

**UNIVERSIDADE FEDERAL DE SANTA CATARINA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO**  
**CONHECIMENTO**

**PEDRO GONZAGA VALENTE**

**Aplicações híbridas para a criação de conteúdo jornalístico na internet**

Dissertação submetida à Universidade Federal de Santa Catarina  
para a obtenção do Grau de Mestre em Engenharia e Gestão do Conhecimento.

**Orientador:**

Prof. Nilson Lemos Lage

Florianópolis

2007

**PEDRO GONZAGA VALENTE**

**Aplicações híbridas para a criação de conteúdo jornalístico na internet**

Esta Dissertação foi julgada adequada para obtenção do Título de “Mestre em Engenharia”, Especialidade em Engenharia e Gestão do Conhecimento e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento.

Florianópolis, 30 de maio de 2007.

---

Prof. Roberto Pacheco, Dr.  
Coordenador do Programa – UFSC

**Banca Examinadora:**

---

Orientador Prof. Nilson Lemos Lage, Dr.  
Engenharia e Gestão do Conhecimento – UFSC.

---

Prof. Eduardo Barreto Vianna Meditsch, Dr.  
Engenharia e Gestão do Conhecimento – UFSC.

---

Prof. João Bosco da Mota Alves, Dr.  
Engenharia e Gestão do Conhecimento – UFSC.

---

Prof. Maria José Baldessar, Dra.  
Departamento de Jornalismo – UFSC.

## AGRADECIMENTOS

*À Universidade Federal de Santa Catarina, em especial ao Departamento de Jornalismo.*

*Ao orientador Prof. Nilson Lage, por ter aceitado levar este desafio até o fim, apesar das dificuldades.*

*Aos professores Luís Alberto Scotti e Maria José Baldessar, do Departamento de Jornalismo, pelo incentivo constante.*

*Aos professores e colegas do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento.*

*Aos meus familiares e à Sara.*

*A todos aqueles que direta ou indiretamente contribuíram para a realização desta pesquisa.*

## SUMÁRIO

LISTA DE FIGURAS.....	5
LISTA DE FIGURAS.....	5
LISTA DE TABELAS.....	6
RESUMO.....	7
ABSTRACT.....	8
CAPÍTULO 1 – INTRODUÇÃO.....	9
1.1 Apresentação do Problema de Pesquisa.....	9
1.2 Objetivo geral do trabalho.....	10
1.3 Objetivos específicos do trabalho.....	10
1.4 Justificativa Teórica.....	10
1.5 Escopo do Trabalho.....	11
1.6 Estrutura do trabalho.....	11
CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA.....	13
2.1 Teoria da Informação.....	13
2.2 Jornalismo de precisão.....	15
2.2.3 História e Conceito .....	16
2.2.4 RAC.....	18
2.3 Cognição e Entendimento.....	19
2.3.1 O cérebro.....	19
2.3.2 O custo cognitivo.....	21
2.4 Economia da Informação em Rede.....	26
2.5 A Emergência da Web Semântica.....	30
2.5.1 O que é web semântica.....	30
2.5.2 Folksonomias e alternativas à web semântica.....	32
2.5.3 Emergência.....	34
2.5.4 Entendimento máquina-a-máquina.....	35
2.6 Web 2.0.....	37
2.7 Perfil do Novo Profissional.....	39
2.8 Jornalismo Online.....	42
2.8.1 Histórico.....	42
2.8.2 O modelo centrado na matéria.....	44
CAPÍTULO 3 - O FERRAMENTAL.....	49
3.1 Metodologia.....	49
3.2 Tecnologias.....	49
3.2.1 APIs e Web Services.....	50
3.2.2 Governo eletrônico.....	55
CAPÍTULO 4 - MODELO PROPOSTO.....	58
4.1 Apresentação do Modelo.....	58
4.2 Descrição das Etapas do Modelo.....	60
4.2.1 Extração de dados dos fornecedores.....	60
4.2.2 Cruzamento e organização.....	60
4.2.3 Apresentação e interação.....	61
5.1 Conclusões gerais.....	62
5.2 Sugestões para trabalhos futuros.....	63
REFERÊNCIAS.....	65
BIBLIOGRAFIA.....	68

## LISTA DE FIGURAS

Figura 2.1: Página principal do Google: baixo custo cognitivo.....	24
Figura 2.2: Página principal do UOL: alto custo cognitivo.....	25
Figura 2.3: Mapa conceitual da Web 2.0.....	38
Figura 3.1: Matriz de combinação de APIs do <i>site</i> Programmable Web.....	54
Figura 3.2: Exemplo de uso de APIs em uma Aplicação Web Híbrida.....	56
Figura 4.1: Modelo genérico de Aplicação Web Híbrida.....	61

## LISTA DE TABELAS

Tabela 2.1: As gerações do jornalismo na internet.....	44
--	----

## RESUMO

A evolução das tecnologias de compartilhamento de dados como APIs e *Web Services* tem ocorrido em paralelo à crescente cultura de colaboração dos usuários e aos níveis cada vez maiores de personalização de serviços *online*. Esta dissertação tem como objetivo descrever como fazer jornalismo cruzando dinamicamente dados e informações de diversas origens na internet, de modo a transformá-los em conteúdo jornalístico relevante. Para isso, define o modelo de Aplicação Web Híbrida, que é composto pelos seguintes elementos: dados (números, listagens de nomes, endereços ou qualquer resultado de uma consulta a bases de dados) recursos (mapas, vídeos, fotografias, áudio ou texto noticioso) – e a retroalimentação proveniente da interação ativa ou passiva dos leitores com a aplicação. Define-se os conceitos envolvidos, modelos de utilização e as tecnologias empregadas.

### Palavras-chave:

1. Jornalismo *online*
2. *Web Services*
3. Cruzamento de dados.

## ABSTRACT

*The evolution of data sharing technologies such as APIs and Web Services happens in parallel with the growth of an user collaboration culture and of the customization levels of online services. This research's objective is to describe how to make online journalism by dynamically merging data and information from different sources, transforming it in relevant journalistic content. To accomplish this, a model for a Hybrid Web Application is described, that has the following elements: data (numbers, name listings, addresses or any result from a database request), resources (maps, videos, photos, audio or news copy) – and the feedback provided by the active or passive interaction of the readers with the application. There's the description of the underlying concepts, usage models and technologies involved.*

### **Keywords:**

1. *Online journalism*
2. *Web Services*
3. *Data merging.*

## CAPÍTULO 1 – INTRODUÇÃO

### 1.1 Apresentação do Problema de Pesquisa

De acordo com Sperber e Wilson (1995, p.48), informações relevantes para um indivíduo são aquelas que, somadas às informações já disponíveis, produzem informações novas, até então não disponíveis. O conceito de conteúdo jornalístico relevante também pode ser descrito desta maneira. Para escrever uma notícia é preciso levar em conta a novidade dos fatos, a proximidade deles com o leitor, a abrangência do acontecimento, sua importância e as consequências que dele podem advir.

Philip Meyer, jornalista e professor da Universidade de Carolina do Norte, nos Estados Unidos incentiva o uso de bases de dados informatizadas por jornalistas (MEYER, 1991). Ele defende que métodos estatísticos e cruzamentos de dados podem ajudar os jornalistas a obter informações relevantes para a produção de reportagens.

Na internet, aos poucos nos aproximamos da integração total entre pessoas e informação, visualizada primeiro por Vannevar Bush (1945) e mais tarde na web semântica de Tim Berners-Lee (2001). A idéia principal de Lee é colocar a comunicação inteligente entre máquinas a serviço dos usuários da rede. Desenvolvedores já trabalham com isso atualmente: criam *sites* que captam informações provenientes de fontes diferentes, recombina as informações e obtêm conteúdo relevante para o usuário.

Em grande parte, as informações utilizadas por desenvolvedores provêm de APIs<sup>1</sup> de órgãos públicos e empresas privadas. Elas colocam à disposição – em padrões aceitos internacionalmente – informações tais como dados estatísticos, listagens de nomes, endereços ou resultados de consultas a bases de dados. Estão ainda ao alcance do usuário recursos mais elaborados: mapas, vídeos, fotografias, áudio, texto noticioso ou informativo. Uma aplicação desse tipo, que combina dados de diversas fontes, é chamada de Aplicação Web Híbrida.

Outro item importante é a disposição de superar a convicção de que o conteúdo jornalístico deve ser apresentado apenas como um grande bloco de texto. “Os avanços tecnológicos e os sistemas sociais para lidar com eles não evoluem no mesmo ritmo. Quando velhos sistemas sociais e culturais são aplicados a novas maneiras de fazer as coisas, o

---

<sup>1</sup> API – Application Programming Interface. É uma camada de abstração que permite acesso a funções internas de um software por agentes externos.

encaixe é por vezes desconfortável e até doloroso” (MEYER, 1991). A internet apresenta novas possibilidades de apuração, combinação e apresentação do conteúdo jornalístico. Cabe aos jornalistas/programadores explorar toda essa potencialidade.

Tendo em vista o cenário exposto acima, a pergunta que este trabalho pretende responder é a seguinte: Como tirar proveito dos recursos da internet para produzir conteúdo jornalístico na forma de Aplicações Web Híbridas?

## **1.2 Objetivo geral do trabalho**

Descrever como é possível fazer jornalismo cruzando dinamicamente dados e informações de diversas origens na internet de modo a transformá-los em conteúdo jornalístico relevante.

## **1.3 Objetivos específicos do trabalho**

- Descrever o estado da arte das tecnologias envolvidas
- Propor um modelo de Aplicação Web Híbrida jornalística

## **1.4 Justificativa Teórica**

Diversos autores, entre eles Baldessar, questionam o papel do jornalista diante das novas tecnologias:

As novas tecnologias da informação desencadearam uma discussão sobre a identidade e sobrevivência das profissões que eram responsáveis pela mediação simbólica. Nesse contexto, o que é ser jornalista na atualidade? (...) considerando o Jornalismo online como uma transposição de uma certa forma de olhar a realidade (jornalístico) para o suporte informático é possível afirmar que a especificidade do meio não altera a especificidade da mensagem? (BALDESSAR, 2003)

Ao responder o problema de pesquisa proposto, busca-se contribuir também com respostas a estas perguntas recorrentes na bibliografia atual.

O tema é interdisciplinar; exige o entendimento e a interligação de conceitos clássicos de comunicação, jornalismo e computação, integrando-os com a gestão do conhecimento e as redes sociais.

## **1.5 Escopo do Trabalho**

Seria impossível prever todos os tipos de aplicações híbridas que podem surgir com as fontes de informação que temos à disposição. É uma área emergente (JOHNSON, 2001) e espontânea, e esses fatores multiplicam suas potencialidades.

Sendo assim, este trabalho apenas busca apontar algumas possibilidades relacionadas com o jornalismo. A intenção é reunir ferramental teórico suficiente para que se entenda melhor o papel do jornalista e como ele pode criar conteúdo jornalístico empregando tal recurso.

O modelo proposto é um exemplo do que se pode fazer, e tem o objetivo de inspirar novas aplicações e novos modelos mais complexos e mais completos do que este. Toda a discussão sobre a interface do usuário, desenho das aplicações e/ou detalhes estéticos foi intencionalmente suprimida por não estar no foco desta dissertação.

## **1.6 Estrutura do trabalho**

A fundamentação teórica é iniciada no capítulo 2 com conceitos da teoria da informação, buscando interligar notícia, relevância, alguns elementos históricos da prática jornalística até chegar na personalização do conteúdo.

Em seguida o jornalismo de precisão é apresentado, mostrando como seus métodos buscam dar ao jornalismo o mesmo rigor do método científico. Sugere-se caminhos para aplicar as noções de jornalismo de precisão hoje, com a internet e dados dinâmicos.

Entra-se também nas ciências da cognição e no estudo do cérebro. Desenvolve-se o conceito de custo cognitivo e analisa-se como ele influencia o projeto de aplicações na internet.

O momento atual da cultura e da economia da informação em rede é discutido. Descreve-se o processo que levou a informação a se tornar um bem valioso e como a cultura da participação e da colaboração vem crescendo no mundo.

Trata-se da web semântica, descrevendo-a desde a sua idealização até o estado da arte das tecnologias relacionada a ela. Processos alternativos aos ideais propostos pela web

semântica também são examinados. Busca-se explicar a polêmica Web 2.0, seu surgimento e as idéias que movem as empresas de internet de maior sucesso atualmente.

O perfil do profissional que produz a Aplicação Web Híbrida é definido. Procura-se no jornalismo e na computação subsídios para descrever como deve ser este profissional.

Um breve histórico do jornalismo online é traçado e identifica-se o modelo jornalístico centrado na matéria, alvo de críticas.

No capítulo 3, que se refere ao ferramental, explica-se a metodologia da pesquisa e detalha-se as tecnologias necessárias para a aplicação do modelo.

No capítulo 4 descreve-se o modelo de Aplicação Web Híbrida.

O capítulo 5 traz as conclusões desta pesquisa e sugestões para trabalhos futuros.

Por fim, são apresentadas as referências bibliográficas e os anexos.

## CAPÍTULO 2 – FUNDAMENTAÇÃO TEÓRICA

### 2.1 Teoria da Informação

Sperber e Wilson analisam como a mente humana lida com a tarefa de processar informação em curto prazo de maneira eficaz:

“Recursos devem ser alocados para o processamento da informação que pode trazer a maior contribuição para os objetivos cognitivos gerais da mente com o menor custo de processamento.

Alguma informação é antiga: já se encontra presente na representação de mundo do indivíduo. A não ser que seja necessária para a execução de uma tarefa cognitiva em particular, e também seja mais fácil de acessar a partir do ambiente do que da memória, tal informação não vale a pena ser processada em absoluto. Outra informação não é apenas nova, mas inteiramente desconectada de tudo na representação de mundo do indivíduo. Ela somente pode ser adicionada a esta representação como pedaços isolados, e isso geralmente representa muito custo de processamento para pouco benefício. Outra informação é nova, porém conectada com informação antiga. Quando estes itens de informação novos e antigos são usados em conjunto como premissas em um processo de inferência, mais novas informações podem ser derivadas: informação que não poderia ter sido inferida sem a combinação de premissas novas e antigas. Quando o processamento de nova informação faz emergir tal efeito multiplicador, chamamos-na de *relevante*. Quanto maior o efeito multiplicador, maior a relevância.” (t.A.) (SPERBER & WILSON, 1995, p.48)<sup>2</sup>

Na comunicação humana, o receptor infere a mensagem a partir das informações que recebe. E sendo assim, qualquer fragmento de informação nova, somado às informações já disponíveis, se torna essencial para o entendimento da mensagem.

---

2 “Resources have to be allocated to the processing of information which is likely to bring about the greatest contribution to the mind's general cognitive goals at the smallest processing cost. Some information is old: it is already present in the individual's representation of the world. Unless it is needed for the performance of a particular cognitive task, and is easier to access from the environment than from memory, such information is not worth processing at all. Other information is not only new but entirely unconnected with anything in the individual's representation of the world. It can only be added to this representation as isolated bits and pieces, and this usually means too much processing cost for too little benefit. Still other information is new, but connected with old information. When this interconnected new and old items of information are used together as premises in an inference process, further new information can be derived: information which could not have been inferred without this combination of old and new premises. When the processing of new information gives rise to such a multiplication effect, we call it relevant. The greater the multiplication effect, the greater the relevance.”

Como a citação acima descreve, uma informação já armazenada na memória, um fato já conhecido, tem baixo custo cognitivo para o observador, ou seja, não exige uma nova “entrada” na memória.

Informações que o cérebro julga relevantes são aquelas que podem ser compreendidas com base em experiências e memórias anteriores. Caso não haja nada parecido com a experiência atual, não há parâmetro para inferir o que o novo dado significa, não há informações preexistentes que, somadas às novas, resultem em aprendizado, em conhecimento, em novos registros na memória.

O conceito de relevância de Sperber e Wilson é essencial em qualquer teoria do jornalismo: é desta maneira que se decide o que é notícia. Editores e repórteres são treinados para raciocinar que o conteúdo jornalístico relevante é aquele que, somado aos conhecimentos do leitor, quebra a normalidade e, por causa disso, interessa às pessoas.

A rotina e a normalidade não têm valor de notícia, salvo quando o evento dá conta de um serviço de grande utilidade pública, como as campanhas de vacinação e o anúncio de datas limite para o pagamento de impostos. Em regra, “se um cachorro morde um homem, não é notícia. Mas se um homem morde um cachorro, vai para a primeira página,” como observou no século XIX o jornalista americano Charles Anderson Dana, editor do *New York Tribune*.

Mas como se faz para decidir o que é informação relevante para todos? O jornalismo considera alguns princípios básicos, comprovados pela experiência de mercado, peculiaridade do público e a relevância dos acontecimentos.

Desde o início do século XX essas regras do jornalismo têm permanecido praticamente inalteradas, e uma das razões para isso foi, até recentemente, a limitação tecnológica. A tradição se estabeleceu e, hoje em dia, mesmo com a internet, muitos ainda conceituam a produção jornalística como antigamente, quando só era possível imprimir uma edição do jornal e distribuí-la uma ou duas vezes por dia.

A racionalização dos custos de impressão e transporte criou todo um fluxo de trabalho baseado no “fechamento”, na “impressão” e segmentado em “edições”.

Esse mínimo denominador comum de informação que, possivelmente, interessa a todos os leitores, ouvintes, espectadores e usuários – ou a parte significativa deles – é intuído por quem a faz. É resultado da experiência dos editores, de tentativa e erro e de pesquisas de mercado. E, evidentemente, não se consegue atender às necessidades ou aspirações de todos.

Se cada pessoa tem sua própria história de vida, sua própria trajetória intelectual, pessoal e social, pode-se dizer que cada uma tem em sua memória um acervo único. Cada um interpreta o mundo e entende as coisas à sua maneira. Por este raciocínio, a informação-

gatilho, que somada às já memorizadas, produz algo novo no cérebro, é diferente em cada pessoa. Embora existam dados comuns que interessem a todos, o verdadeiramente relevante é algo individual, e não “de massa”: distribui-se em diferentes conjuntos humanos, concentrados ou dispersos, e estes em subconjuntos. A evolução da imprensa para nichos especializados é um fato que corrobora este ponto de vista. Busca-se atingir grupos de pessoas com experiências de vida comuns, sejam aficionados por atividades físicas, motocicletas, música clássica, eletrônica ou tatuagens e *piercings*.

O fenômeno da segmentação não é novo, mas é uma tendência que se acentuou com a internet e caminha para o seu extremo: a individualização da informação. A tecnologia permite identificar e direcionar conteúdo diferente para cada pessoa conectada à rede. Um exemplo é da loja Amazon.com, pioneira em sistemas de recomendação de produtos; a partir de sugestões dos compradores que refletem hábitos ou preferências, as ofertas se organizam para aquele cliente em particular, de forma a atender a seus interesses específicos.

Chris Anderson estudou este assunto e chegou a um gráfico que representa a enorme quantidade de micro-interesses compartilhados por poucas pessoas. Anderson argumenta que - se a loja ou o canal de distribuição forem grandes o suficiente - produtos com baixa demanda ou baixo volume de vendas podem coletivamente alcançar uma fatia de mercado que se equipare ou até exceda os relativamente poucos itens mais vendidos e mais conhecidos. O gráfico toma a forma de uma “cauda longa” (*Long Tail*), termo que deu o título ao artigo que escreveu para o número de outubro de 2004 na revista *Wired* e ao livro que publicou dois anos depois, com título sugestivo: *A Cauda Longa: porque o futuro dos negócios está em vender mais de menos* (*The Long Tail: why the Future of Business is Selling Less of More*).

## 2.2 Jornalismo de Precisão

Há mais de 15 anos, Philip Meyer publicou o livro *O Novo Jornalismo de Precisão – The New Precision Journalism* (MEYER, 1991), no qual incentiva os jornalistas a usar bases de dados informatizadas para cruzar informações, e com isso chegar a conclusões jornalisticamente relevantes das quais possam partir para a produção de reportagens.

Segundo Meyer, Jornalismo de Precisão é a aplicação de métodos científicos de investigação social e comportamental à prática do jornalismo. Lara de Lima (LIMA, 2000) cita observações do jornalista José Luis Dader, tradutor de *The New Precision Journalism* para o espanhol. Ele explica que os métodos referidos por Meyer são a sondagem ou pesquisa de

opinião, o experimento psicossocial e a análise de conteúdo. No ponto de vista do jornalista espanhol, também autor de trabalhos sobre o assunto, o Jornalismo de Precisão excede o campo da sociologia. Luis Dader resume que “é o controle e a indagação sobre o método, em definitivo, o que permite falar de Jornalismo de Precisão” (LIMA, 2000).

### 2.2.3 História e Conceito

A partir da década de 80, jornais americanos passaram a basear suas notícias e reportagens em pesquisas próprias, em parte por não confiar nas pesquisas solicitadas por políticos. Baixas tiragens também levaram os editores desses jornais a procurar o aprimoramento do produto jornalístico por meio de cobertura mais científica. O começo dessa busca coincidiu com o acesso, pelos jornais, a computadores e bases de dados, nos anos 70. (MEYER, 1991)

Meyer publicou o primeiro livro sobre o assunto, *Precision Journalism. A Reporter's Introduction to Social Science Methods*, em 1973. Em 1989, os jornais *The Washington Post*, *USA Today*, *Los Angeles Times* e *The New York Times* inauguraram suas seções de bases de dados, conforme relata José Luis Dader (MEYER, 1993 apud LIMA, 2000).

Aplicando ferramentas do Jornalismo de Precisão, alguns jornais tiveram suas reportagens premiadas com o Pulitzer, como o *Dallas Morning News* em 1985, por reportagem em que denunciou a segregação racial em moradias públicas do Texas e, três anos depois, o *Atlanta Constitution*, que provou haver discriminação entre raças nos empréstimos hipotecários feitos pelo governo federal norte-americano. (MEYER, 1993 apud LIMA, 2000)

Antes de influenciar a imprensa americana, as idéias de Meyer provocaram mudanças no meio acadêmico e nos jornais de médio porte onde foram testadas. Em um dos testes, provou-se que, ao contrário do que se imaginava, atos de vandalismo na cidade de Dallas não partiam predominantemente de pessoas com baixo nível de instrução e de negros oriundos do Sul. Com o cruzamento de dados, o *Detroit Free Press* “descobriu que as pessoas com nível superior haviam participado dos distúrbios em percentagens similares às que não tinham chegado a completar o segundo grau” (MEYER, 1993 apud LIMA, 2000). A experiência havia mostrado que os métodos de investigação social são aplicáveis ao jornalismo.

Tal como foi originalmente concebido, durante os movimentos de protesto social dos anos 60, o Jornalismo de Precisão era uma via de ampliação do equipamento instrumental para que o repórter convertesse em material de indagação minuciosa os

assuntos até então inacessíveis ou somente acessíveis de maneira muito vaga. Esta forma jornalística resultou de especial utilidade para escutar a voz dos grupos dissidentes e minoritários que estavam lutando pelo reconhecimento de uma representação (MEYER, 1991).

O termo “Jornalismo de Precisão” foi usado em 1971 por Everette Dennis, para explicar a seus estudantes o “novo jornalismo” proposto por Meyer. Segundo o próprio Meyer, Dennis usou o adjetivo “de precisão” para diferenciar este jornalismo, baseado no método científico, daquele “novo jornalismo” de ênfase literária que tornou famosos jornalistas como Tom Wolfe nos anos 60 (MEYER, 1993 apud LIMA, 2000).

Bem aceito nas universidades, entre os profissionais, contudo, o Jornalismo de Precisão encontrou resistência, principalmente por uma compreensão estreita dos ideais de objetividade dos jornalistas. Para os que defendem a objetividade no jornalismo, não cabe aos repórteres e editores assumir posição diante dos fatos, mas apenas apresentar opiniões sobre os temas contraditórios. Partindo dessa idéia, jornalistas concluíram que os meios de comunicação não devem fazer pesquisas de opinião, e sim publicar as sondagens feitas por outros órgãos (MEYER, 1993 apud LIMA, 2000).

Meyer contra-argumenta: “o modelo da objetividade foi desenhado para um mundo muito mais simples, onde os fatos desnudos poderiam falar por si mesmos” (MEYER, 1993 apud LIMA, 2000). Tanto esse modelo é inadequado que, já nos anos 60, “a frustração com o inalcançável ideal da objetividade” levou parte dos jornalistas a aderirem ao “novo jornalismo”, aquele de caráter literário. Na opinião do autor, apesar de os esforços nesse sentido serem válidos, a literatura não oferece a disciplina que o jornalismo requer:

Uma solução melhor consiste em aproximar o jornalismo do método científico, incorporando os poderosos instrumentos de que a ciência dispõe, tanto para a coleta como para a análise de dados, assim como sua busca sistematizada de uma verdade verificável (MEYER, 1991).

Após ter ministrado 16 cursos acadêmicos sobre o tema, Meyer publicou, em 1991, seu segundo livro sobre Jornalismo de Precisão, *The New Precision Journalism*, que traz exemplos práticos da aplicação de metodologias das ciências sociais, como a estatística, no jornalismo. Meyer contou com o apoio do jornal USA Today e com os serviços informativos da CBS para testar a teoria em experiências jornalísticas (LIMA, 2000).

## 2.2.4 RAC

No livro de 1991, o autor trata também da Reportagem Assistida por Computador (*Computer Assisted Reporting*), variante do Jornalismo de Precisão que prevê a realização de reportagens a partir de informações de bases de dados. Esta modalidade, também chamada de Reportagem com o Auxílio do Computador, exige que os jornalistas saibam de que forma acessar e como interpretar informações dessa procedência. Segundo José Luis Dader (apud LIMA, 2000), “a Reportagem Assistida por Computador (RAC) (...) é, sem dúvida, o que mais espetacularmente está crescendo, dentro da ampla gama de atuações de precisão”.

Os principais argumentos dos que desaconselham a adoção do Jornalismo de Precisão pelos países iberoamericanos são as legislações que dificultam o acesso às informações e a inexistência de bancos informatizados de informação pública (DADER apud LIMA, 2000 , p. 16). “Essa é grande cartada para dizer que se passarão décadas antes que possamos imitar trabalhos como os citados no livro apresentado”, antecipa Luis Dader. (DADER apud LIMA, 2000)

Ele esclarece que:

A primeira e fundamental ferramenta do Jornalismo de Precisão é a imaginação e a segunda, a aprendizagem de certas regras – tampouco demasiadas – da metodologia científica. Só com ambas pode-se abordar uma infinidade de projetos de pressuposto insuficiente, reduzido volume de dados e acesso aberto a qualquer curioso. O Jornalismo de Precisão não é só para empresas jornalísticas ricas e sociedades ultratecnologizadas, mas também para qualquer jornalista anticonvencional e anti-rotineiro com um mínimo de treinamento nos rigores da análise sistemática de dados objetivados (MEYER, 1993 apud LIMA, 2000).

Embora na prática o jornalista de precisão ainda esbarre na burocracia e na falta de transparência das fontes oficiais de dados, a posição oficial do governo brasileiro dá indícios de mudanças neste panorama exposto por Dader. No capítulo sobre Governo Eletrônico este assunto é discutido em mais detalhes, com base em documentos oficiais do governo.

Trazendo a discussão para os dias de hoje, pode-se constatar que o papel de “jornalista anticonvencional e anti-rotineiro” citado por Dader pode ser assumido por qualquer usuário da internet que tenha disposição e conhecimento para executar este tipo de reportagem. A distribuição da informação a um custo virtualmente nulo proporcionada pela internet abre esse espaço.

Para Lima, a realidade que Luis Dader observa nos países europeus é comparável à brasileira. Ele identifica “evidências isoladas de trabalhos de precisão em diferentes meios europeus, mas, salvo o que pudesse contribuir alguma investigação hipoteticamente em curso, o panorama europeu ainda está muito distante de oferecer um movimento de percepção coletiva e atuação generalizada nesta linha” (DADER apud LIMA, 2000, p. 13). Na Espanha, jornalistas publicam reportagens de relativa precisão, mesmo sem ter consciência da sua classificação como Jornalismo de Precisão (DADER apud LIMA, 2000, p. 14).

Lima afirma, em seu texto do ano 2000, que no Brasil esse direcionamento do jornalismo ainda é pouco conhecido. Ela acredita que muitos jornalistas se aproximam dos ideais do Jornalismo de Precisão, na medida em que se empenham em apurar informações com rigor, com o objetivo de melhorar o produto final. Alguns deles certamente usaram métodos científicos na elaboração de suas matérias antes do surgimento de qualquer teoria a respeito.

O jornalismo de precisão de Meyer, descrito numa época em que o mundo não estava tão conectado quanto hoje, pode ser trazido para um novo contexto, levando adiante a integração à atividade jornalística elementos originários das ciências exatas. Uma das propostas desta pesquisa é usar os agentes inteligentes da web semântica e bancos de dados de acesso público – disponíveis instantaneamente, no momento da atualização – para ampliar o poder do jornalista/desenvolvedor independente.

## **2.3 Cognição e Entendimento**

### **2.3.1 O cérebro**

Atualmente, com o avanço das técnicas de prospecção, é possível ter uma idéia de como o cérebro funciona melhor do que se tinha há poucas décadas. É possível entender como a mente humana percebe as coisas, processa informações e, com base neste processamento, determina a realização de ações complexas. Quando se projeta interfaces que seres humanos vão utilizar, é necessário imaginar como os outros vão entender o que está na tela. Mas a imaginação não é suficiente: pressupostos científicos devem embasar as decisões de arquitetura da informação e desenho de interface.

A Teoria Computacional da mente é um ponto de partida que leva a reflexões interessantes. Pinker explica que os seres humanos muitas vezes induzem sensações primitivas em ambientes controlados como forma de diversão. É o que faz uma montanha-

rusa, por exemplo. Desperta o medo e a vertigem, faz o organismo descarregar adrenalina no sangue, faz a pessoa se sentir viva:

A mente é um sistema de órgãos de computação, projetados pela seleção natural para resolver os tipos de problemas que nossos ancestrais enfrentavam em sua vida de coletores de alimentos, em especial entender e superar em estratégia os objetos, animais, plantas e outras pessoas. Essa síntese pode ser desdobrada em várias afirmações. A mente é o que o cérebro faz; especificamente: o cérebro processa informações, e pensar é um tipo de computação. A mente é organizada em módulos ou órgãos mentais, cada qual com um *design* especializado que faz desse módulo um perito em uma área de interação com o mundo. A lógica básica dos módulos é especificada por nosso programa genético. O funcionamento dos módulos foi moldado pela seleção natural para resolver os problemas da vida de caça e extrativismo vivida por nossos ancestrais durante a maior parte de nossa história evolutiva. Os vários problemas de nossos ancestrais eram subtarefas de um grande problema para seus genes: maximizar o número de cópias que chegariam com êxito à geração seguinte.(PINKER, 1998: p.32)

O Autor sugere que as partes do cérebro responsáveis pela imaginação têm similaridades com as áreas visuais do cérebro; define-os como órgãos vestigiais de visão. Por isso, o ser humano tem facilidade de imaginar visualmente, de interpretar símbolos em uma folha de papel e de convertê-los para uma imagem mental vívida.

Outra característica importante do pensamento é o uso de estereótipos, ou protótipos:

Nosso pensamento faz uso de um princípio engenhoso, que pode ser denominado princípio do protótipo: O evento mais específico pode servir de exemplo geral de uma classe de eventos. Todos sabem que eventos específicos possuem um fulgor que os imprime tão fortemente na memória; que eles podem ser, posteriormente, usados como modelos para outros eventos a eles semelhantes, de alguma forma. Assim, em cada evento específico, existe o embrião de toda uma classe de eventos semelhantes. Essa idéia de que há generalidade no específico é de importância transcendental. (HOFSTADTER, 2001, p.383)

As pessoas buscam padrões familiares em todas as coisas que vêem. Sabem o que é um chifre e o que é um cavalo, e por associação conseguem montar um unicórnio genérico no olho da mente. O unicórnio de cada um é diferente, mas ao mesmo tempo, suas características básicas são iguais.

A busca de padrões e estereótipos e sua combinação é o que possibilita a leitura, por exemplo. Quanto mais se pratica, mais rápida é a associação. De acordo com Pinker, quanto mais as trilhas neuronais são ativadas por determinado símbolo ou padrão, seja visual, auditivo, olfativo, motor ou de comportamento, mais fortes essas trilhas se tornam, e mais veloz é a sua recuperação da memória.

Pode-se concluir então que o aprendizado e a automatização de reflexos seriam como o asfaltamento do que um dia foram apenas trilhas em mata fechada.

### 2.3.2 O custo cognitivo

Levando em conta o princípio do protótipo e as considerações de Pinker, pode-se dizer que o ser humano primeiro raciocina em termos de estereótipos e, em seguida, individualiza o evento ou a entidade com base em detalhes e nas relações com o contexto, o que pode até fazê-lo rever o modelo esperado para aquela situação.

Walter Lippman (1922) introduziu o termo “estereótipo” no campo das ciências sociais, dizendo que eles são figuras simplificadas das pessoas e eventos do mundo dentro das nossas cabeças:

“Na grande confusão do mundo exterior, selecionamos o que a nossa cultura já definiu para nós, e tendemos a perceber o que selecionamos na forma estereotipada para nós por nossa cultura.” (t.A.) (LIPPMAN, 1922, p. 81)<sup>3</sup>

Um sinal de trânsito vermelho significa “pare”. Esta noção já está bem assentada na mente de quem dirige automóveis. Mesmo com variações na localização e formato dos semáforos, os motoristas – pelo menos os que obedecem às leis de trânsito – continuam parando no sinal vermelho. É uma tarefa cognitivamente simples, comportamento tão automatizado que não exige muito do cérebro.

Outras reações instantâneas, mas muito mais complexas, também acontecem rotineiramente. Gladwell (GLADWELL, 2005) relata a história de um experiente especialista em artes plásticas que, num piscar de olhos, teve a sensação de que uma estátua dada como antiga era falsa. Nenhum teste de laboratório conseguiu comprovar o palpite dele, que se rendeu aos fatos e deu o seu aval ao objeto como obra verdadeira. Algum tempo depois, no entanto, ao se analisar o objeto com equipamentos mais modernos descobriu-se que a estátua era mesmo falsa, e que a intuição do especialista estava correta desde o início.

<sup>3</sup> *“In the great blooming, buzzing confusion of the outer world we pick out what our culture has already defined for us, and we tend to perceive that which we have picked out in the form stereotyped for us by our culture”*

O responsável pela intuição do especialista seria, segundo Gladwell, seu inconsciente adaptativo. Isso significa que a mente processa instantaneamente os estímulos que recebe do ambiente, e faz a pessoa reagir de determinada maneira. Quando alguém vê um caminhão se aproximando em alta velocidade e a pouca distância, não dispõe de tempo para analisar todas as possibilidades e decidir qual o melhor curso de ação. O cérebro analisa a situação e emite para o corpo uma mensagem clara: “fuja!”.

Pela tese de Gladwell, muitas vezes as decisões tomadas instantaneamente são mais acertadas do que as tomadas depois de demoradas análises. Ele ressalta, no entanto, que é muito fácil se enganar, especialmente quando não se tem muita familiaridade com os estímulos que levam à decisão. Se qualquer outra pessoa, e não o especialista que convive todos os dias com objetos de arte do mesmo período e procedência, fosse analisar a estátua falsa, daria apenas um palpite sem validade.

Os estereótipos e preconceitos têm papel fundamental no julgamento instantâneo do que se vê. É freqüente referir ao “sujeito mal-encarado” ou à “dondoca” quando não se sabem detalhes sobre a pessoa que possam contradizer os estereótipos.

Quando o objetivo é fazer as pessoas perceberem determinada informação em tempo reduzido, os estereótipos e preconceitos podem ser usados como ferramentas. Um romancista descreve suas personagens-tipo (*flat*) relacionando características pessoais estereotipadas que os leitores têm armazenadas em suas mentes; não o faria se se tratasse de personagens complexos (*round*).

É clássico o exemplo da dificuldade de entendimento que se tem quando alguém se depara com um texto de alguma área do conhecimento muito oposta à sua. Não há na memória um “vocabulário” estereotipado sobre este novo universo de palavras e noções que permita deduzir significados, e por isso o texto será considerado “difícil”.

Um texto trivial, por outro lado, pode ser definido como algo que “não faz pensar” – mensagem tão facilmente entendida que seu custo cognitivo é imperceptível.

Também na percepção visual de informação há um custo cognitivo que pode estabelecer a diferença entre o entendimento ou não de uma mensagem.

Tufte (2001), em sua Teoria dos Gráficos de Dados (TUFTE, 2004), refere-se a “métodos visuais de representação de informação quantitativa”, tais como os tradicionais infográficos de jornais e revistas; estabelece uma fórmula para calcular a razão dados/tinta (ou, atualizando para o contexto da internet, a razão dados/pixel) de cada gráfico. Assegura que todos os elementos gráficos que não transmitem informação podem ser eliminados sem

prejuízo do entendimento. Chama de *chartjunk*, ou lixo gráfico, tudo que atrapalha ou não contribui para a visualização despojada da informação principal.

Para calcular a razão dados/tinta, Tufte analisa a quantidade de tinta que transmite informação e a divide pelo total de tinta do infográfico, chegando assim a uma “proporção de tinta devotada à representação não-redundante de informação”.

$$\text{Razão dados-tinta} = \frac{\text{tinta que transmite informação}}{\text{total de tinta utilizado para exibir o gráfico}}$$

$$\text{Razão dados-tinta} = \text{proporção de tinta devotada à representação não-redundante de informação}$$

(TUFTE, 2004)

Os princípios de Tufte para os infográficos também servem para o desenho de interfaces:

Acima de tudo, mostre os dados

Maximize a razão dados-tinta (ou dados-*pixel*)

Apague a tinta (*pixels*) que não representa dados

Apague os dados-tinta redundantes

Revise e altere

(TUFTE, 2004)

Para Tufte, enfeites ou flamejamentos aumentam desnecessariamente o custo cognitivo de um infográfico. A razão dados-tinta dele é análoga à razão sinal-ruído das transmissões telefônicas. O ruído visual atrapalha o entendimento do que é importante e faz a mente processar mais do que o necessário.

A edição jornalística segue esta linha, com a finalidade de concentrar a atenção no texto, fotos ou infográficos (estes preferentemente mais discretos), sobretudo nos veículos que valorizam o conteúdo, como os que pretendem atingir elites culturais ou econômicas. O que Tufte fez foi expressar estas diretrizes no universo visual.



*Figura 2.1: Página principal do Google: baixo custo cognitivo*  
 (Fonte: <http://www.google.com.br>, acesso em: 01/05/2007)

Um exemplo de baixo custo cognitivo é o *site* de buscas Google (Fig. 2.2). Uma tela praticamente vazia com um logotipo, um espaço para escrever e um botão “Pesquisa Google”. O número de *pixels* que não dizem respeito à função principal do *site* é mínimo em relação a outros concorrentes. É uma página que tem o propósito singelo de identificar e não de adicionar informações, fazendo o usuário “pensar”. Tem custo cognitivo muito baixo, ao contrário de alguns portais, repletos de anúncios, cores e movimentos, informação desnecessária ou redundante que se pode chamar de poluição visual (Fig. 2.3).

ASSINE BATE-PAPO FOLHA ONLINE SAC Messenger Voip E-Mail grátis Shopping ÍNDICE PRINCIPAL

**UOL**  
O MELHOR CONTEÚDO

O Governo Federal investe para o Brasil crescer mais. **PAC** PROGRAMA DE ACCELERÇÃO DO CRESCIMENTO

BATE-PAPO UOL BUSCA Web Shopping Imagens Vídeos Notícias Dicionários E-MAIL @uol.com.br SENHA

Pen drive 1GB A partir de R\$ 60 Compare preços

Segunda-feira, 21 de maio de 2007

10:43 Itau

EMPREGOS? CATMO extra.com.br

Shopping UOL

TV é no Crocodilo FRETAR GRÁTIS! Todos linha TVs frete grátis Brasil em 12. Só hoje!

Submarino 15.4" 1.86GHz 1GB HD 80GB por R\$ 2.999!

MagazineLuiza.com Monitor LCD 17" multimídia por R\$ 599 em 12!

Extra.com.br LCD 17" Monitor LG com DVI: R\$ 579 ou 12 de R\$ 56,58!

Computador C521 Monitor CRT 17" e Windows Vista por R\$ 1.499!

Sack's Perfumaria Novo antidade Nivea DNAge por R\$ 53,90!

MagazineLuiza.com Epcosm 80GB + gravador DVD: R\$ 1.039 em 12

Novo Nintendo Wii Ache a partir de 12 vezes de R\$ 199!

PontoFrio.com De desconto em um show de ofertas! Só aqui!

Pernambucanas Monitor 17" LCD LG em 12 de R\$ 58,25!

MagazineLuiza.com PC Itautec 512MB monitor 17": R\$ 1.199 em 12 vezes

Fast Shop TV 29" Super Slim plana game LG: R\$ 1.099 em 12!

Computadores! Compare e ache os menores preços!

Computadores Dell A partir de R\$ 799 ou 6 sem juros no cartão!

Extra.com.br Telefone VoIP Santel Z-18: só R\$ 99,90!

Submarino Celular a partir de R\$ 149 em 12 vezes e frete grátis!

Nike Shox Impactável! Ache modelos partir de apenas R\$ 399!

190 mil anúncios de 7 dias grátis! Empregos. Anuncie seu currículo aqui!

Mais lojas

Shopping UOL

Tênis Mizuno Ache modelos em até 12x. É imperdível!

Celular Motorola K1 Com Câmera, Bluetooth e USB a partir de R\$ 899,99.

Sony 5650 Com 7.2MP. Ache aqui pague em até 10x. Confira!

Box 24 horas 3ª temporada. Encontre aqui a partir de R\$ 72,90.

MP3 players de R\$ 89 Compare aqui os preços e escolha o seu! É imperdível!

Gravador de DVD Encontre aqui a partir de apenas R\$ 95. Aproveite!

Fogão Brastemp Único Inox. Ache aqui e pague em até 12x. Confira!

Monitor LG 17" Encontre aqui a partir de apenas R\$ 299. Compre!

Mini System Philips Encontre aqui a partir de apenas 299. Aproveite!

Scanners Encontre aqui. Compare e ache o melhor preço.

Flower by Kenzo Encontre aqui este perfume a partir de R\$ 125. Aproveite!

Arranjos de flores Ache diversos modelos a partir de R\$ 20. Aproveite!

Violões: Encontre aqui. Compare e ache o melhor preço.

Computadores Com Windows Vista. Ache aqui. Compare e compre!

George Foreman Grill Super Jumbo. Ache aqui e pague em até 12x. Confira!

Patins Ache diversos modelos a partir de R\$ 81,50. Compre!

DVD Player Invoxox a partir de R\$ 129.

Bolsa para Câmera a partir de R\$ 15,90.

Boné Nike Ache a partir de R\$ 57.

Maleta para Notebook. Pelo melhor preço.

Playstation Portátil. Pague em até 12x.

Real Madrid Camisa a partir de R\$ 149.

Album de Fotos

Antivírus e Firewall

Aplicativos e Produtos

Banda Larga

Bate-papo

Biblioteca

Bichos

Blog Fotoblog

Busca

Carros

Cartões

Celebridades

Ciência e Saúde

Cinema

Crianças

Diversão e Arte

Economia

Educação

E-Mail

Empregos

Esporte

Estilo

Folha Online

Gay

Horóscopo

Humor

Índice

Jogos Loja de jogos

Jornais

Lição de Casa

Música Megastore

Pan-Americano

Rádio UOL

Revistas

SAC

Serviços Classificados

Sexo

Shopping

Sites Pessoais

Tecnologia

Televisão

Tempo Trânsito Mapas

TV UOL e vídeos

Últimas Notícias

UOL K

UOL News

Vestibular

Viagem

Voip

Wi-Fi

ASSINANTE UOL

Discador UOL

AntiSpam

Troque sua senha

SERVIÇOS GRATUITOS

Barra UOL

E-Mail grátis

LINKS PATROCINADOS

Só pague quando seu anúncio for clicado

Inquérito da PF cita elo entre governador e máfia

Ministério: imagens mostram entrega de verba

Rondeau: governo já discute hipótese de saída

Grupo flagra pagamento a deputado federal

Política externa

Jornal paraguaio compara Lula a colonizadores dos séculos 16 e 17

Internet

Projeto quer 'provedor dedo-duro' e permite justiceiros virtuais na Web

Bombardeio libanês a campo de refugiados mata oito

Blog do Josias Polícia Federal mantém ameaça de greve para esta terça-feira

Violência Adolescente é assassinada com tiro na cabeça no litoral de SP

Tempo Santa Catarina e Rio Grande do Sul devem ter neve na quarta-feira

Música

RESFEST | AGENDA DE SHOWS | FOTOS

The Waiters

Banda de reggae toca em São Paulo; veja as fotos

Hatori mistura eletrônico, pop e forró

Dudu Nobre recebe alta médica no Rio

Show de Tim Maia de 1989 em DVD

Mais! Lobão comenta o CD "Roadwork"

Horóscopo

Gêmeos

O Sol já está quase lá; conheça o perfil deste signo

Como melhorar a vida

Decisões firmes ajudam

Combinação dos signos

Engenheiros

Sandy e Júnior

Dupla marca show de despedida para próximo mês

Stallone é condenado

Tom Cavalcante em férias

Britney improvisa em show

Celebridades

Cannes

FOTOS

Crianças

CARTÕES | LIÇÃO DE CASA

Estrelas

Andie MacDowell, Sharon Stone, Alain Delon e mais

Espectáculo

Público infantil aprova show do 'High School Musical'

Salles explica a relevância dos festivais

Conheça as histórias de Ruth Rocha

Tente descobrir qual é a palavra correta

Aprenda a fazer um envelope de papel

Busca: imagens da atriz Juliette Binoche

Plaidas e mais plaidas sobre cachorros

Zapping 'Pânico' pede R\$ 1 milhão para trocar Rede TV! pela Record

Virada Cultural Evento deverá alcançar mais três cidades em 2008

Jornais e Revistas

AGORA PARANÁ | MANEQUIM | TRÓPICO | VOCÊ S/A

Folha de S. Paulo

Cox Newspapers

Mídia Global

Tamanho do cérebro prevê inteligência de primatas

Cem motivos para cultivar a memória de John Wayne

Em Cuba, igrejas lotam e atraem novos convertidos

National Geographic Favela de Dharavi, na Índia, um lar ou uma aberração?

Produtos UOL

As melhores soluções para você

UOL Voip Tarifas reduzidas a qualquer hora

Firewall A barreira de proteção para sua conexão

Ragnarok Muito mais diversão e aventura!

Apenas R\$ 0,14/minuto nas ligações para todo o Brasil. Compre créditos agora!

Proteja os seus arquivos pessoais de invasores. A partir de R\$ 7,90 mensais

Entre para Ragnarok e participe de grandes guerras. A partir de R\$ 14,30 mensais

Shopping UOL

Compare e ache o menor preço

Extra.com.br Aparelho jantar e chá Schmidt 20 peças: R\$ 199

Dell.com.br Notebooks a partir de R\$ 1.799! Aproveite!

Pernambucanas Motorola V3 Black em 3 vezes de R\$ 193

COMPARE PREÇOS DE digite produto ou marca

ASSINE BATE-PAPO FOLHA ONLINE SAC Messenger Voip E-Mail grátis Shopping ÍNDICE PRINCIPAL

About UOL • Anuncie no UOL • Assine UOL • Relações com Investidores • Segurança e privacidade • Sobre o UOL • Trabalhe no UOL • UOL Argentina Banda Larga • Pesquisa de preços • Antivírus • Música • Voip • Jogos • Wi-Fi • Links patrocinados • E-mail gratuito

RSS O que é isso? © 1996-2007 UOL - O melhor conteúdo. Todos os direitos reservados Faça do UOL a sua casa

Figura 2.2: Página principal do UOL: alto custo cognitivo  
(Fonte: <http://www.uol.com.br>, acesso em: 21/05/2007)

## 2.4 Economia da Informação em Rede

Em seu livro “*The Wealth of Networks*” (A Riqueza das Redes), Yochai Benkler (2006) afirma que o mundo assiste à emergência de um novo estágio na economia da informação, algo que ele chama de “economia da informação em rede”. Ela vem tomar o lugar da economia industrial da informação que moldou a produção informativa a partir da metade do século XIX e por todo o século XX.

O que caracteriza a economia da informação em rede é que a ação descentralizada do indivíduo, – especificamente em ações cooperativas e coordenadas levadas adiante por mecanismos radicalmente distribuídos e fora-de-mercado que não dependem de estratégias proprietárias – tem um papel muito maior do que teve, ou do que poderia ter tido, na economia industrial da informação.

O catalisador para essa mudança, de acordo com Benkler, é o advento da tecnologia de fabricação de computadores e seu efeito-cascata nas tecnologias das comunicações e armazenamento de dados. O declínio do preço da computação, comunicação e armazenamento, na prática, colocou os meios materiais de produção cultural e de informação nas mãos de uma parcela significativa da população mundial – em torno de um bilhão de pessoas pelo mundo.

Benkler afirma que uma característica definidora das comunicações, da informação e da produção cultural desde os meados do século XIX é a necessidade de investimentos cada vez maiores de capital físico para a ampliação do alcance. Só assim se conseguiria uma comunicação eficaz por todas as sociedades e territórios em expansão que compunham as unidades políticas e econômicas relevantes da época.

Impressoras mecânicas para jornais de grande circulação, o sistema de telégrafos, os transmissores de rádio e, mais tarde, de televisão, cabo e satélite, além do computador “*mainframe*”<sup>4</sup>, se tornaram necessários para elaborar informação e transmiti-la em escalas mais abrangentes do que a própria vizinhança. “Querer se comunicar com os outros não era condição suficiente para se conseguir fazê-lo”. (BENKLER, 2006)

Como resultado, a produção informativa e cultural foi feita, neste período, seguindo um modelo mais industrial do que a economia da informação em si teria necessitado, argumenta ele. O crescimento do ambiente das comunicações em rede mediadas por computador mudou este fato.

---

4 Um mainframe é um computador de grande porte, dedicado normalmente ao processamento de um volume grande de informações. Os mainframes são capazes de oferecer serviços de processamento a milhares de usuários através de milhares de terminais conectados diretamente ou através de uma rede.

O autor constata que os requisitos materiais para uma produção e comunicação eficaz da informação agora estão em posse de um número de indivíduos que é muitas ordens de magnitude maior do que o número dos donos dos principais meios de produção e troca de informação há apenas duas décadas. A remoção das limitações físicas para a produção eficaz de informação fez da criatividade humana e da economia da informação em si os fatos-chave na estruturação da nova economia da informação em rede.

Eles têm características bem diferentes do carvão, do aço e do trabalho manual, que caracterizaram a economia industrial e estruturaram o nosso pensamento básico sobre a produção econômica no século passado. (BENKLER, 2006).

Ele enumera três características deste sistema emergente de produção de informação, estratégias não-proprietárias, a produção fora-de mercado e os esforços cooperativos de larga escala:

1) **Estratégias não-proprietárias.** Elas têm sido mais importantes na produção da informação do que na produção de aço ou de automóveis, mesmo quando a economia da comunicação pendia para os modelos industriais.

Educação, artes e ciências, debate político e disputas teológicas têm sido influenciadas com muito mais força por motivações e por atores fora-de-mercado do que, por exemplo, a indústria automobilística.

Conforme a barreira material – que fez grande parte do ambiente informativo passar pelo “funil” das estratégias proprietárias baseadas no mercado – é removida, as motivações e formas organizacionais básicas fora-de-mercado e não proprietárias se tornam ainda mais importantes para o sistema de produção de informações.

2) **Produção fora-de-mercado.** Benkler diz que a produção fora-de-mercado tem atingido uma importância muito grande. Ficou fácil para qualquer um alcançar e informar ou “edificar” milhões mundo afora.

Tal alcance era impossível antes para indivíduos com quaisquer motivações, a não ser que eles “afunilhassem seus esforços” por meio de organizações do mercado, filantrópicas ou estatais.

O fato de tal iniciativa estar disponível para qualquer um conectado à rede, de qualquer lugar, levou à emergência de efeitos coordenados, onde o agregado das ações individuais, mesmo quando não conscientemente cooperativas, produz o efeito coordenado de um ambiente informativo novo e diversificado. (BENKLER, 2006; JOHNSON, 2001)

Basta fazer uma busca no Google sobre qualquer tópico de interesse para ver como o 'bem informativo' – que é a resposta à consulta de alguém – é produzido pelos efeitos coordenados das ações descoordenadas de uma ampla gama de indivíduos e organizações que agem com uma ampla gama de motivações – tanto de mercado quanto fora-de-mercado, tanto estatais quanto privadas. (BENKLER, 2006)

3) **Esforços cooperativos de larga escala.** Benkler considera o terceiro item de sua análise o mais radical e difícil para observadores acreditarem. Refere-se ao surgimento de esforços cooperativos eficazes de larga-escala – a produção de informação, conhecimento e cultura pelos pares. Esforços tipificados pela emergência do software livre e de código aberto.

Começou-se a ver a expansão deste modelo não apenas nas plataformas-base de software, mas além delas, por todos os domínios da produção cultural e informativa, incluindo a produção de enciclopédias pelos pares, notícias, opiniões e entretenimento imersivo.

Para Benkler, é fácil não perceber essas mudanças. Elas vão contra a corrente das nossas intuições mais básicas sobre economia. Intuições fundadas na economia industrial de uma época na qual a única alternativa séria era o comunismo de Estado – uma opção que hoje é quase universalmente considerada inviável.

O inegável sucesso econômico do software livre fez alguns economistas de vanguarda tentar entender como milhares de desenvolvedores de software livre em rede conseguem competir com a Microsoft em seu próprio jogo e produzir um sistema operacional – o GNU/Linux.

A crescente literatura sobre o assunto se concentra no software e nas particularidades das comunidades de desenvolvimento livres e de código aberto. No entanto, Benkler destaca o trabalho de Eric von Hippel, e relata que sua noção de “inovação conduzida pelo usuário” (*user-driven innovation*) começou a expandir o foco da área e a fazer pensar sobre como a necessidade e a criatividade individuais direcionam a inovação e como acontece sua difusão por redes de indivíduos com idéias parecidas.

Benkler chega à parte final de sua argumentação dizendo que o software livre é apenas um exemplo saliente de um fenômeno muito mais amplo. E questiona:

Por que 50 mil voluntários, co-autores da Wikipedia, a alternativa mais séria à Enciclopédia Britannica, podem virar as costas e entregar tudo de graça?

Por que 4,5 milhões de voluntários contribuem com os ciclos de processamento de seus computadores inativos para criar o supercomputador mais poderoso da Terra, o SETI@Home? (BENKLER, 2006)

Hoje a iniciativa Folding@Home<sup>5</sup>, na qual usuários doam tempo de processamento de seus computadores para a análise de proteínas, acumula cerca de 200 mil CPUs ativas.

Em resposta às perguntas, o autor pondera que sem um modelo analítico amplamente aceito para explicar estes fenômenos, tendemos a tratá-los como curiosidades, talvez modas passageiras, possivelmente significativas para um segmento de mercado ou outro. E conclui:

Devemos tentar, ao invés, vê-los pelo que são: um novo modo de produção emergindo no meio das economias mais avançadas do mundo – aquelas que têm mais computadores conectados e para as quais os bens e serviços da informação vieram para ocupar os papéis de maior valor. (BENKLER, 2006)

Na economia da informação em rede, o capital necessário para a produção é distribuído pela sociedade. Caso um indivíduo não tenha capacidade para levar sozinho um projeto adiante, pode fazer isso em cooperação com outras pessoas que tenham interesses complementares.

O resultado é um setor de produção de informação, conhecimento e cultura fora-de-mercado que floresce baseado no ambiente em rede, aplicado a qualquer coisa que os indivíduos conectados consigam imaginar. E o produto da cooperação, em consequência, não é tratado como propriedade exclusiva, é sujeito à ética crescentemente robusta do compartilhamento aberto, aberto para outros usarem como base, estenderem e personalizarem.

Em suma: na economia da informação de hoje, os modelos industriais fechados perdem espaço para os colaborativos, abertos. E mesmo empresas da “antiga” economia podem adaptar-se ao novo ambiente que surge. Seja abrindo seu código ou permitindo o uso de suas informações pela comunidade, aproveitando em contrapartida as melhorias advindas desta colaboração.

Conseguir que a multidão colabore de forma produtiva, no entanto, não é tão fácil quanto parece. Surowiecki (2004) explica como muitas vezes a sabedoria das multidões supera as decisões das pessoas mais inteligentes dentro do grupo. Ele também mostra que não basta juntar um grupo de pessoas de qualquer maneira para que atinjam grandes resultados. São essenciais ferramentas de mediação e organização da informação para gerenciar com eficácia a inteligência coletiva.

---

5 <http://folding.stanford.edu>

## 2.5 A Emergência da Web Semântica

### 2.5.1 O que é web semântica

Berners-Lee (BERNERS-LEE, 2001) visualiza um mundo em que dispositivos e programas personalizados e especializados, chamados agentes, possam interagir por meio da infra-estrutura de dados da internet trocando informações entre si, de forma a automatizar tarefas rotineiras dos usuários. O projeto da web semântica, em sua essência, é a criação e implantação de padrões tecnológicos para permitir estabelecer panorama que não somente facilite as trocas de informações entre agentes pessoais, mas, principalmente, crie uma língua franca para o compartilhamento mais significativo de dados entre dispositivos e sistemas de informação de uma maneira geral.

Para atingir este propósito, é necessária uma padronização de tecnologias, de linguagens de programação e marcação e de metadados descritivos, de forma que todos os usuários da web obedeçam a determinadas regras compartilhadas sobre como armazenar dados e descrever a informação armazenada e que esta possa ser consumida por outros usuários, humanos ou não, de maneira automática e não ambígua. Com a existência da infra-estrutura tecnológica comum da internet, o primeiro passo para esse objetivo é a dupla proposição de padrões para descrição de dados e de uma linguagem que permita a construção e codificação de significados compartilhados. (SOUZA e ALVARENGA, 2004)

O padrão da web semântica para codificar as relações e significados dos recursos presentes na rede é o RDF (*Resource Description Framework*)<sup>6</sup>. Usando esta linguagem, os dados seriam descritos e referenciados por meio de endereços únicos, chamados URIs (*Universal Resource Identifier*)<sup>7</sup>. Isso significa que, em determinado endereço na internet, fica um arquivo RDF que atua como representante virtual da pessoa “José da Silva”, por exemplo. Esse arquivo, seguindo padrões estabelecidos, diz que o nome da pessoa é José da Silva, que sua data de nascimento é 10/07/1962, que seu e-mail é [jose@provedor.com.br](mailto:jose@provedor.com.br) e seu endereço é Rua das Palmeiras, nº 15. Mais importante que isso, o arquivo RDF pode dizer também, segundo os princípios da web semântica, que José tem uma profissão, a de Engenheiro Civil, e que tem uma relação de sócio com outro cidadão, digamos Marcos, na empresa em que trabalham.

---

6 Arquivos RDF são modelos ou fontes de dados, também conhecidos como metadados, cujos principais objetivos são criar um modelo simples de dados, com uma semântica formal na internet.

7 Cada recurso disponível na Web — documento HTML, imagem, videoclipe, programa, etc. — tem um endereço que pode ser codificado por um Identificador Uniforme de Recursos, ou "URI", iniciais em inglês de Uniform Resource Identifier.

Cada elemento descrito acima pode ter uma representação virtual, uma URI. Com isso, ao acessar “José da Silva”, um agente pode chegar a “Marcos”, à empresa deles ou ao endereço da casa em que ele mora. A rede FOAF (*Friend-Of-A-Friend*)<sup>8</sup> descreve relações de amizade de forma semântica

Dados também podem ser representados dessa maneira. Um elemento da tabela periódica, por exemplo, pode ter a sua própria URI. Acessando este arquivo, um agente poderia descobrir que o elemento sódio “tem símbolo” “Na”, que “pertence à classe” dos “metais alcalinos” e que seu “ponto de ebulição é” “370,87 graus Kelvin”.

Por mais inteligentes que sejam estes agentes virtuais, eles precisam de alguma referência no momento de analisar as relações e “entender” as coisas. Aí entram as ontologias. Usando o padrão OWL (*Web Ontology Language*)<sup>9</sup> recomendado pelo consórcio W3C<sup>10</sup>, os arquivos que expõem metadados podem explicitar qual o “dicionário” do mundo que estão usando, ou seja, que ontologia estão seguindo.

Uma ontologia serve para definir que objetos existem e quais as relações possíveis entre eles. Uma ontologia familiar, por exemplo, poderia dizer que existem “pessoas” que se relacionam por graus de parentesco. Para ligar uma pessoa à outra, deve-se usar A “é filho de” B, ou C “é mãe de” B, ou G “casou-se com” H. A ontologia também pode afirmar que “o filho de um filho é um neto”. A partir disso, ao consultar uma rede de URIs da mesma família, um agente pode coletar não apenas as relações imediatas, mas também as derivações delas.

Embora existam tentativas de criar uma ontologia universal, não é difícil perceber que é inviável a tarefa de englobar todas as coisas, idéias e informações do mundo em um conjunto não ambíguo de denominações no modelo aristotélico de condições necessárias e suficientes. O uso principal das ontologias hoje em dia se faz em comunidades científicas, onde a necessidade de conceitos comuns e da estruturação dos dados é grande. A segmentação do mundo em domínios de conhecimento específicos, sejam científicos, organizacionais ou sociais, facilita o mapeamento virtual dos objetos destes domínios e de suas inter-relações.

---

8 FOAF (um acrônimo para *Friend of a Friend*, ou Amigo de um Amigo) é uma ontologia legível por máquinas que descreve pessoas, suas atividades e suas relações com outras pessoas e objetos. Qualquer um pode usar FOAF para se descrever na internet.

9 A OWL (*Web Ontology Language*) é uma linguagem para definir e instanciar ontologias na Web. Uma ontologia OWL pode incluir descrições de classes e suas respectivas propriedades e seus relacionamentos. OWL foi projetada para o uso por aplicações que precisam processar o conteúdo da informação ao invés de apenas apresentá-la aos humanos.

10 O W3C, ou World Wide Web Consortium, é um consórcio de empresas de tecnologia, atualmente com cerca de 500 membros. Foi fundado por Tim Berners-Lee em 1994 para levar a Web ao seu potencial máximo, por meio do desenvolvimento de protocolos comuns e fóruns abertos que promovem sua evolução e asseguram a sua interoperabilidade. O W3C desenvolve tecnologias denominadas padrões da web (*web standards*) para a criação e a interpretação dos conteúdos para a Web.

O caráter distribuído da web semântica é garantido pela possibilidade de uma ontologia se ligar a outra. Assim, numa ontologia sobre o universo das proteínas, ao me referir a elementos químicos, posso indicar que a descrição deles está em uma outra ontologia, feita por outra instituição de pesquisa. Com isso, áreas de conhecimento podem aproveitar o trabalho anterior realizado por outras.

### 2.5.2 *Folksonomias* e alternativas à web semântica

A web semântica, na forma em que foi idealizada por Tim Berners-Lee, é alvo de críticos como Clay Shirky:

Grande parte do valor proposto pela Web Semântica está chegando, mas não está chegando por causa da Web Semântica. A quantidade de metadados que geramos aumenta dramaticamente e é exposta ao consumo tanto de máquinas quanto – ou em vez – de pessoas. Mas está sendo desenhada um pedaço de cada vez, por interesses próprios e sem consideração pela ontologia global. (...) Há desvantagens significativas para este processo em relação à reluzente visão da Web Semântica, mas a grande vantagem deste design e adoção bottom-up é que está realmente funcionando agora.” (trad. minha).<sup>11</sup> (SHIRKY, 2006)

Um dos fenômenos que Shirky menciona é a *folksonomia*, termo que se refere à categorização espontânea da informação, feita em cooperação por um grupo de pessoas, diferente dos métodos tradicionais de classificação. Ela surge tipicamente em comunidades não-hierárquicas – *sites* de acesso público, por exemplo. Como os próprios usuários são os organizadores da informação, o resultado reflete o modelo conceitual de informação desta população. O neologismo, que numa tradução literal significaria “taxonomia popular” foi criado por Thomas Vander Wal (2004).

Um sistema que usa a *folksonomia* deve ser baseado em *tags*, mas não apenas isso. Se fôssemos determinar regras para identificá-lo, poderíamos considerar que deve conter o seguinte:

- **O objeto** que recebe as *tags*. Simplificando, podemos considerar este objeto como uma URI (Identificador Universal de Recursos). Ela pode representar um *site*, uma

---

<sup>11</sup> “*Much of the proposed value of the Semantic Web is coming, but it is not coming because of the Semantic Web. The amount of meta-data we generate is increasing dramatically, and it is being exposed for consumption by machines as well as, or instead of, people. But it is being designed a bit at a time, out of self-interest and without regard for global ontology. (...) There are significant disadvantages to this process relative to the shining vision of the Semantic Web, but the big advantage of this bottom-up design and adoption is that it is actually working now*”

página específica, um arquivo de texto, uma imagem, um vídeo, uma música, uma notícia ou qualquer outra coisa que se possa acessar via internet ou que tenha uma representação na rede (numa rede de relacionamentos, por exemplo, as *tags* podem se referir a uma pessoa, mas estarão vinculadas a uma URI que representa esta pessoa).

- **As *tags*** – palavras, siglas ou qualquer outro código proposto livremente pelo usuário, de acordo com a sua conveniência. Um objeto pode receber número ilimitado de *tags*.
- **Os usuários** que atribuem *tags* aos objetos. A princípio, sua ação é de interesse pessoal, mas dessa ação emergem resultados interessantes. Quando vários usuários usam a mesma *tag*, por exemplo, isso permite utilizar, tomando-as como referência, algoritmos de recomendação baseados em modelo de usuário<sup>12</sup> ou filtragem colaborativa<sup>13</sup>, por exemplo.

As primeiras referências feitas por Tim Berners-Lee à web semântica remontam uma conferência em 1994 e um artigo publicado na revista *Scientific American*, de 2001. Em um artigo de 2006 que escreveu em parceria com Nigel Shadbolt e Wendy Hall, Tim Berners-Lee revisita a web semântica, responde a críticas e faz uma comparação entre folksonomias e ontologias:

“O uso de *tags* na escala da web é com certeza um desenvolvimento interessante. Provém uma fonte potencial de metadados. As folksonomias que emergem são uma variação das buscas por palavra-chave. São uma tentativa emergente interessante de recuperação de informações. No entanto, folksonomias servem a objetivos muito diferentes dos das ontologias. Ontologias são tentativas de definir com mais cuidado partes do mundo dos dados e permitir mapeamentos e interações entre dados armazenados em diferentes formatos. Ontologias referem por meio de URIs; *tags* usam palavras. Ontologias são definidas por um processo explícito e cuidadoso que tenta eliminar a ambiguidade. A definição de uma *tag* é um processo solto e implícito onde a ambiguidade pode muito bem permanecer. O processo inferencial aplicado às ontologias é baseado em lógica e usa operações como “junção” (join). O processo inferencial usado nas *tags* é de natureza estatística e emprega técnicas como o agrupamento (clustering).” (t.A.)<sup>14</sup> (BERNERS-LEE, SHADBOLT e HALL, 2006)

12 O modelo de usuário padrão é um modelo de dados para o usuário de um sistema. Em teoria, permite que estruturas adaptativas, especialmente software, adaptem-se às características do usuário humano. Ex.: idioma preferido, tamanho das letras, adaptação para daltonismo, volume do áudio entre outros.

13 Filtragem colaborativa é o método de se fazer previsões automáticas (filtragem) sobre os interesses de um usuário ao coletar informação de preferências de muitos usuários (colaboração). A premissa dessa abordagem é: aqueles que concordaram no passado tendem a concordar novamente no futuro.

14 “*Tagging on a Web scale is certainly an interesting development. It provides a potential source of metadata. The folksonomies that emerge are a variant on keyword searches. They’re an interesting emergent attempt at information retrieval. But folksonomies serve very different purposes from ontologies. Ontologies are attempts to more carefully define parts of the data world and to allow mappings and interactions between data held in different formats. Ontologies refer by virtue of URIs; tags use words. Ontologies are defined through a careful, explicit process that attempts to remove ambiguity. The definition of a tag is a loose and implicit process where ambiguity might well remain. The inferential process applied to ontologies is logic based and uses operations such as join. The inferential process used on tags is statistical in nature and employs techniques such as clustering*”

### 2.5.3 Emergência

A facilidade de comunicação entre agentes inteligentes que os padrões de descrição de metadados e também que uma API pode trazer são passos importantes em direção à web semântica em grande escala. Quanto mais fornecedores de informação liberarem acesso às suas bases de dados, mais se potencializa a emergência (JOHNSON, 2001) de soluções criativas e inovadoras utilizando estas informações.

A Teoria Geral de Sistemas permite analisar a web:

A web, com todas as suas páginas, *sites* e aplicações, é um exemplo clássico de sistema na definição proposta por Bertalanffy (1977), já que é composta por um conjunto de elementos (as informações) inter-relacionados (através de hiperlinks). Também é fácil verificar que trata-se de um sistema complexo e emergente: a criação de links entre as páginas não obedece a nenhuma ordem superior, sendo que a “forma” atual da web é resultado da soma de todas as ações dos seus componentes. (LACERDA e VALENTE, 2005, p.1)

Com base no trecho acima, pode-se levar a argumentação adiante, entendendo o inter-relacionamento dos elementos do sistema também como as ligações dinâmicas entre fornecedores e processadores de informação. Como não existe determinação centralizada sobre o que deve ser feito, mas apenas regras de conduta e utilização, o ambiente é ideal para a emergência de novos padrões. O sistema é definido de baixo para cima (*bottom-up*):

Que características comuns têm esses sistemas [emergentes]? Em termos simples, eles resolvem problemas com o auxílio de massas de elementos relativamente simplórios, em vez de contar com uma única “divisão executiva” inteligente. São sistemas *bottom-up*, não *top-down*. Pegam seus conhecimentos a partir de baixo. Em uma linguagem mais técnica, são complexos sistemas adaptativos que mostram comportamento emergente. Neles, os agentes que residem em uma escala começam a produzir comportamento que reside em uma escala acima deles: formigas criam colônias. Cidadãos criam comunidades; Um software simples de reconhecimento de padrões aprende como recomendar novos livros. O movimento de regras de nível baixo para a sofisticação do nível mais alto é o que chamamos de emergência. (JOHNSON, 2001)

Se os computadores conseguem se entender, trocar informações e estão em rede, logo, agentes inteligentes são capazes de exibir comportamento emergente, encontrando informações ou caminhos para resolver problemas não imaginados originalmente por seus programadores. A ampliação do entendimento máquina-a-máquina é essencial para a viabilização desses agentes. Esse raciocínio estabelece a ligação entre a emergência e a web semântica.

#### 2.5.4 Entendimento máquina-a-máquina

Além do padrão RDF utilizado pela web semântica, o entendimento máquina-a-máquina também vem sendo alcançado em outras frentes. O movimento dos padrões de programação na web (*web standards*)<sup>15</sup> prega a separação cada vez maior entre forma e conteúdo na internet. Tirar o código HTML<sup>16</sup> que determina o layout do caminho da informação ajuda muito esse entendimento. Além disso, serviços como RSS<sup>17</sup> e Atom<sup>18</sup> – que são variações de XML<sup>19</sup>, assim como o RDF – transmitem apenas dados estruturados, sem formatação alguma.

Hoje a web se encontra em estado intermediário, onde convivem a bagunça desestruturada das páginas fora de padrão e a estrutura formal e bem definida da web semântica.

Práticas de desenvolvimento para web que vêm ganhando popularidade buscam avançar aos poucos no entendimento entre máquinas, sem mudar radicalmente a maneira com que as pessoas trabalham e escrevem código. Os *web standards* são um avanço, ao pregar o uso das estruturas HTML para os fins que realmente foram criadas.

Durante muito tempo o elemento `<table>`, pensado para exibir uma tabela de dados, acabou sendo usado como grade para a disposição dos elementos na página. A abolição de improvisos como este é o alvo de quem defende a programação baseada em padrões. Vantagem imediata desta prática é a melhoria na acessibilidade do *site*. Programas leitores para cegos, por exemplo, conseguiriam transmitir o conteúdo de forma mais coerente.

---

15 *Web Standards*, ou Normas da Web, são um conjunto de normas, diretrizes, recomendações, notas, artigos, tutoriais e afins de caráter técnico, produzidos pelo W3C e destinados a orientar fabricantes, desenvolvedores e projetistas para o uso de práticas que possibilitem a criação de uma Web acessível a todos.

16 HTML (acrônimo para a expressão inglesa HyperText Markup Language, que significa Linguagem de Marcação de Hipertexto) é uma linguagem de marcação utilizada para produzir páginas na Web. Documentos HTML podem ser interpretados por navegadores como Mozilla Firefox e Internet Explorer.

17 RSS é um subconjunto de "dialetos" XML que servem para agregar conteúdo. Pode ser acessado via programas/sites agregadores. É usado principalmente em sites de notícias e blogs.

18 Ao contrário do RSS, Atom não é uma sigla, mas também é um formato para divulgação de notícias. Há quem diga que esse projeto é, inicialmente, uma proposta de unificação do RSS 1.0 e do RSS 2.0. O Atom também é baseado em XML, mas seu desenvolvimento é tido como mais sofisticado. O grupo que nele trabalha tem até o apoio de grandes corporações, como o Google.

19 XML (eXtensible Markup Language) é uma recomendação da W3C para gerar linguagens de marcação para necessidades especiais. O objetivo do projeto era criar uma linguagem que pudesse ser lida por software, e integrar-se com as demais linguagens. Sua filosofia seria incorporada por várias diretrizes como separação do conteúdo da formatação; simplicidade e legibilidade, tanto para humanos quanto para computadores; possibilidade de criação de tags sem limitação; criação de arquivos para validação de estrutura (chamados DTDs); interligação de bancos de dados distintos; e concentração na estrutura da informação, e não na sua aparência.

Ao escrever uma sigla no texto, por exemplo, pode-se usar o elemento `abbr` (abreviação):

```
<abbr title="Instituto Brasileiro de Geografia e Estatística">IBGE</abbr>
```

Criar um código semântico é uma maneira de – sem separar totalmente a forma do conteúdo – conseguir melhorar a legibilidade e o entendimento para as máquinas. Um exemplo seria o elemento `<p>`, que denota um parágrafo de texto. Pode-se criar para ele uma classe CSS<sup>20</sup>, que nada mais é do que um estilo visual, e aplicá-la assim:

```
<p class="grande_e_azul">
```

Esse é um código que vai ser lido por mecanismos de busca, por agentes virtuais e por leitores para cegos. No entanto, não informa qual o sentido deste parágrafo, e é isso que o código semântico busca alterar. Uma alternativa seria criar o nome do estilo com a finalidade daquele bloco de texto, assim:

```
<p class="dados_pessoais">
```

Microformatos<sup>21</sup> são outra iniciativa, também baseada em dar sentido ao código de uma página, de forma que ele seja inteligível tanto para pessoas que a acessam por um navegador quanto para outras máquinas que tentam entender a informação contida ali. Alguns exemplos de microformatos são os padrões `hCard`<sup>22</sup>, com dados equivalentes a um cartão de visita, e `hCalendar`<sup>23</sup>, com local, dia e horário de algum evento. Caso um agente detecte um `hCard` na página, a pessoa pode facilmente adicioná-lo à sua lista de contatos; caso detecte um evento `hCalendar`, pode adicioná-lo ao seu calendário.

As formas de expressão e leitura das informações variam, no entanto, pode-se observar facilmente a tendência de integração entre as máquinas e o aumento da inteligência dos agentes. Mesmo não atingindo o ideal das ontologias formais da web semântica de Berners-Lee, a evolução em direção a este objetivo tem gerado soluções e inovações tecnológicas em paralelo e melhorado muito a relação das máquinas com a informação e com as pessoas.

---

20 Cascading Style Sheets, ou simplesmente CSS, é uma linguagem de estilo utilizada para definir a apresentação de documentos escritos em uma linguagem de marcação, como HTML ou XML. Seu principal benefício é prover a separação entre o formato e o conteúdo de um documento.

21 Microformatos são um conjunto de formatos abertos projetados para adicionar semântica em qualquer documento XML, especialmente HTML/XHTML. Aplicações, como buscadores, podem extrair informações específicas de páginas que usam microformatos, como informações de contato, evento, licença, etc.

22 `hCard` (abreviação de HTML `vCard`) é um padrão de microformato para publicação de detalhes de contato para pessoas, empresas, organizações e lugares. Pode ser escrito em (X)HTML, Atom, RSS ou XML.

23 `hCalendar` (abreviação de HTML `iCalendar`) é um padrão de microformato para exibir uma representação semântica em (X)HTML de informações de uma agenda em formato `iCalendar` sobre um evento e páginas da Web.

## 2.6 Web 2.0

Termos da moda não faltam quando se fala de internet. Critica-se muito o *hype* (aumento artificial da importância de algo) criado sobre algumas idéias ou empresas. Desde sua criação, em 2005, termo “Web 2.0” é um dos alvos preferidos das críticas, principalmente por não ter uma definição clara.

O que Tim O'Reilly (2005) buscou ao criar o termo foi identificar novas tendências na rede, idéias e técnicas – às vezes nem tão novas assim, mas que passaram a ser adotadas em ritmo crescente por desenvolvedores e empresas de sucesso.

Para se entender a gênese da Web 2.0, é preciso voltar à Web 1.0, a internet do final dos anos 90 e início dos anos 2000. Nesta época houve o colapso da bolsa de valores causado pela especulação em cima das empresas “ponto com” que haviam aberto seu capital. Esse fato ficou conhecido como o “estouro da bolha especulativa” das empresas de internet. Muitos investidores perderam dinheiro em empresas cujo valor futuro nunca iria se equiparar às expectativas infladas do mercado.

Nesse processo, as empresas com dinheiro em caixa investiram pesado – entre outras coisas – na qualificação de seus funcionários, que recebiam a melhor educação possível e salários altíssimos até o estouro da bolha, o fim da maioria destas empresas e, conseqüentemente, dos empregos dentro destas empresas.

Nos anos seguintes, uma força de trabalho qualificada e com experiência na área não tinha mais uma empresa para servir, estava por conta própria. Essas pessoas passaram a criar projetos que julgavam úteis ou interessantes. E aí começou a Web 2.0.

As iniciativas de maior sucesso a partir de 2003 têm por trás desenvolvedores com este histórico. Ao contrário das grandes empresas, que buscavam produtos rentáveis, esses indivíduos estavam concentrados principalmente na utilidade e na praticidade. Ao invés de tentar manter o usuário no grande portal gerando acessos, as novas aplicações não tinham problema em referenciar conteúdos externos, aproveitando ao máximo as vantagens de estar em uma rede mundial.

Emergiu uma nova geração de *sites* que começou a entrar no mercado das grandes empresas já estabelecidas, aplicações com uma filosofia que agradou ao público e que fazia sentido do ponto de vista da usabilidade.

Acentuou-se – e continua forte até hoje – a fórmula de aquisição e integração de pequenas iniciativas por grandes empresas. A Google é a empresa mais proeminente nesta

área, e pagou US\$ 1,65 bilhão pelo *site* de vídeos YouTube<sup>24</sup>. Yahoo!, Microsoft e outras também partiram para essa estratégia.

Como o processo foi emergente, aos poucos foi-se identificando as características de serviços Web 2.0. Tim O'Reilly criou um diagrama para tentar definir as idéias que sustentam esse movimento:

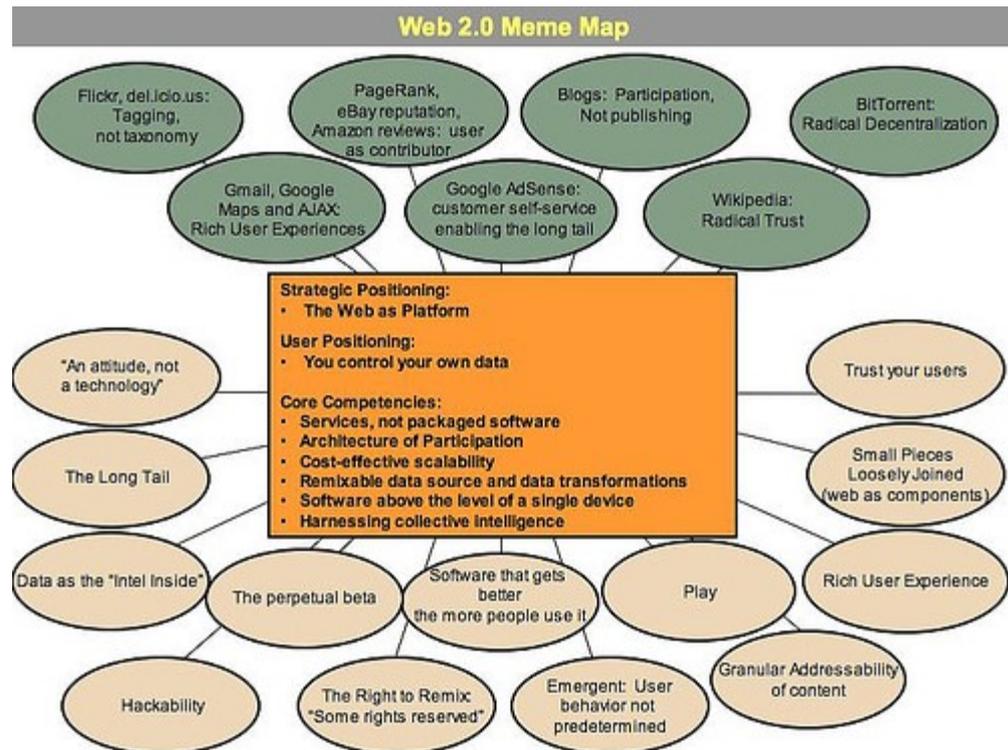


Figura 2.4: Mapa conceitual da Web 2.0

Abaixo, um transcrição dos principais tópicos do artigo de O'Reilly, que aconselha quem deseja se encaixar na nova filosofia:

*"A cauda longa* – Pequenos *sites* são a maioria do conteúdo da internet; pequenos nichos abrigam a maioria das aplicações possíveis na internet. (...)"

"Aplicações são cada vez mais direcionadas a dados. (...) Para vantagem competitiva, busque ser dono de uma fonte de dados única, difícil de recriar."

*"Usuários adicionam valor* – A chave para vantagem competitiva em aplicações na internet é a extensão na qual os usuários adicionam dados ao que você os provém. (..) não

24 <http://www.youtube.com>

restringa sua “arquitetura da participação” ao desenvolvimento de software. Envolve seus usuários implicitamente e explicitamente na adição de valor à sua aplicação.”

“Apenas uma pequena porcentagem dos usuários vai se dar ao trabalho de adicionar valor à sua aplicação. (...) defina padrões inclusivos para agregar dados de usuários como um efeito colateral do seu uso da aplicação.”

“Proteção da propriedade intelectual limita a reutilização e previne a experimentação. (...) quando benefícios vierem da utilização coletiva, e não da restrição privada, permita que as barreiras para a utilização sejam baixas. Siga padrões existentes e use licenças com o menor número de restrições possível. Projete para a 'hackabilidade' e a 'remixagem'.”

“*O beta perpétuo* – Quando aparelhos e programas estão conectados à internet, as aplicações não são mais artefatos de software, são serviços em andamento. (...) Encare seus usuários como testadores em tempo real e instrumentalize o serviço de modo que você saiba como as pessoas usam as novas funcionalidades.”

“*Coopere, não controle* – Aplicações da Web 2.0 são feitas de uma rede de serviços de dados cooperativos. (...) ofereça interfaces de Web Services e distribuição de conteúdo, e reutilize os serviços de dados dos outros. Apóie modelos leves de programação que permitam sistemas mais distribuídos.”

“*Software acima do nível da máquina única* – O PC não é mais a única máquina de acesso a aplicações da internet, e aplicações limitadas a um tipo de máquina são menos valiosas que as sem esse limite. (...) projete sua aplicação desde o início para integrar serviços em computadores de mão, PCs e servidores.”

## **2.7 Perfil do Novo Profissional**

Historicamente os jornalistas têm resistência às novas tecnologias. Como regra, o dia-a-dia da redação não oferece margem para o aprendizado de novas maneiras de se contar uma notícia. Além disso, os avanços tecnológicos são vistos como ameaça ao emprego, uma maneira de cortar postos de trabalho. (BALDESSAR, 2003)

A evolução tecnológica, mesmo assim, tem aproximado o jornalista do desenvolvedor de software de forma inexorável. Seja obrigando o profissional da comunicação a saber consultar bancos de dados para escrever reportagens no modelo do jornalismo de precisão; seja obrigando o profissional da computação a entender o fluxo de trabalho e as peculiaridades de uma empresa jornalística para criar um sistema de publicação de notícias.

Essa é uma das coisas mais difíceis para um programador, entrar no domínio de conhecimento do “cliente”, a pessoa para quem ele produz o software.

O profissional que seria responsável pela produção das aplicações híbridas discutidas nesta dissertação não tem um perfil que se encaixe nos rótulos tradicionais do jornalismo ou da computação. É um profissional também híbrido, de formação interdisciplinar, que consegue dialogar tanto com jornalistas quanto com programadores.

Autores da área de jornalismo defendem a atualização das habilidades necessárias para o exercício da função. Meyer (apud LIMA, 2000) enumera o que um jornalista deve saber para que consiga desempenhar satisfatoriamente o Jornalismo de Precisão:

Recompilar (1), armazenar (2), recuperar (3), analisar (4), resumir (5) e comunicar (6). Normalmente, é da natureza da profissão realizar bem a quinta e a sexta tarefas; a primeira e quarta sempre são desempenhadas, ainda que em graus variados e de maneiras distintas. Já a segunda e a terceira são mais raras; em geral os jornalistas armazenam apenas temporariamente os dados necessários para escrever uma notícia ou reportagem e com pouca frequência resgatam informações em bancos de dados. (LIMA, 2000, p.50)

O resultado da prática do Jornalismo de Precisão é a reportagem. Assim como o objeto principal de todo o jornalismo que é praticado na maioria das empresas de comunicação. Via de regra, o texto da reportagem ou da notícia é um conteúdo estático. Pode se dizer, simplificando, que o texto é um recorte da realidade em um determinado momento, transformado em relato jornalístico pelo repórter.

O que a internet proporciona é a passagem do estático para o dinâmico. O resultado do trabalho jornalístico não está mais eternizado numa folha de papel, não sofre com o fluxo temporal que exige a presença do telespectador ou do ouvinte em frente ao rádio e à TV no momento da transmissão da informação. O resultado do trabalho é visto pelo leitor direto na fonte, direto onde a notícia foi produzida, na hora que lhe convier.

Essa constatação, por revolucionária que possa parecer, não teve implicações proporcionais na prática profissional do jornalista. O mesmo modelo de trabalho do mundo desconectado foi adotado no mundo em rede, e as características do novo meio tratadas como meras conveniências: a facilidade de correção de um texto publicado, a rapidez para se colocar notas no ar ou a possibilidade de trabalhar à distância.

Os avanços tecnológicos e os sistemas sociais para lidar com eles não evoluem no mesmo ritmo. Quando velhos sistemas sociais e culturais são aplicados a novas

maneiras de fazer as coisas, o encaixe é por vezes desconfortável e até doloroso (MEYER, 1991).

Enquanto o jornalismo online busca encontrar sua identidade, o jornalista que publica na internet tenta adaptar o meio à sua identidade pré-concebida. Baldessar observa em outros autores a preocupação com as habilidades necessárias para o exercício da profissão de jornalista nos dias de hoje:

Tanto Lage quanto Corrêa congregam algumas percepções comuns: domínio de língua estrangeira, capacidade de análise do mundo globalizado e, domínio de técnicas e softwares que respondam ao momento. No entanto, esta visão não encontra ressonância nos currículos das escolas de Jornalismo e, conseqüentemente, na formação profissional (BALDESSAR, 2003)

Pela lógica, se as escolas treinam jornalistas para produzir conteúdo estático, ou seja, notícias e reportagens no modelo clássico, é isso que estes profissionais vão fazer ao se depararem com um meio de comunicação dinâmico.

Os profissionais da computação, por outro lado, são formados para trabalhar com um universo de dados dinâmicos, em constante atualização e interligação. Mas falta a eles o corpo de conhecimento que permita criar softwares de cunho jornalístico. Ferramentas que sirvam para “transmitir a informação de forma que supere a sobrecarga informacional e chegue ao público que a necessita e deseja” (LIMA, 2000).

Seria possível inculcar estas habilidades técnicas no jornalista? Ou então treinar o programador para ver o mundo com um olhar jornalístico? “Como formar um jornalista que saiba aliar a capacidade técnica de produção com um olhar crítico da realidade? Para muitos essa parceria é inviável.” (BALDESSAR, 2003)

A autora segue questionando:

- (1) as novas tecnologias da informação desencadearam uma discussão sobre a identidade e a sobrevivência das profissões que eram responsáveis pela mediação simbólica. Nesse contexto, o que é ser (...) jornalista na atualidade?
- (2) sendo as ciências da Comunicação e o Jornalismo, e os estudos teóricos relacionados a ambas – como os estudos culturais, um dos locais onde se procede uma reflexão multifacetada e transdisciplinar sobre o mundo de hoje, como deve ser a formação de um profissional que dê conta dessa realidade, levando em conta questões éticas, estéticas e de linguagem que as especificidades do Jornalismo exigem?

(3) considerando o Jornalismo online como uma transposição de uma certa forma de olhar a realidade (jornalístico) para o suporte informático é possível afirmar que a especificidade do meio não altera a mensagem?

(4) até onde a construção desse profissional deve aprofundar saberes específicos ou mesclá-los com generalidades e saberes localizados? (BALDESSAR, 2003)

As respostas a estas questões, diz ela, talvez possam ser facilitadas se tivermos claro que o jornalismo sempre teve seu fazer cotidiano ligado à tecnologia. A cada novo invento a profissão modificou suas práticas, desenvolveu linguagens, criou novas formas de mostrar o mundo através da informação. (BALDESSAR, 2003)

A essência do jornalismo – sua mensagem – é algo que independe do meio. Pode ser transmitida em que suporte for, mas seguindo sempre sua razão de ser: informar o leitor com conteúdo relevante. Isso significa que organizar a informação já existente e criar maneiras para acessá-la mais facilmente também são tarefas do jornalista. São alternativas para se explorar verdadeiramente as potencialidades do conteúdo dinâmico.

## **2.8 Jornalismo *Online***

### 2.8.1 Histórico

Diversos autores, entre eles Castilho (2006), constataam que revolução tecnológica contemporânea teve um impacto enorme em diversas áreas, com destaque para as mudanças no campo da comunicação e, em especial, o jornalismo.

As mudanças ocorrem na base econômica da comunicação, passando pelas funções que exercem os protagonistas da informação e pelo no modelo de negócios das empresas. Surgem como conseqüência novos conceitos de jornalismo e de notícia. Acontece a convergência de canais de comunicação e a emergência do usuário como produtor e *publisher* em pouco mais de uma década. (CASTILHO, 2006)

Trata-se da mais profunda mudança registrada na imprensa desde que o ourives alemão Johannes Gutenberg inventou, em 1450, o sistema de impressão que deu origem à imprensa. (CASTILHO, 2006)

A combinação de informática (desenvolvimento da digitalização da informação por meio de computadores) e da telemática (desenvolvimento de redes de transmissão de dados digitalizados) deu origem ao que Clayton Christensen chamou de tecnologias de ruptura (*disruptive technologies*) que, além de mudarem os paradigmas infraestruturais, alteram também a estrutura social (CHRISTENSEN, 1997 apud CASTILHO, 2006).

Sistemas convencionais entram em crise porque já não conseguem mais atender às demandas materiais e sociais criadas pela inovação tecnológica. Novas práticas e novas realidades provocam o surgimento de novos valores e comportamentos humanos, que entram em conflito com os que ainda estão em vigor, provocando um ambiente de insegurança e incerteza quanto ao futuro. (CASTILHO, 2006)

Castilho explica que economia baseada em *commodities* sofre o impacto da emergência da informação como bem mais valorizado. Em vez de se esgotar com o uso, a informação se multiplica à medida em que circula entre as pessoas. Em contraste com operações econômicas tradicionais, onde alguém troca um bem ou serviço por dinheiro, ou com o escambo, onde há a troca de um bem por outro; com a troca de informação ambas as partes continuam com o que tinham antes e ganham uma informação nova.

O autor diz que quanto mais intensa for a troca, maior será o ganho de ambas as partes, “o que indica um potencial de geração de riquezas inédito na história da economia mundial”. Para ele, o corolário desta mudança é uma transformação radical dos valores humanos associados a questões como o individualismo, privacidade, direito autoral e autoria compartilhada entre outros.

É neste contexto que o jornalismo e a imprensa começam a sofrer as conseqüências da revolução tecnológica e são forçados a adaptar-se às novas circunstâncias. (CASTILHO, 2006)

A adaptação do jornalismo vem acontecendo aos poucos. Os primeiros jornais que se aventuraram no terreno digital começaram reproduzindo ali o mesmo formato e o mesmo conteúdo que utilizavam no papel.

Mielniczuk (2003) classifica a evolução do jornalismo *online* em três fases, ou gerações, que podem ser identificadas no Brasil desde a implantação da internet (Tabela 2.1).

A postura de grande parte dos jornalistas e dos pesquisadores da área de comunicação, no entanto, é superficial em relação à tecnologia, o que prejudica uma análise mais profunda sobre o impacto que ela pode ter na atividade informativa.

Um exemplo é a extensa literatura dedicada ao estudo do *hiperlink* e suas implicações narrativas, que nada mais é do que a ponta do *iceberg*, é o que está visível sobre todo o sistema emergente da rede mundial de computadores. A cultura da transposição – identificada por Ribas no jornalismo – também se aplica nos estudos de comunicação sobre a internet.

Tabela 2.1: As gerações do jornalismo na internet

GERAÇÕES	CARACTERÍSTICAS
<b>Primeira geração ou Fase da transposição</b>	<ul style="list-style-type: none"> <li>- reproduções de partes dos jornais impressos</li> <li>- cópias do conteúdo de jornais existentes no papel</li> <li>- material atualizado a cada 24 horas</li> <li>- disponibilização de conteúdo de alguns cadernos semanais</li> <li>- rotina de produção de notícias atrelada ao modelo estabelecido nos jornais impressos</li> <li>- nenhuma preocupação com relação a uma possível forma inovadora de apresentação das narrativas jornalísticas</li> <li>- ocupação de um espaço, sem explorá-lo, enquanto um meio que apresenta características específicas.</li> </ul>
<b>Segunda geração ou Fase da metáfora</b>	<ul style="list-style-type: none"> <li>- o jornal impresso funciona como uma referência para a elaboração das interfaces dos produtos</li> <li>- as publicações começam a explorar as potencialidades do novo ambiente</li> <li>- uso de e-mail e fórum de debate para contato do usuário com o produtor ou com outros usuários</li> <li>- surgimento da seção 'últimas notícias'</li> <li>- exploração do uso mais elaborado do hipertexto.</li> </ul>
<b>Terceira geração ou Fase do Webjornalismo</b>	<ul style="list-style-type: none"> <li>- <i>sites</i> jornalísticos que extrapolam a idéia de uma versão para a Web de um jornal impresso já existente</li> <li>- exploração e aplicação das potencialidades oferecidas pela Web para fins jornalísticos: hipertextualidade, interatividade, multimídia, personalização, memória.</li> </ul>

Fonte: MIELNICZUK, 2003 apud RIBAS, 2005

### 2.8.2 O modelo centrado na matéria

O modelo centrado na matéria deriva da limitação do entendimento entre a área tecnológica e os jornalistas. Quem desenvolve o argumento é Adrian Holovaty, criador do *site* Chicago Crime<sup>25</sup> e diretor de projetos especiais na internet do jornal Washington Post<sup>26</sup>. Ele critica principalmente a maneira como os jornais tratam a informação na internet.

Para ele, mudanças fundamentais devem acontecer para que as empresas jornalísticas continuem sendo fontes essenciais de informação para suas comunidades. E uma das mudanças mais importantes é: jornais precisam acabar com a visão de mundo centrada na matéria.

<sup>25</sup> <http://www.chicagocrime.org>

<sup>26</sup> <http://washingtonpost.com>

Conicionados por décadas de um estilo consolidado de jornalismo, os repórteres de jornal tendem a perceber suas funções assim:

1. Coletar informação
2. Escrever uma matéria de jornal

O problema é que, para muitos tipos de notícias e informação, as matérias de jornal não servem mais. O autor justifica dizendo que muito do que os jornalistas locais coletam no dia-a-dia é informação estruturada: o tipo de informação que pode ser quebrada e fatiada de maneira automática por computadores. Ainda assim, a informação é transformada num grande bloco de texto – a matéria de jornal – que não tem chance de ser reutilizada.

Holovaty explica que o termo “reutilizar” não se refere a “mostrar a matéria em um telefone celular”; não significa “mostrar a matéria em RSS” ou “mostrar a matéria no seu computador de mão”. Estes são objetivos louváveis, diz ele, mas são exemplos de alteração no formato, e não na informação em si. Reutilizar e agregar informação é outra história, e requer que a informação seja armazenada atômica e em formatos legíveis por máquinas.

Por exemplo, imagine que um jornal escreveu uma matéria sobre um incêndio local. Poder ler esta matéria no celular é ótimo. Viva a tecnologia! Mas o que eu realmente quero poder fazer é explorar os fatos “crus” dessa matéria, um por um, com camadas de atribuições e uma infraestrutura para comparar os detalhes do incêndio – data, hora, lugar, vítimas, número da viatura dos bombeiros, distância até o quartel, nomes e anos de experiência dos bombeiros no local do incêndio, tempo que levaram para chegar – com detalhes de incêndios anteriores. E incêndios subsequentes, quando acontecerem (HOLOVATY, 2006)

Para ele, esta é a definição de dados estruturados: informação com atributos que sejam consistentes em um domínio. Todo incêndio possui aqueles atributos, assim como todo crime denunciado tem vários atributos, assim como todo jogo de futebol tem vários atributos.

Estes três exemplos são candidatos óbvios para estrutura, principalmente por sua ubiquidade. Pessoas têm quebrado e fatiado os dados de esportes por anos. Pessoas têm analisado as estatísticas de criminalidade por anos.

Mas esses exemplos óbvios não são os únicos. Holovaty explica que se alguém analisar o tipo de informação que o repórter de jornal coleta, a quantidade de estrutura vai saltar aos olhos. Ele exemplifica:

- Um obituário é sobre uma pessoa, envolve datas e funerárias.
- Um anúncio de casamento é sobre um casal, com uma data de casamento, data de noivado, cidade-natal da noiva, cidade-natal do noivo e vários outros felizes pedaços de informação.
- Um nascimento tem pais, uma criança (ou crianças) e uma data.
- Um formando tem um estado-natal, uma cidade-natal, um diploma, uma pareia de graduação e um ano de graduação.
- Uma enquete tem entrevistados, respostas e uma data de publicação.
- Uma bebida especial tem um dia da semana em que é servida e um bar.
- A agenda do congresso tem um dia e itens a serem cumpridos.
- Uma propaganda política tem um candidato, um estado, um partido, vários assuntos, personagens, etc.
- Cada eleição para o senado, câmara de deputados e governo tem uma localização, análise, informação demográfica, resultados de eleições anteriores, informação de financiamento de campanha e mais...
- Cada preso na baía de Guantânamo tem uma idade aproximada, um local de nascimento, indiciamento formal e etc.

(HOLOVATY, 2006)

Percebe-se um padrão na lista. Muitas das informações que os jornais coletam são inexoravelmente estruturadas. Só é necessário que alguém perceba a estrutura (a parte fácil) e comece a armazenar em formato estruturado (a parte difícil).

O autor diz entender porque jornais são lentos em aceitar este tipo de raciocínio. “Jornalistas não são o grupo mais ligado em tecnologia, não são o grupo mais inovador, e são resistentes a mudanças.”

Para ele, uma barreira a esse raciocínio é um tipo de arrogância jornalística: “Como isso é jornalismo? Nós somos jornalistas e fomos treinados para explicar informações complexas para o público de maneiras que ele possa entender. Mostrar os dados 'crus' não ajuda as pessoas; escrever uma matéria ajuda as pessoas, porque está em inglês [ou português] claro”. Holovaty conta que apresentou estes conceitos (jornalismo via programação de computador, a importância de dados legíveis por máquinas e etc.) em vários eventos relacionados a jornalismo e inevitavelmente ouviu as perguntas acima.

Ele responde dizendo que a questão sobre “Como isso é jornalismo?” é acadêmica. Jornalistas deveriam se importar menos sobre o que é e o que não é “jornalismo” e se importar mais com informação importante, concentrada que seja útil para a vida das pessoas e ajude-as

a entender o mundo. Um jornal deve ser isso, na opinião dele: um olhar justo sobre as informações importantes e atuais para seus leitores.

Em seguida, diz que não faz uma proposta “tudo ou nada”; não sugere que os jornais devam converter totalmente vastas coleções de dados, abandonando completamente o formato da matéria jornalística. Matérias são ótimas para contar histórias, analisar assuntos complexos e vários tipos de outras coisas. Uma matéria – grande bloco de texto – é geralmente a melhor maneira de se explicar conceitos. As sutilezas do idioma não têm correspondência em fontes de dados manipuláveis por máquinas. Quando Holovaty diz que “jornais devem acabar com a visão de mundo centrada na matéria”, ele não quer dizer que “jornais devem abolir matérias”. As duas formas de disseminação de informações podem coexistir e complementar uma à outra.

Para o autor, além da arrogância jornalística, outro problema é que a estrutura organizacional e de software das empresas jornalísticas desencoraja fortemente qualquer tipo de “embalagem especial da informação”. Quase todos os sistemas de gerenciamento de conteúdo (CMS – *Content Management System*)<sup>27</sup> para os *sites* de jornais são “desavergonhadamente” centrados na matéria.

Para inserir uma informação de agenda em alguns dos CMS mais usados, é preciso postá-la como um objeto “matéria”. Para publicar listas de crimes recentes na cidade deve-se usar o formato “matéria”. Não há muito que um repórter ou um editor *online* possam fazer sobre isso. A justificativa dada pelas empresas, diz Holovaty, é que muito foi investido neste sistema e/ou o *site* do jornal não emprega nenhum programador. Ele compara a última situação com um diretor de cinema se recusando a empregar operadores de câmera ou editores de vídeo.

O que acontece com os sistemas de publicação de conteúdo utilizados pelas empresas jornalísticas, é que os jornalistas se adaptam às suas limitações. Em alguns ambientes, para publicar apenas uma foto desvinculada de uma matéria, mesmo assim é preciso criar um objeto “matéria” sem texto algum, apenas com uma fotografia. Para se criar uma enquete, com perguntas e respostas de diversos entrevistados, abre-se o modelo de “matéria” e coloca-se tudo ali dentro. Qualquer coisa que se deseja inserir no *site* tem que entrar como “matéria”, porque é a única coisa que o sistema sabe fazer.

Este problema é sutil, e Holovaty teve dificuldades em explicá-lo aos jornalistas. Inicialmente eles não entendem porque inserir tudo como uma “matéria” é ruim. Para eles, o

---

<sup>27</sup> Sistema de Gerenciamento de Conteúdo, do inglês Content Management Systems - CMS, é um framework, “um esqueleto” de website pré-programado, com recursos básicos e de manutenção e administração já prontamente disponíveis. É um sistema que permite a criação, armazenamento e administração de conteúdo de forma dinâmica, através de uma interface de usuário via internet.

sistema de publicação é apenas o meio para um fim: levar a informação até o público. Eles querem que seja rápido e fácil levar a informação X até o *site* Y. O objetivo não é ter dados limpos, é publicar os dados rapidamente, com bônus por uma interface agradável, ironiza o autor.

Para ele, que diz se concentrar no longo prazo, o objetivo é armazenar a informação no formato mais valioso possível. O problema é frustrante de ser explicado porque não é necessariamente óbvio; se a empresa armazena tudo no seu *web site* como uma “matéria”, o *site* não é necessariamente difícil de usar. No entanto, é um problema de oportunidade perdida. Se toda a informação está na vala comum das “matérias”, não é possível tirar com facilidade somente os crimes e pintá-los sobre o mapa da cidade. Não dá para pinçar os eventos com facilidade e criar uma agenda. A empresa acaba se acomodando no menor denominador comum: um *site* que sabe como mostrar apenas um tipo de conteúdo, um grande bloco de texto. Esse *site* não sabe fazer as coisas que os usuários estão começando a exigir.

Além disso, na nova abordagem há a capacidade de se descobrir coisas por acaso, o serendipismo<sup>28</sup>. Com o armazenamento de dados estruturados, é comum que emergjam usos para estes dados que não foram imaginados no início. Esses usos podem aparecer dentro da redação ou surgir a partir dos leitores, dependendo do nível de colaboração permitido pelo *site*.

Holovaty conclui dizendo que muitos jornais têm procurado mudar o panorama atual e têm projetos para armazenar dados de forma reutilizável. A mentalidade de software de código aberto é uma aliada no esforço de mudança.

Em contraste com o software proprietário, que é comprado em um pacote numa loja e não pode ser alterado, os programas de código aberto são livres e personalizáveis. Nesse ponto está a mudança de mentalidade: não é mais necessário esperar que a nova versão do software proprietário – por sorte – resolva as necessidades específicas de uma determinada empresa. Com o código aberto, é possível contratar um desenvolvedor para deixar o software exatamente como a redação deseja. Não é mais o usuário que se adapta às limitações do software, é o software que se adapta às necessidades do usuário.

---

<sup>28</sup> Serendipismo, do inglês Serendipity, é um termo científico para designar o processo gradual de acúmulo de conhecimento científico ou de resultados de pesquisadores ou inventores até se chegar a um resultado mais atualizado ou uma nova invenção. Geralmente, o termo é empregado para indicar descobertas feitas por acaso, ou seja, ao se procurar uma coisa, encontra-se outra. O psicólogo behaviorista B. F. Skinner afirmou que o serendipismo foi um dos poucos princípios metodológico-científicos que seguiu durante suas investigações.

## CAPÍTULO 3 – O FERRAMENTAL

### 3.1 Metodologia

A abordagem do problema de pesquisa apresentado neste estudo se encaixa na definição de pesquisa qualitativa descrita por Gil (GIL, 1999). Segundo ele, há na pesquisa qualitativa relação dinâmica entre o mundo real e o sujeito, ou seja, uma ligação estreita entre o mundo objetivo e a subjetividade do sujeito. Essa ligação, pelo que tem de subjetividade, não pode ser descrita em números como em muitos processos das ciências exatas. A atribuição de significados aos fenômenos relatados na pesquisa, bem como a interpretação deles, são essenciais no decorrer da pesquisa qualitativa.

É importante destacar que pesquisadores que optam por utilizar os métodos qualitativos procuram esclarecer as razões das coisas, expressando o que deve ser feito, mas não enumeram os valores e as trocas simbólicas nem se submetem à prova de fatos, porque os dados analisados são não-métricos e se valem de diversas abordagens.

Considerando os objetivos, esta dissertação caracteriza-se como exploratória, porque busca dar visibilidade ao problema do desenvolvimento e concepção das Aplicações Web Híbridas no jornalismo. A pesquisa faz isso com o objetivo de torná-lo explícito, construir hipóteses e modelos.

O levantamento bibliográfico também é parte da pesquisa exploratória, conforme mostra Gil (GIL, 1999). Além disso, este tipo de estudo pode incluir tanto entrevistas com pessoas que tiveram experiências práticas com o problema pesquisado quanto a análise de exemplos que incentivem o entendimento do assunto abordado. A dissertação, por consequência, assume as formas de pesquisa bibliográfica e estudo de caso.

### 3.2 Tecnologias

#### 3.2.1 APIs e *Web Services*

As páginas web vistas por milhões de usuários em seus navegadores de internet são o padrão para distribuição de informação e para a execução de serviços – transações bancárias, compras ou buscas em bases de dados. Mas são um padrão para consumo humano. As páginas que se recebe no navegador trazem, além da informação, elementos visuais e estruturais em

uma variedade de linguagens de programação e níveis de clareza de descrição. Estes são fatores que impedem as máquinas de entenderem a informação trazida pela página.

Quando o objetivo da comunicação via internet ou entre aplicações não é o consumo humano, mas sim o entendimento entre diferentes sistemas, uma “linguagem de máquina” é usada. Surge então o termo API (*Application Programming Interface* ou Interface de Programação de Aplicações). É uma porta de entrada personalizada para um software ou um sistema. Ela é criada por quem programa e contém regras sobre o que é possível obter por meio dela e também sobre como se deve fazer isso.

Qualquer *software* pode ter uma interface deste tipo, permitindo que aplicações de terceiros *conversem* com ele e ambos se integrem para a execução de determinadas tarefas. Um exemplo claro disso é quando um programa de mensagens instantâneas *conversa* com o tocador de música e exibe o nome da canção que está sendo ouvida naquele momento. Duas aplicações diferentes, criadas por desenvolvedores diferentes e até em linguagens diferentes, encontram na API um ponto em comum para trocar informações.

Cada vez mais o conceito de páginas estáticas na internet vem sendo deixado de lado e dando lugar para as aplicações web dinâmicas, que são *sites* com características de software (CONALLEN, 1999). Além de ter funcionalidades de um *software offline*, elas aproveitam as características da rede para ampliar seu potencial. Estas aplicações também podem ter interfaces de programação acessíveis pela internet.

Na internet já existem padrões de comunicação máquina-a-máquina, conhecidos como *Web Services*, que formam uma camada de abstração, útil na integração de diferentes sistemas. São como um “balcão de serviços” que faz sistemas diferentes trocarem informações entre si:

“*Web Services* fornecem uma nova camada de abstração sobre os sistemas de software existentes, capaz de interligar qualquer sistema operacional, plataforma de hardware ou linguagem de programação. Enquanto a *Web* é principalmente para usuários humanos, os *Web Services* fornecem um framework para comunicação programa-a-programa. *Web Services* são basicamente adaptadores entre aplicações distribuídas, que permitem codificar mensagens em um formato canônico e mandá-las através da internet. Pela adoção ampla desta tecnologia, aplicações em vários locais da internet podem estar diretamente interconectadas como se fossem parte de um único, grande sistema de informação.” (trad. minha) (GAMPER e AUGSTEN, 2005)<sup>29</sup>

---

29 “Webservices provide a new layer of abstraction above existing software systems, capable of bridging any operating system, hardware platform or programming language. While the Web is mainly for human users, Web services provide a framework for program-to-program communication. Web services are basically adapters between distributed applications, which allow to map messages into a canonical format and to send them across the internet. Through the widespread adoption of this technology, applications at various internet locations can be directly interconnected as if they were part of a single, large information system.”

Logicamente, para aproveitar a disseminação desses padrões, as APIs de aplicações web têm usado em grande parte, mas não exclusivamente, os *Web Services* em sua arquitetura.

Para se entender melhor como funcionam as interfaces de programação via internet, basta olhar para as cadeias produtivas da indústria ou de serviços. O dono de uma padaria não precisa também ser dono de um moinho de farinha. Ele faz o pedido e a farinha é entregue na quantidade especificada. O que é possível com as APIs é a terceirização do armazenamento, do processamento e da organização da informação. Por meio delas, o desenvolvedor faz pedidos pontuais que são atendidos pelo sistema “terceirizado”.

A Amazon.com é um exemplo claro disso. Por meio de sua subsidiária Amazon Web Services, ela coloca à disposição de desenvolvedores alguns serviços como o *Amazon Simple Storage Service (S3)*<sup>30</sup>, iniciado em março de 2006, que possibilita aproveitar a estrutura da gigante do comércio eletrônico e hospedar em seus servidores qualquer arquivo, terceirizando o espaço em disco e a banda necessária para que os usuários acessem os arquivos. O serviço é pago de acordo com a intensidade de uso. Outro serviço um pouco mais abstrato é o *Elastic Compute Cloud (EC2)*<sup>31</sup>, introduzido em agosto de 2006, que serve para terceirizar a capacidade de processamento e armazenamento em banco de dados. Segundo a empresa, no final de 2006 o programa de ferramentas para desenvolvedores da Amazon tinha cerca de 200 mil membros registrados.

Outro gigante do ramo dos leilões online, o eBay, também tem um programa para desenvolvedores, mas não cobra pela sua utilização. Mais de 20% dos produtos listados no *site* vêm de aplicações de terceiros, que automatizaram o cadastramento de múltiplos itens através de um interface de programação. Segundo o diretor do eBay, Greg Isaacs, o programa cadastrara em 2006 cerca de 18 mil desenvolvedores.

O cadastro de desenvolvedores é uma maneira de impedir o uso indiscriminado dos recursos e controlar em detalhes o que cada um está fazendo. Cada desenvolvedor recebe uma chave de acesso única que deve ser usada para fazer qualquer pedido e bloquear o acesso de quem não cumpre os termos de uso.

Joshua Bloch (2005), principal engenheiro de software do Google, lembra que uma API deve ser projetada com muito cuidado; adverte que ela tanto pode ser um dos principais ativos de uma empresa quanto uma de suas maiores fraquezas. Pode ser um ativo importante porque os clientes investem pesado nela ao pagar pela sua utilização, ao escrever códigos para

---

30 <http://aws.amazon.com/s3>

31 <http://aws.amazon.com/ec2>

integrá-la aos seus sistemas e ao dedicar tempo para aprender como usá-la. Depois disso, o custo de deixar de usá-la pode ser proibitivo, o que beneficia o fornecedor do serviço.

Bloch também argumenta que a publicidade criada por uma API pública de sucesso pode trazer novos clientes. Do lado negativo, ressalta que uma mal feita pode resultar em seqüência infindável de ligações para o suporte técnico: “Uma API pública é para sempre, e só se tem uma chance para fazê-la direito”

Para os fornecedores de informação, Bloch aconselha:

- Se você programa, é um designer de APIs
- O bom código é modular – cada módulo tem uma API
- Módulos úteis tendem a ser reutilizados
- Uma vez que um módulo tem usuários, a API não pode ser modificada ao bel-prazer
- Bons módulos reutilizáveis são ativos da empresa
- Pensar em termos de APIs aumenta a qualidade do código.

(BLOCH, 2005)

Essa descrição do engenheiro do Google se encaixa bem no que a Amazon fez: utilizou seus módulos de programação e sua estrutura interna, que tinha potencial para expansão, e compartilhou-os com o público externo, mediante pagamento.

Joshua Bloch enumera algumas características que uma boa API deve ter:

- Fácil de aprender;
- Fácil de usar, mesmo sem documentação
- Difícil de usar erroneamente
- Utilizando código fácil de ler e de manter
- Suficientemente poderosa para satisfazer as exigências
- Fácil de estender
- Apropriada para o público

(BLOCH, 2005)

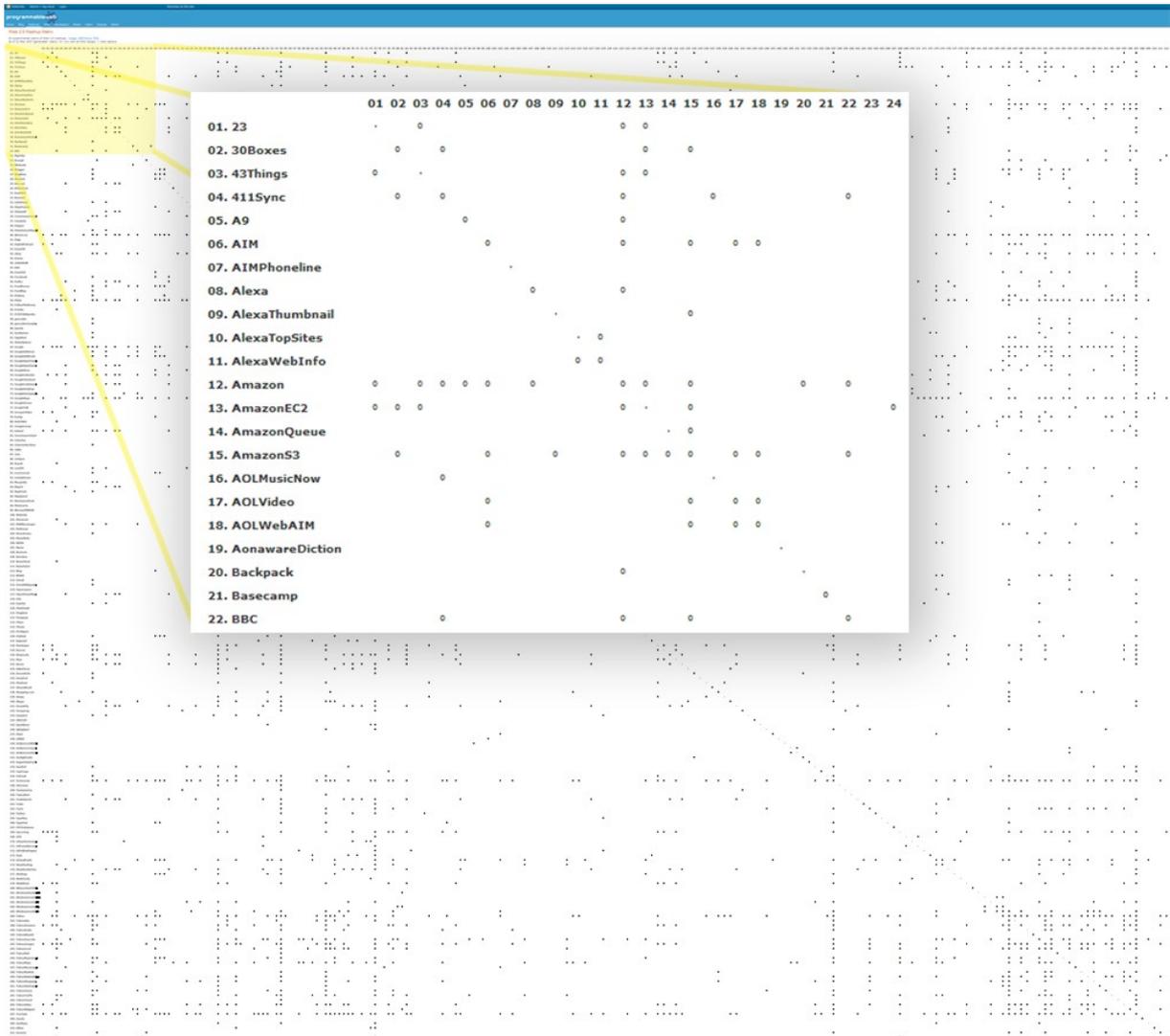


Figura 3.1: Matriz de combinação de APIs do site Programmable Web

O site Programmable Web<sup>32</sup>, especializado no assunto, contabilizava, em novembro de 2006, 312 interfaces de programação disponíveis na internet. Segundo o site, as mais utilizadas por terceiros são as de aplicações do Google, seguidas pelas da Amazon.com, Microsoft, eBay, Yahoo!, AOL e outras. As categorias englobadas são as mais diversas, incluindo buscas, publicidade, comércio, bate-papo, música, jogos, previsão do tempo, listas de favoritos, eventos, e mapas.

A interface mais utilizada por terceiros, de acordo com a contagem do site, é a do Google Maps<sup>33</sup>, presente em mais de 600 aplicações, seguida pela do Flickr<sup>34</sup>, de fotografias, com 134 utilizações. A API do Google Maps dá ao desenvolvedor a possibilidade de trabalhar

<sup>32</sup> <http://www.programmableweb.com>

<sup>33</sup> <http://maps.google.com>

<sup>34</sup> <http://www.flickr.com>

com mapas do mundo todo na forma de foto de satélite ou tradicional, e colocar sobre este mapa elementos personalizados baseados na localização geográfica.

A Figura 3.1 mostra a matriz de *mash-ups* do *site* Programmable Web. Um *mash-up* é uma Aplicação Web Híbrida não necessariamente jornalística, mas que cruza dados de fontes distintas por meio de interfaces de programação de aplicações. Na Figura 3.1, um detalhe foi ampliado para dar a dimensão das combinações que podem ocorrer. Cada ponto preenchido da matriz pode denotar um ou mais *mash-ups* que combinam dados das APIs que são mostradas na horizontal e na vertical.

Quando não há uma interface que forneça as informações de determinado *site* para desenvolvedores independentes, mas há esta informação em páginas da internet acessíveis pelo navegador, cria-se o ambiente propício para a “raspagem” ou *scraping*.

A raspagem é defendida por seus praticantes como a única saída para utilizar informações que são acessíveis exclusivamente nas páginas da internet e não via *Web Services*. Ela consiste em criar um “robô” que varre um *site* e pinça determinadas informações para uso em outra aplicação. Esta técnica é frágil e arriscada, por ser aplicada geralmente sem autorização da fonte e por depender da diagramação da página – qualquer alteração de layout pode desorientar o “robô”.

Do ponto de vista do *site* que é “raspado”, haveria economia de recursos e maior controle sobre a liberação dos dados se houvesse uma API para os desenvolvedores externos. Isso reduziria o custo de banda e de processamento, tendo em vista que apenas requisições pontuais de informação seriam feitas, e não o descarregamento de páginas inteiras com imagens e elementos estruturais.

O prêmio Batten de 2005 para inovação no jornalismo, no valor de US\$ 10 mil, foi entregue a Adrian Holovaty por seu trabalho no *site* Chicago Crime<sup>35</sup> (Figura 3.2). Os juízes descreveram o vencedor como “uma integração pioneira entre geomapeamento e uma base de dados pública, que resulta em um dos mais abrangentes *sites* sobre crime (...)”<sup>36</sup>. O *site* mostra, sobre o mapa da cidade de Chicago, as ocorrências policiais, seus tipos e sua frequência, permitindo que se interaja com a informação de várias formas.

O interessante desse projeto é que ele usa a API do Google Maps combinada com a “raspagem” das informações sobre crime disponíveis *online* no *site* da polícia de Chicago. O próprio Holovaty admitiu, em declaração à imprensa, que “se a polícia resolver mudar o layout do *site*, seu *raspador* de informações precisará ser reescrito”.

---

35 <http://www.chicagocrime.org>

36 <http://www.j-lab.org/batten05winners.shtml>

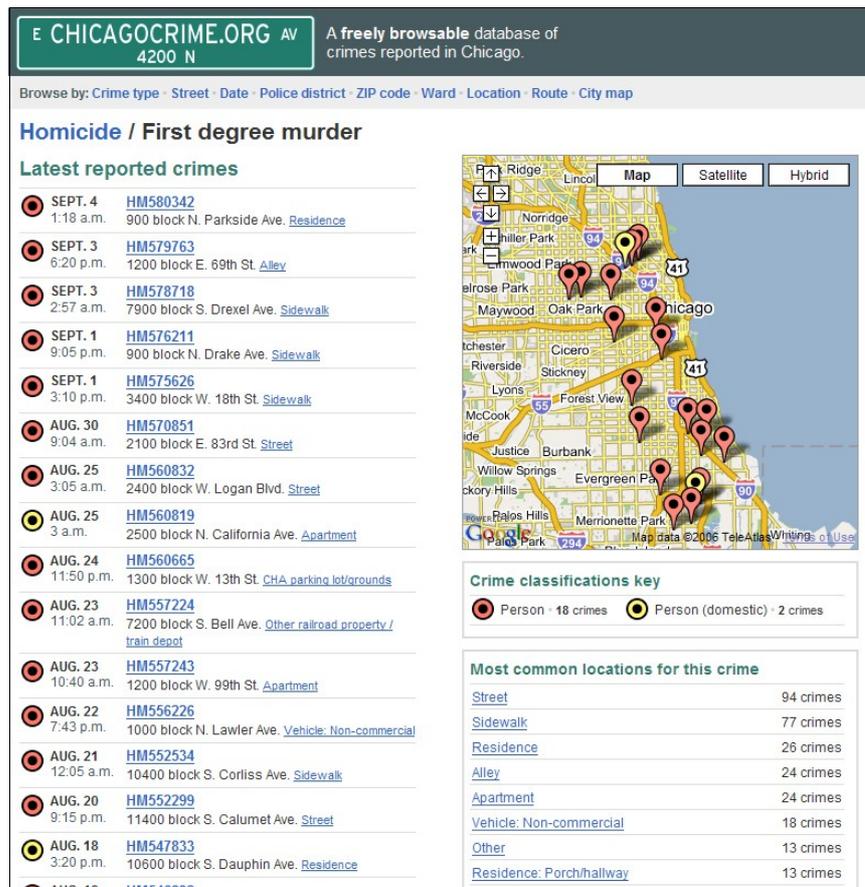


Figura 3.2: Exemplo de uso de APIs em uma Aplicação Web Híbrida.

### 3.2.2 Governo eletrônico

Fontes oficiais de informação fornecem matéria-prima valiosa para o trabalho jornalístico. A burocracia para conseguir alguma destas informações, o subterfúgio do segredo sob o pretexto de “segurança nacional” e outros empecilhos historicamente dificultaram a garimpagem de dados nesta área. A transparência do governo é turvada por obstáculos à divulgação pela imprensa do que acontece nas entranhas do Estado.

A tecnologia, hoje, chegou a um ponto em que a transparência total é possível e viável. No entanto, embora o acesso a dados do governo e de órgãos oficiais tenha sido facilitado em decorrência dos avanços tecnológicos, a adoção de novos métodos de divulgação de informação segue em passos lentos. As fontes oficiais, e entre elas o governo, são consideradas no meio jornalístico como fontes primárias e relativamente confiáveis de informação. (LAGE, 2001)

Fontes oficiais são mantidas pelo Estado; por instituições que preservam algum poder de Estado, como as juntas comerciais e os cartórios de ofício; e por empresas e organizações, como sindicatos, associações, fundações, etc. (LAGE, 2001)

Na produção de reportagens, a confiança na fonte e a credibilidade dela fazem os jornalistas assumirem os dados fornecidos por elas como verdade:

(...) as fontes oficiais são tidas como as mais confiáveis e é comum não serem mencionadas: os dados que propõem são tomados como verdadeiros. Assim, acontece de citarmos a população de uma cidade brasileira sem mencionar que ela foi estimada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) ou divulgarmos o Produto Interno Bruto (PIB) sem nos referirmos à repartição que o calculou. (LAGE, 2001)

No entanto, Lage condena este comportamento:

Trata-se de um mau hábito (...). Devem-se citar, sempre que possível, as fontes, sobretudo de dado numéricos, e questionar informantes sobre a origem dos números que citam. (LAGE, 2001)

A informatização em todas as esferas de governo vem acontecendo e a transparência aumenta. Para que essa atualização dos sistemas aconteça de forma ordenada, foi criado o e-PING, documento produzido pelo governo federal, que especifica Padrões de Interoperabilidade de Governo Eletrônico.

Este documento se refere, entre outras coisas, à diversidade de sistemas presentes no governo e determina que a integração entre eles deve ser feita por meio de Web Services. No capítulo anterior foi visto que Web Services formam uma camada de abstração, que serve para que diferentes sistemas possam ser integrados. São uma maneira de se utilizar a funcionalidade de um sistema estando fora dele, como se houvesse um “balcão de serviços” que faz sistemas diferentes trocarem informações entre si.

Nesse contexto, considerando as tendências mundiais e a integração crescente de sistemas distintos, é possível criticar o documento do governo e defender que os *Web Services* também sejam estendidos para o público em geral, para que desenvolvedores fora do governo possam trabalhar com os dados e funcionalidades desses sistemas.

As informações colocadas à disposição dos programadores independentes seriam apenas as de domínio público, a que, por lei, a população deve ter acesso. As funcionalidades dos sistemas, no entanto, são pontos a ser discutidos caso a caso, cuidando-se que o fornecimento dos serviços não prejudique a qualidade do serviço (QoS) interno.

No documento da e-PING, algumas considerações sobre e-Gov são feitas:

- Governo eletrônico significa, essencialmente, o governo servir melhor às necessidades do cidadão utilizando os recursos de Tecnologia, Informação e Comunicação. A arquitetura e-PING possibilita a integração e torna disponíveis serviços de forma íntegra, segura e coerente, permitindo obter melhores níveis de eficiência no governo.

- O governo deve incentivar a sociedade a opinar, comentar, e contribuir com sugestões de inovações que possam ajudá-lo a melhorar o acesso à informação e a prestação de seus serviços.
- Todos os processos de divulgação e de inter-relacionamento da e-PING prevêm a participação ativa do cidadão e da sociedade em geral, no processo de construção e gestão da arquitetura. (CEGE, 2005)

Para que esta visão se concretize plenamente, é essencial que haja algum tipo de interface de programação ao alcance da sociedade.

Projetos de código aberto no mundo todo provam que a capacidade do desenvolvedor independente é muito grande. Ao se prover o cidadão com informações confiáveis e Web Services de qualidade, emerge uma rede de aplicações que podem servir de base para melhorias no próprio governo.

A priorização do navegador como interface para divulgação de informações e serviços ao cidadão foi o caminho escolhido pelo governo (CEGE, 2005). Sem dúvida, essa é uma estratégia que avança na transparência e melhora o poder de vigilância do cidadão, além de prover serviços básicos para a sociedade.

O documento explicita o uso de *Web Services* para garantir a interoperabilidade interna dos sistemas. Do ponto de vista tecnológico, há neste quadro um potencial de avanço não contemplado, que é colocar à disposição da sociedade, por meio de *Web Services*, informações de domínio público das bases de dados do governo; do ponto de vista do governo, haveria economia de recursos e maior controle sobre a liberação dos dados.

Outro benefício trazido por uma API, é a possibilidade de se cadastrar os desenvolvedores e dar a cada um deles uma chave de acesso única, como é feito pelo Google ou o Yahoo!. Isso permitiria controlar o que cada um está fazendo e bloquear o acesso de quem não cumpre os termos de uso.

Para jornalistas e desenvolvedores independentes seria vantajoso o acesso a uma informação de qualidade, confiável e atualizada, de forma relativamente simples. Com isso, podem ser criadas aplicações sofisticadas e de baixo custo. O próprio governo, ao permitir que terceiros criem aplicações com seus dados, pode receber a ajuda da comunidade de desenvolvedores em seus projetos. A sociedade se beneficiaria com a maior transparência na administração pública.

## CAPÍTULO 4 – MODELO PROPOSTO

### 4.1 Apresentação do Modelo

Tendo como base toda a argumentação desenvolvida até aqui, chegamos a um modelo de Aplicação Web Híbrida voltada para o jornalismo. Ele leva em consideração os princípios da relevância individualizada de conteúdo, do jornalismo de precisão, da cognição humana, do novo panorama cultural e econômico mundial, além de contemplar as possibilidades abertas pela busca da web semântica, pelas características que deram origem à Web 2.0 e pelas habilidades dos novos profissionais da informação em rede.

Procura fugir do modelo de jornalismo online centrado na matéria e utiliza tecnologias como APIs e *Web Services* para apresentar uma alternativa complementar – e não excludente – para organização e exibição de conteúdo estruturado.

Este modelo foi elaborado em parte a partir da observação de aplicações em funcionamento na internet e a partir do arcabouço teórico explicitado no decorrer da pesquisa.

Existem, neste modelo, participantes que podem ser classificados por sua relação com a informação da seguinte forma:

#### a) Fornecedor

Assume o papel de fornecedor de informação aquele que se responsabiliza por:

- Processar dados. Ex.: Serviço de conversão automática de vídeos de um formato para outro.
- Armazenar dados. Ex.: Serviço de hospedagem de fotos.
- Produzir conteúdo. Ex.: Jornais online que produzem textos noticiosos ou institutos de pesquisa que produzem dados estruturados.

#### b) Organizador

A Aplicação Web Híbrida exerce prioritariamente este papel. Ela se responsabiliza por:

- Conjuguar dados de diferentes fornecedores. Ex.: Agregador de notícias RSS.
- Permitir a consulta. Ex.: Aproximação e detalhamento em aplicações com mapas.
- Contextualizar e selecionar informação. Ex.: Áreas de “conteúdo recomendado” em *sites* de compartilhamento de vídeo.

- Gerenciar o retorno dado pelos consumidores. Ex.: Organização das tags utilizadas pelos usuários.

### c) Consumidor

O consumidor da informação é responsável por:

- Consultar a informação. Ex.: Procurar alguma mercadoria de seu interesse em *sites* de compras.
- Retorno (*feedback*) ativo ou involuntário. Ex.: Gerar estatísticas de acesso em determinadas áreas (involuntário) ou dar uma nota para um determinado item (voluntário).
- Distribuição. Ex.: Recomendar para colegas um item que julgou interessante.

Estes papéis podem se sobrepor, dependendo da ação exercida no momento. Um consumidor pode exercer também o papel de produtor de conteúdo dentro da aplicação. No YouTube, por exemplo, quem produz os vídeos são os próprios consumidores. Cria-se um círculo onde, dependendo da circunstância, a pessoa assume um ou outro papel.

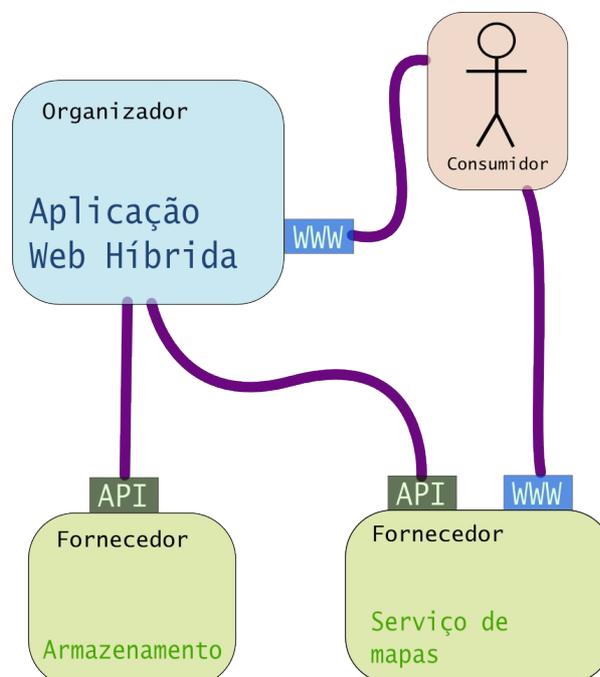


Figura 4.1: Modelo genérico de Aplicação Web Híbrida

## 4.2 Descrição das Etapas do Modelo

As etapas necessárias para a construção de uma Aplicação Web Híbrida que tire proveito de diferentes fontes de dados são: a extração de dados dos fornecedores, o cruzamento e organização da informação e a apresentação da informação de maneira interativa para o usuário.

### 4.2.1 Extração de dados dos fornecedores

Nessa etapa, deve-se avaliar como os fornecedores disponibilizam os dados. É possível que a mesma empresa exerça o papel de fornecedor e organizador simultaneamente, caso utilize dados próprios em sua aplicação.

O IBGE, por exemplo, não libera seus dados por meio de *Web Services*, mas é fornecedor na medida em que produz os dados de pesquisas e os disponibiliza em formato *Excel*. A aplicação que utilizar esses dados terá que ser responsável pelo seu armazenamento. O entendimento máquina-a-máquina traria a vantagem de acelerar o desenvolvimento e eliminar toda a etapa de conversão e armazenamento de um banco de dados que pode ser proibitivamente complexo.

É essencial observar os termos de uso de APIs que forem utilizadas. Muitas limitam o número de requisições que se pode fazer em um dia, outras permitem apenas o uso para fins não lucrativos. Um procedimento relativamente simples que alivia a carga sobre os fornecedores é a utilização de *cache*, ou seja, de uma cópia local do dado que foi trazido do fornecedor. Assim não é necessário buscar na fonte o mesmo dado em cada acesso, arriscando a sobrecarga, mas apenas periodicamente para checar se alguma alteração ocorreu.

A produção de conteúdo também é feita nessa etapa. Seja no ambiente da própria aplicação ou em sistemas externos, é importante que os dados sejam armazenados de forma estruturada. Ex.: Datas, nomes, empresas, resultados de jogos, valores.

Adrian Holovaty (apud BARBOSA, 2007, p. 174) chama esta etapa de obtenção/coleta da informação (*gathering*).

### 4.2.2 Cruzamento e organização

Um *site* já citado é o Chicago Crime, de Adrian Holovaty. Ele é uma aplicação que organiza dados coletados dinamicamente da polícia de Chicago sobre mapas gerados na hora via API do Google Maps. O usuário pode filtrar as informações de diversas formas, tanto pelo

tipo de ocorrência, gravidade do crime ou frequência com que ele ocorre. Pode também exibir apenas o que se relaciona à parte da cidade que seja do seu interesse.

A interface dessa aplicação, assim como suas funcionalidades e a sua arquitetura são essenciais para o seu sucesso. Não basta reproduzir um sistema de consulta a banco de dados, é preciso criar a ferramenta com um olhar jornalístico, potencializando a emergência de relevância para qualquer tipo de usuário.

A filtragem de informação e a busca são ferramentas essenciais neste tipo de aplicação. O sistema deve ser capaz de reordenar e reagrupar os dados das maneiras mais diversificadas possíveis para atender às necessidades do maior número de pessoas.

Holovaty (apud BARBOSA, 2007, p. 175) se refere a esta etapa como a de edição e tratamento da informação (*distilling*).

#### 4.2.3 Apresentação e interação

A interface deve ser ágil e intuitiva. Para isso, é possível aproveitar a tecnologia AJAX (*Asynchronous Javascript and XMLHttpRequest*). Ela provém a funcionalidade para a criação de aplicações de tela única, que não precisam recarregar toda a tela a cada operação. Desta maneira é possível atualizar uma área específica da tela com conteúdo dinâmico do servidor, acelerando a navegação e facilitando a interação.

Sugestões de consultas podem ser feitas baseadas no comportamento do usuário e suas preferências. As recomendações podem ser feitas baseando-se no comportamento de usuários similares ou nas declarações explícitas do usuário ao selecionar assuntos de interesse, utilizar *tags* ou dar nota para algum conteúdo.

Apresentação é o nome também usado por Holovaty (apud BARBOSA, 2007, p. 175) para se referir à esta etapa. Para ele, é a mais complexa das três, pois envolve o desenvolvimento de uma interface que preveja todo tipo de cenário que o usuário da aplicação pode encontrar. Ele ressalta que a dificuldade de se executar uma das etapas pode resultar em aplicações que automatizem apenas uma parte do processo, o que não inviabiliza o projeto.

## CAPÍTULO 5 – CONCLUSÕES

### 5.1 Conclusões Gerais

Diante do exposto, pode-se concluir que há vantagens para todos os envolvidos na utilização da Aplicação Web Híbrida.

Para os jornalistas e programadores que as desenvolvem, abrem possibilidades de inovação tanto no conteúdo quanto na interface e na relação com os outros agentes do processo. As APIs e Web Services permitem o acesso a uma informação de qualidade, confiável e atualizada, de forma relativamente simples. Com isso, podem ser criadas aplicações sofisticadas e de baixo custo.

Para os fornecedores de informação, melhora o controle sobre a divulgação e multiplicam-se as formas de levar as informações até a população.

Para os usuários das aplicações, a principal vantagem é poder interagir e decidir que parte do conteúdo da aplicação lhes interessa. Neste modelo, que mais parece uma ferramenta de consulta de dados, o usuário tem a oportunidade de escolher que informação quer ver e de que maneira ela deve ser apresentada. Além disso, em alguns casos pode colaborar com conteúdo para a aplicação.

Fontes de dados distintas são essenciais no objetivo de se criar conteúdo jornalístico em rede. Por tudo isso, conclui-se que no modelo de Aplicação Web Híbrida é possível e viável cruzar dinamicamente dados e informações de diversas origens de modo a transformá-los em conteúdo jornalístico relevante.

Durante este trabalho descrevemos o estado da arte das tecnologias envolvidas, mostrando como funcionam as APIs e Web Services, e como uma web semântica intermediária vem emergindo.

Por fim, ao propor um modelo de Aplicação Web Híbrida jornalística, concluímos a descrição de como se pode tirar proveito dos recursos da internet para produzir conteúdo jornalístico desta forma.

Um aprendizado essencial proporcionado por esta pesquisa foi o da importância da interdisciplinaridade no desenvolvimento de aplicações para a internet. Foi necessário transitar pela computação, comunicação, jornalismo e até fazer incursões pela estatística, neurofisiologia, psicologia, design e economia.

Percebe-se que há muitas intersecções entre as áreas de conhecimento que podem fomentar a inovação umas nas outras. A ideia de se armazenar conteúdo jornalístico na forma de dados estruturados é um exemplo disso.

## 5.2 Sugestões para Trabalhos Futuros

- APIs como estratégias de disseminação de informação. Como instituições podem abrir seus dados para o público? É importante conscientizar os desenvolvedores de organizações governamentais ou não de como é possível liberar alguns de seus dados para consumo externo e de como isso pode trazer benefícios para a própria organização. Mostrar como a filosofia do software livre e de código aberto pode incentivar a emergência de aplicações que utilizam estes dados.
- Formas de representação visual de dados estruturados (gráficos, tabelas, mapas...). Um desafio que se coloca para os profissionais que desenham interfaces é o de lidar com conteúdo sempre em transformação. Um trabalho que identificasse técnicas e apontasse caminhos para o design dinâmico e emergente seria de grande valia. As áreas de conhecimento que se cruzam neste tópico incluem design, geoprocessamento, estatística, matemática, jornalismo e muitas outras.
- A evolução do conceito de jornalismo e da definição de jornalista. Estudos teóricos sobre o que é jornalismo nesse novo contexto e se o que se produz com tecnologias de ponta – as Aplicações Web Híbridas apresentadas aqui, por exemplo – pode ser classificado como jornalismo.
- Estratégias para o armazenamento de conteúdo jornalístico na forma de dados estruturados. Como lidar com as redações que não reaproveitam o conteúdo, que não armazenam metadados sobre suas notícias e reportagens? Como extrair sentido dos textos já escritos? Trabalhos interdisciplinares que unam processamento de linguagem natural com conteúdo jornalístico podem ajudar bastante nestas tarefas.

- Participação dos usuários na construção de aplicações jornalísticas. Quais os resultados das experiências com a participação do leitor? Uma comparação entre iniciativas deste tipo pode dar informações essenciais para futuros projetos. Qual era o resultado esperado? Qual foi o resultado obtido? Que problemas surgiram no caminho?

## REFERÊNCIAS

- BALDESSAR, Maria José. **A mudança anunciada:** o cotidiano dos jornalistas com o computador na redação. Florianópolis: Insular, 2003. 104 p.
- BARBOSA, Suzana. **Jornalismo Digital em Base de Dados (JDBD):** Um paradigma para produtos jornalísticos digitais dinâmicos. (Tese de Doutorado). FACOM/UFBA, Salvador 2007.
- BENKLER, Yochai. ***The Wealth of Networks: how social production transforms markets and freedom.*** Newhaven e Londres: Yale University Press, 2006. 515 p.
- BERNERS-LEE, Tim; HENDLER, J.; LASSILA, O.. ***The Semantic Web.*** Scientific American, Nova York, n., p.34-43, 01 maio 2001.
- BERNERS-LEE, Tim; SHADBOLT, Nigel; HALL, Wendy. ***Semantic Web Revisited.*** Ieee Intelligent Systems, Nova York, n., p.96-101, 01 maio 2006. Disponível em: <<http://www.computer.org/intelligent>>. Acesso em: 01 maio 2007.
- CEGE – Comitê Executivo de Governo Eletrônico. Governo Brasileiro. **E-PING:** Padrões de Interoperabilidade de Governo Eletrônico. Versão 1.5 Brasília, 2005. 66 p.
- BUSH, Vannevar. ***As We May Think.*** Atlantic Monthly, Nova York, v. 176, n. 1, p.101-108, 01 ago. 1945.
- CASTILHO, Carlos Albano Volkmer de. **Jornalismo Cidadão:** A relação entre amadores e profissionais no jornalismo via internet. 2006. 36 f. Tcc (Graduação) – Curso de Comunicação Social – Mídia Eletrônica, Faculdades Assesc, Florianópolis, 2006.
- GAMPER, Johann; AUGSTEN, Nikolaus. The Role of Web Services in Digital Government. ***Lecture Notes In Computer Science: Electronic Government,*** Berlim, p. 161-166. 24 jan. 2004. Disponível em: <<http://www.springerlink.com/content/330vdcth6vyh5mm0/>>. Acesso em: 17 ago. 2006.
- GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social.** 5. ed. São Paulo: Atlas, 1999.
- GLADWELL, Malcolm. ***Blink: The power of thinking without thinking.*** Nova York: Little, Brown, 2005. 277 p.
- HOFSTADTER, Douglas R.. **Gödel, Escher, Bach:** Um entrelaçamento de gênios brilhantes. Brasília: Universidade de Brasília, 2001. 892 p.

HOLOVATY, Adrian. *A fundamental way newspaper sites need to change*. Disponível em: <<http://www.holovaty.com/blog/archive/2006/09/06/0307>>. Acesso em: 06 jul. 2006.

JOHNSON, Steven. *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. Nova York: Scribner, 2001. 288 p.

LAGE, Nilson. **Teoria e técnica do texto jornalístico**. Rio de Janeiro: Elsevier, 2005. 188 p.

LIMA, Lara Viviane Silva de. **Jornalismo de precisão e jornalismo científico: Estudo da aplicabilidade**. 2000. 123 f. Dissertação (Mestrado) – Curso de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2000.

MEYER, Philip. *The new precision journalism*. Disponível em: <<http://www.unc.edu/~pmeyer/book/>>. Acesso em: 07 ago. 2006.

MIELNICZUK, Luciana. **Jornalismo na Web: uma contribuição para o estudo do formato da notícia na escrita hipertextual**. (Tese de Doutorado). FACOM/UFBA, Salvador, 2003.

PINKER, Steven. **Como a mente funciona**. 2. ed. São Paulo: Companhia Das Letras, 1998. 666 p.

RIBAS, Beatriz. **A Narrativa Webjornalística – um estudo sobre modelos de composição no ciberespaço**. (Dissertação de Mestrado). FACOM/UFBA, Salvador 2005.

SHIRKY, Clay. *The Semantic Web, Syllogism, and Worldview*. Disponível em: <[http://www.shirky.com/writings/semantic\\_syllogism.html](http://www.shirky.com/writings/semantic_syllogism.html)>. Acesso em: 06 jul. 2006.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação**. Florianópolis: Laboratório de Ensino A Distância da Ufsc, 2000. 118 p.

SOUZA, Renato; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, Df, v. 1, n. 33, p.1-10, 18 jun. 2004. Disponível em: <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=71>>. Acesso em: 05 maio 2007.

SPERBER, Dan; WILSON, Deirdre. *Relevance: Communication and Cognition*. 2ª Oxford, Inglaterra: Blackwell, 1995.

SUROWIECKI, James. *The Wisdom of Crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. Nova York: Doubleday, 2004. 297 p.

TUFTE, Edward. *The Visual Display of Quantitative Information*. 2. ed. Cheshire, Connecticut, Eua: Graphic Press, 2004. 197 p.

WAL, Thomas Vander. *Feed on this*. Disponível em:  
<<http://www.vanderwal.net/random/entrysel.php?blog=1562>>. Acesso em: 05 nov. 2004.

## BIBLIOGRAFIA

ALVES, J. B. M. **Teoria Geral de Sistemas**, 2004. Disponível em: <http://www.inf.ufsc.br/~jbosco/Egc.html>

BERTALANFFY, Ludwig v. **Teoria Geral dos Sistemas**. 3ª edição, ed. Vozes, Petrópolis, 1977.

BRIN S., Page L. **The anatomy of a large-scale hypertextual Web search engine**. Computer Networks and ISDN Systems, Vol. 30, No. 1—7, 1998. pp. 107-117

CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 28., 2005, Rio de Janeiro. **A qualidade de informação em webjornais – conceitos norteadores**. Rio de Janeiro: Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação, 2005. p. Disponível em: <<http://repositorio.portcom.intercom.org.br/bitstream/1904/16804/1/R1699-1.pdf>>. Acesso em: 25 jul. 2006.

CROCOMO, Fernando Antonio. **O USO DA EDIÇÃO NÃO-LINEAR DIGITAL: AS NOVAS ROTINAS NO TELEJORNALISMO E A DEMOCRATIZAÇÃO DE ACESSO À PRODUÇÃO DE VÍDEO**. 2001. 107 f. Dissertação (Mestrado) – Curso de Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2001. Disponível em: <<http://teses.eps.ufsc.br/defesa/pdf/6773.pdf>>. Acesso em: 25 jul. 2006.

DAWKINS, R. **The Selfish Gene**. Oxford University Press, 1976. 352pp.

FRASER, J. **It's a Whole New Internet**, 2005. Disponível em: <<http://www.adaptivepath.com/publications/essays/archives/000430.php>> . Acesso em: nov. 2005.

GIBSON, B. **IBM's Intranet and Folksonomy**. Disponível em: <[http://thecommunityengine.com/home/archives/2005/03/ibms\\_intranet\\_a.html](http://thecommunityengine.com/home/archives/2005/03/ibms_intranet_a.html)> . Acesso em: dez. 2005.

HAMMOND, T., HANNAY, T., LUND, B, SCOTT, J. **Social Bookmarking Tools (I): A General Review**. In D-Lib Magazine, Vol. 11, No. 4, 2005. Disponível em: <<http://www.dlib.org/dlib/april05/hammond/04hammond.html>>. Acesso em: nov. 2005.

MAZZOCCHI, S. **Folksologies: de-idealizing ontologies**, 2005. Disponível em: <<http://www.betaversion.org/~stefano/linotype/news/85>>. Acesso em: fev. 2006.

ORDMAN, N. *Et alii*. **Traversing the Corporate Web – IBM's Information Management Workflow**. Information Architecture Summit, Canadá, 2005. Disponível em: <<http://subwaylove.com/IA2005.ppt>>. Acesso em: jan. 2006.

O'REILLY, T., **The Architecture of Participation**. In O'Reilly Developer Weblogs, 2003. Disponível em: <<http://www.oreillynet.com/pub/wlg/3017>>. Acesso em: mar. 2006.

PFAFF, Donovan; SIMON, Bernd. New Services through Integrated e-Government. **Lecture Notes In Computer Science: Electronic Government: First International Conference, EGOV 2002, Aix-en-Provence, France, September 2-5, 2002. Proceedings**, Aix-en-provence, p. 391-394. 01 ago. 2003. Disponível em: <<http://www.springerlink.com/content/p518k6l2aq8qyjp/>>. Acesso em: 17 ago. 2006.

UDELL, J. *Collaborative knowledge gardening*, 2004Disponível em: <[http://www.infoworld.com/article/04/08/20/34OPstrategic\\_1.html](http://www.infoworld.com/article/04/08/20/34OPstrategic_1.html)>. Acesso em: ago 2006.

WAL, Thomas V. *Explaining and Showing Broad and Narrow Folksonomies*, 2005. Disponível em: <[http://www.personalinfocloud.com/2005/02/explaining\\_and\\_.html](http://www.personalinfocloud.com/2005/02/explaining_and_.html)>. Acesso em: ago. 2006.

WILSON, Deirdre. *Relevance and Relevance Theory*. In: WILSON, Robert A.; KEIL, Frank C. (Comp.). *The MIT Encyclopedia of the Cognitive Sciences*. [s. L.]: Mit Press, 1999. p. 719-720.