

---

# **Utilização de técnicas de hipermídia adaptativa para seleção de amostras em pesquisa digital de opinião: um modelo organizacional**

Fausto Neri da Silva Vanin (Mestre)

Curso de Bacharelado em Ciência da Computação – Universidade Tuiuti do Paraná

Caio César Ferreira (Graduando)

Curso de Bacharelado em Ciência da Computação – Universidade Tuiuti do Paraná

---

---

## Resumo

Este trabalho faz uma análise da representação de dados e metadados no contexto da representação digital de pesquisas de opinião. Esta representação leva em consideração um ambiente que utiliza a pesquisa de dados por pesquisa de opinião com um mecanismo para obtenção do perfil do respondente. A tarefa de resolução do perfil do respondente, neste caso, resolve-se de forma análoga às utilizadas em Hipermídia Adaptativa, sendo que a obtenção do perfil do respondente pode ser comparada à modelagem da Base de Modelos de Usuários (BMU) e a escolha do instrumento para o respondente pode ser comparada à adaptação de apresentação e navegação da HA. Como resultado, tem-se um novo padrão XML de dados que leva em consideração três fatores importantes e suas relações: pesquisa de opinião, ambiente hipermídia, e extração de padrões de dados.

**Palavras-chave:** pesquisa de opinião; XML; hipermídia adaptativa.

## Abstract

This work is an analysis on data and metadata in digital social research context. This representation considers an environment that uses the social research as a mechanism to obtain the respondents profile. The task of obtain the user profile, in this case, is similar to the Adaptive Hypermedia adaptation techniques. The respondent profile obtaining can be related to User Base Modeling and the choose of the social survey tool can be related to the adaptation of presentation and navigation. As result we propose a new XML data format that takes three major factors in account: social survey, hypermedia environment and pattern extraction.

**Key words:** social survey, XML, adaptive hypermedia.

---

---

## Introdução

A pesquisa digital de opinião possui, entre as suas principais questões, a representação dos instrumentos, a seleção dos respondentes e o modelo de interação com o usuário.

A organização dos dados envolvidos em um ambiente digital torna-se, cada vez mais, importante hoje em dia devido à interoperabilidade de ambientes e ferramentas. Esta é inclusive uma tendência apontada pela Web 2.0 (O'Reilly, 2005; Musser e O'Reilly, 2006) e pela SOA (*Service Oriented Architecture*) (Durvasula, 2006a; Durvasula, 2006b; Durvasula, 2006c). Outros estudos visam propôr estruturas organizacionais para dados e metadados.

Em Cannataro e Pugliese (2001), é proposta uma linguagem XML (*eXtensible Markup Language*) para representação do ambiente hipermediático chamada XAHM (*XML-based Adaptive Hypermedia Model*). Este modelo permite descrever: a) a estrutura lógica e conteúdos de um ambiente de hipermídia adaptativa, definindo que partes deste ambiente devem ser adaptadas; e b) a lógica do processo de adaptação, distinguindo a adaptação feita por restrições tecnológica ou por características do usuário.

---

Neste trabalho, o ambiente é visto como um grafo ponderado (dígrafo) em que os nós representam o conteúdo; os links, as arestas; e os pesos das arestas representam uma medida de correlação entre os nós. O domínio da aplicação é modelado como um espaço ortogonal composto por três dimensões: Comportamento do usuário, ambiente externo (localização, peculiaridades culturais, etc.) e tecnologia (tipo de rede, velocidade de conexão, etc.).

Em Barclay et al. (2002), é apresentada a linguagem SuML (*Survey Markup Language*) que é uma aplicação XML para representar questionários digitais. A linguagem compõe uma suite de desenvolvimento que também incorpora uma biblioteca de programação para a linguagem Perl e um conjunto de transformações XSL para os questionários. A linguagem suporta apenas questões de múltipla escolha e prevê estruturas para armazenamento das perguntas e também das respostas, sendo que a visualização dos resultados deve ser feita pela utilização de *XPath*<sup>1</sup> para acessar os elementos contidos nos arquivos de resposta. A linguagem também incorpora suporte para a inclusão de elementos gráficos às questões.

Em DMG (2008), é apresentada a linguagem de marcação *Predictive Model Markup Language*. Esta linguagem já se tornou um padrão de mercado e tem

<sup>1</sup> XPath: Linguagem para inspeção de dados em arquivos XML.

sido suportada por diversos aplicativos de mineração de dados. Ela permite definir, além dos cabeçalhos específicos de cada contexto, os dicionários de dados, transformações de dados e a modelagem de conhecimento que será utilizada, suportando diversas formas de representação como Árvores de Decisão, Algoritmos de Agrupamento, Redes Neurais, por exemplo. Esta abrangência torna a linguagem muito útil a uma série de contextos.

O trabalho de Frias-Martinez et al. (2006) faz um estudo das técnicas de Mineração de Dados e Aprendizagem de Máquina para a criação automática de Modelos de Usuários como suporte a ambientes de Hipermídia Adaptativa. As técnicas foram divididas em supervisionadas e não supervisionadas, sendo que o processo de modelagem automática de usuários é dividido nas seguintes fases: a) coleta de dados; b) pré-processamento; c) extração de padrões; e d) validação e interpretação.

Entre as principais técnicas avaliadas estão os algoritmos de k-médias, SOM (*Self-organizing Maps*), agrupamentos difusos e regras de associação entre as não supervisionadas. Entre as supervisionadas, foram avaliadas as técnicas de árvores de decisão, k-NN, redes neurais e SVM (*Support Vector Machines*).

O autor aponta três critérios importantes para decidir

qual técnica utilizar para modelar o conhecimento: 1) Ter ou não dados rotulados; 2) o tipo de atividade a ser realizada (classificação ou recomendação); 3) a necessidade ou não de representar o conhecimento de uma forma que seja amigável ao ser humano.

A partir destes estudos, é possível relacionar a estrutura de um ambiente de pesquisa digital de opinião com o de um ambiente de HA. Esta é a proposta deste trabalho, e será apresentado nas próximas seções.

### *Objetivos*

O objetivo principal deste trabalho é a criação de um padrão de representação de pesquisa digital que permita:

- Representar os seguintes instrumentos de pesquisa: questionário, teste e entrevista;
- Suportar a geração de dados adequados à extração de padrões;
- Apresentar um modelo hipermídia de representação de instrumentos que permita e sua operacionalização em ambientes hipermídia arbitrários.

## 2 Metodologia

A partir da literatura consultada, é possível fazer uma analogia entre os sistemas de HA e a pesquisa de opinião. Nesta analogia, a extração do perfil do respondente pode

ser diretamente relacionada à montagem da Base de Modelos de Usuário, sendo que os resultados apontados por Frias-Martinez (2006) podem ser utilizados como base para os testes a serem executados nesta pesquisa. A apresentação do instrumento de pesquisa ao respondente, neste caso, pode ser resolvida da mesma forma que, na HA, é feita a adaptação de apresentação e de navegação. Assim, teríamos o instrumento de pesquisa (entrevista, questionário, teste) como um elemento de mídia dentro do ambiente hipermediático. Exemplo, a pesquisa de opinião.

Neste caso, o que será adaptado para apresentação ao usuário serão os instrumentos de pesquisa e, baseado na interação do usuário com este instrumento e das respostas obtidas, será possível completar cada vez mais o perfil deste usuário.

Entre os padrões de dados estudados, o padrão de representação dos instrumentos de pesquisa, SuML, é o que possui menor aderência aos objetivos da pesquisa. Entre as principais fragilidades do modelo estão:

- Suporta apenas questões de múltipla escolha;
- Representa apenas o questionário como instrumento;
- Não especifica metadados de avaliação (*score*) aos resultados obtidos;
- Não define os tipos de dados das respostas

obtidas. Esta característica é importante por oferecer a interoperabilidade com PMML, por exemplo.

Portanto, um novo padrão para representação dos instrumentos será criado. Este padrão será uma aplicação de XML organizada em três níveis:

1. Pergunta: elemento mínimo da pesquisa. Pode ser:
  - Dissertativa;
  - Verdadeiro/Falso;
  - Múltipla escolha.
2. Instrumento: o instrumento que irá aplicar a pergunta. Esta camada também é responsável por conter os possíveis encadeamentos<sup>2</sup> entre respostas. Inicialmente, os instrumentos suportados serão:
  - Questionário;
  - Entrevista;
  - Teste.
3. Pesquisa: a pesquisa em si, onde os dados das respostas podem ser pontuados. Separar a pesquisa do instrumento permite a aplicação de vários instrumentos em uma mesma pesquisa e que uma mesma pesquisa seja avaliada sob diferentes contextos, aumentando a flexibilidade.

A estrutura proposta se relaciona com as linguagens PMML e XAHM como mostrado na figura 1. O modelo de interação com o usuário é coberto pela interação entre estes três padrões, sendo que toda a relação estímulo-resposta é gerenciada pela XAHM, enquanto a PMML contém os modelos de dados

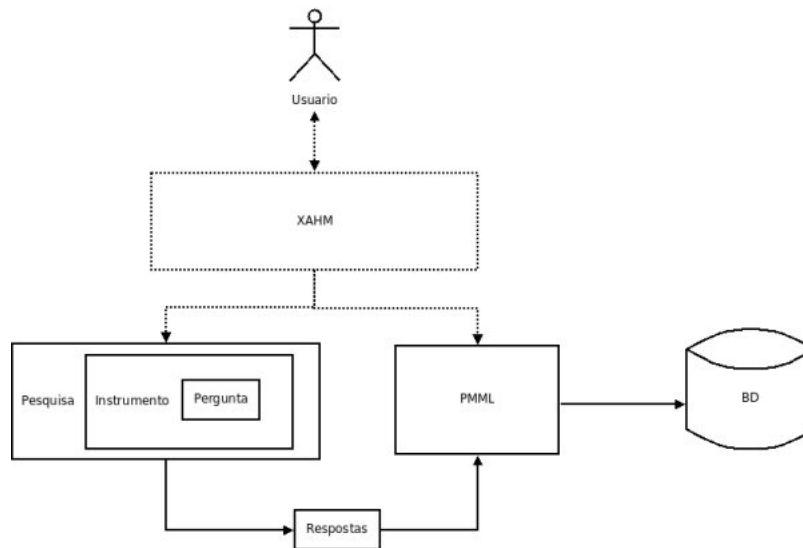


Figura 1: Estrutura de dados proposta.

dos usuários para montagem do perfil e a estrutura sugerida agrega a representação da pesquisa de opinião como mecanismo tanto de coleta de dados, como de adaptação do perfil do usuário à medida que este com

2 O encadeamento de perguntas se dá quando uma determinada pergunta está atrelada a uma resposta da pergunta anterior.

ela interage.

A seguir, cada elemento da estrutura proposta será descrito em separado.

### 2.1 Perguntas

Este elemento é composto pelos seguintes itens:

- question: escopo mais externo que contém os seguintes atributos:
  - type: o tipo da pergunta. Pode ser *single*, *multiple* e *boolean*
  - description: descrição da pergunta
  - author: autor da pergunta
  - input: a entrada de dados do usuário. Possui os seguintes atributos:
    - type: pode ser *radio*, *checkbox*, *text*
    - id: identificação única da entrada
    - implies: representa a associação entre opções de uma mesma questão.
    - value: valor padrão para a pergunta
    - maxlength: tamanho máximo para resposta do tipo *text*

O quadro 1 contém um exemplo de questão.

#### Quadro 1: Exemplo de Pergunta

```
<?xml version='1.0' encoding='utf-8' ?>
<question type='multiple'
title='Em que área você considera a pesquisa mais útil?'>
  <author>Fausto Vanin</author>
  <description>
    Marque quantas alternativas jogar importantes.
    Caso queira acrescentar mais algum item, marque a opção Outros.
  </description>
  <input type='checkbox' id='produtos_novos' implies='none'
caption='Produtos novos' />
  <input type='checkbox' id='avaliacao_impacto' implies='none'
caption='Avaliação de impacto' />
  <input type='checkbox' id='ident_tendencias' implies='none'
caption='Identificar Tendências' />
  <input type='checkbox' id='pos_venda' implies='none' caption='Pós-
venda' />
  <input type='checkbox' id='outros' implies='5' caption='Outros' />
  <input type='text' id='outros' value='' implies='none' maxLength='30' />
</question>
```

### 2.2 Instrumento

Os padrões para instrumento possuem os seguintes itens:

- survey: escopo mais externo que contém o restante dos elementos e possui os seguintes atributos:
  - type: O tipo do instrumento. Pode ser: *questionnaire*, *test* ou *interview*
  - title: Título
  - date-of-creation: Data de criação do instrumento
- owner: O proprietário do questionário

- description: Descrição
  - disclaimer: Espaço para alguma mensagem a ser mostrada ao usuário antes de responder
  - question: Uma questão que compõe o instrumento. Possui os seguintes atributos:
    - file: Arquivo que contém os dados da pergunta, como mostrado no quadro 1
    - id: Identificação da questão
    - depends-on: Relação de dependência entre as questões. Valor na forma *id-pergunta@id-resposta=valor*
- O quadro 2 mostra um exemplo de instrumento.

### 2.3 Pesquisa

O instrumento de pesquisa é um elemento que é composto pelos seguintes itens:

- survey-meta-data: Escopo mais externo que contém os dados da pesquisa
  - group: escopo que contém diversos instrumentos. Possui os seguintes atributos:
    - id: identificador único da pesquisa
- survey: um dos instrumentos do grupo. Possui os seguintes atributos:
  - file: arquivo do instrumento
  - id: identificador único do instrumento

#### Quadro 2: Exemplo de Instrumento de Pesquisa

```
<survey type='questionnaire' title='Levantamento de dados'
date-of-creation='09/19/2008'>
  <owner>Fausto Vanin</owner>
  <description>
    Este questionário tem por objetivo saber a sua opinião sobre a tv
    digital.
  </description>
  <disclaimer>
    Os dados informados nesta pesquisa serão armazenados de forma anônima
    e têm
    apenas fins de pesquisa, sendo que os autores se comprometem a
    informar os
    respondentes interessados sobre o andamento do processo de pesquisa e
    do
    uso feito dos dados informados.
  </disclaimer>
  <question file='opinioao01.xml' id='p1' dependsOn='none' />
  <question file='opinioao02.xml' id='p2' dependsOn='p1@conhece=sim' />
  <question file='opinioao03.xml' id='p3' dependsOn='p1@conhece=nao' />
</survey>
```

- value: peso do instrumento na pesquisa
- keyword: palavra-chave associada ao instrumento
- question: escopo que representa uma das questões do instrumento. Possui os atributos:
  - id: identificador da questão no instrumento
  - value: peso da questão no contexto do instrumento
- answer: resposta esperada. Possui os seguintes atributos:
  - value: o conteúdo esperado da resposta
  - score: o peso da resposta para a questão



O quadro 3 mostra um exemplo de pesquisa.

#### 2.4 Produtos

Para operar de forma complementar aos padrões projetados, foram criados os seguintes produtos:

1. Transformações XSL (XML Stylesheet Language): Estas transformações foram criadas para transformar perguntas da estrutura padrão para a estrutura HTML (Hypertext Markup Language), o que viabiliza a veiculação online da pesquisa.

2. Algoritmos para coleta de dados: Estes algoritmos foram desenvolvidos na linguagem PHP (Personal Home Page) e permitem fazer a coleta dos dados do ambiente online para o armazenamento em arquivo.

#### 2.5 Resultados Obtidos

Em relação aos objetivos estabelecidos, os resultados obtidos foram:

#### Quadro 3: Exemplo de pesquisa

```
<?xml version='1.0' encoding='utf-8' ?>
<survey-meta-data>
  <group id='1'>
    <survey file='perguntas.xml' id='1' value='20'>
      <keyword>análise</keyword>
      <keyword>mercado</keyword>
      <question id='1' value='16'>
        <answer value='sim' score='10' />
        <answer value='nao' score='6' />
      </question>
      <question id='2' value='4'>
        <answer value='jornais' score='1' />
        <answer value='revistas' score='1' />
        <answer value='tv' score='1' />
        <answer value='internet' value='1' />
      </question>
    </survey>
  </group>
</survey-meta-data>
```

- Representar os seguintes instrumentos de pesquisa: Este objetivo foi cumprido com a criação do padrão XML para representação em três níveis: 1) pergunta; 2) instrumento; e 3) pesquisa. Este padrão contempla três instrumentos diferentes (testes, questionários e entrevistas), visando a integração com ambientes de hipermídia adaptativa.
- Suportar a geração de dados adequados à extração de padrões: A existência dos metadados no nível de pesquisa, para os padrões XML criados, permitem

que pesos sejam associados às respostas obtidas de cada instrumento. A combinação entre os pesos e as respostas deverá gerar dados no formato PMML para aplicação de algoritmos de Reconhecimento de Padrões.

- Apresentar um modelo hipermídia de representação de instrumentos que permita e sua operacionalização em ambientes hipermídia arbitrários: O modelo criado, pela utilização de linguagem XML, permite a sua combinação com transformações XSL, por exemplo, para geração de dados de mídia em diversos formatos. Além disto, este tipo de arquivo pode ser lido facilmente por bibliotecas padrão de programação que acessam, de forma estruturada, o conteúdo<sup>3</sup> e permitem a geração dos elementos de pesquisa de opinião em diversos formatos de apresentação.

É facilmente perceptível a necessidade de uma aplicação computacional que faça a leitura dos dados e a geração dos devidos formulários de resposta a serem inseridos no ambiente hipermídia. Apenas desta forma será possível avaliar totalmente a adequação do padrão no ambiente e a geração de dados para o reconhecimento de padrões.

Um ambiente hipermídia com estas características está em fase de desenvolvimento e irá servir de base

para estas análises e também para trabalhos futuros.

## Conclusões

Este trabalho apresentou uma estrutura organizacional para pesquisa digital de opinião. Esta estrutura provém padrões de dados e metadados em três níveis e está estruturada de forma análoga às aplicações de HA.

Grande parte do desenvolvimento deste trabalho foi no sentido de prover uma estrutura organizacional adequada a outras pesquisas, desenvolvendo padrões de representação que possam se tornar referência e que funcionem adequadamente com outros padrões já existentes.

Mesmo, assim, acredita-se que uma boa contribuição científica está presente na representação dos instrumentos de pesquisa como elementos multimídia. Relacionando a pesquisa de opinião à HA, abre diversas possibilidades de desenvolvimento no campo da interação homem-máquina e dos sistemas inteligentes, provendo mecanismos de interação hipermediáticos adaptativos, coletando, neste caso, a opinião do respondente de diversas formas diferentes.

Este trabalho dá início a diversas possibilidades de estudo futuras, dando a sua grande contribuição neste momento às representações de dados, metadados

<sup>3</sup> Entre estas bibliotecas padrão podemos citar o DOM (Document Object Model) que é uma padronização para acesso a arquivos XML implementado para diversas linguagens populares como PHP, Python e Java, por exemplo:

e algoritmos de extração de padrões, mas servindo de suporte para futuros desenvolvimentos, como apresentado a seguir.

### *Trabalhos Futuros*

No momento atual da pesquisa, é possível perceber novas possibilidades de desenvolvimento para esta pesquisa, caminhando em áreas que possam proporcionar estudos interessantes tanto no ramo da HA quando de Reconhecimento de Padrões. A seguir, apresentamos algumas destas possibilidades.

- Testar as técnicas com um tema de pesquisa que seja pertinente a uma determinada área de pesquisa já consolidada: Este fator contribuiria muito para a qualidade dos testes realizados por ter a necessidade de atender uma demanda real.
- Suportar diferentes elementos de mídia nos padrões de dados estipulados: Assim como em Barclay et al. (2002), é interessante que os instrumentos, desde a sua primeira versão já suporte diferentes elementos

de mídia, até para se relacionarem de forma mais adequada aos padrões do ambiente hipermídia, como o XAHM (Cannataro e Pugliese, 2001).

- Criação de instrumentos de coleta que explorem melhor os elementos multimídia e extraiam as informações baseados em um modelo de interação: Esta característica permite o desenvolvimento de estudo na área de *advergames*, aprendizagem auxiliada por computador e realidade virtual, por exemplo.
- Estudos de hipermídia adaptativa: a partir de um modelo, que relaciona os instrumentos de coleta com o modelo de interação, torna-se interessante a criação de ambientes de Hipermídia Adaptativa que permita colocar estas técnicas em prática e avaliar o ganho de desempenho proporcionado pelo modelo.
- Extração de padrões: Novas formas de representação de conhecimento podem ser avaliadas em relação aos já estudados em Frias-Martinez et al. (2006), como modelos Neuro-fuzzy e Redes Bayesianas, por exemplo.

## Referências

- BARCLAY, M., *et al.* *Survey Markup Language (SuML) – How To*, 2002.
- BRUSILOVSKY, P. Methods and techniques of adaptive hypermedia. In: *Adaptive Hypertext and Hypermedia*. [S.l.]: Kluwer Academic Publishers, 1996. p. 87–129.
- BRUSILOVSKY, P.; MILLÁN, E. The adaptive web: Methods and strategies of web personalization. lecture notes in computer science. In: New York: Springer-Verlag, 2007. v. 4321, cap. User Models for Adaptive Hypermedia and Adaptive Educational Systems, p. 3–53.
- CANNATARO, M.; PUGLIESE, A. Xahm: an xml-based adaptive hypermedia model and its implementation. In: *In Proc. of 3rd Workshop on Adaptive Hypertext and Hypermedia, Arhus*. [S.l.: s.n.], 2001. p. 252–263.
- DMG, D. M. G. *Predictive Modeling Markup Language*. 8, 2008. On line.
- DURVASULA, S. e. a. Part 1: Why service oriented architecture. In: *SOA Practitioners Guide*. [S.l.: s.n.], 2006a.
- \_\_\_\_\_. e. a. Part 2: Soa reference architecture. In: *SOA Practitioners Guide*. [S.l.: s.n.], 2006b.
- \_\_\_\_\_. e. a. Part 3: Introduction to services lifecycle. In: *SOA Practitioners Guide*. [S.l.: s.n.], 2006c.
- FRIAS-MARTINEZ, E.; CHEN, S. Y.; LIU, X. Survey of data mining approach to user modeling for adaptive hypermedia. *IEEE Transactions On Systems, Man and Cybernetics. Part C: Applications and Reviews*, v. 36, n. 6, 2006.
- GIL, A. C. *Métodos e técnicas de pesquisa social*. 5. ed. São Paulo: Editora Atlas, 1999.
- JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, p. 4–37, 2000.
- LÉVY, P. *L'intelligence collective - Pour une anthropologie du cyberspace*. [S.l.]: La Découverte Éditions, 1994.
- MUSSER, J.; O'REILLY, T. *Web 2.0 Principles and Best Practices*. [S.l.], 2006.
- O'REILLY, T. *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. 9 2005. Online. Disponível em: <<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>.
- SELLTIZ, C. *et al.* *Métodos de Pesquisa nas Relações Sociais*. São Paulo: Herder, 1967.