

Ministério da Saúde  
**FIOCRUZ**  
Fundação Oswaldo Cruz



**ICICT**  
Instituto de Comunicação e Informação  
Científica e Tecnológica em Saúde

**Sistema de Recuperação da Informação (SRI) da plataforma CNPq no  
Diretório do Grupo de Pesquisa**

**por**

**FLÁVIO AUGUSTO DE MENEZES ALENCAR**

**FIOCRUZ**

Projeto apresentado ao Instituto de Comunicação e Informação Científica e Tecnológica em Saúde da Fundação Oswaldo Cruz como requisito parcial para obtenção do título de Especialista em Informação Científica e Tecnológica em Saúde.

**Orientador: Leonardo Cruz da Costa**, doutorado em  
Ciência da Informação pela UFF/IBICT

**Rio de Janeiro, Dezembro/2011**

## **1 RESUMO**

O Diretório dos Grupos de Pesquisa no Brasil é uma base de dados que contém informações referentes a pesquisas em atividade no País, essa base é atualizada pelos integrantes do Grupo e pelas instituições integrantes e mantém sua memória histórica por ter seus dados acumulados ao longo do tempo. Porém, apesar de possuir um volume de informações muito ricas, apresenta limitações estratégicas nas suas buscas e inter-relações. Quanto maior o universo a ser estudado, maiores são as restrições de correlação das informações por necessitar de tratamento de um grande volume de dados não tratados.

O Grupo de Ciências da Saúde é o segundo maior cadastrado nessa base e, devido ao seu grande volume de informações, demanda mais recursos humanos e de tempo para que os dados e informações possam ser garimpados, analisados e processados. Esse fenômeno ocorre principalmente se tivermos como objetivo critérios de busca não previstos pelo sistema da base ou quisermos relacionar dados e informações que não foram usadas como critérios de filtros pelo mesmo.

È proposto diversos itens de melhora e otimizar o trabalho do pesquisador no DGP do CNPq, é mais um esforço para qualificar a interação da pesquisa com a informação, dando mais um salto nos índices de produtividade.

## **2 PALAVRAS CHAVE**

Recuperação de dados, Informação, tecnologia, Grupos de Pesquisa, análise, busca

### **3 INTRODUÇÃO**

Um projeto criado em 1991 e iniciado em 1992 pelo CNPq, o Diretório dos Grupos de Pesquisa no Brasil, é uma base de dados que contém as informações das atividades no País referentes a pesquisas em atividade, essa base é atualizada pelos integrantes do Grupo e pelas instituições integrantes. Estas informações dizem sobre os recursos humanos constituintes dos grupos, as linhas de pesquisas em andamento, às especialidades dos conhecimentos, aos setores de aplicação envolvidos, à produção científica e tecnológica e aos padrões de interação com os setores produtivos.

As instituições estão localizadas em universidades, instituições isoladas de ensino superior, institutos de pesquisa científica, institutos tecnológicos e laboratórios de pesquisa e desenvolvimento de empresas estatais ou ex-estatais, exceto os grupos localizados nas empresas do setor produtivo. Os grupos se organizam em torno de uma liderança, abrigados em uma instituição previamente autorizada pelo CNPq, pois a relação do Diretório com o Grupo é institucional.

Cada participante do Grupo possui uma página individual acessada através de senha onde são imputados os dados e informações referentes a dados pessoais, institucionais e acadêmicos (Currículo Lattes), e sua atuação na instância onde se encontra no grupo.

Esta base é um eficiente instrumento para intercâmbio e a troca de informações, convidando ao aprofundamento do conhecimento por meio de algumas possibilidades e tem um papel importante na preservação da memória da atividade científica e tecnológica no Brasil.

A Base Corrente (atual) é atualizada diariamente, refletindo os dados dos grupos, líderes, pesquisadores e estudantes presentes e só podem ser recuperados os trabalhos certificados pelos dirigentes institucionais.

As bases censitárias extraídas periodicamente (a partir do ano 2000), são conteúdos estáticos coletados da base corrente bianualmente, eles retratam o estado da pesquisa no Brasil e para que tal processo ocorra um maior grau de

fidelidade ele e anunciados previamente pelo CNPq para que os grupos atualizem seus dados na Plataforma Lattes.

Para participar do DGP é preciso que a instituição que o abriga seja previamente autorizada pelo CNPq. Os participantes têm que se enquadrar nas regras vigentes desde janeiro/2005 conforme itens abaixo:

1. Pertencer a:

- Universidades federais, estaduais, municipais e privadas;
- Instituições de educação superior – IES não universitárias (centros universitários, faculdades integradas, faculdades isoladas, institutos, escolas, centros de educação tecnológica, etc.) que possuam pelo menos um curso de pós-graduação – mestrado ou doutorado – reconhecido pela CAPES/MEC;
- Institutos públicos de pesquisa científica;
- Institutos tecnológicos públicos e centros federais de educação tecnológica;
- Laboratórios de pesquisa e desenvolvimento de empresas estatais.

2. Ter solicitado seu ingresso pela autoridade institucional de pesquisa ou superior por meio do Fale Conosco do site do CNPq, selecionando o assunto Plataforma Lattes, Diretório dos Grupos de Pesquisa.

A recuperação de dados do Diretório do Grupo de Pesquisa (DGP) do CNPq, pela identificação das informações, segundo o critério de pesquisa escolhido e filtrado pelo usuário, tem como objetivo, criar um banco de dados que contemplará inferências entre os diversos dados do tipo:

- Informações sobre recursos humanos;
- As linhas de pesquisa em andamento;
- As especialidades do conhecimento;

- Os setores de atividades envolvidos;
- A produção científica e tecnológica e artística dos pesquisadores e estudantes que integram os grupos .
- Os padrões de interação com o setor produtivo.

O DGP é uma ferramenta de planejamento e gestão das atividades de ciências e tecnologia, sendo um eficiente instrumento para intercâmbio e troca de informações, capaz de responder quem é quem, onde se encontra os setores do grupo de pesquisa, o que estão fazendo, e o que produziram recentemente. Nas várias instâncias de organização político-administrativa, a base de dados é uma fonte rica de informações, além das já disponíveis sobre os grupos, seu caráter censitário possibilita várias possibilidades de estudos, e a construção de amostras permitira dar resposta sobre campos não cobertos pelos dados, por exemplo, financiamentos, avaliação quantitativa da produção científica e tecnológica, interações entre grupos de pesquisa e o setor produtivo, e outras.

(Fonte: <http://dgp.cnpq.br/diretorioc/>)

## 4 JUSTIFICATIVA

Desde a criação do DGP do CNPq, só a partir de 2002 foi disponibilizado censos bianuais só então com visões sistêmicas e com dados estáticos. A comunidade científica tem garimpado informações para estudos de diversas naturezas no Base Corrente online. Esta recuperação de dados tem se mostrado custosa, pois, leva muito tempo para extrair os dados das paginas filtradas, a atividade de copiar e colar, torna a pesquisa mais lenta e aumenta seu custo tanto quanto for o seu volume. Além deste trabalho, estes dados poderão ter redundâncias que possivelmente poderá ser identificada, só depois da recuperação e análise feita, Estes dados também não estão inter-relacionados, o que não permite uma visão sistêmica que seria de grande utilidade para uma análise mais criteriosa. Uma prospecção em uma base com relacionamentos e sem repetições de informações, modelada, daria maior integridade, permitindo identificação de padrões e uma visualização efetiva de redes de interação entre autores, linhas de pesquisa e outras relações, permitindo identificar aspectos relevantes ao trabalho.

O conjunto dos grupos de pesquisa na área de Ciência da Saúde é o segundo maior em quantidade de pesquisadores e linhas de pesquisa, segundo o censo de 2010, e vem crescendo anualmente, por isso é de suma importância desenvolver melhores interfaces para este diretório.

Grande área	Grupos G	Linhas de Pesquisa L	Pesquisadores P	Estudantes E	Técnicos T	L/G	P/G
<u>Ciências Agrárias</u>	<u>2.699</u>	<u>13.609</u>	<u>25.547</u>	<u>27.249</u>	4.781	5	9,5
<u>Ciências Biológicas</u>	<u>3.108</u>	<u>14.585</u>	<u>23.390</u>	<u>32.081</u>	4.638	4,7	7,5
<u>Ciências da Saúde</u>	<u>4.573</u>	<u>16.728</u>	<u>34.375</u>	<u>41.446</u>	6.681	3,7	7,5
<u>Ciências Exatas e da Terra</u>	<u>2.934</u>	<u>13.418</u>	<u>19.645</u>	<u>22.140</u>	2.385	4,6	6,7
<u>Ciências Humanas</u>	<u>5.387</u>	<u>16.813</u>	<u>41.196</u>	<u>47.939</u>	3.214	3,1	7,6
<u>Ciências Sociais Aplicadas</u>	<u>3.438</u>	<u>10.422</u>	<u>23.877</u>	<u>23.569</u>	1.891	3	6,9
<u>Engenharias</u>	<u>3.548</u>	<u>16.115</u>	<u>25.275</u>	<u>30.274</u>	3.324	4,5	7,1
<u>Linguística, Letras e Artes</u>	<u>1.836</u>	<u>5.025</u>	<u>12.140</u>	<u>14.157</u>	590	2,7	6,6
<b>TOTAIS</b>	<b>27.523</b>	<b>106.715</b>	<b>205.445</b>	<b>238.855</b>	<b>27.484</b>	<b>3,9</b>	<b>7,5</b>

\* Em geral há dupla contagem no número de pesquisadores, estudantes e técnicos, tendo em vista que o indivíduo que participa de mais de um grupo de pesquisa foi computado mais de uma vez (ver técnica 1).

O DGP do CNPq só permite que seja consultado por palavra(s) chave(s) junto com outros critérios, impedindo que se verifique, por exemplo, quantos Grupos existem em determinado estado e/ou instituição ou qualquer outro critério de pesquisa que não inclua palavra chave. Não leva em consideração uma ordenação relevante na listagem, quando poderia considerar as informações de Estado, Instituição, Pesquisador ou outra informação dentre outros critérios.

Podemos elencar algumas outras deficiências na utilização das buscas:

Nos estudos em Ciências da Saúde, por exemplo, poderia ser feito diversos trabalhos de análise com determinado critério de forma mais rica. Quando pesquisamos a 'Malaria' como critério apenas, temos 118 grupos, não sabendo de quais Estados são eles e poderia vir agrupado por instituição e/ou aplicação, mas, vem agrupado sempre por ordem alfabética do nome do grupo, o que não permite uma análise mais relevante.

## 5 REFERENCIAL TEORICO:

O termo “recuperação de informação” (RI) remete a um processo pelo qual se selecionam documentos, a partir de um acervo, em função de uma necessidade do usuário. Os Sistemas de Recuperação da Informação (SRI), são considerados como uma interface entre uma coleção de recursos informacionais, em meio impresso ou não, e uma população de usuários, constituindo, segundo Hjørland (2003) como um dos principais atores na Ciência da Informação.

Um SRI tem por objetivo prover acesso fácil aos documentos, satisfazendo a necessidade de informação de seus usuários. Segundo Lancaster (1991) é um fenômeno complexo que engloba documentos e sua descrição, perguntas e um mecanismo que permita o “casamento” dessa descrição com as perguntas. Um SRI opera recuperando um conjunto de registro de documentos que reflete um relacionamento com a consulta apresentada pelo usuário. A recuperação é baseada na similaridade de conteúdo que “casa” a palavra chave ou termos que representam o assunto do documento com a questão do usuário.

Sob um aspecto mais amplo, a RI pode compreender procurar informações específicas em documentos e encontrar metadados<sup>1</sup> que descrevam documentos. Engloba ainda a busca de textos, sons, imagens ou dados em coleções de dados estruturadas – tais como bancos de dados relacionais – e não estruturadas – tais como Intranets e a Internet.

A Web, cada vez mais, se torna um repositório universal do conhecimento e cultura humana permitindo o compartilhamento de ideias e informações em uma escala nunca vista anteriormente (BAEZA-YATES E RIBEIRO-NETO, 1999). A visão da recuperação da informação é igualmente aplicada para WEB, mas de forma diferente do que ocorre em uma biblioteca ou centro de documentação onde conta-se com serviços de análise, indexação, catalogação e armazenamento das informações sobre sua descrição de forma a representar adequadamente os documentos. A falta de um padrão indexatório a ser seguido prejudica a recuperação de documentos na Web, devido à dificuldade de

---

<sup>1</sup> **Metadados**, ou **Metainformação**, são dados sobre outros dados. Um item de um metadado pode dizer do que se trata aquele dado, geralmente uma informação inteligível por um computador. Os metadados facilitam o entendimento dos relacionamentos e a utilidade das informações dos dados.

identificação de quais termos são pertinentes e relevantes para a uma determinada consulta.

Em RI faz-se também necessário estudar e estabelecer critérios para as formas de representação, organização e apresentação da informação com intuito de facilitar o acesso do usuário à informação que lhe interessa.

Em suma, é uma área de natureza interdisciplinar que lida com a representação, o armazenamento, a organização e o acesso à informação, valendo-se de técnicas e conhecimentos derivados de áreas tão diversas como a ciência da computação, a matemática, a biblioteconomia, a ciência da informação, a psicologia cognitiva, a linguística e a estatística entre outras.

Dada à tarefa de busca, pode-se dizer que sob um ponto de vista funcional, um SRI lida com objetos e consultas. Uma consulta é a descrição da necessidade de informação de um usuário, formulada e submetida a um SRI, normalmente, pelo próprio usuário. Um objeto é uma entidade que mantém ou armazena informação em um banco de dados – por exemplo, um documento. Consultas do usuário são casadas com objetos armazenados no banco de dados, de modo que objetos que atendem à necessidade de informação do usuário – objetos relevantes – sejam selecionados. Como o foco de um sistema de RI é a informação, uma certa imprecisão nos resultados é tolerada dada à dificuldade em interpretar e representar a necessidade do usuário e o conteúdo dos objetos.

Apesar de sua abrangência e importância, a área de RI foi vista, por muitos anos, como de menor interesse, restrita aos bibliotecários e aos especialistas em informação. Esta visão, entretanto, modificou-se completamente com o surgimento do World Wide Web, em meados dos anos 1990. A Web tornou-se um imenso repositório de conhecimentos e cultura, tornando-se rapidamente acessível a todos em todos os lugares. Sua adoção modificou profundamente o estilo de vida das pessoas, a forma como se comunicam, interagem e buscam informações. Entretanto, sua vastidão e sua natureza livre e desestruturada tornam extremamente difícil para os usuário a tarefa de encontrar informação útil.

Desde que os objetos primários da área de RI é a redução dos problemas associados ao excesso de informação, ela, tornou-se o centro das atenções. Suas tecnologias passaram a ser vistas como soluções promissoras para se lidar com a imensidão de informações na Web. Como resultado, muitas destas tecnologias ganharam enorme visibilidade, especialmente aquelas utilizadas em máquinas de busca, o produto mais conhecido da moderna RI. O impacto social e econômico dos avanços nesta área foi tal que as empresas dominantes na era da informação – Google, Yahoo!, Microsoft e outros – oferecem aos usuários máquinas de busca para a Internet como parte essencial de sua estratégia de negócios.

Atualmente, a pesquisa na área de RI evoluiu para incluir tópicos tão diversos quanto modelagem de informação, classificação e agrupamento de documentos, arquiteturas de sistemas, interfaces de visualização, filtragem de informação, linguagens, etc..

Nos últimos anos, vários trabalhos dedicaram-se ao problema de extração de dados de páginas da Web. A importância do tema deve-se ao fato de os dados, uma vez extraídos, poderem ser manipulados de forma semelhante a instâncias armazenadas em um banco de dados tradicional.

Existem hoje os mais variados tipos de ferramentas para extração de dados da web, e cada vez mais, novas ferramentas continuam aparecendo. Cada uma dessas ferramentas usa sua estratégia própria de extração de dados. Dentre elas, podemos destacar: linguagem natural, linguagens e gramáticas, aprendizagem de máquina, recuperação de informação, ontologias etc. Como consequência, essas abordagens apresentam características e capacidades diferentes, tornando difícil uma comparação direta entre elas. Existem linguagens especializadas de auxílio na construção de wrappers (uma classe em programação orientada a objetos) desenvolvidas como alternativa às linguagens de propósito geral, como Perl e Java, que eram usadas para esta tarefa.

Ferramentas baseadas na estrutura HTML são ferramentas que dependem de características estruturais inerentes dos documentos HTML para realizar a extração de dados. Antes de executar o processo de extração, essas ferramentas transformam o documento em uma árvore que reflete a hierarquia das tags HTML.

Em seguida, são geradas regras de extração, de forma automática ou parcialmente automática., que são aplicadas à árvore.

Ferramentas baseadas em linguagem natural (NLP – Processamento em Linguagem Natural) buscam capturar a semântica de um texto, de forma a extrair os dados relevantes dos documentos.

Este trabalho toca apenas na ponta do iceberg, pois, a internet, pelo tamanho que tem, carece de mais indexações, padronização e melhoramento em suas máquinas de busca (web semântica). Além disso existe o agravante de ter diversos tipos de mídias e linguagens o que intensifica a necessidade de uma padronização.

## **6 OBJETIVOS:**

Tem a intenção de suprir a comunidade científica e tecnológica com instrumento capaz informar quem, onde, quando e o que esta sendo feito com relação a um determinado trabalho de pesquisa em diversas instâncias, financiamento, região,

setores produtivos, abrangência e outras. Estas informações permitirão uma avaliação da produção e desenvolvimento do país em diversas áreas. Estes dados por serem recorrentes permitirão também um monitoramento do desenvolvimento das pesquisas no Brasil em quaisquer período.

Desenvolveremos uma interface (software livre) interativa para recuperação de informação disponibilizada no DGP do CNPq que gere um banco de dados estruturado com todas as informações contidas na base online, permitindo extrair e representar conhecimentos, identificar e validar padrões das atividades de pesquisas, trazendo informações sobre um grupo com interesses comuns.

As vantagens provenientes da utilização desta ferramenta, cuja relevância esta na consolidação das produções científicas no país.

#### **OBJETIVOS ESPECÍFICOS:**

- Nova interface, rica em critérios de pesquisa e opções para ordenação
- Opção para criação de um Banco de Dados relacionado com critério de hierarquia.
- Recuperação de filtros anteriormente submetidos.
- Recuperação de Banco de Dados criado.
- Gerador de relatórios diversos com base no Banco de Dados.

#### **7 METODOLOGIA:**

A informação na web carece de estruturação e definição de critérios de relevância do conhecimento. Embora subjetivos, diversos esquemas conceituais e modelos teóricos vêm tentando explicar esta estruturação, utilizando-se das modernas

tecnologias baseadas em indução, em ontologias e outras. Tanto o método automático como o semi-automático, referenciado anteriormente em documentos HTML, podem ser representados como árvores através do modelo Document Object Model – DOM.

Árvores DOM são usadas para representar e manipular as informações de páginas HTML. Este modelo é essencial para os métodos de extração, é um modelo orientado a objeto especificado pelo WC3 e utilizado para representar documentos HTML e XML. Este modelo define uma estrutura lógica dos documentos e o meio pelo qual um documento pode ser acessado e manipulado. Com DOM é possível criar documentos, navegar por sua estrutura e adicionar, modificar ou apagar elementos e conteúdos. Todos os componentes que podem ser encontrados em um documento HTML (ou XML) podem ser manipulados através deste modelo.

O DOM trata a informação contida dentro dos documentos como um modelo de objetos hierárquico, de forma semelhante a uma árvore, cujo acesso à informação desses documentos pode ser realizado através de interações com a referida árvore.

Como toda árvore, as árvores DOM apresentam uma estrutura constituída por vários tipos de nós. Essa estrutura possui um único nó na raiz, sendo que todos os demais nós são seus descendentes. Todos os nós da árvore DOM, com exceção do nó raiz, têm um pai. Além disso, cada nó possui uma lista de nós filhos, sendo que, em alguns casos, esta lista pode ser vazia.

Como exemplo, consideremos um pequeno fragmento de código de um documento HTML, que representa uma tabela, a qual está sendo apresentada na figura 1. A partir deste código obtemos a sua árvore DOM ilustrada pela figura 2.

**Figura 1:** Fragmento do código de uma página HTML filtrada pelo usuário.

Identificação	Recursos Humanos	Linhas de Pesquisa	Indicadores do Grupo
---------------	------------------	--------------------	----------------------

**Identificação**

**Dados básicos**

Nome do grupo: Ciência da Informação: Cognição, Mediação e Construção do Conhecimento

Status do grupo: **certificado pela instituição**

Ano de formação: 2003

Data da última atualização: 04/12/2010 02:01

Lider(es) do grupo: Aida Varela Varela -

Maria Isabel de Jesus Sousa -

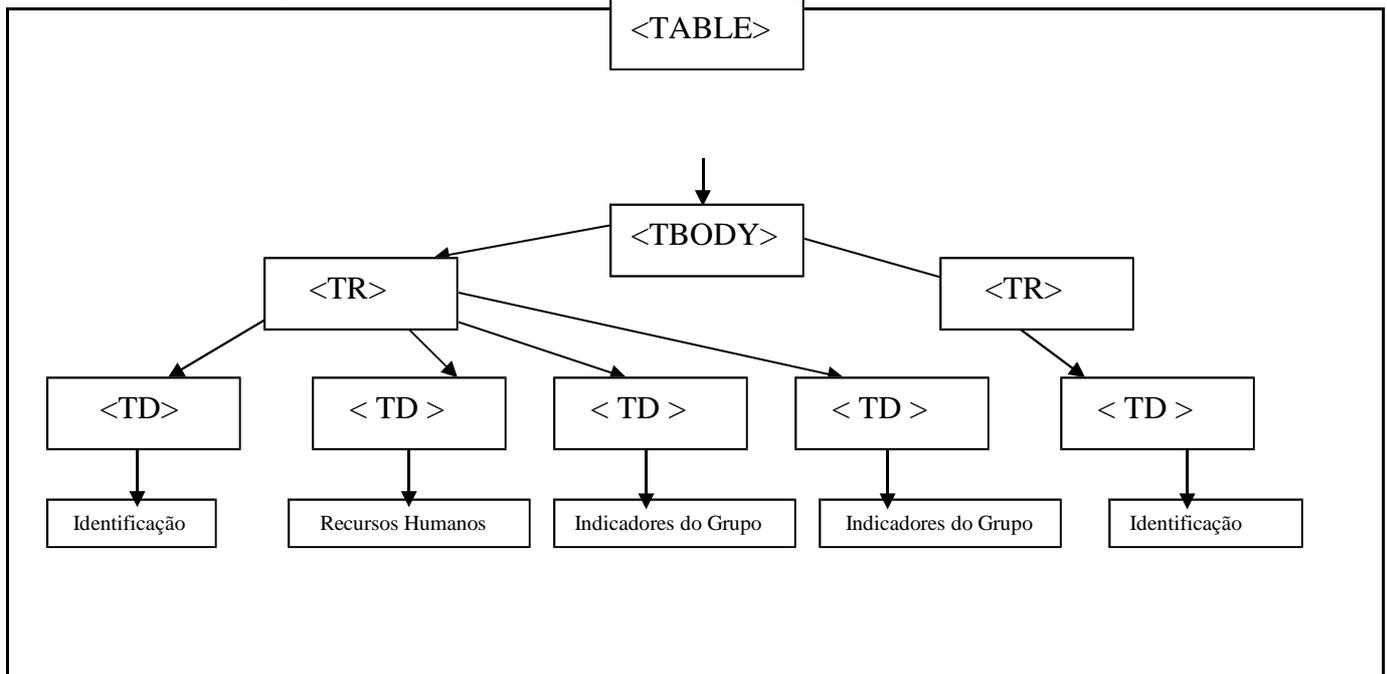
Área predominante: Ciências Sociais Aplicadas: Ciência da Informação

```

<table border="0" width="100%">
  <tbody><tr bgcolor="#FFDC75">
    <td align="Center" width="25%">
      <a href="#identificacao"><font style="Font-Family: Arial; Font-Size: 8.5pt"><b>Identificação</b></font></a>
    </td>
    <td align="Center" width="25%">
      <a href="#rh"><font style="Font-Family: Arial; Font-Size: 8.5pt"><b>Recursos Humanos</b></font></a>
    </td>
    <td align="Center" width="25%">
      <a href="#linhapesq"><font style="Font-Family: Arial; Font-Size: 8.5pt"><b>Linhas de Pesquisa</b></font></a>
    </td>
    <td align="Center" width="25%">
      <a href="#indicadores"><font style="Font-Family: Arial; Font-Size: 8.5pt"><b>Indicadores do Grupo</b></font></a>
    </td>
  </tr>
  <tr>
    <td colspan="4">&nbsp;</td>
  </tr>
  <tr bgcolor="#99CCCC">
    <td colspan="4">
      &nbsp;<a name="identificacao"><font style="Font-Family: Arial; Font-Size: 8.5pt"><b>Identificação</b></font></a>
    </td>
  </tr>
  <tr bgcolor="#FFDC75">
    <td colspan="4">
      &nbsp;<font style="Font-Family: Arial; Font-Size: 8.5pt"><b>Dados básicos</b></font>
    </td>
  </tr>
  <tr>
    .. 1 ..
  </tr>

```

**Figura 2: Representação DOM de uma tabela exemplo.**



A partir dessas informações fornecidas pelo usuário, podemos, através da árvore DOM, gerar caminhos nessa árvore que começam na sua raiz e vão até as folhas onde se encontram os dados desejados. Cada um destes caminhos indica a localização do dado na página, portanto, pode ser usado como regra para extração dos dados selecionados pelos usuários.

Outro método baseia-se na extração da informação tomando como base a estrutura do HTML das páginas que as contem após a filtragem do usuário. Esta análise da estrutura deverá ser realizada em conjunto com as expressões regulares que determinara o seguimento de texto que contem a informação relevante, indicando a sua localização e extração e assim sucessivamente onde será identificado as informações que comporá a base de dados.

Estes esquemas e modelos estão relacionados entre si com interseções dinâmicas não hierarquizadas com vários olhares

Mas o importante é o significado do armazenamento, distribuição e recuperação e principalmente o estudo, treinamento e capacitação para evolução teste novo paradigma de pensar o pensamento com as inteligências artificiais.

## **8 RESULTADOS ESPERADOS:**

Busca-se reduzir as redundâncias, criando assim uma interface evoluída que permita ao usuário perceber ligações que antes estavam engessadas em função da forma de apresentação e limitação em volume. Permitir ao usuário perceber tendências, incidências e inferências que uma visão abrangente proporciona, e assim qualificar melhor a informação.

Construir vantagens competitivas significa criar e obter novos saberes, e divulgar nas áreas pertinentes, interpretar e integrar com outros já existentes e, principalmente, utilizar seus benefícios para alcançar eficiência e eficácia.

Serão os mais bem sucedidos nesses novos ambientes organizacionais aqueles que têm habilidades para obter e integrar informações e é isso que pretendemos com essa nova ferramenta, disponível à todos livremente.

Este projeto procura apresentar uma forma de aplicação das técnicas de “recuperação da informação” - RI de modo a otimizar a seleção e apreensão da informação para o usuário, tornando mais rica as perspectivas de visão da informação em uma dimensão mais abrangente e inter-relacionada.

## 9 REFERÊNCIAS CONSULTADAS:

DEV MEDIA. **Criando uma árvore DOM**. Disponível em:

<<http://www.devmedia.com.br/post-1831-Criando-uma-arvore-DOM.html>>.

Acesso em: 15 out. 2011.

LEARNING, Quick & Easy. **Introdução a HTML DOM**. Disponível em:

<<http://www.academictutorials.com/Pt/html-dom/html-dom-introduction.asp>>.

Acesso em: 15 out. 2011.

PRESSMAN, Roger S.; LOWE, David. **Engenharia Web**. Rio de Janeiro:

Ltc, 2009. 416 p. Capítulos 1, 7 e 8.

CNPQ (Brasil/rj/rio de Janeiro) (Org.). **Portal CNPq**. Disponível em:

<<http://dgp.cnpq.br/diretorioc>>. Acesso em: 10 out. 2011.

JOHNSON, J. David. **Gestão de redes de conhecimento**. São Paulo: Senac,

2009. 440 p. Tradução: Janaina Marcoantonio.

FREIRE, Gustavo Henrique de Araujo; FREIRE, Isa Maria. **Introdução a Ciência da Informação**. João Pessoa: Editora Universitária da Ufpb, 2010. 128 p.

SILVA, Eduardo Fraga do Amaral e. **UM SISTEMA PARA EXTRAÇÃO**

**DE INFORMAÇÃO EM REFERÊNCIA BIBLIOGRÁFICA BASEADO EM**

**APRENDIZAGEM DE MÁQUINA**. 2004. 104 f. Discertação (Mestrado) -

Curso de Pós-graduação em Ciência da Informação, Departamento de

Centro de Informática, Universidade Federal de Pernambuco, Recife, 2004.

CARDOSO, Olinda Nogueira Paes. **RECUPERAÇÃO DE**

**INFORMAÇÃO**. Disponível em:

<<http://www.dcc.ufla.br/infocomp/artigos/v2.1/olinda.pdf>>. Acesso em: 10

de outubro 2011.

MORENO, Fernanda Passini; LEITE, Fernando César Lima; ARELLANO, Miguel

Ángel Márdero. Acesso livre a publicações e repositórios digitais em ciência

da informação no Brasil. **Perspect. Ciênc. Inf.**, Belo Horizonte, v. 11, n. 1,

p.82-94, 2006. Quadrimestral.

<b>10. CRONOGRAMA:</b>	Meses
1. Estudar as linguagens;	2
2. Montar os layouts de entrada;	2
3. Validação das variáveis;	2
4. Montar os layouts de saída;	2
5. Teste de usabilidade e resultados.	2
6. Gerenciamento de evolução e mudanças.	1
7. Relatório final.	1

**11. ORÇAMENTO:**

Curso da linguagem.....	R\$ 2.000,00
Infra estrutura oferecida pela Instituição.....	R\$ 0,0
Atividade Mensal (12 meses).....	R\$ 1.200,00
<b>TOTAL .....</b>	<b>R\$16.400,00</b>