets with different characteristics, as well as on synthetic data, showing that our algorithms outperform the existing state-of-the-art techniques.

Many applications rely on Web data and extraction systems to accomplish knowledge-driven tasks. Web information is not curated, so many sources provide inaccurate, or conflicting information. Moreover, extraction systems introduce additional noise to the data. **We wish to automatically distinguish correct data and erroneous data for creating a cleaner set of integrated data.** Previous work has shown that a naïve voting that trusts data provided by the majority or at least a certain number of sources may not work well in the presence of copying between the sources. However, correlation between sources can be much broader than copying: sources may provide data from complementary domains (negative correlation), extractors may focus on different types of information (negative correlation), and extractors may apply common rules in extraction (positive correlation, without copying). In this paper we present novel techniques modeling correlations between sources and applying it in truth finding. We provide a comprehensive evaluation of our approach on three real-world datasets with different characteristics, as well as on synthetic data, showing that our algorithms outperform the existing state-of-the-art techniques.

Many applications rely on Web data and extraction systems to accomplish knowledge-driven tasks. Web information is not curated, so many sources provide inaccurate, or conflicting information. Moreover, extraction systems introduce additional noise to the data. We wish to automatically distinguish correct data and erroneous data for creating a cleaner set of integrated data. **Previous work has shown that a naïve voting that trusts data provided by the majority or at least a certain number of sources may not work well in the presence of copying between the sources. However, correlation between sources can be much broader than copying: sources may provide data from complementary domains (negative correlation), extractors may focus on different types of information (negative correlation), and extractors may apply common rules in extraction (positive correlation, without copying).** In this paper we present novel techniques modeling correlations between sources and applying it in truth finding. We provide a comprehensive evaluation of our approach on three real-world datasets with different characteristics, as well as on synthetic data, showing that our algorithms outperform the existing state-of-the-art techniques.

Many applications rely on Web data and extraction systems to accomplish knowledge-driven tasks. Web information is not curated, so many sources provide inaccurate, or conflicting information. Moreover, extraction systems introduce additional noise to the data. We wish to automatically distinguish correct data and erroneous data for creating a cleaner set of integrated data. Previous work has shown that a naïve voting that trusts data provided by the majority or at least a certain number of sources may not work well in the presence of copying between the sources. However, correlation between sources can be much broader than copying: sources may provide data from complementary domains (negative correlation), extractors may focus on different types of information (negative correlation), and extractors may apply common rules in extraction (positive correlation, without copying). **In this paper we present novel techniques modeling correlations between sources and applying it in truth finding.** We provide a comprehensive evaluation of our approach on three real-world datasets with different characteristics, as well as on synthetic data, showing that our algorithms outperform the existing state-of-the-art techniques.



Many applications rely on Web data and extraction systems to accomplish knowledge-driven tasks. Web information is not curated, so many sources provide inaccurate, or conflicting information. Moreover, extraction systems introduce additional noise to the data. We wish to automatically distinguish correct data and erroneous data for creating a cleaner set of integrated data. Previous work has shown that a naïve voting that trusts data provided by the majority or at least a certain number of sources may not work well in the presence of copying between the sources. However, correlation between sources can be much broader than copying: sources may provide data from complementary domains (negative correlation), extractors may focus on different types of information (negative correlation), and extractors may apply common rules in extraction (positive correlation, without copying). In this paper we present novel techniques modeling correlations between sources and applying it in truth finding. **We provide a comprehensive evaluation of our approach on three real-world datasets with different characteristics, as well as on synthetic data, showing that our algorithms outperform the existing state-of-the-art techniques.**

**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.

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

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.

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

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.

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

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.

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

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.**

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

# 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**



# 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**?

Fonte: Leonardo Murta

# Análise de Introdução de Artigos

- Pochampally, Das Sarma, Luna Dong, Meliou, Srivastava. [Fusing data with correlations.](#) SIGMOD 2014: 433-444
- Zhang, Ramakrishnan, Livny. [BIRCH: An Efficient Data Clustering Method for Very Large Databases.](#) SIGMOD 1996: 103-114

# 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

# Estrutura

1. Introdução

2. Background (\*)

3. Trabalhos Relacionados

4. Abordagem

5. Avaliação (\*)

6. Conclusão

Agradecimentos (\*)

Referências



# Análise de Estrutura de Artigos

- Pochampally, Das Sarma, Luna Dong, Meliou, Srivastava. [Fusing data with correlations.](#) SIGMOD 2014: 433-444
- Zhang, Ramakrishnan, Livny. [BIRCH: An Efficient Data Clustering Method for Very Large Databases.](#) SIGMOD 1996: 103-114

# Onde Vender o Peixe?



# Conclusão

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



# Análise de Conclusão de Artigos

- Pochampally, Das Sarma, Luna Dong, Meliou, Srivastava. [Fusing data with correlations.](#) SIGMOD 2014: 433-444
- Zhang, Ramakrishnan, Livny. [BIRCH: An Efficient Data Clustering Method for Very Large Databases.](#) SIGMOD 1996: 103-114

# 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
  - CiteSeer



# COLLECTING

THEN



JORGE CHAM © 2011

NOW



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

# Tipos de Revisão da Literatura a Serem Evitados

- *Summa*
  - Apresenta “toda” a produção científica da cultura ocidental e oriental
  - Tentativa de esgotar o assunto
- *Patchwork*
  - Colagem de conceitos, pesquisas de diversos autores sem um fio condutor
  - Não se percebe planejamento ou sistematização
  - Autor Perdido



# Tipos de Revisão da Literatura a Serem Evitados

- Suspense
  - Há um roteiro, mas não se sabe onde o autor quer chegar, nem qual a ligação dos fatos com o tema do estudo
  - Cortina de fumaça
- Caderno B
  - Texto leve que trata os assuntos mais complexos de modo ligeiro, sem aprofundamentos cansativos
  - Coleção Primeiros Passos

# Tipos de Revisão da Literatura a Serem Evitados

- Cronista Social
  - Independente de sua importância no contexto do trabalho, cita autores da moda
- *Off the records*
  - Autor garante o anonimato das fontes: não revela os autores
    - sabe-se...
    - muitos autores...
    - vários estudos...
    - tem sido observado

# Tipos de Revisão da Literatura a Serem Evitados

- Ventriloquo
  - Autor só fala pela boca dos outros
    - Para Fulano...
    - Segundo Beltrano...
    - Fulano afirma...
    - Sicrano observa...
  - Não há críticas ou posicionamento

# Durante a Redação

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

# 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



# Demonstração

- Zotero

# Questões Secundárias

- Correção Gramatical
- Figuras
- Tabelas
- Formatação

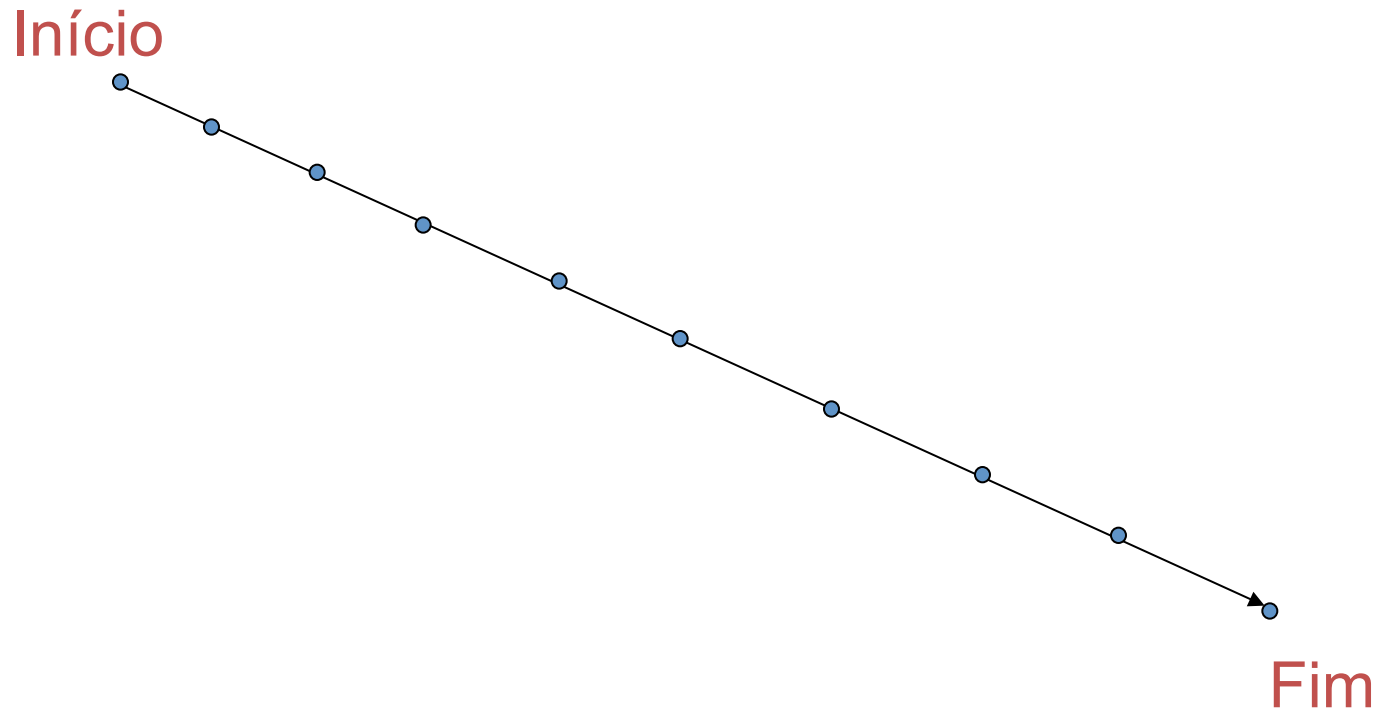
# Planejamento

- Diversas formas de se escrever um artigo
- Escolha a que funciona melhor para você
  - Sequencia sugerida aqui; ou
  - Começar pela estrutura do artigo e ir construindo o todo a partir daí (técnica de redação através de scripts); ou
  - ...

# Agenda

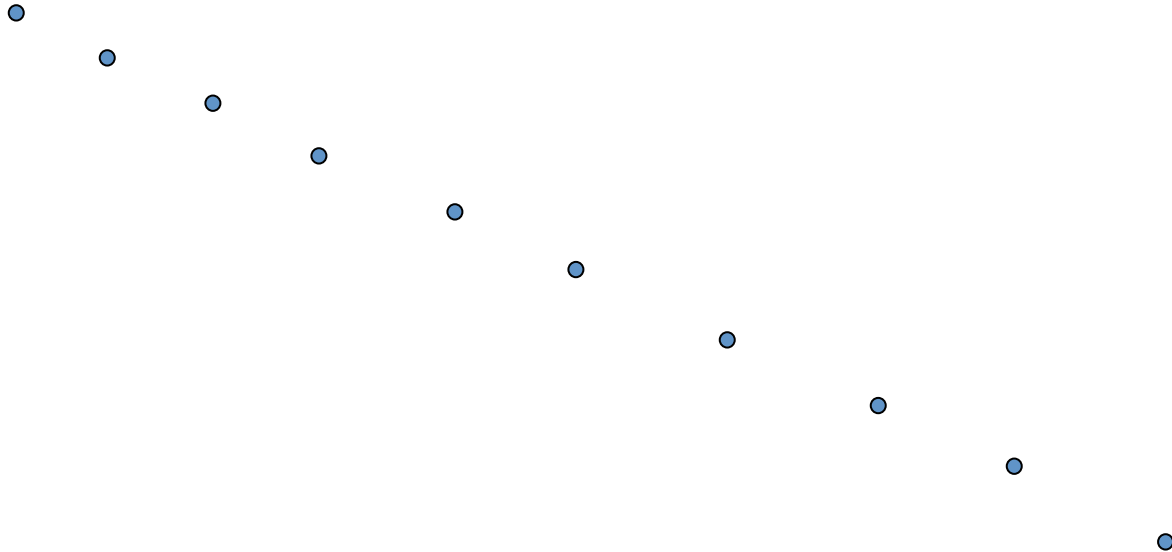
- Importância de Publicar
- Fluxo de Publicação: Congresso x Periódico
- Carta de Resposta aos Revisores
- Redação de Artigos
- **Considerações Finais**

# Um bom artigo



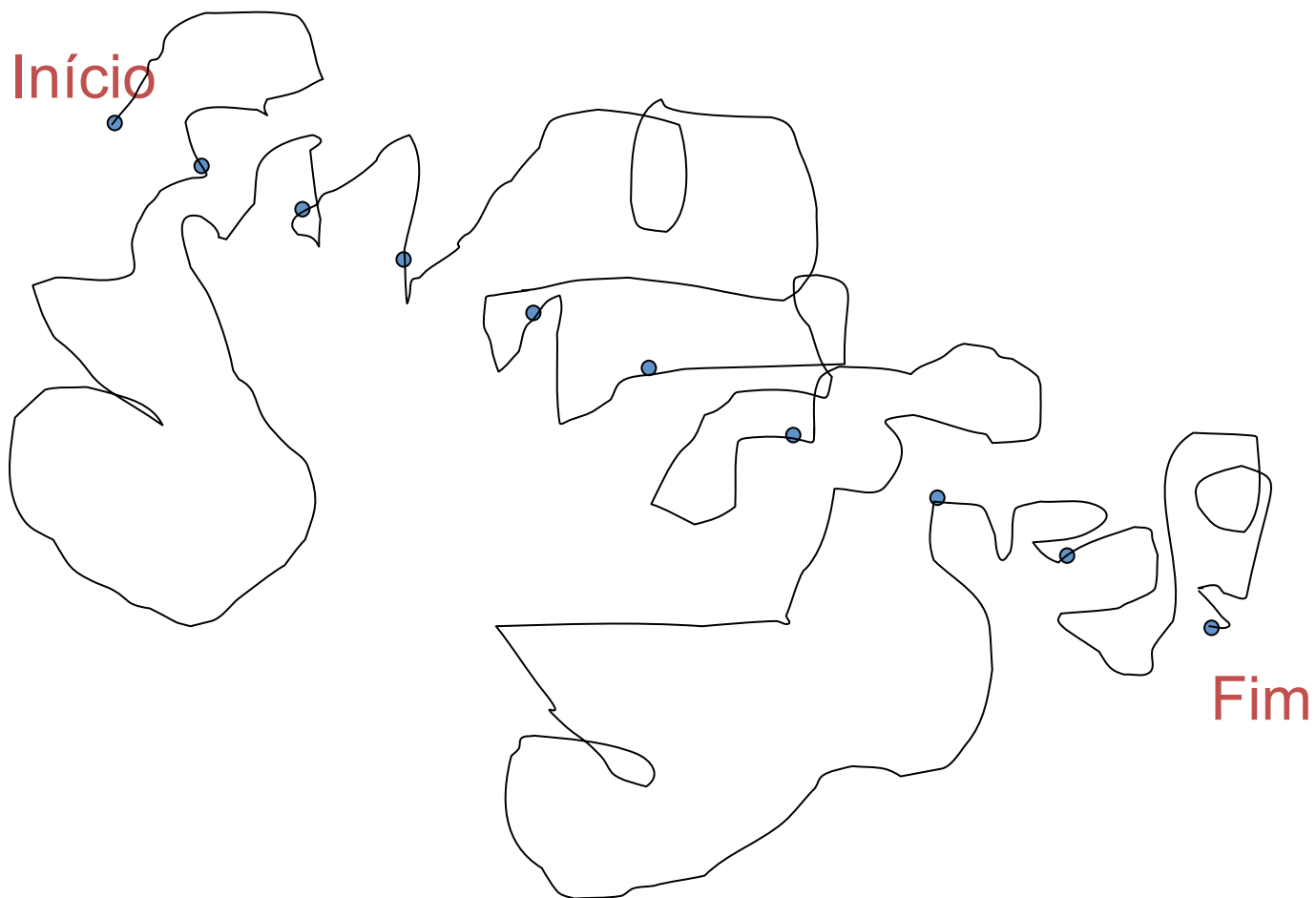
# Um Artigo Ruim

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

# Não Perca!

Como Apresentar Resultados de  
Pesquisa

**AMANHÃ, 16h**

<https://www.facebook.com/gpbduff>



# Venha fazer Mestrado e Doutorado na UFF!!

