

# Análise de Critérios de Qualidade da Informação em um PDMS

***Abstract.** This meta-paper describes the analysis and implementation of Information Quality criteria in the recasting of queries in a Peer Data Management Systems (PDMS). This type of system is characterized by having a decentralized infrastructure and dynamic. Therefore there will be an enrichment in the answer to the user who requested the consultation on this architecture.*

***Resumo.** Este meta-artigo descreve a análise e implementação de critérios de Qualidade da Informação no processo de reformulação de consultas em um Peer Data Management Systems (PDMS). Esse tipo de sistema caracteriza-se por possuir uma infra-estrutura descentralizada e dinâmica. Conseqüentemente haverá um enriquecimento na resposta dada ao usuário que solicitou a consulta nessa arquitetura.*

## 1. Introdução

O desenvolvimento de aplicações na Web, com a maior participação dos usuários na produção de informações (Web 3.0), fez crescer a necessidade por novas arquiteturas de sistemas distribuídos. Uma delas é a arquitetura Peer-to-Peer (P2P) ou Peer Data Management Systems (PDMS) que vem sendo bastante utilizada para trocas de informações, especialmente arquivos, na Web. O interesse do projeto concentra-se no problema de gerenciamento de dados, acessados a partir de fontes de dados distribuídas, e a integração dos mesmos para responder às consultas efetuadas nos diversos pontos que compõem um sistema P2P.

O processamento de dados distribuídos figura atualmente como uma das principais arquiteturas de disseminação de informações, visto que a cada dia a massa populacional na internet tende a crescer e, conseqüentemente, as trocas de informações aumentam de forma exponencial. As pessoas estão cada vez mais exigentes por informações que realmente atendam suas necessidades, estejam elas em banco de dados local, distribuídos em diferentes fontes de dados ou até mesmo compartilhadas entre projetos institucionais [Souza 2009]. Devido a essa necessidade crescente, é indispensável um estudo de critérios da qualidade da informação, afinal, esse é o principal fator discriminador de fonte de dados na web [Naumann 2001].

Qualidade da Informação (QI) é comumente definida como um conjunto de critérios para indicar o grau de qualidade geral de uma informação obtida por um sistema [Batista 2008]. A partir de artigos sobre qualidade da informação e qualidade dos dados (*data quality*) [Naumann 2001] chegou a uma verdade indiscutível afirmando que qualidade da informação é o principal fator discriminador de fonte de dados na Web e as métricas de qualidade da informação devem ser levadas em consideração para melhorar os resultados de consultas integradas. Com esta definição, os critérios de QI

tornam-se uma ferramenta bastante útil no enriquecimento de consultas através de sua reformulação.

Este documento está organizado da seguinte forma: A seção 2 mostrará importantes critérios de QI. A descrição de como será a reformulação de consultas em um PDMS com critérios de qualidade da informação é apresentada na seção 3. Na seção 4 encontra-se o funcionamento do módulo de software implementado com critério de QI. A conclusão foi elaborada na seção 5. Finalizando encontram-se as referências bibliográficas.

## 2. Critérios de Qualidade da Informação

Qualidade da Informação (QI) é comumente definida como um conjunto de critérios para indicar o grau de qualidade geral de uma informação em um sistema. [Roth 2005], [Wang 1998], [Nauman 2001]. O crescimento do uso de QI em sistemas de acesso direto a informações tende a necessitar cada vez mais do tratamento de critérios de qualidade no desenvolvimento de sistemas em geral [Batista 2003].

Informação de baixa qualidade é um dos problemas que mais perseguem os usuários de informações distribuídas por fontes de dados autônomas. Este cenário se torna mais forte para a variedade de tipos de usuários de informações da WWW [Nauman 2000].

Alguns trabalhos de pesquisa consideram a QI como sendo um dos aspectos mais importantes para integração de informações, uma vez que a baixa qualidade da informação acarreta problemas para os usuários das informações distribuídas entre as fontes de dados autônomas. Muitos projetos são direcionados a esse fato através da coleta e classificação de critérios de qualidade. Assim, a Qualidade da Informação é um critério multidimensional que depende no todo, da análise de aspectos individuais [Wang 1998]. Em sistemas de integração de dados, QI depende de três fatores maiores: a percepção do usuário, a informação por si só e o acesso às informações. Os três fatores são classificados como o sujeito, objeto e predicado de uma consulta e cada um deles é uma fonte para os metadados de QI ou escores dos critérios de QI. Um escore de QI é um valor associado a um determinado critério de QI [Naumann 2000].

Dessa maneira pode-se observar o quanto é necessário uma reformulação de consultas usando os critérios de QI. Por isso, é preciso o estudo prévio da relação entre elementos em um PDMS e critérios de QI envolvidos. Abaixo segue uma tabela definida por [Souza 2011]:

Tabela 1 – Elementos em um PDMS e critérios de QI.

<b>Elementos de um PDMS</b>	<b>Critérios de QI</b>
Fonte de dados (peers)	Reputação, Fácil Acesso, Tempo de Resposta, Disponibilidade

Esquema (compartilhado por cada peer)	Integridade do Esquema, Minimalidade, Representação Consistente
Mapeamentos (entre peers)	Disponibilidade, Tempo de Resposta, Latência (carga na rede)
Dados	Confiabilidade, Objetividade, Precisão, Integridade

### **3. Reformulação de Consultas em um PDMS com Critérios de Qualidade da Informação**

Em desenvolvimento.

### **4. Funcionamento do Módulo de Software Implementado com Critérios de QI.**

Em desenvolvimento.

### **5. Conclusão**

Em desenvolvimento

### **Referências**

- Batista, M. C. M. Otimização de Acesso em um Sistema de Integração de Dados através do uso de Caching e Materialização de Dados, Dissertação de Mestrado, Centro de Informática – UFPE, 2003.
- Batista, M. C. M. Schema Quality Analysis in a Data Integration System. Tese de Doutorado, Centro de Informática – UFPE, 2008.
- Naumann, F. From Databases to Information Systems – Information Quality Makes the Difference. 6th International Conference on Information Quality (IQ01) Boston, 2001.
- Naumann, F. and Rolker, C. Assessment Methods for Information Quality Criteria. In Proceedings of the Conference on International Quality (IQ00) Boston, 2000.
- Roth A. and Naumann F. Benefit and Cost of Query Answering in PDMS. In Proc. of the Int. Workshop on Databases, Information Systems and Peer-to-Peer Computing (DBISP2P), 2005.

Souza, B. F. F. ; Batista, M. C. M. ; Salgado, A. C. . Critérios de Qualidade da Informação em Reformulação de Consultas em um PDMS: Uma Perspectiva. In: Escola Paraibana de Informática, 2011, João Pessoa. Escola Paraibana de Informática, 2011.

Souza D. Using Semantics to Enhance Query Reformulation in Dynamic Distributed Environments. Tese de Doutorado, Centro de Informática - UFPE, 2009.

Wang, R.Y. A Product Perspective on Total Data Quality Management. Communications of the ACM, Vol. 41, N. 2, 1998.