

# **CONIC-SEMESP**

## 13º Congresso Nacional de Iniciação Científica

Anais do Conic-Semesp. Volume 1, 2013 - Faculdade Anhanguera de Campinas - Unidade 3. ISSN 2357-8904

**TÍTULO:** ANÁLISE DE DADOS ABERTOS: UM ESTUDO DA DISPONIBILIDADE DA INFORMAÇÃO NA ÁREA DA EDUCAÇÃO

**CATEGORIA:** CONCLUÍDO

**ÁREA:** ENGENHARIAS E TECNOLOGIAS

**SUBÁREA:** COMPUTAÇÃO E INFORMÁTICA

**INSTITUIÇÃO:** FACULDADE DE TECNOLOGIA DE CARAPICUIBA

**AUTOR(ES):** DIEGO CORREIA ELIAS, DANIEL PIMENTA TITO, GUILHERME ZAMBELLI DA SILVA, LUCAS CORREA

**ORIENTADOR(ES):** MARIO MARQUES

Realização:



Apoio:



## **ANÁLISE DE DADOS ABERTOS: UM ESTUDO DA DISPONIBILIDADE DA INFORMAÇÃO NA ÁREA DA EDUCAÇÃO**

### **1. RESUMO**

Este projeto de pesquisa tem o objetivo de solucionar o problema de disponibilidade e disseminação da informação para uso coletivo.

Especificamente, realizaremos uma análise sobre os dados na área da educação disponibilizado pelo governo federal através do site dados.gov.br. Desejamos aproximar essas informações à sociedade em geral, principalmente aos envolvidos na área de educação. A proposta é idealizar uma solução que possua grande usabilidade com menor curva de aprendizado possível.

**Palavras-chave: Informação, Educação, Dados**

### **2. INTRODUÇÃO**

Vivemos na geração da informação. De acordo com Brad Frost, designer de front-end, consultor e palestrante de Pittsburgh (EUA), no ano passado foram lançados quase 10 vezes mais livros do que uma década e o número de fotos tiradas no ano de 2012 foi o equivalente a quase 10% de todas as fotos tiradas na história. Temos contato com o equivalente a 174 jornais impressos por dia.

São números impressionantes, mas que revelam a problemática atual: grande parte desse mar de informações não é útil e sim ruído. Descartamos uma significativa porção de dados e não os concentramos no que pode ser valioso. Mesmo com a tecnologia atual nos permitindo acompanhar informações a nível mundial e em tempo real, ainda temos problemas em filtrá-las de forma correta. Por vezes, não conseguimos utilizá-la de forma fácil para tomar decisões confiáveis ou aprimorar processos.

Em novembro de 2011 foi criada a Lei N°12.527/2011 - Lei de Acesso a informações públicas que entrou em vigor dia 16 de maio de 2012. Ela descreve que todas as informações não sigilosas, produzidas pelo poder público, estarão acessíveis à população em geral. A lei fará com que o Brasil entre no grupo dos 91 países que reconhecem a informação gerada e administrada pelos estados um bem

público de acesso livre. Para que fosse possível a disponibilização dessas informações por parte do governo foi criado um site com o endereço dados.gov.br, desenvolvido pela INDA (Infraestrutura Nacional de Dados Abertos). O site possui uma imensa quantidade de dados estatísticos relacionados a diversos tipos de setores brasileiros, como: educação, administração, agricultura, assistência social, ciência e tecnologia, comércio, serviços, entre outros.

O grande problema é a forma como os dados são disponibilizados. Ao entrarmos no site temos que baixar um arquivo grande no formato de texto com as informações pesquisadas. Esses arquivos são disponibilizados, porém são pouco intuitivos para se realizar uma análise objetiva por parte da maioria das pessoas. Dentro desse arquivo existem instruções para utilizar os dados ali contidos, porém é necessário o uso de ferramentas pagas para isto ou ter experiência técnica com base de dados. Perante esse problema, pretendemos apresentar a ideia de uma solução que realiza a síntese dessas informações e permite a sua visualização de forma bem simples e clara.

O enfoque deste trabalho será a educação no Brasil, que é um dos grandes entraves do desenvolvimento do país. Desta forma, condensaremos os dados que já são disponibilizados para transformá-los em informações de grande valia em tomada de decisões.

Com o uso de uma ferramenta de fácil acesso a todos, com gráficos e interface intuitiva, esperamos a melhoria da abstração dos problemas no país em áreas específicas. A informação extraída desse grande volume de dados será útil para diversos setores como: empresas, os cidadãos, ONGs e o próprio governo.

### **3. OBJETIVOS**

#### **3.1. GERAL**

Elaborar uma solução que permita organizar e disponibilizar os dados fornecidos pelo Governo Federal, na área da educação, usando uma interface simples e intuitiva que permita o acesso por um número maior de pessoas.

### 3.2. ESPECÍFICO

- Definir o que são dados abertos;
- Relacionar os dados abertos com as maiores necessidades da sociedade brasileira;
- Identificar a melhor forma de organizar os dados;
- Analisar os dados abertos filtrando-os com foco na educação;
- Apresentar um protótipo de uma solução para exibição do conjunto de indicadores analisados que ilustrarão diversos panoramas acadêmicos do Brasil.

### 4. METODOLOGIA

Pesquisar significa, de forma bem simples, procurar respostas para indagações propostas. Minayo (1993), analisando por um ponto de vista mais filosófico, considera a pesquisa como uma atividade básica das ciências na sua indagação e descoberta da realidade. "É uma atividade de aproximação sucessiva da realidade que nunca se esgota, fazendo uma combinação particular entre teoria e dados" (pág. 23).

Demo (1996) insere a pesquisa como atividade cotidiana considerando-a como uma atitude, "um questionamento sistemático crítico e criativo, mais a intervenção competente na realidade ou o diálogo crítico permanente com a realidade em sentido teórico e prático" (pág.34).

Para Gil (1999, pág.42), a pesquisa tem um caráter pragmático. "É um processo formal e sistemático de desenvolvimento do método científico. O objetivo fundamental da pesquisa é descobrir respostas para problemas mediante o emprego de procedimentos científicos". Nesse sentido, pesquisa é um conjunto de ações, propostas para encontrar a solução de um problema, que tem por base procedimentos racionais e sistemáticos.

Nossa pesquisa é de caráter eminentemente quantitativo. Usaremos dados já disponíveis para a consolidação do objetivo final. Faremos também as descrições e análises dos dados de pesquisa, o processo e seus significados será o foco principal de nossa abordagem. Faremos uma pesquisa telematizada a qual será realizada no

Brasil em que serão observados os dados fornecidos pelo site do governo federal [www.dados.gov.br](http://www.dados.gov.br) referentes aos indicadores no setor de educação. Através desses dados será realizada uma análise e, a partir desta, definiremos uma solução que venha suprir as necessidades. Expondo, de forma mais clara, as informações aos atuantes e interessados no setor analisado.

## 5. DESENVOLVIMENTO

Segundo Fernando Rebouças, no mundo atual, “o volume de informações e dados que são publicados e transmitidos é multiplicado por dez a cada cinco anos.” Isso gera um mar de informações que tende a crescer exponencialmente a cada minuto. Dentro dessa realidade, filtrar a informação útil se torna essencial em meio ao infinito de dados disponibilizados. As ferramentas cada vez mais procuram distinguir a informação útil do ruído criado.

Dentro da realidade em que vivemos filtrar a informação útil se torna essencial em meio ao infinito de dados disponibilizados. As ferramentas cada vez mais procuram distinguir a informação útil do ruído criado.

Segundo a definição da Open Knowledge Foundation (OKF), uma organização sem fins lucrativos que promove conhecimento livre, “dados são abertos quando qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los, estando sujeito a, no máximo, exigência de creditar a sua autoria e compartilhar pela mesma licença”. Estes dados são pautados por três leis e oito princípios. As leis foram propostas para os Dados Abertos Governamentais.

As três leis:

1. Se o dado não pode ser encontrado e indexado na Web, ele não existe;
2. Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado;
3. Se algum dispositivo legal não permitir sua replicação, ele não é útil.

Em 2007 foram definidos na Califórnia, Estados Unidos da América, os 8 princípios:

1. **Completo.** Todos os dados públicos são disponibilizados. Dados são informações eletronicamente gravados, incluindo, mas não se limitando a documentos, bancos de dados, transcrições e gravações audiovisuais.

Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou controle de acesso, reguladas por estatutos.

2. **Primários.** Os dados são publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.
3. **Atuais.** Os dados são disponibilizados o quão rapidamente seja necessário para preservar o seu valor.
4. **Acessíveis.** Os dados são disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.
5. **Processáveis por máquina.** Os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.
6. **Acesso não discriminatório.** Os dados estão disponíveis a todos, sem que seja necessária identificação ou registro.
7. **Formatos não proprietários.** Os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo.
8. **Livres de licenças.** Os dados não estão sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos.

Com base nas leis e fundamentos, teremos que aprimorar a disponibilidade e facilidade do acesso aos dados abertos para que possamos usá-la de forma eficiente. Já a eliminação das informações incorretas e desatualizadas ocorre por meio de uma fonte atualizada e confiável, neste caso o site dados.gov.br.

O site já disponibiliza uma vasta quantidade de dados, mas que ainda continuam presos a dificuldade de entendimento por parte das pessoas. São liberados os dados que necessitam de alguma ferramenta para sua abstração e por fim a sua utilização de maneira adequada.

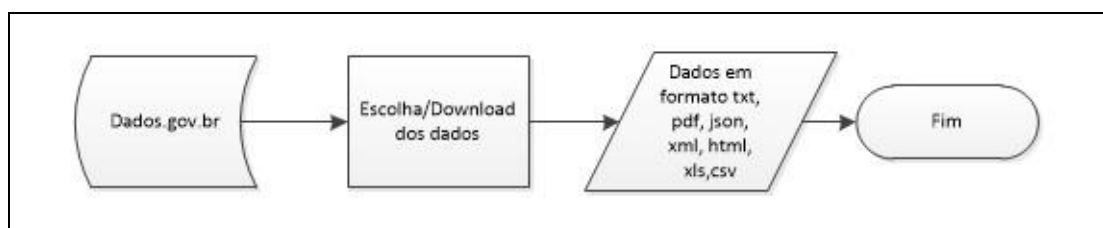


Figura 1 – Fluxo do panorama atual na disponibilização de dados abertos pelo site dados.gov.br

A solução visa, na primeira parte, obter essas informações diretamente do site supracitado na categoria educação que estarão em forma bruta sem viabilidade de compreensão. Estes são disponibilizados nos formatos: texto, pdf, json, xml, html, xls, csv.

A ferramenta fará o download do arquivo específico que utilizaremos como fonte de dados. Os arquivos alimentarão a base de dados MySQL, de acordo com sua categoria e relacionamento. A população dessa base é vital para que a interface funcione, pois todas as informações serão extraídas de lá e não dos arquivos. Isso permite a utilização de uma base consistente para extração de relatórios, históricos e consultas mais aprimoradas para o usuário, com muita confiabilidade.

A aplicação será baseada na linguagem Java (J2EE) que possui grande visibilidade no mercado atual integrado com o contêiner Apache Tomcat que atua com a base de dados através do framework de persistência Hibernate.

O projeto se fundamenta no Design Patterns Data Access Object (DAO), separando as regras de negócio das regras de acesso ao banco de dados. Usaremos o modelo de arquitetura Model-View-Controller (MVC) e apresentaremos esses dados em uma interface intuitiva de extensão .xhtml com o auxílio do framework Java Server Faces (JSF) e implementação da tecnologia PrimeFaces. Utilizando em seu desenvolvimento a IDE Eclipse Indigo. O objetivo é a simplicidade da tela sem poluição de informações.

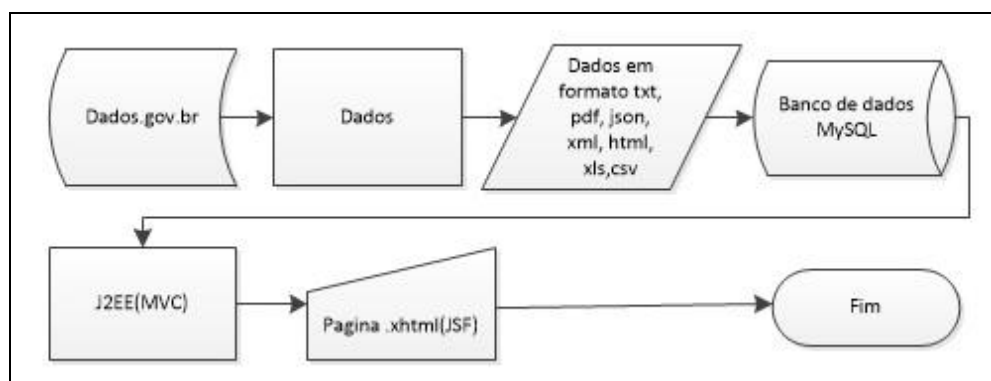


Figura 2– Fluxo do panorama projetado para o tratamento dos dados abertos.

## 6. RESULTADOS

Baseado nos princípios levantados pela pesquisa, desenvolvemos um protótipo para demonstração da interface.



Figura 3- Exemplo da interface gráfica idealizada pela pesquisa

Os arquivos abastecem a base de dados com as informações. As informações são tratadas e disponibilizadas via interface gráfica que formata dois tipos de gráficos. A interface é bem limpa e livre de excesso de informação. Isso torna o processo de análise de dados inegavelmente mais simples. Por meio de gráficos qualquer pessoa ou instituição poderá usufruir dos dados para informação o tomada de decisão.

Conforme a imagem, a interface possui dois tipos de gráficos. Um de setores e outro em barras laterais. As abas superiores contêm os filtros por ano a escolha do usuário. O filtro localizado abaixo da aba refina o filtro do gráfico de setores. O gráfico do lado direito visa complementar a pesquisa com dados pertinentes. A tela também se adéqua a resolução da tela usada.



O resultado demonstra que o protótipo se atende as regras de ergonomia de software e usabilidade. Com apenas dois filtros é possível realizar a consulta, portanto a solução se mostrou eficaz como forma de disponibilizar dados.



Figura 4 - Outro exemplo da interface gráfica idealizada pela pesquisa

O site dados.gov.br, possui uma área para disponibilização de projetos que usam de seus dados para fins específicos, nenhum deles possui foco na área da educação. Com isso, especificamos uma solução que atendesse essa demanda.

## 7. CONSIDERAÇÕES FINAIS

A grande quantidade de dados disponibilizada atualmente é incrível. Em todos os lugares temos acesso a uma grande quantidade de informações sem mesmo nos darmos conta. Mas até onde essas informações nos são apresentadas de forma produtiva e simples?

Nosso trabalho abordou essa questão trazendo a grave deficiência no acesso e disponibilização dos dados abertos para as partes interessadas. Após a análise da

interface utilizada no site [www.dados.gov.br](http://www.dados.gov.br) podemos concluir que o processo para a disponibilização dessa informação se torna muito burocrático e desconfortável para o usuário final. Ela destoa da realidade na sociedade atual onde a informação é essencial e a facilidade de acesso é imprescindível. A melhor forma de contornar esse problema é a recriação dessa interface fazendo com que a interação do usuário seja mais dinâmica e amigável. O trabalho teve o objetivo de solucionar esses problemas projetando uma solução que atenda às necessidades do usuário.

## 8. FONTES CONSULTADAS

O que são dados abertos? Disponível em < <http://dados.gov.br/dados-abertos/>>. Data de Acesso: 18/04/2013

Domingos. Educação no Brasil – Tragédia o Desafio?. Disponível em <<http://profdomingos.blogspot.com.br/2011/07/opinio-educacao-no-brasil-tragedia-ou.html>>. Data de Acesso: 11/04/2013

ABRUCIO Fernando. A lei de acesso a informação e a cidadania. Disponível em< <http://revistaepoca.globo.com/opinio/fernando-abrucio/noticia/2012/06/lei-de-acesso-informacao-e-cidadania.html>>. Data de Acesso: 15/04/2012.

DEMO, Pedro. Pesquisa e construção de conhecimento. Rio de Janeiro: TempoBrasileiro, 1996.

REBOUÇAS, Fernando. Excesso de Informação. Disponível em <<http://www.infoescola.com/comunicacao/excesso-de-informacao/>>. Data de Acesso: 12/08/2013.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 1999.

MINAYO, Maria Cecília de Souza. **O desafio do conhecimento**. São Paulo: Hucitec, 1993

ALTAIR, Alexandre e HEINZELMANN Décio. **Programação Java para a Web**. São Paulo: Novatec, 2010.

Que são dados abertos? Disponível em: <http://br.okfn.org/dados-abertos/>.  
Data de Acesso: 10/07/2013.

GRANJEIRO, W. José. Conheça a lei de acesso a informação. Ela vai cair na prova. Disponível em:  
<<http://congressoemfoco.uol.com.br/opinia0/c0lunistas/conheca-a-lei-de-acesso-a-informacao-ela-vai-cair-na-prova/>> Data de Acesso: 14/08/2013.

EAVES, Daves. The Three Laws of Open Government Data. Disponível em: <  
<http://eaves.ca/2009/09/30/three-law-of-open-government-data/>> Data de Acesso:  
10/07/2013.