

# Escrita e Apresentação de Artigos Científicos

Mirella M. Moro  
[mirella@dcc.ufmg.br](mailto:mirella@dcc.ufmg.br)

Realização



DEPARTAMENTO DE CIÊNCIA DA  
COMPUTAÇÃO



PROGRAMA DE POS-GRADUAÇÃO  
EM CIÊNCIA DA COMPUTAÇÃO

**UFMG**

Apoio



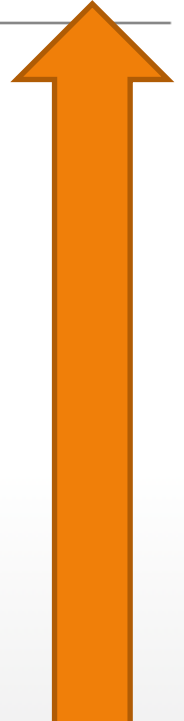
## MIRELLA M. MORO

HOME / CALL FOR PAPERS / ALUNOS

### CONTACT INFORMATION

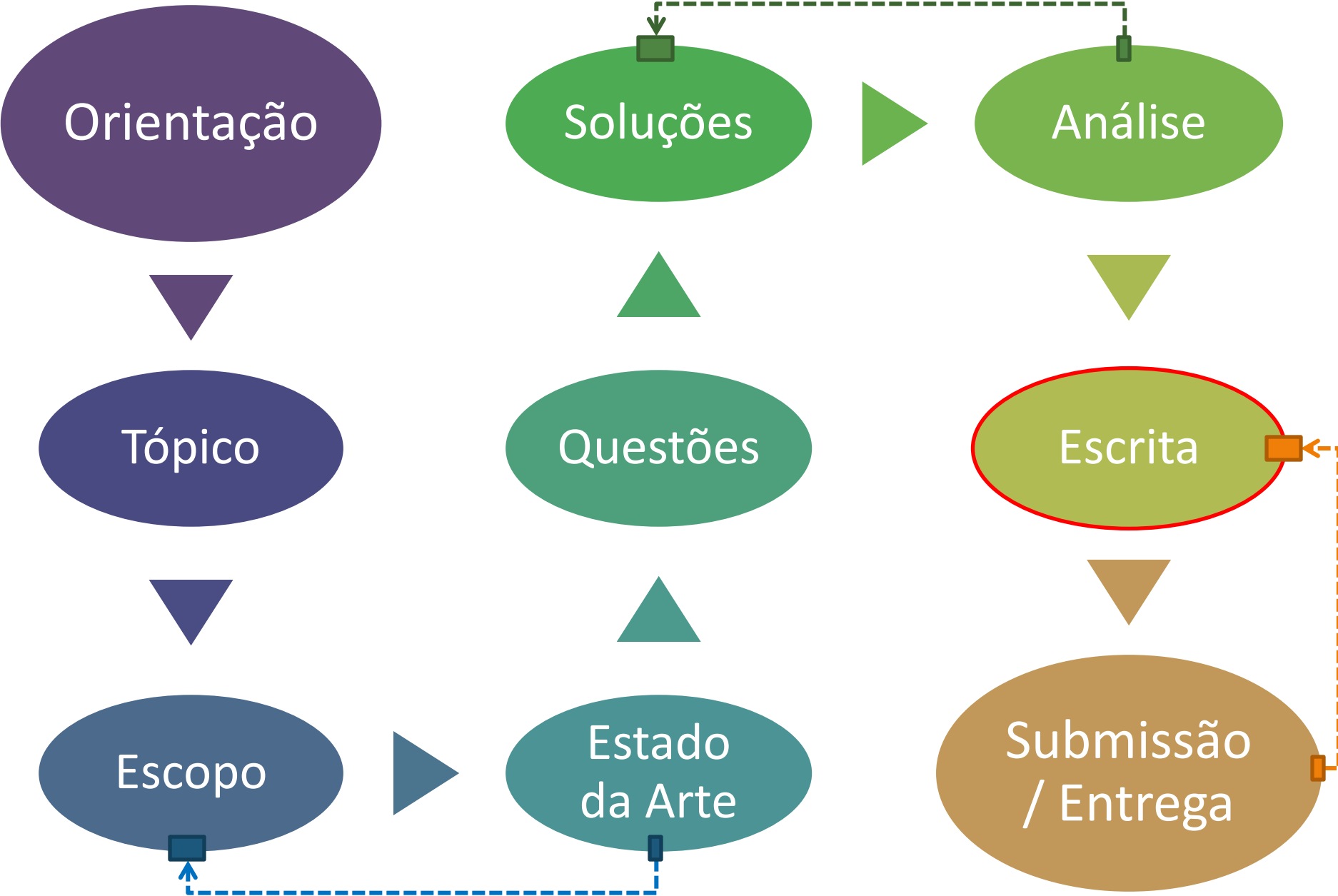
**Mirella M. Moro**

Departamento de Ciência da Computação  
Universidade Federal de Minas Gerais - UFMG  
Av. Antônio Carlos, 6627  
Pampulha, Belo Horizonte - MG, Brazil  
CEP 31270-010  
Office: Building ICEX, room 4012  
Phone numbers:  
+55 (31) 3409-5849 (Office)  
+55 (31) 3409-5858 (Fax)  
Email: mirella [AT] dcc [DOT] ufmg [DOT] br



PARTE 1

# **ESCRITA CIENTÍFICA**

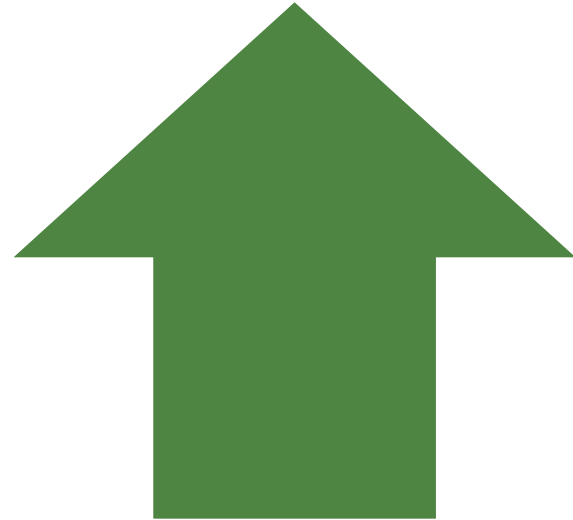




Pesquisa  
+ Escrita



Planejamento



# As complexidades da tarefa

- Como tornar um interesse vago em um **problema merecedor** de apresentação e solução
- Como construir um argumento que **motiva leitores** a aceitar o que você diz
- Como antecipar dúvidas de atentos mas **críticos leitores** e como respondê-las apropriadamente
- Como criar uma introdução e conclusão que respondam à **pergunta mais difícil delas, *E eu com isso?***
- Como ler sua própria escrita como outros irão, e então **aprender quando e como revisá-la**

# Receita de Planejamento

## A. Defina o OBJETIVO

- Tópico: *Estou pesquisando* \_\_\_\_
- Questão: *a fim de descobrir o que/como/por que* \_\_\_\_
- Relevância: *para* \_\_\_\_
- Aplicação prática potencial: *que pode ser utilizada em* \_\_\_\_

## B. Defina o ESPECTRO

- < 4 palavras: muito abrangente

## C. Entenda o PÚBLICO ALVO

- Quem? Background? Expectativas?

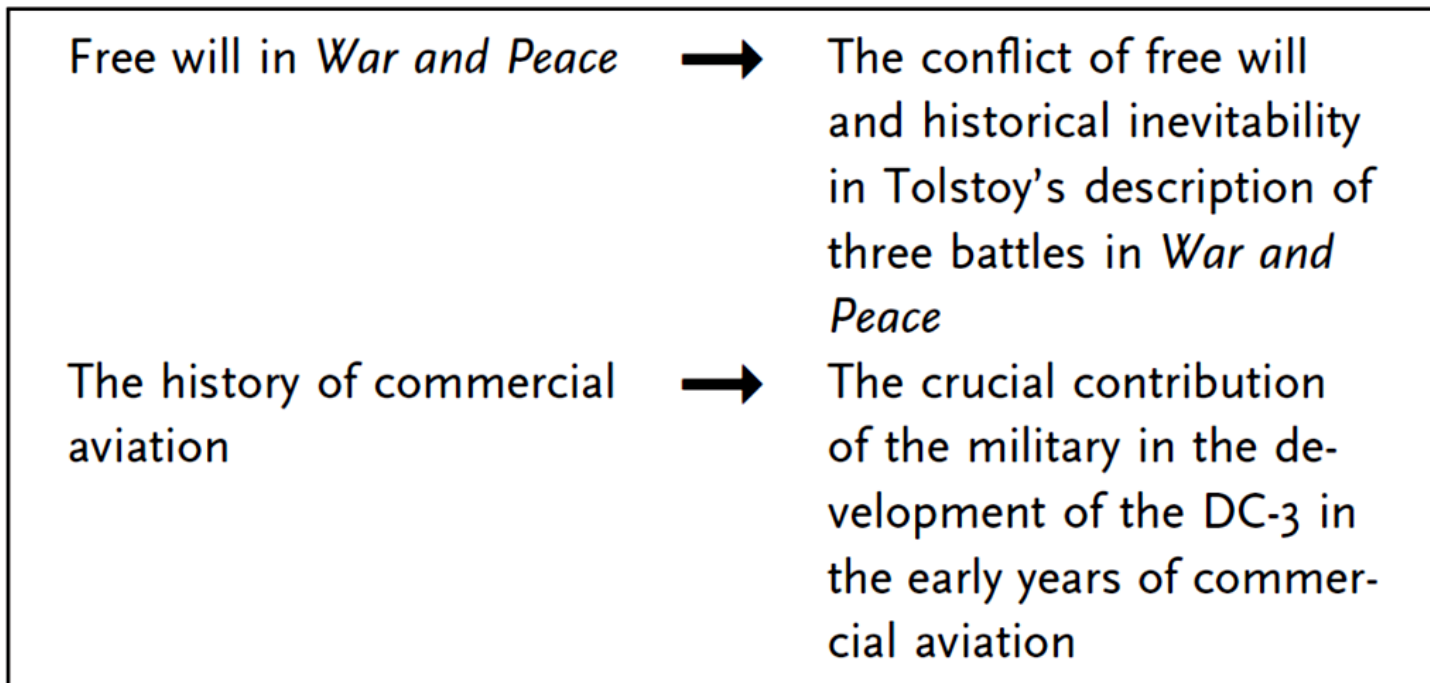
# Exemplo: A. Defina o OBJETIVO

1. **Topic:** I am studying the differences among various nineteenth-century versions of the story of the Alamo
2. **Question:** because I want to find out how politicians used stories of great events to shape public opinion,
3. **Conceptual Significance:** in order to help readers understand how politicians use elements of popular culture to advance their political goals,
4. **Potential Practical Application:** so that readers can better protect themselves from unscrupulous politicians.



# Exemplo: B. Defina o ESPECTRO

- Com um tópico muito abrangente, você pode ficar intimidado pela ideia de encontrar, muito menos ler, sequer **uma fração das fontes disponíveis**
- Então, deve-se reduzi-lo



# Exemplo: C. Entenda o Público Alvo

- **Quem lerá seu texto?**

Profissionais? Alunos? Pesquisadores? Torcida do Atlético?

- **A expectativa é**

Diverti-los? Apresentar novos conhecimentos? Ajudar a entender algo melhor?

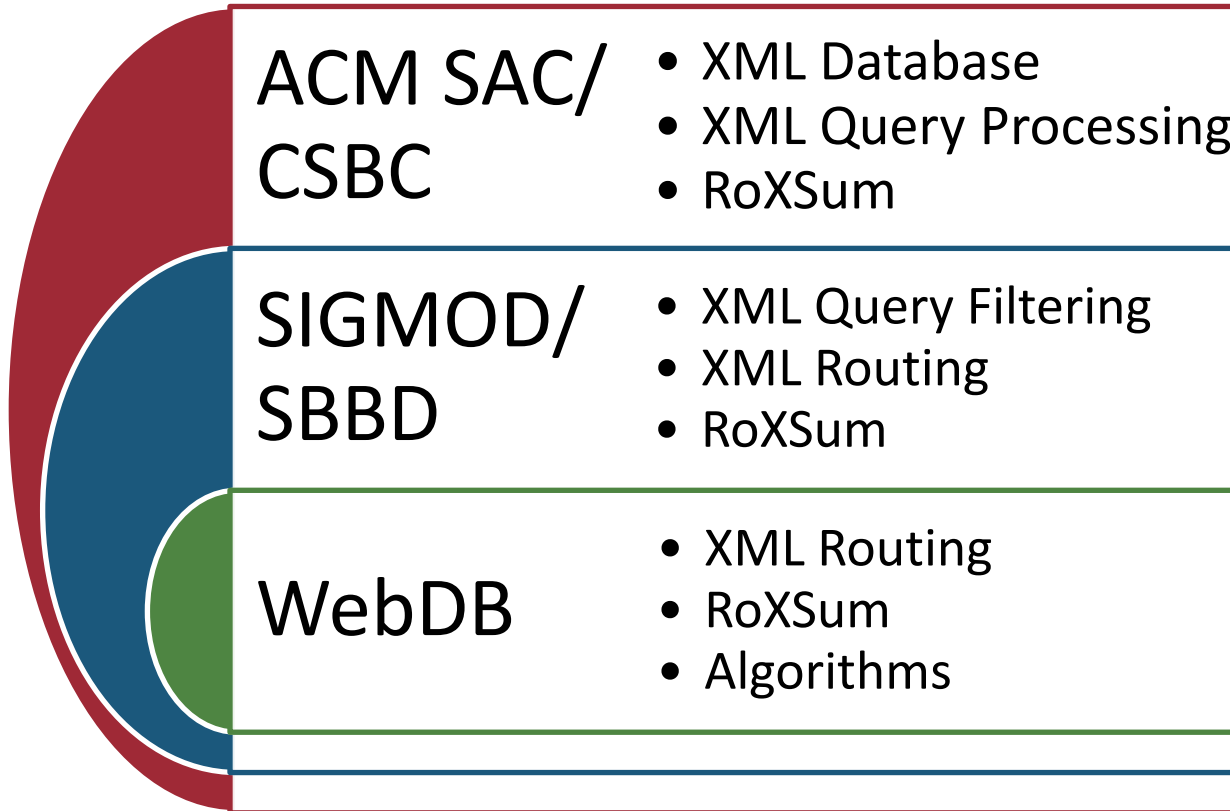
- **O que eles sabem sobre o tópico?**

Eles têm algum interesse especial? O problema já é conhecido? É um problema existente mas que ainda não reconhecem? É um problema seu e não deles?

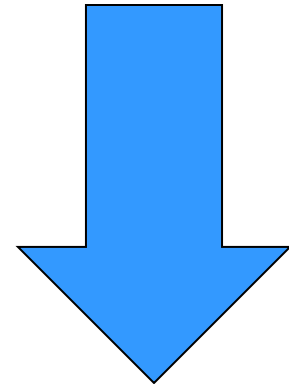
- **Como os leitores responderão ao seu trabalho?**

Irá contradizer o que eles já acreditam? Como? Eles terão argumento padrão contra a solução?

# Objetivo+Espectro+Público: Exemplo Real



**+GENÉRICO**  
**+BACKGROUND**



**+ESPECÍFICO**  
**+DETALHES**

# Planejamento em 1 slide

## PRONTO

- Título
- Autor (es)
- Resumo
- Introdução
- Corpo
- Conclusão
- Referências

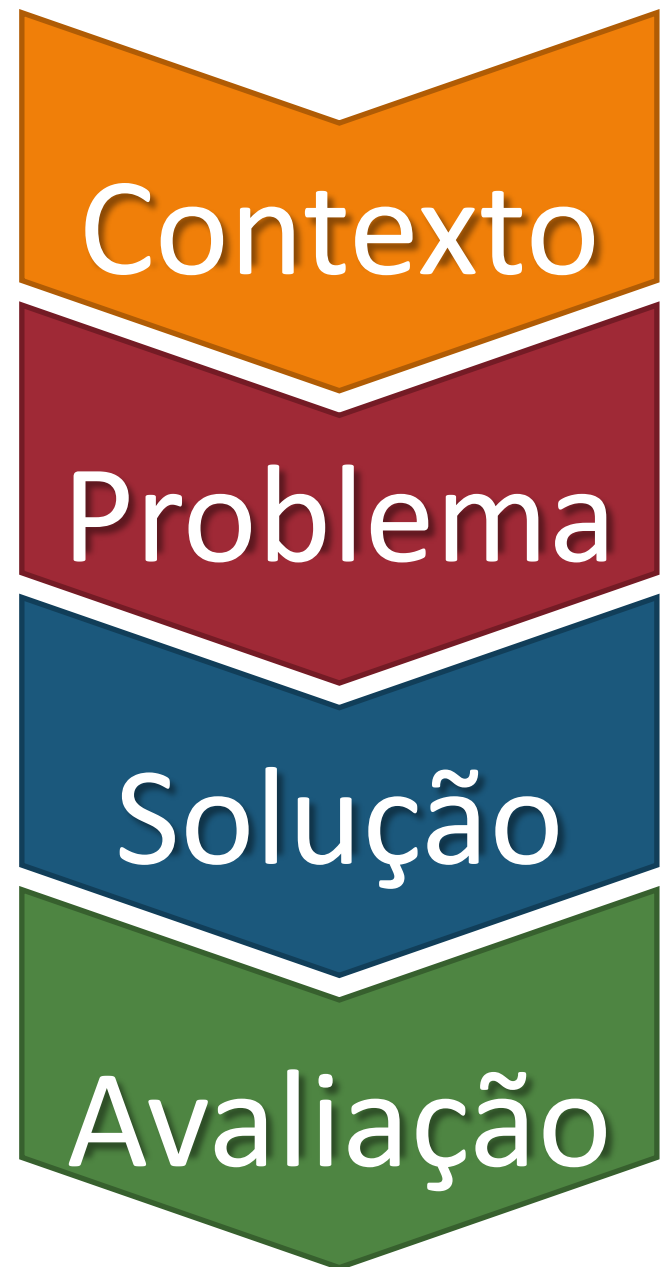
## A FAZER

- **Onde? O que? Por quê? Para quê? Como?**
- **Quanto? “largura e fundura”**
- **Quem irá ler seu texto?**
  - evento, periódico, orientador, alunos, banca
  - **Eles esperam que faça o que pretende fazer?**  
contexto, novidade, contribuição, surpresa (*twist*)
  - **Quanto se espera que eles saibam?**  
profundidade/amplitude, background, público
  - **Como responderão aos problemas e soluções apresentados?**

# Enquanto estiver

- Escrevendo
- Apresentando
- Pensando
- Planejando
- Discutindo
- Reunindo
- Argumentando

**CON-PROSA:**



## Contexto

- Contexto geral
- Contexto específico [estado da arte]

## Problema

- O que \*não\* existe/funciona?
- O que pode melhorar?

## Solução

- Contribuição
- Como resolver o problema

## Avaliação

- É bom? É melhor?
- [Nem] Sempre: Quando? Por quê?

# Contribuições por TUDO!!!!

Contexto

- $X \text{ atual} > \text{ novo contexto}$

Problema

- $X \text{ atual} > \text{ novo problema}$

Solução

- $X \text{ atual} > \text{ nova solução}$
- $\text{Solução atual} > \text{ melhorias}$

Avaliação

- $\text{Diferentes avaliações} > X \text{ atual}$

Onde  $X \in \{ \text{contexto, problema, solução, avaliação} \}$

# Bazinga





# Contribuições por TUDO!!!!

## Contexto

- Dados relacionais > streams
- Dados relacionais > móveis

## Problema

- Dados relacionais > semiestruturado
- Modelo relacional > big data analysis

## Solução

- Mais comum

## Avaliação

- Escolha qualquer um e avalie em: paralelo, distribuído, grandes volumes, map-reduce, novos hardwares, ...

# Exemplo [ ZHANG et al @ SIGMOD 1996 ]

- Contexto** Finding useful patterns in large datasets has attracted considerable interest recently,
- Problema** 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 adequately address the problem of large datasets and minimization of I/O costs.
- Solução** 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.
- (Detalhes)** 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.
- Avaliação** 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.**

# Não é só em BD

## RAGHAVAN et al @ SIGCOMM 2007

**Contexto** Today's cloud-based services integrate globally distributed resources into seamless computing platforms.

**Problema** Provisioning and accounting for the resource usage of these Internet-scale applications presents a challenging technical problem.

**Solução** This paper presents the design and implementation of distributed rate limiters, which work together to enforce a global rate limit across traffic aggregates at multiple sites, enabling the coordinated policing of a cloud-based service's network traffic.

**(Detalhes)** Our abstraction not only enforces a global limit, but also ensures that congestion-responsive transport-layer flows behave as if they traversed a single, shared limiter. We present two designs—one general purpose, and one optimized for TCP—that allow service operators to explicitly trade off between communication costs and system accuracy, efficiency, and scalability.

**Avaliação** Both designs are capable of rate limiting thousands of flows with negligible overhead (less than 3% in the tested configuration). We demonstrate that our TCP-centric design is scalable to hundreds of nodes while robust to both loss and communication delay, making it practical for deployment in nationwide service providers.

# Escrita Científica

## PRONTO

- Título
- Autor (es)
- Resumo
- Introdução
- Corpo
- Conclusão
- Referências

## A FAZER

- Título
- Autor (es)
- Resumo
- Introdução
- Corpo
- Conclusão
- Referências

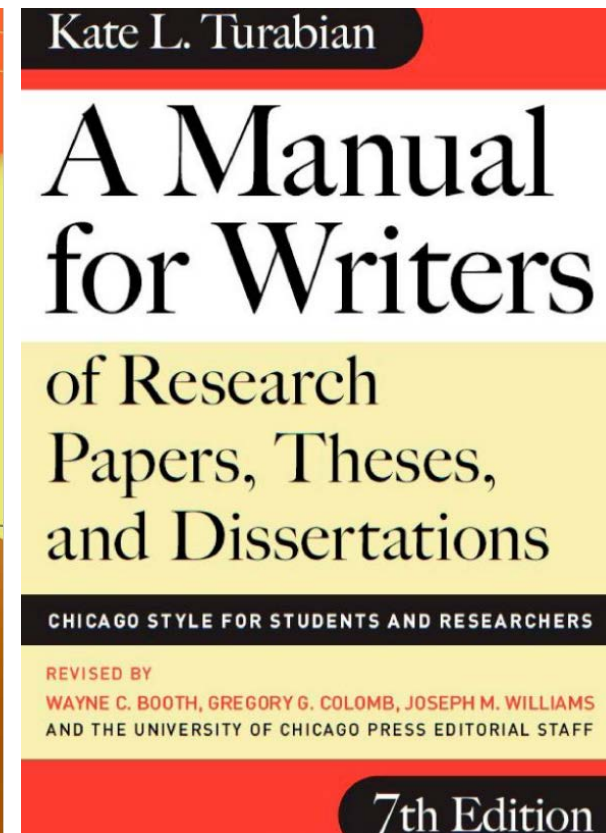
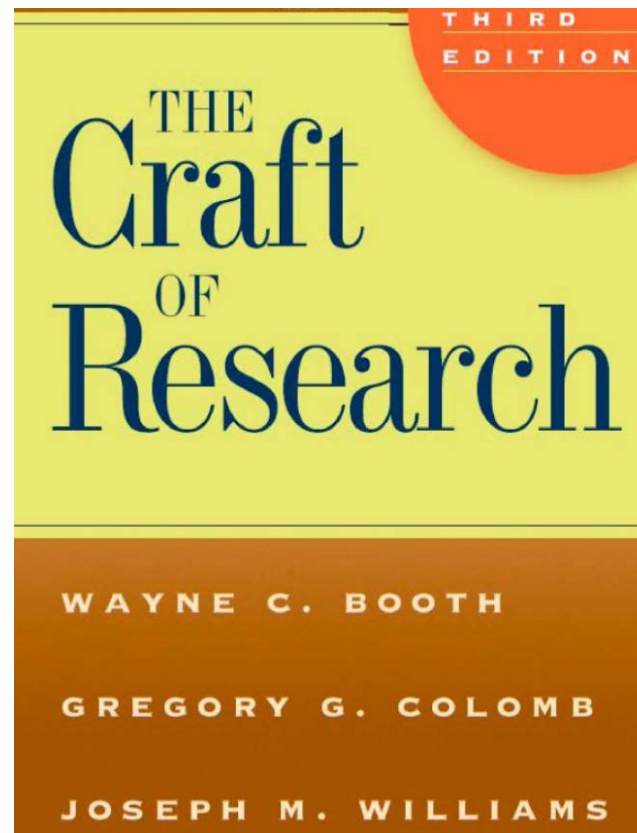
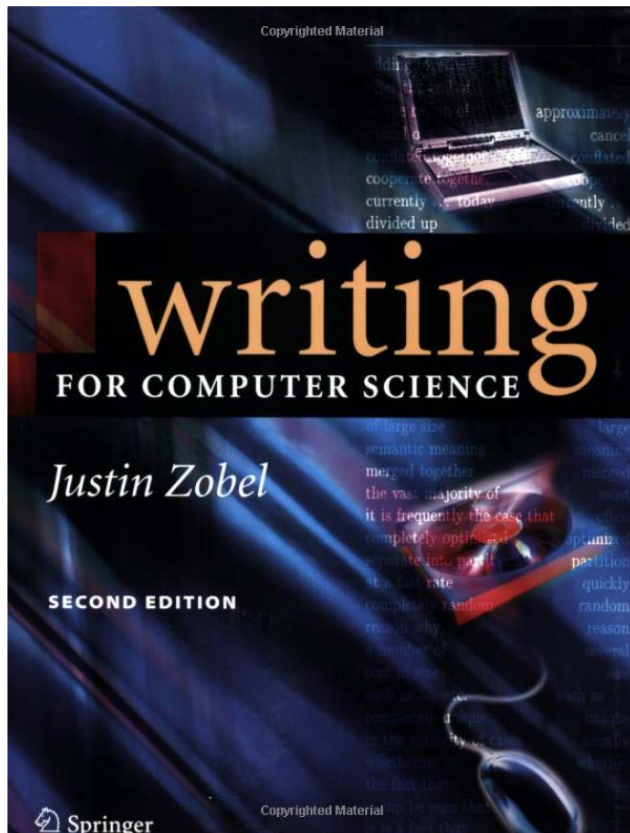


**ERGO**

**ESTE CURSO**



# PS: vá à (minha) biblioteca



## ROTEIRO

- 2. ESCRITA: INÍCIO**
- 3. ESCRITA: CENTRO**
- 4. ESCRITA: FIM**
- 5. APRESENTAÇÃO DE ARTIGOS**
- 6. EXEMPLOS SBBD 2014**
- 7. FINALMENTE**

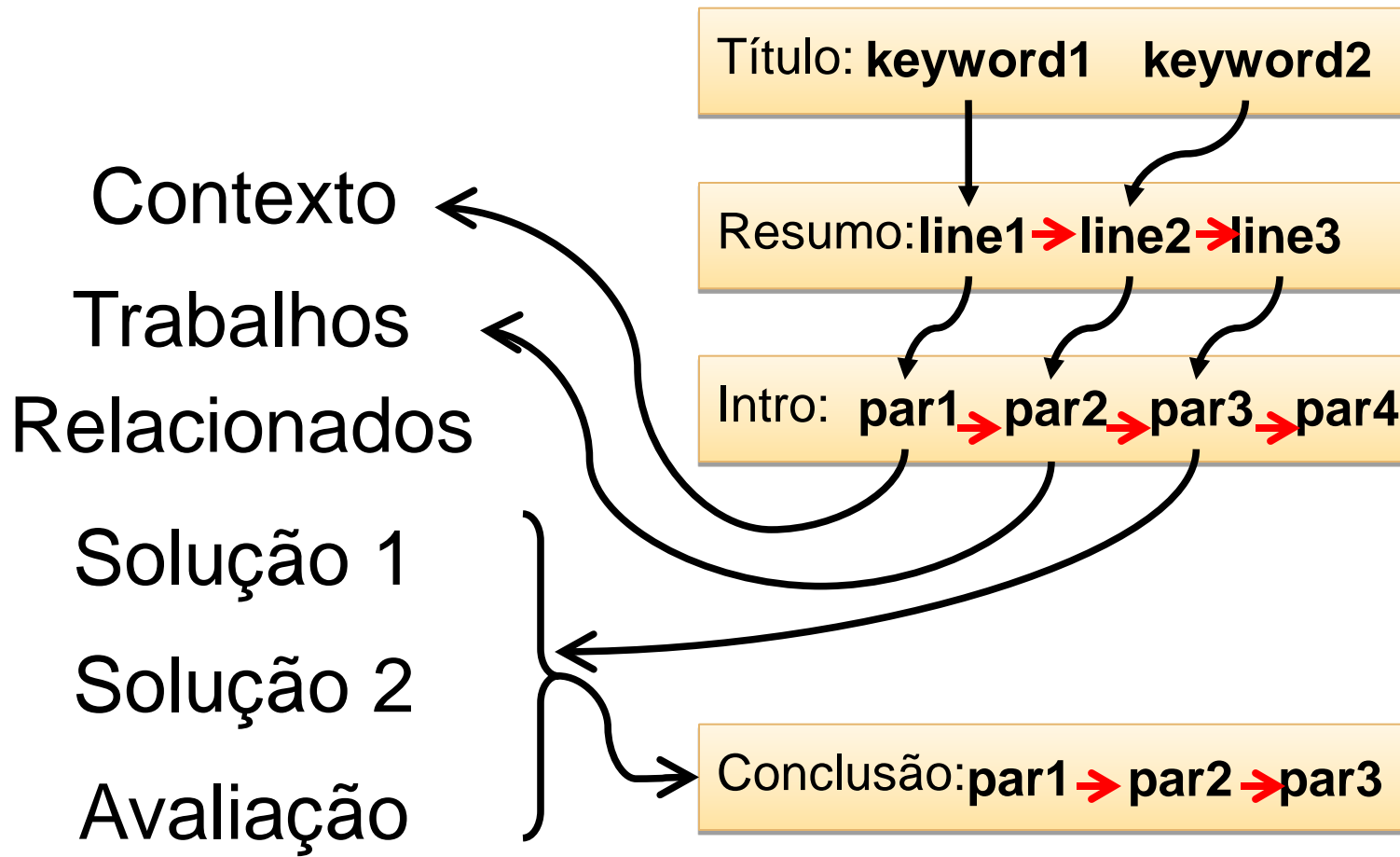
Artigos Científicos  
Conferências  
Periódicos  
Monografia  
Dissertação  
Tese  
Relatório Técnico  
Trabalho p/ Disciplina



PARTE 2

# **COMPONENTES INICIAIS DO ARTIGO CIENTÍFICO**

# Componentes Fluxo



# TÍTULO



- Referência principal ao trabalho
- Chave para ser referenciado
- Claro, curto, correto
  - Nome, não uma frase, original
  - Primeira coisa a se escrever??

# Título: exemplos brasileiros

- On the automatic design of decision-tree induction algorithms  
R.C. BARROS @ SBC CTD 2014 – 1º lugar doutorado
- Unsupervised Information Extraction by Text Segmentation  
E. CORTEZ @ SBC CTD 2013 – 1º lugar doutorado
- Data Mining in Large Sets of Complex Data  
R. CORDEIRO @ SBC CTD 2012 – 1º lugar doutorado
- Análise de Dados de expressão gênica: normalização de microarrays e modelagem de redes regulatórias  
A.FUJITA @ SBC CTD 2008 – 1º lugar doutorado
- Low Cost BIST Techniques for Linear and Non-Linear Analog Circuits  
M. NEGREIROS @ DATE 2006 – dissertation award
- Updating relations through XML Views  
V. BRAGANHOLO @ SBC CTD 2005 – 1º lugar doutorado

# Utilizar um nome ajuda...

Jon's Blog:

Today my cat viciously attacked and devoured my precious pet goldfish.



Who knew that my cat was capable of such an evil and unspeakable deed?



I DO HAVE A NAME, YOU KNOW

SMACK!



# Título: exemplos

## Clássicos e Novos Clássicos

- **The Entity-Relationship Model**: Toward a Unified View of Data PETER CHEN @ VLDB 1975
- **Marching cubes**: A high resolution 3D surface construction algorithm W.E. LORENSEN @ SIGGRAPH 1987
- **BIRCH**: An Efficient Data Clustering Method for Very Large Databases T. ZHANG et al @ SIGMOD 2006
- **Bigtable**: A distributed storage system for structured data F. CHANG et al @ ACM TOCS 2008
- **MapReduce**: simplified data processing on large clusters J. DEAN & S. GHEMAWAT @ CACM 2008

# Trabalhar em grupo é uma faca de dois gumes



# AUTOR (es)

- Nome completo (“artístico”) + filiação + email
- Ordem dos autores??
  - Não há regra padrão aceita globalmente

- **André Silva, Bento Muniz, Carlos Costa**
- **Carlos Costa, Bento Muniz, André Silva**
- **Bento Muniz, Carlos Costa, André Silva**

Aluno “dono” tese

Bolsista

Orientador



# www.phdcomics.com

## THE AUTHOR LIST: GIVING CREDIT WHERE CREDIT IS DUE

**The first author**  
Senior grad student on the project. Made the figures.

**The third author**  
First year student who actually did the experiments, performed the analysis and wrote the whole paper. Thinks being third author is "fair".

**The second-to-last author**  
Ambitious assistant professor or post-doc who instigated the paper.

Michaels, C., Lee, E. F., Sap, P. S., Nichols, S. T., Oliveira, L., Smith, B. S.

**The second author**  
Grad student in the lab that has nothing to do with this project, but was included because he/she hung around the group meetings (usually for the food).

**The middle authors**  
Author names nobody really reads. Reserved for undergrads and technical staff.

**The last author**  
The head honcho. Hasn't even read the paper but, hey, he got the funding, and his famous name will get the paper accepted.

# RESUMO

Um parágrafo 150-250 palavras

– Propaganda ou trailer do artigo

– Atrai (?) a atenção e o interesse do leitor

- Sempre menciona informações ou conclusões do texto
- Sem referências bibliográficas (exceto em ocasiões raras, como modificações a um método publicado previamente)
- Primeiro a ser escrito: planejamento
- Último a ser escrito: ideia melhor

# Resumo: ESTILO CURTO

1 OU 2 LINHAS PARA CADA ITEM

<b>Contexto</b>	<u>Structural summaries</u> are data structures that preserve all structural features of XML documents in a compact form. <u>We investigate</u>
<b>Problema/ Solução</b>	<u>the applicability</u> of the most popular summaries as access methods within XML query processing. In this context, <u>issues</u> like space and false positives introduced by the summaries need to
<b>Avaliação</b>	<u>be examined</u> . <u>Our evaluation</u> reveals that the additional space required by the more precise structures is usually small and justified by the considerable performance gains that they achieve.

MORO et al – WWW 2006

# Resumo: ESTILO LONGO

## CONPROSA + outras informações

**Contexto** A Web é abundante em páginas que armazenam dados de forma implícita. Em muitos casos, estes dados estão presentes em textos

**Problema** semiestruturados sem a presença de delimitadores explícitos e organizados em uma estrutura também implícita. Neste artigo apresentamos uma nova abordagem para extração em textos semi-

**Solução** estruturados baseada em Modelos de Markov Ocultos (Hidden Markov Models - HMM). Ao contrário de outros trabalhos baseados em HMM, nossa abordagem dá ênfase à extração de metadados

**Estado-da-Arte** além dos dados propriamente ditos. Esta abordagem consiste no uso de uma estrutura aninhada de HMMs, onde um HMM principal identifica os atributos no texto e HMMs internos, um para cada atributo, identificam os dados e metadados. Os HMMs são gerados a partir de um treinamento com uma fração de amostras da base a ser extraída. Nossos experimentos com anúncios de classificados retirados da Web mostram que o processo de extração alcançáveis

**Avaliação** de qualidade acima de 0,97 com a medida F, mesmo se esta fração de treinamento é pequena.

SANTOS et al @ SBBD 2006

**Contexto Geral** Publish-subscribe applications are an important class of content-based dissemination systems where the message transmission is defined by the message content, rather than its destination IP address. **Contexto Específico** With the increasing use of XML as the standard format on many Internet-based applications, XML aware pub-sub applications become necessary. In such systems, the messages (generated by publishers) are encoded as XML documents, and the profiles (defined by subscribers) as **Problema** XML query statements. As the number of documents and query requests grow, the performance and scalability of the matching phase (i.e. matching of queries to incoming documents) become vital. **Estado-da-arte** Current solutions have limited or no flexibility to prune out queries in advance. In this paper, we overcome such **Solução** limitation by proposing a novel early pruning approach called Bounding-based XML Filtering or BoXFilter. The BoXFilter is based on a new tree-like indexing structure that organizes the queries based on their similarity and provides lower and upper **Solução c/ detalhes** bound estimations needed to prune queries not related to the incoming documents. Our experimental evaluation shows that the early profile pruning approach offers drastic performance **Avaliação** improvements over the current state-of-the-art in XML filtering.

MORO et al @ VLDB 2007

**Contexto e Problema**

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 adequately address the problem of**

**Estado-da-arte**

**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**

**Solução**

**multi-dimensional metric data points to try to produce the best quality**

**Solução c/ detalhes**

**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*

**Vantagens**

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**

**Avaliação**

**clustering quality through several experiments. We also present a**

**performance comparisons of *BIRCH* versus *CLARANS*, a clustering method proposed recently for large datasets, and show that *BIRCH* is consistently superior.**

**ZHANG et al – SIGMOD 1996**

**Contexto** Today's cloud-based services integrate globally distributed resources into seamless computing platforms. **Provisioning and accounting for the**

**Problema** **resource usage of these Internet-scale applications presents a**

**Solução** **challenging technical problem.** This paper presents the design and implementation of distributed rate limiters, which work together to enforce a global rate limit across traffic aggregates at multiple sites, enabling the coordinated policing of a cloud-based service's network traffic. **Our abstraction not only enforces a global limit, but also ensures**

**Solução c/ detalhes** **that congestion-responsive transport-layer flows behave as if they**

**Avaliação** **traversed a single, shared limiter. We present two designs—one general purpose, and one optimized for TCP—that allow service operators to explicitly trade off between communication costs and system accuracy, efficiency, and scalability.** Both designs are capable of rate limiting thousands of flows with negligible overhead (less than 3% in the tested configuration). We demonstrate that our TCP-centric design is scalable to hundreds of nodes while robust to both loss and communication delay, making it practical for deployment in nationwide service providers.

**RAGHAVAN et al – SIGCOMM 2007**

**Contexto** The Elias-Fano representation of monotone sequences has been recently applied to the compression of inverted indexes, showing excellent query performance thanks to its efficient random access and search operations. While its space occupancy is competitive with some state-of-the-art

**Problema** methods such as gamma-delta-Golomb codes and PForDelta, it fails to exploit the local clustering that inverted lists usually exhibit, namely the presence of long subsequences of close identifiers. In this paper we

**Solução** describe a new representation based on partitioning the list into chunks and encoding both the chunks and their endpoints with Elias-Fano, hence forming a two-level data structure. This partitioning enables the encoding to better adapt to the local statistics of the chunk, thus exploiting

**Solução c/ detalhes** clustering and improving compression. We present two partition strategies, respectively with fixed and variable-length chunks. For the latter case we introduce a linear-time optimization algorithm which identifies the minimum-space partition up to an arbitrarily small approximation factor. We show that our partitioned Elias-Fano indexes offer significantly better compression than plain Elias-Fano, while preserving their query time efficiency. Furthermore, compared with other state-of-the-art compressed encodings, our indexes exhibit the best compression ratio/query time trade-off.

**Avaliação**



**Contexto** XML has been explored by both research and industry communities. More than 5500 papers were published on different aspects of XML. **With**

**Problema** so many publications, it is hard for someone to decide where to start. Hence, this paper presents some of the research topics on XML, namely:

**Contribuição** XML on relational databases, query processing, views, data matching, and schema evolution. It then summarizes some (some!) of the most relevant or traditional papers on those subjects.

**MORO et al – SIGMOD RECORD 2009**

**Contexto é um tópico**

We have had to wait over 30 years since the naive Bayes model was first introduced in 1960 for the so-called Bayesian network classifiers to resurge. **Based on Bayesian networks, these classifiers have many strengths, like model interpretability, accommodation to complex data and classification problem settings, existence of efficient algorithms for learning and classification tasks, and successful applicability in real-world problems.**

**Relevância**

**In this article, we survey the whole set of discrete Bayesian network classifiers devised to date, organized in increasing order of structure complexity: naive Bayes, selective naive Bayes, seminaive Bayes, one-dependence Bayesian classifiers,  $k$ -dependence Bayesian classifiers, Bayesian network-augmented naive Bayes, Markov blanket-based Bayesian classifier, unrestricted Bayesian classifiers, and Bayesian multinets. Issues of feature subset selection and generative and discriminative structure and parameter learning are also covered.**

**BIELZA & LARRAÑAGA – COMP. SURVEYS 2014**

**Contribuição ( topologia! )**

**Contexto** Efficient document processing is a must when large volumes of XML data are involved. In such critical scenarios, a well-known solution to this problem is to distribute (map) the data among several processing nodes, and then distribute the processing accordingly, taking advantage of parallelism. **This is the approach taken by distributed databases and MapReduce environments.**

**Relevância** Fragmentation techniques play an important role in these scenarios. They provide a way to "cut" the database into pieces and distribute the pieces over a network.

**Problema** This way, queries can also be "cut" into sub-queries that run in parallel, thus achieving better performance when compared to the centralized environment. However, there is no consensus in the database community as to what an XML fragment is. In fact, several approaches in literature present definitions of XML fragments.

**Aplicações** In addition to query processing, using XML fragmentation techniques may also be helpful when managing XML documents distributed along the web or clouds. This paper surveys the existing XML

**Contribuição** fragmentation approaches in literature, comparing their features and highlighting their drawbacks. Our contribution resides in establishing a map of the area. **BRAGANHOLO & MATTOSO - SIGMOD RECORD 2014**

**Contexto** Middleboxes are ubiquitous in today's networks and perform a variety of important functions, including IDS, VPN, firewalling, and WAN optimization. These functions differ vastly in their requirements for hardware resources (*e.g.*, CPU cycles and memory bandwidth).

**Problema** Thus, depending on the functions they go through, different flows can consume different amounts of a middlebox's resources. While there is much literature on

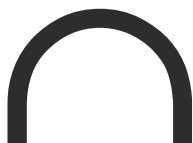
**Estado da arte** weighted fair sharing of link bandwidth to isolate flows, it is unclear how to schedule *multiple* resources in a middlebox to achieve similar guarantees. In this paper, we analyze several natural packet scheduling algorithms for multiple resources and show that they have undesirable properties. We propose a new algorithm, Dominant Resource Fair Queuing (DRFQ), that retains the attractive properties that fair sharing provides for one resource. In doing so, we generalize the concept of virtual time in classical fair queuing to multi-resource settings. The resulting algorithm is also applicable in other contexts where several resources need to be multiplexed in the time domain.

**Contribuição**

**Relevância**

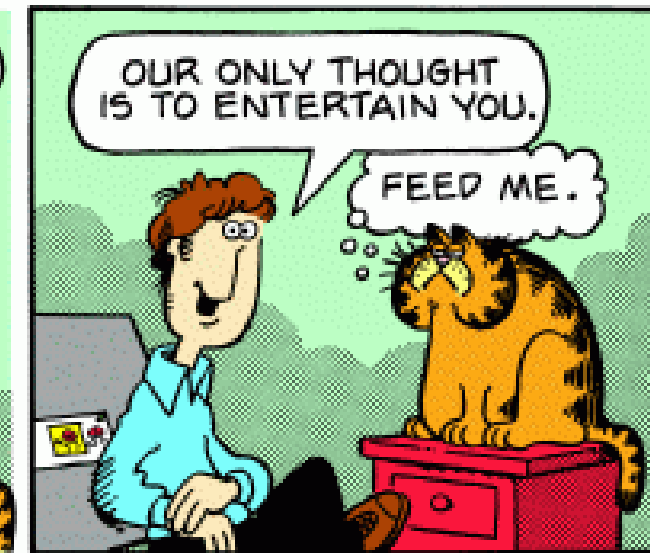
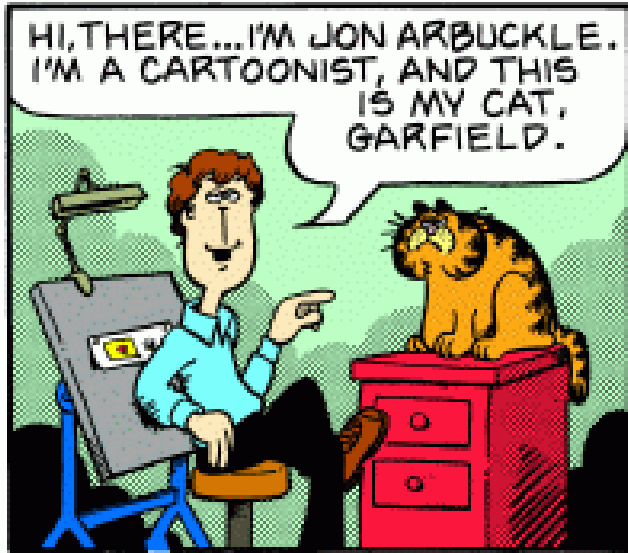
**GHODSI et al – SIGCOMM 2012**

# INTRODUÇÃO



APESAR de serem ambos bem escritos

# Introdução *introduz* 😊



# Introdução



**Um artigo científico não é um livro de suspense** no qual o leitor só descobre o que está realmente acontecendo no capítulo final

- Uma introdução bem escrita é fundamental!!
- O leitor deve estar ciente do que acontece desde o início, desde a introdução

# Introdução

A introdução é uma reafirmação estendida do conteúdo do Resumo (linhas → parágrafos)

ERRO MAIS COMUM

Não colocar as contribuições

Deixar tudo “implícito”

Achar que é *óbvio*

***Óbvio não é ciência,  
é senso comum***



# Introdução

Em algum lugar (na introdução), adicionar exemplos (1 ou mais):

- Como o seu trabalho pode ser empregado
- Onde o seu trabalho pode ser empregado, quais os contextos, quais aplicações podem se beneficiar
- Qual a contribuição social
- Quais problemas práticos o seu trabalho resolve

# Introdução

- Contexto
- Problema
- Relacionados
- Solução  
(Organização)

Contexto, motivação

Evolução de um contexto

“**Yesterday’s** version of distributed computing was a self-contained, colocated server farm. **Today**, applications are increasingly deployed on third-party resources hosted across the Internet. Indeed, the rapid spread of open protocols and standards like Web 2.0 has fueled an explosion of compound services that script together third-party components to deliver a sophisticated service [27, 29]. These specialized services are just the **beginning**: flagship consumer and enterprise applications are increasingly being delivered in the software-as-a-service model [9]. **For example**, Google Documents, Groove Office, and Windows Live are early examples of desktop applications provided in a hosted environment, and represent **the beginning of a much larger trend.**”

RAGHAVAN et al – SIGCOMM 2007

# Introdução

- Contexto
- Problema
- Relacionados
- Solução  
(Organização)

## O problema em questão

- Definição do problema
- Sua importância, relevância, aplicações práticas

“One of the **key barriers** to moving traditional applications to the cloud, however, is the loss of cost control [17]. In the cloud-based services model, cost recovery is typically accomplished through metered pricing. Indeed, **Amazon’s EC2** charges incrementally per gigabyte of traffic consumed [3] [...] Limiting global resource consumption in a distributed environment, however, presents a **significant technical challenge**. **Ideally**, resource providers would not require services to specify the resource demands of each distributed component *a priori*; *such fine-grained measurement and modeling can be challenging* for rapidly evolving services. **Instead**, they should provide a fixed price for an aggregate, global usage, and allow services to consume resources dynamically across various locations, subject to the specified aggregate limit.”

RAGHAVAN et al – SIGCOMM 2007

# Introdução

- Contexto
- Problema
- Relacionados
- Solução  
(Organização)

## Estado-da-arte / Trabalhos relacionados

- No mesmo contexto, não resolvem o problema ou apresentam apenas soluções parciais
- Extensão ou continuação de um trabalho anterior: deve ser mencionado na introdução
- Uma frase sobre cada trabalho ou
- Agrupar trabalhos similares e detalhar um ou dois:  
“... Como resposta a tal requisito, alguns trabalhos têm focado a questão do suporte a versões [2,4,9,13,23,27]. Entre esses, Golendziner propõe o Modelo de Versões: uma extensão aplicável a modelos de dados orientado a objetos ...  
... [9]”

# Introdução

- Contexto
- Problema
- Relacionados
- ➔ Solução  
(Organização)

Solução: Contribuições principais do artigo

- Por extenso em um parágrafo  
“Considerando o contexto atual, esse trabalho propõe ...”
- Delimitado por itens

“This paper makes three primary contributions:

- **Rate Limiting Cloud-based Services**. We identify a key challenge...
- **Distributed Rate Limiter Design**. We present the design and ...
- **Evaluation and Methodology**. We develop a methodology...”

RAGHAVAN et al – SIGCOMM 2007



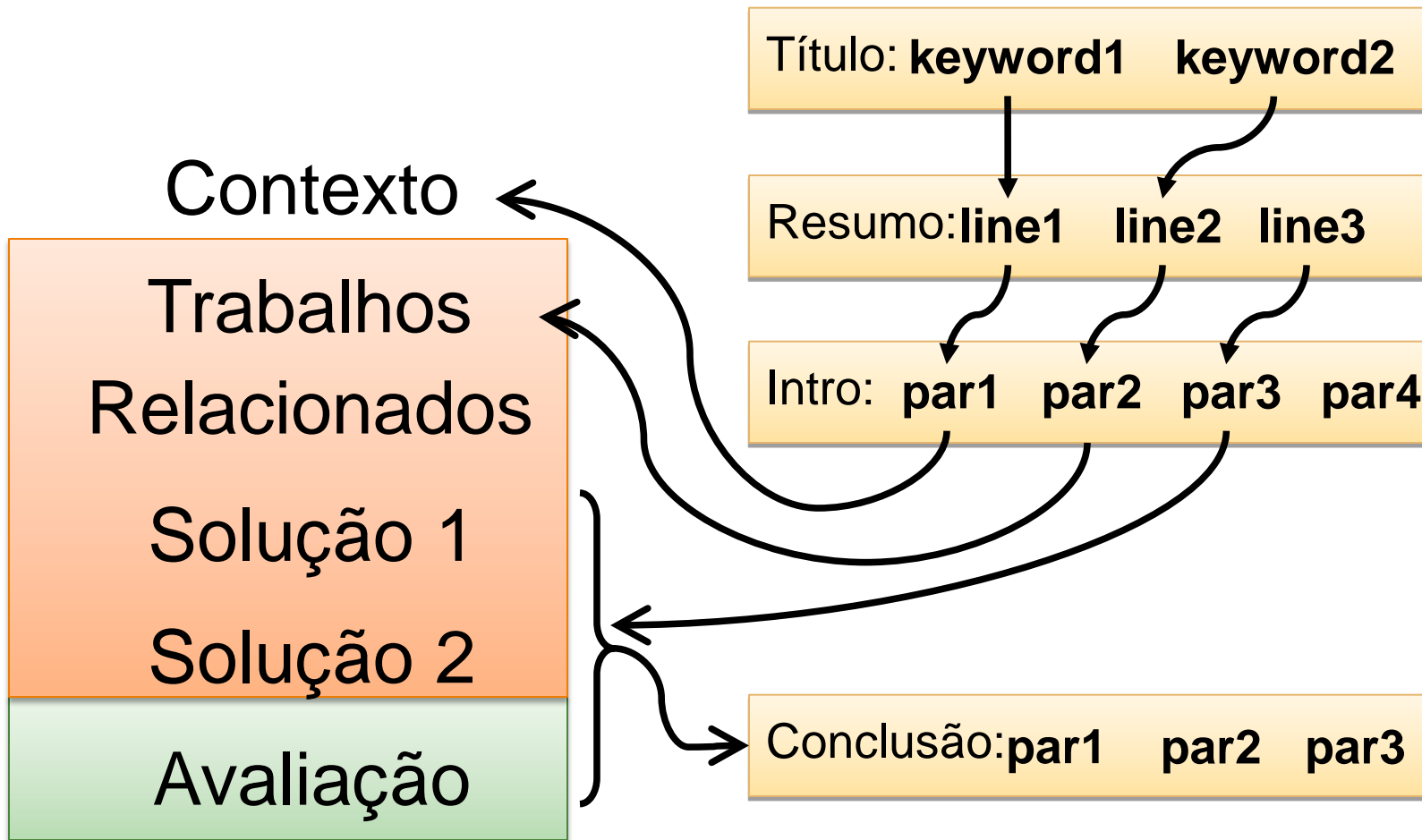
# Introdução

- Organização [opcional?]

- Contexto
- Problema
- Relacionados
- Solução
- Organização

“O restante do artigo está organizado da seguinte maneira. A seção 2 apresenta alguns conceitos básicos e discute trabalhos relacionados. A seção 3 detalha o modelo proposto. A seção 4 apresenta um estudo comparativo através de experimentos, enquanto a seção 5 conclui o trabalho.”

# ATÉ AQUI



PARTE 3

**COMPONENTE CENTRAL  
DO ARTIGO CIENTÍFICO**



# Corpo

- N maneiras
- Rascunhos com ordens diferentes
- Revise com seu orientador
- Alterável

# Escrever Artigos como Arte



# Corpo

## Exemplos de Organização

### **ARANTES @ SBBD 2003**

1. Introdução
2. Trabalhos Relacionados
3. Motivação e Conceitos Fundamentais
4. Composição de Operadores por Similaridade: os Novos Algoritmos
5. Experimentos Realizados
6. Conclusões e Trabalhos Futuros

### **BRAGANHOLO @ CTD 2005**

1. Introduction
2. Related Work
3. Query Trees
4. Update Language
5. Mapping
6. Summary and Concluding Remarks

# Corpo

## Exemplos de Organização

### **LORENSEN @ SIGGRAPH 87**

1. Introduction
2. Information flow for 3D medical algorithms
3. Related work
4. Marching cube algorithm
5. Enhancements of the basic algorithm
6. Implementation
7. Results
8. Conclusions

### **OTTAVIANO @ SIGIR 2014**

1. Introduction
2. Background and Notation
3. Related Work
4. Searchable Sequences
5. Experimental Analysis
6. Conclusion and Future Work

# Corpo

## Exemplos de Organização

### **ZHANG @ SIGMOD 1996**

1. Introduction
2. Summary of Relevant Research
3. Background
4. Clustering Feature and CF Tree
5. The BIRCH Clustering Algorithm
6. Performance Studies
7. Summary and Future Research

### **RAGHAVAN @ SIGCOMM07**

1. Introduction
2. Classes of Clouds
3. Limiter Design
4. Evaluation Methodology
5. Evaluation
6. Related Work
7. Conclusion

# Corpo

## Exemplos de Organização

### **CHANG et al@ ACM TOCS 08**

1. Introduction
2. Data Model
3. API
4. Building Blocks
5. Implementation
6. Refinements
7. Performance Evaluation
8. Real Applications
9. Lessons
10. Related Work
11. Conclusions
12. Acknowledgements
13. References

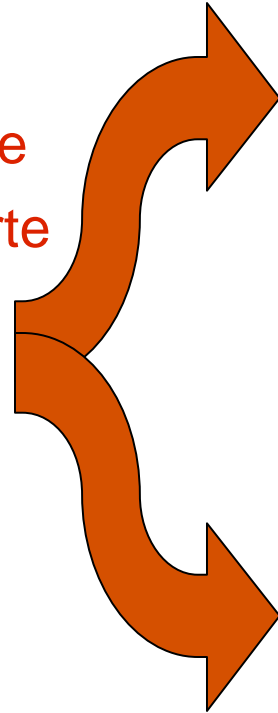
### **DEAN,GHEMAWAT @ CACM 08**

1. Introduction
2. Programming Model
3. Implementation
4. Refinements
5. Performance
6. Experience
7. Related Work
8. Conclusions
9. Acknowledgements
10. References

# Corpo

## Exemplos de Organização

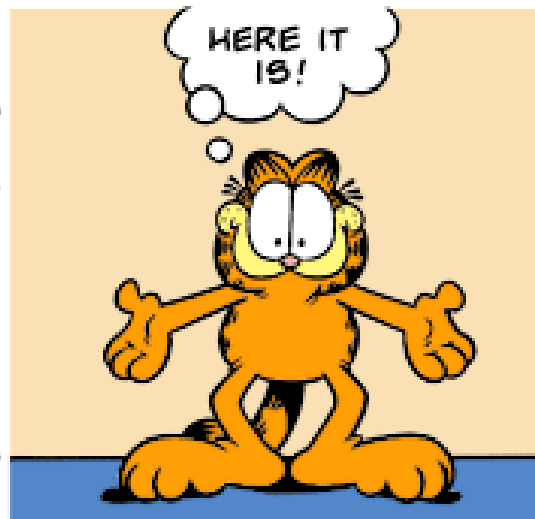
1. Introdução
2. O que já existe  
Estado-da-arte
3. **NOVIDADE**
4. **VALIDAÇÃO**
5. Conclusão



- A. Conceitos Básicos
- B. MODELO**
- C. LINGUAGEM; IMPLEMENTAÇÃO**
- D. ESTUDO DE CASO**
- E. Trabalhos Relacionados

- A. Conceitos Básicos
- B. Trabalhos Relacionados
- C. ARQUITETURA; ALGORITMOS**
- D. ANÁLISE COMPARATIVA**
- E. EXPERIMENTOS**
- F. Discussão

# *Verificar o que já existe evita reinventar o passado*





# O que já Existe

- Já existe
- Novidade
  - Validação
  - Discussão



- Conceitos Básicos e Trabalhos Relacionados
- Apresentados juntos/não
- Conceitos Básicos **antes** da contribuição principal
- Trabalhos Relacionados no início ou fim

# O que já Existe: Conceitos Básicos

- Definições
  - Notações
  - Modelos
  - Arquiteturas
  - Linguagens
  - Cenários
  - Padrões
- } necessários  
para entender o  
trabalho

+ Referências para trabalhos onde os conceitos são introduzidos ou melhor detalhados

# O que já Existe: Trabalhos Relacionados



Trabalhos em temas relacionados 😊

```
graph TD; A[Trabalhos em temas relacionados 😊] --> B[Cobrem temas relacionados mas não exatamente; ou contexto]; B --> C[Detalhes deles evidenciam que o seu é melhor/mais completo]; C --> D[Desvantagens/pontos fracos desses são aprimorados]; D --> E[Texto <-> Referências];
```

Cobrem temas relacionados mas não exatamente; ou contexto

Detalhes deles evidenciam que o seu é melhor/mais completo

Desvantagens/pontos fracos desses são aprimorados

Texto <-> Referências

# O que já Existe: Trabalhos Relacionados

- **Seção de Trabalhos Relacionados é praticamente obrigatória**
- **Como o seu trabalho se relaciona a todos os mencionados**
- Por exemplo, o título é:

“Resumos Estruturais em Processamento de Consulta XML”

Exemplos de temas relacionados:

- Resumos estruturais em outros contextos (outros processamentos)
- Resumos estruturais para processar outros tipos de dados
- Outras formas de processar consultas XML
- ...

# O que já Existe: Trabalhos Relacionados

- Mas o artigo é um survey, precisa de trabalhos relacionados??
  - **SIM**
  - É impossível que um artigo referencie **todos** os outros
  - Que outros trabalhos ficaram fora do survey e por quê?
- Precisa ter uma seção para Trabalhos Relacionados?
- Não pode ir simplesmente mencionando os artigos aqui e ali no meio do artigo?
  - **Melhor não**
  - Melhor ter todos os artigos em uma seção
  - **COMPARANDO** com o trabalho em questão

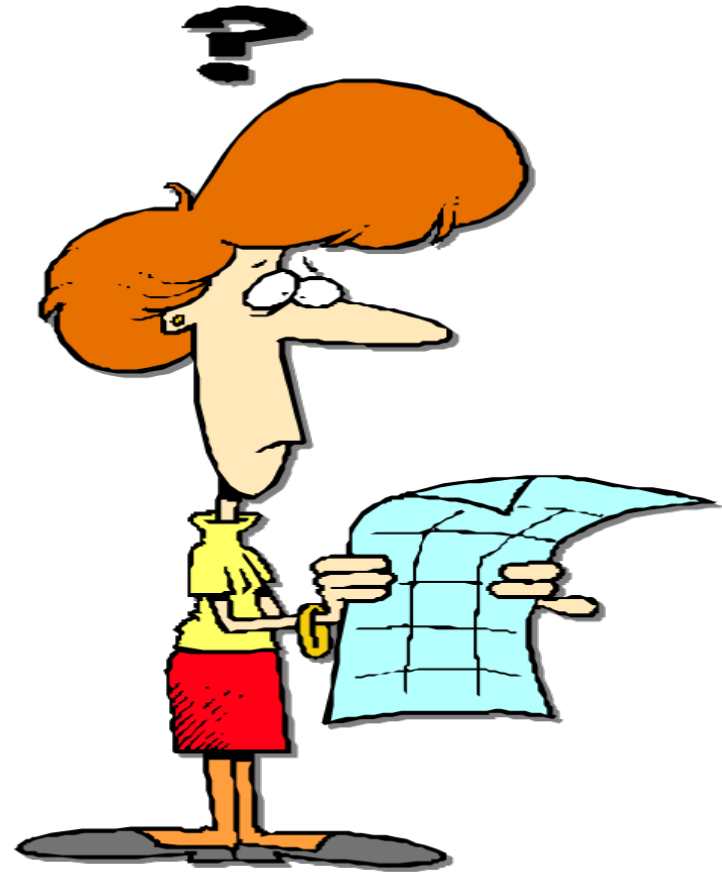
# Cuidado!!!!!!!!!!!!



# Novidade: Contribuições

- Já existe  
→ Novidade
- Validação
- Discussão

- Parte principal do artigo!!!!
- A que veio
- Adiciona o que
- Tem de estar claríssimo





# Novidade: Contribuições

- Um parágrafo com ideia geral da proposta
- Esclareça novas definições (escreva claramente que são novas definições propostas no artigo)
- Adicione quantos parágrafos necessários para apresentar:
  - o que é o trabalho
  - como funciona a proposta
  - o que é novidade, por que
  - detalhes e explicações sobre partes principais do funcionamento da proposta

# EXPLICITAMENTE

O que

Como

Por que

Onde

Quanto

PARTE 4

**FINALIZANDO A ESCRITA**

# Validação / Avaliação

- Já existe
- Novidade
- Validação
- Discussão

- Mostrar que a solução proposta funciona e seus benefícios
- Análise
- Estudo de caso
- Experimentos



# Validação: **Análise**

- **Proposta é correta (demonstração e provas)**
- **Inclua (conforme necessário):**
  - Um parágrafo com o resumo do que é provado nessa seção
  - Um parágrafo com definições específicas usadas na análise (ex. estruturas usadas nas provas)
  - Provas e análises
  - Comentários finais sobre o significado das provas de um modo intuitivo ou num nível mais prático

# Validação: **Estudo de Caso**

- **Proposta é aplicável, implementável**
- Inclua (conforme necessário):
  - Contexto geral
  - Regras ou condições específicas necessárias nesse estudo de caso caso
  - Modelagem / Implementação
  - Funcionamento
  - Vantagens e desvantagens de usar o modelo proposto nesse estudo de caso

# Validação: **Experimentos**

- **Proposta funciona, desempenho bom/superior**
- Inclua (conforme necessário):
  - Contexto: o que é medido, o que não é, por que
  - Modelo de simulação ou infraestrutura das medições: configuração do sistema, tipo de máquinas usadas, linguagens, ...
  - Resultados dos experimentos [próximo slide]
  - Comentários finais, discussões, explicações adicionais

# Validação: Experimentos

## Descrição de resultados

- **Subtítulo** (em negrito): para diferenciar experimentos (ex: avaliando tamanho da entrada, variando a quantidade de consultas, usando dados sintéticos, usando dados reais, ...)
- **Razão clara**: por que cada gráfico aparece no seu artigo (ex: conforme o tamanho dos arquivos de entrada aumenta, o throughput diminui, ...)
- **Explicar**: os eixos dos gráficos, o que o gráfico mostra, qual é a tendência, por que a tendência aparece, por que um algoritmo apresenta melhores resultados que outro, ...
- **Auto-contido**: legendas devem ser legíveis e compreensíveis e suficientes para entender o gráfico



# Experimentos: Exemplo

## 6. Performance studies

6.1 Analysis

6.2 Synthetic Dataset Generator

6.3 Parameters and Default Setting

6.4 Base Workload Performance

6.5 Sensitivity to Parameters

6.6 Time Scalability

6.7 Comparison of BIRCH and CLARANS

6.8 Application to Real Datasets

**ZHANG @ SIGMOD 1996**

# Experimentos: Moral da História

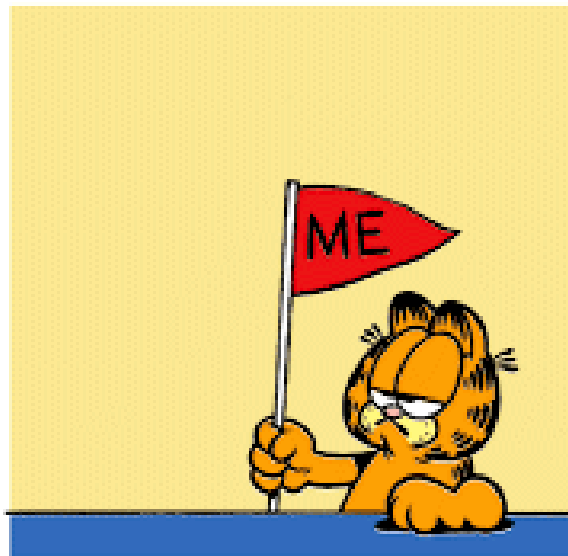


# Discussão

- Já existe
- Novidade
- Validação
- ➔ Discussão

- Discussão
  - Pode ser incluída como subseção final de Experimentos ou na Conclusão
  - Relacionamentos entre os fatos e resultados observados
  - Princípios, relações, generalizações mostrados nos Experimentos
  - Exceção ou falta de relação, pontos incertos
  - Mostrar que resultados e interpretações concordam (ou contrastam) com trabalhos previamente publicados
  - Implicações teóricas e possíveis aplicações práticas
- Conclusão
  - Principal contribuição
  - **Evidências para cada conclusão (não assuma que o leitor é super capaz de juntar todos os pontos sozinho)**

# *Você = maior fã do seu trabalho*



# CONCLUSÃO

Sugestão: ser mais específico que na introdução e informar (um parágrafo/linha por item)

- resumo do que o artigo apresentou
- principais resultados e contribuições, com devidas evidências
- comentários sobre a importância, relevância ou
- dicas para o uso prático do seu trabalho (como os resultados dos experimentos podem ajudar na prática...)
- trabalhos futuros (**evite entregar suas ideias de trabalhos mais inovadores de graça!!**)

# Conclusão: exemplo

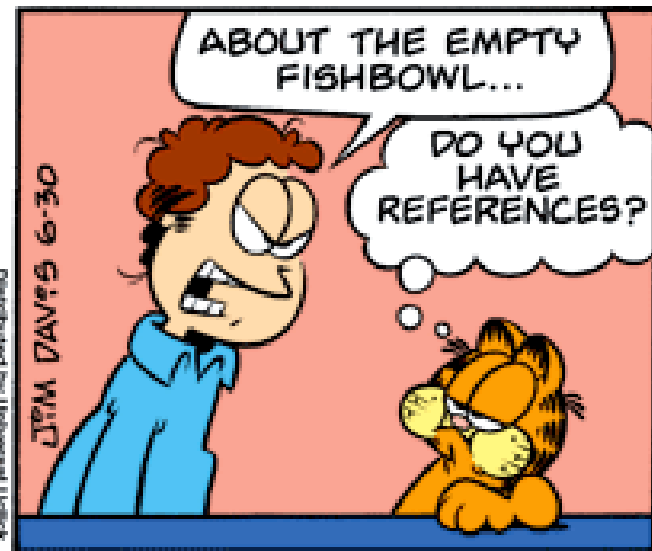
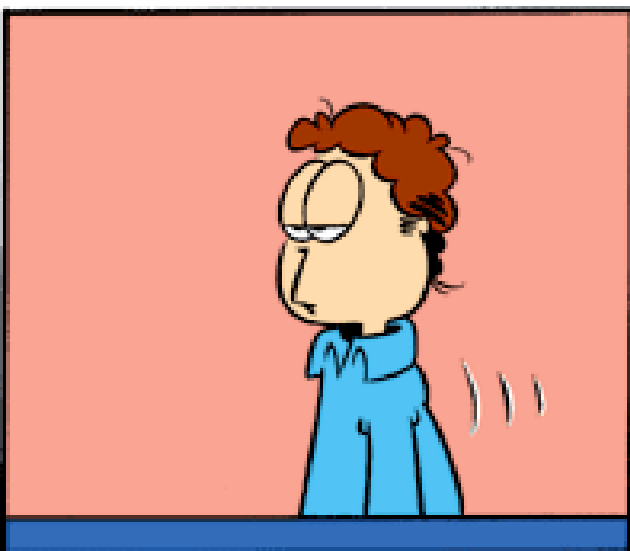
“As cloud-based services transition from marketing vaporware to real, deployed systems, the demands on traditional Web-hosting and Internet service providers are likely to shift dramatically. **In particular**, current models of resource provisioning and accounting lack the flexibility to effectively support the dynamic composition and rapidly shifting load enabled by the software as a service paradigm. **We have identified one key** aspect of this problem, namely the need to rate limit network traffic in a distributed fashion, and provided two novel algorithms to address this pressing need.

**Our experiments show** that naive implementations based on packet arrival information are unable to deliver adequate levels [...]

**Our results demonstrate** that it is possible to recreate, at distributed points in the network, the flow behavior that end users and network operators expect from a single centralized rate limiter. Moreover, it is possible [...]

RAGHAVAN et al - SIGCOMM 2007

# Referências são importantes, sempre



# REFERÊNCIAS

- Corretas, completas, específicas
- Informações obrigatórias: autores, título, nome do evento ou periódico (editora), volume e número se necessário, ano
- Referências relevantes
  - Do mesmo ano (ou ano anterior) para ilustrar que o tópico é atual e de interesse da comunidade
  - Artigos de conferências, periódicos, livros (não apenas sites da Internet!)
  - Todas as obras listadas no conjunto de referências devem ser mencionadas no texto, e vice-versa



# SBC Horizontes: Agosto'2009

- [HOW TO]

## **Como fazer pesquisa bibliográfica**

Agma Juci Machado Traina, Caetano Traina Jr., ICMC-USP São Carlos

*Por onde devo começar quando quero iniciar uma pesquisa bibliográfica? Este artigo visa responder a essa questão, apresentando ao jovem pesquisador os passos principais para realizar uma pesquisa bibliográfica que contemple seu objetivo, bem como as ferramentas básicas amplamente disponíveis para ser bem sucedido.*

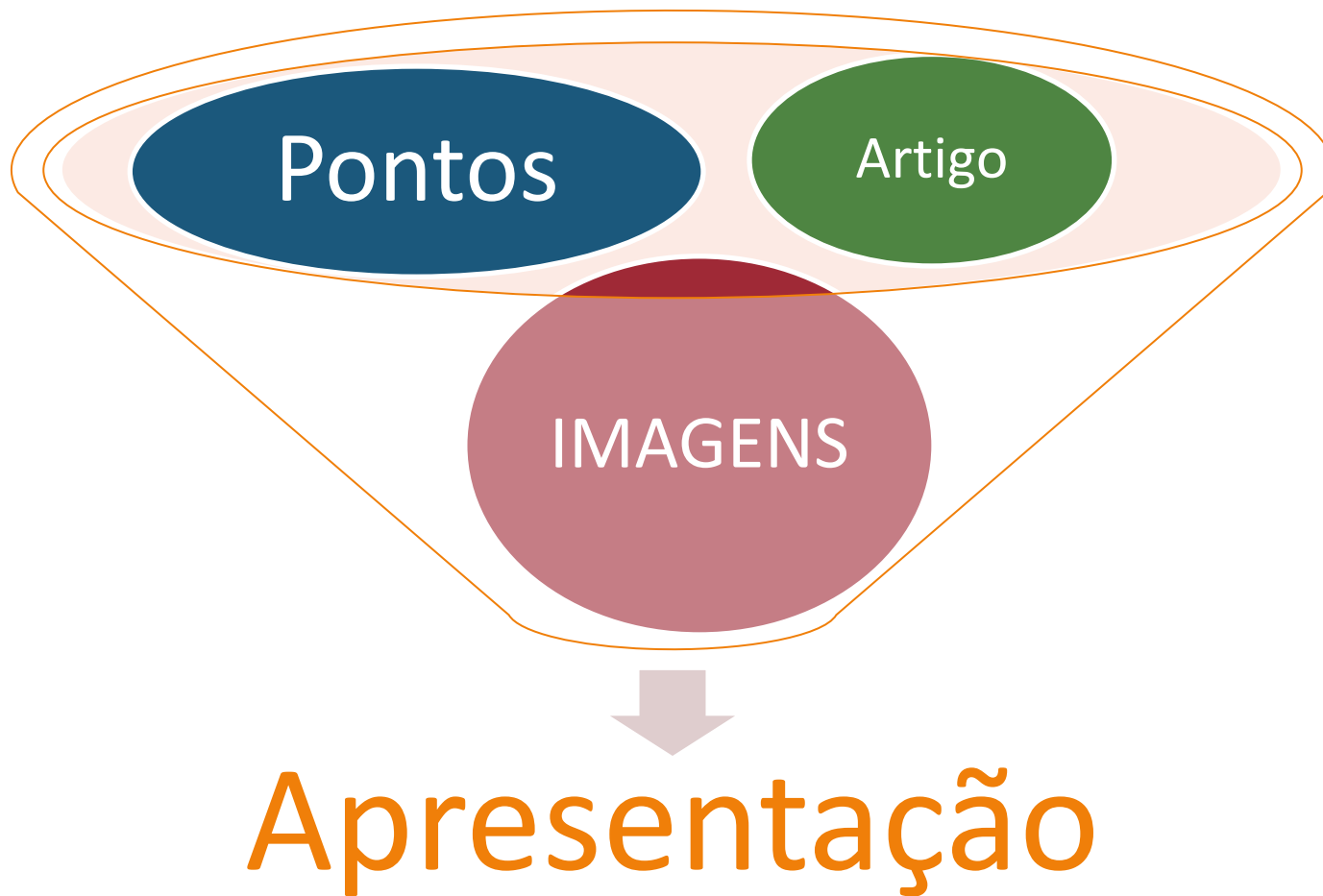
PARTE 5

# APRESENTAÇÃO DE ARTIGOS

# PRINCÍPIO SIMPLES 1

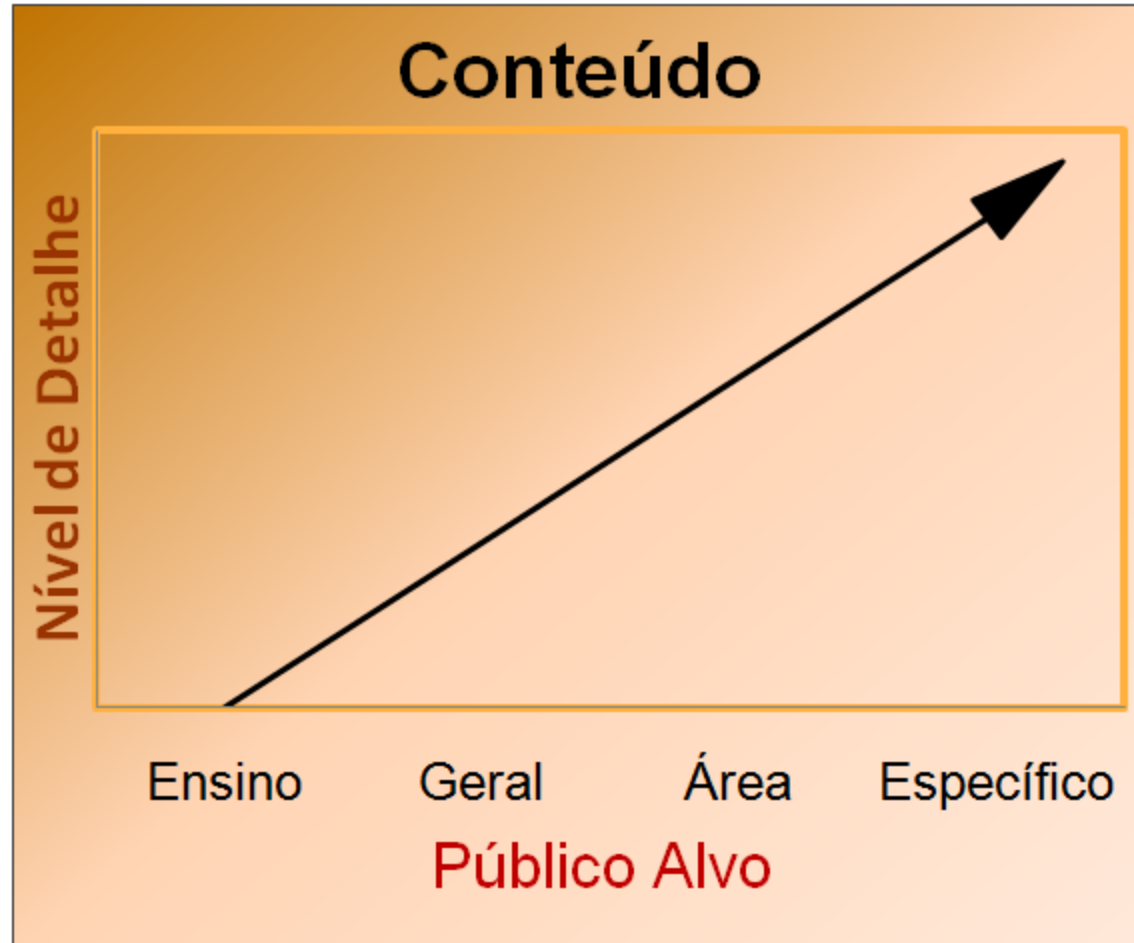
Detalhes estão  
no artigo

# PRINCÍPIO SIMPLES 2



# Planejamento

## O que apresentar

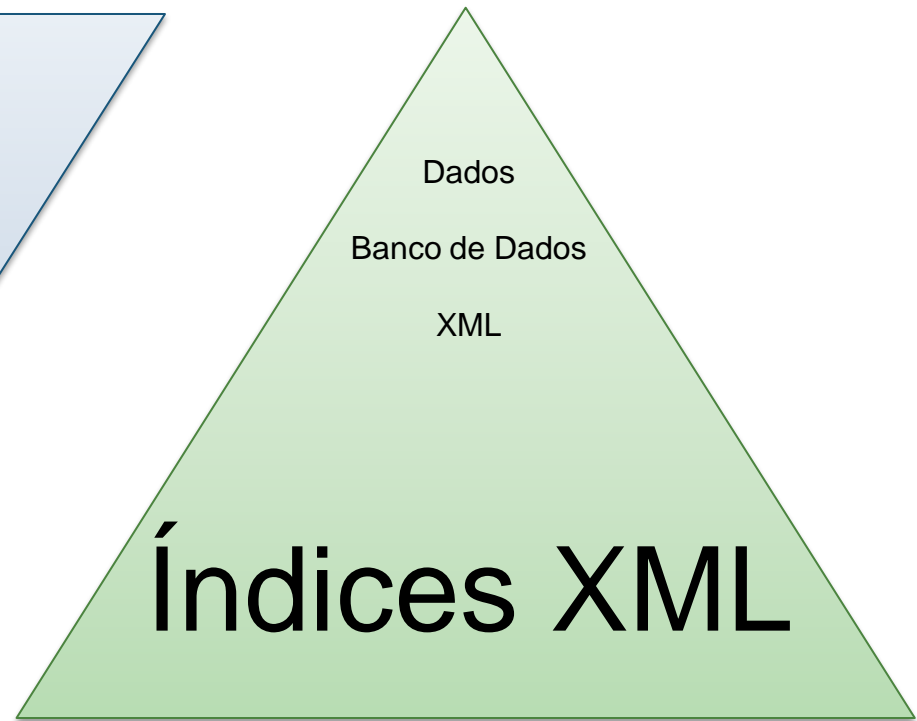
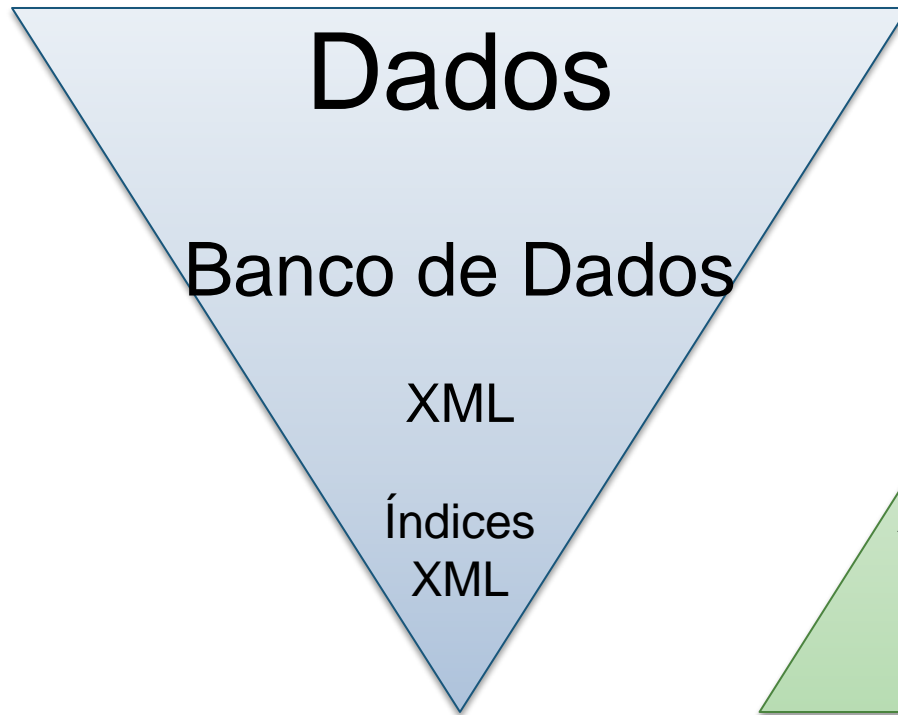


# Planejamento

## O que apresentar

### “XHI – Índice Híbrido para Dados XML”

Aula na Graduação



XSym, XYME-P

# SETE pecados capitais do(a) APRESENTADOR(A)

1. Monotonia, falar baixo
2. Falta de ensaio, se perder, ããã
3. Ticks: cabelos, pés, braços, ...
4. RELÓGIO
5. Exagerar: humor, gestos...
6. Falar rápido, atropelar informação/  
transição importante
7. Falta de contato visual, costas



# 7 pecados capitais da APRESENTAÇÃO

1. Frases completas, longas
2. Slides poluídos (cores, figuras)
3. Fonte pequenas ou Fontes difíceis
4. Imagens/tabelas ilegíveis  
(copy/paste do artigo)
5. Animações malucas
6. Eros otográficos
7. Falta  
de fluxo/tempo





# Corrigindo os 7 pecados capitais

1. Frases completas
  2. Slides poluídos
  3. Fonte pequenas ou Fontes difíceis
  4. Imagens ilegíveis
  5. Animações malucas
  6. Eros otográficos
  7. Falta de fluxo/tempo
1. Palavras-chave
  2. 1 slide = 1 informação
  3. Fonte tamanho  $\geq 24$   
*Arial, sans-serif*
  4. Novas imagens
  5. Animações necessárias
  6. Revisão
  7. Revisão, revisão, revisão +  
ensaio, ensaio, ensaio

EXEMPLOS reais  
desses pecados  
e outros

# [1] Muita informação + fonte pequena

- The widespread employment of XML motivates the development of efficient methods for manipulating XML data.
- Query languages (e.g. XQuery, XPath) take into consideration the structure of the data and enable querying both on its structure and simple values.
- Tree-pattern queries are considered among the most important and yet more complex queries to process.
- Holistic processing techniques have outperformed more conventional solutions.
- Many techniques have been proposed to handle tree-pattern queries holistically (e.g. TwigStack, XML filtering, ViST, PRiX).
- Index structures (e.g. B+-tree, XR-tree, XB-tree) have been introduced to further improve performance.
- A common characteristic for all holistic approaches is that some preprocessing is required, either on the data (e.g. TwigStack) or both the data and query (e.g. ViST, PRiX).
- Thus the need for a “dedicated” (native or relational) XML storage manager where all data resides and can be preprocessed.
- Problem: While many processing methods have already been proposed for such queries (and were shown to have promising performance and potential), none of them has found its way to any of the existing lightweight XML engines (e.g. Galax XQuery processor, eXist).

# Muita informação+fonte pequena

- Solução 1: dividir slide
- 1 slide = 1 informação

- The widespread employment of XML motivates the development of efficient methods for manipulating XML data.
- Query languages (e.g. XQuery, XPath) take into consideration the structure of the data and enable querying both on its structure and simple values.
- Tree-pattern queries are considered among the most important and yet more complex queries to process.
- Holistic processing techniques have outperformed more conventional solutions.
- Many techniques have been proposed to handle tree-pattern queries holistically (e.g. TwigStack, XML filtering, ViST, PRiX).
- Index structures (e.g. B+-tree, XR-tree, XB-tree) have been introduced to further improve performance.
- A common characteristic for all holistic approaches is that some preprocessing is required, either on the data (e.g. TwigStack) or both the data and query (e.g. ViST, PRiX).
- Thus the need for a “dedicated” (native or relational) XML storage manager where all data resides and can be preprocessed.
- Problem: While many processing methods have already been proposed for such queries (and were shown to have promising performance and potential), none of them has found its way to any of the existing lightweight XML engines (e.g. Galax XQuery processor, eXist).

- XML everywhere
- XQuery, XPath  
structure + values
- Tree-pattern queries

## Holistic Tree Pattern Query

- Better than conventional solutions
- E.g.: TwigStack, filtering, ViST, PRiX
- Index: B+-tree, XR-tree, XB-tree
- Preprocessing required
  - Data
  - Data + query

## Problems

- Need for a “dedicated” XML storage manager (native or relational)
- No algorithm on lightweight XML engines

# Muita informação+fonte pequena

- Solução 2: usar cores
- 1 informação = 1 cor

```
SELECT t2.nome, t3.salario
FROM PESSOA t0, MEDICO t1,
     PESSOA_NOME t2, MEDICO_SALARIO t3
WHERE t0.oid = t1.oid AND t0.oid = t2.oid AND
     t1.rid = t3.rid AND t3.salario > 2000 AND
     t2.t_timei <= '01/01/2000 AND
     t3.t_timei <= '01/01/2000'
```

**MAPEAMENTO:**  
Tabelas  
Relacionamentos  
Restrições dados  
Restrições tempo

## [2] Cor em exagero

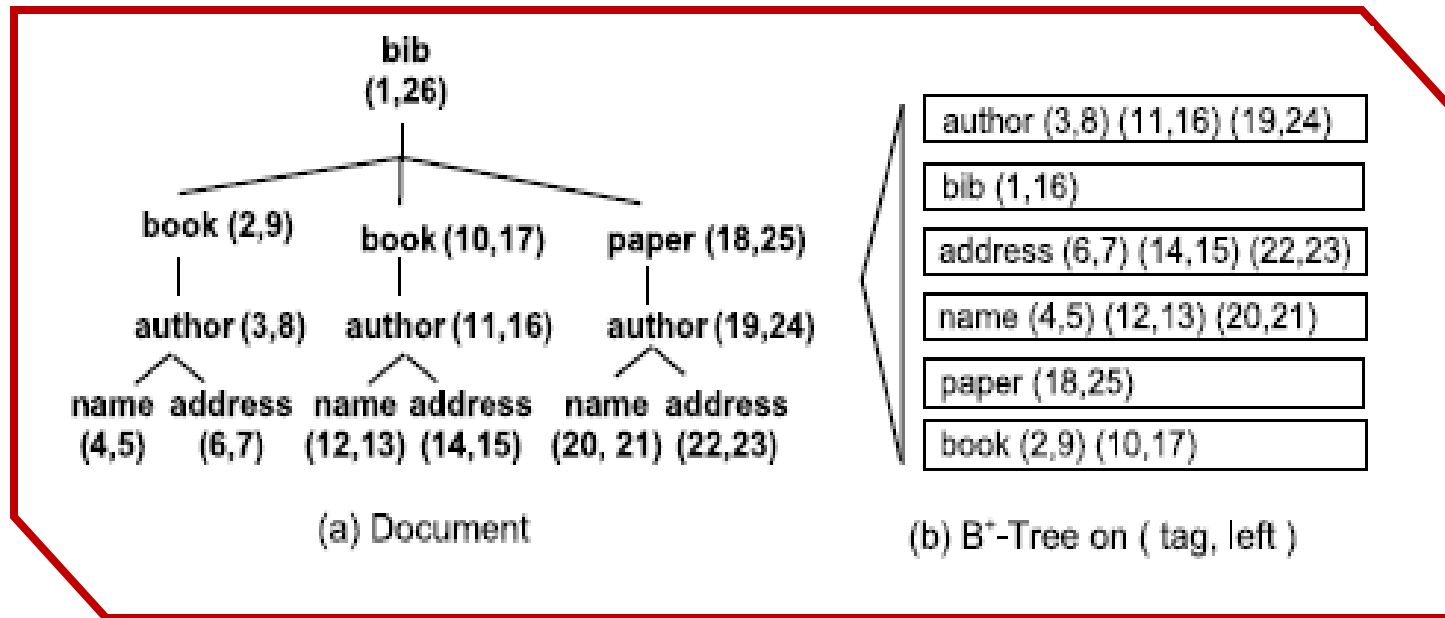
```
SELECT t2.nome, t3.salario
FROM PESSOA t0, MEDICO t1,
     PESSOA_NOME t2, MEDICO_SALARIO t3
WHERE t0.oid = t1.oid AND t0.oid = t2.oid AND
     t1.rid = t3.rid AND t3.salario > 2000 AND
     t2.t_timei <= '01/01/2000 AND
     t3.t_timei <= '01/01/2000'
```

```
SELECT t2.nome, t3.salario
FROM PESSOA t0, MEDICO t1,
     PESSOA_NOME t2, MEDICO_SALARIO t3
WHERE t0.oid = t1.oid AND t0.oid = t2.oid AND
     t1.rid = t3.rid AND t3.salario > 2000 AND
     t2.t_timei <= '01/01/2000 AND
     t3.t_timei <= '01/01/2000'
```

# [3] copy/paste imagem

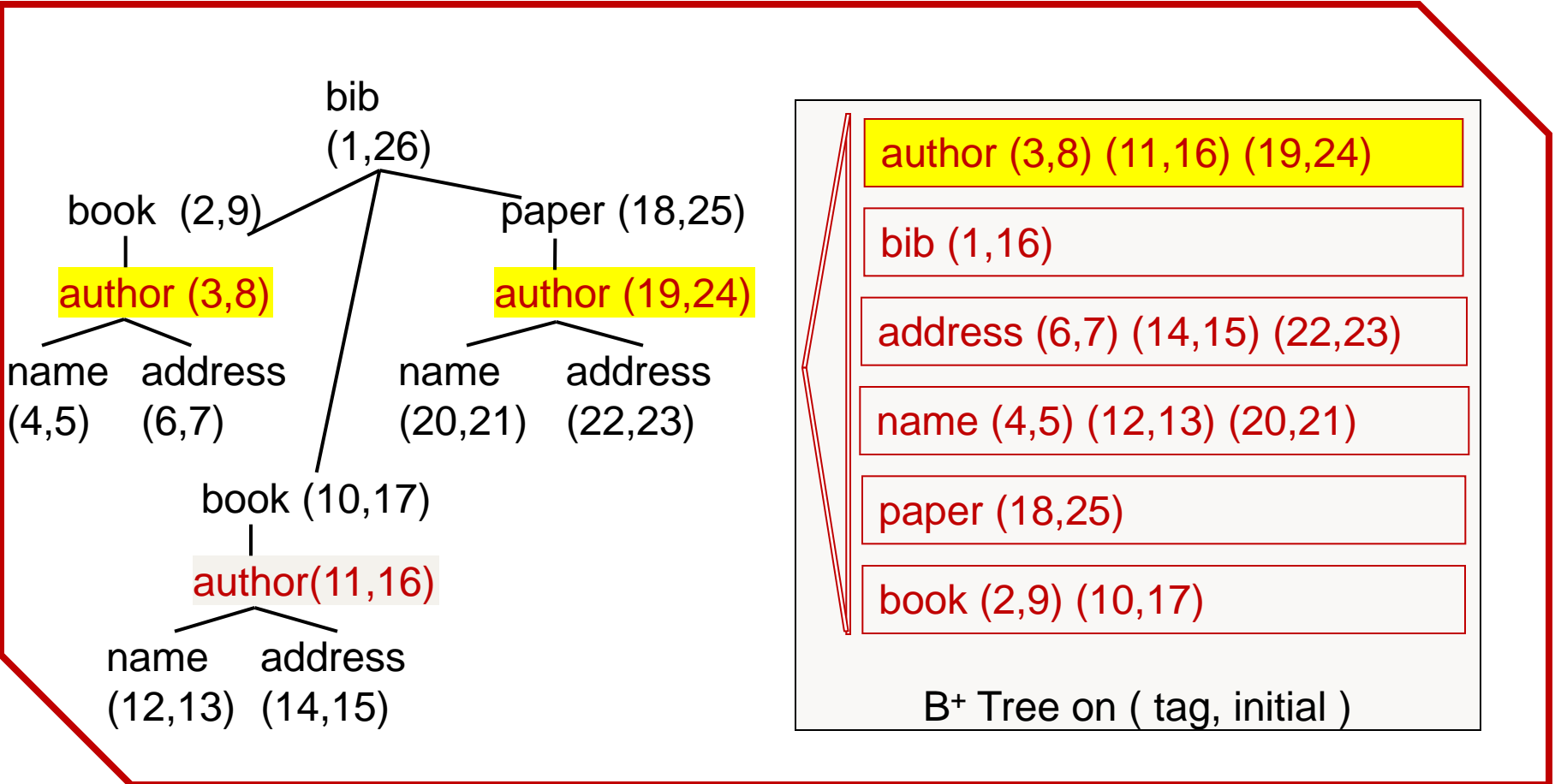
## Copy/paste imagem do artigo

- Nem sempre legível
- Horrível, informação demais



# copy/paste imagem

Solução: refazer imagem usando software de apresentação





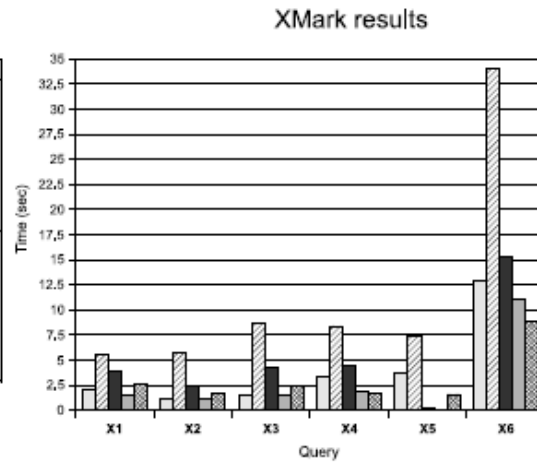
# copy/paste imagem

## Solução: refazer imagem usando software de apresentação

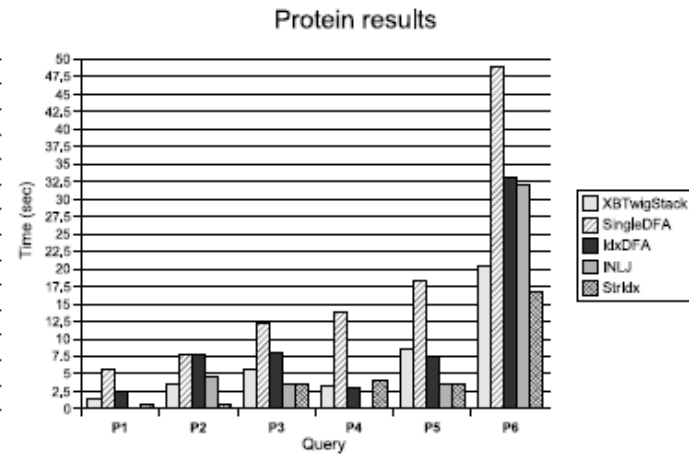
	Twig	Single	IdxDFA	INLJ	StrIdx	Sel(%)	#Nodes
X1	100	100	100	140	100	100	600002
X2	100	100	100	119,93	79,95	78,95	636944
X3	99,66	100	100	114,47	61,37	80	891372
X4	100	100	78,45	42,03	42,14	30,02	1055512
X5	100	100	2,73	4,7	40,06	2,7	1364519
X6	85,82	100	73,35	135,4	78,89	60,35	3099331
P1	67,48	100	100	3,82	34,6	3,05	793637
P2	100	100	100	128,83	21,55	0,16	1223068
P3	96,49	100	96,49	114,93	78,04	78,04	1423126
P4	39	100	100	4,57	73,82	2,91	1714187
P5	47,9	100	47,9	57,55	38,37	38,37	2736993
P6	96,53	100	95,91	154,88	91,02	91,02	5497746

Percentage of nodes accessed per method, query selectivity, number of nodes in document

(a)



(b)



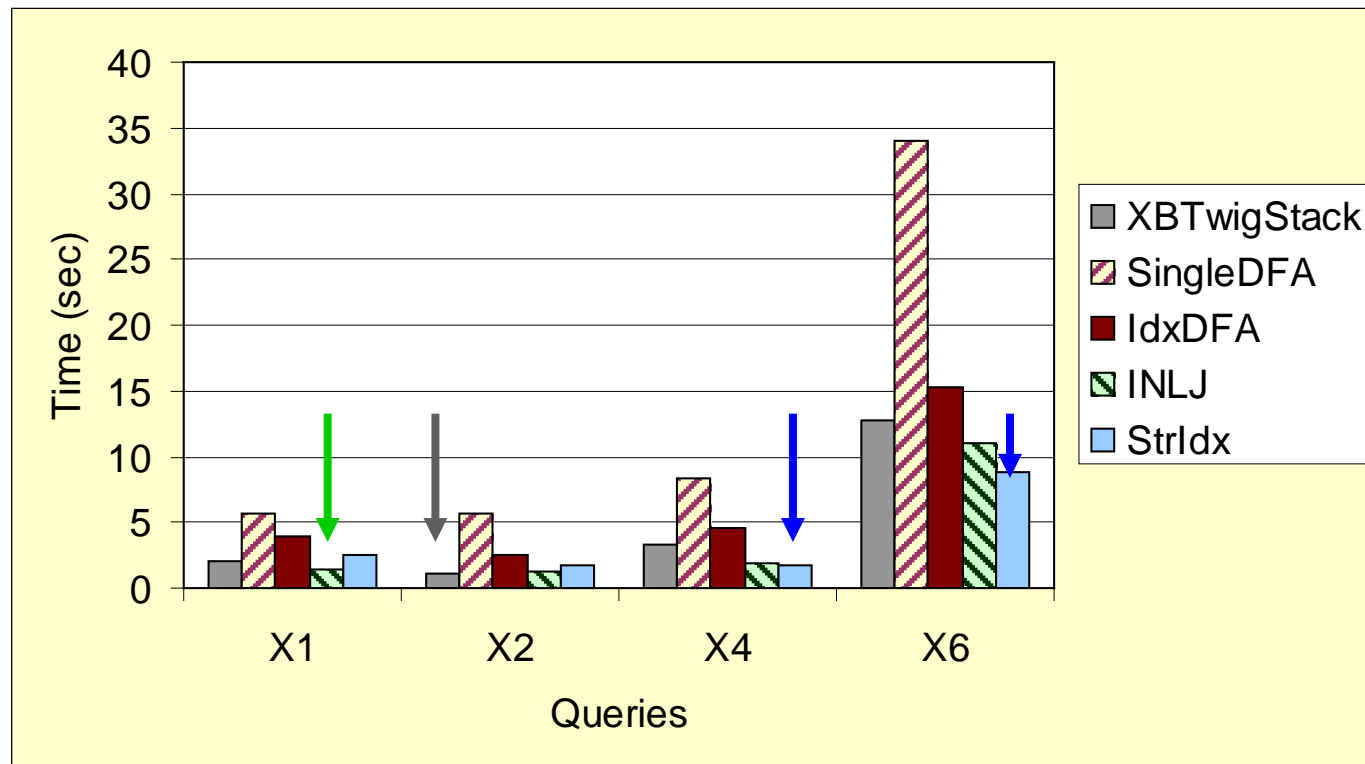
(c)

Figure 6: Results for XMark and Protein datasets

- Resultados XMARK
  - Não tem classificação fixa
  - SingleDFA tem pior desempenho
  - StrIdx melhor ou comparável a melhor solução

# copy/paste imagem

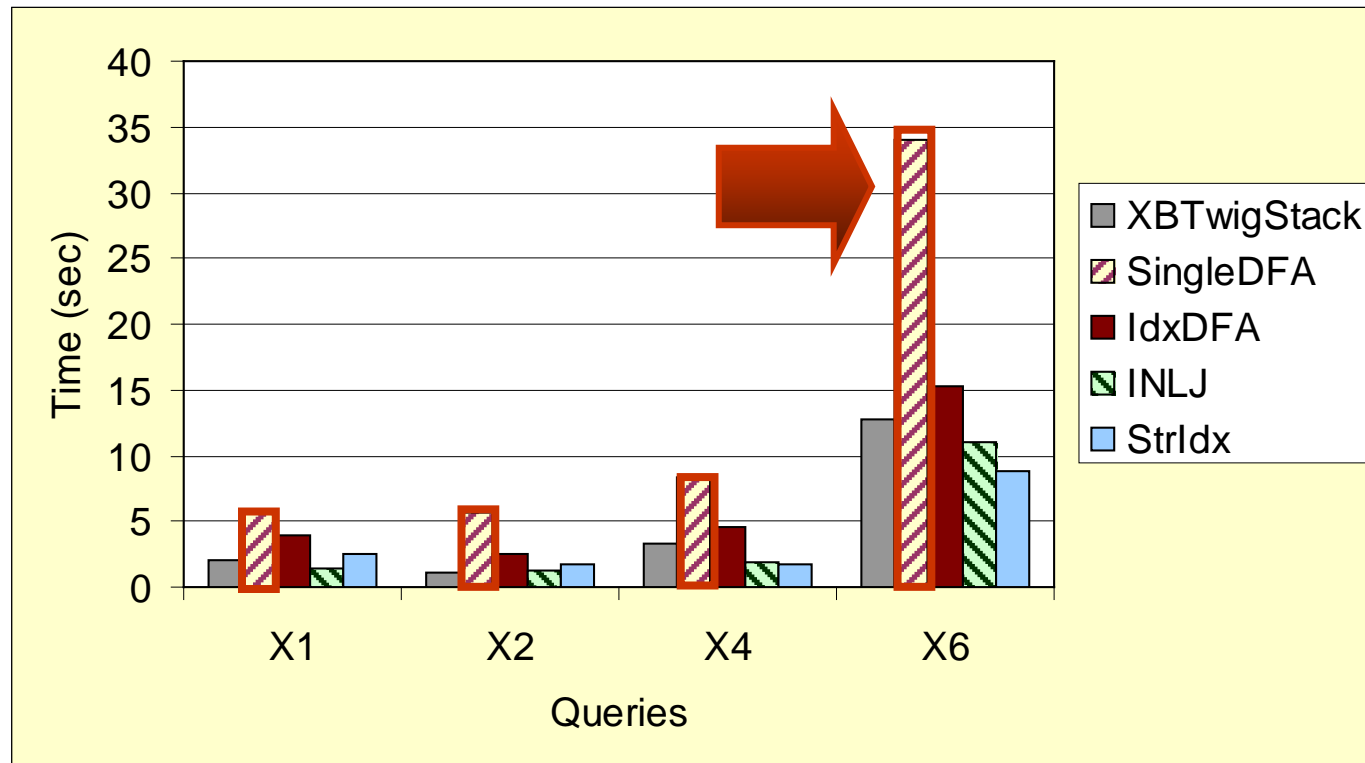
Solução: refazer imagem usando software de apresentação + animação



- Resultados XMARK: 1. não tem classificação fixa

# copy/paste imagem

Solução: refazer imagem usando software de apresentação + animação



- Resultados XMARK: 2 SingleDFA pior desempenho

# Começando o Curso

## Atualize seus *bookmarks*

- **BDBCComp**
  - <http://www.lbd.dcc.ufmg.br/bdbcomp/>
- **Portal Periódicos/CAPEs**
  - <http://www.periodicos.capes.gov.br>
- **DBLP** – biblioteca digital Ciência da Computação
  - <http://www.informatik.uni-trier.de/~ley/db>
- **ACM Digital Library** – *Assoc. for Computing Machinery*
  - <http://portal.acm.org/dl.cfm>
- **IEEE Xplore** – *Inst. of Electrical and Electronics Eng.*
  - <http://ieeexplore.ieee.org>
- **CiteSeer** – biblioteca digital de literatura científica
  - <http://citeseer.nj.nec.com/cs>
- **Science Direct** – journals publicados pela *Elsevier*
  - <http://www.sciencedirect.com/science/journals/computerscience>
- **Google** <http://www.google.com>



# Como se faz Pesquisa

## 1) Tema em Aberto

LEIA LEIA LEIA LEIA

Mas ler o que mesmo?

- Tuuuuuuuuuuuudo! (cientificamente falando)
- Inicie pelos trabalhos recentes do seu orientador (facilita a tarefa de conhecer o seu trabalho)
- A partir desses trabalhos, selecione algumas referências bibliográficas e leia também
- Leia os *proceedings* das conferências mais importantes da sua área dos últimos três anos
  - SBBD, VLDB, ICDE, SIGMOD
- Aluno de doutorado : um artigo por dia

# Ler é a ÚNICA solução

## Orientador

- Seus artigos recentes
- Dissertações/Teses recentes

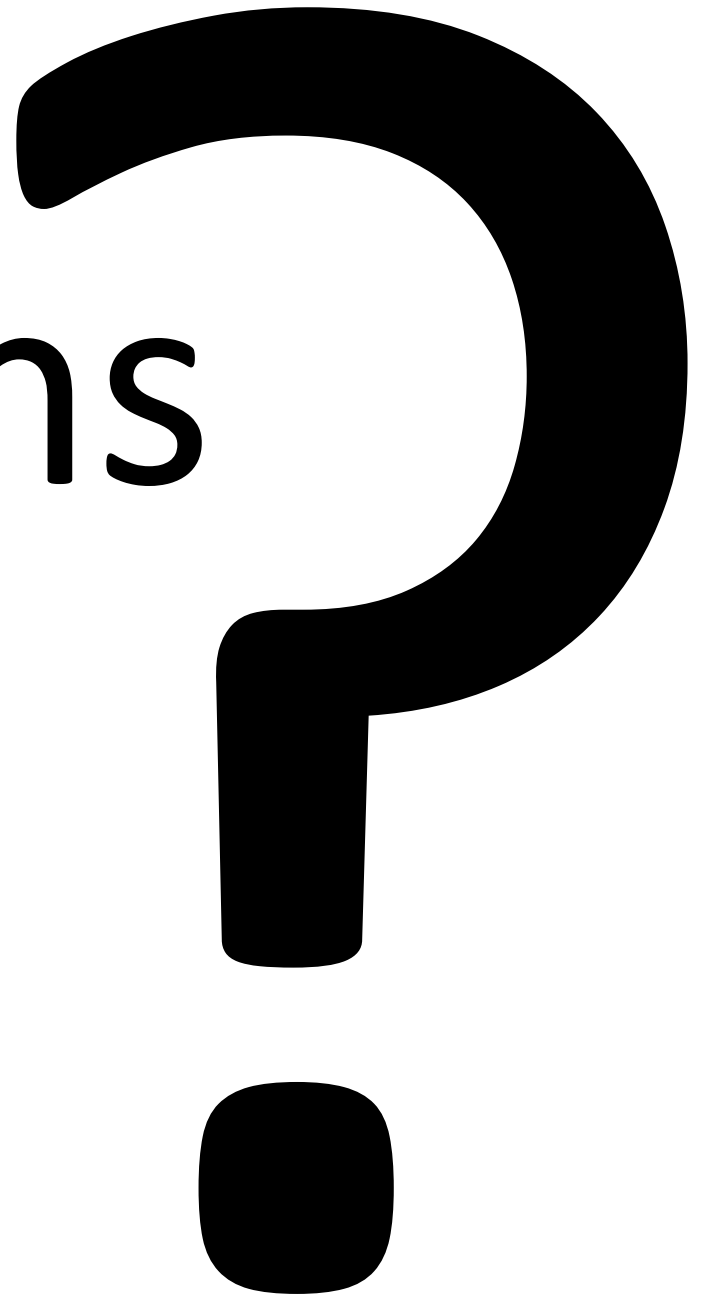
## Relacionados

- A partir das referências
- A partir de buscas

## Eventos Periódicos

- Principais (VLDB, SIGMOD, SBBD)
- De 2014, 2013, ...

Questions





# Escrita e Apresentação de Artigos Científicos

Mirella M. Moro, UFMG

[mirella@dcc.ufmg.br](mailto:mirella@dcc.ufmg.br)



**SOLUÇÃO:** Repita o primeiro slide com título, autores, email, fomento

Data Mining

Career

Data Mining

Graph Mining

Structural  
Correlation  
Patterns

Knowledge  
Transfer

Web  
Observatory

Analytics  
Systems

Summary

Research education is a long process.

Every work should present clear contributions.

Cooperation is a key research mechanism.

Target real problems, beyond CS, whenever possible.

Research results should become technology and be transferred.



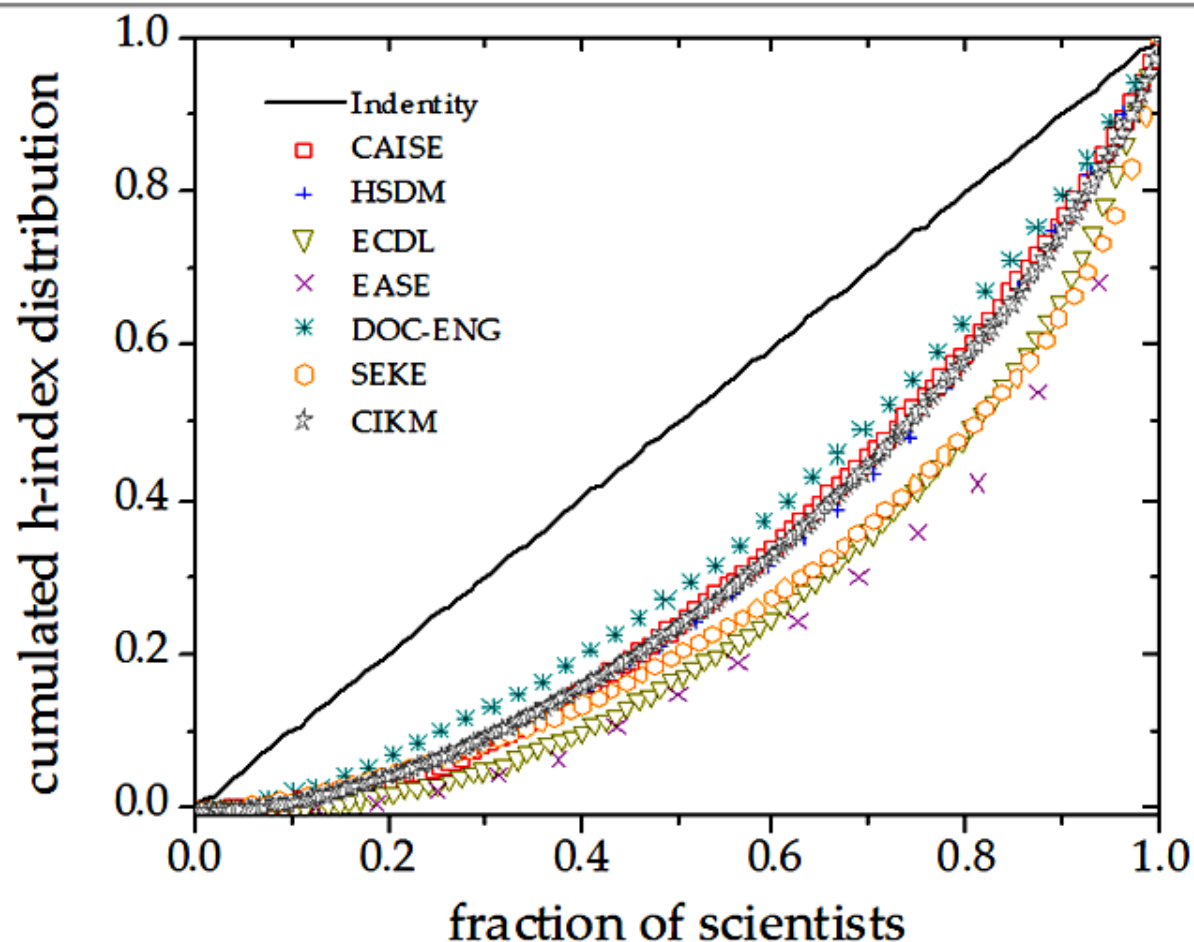


Fig. 2. Lorenz curves for the h-index distributions of researches in conferences of Engineering Software.

The Lorenz curve is a graph that represents the cumulative distribution of a probability density function.

The Gini coefficient is calculated as the area between the *perfect equality line* and the observed Lorenz curve.

This coefficient is directly proportional to the inequality of the distribution.

A low value of Gini coefficient indicates a more equal distribution among the parts, and a high value indicates a more unequal distribution.

PARTE 6

**EXEMPLOS DO SBBD 2014**

# Outline

- Introduction.
- Similarity join.
- The similarity *Wide-join* operator.
- Experiments.
- Conclusion.

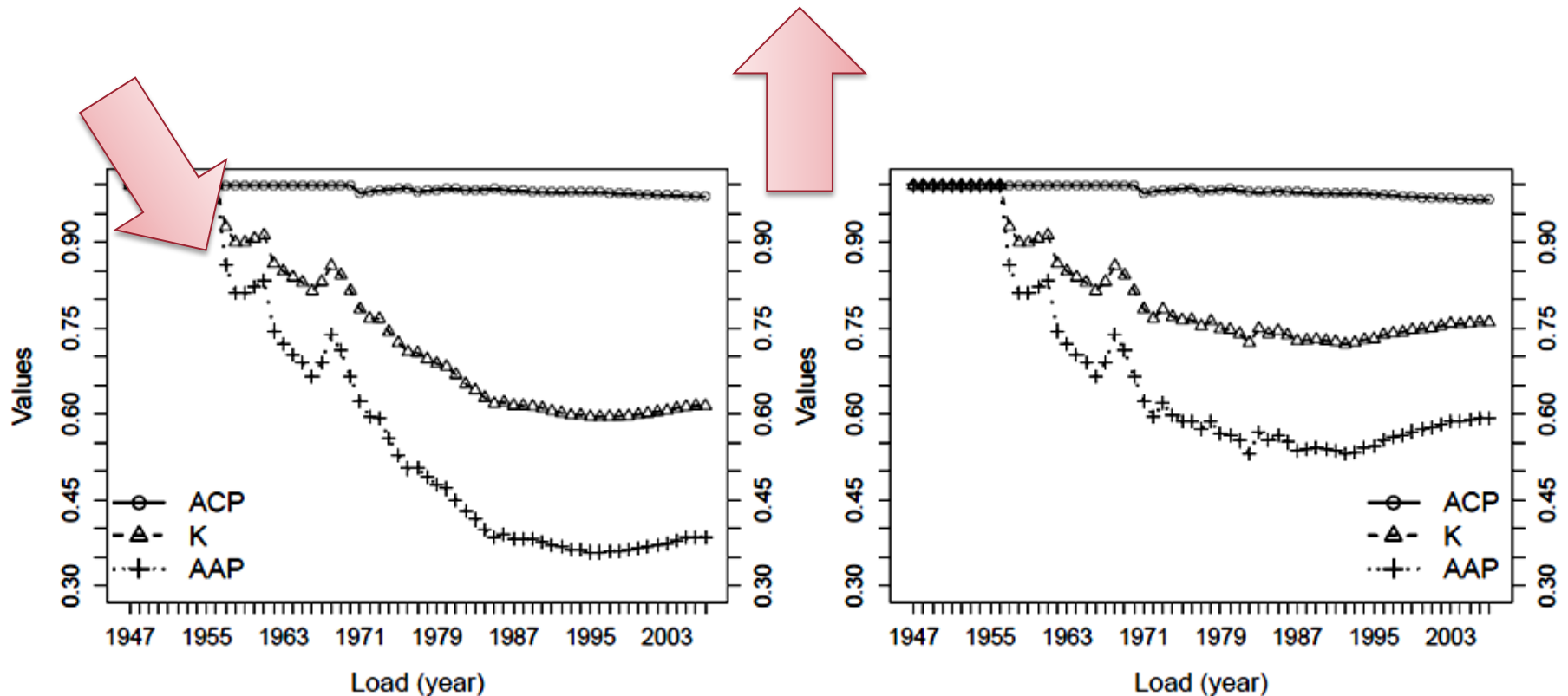
Em uma apresentação de 10-15-20-25 mins?!

Em uma conferência em que *\*todas\** apresentações seguem o *\*mesmo\** padrão? Cada item com ponto final?!

# SAND Design

## The Author Assignment Step

- We propose to use the lack of rules supporting any already seen author as evidence indicating the appearance of an unseen author.
- The number of rules that is necessary to consider an author as an already seen one is controlled by a parameter,  $\gamma_{min}$ .
- For an reference  $x \in T$ , if the number of rules extracted from  $D^x(\gamma(x))$  is smaller than  $\gamma_{min}$ , then the author of  $x$  is considered as a new/unseen author and a new label  $a_k$  is created to identify such author.
  - This prediction is considered as a new example and included into  $D$ .
- An appropriate value for  $\gamma_{min}$  can be obtained by performing cross-validation in  $D$ .



(a) INDi

(b) ALL-ALL

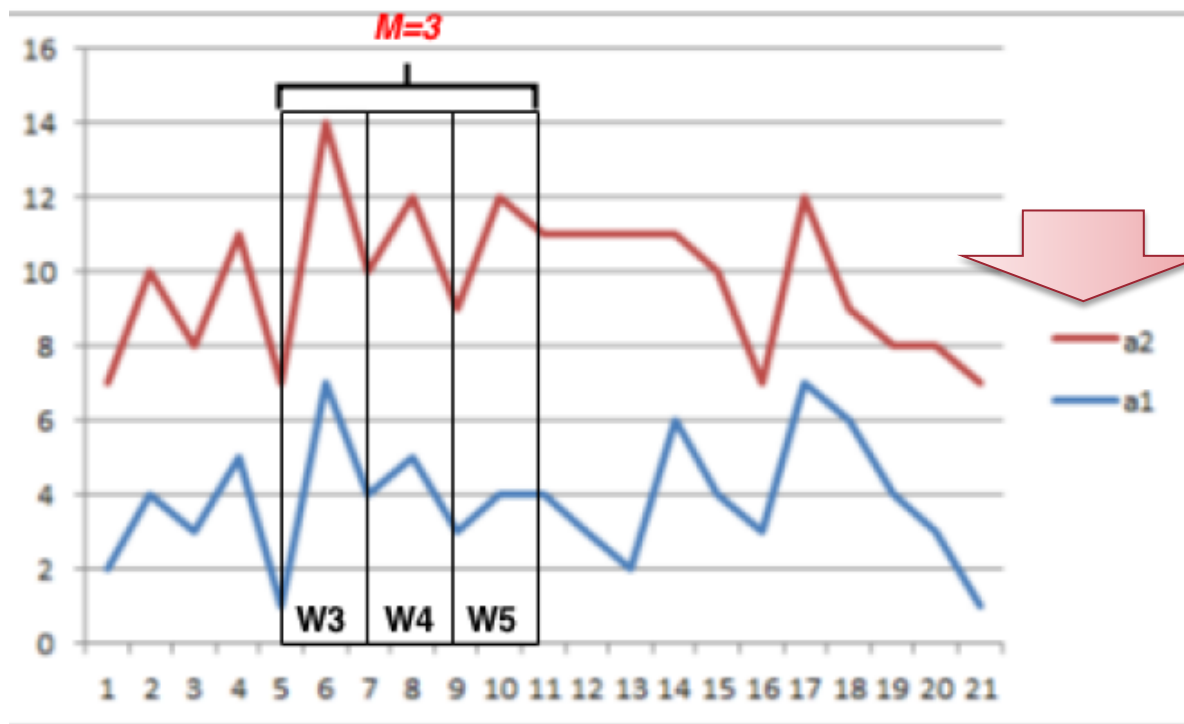
Figure : KISTI collection

**ALL-ALL**Outperforms INDi around 16,6%, under the  $K$  metric.



# *Halite*<sub>ds</sub> algorithm

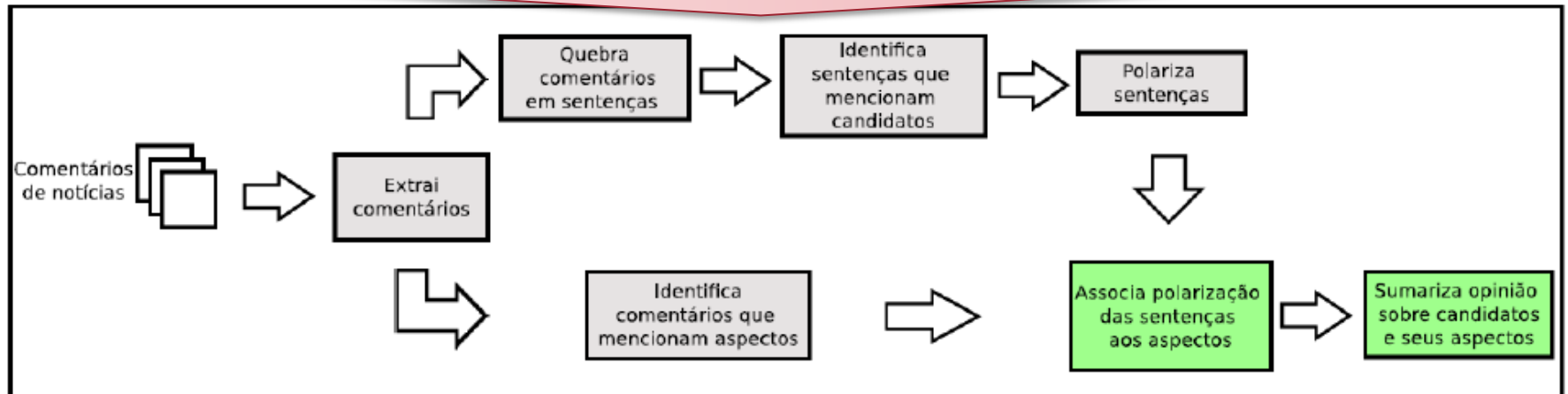
- How to analyze the data evolution using Halite?
  - Halite: Subspace Clustering in Static Databases.



An Evolutionary  
Quad-Tree!

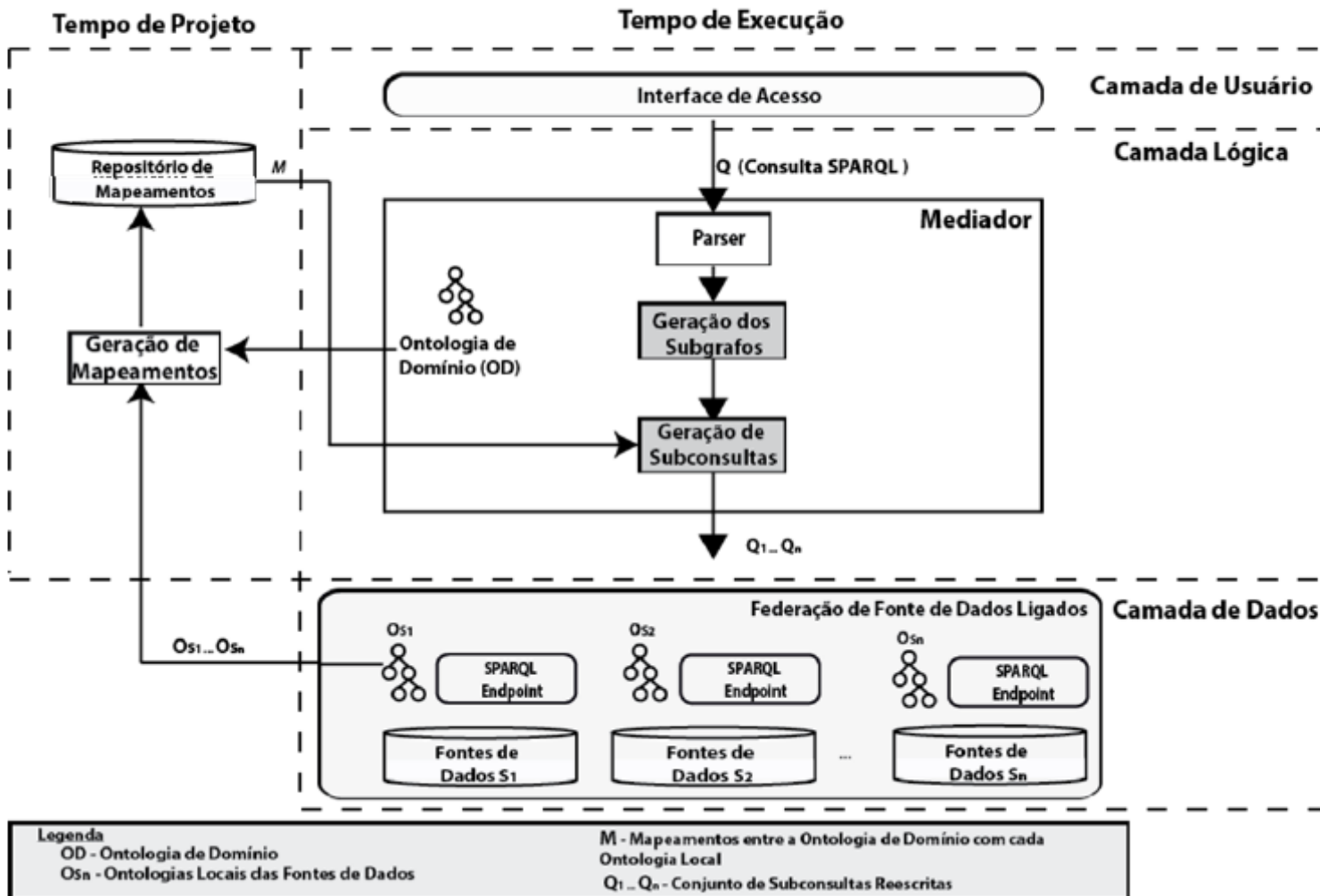


# Mineração da opinião sobre aspectos de candidatos a eleições em comentários de notícias



Processo para mineração das opiniões em nível de aspecto

oLinDa: *Query Decomposition over Linked Data Federation*



# Problema e Motivação

- ▶ Mas, seria possível olhar pra dentro desta caixa e capturar as suas informações internas a ela?
- ▶ Se isto fosse possível, como relacionaríamos este tipo de informação com a especificação e a execução do workflow?
- ▶ Seria possível relacionar estes dados sem a realização de algum pós-processamento?

# Chameleon

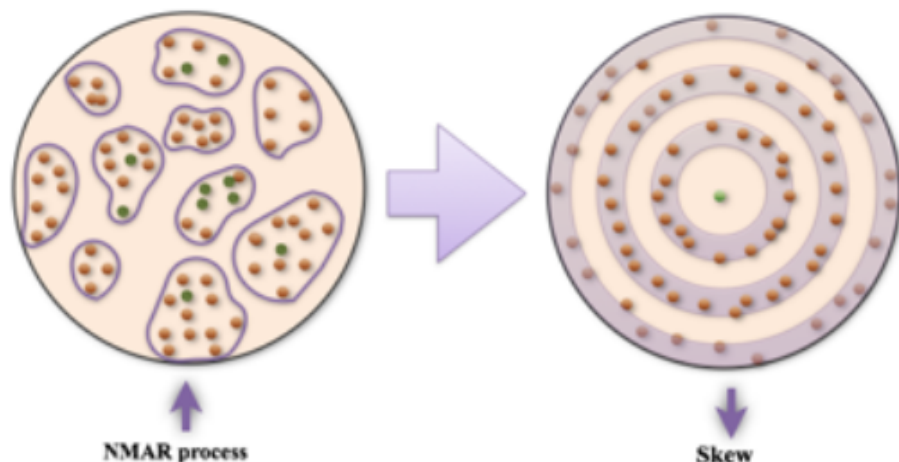
- Otimização de sistemas MapReduce tornou-se comum;
- Sistemas com algum tipo de otimização ganham sucesso no mercado;
- Análise do histórico de execução auxilia na otimização de novas tarefas;
- Utilizando ajuste automático de parâmetros, o Chameleon auxilia na otimização de consultas do Hive em tempo real.

# Analyzing Missing Data in Metric Spaces

Large databases may consist of thousands of attributes and may be a subject of **missing values**

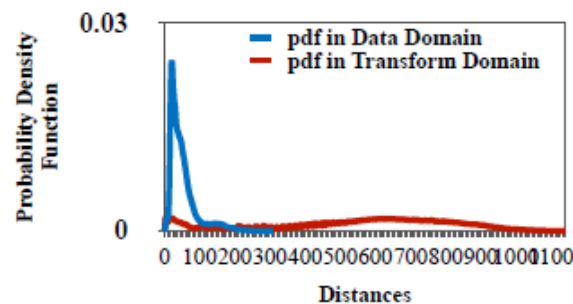
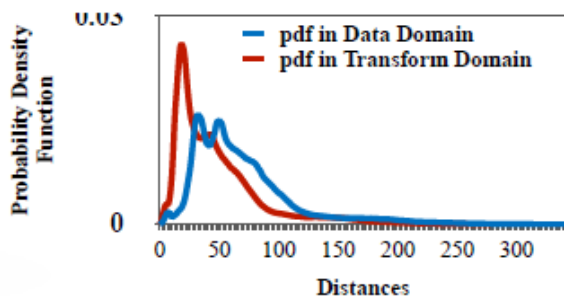


MAR or NMAR?



NMAR mechanism causes **skew** in the metric space

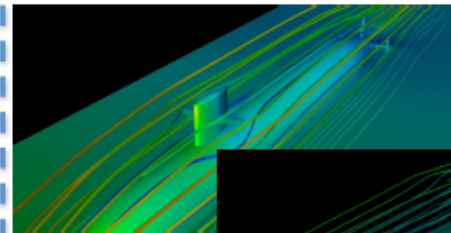
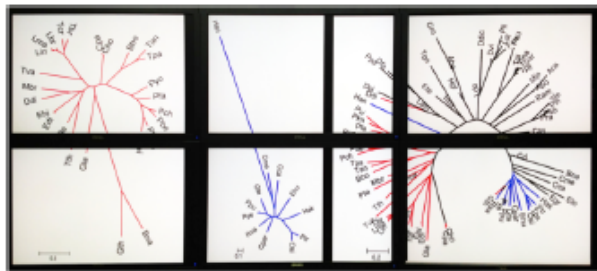
NMAR process causes **distance concentration**



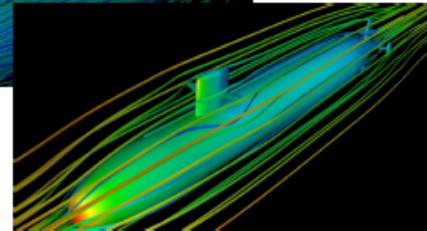
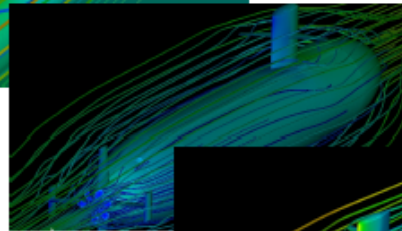
- Máquina paralela de *workflows* científicos – *dataflow*
  - Abordagem baseada em álgebra relacional para workflows
- Algoritmo adaptativo
  - Elasticidade em nuvens, tolerância a falhas, ...
- Consulta a dados de proveniência em tempo de execução
- Workflow dinâmico

SciPhy  
Phylogenetic analysis

SciEvol  
Evolutionary analysis



Computational Fluid Dynamics (CFD)



Tiled wall



## Metrics

$$ACP = \frac{1}{N} \sum_{i=1}^e \sum_{j=1}^t \frac{n_{ij}^2}{n_i} \quad (1)$$

$$AAP = \frac{1}{N} \sum_{j=1}^t \sum_{i=1}^e \frac{n_{ij}^2}{n_j} \quad (2)$$

$$K = \sqrt{ACP \times AAP} \quad (3)$$



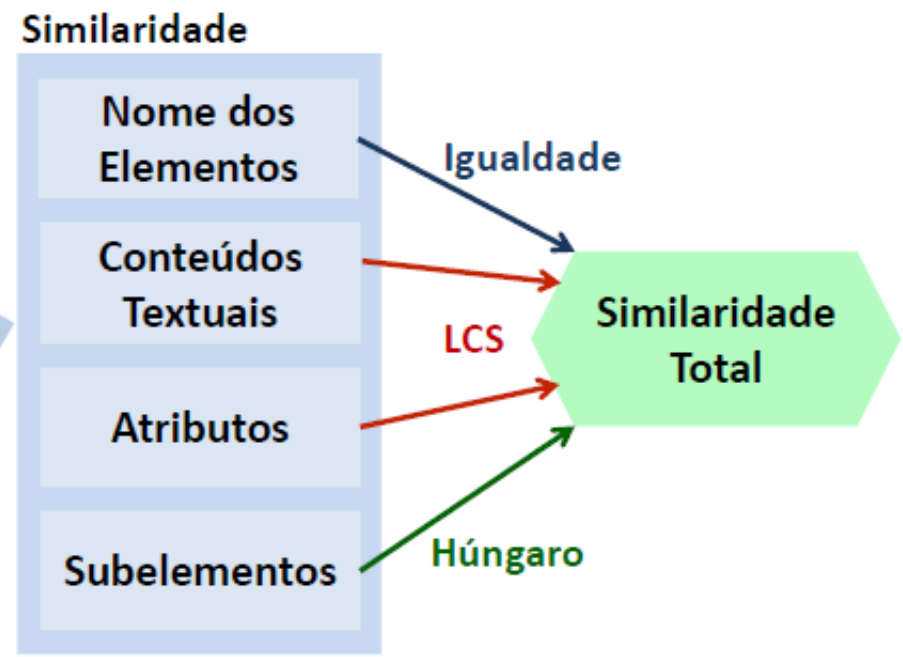
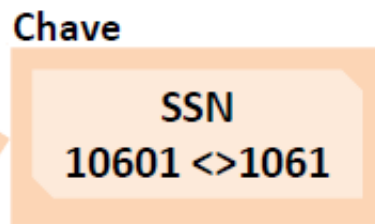
Table : Results in BDBComp. Best results, including statistical ties (highlighted).

Strategies	# of merges	# wrong merges	% wrong merges	ACP	AAP	K
INDi	0	0	-	<b>0,994 ± 0,005</b>	0,765 ± 0,013	0,872 ± 0,005
ALL-ALL	5	1	20,0	<b>0,995 ± 0,003</b>	<b>0,793 ± 0,003</b>	<b>0,888 ± 0,002</b>
ALL-DBS	6	1	16,7	<b>0,995 ± 0,003</b>	0,766 ± 0,004	0,873 ± 0,003
ALL-KMS	4	1	25,0	<b>0,995 ± 0,003</b>	<b>0,790 ± 0,002</b>	<b>0,886 ± 0,002</b>
ALL-TIW	11	1	09,1	<b>0,995 ± 0,003</b>	0,723 ± 0,005	0,848 ± 0,003
ALL-CEN	7	1	14,3	<b>0,993 ± 0,005</b>	0,766 ± 0,014	0,872 ± 0,006
TMC-ALL	4	1	25,0	<b>0,995 ± 0,003</b>	0,785 ± 0,004	0,884 ± 0,002
TMC-DBS	4	1	25,0	<b>0,995 ± 0,003</b>	0,759 ± 0,008	0,869 ± 0,005
TMC-KMS	4	1	25,0	<b>0,995 ± 0,003</b>	0,788 ± 0,002	0,885 ± 0,002
TMC-TIW	11	1	09,1	<b>0,995 ± 0,003</b>	0,723 ± 0,005	0,848 ± 0,003
TMC-CEN	4	1	25,0	<b>0,993 ± 0,005</b>	0,756 ± 0,019	0,866 ± 0,009

ALL-ALL e ALL-KMS

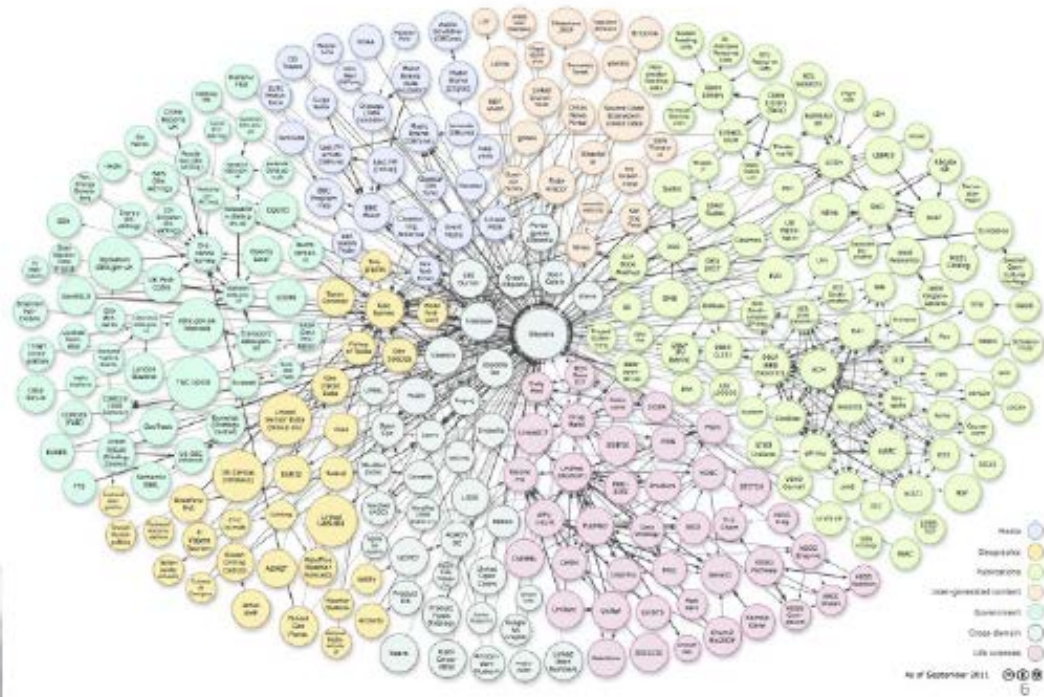
Outperform INDi around 1.8% and 1.4%, respectively, under the *K* metric.

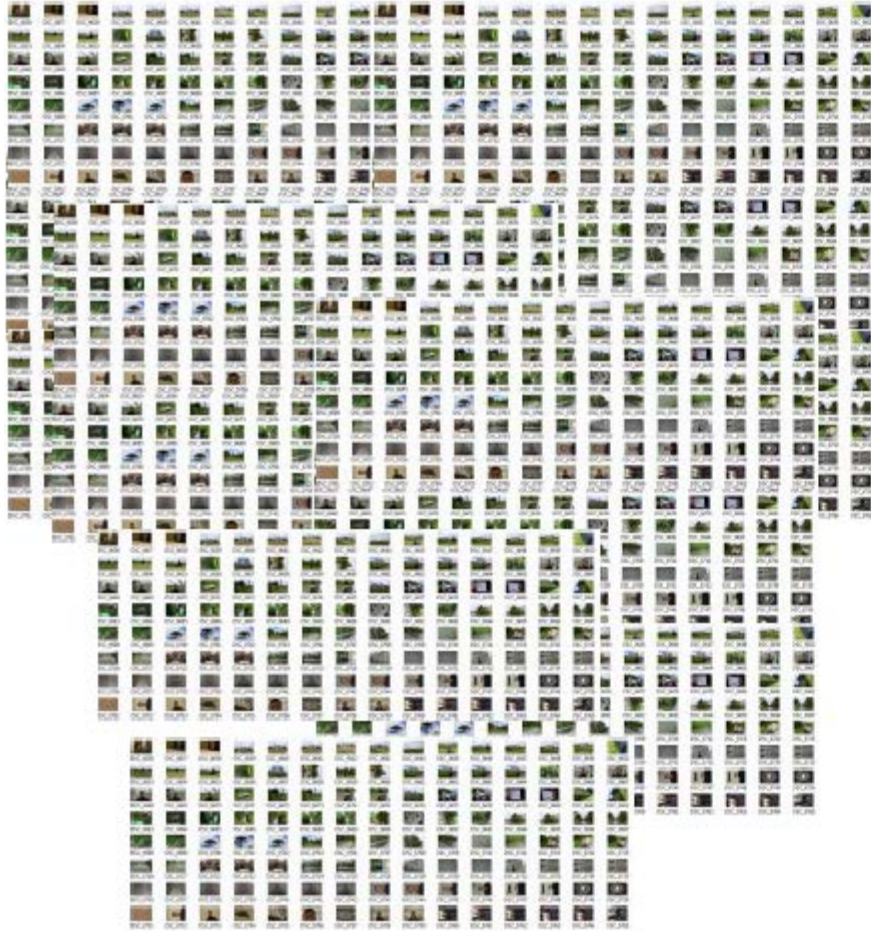
os BONS exemplos



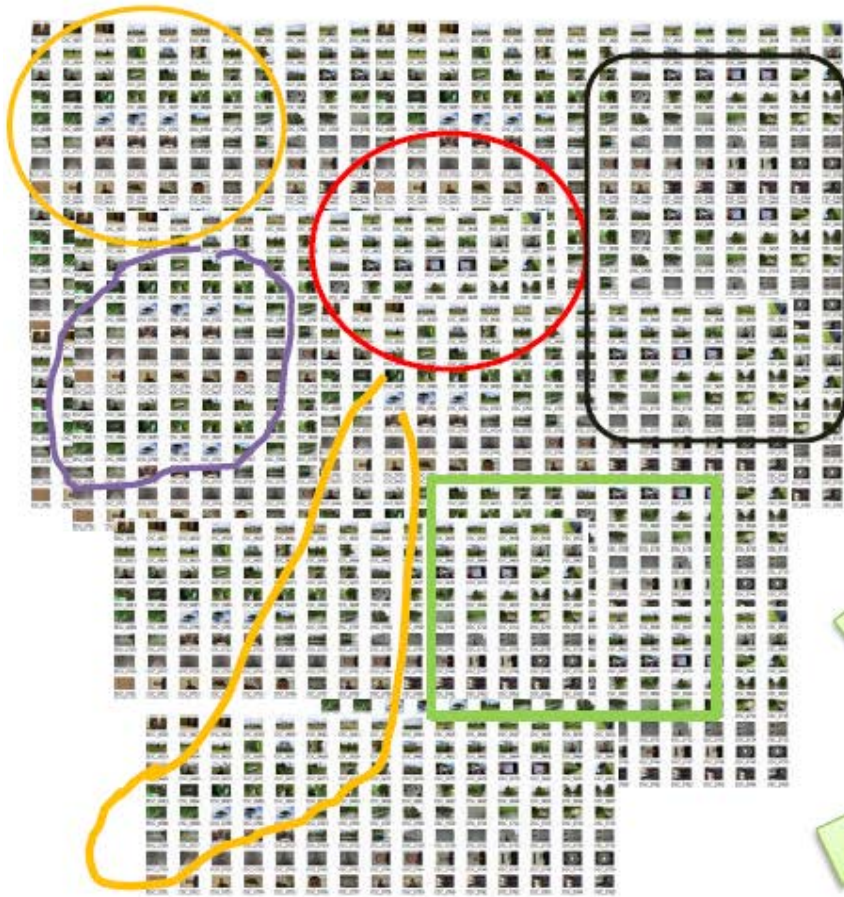
# Quer saber como encontrar os melhores *linked datasets* para sua aplicação?

Aplicação  
*Linked Data*





Novas aplicações  
de descoberta de  
conhecimento /  
Data Mining



Organizar Fotografias  
Pessoais

Encontrar Lugares  
Legais

Agrupar por Padrões

Investigações Criminais

# Recomendação de objetos

- \* Sugerir objetos adequados
- \* Para a pessoa certa
- \* Na hora certa
- \* No contexto certo

# No futebol

*Onde está o gol?*



Fred





PARTE 7

**FINALMENTE**

# ESCRITA: REVISÃO FINAL

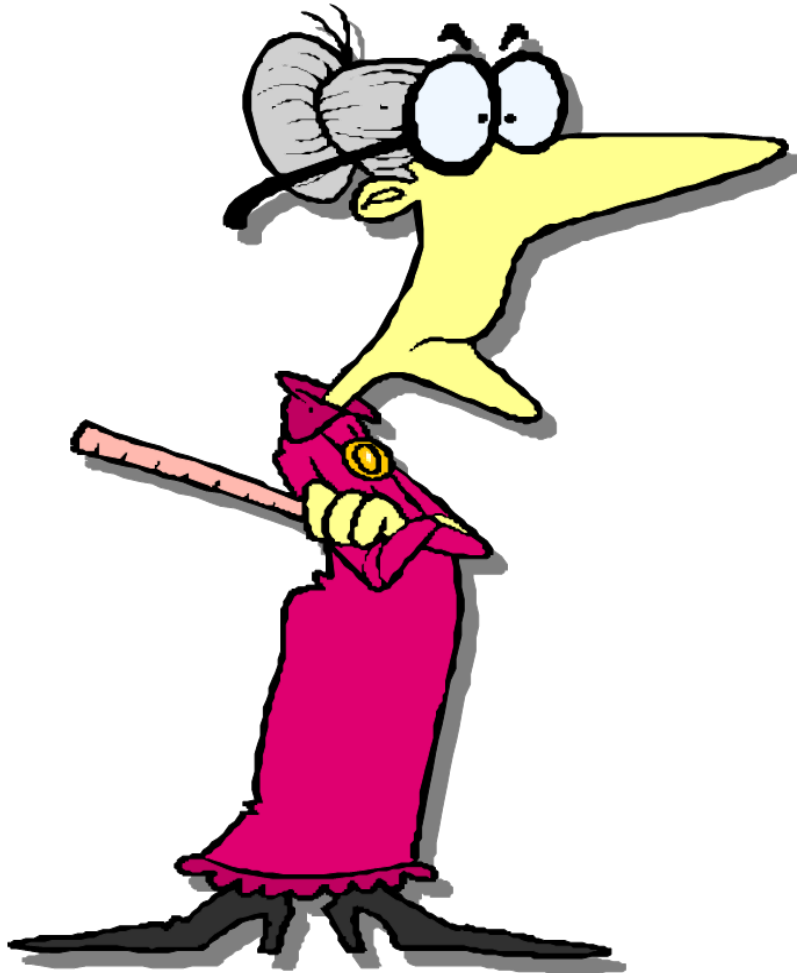
## Verificar antes da Submissão

### **SE TODOS AUTORES LERAM O ARTIGO INTEIRO E ESTÃO DE ACORDO**

- Ortografia de título, nomes dos autores e filiação
- Imprima o artigo (no formato final de submissão): tudo legível
- Tenha certeza absoluta da data e do horário limites para submissão de trabalhos
- MS Word:
  - Numeração das seções e subseções
  - Numeração no texto concorda com a numeração usada em figuras e tabelas
  - Referências cruzadas não foram perdidas dentro do editor

# ESCRITA: Revisão Final

## Exemplo de Formulário de Avaliação



- Relevância (enquadramento no evento)
- Originalidade
- Mérito técnico-científico
- Apresentação
- Organização
- Legibilidade (*readability*)
- Referências

# ESCRITA: EM RESUMO

Título	Título inicial do trabalho
Autores	Aluno + Orientador(es)
Resumo / Abstract	Contexto + problema + objetivos
1. Introdução	Contexto + motivação + problema + estado-da-arte + contribuições + organização
2. O que já existe	Estado-da-arte: comparação
3. NOVIDADE	Contribuições + <b>trabalho desenvolvido</b>
4. Validação	Validação: análise, estudo, experimentos
5. Conclusão	Estado atual + próximos passos
6. Referências	Local (IES) + nacional + internacional

# APRESENTAÇÃO: REVISÃO FINAL

## ENSAIE COM OS DEMAIS AUTORES

- Ortografia de título, nomes dos autores e filiação
- Verifique cada slide: tudo legível
- Tenha certeza absoluta data, horário e local da apresentação
- Chegue **\*ANTES\*** para testar
- **Regras de ouro:**
  - **1 slide por minuto (não corra!)**
  - **Menos tempo = menos texto mais figura/palavra-chave**
  - Exceção: esta apresentação 😊 a qual tem de ser compreensível online

**EVCOMP 2015**

**Escrita e Apresentação de Artigos**

**[mirella@dcc.ufmg.br](mailto:mirella@dcc.ufmg.br)**  
**[www.dcc.ufmg.br/~mirella](http://www.dcc.ufmg.br/~mirella)**

