UNIVERSIDADE FEDERAL DO ESPÍRITO

DEPARTAMENTO DE DESENHO INDUSTRIAL

RÔMULO GUIMARÃES DE OLIVEIRA

**UMA NOVA FERRAMENTA DE AGREGAÇÃO E RECOMENDAÇÃO DE CONTEÚDO BASEADO EM FILTRAGEM COLABORATIVA**

Vitória

2012

RÔMULO GUIMARÃES DE OLIVEIRA

**Uma nova ferramenta de agregação e recomendação de conteúdo baseado em filtragem colaborativa**

Monografia apresentada ao Programa de Graduação em Desenho Industrial (Programação Visual) do Centro de Artes da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Desenhista Industrial, sob a orientação do Profº Ms. **Hugo Cristo**.

Vitória

2010

**Uma nova ferramenta de agregação e recomendação de conteúdo baseado em filtragem colaborativa**

Rômulo Guimarães de Oliveira

**Resumo:** Este projeto tem como objetivo desenvolver uma ferramenta de agregação de conteúdo advindo de sites e blogs de notícias, e através da avaliação dos usuários gerar recomendações personalizadas tendo em vista o conceito de classificação colaborativa. A ferramenta utiliza um algoritmo baseado no coeficiente de pearson para calcular as similaridades entre os usuários através das avaliações em comum, retornando recomendações de notícias ainda não avaliadas pelo usuário em questão. Para o projeto, foi feita uma pesquisa com blogueiros e leitores de notícias oriundas de blogs e sites de notícias, com analise qualitativa dos dados para obtenção de expectativas relacionadas com a proposta de uma nova ferramenta que atingisse o perfil do entrevistado. Em um segundo momento, foi feita uma análise de sites que realizam a recomendação de conteúdo, com o intuito de reunir características em comum entre eles, e apontar deficiências e ausências de funcionalidades, para então poder propor uma nova ferramenta capaz de oferecer algo que contemple as possíveis ausências e deficiências.  
  
**Palavras-chave:** Filtragem colaborativa. recomendação. similaridade.

**SUMÁRIO**

**INTRODUÇÃO............................................................................................................5**

**1 - WEB 2.0.................................................................................................................6**

1.1 - CONCEITO E PROPRIEDADES........................................................................6

1.2 - O TERMO WEB 2.0............................................................................................6

1.3 - PRÓXIMO INTEL INSIDE: A IMPORTÂNCIA DOS DADOS.............................11

1.4 - INTELIGÊNCIA COLETIVA...............................................................................12

1.4.1 - O QUE É INTELIGÊNCIA COLETIVA............................................................13

1.4.2 - TIRANDO PARTIDO DA INTELIGÊNCIA COLETIVA....................................14

1.5 - BLOGS E MICROBLOGS..................................................................................15

1.5.1 - A ASCENSÃO DOS BLOGS...........................................................................16

1.5.2 - DOS SITES PESSOAIS AOS BLOGS............................................................17

1.5.3 - CARACTERÍSTICAS DOS BLOGS................................................................18

1.5.4 – MICROBLOG..................................................................................................19

1.5.5 - A SABEDORIA DAS MASSAS........................................................................21

**2 - CLASSIFICAÇÃO COLABORATIVA...................................................................22**

2.1 – FOLKSONOMIA................................................................................................23

2.1.1 - FOLKSONOMIA EM BLOGS E MICROBLOGS.............................................23

**3 - O PROJETO DE UM SISTEMA DE PUBLICAÇÃO DE BLOGS COM CLASSIFICAÇÃO COLABORATIVA........................................................................26**

**4 – OBJETIVO...........................................................................................................27**

4.1 - OBJETIVO GERAL............................................................................................27

4.2 - OBJETIVOS ESPECÍFICOS..............................................................................27

**5 – JUSTIFICATIVAS................................................................................................27**

**6 – METODOLOGIAS................................................................................................29**

6.1 - METODOLOGIA DE PESQUISA.......................................................................29

6.2 - METODOLOGIA DE PROJETO.........................................................................30

6.3 - APLICANDO AS METODOLOGIAS...................................................................33

**7 - PESQUISA INICIAL..............................................................................................33**

7.1 - RESULTADOS INICIAIS....................................................................................34

**8 - DESENVOLVIMENTO DA FERRAMENTA..........................................................35**

8.1 – ESTRATÉGIA....................................................................................................36

8.2 – ESCOPO...........................................................................................................37

**9 – CRONOGRAMA...................................................................................................38**

**REFERÊNCIAS BIBLIOGRÁFICAS..........................................................................39**

**1 - INTRODUÇÃO**

Com a popularização da internet, pudemos perceber a utilização de um novo espaço de produção da informação, e por consequência a sua disseminação em rede. O enorme fluxo dessa informação transforma a internet em um ambiente rico em quantidade de conteúdo.

Uma ferramenta muito usada nos dias de hoje, para publicação de qualquer tipo de conteúdo, são os blogs. Elas oferecem para qualquer usuário que tenha acesso a internet, a possibilidade de escrever e tornar público o que lhe interessa. Essa possibilidade, a priori sem fronteiras, injeta uma quantidade crescente de informação na rede, e quanto mais o ciberespaço se amplia, mais ele se torna “universal” (LÉVY, 1999, p. 111), e essa universalidade aceita todos, contentando-se apenas em tornar possível o contato de determinada informação com qualquer usuário da rede.

Segundo o Technorati, famoso indexador de diários virtuais, o número de blogs ativos no mundo em 2007 chegava a quase um bilhão (SCHNOOR, 2007), com um aumento cada vez mais crescente, já que na mesma época, a proporção média diária de novos blogs era de 175 mil.

À medida que a crescente comunidade de usuários de blogs publica os mais diversos conteúdos, os portais de notícias abordam novas estratégias para acompanhar essa crescente disseminação de informação, a ponto desses dois meios, em muitos momentos, equivalerem-se na forma de publicação dos mais variados tipos de conteúdos.

De outro lado, surgem ferramentas que prometem aos consumidores desta imensidão de informação, formas particulares de se filtrar o conteúdo, recebendo apenas aquilo que mais lhe interessa, baseado em suas preferências.

É deste momento em particular que este projeto propõe uma nova maneira tanto de filtrar o conteúdo das mais diversas fontes de notícias, sejam elas blogs ou portais de notícias, como de promover a recomendação através da simples classificação colaborativa entre os próprios usuários desta nova ferramenta proposta.

Essa ferramenta se faz necessária em um cenário aonde a informação chega com maior velocidade a todos os usuários da web, e a filtragem com base nas preferências se torna crucial para se absorver apenas os conteúdos que interessam a determinado usuário. No mesmo cenário, oferecer a possibilidade de receber recomendações levando em consideração a classificação colaborativa é de fato enriquecedor, e unir essas duas funcionalidades centradas nos usuários é possibilitar uma nova experiência na forma de consumir informação na internet.

Este projeto tem como objetivo geral desenvolver uma ferramenta web de agregação de notícias de qualquer site que disponibiliza seu conteúdo no formato de feeds, e através da avaliação destas informações por parte dos usuários, recomendar novos conteúdos agregados pelos próprios usuários baseado nas similaridades entre eles.

São componentes específicos dos objetivos deste projeto:

* Desenvolver uma ferramenta web, hospedada através de um site, onde permita o cadastro público de usuários;
* Permitir o cadastro de fontes de notícias no formato de feeds;
* Possibilitar a leitura nas notícias oriundas das fontes previamente cadastradas;
* Permitir a avaliação das notícias listadas pelas fontes cadastradas;
* Oferecer recomendações aos usuários baseado no cruzamento das avaliações dos usuários, utilizando o coeficiente de correlação de Pearson.

Uma ferramenta desse caráter, quem tem como ponto forte a classificação colaborativa, pode trazer grandes benefícios para a grande massa de usuários que consomem conteúdo na internet.

Num primeiro momento, filtrar o conteúdo de seu interesse nem sempre é conseguido ao navegar por inúmeros sites, e permitir cadastrar fontes de notícias em um único espaço permite absorver apenas o que é de sua preferência. Essa agregação de conteúdo só é permitida com facilidade em um momento onde novas tecnologias como a disponibilização de feeds são oferecidas publicamente. E é tomando partido desta funcionalidade de inúmeros sites que este projeto se faz possível.

Ainda hoje navegamos incansavelmente pelos sites que consideramos mais relevantes, e a grande quantidade de informação disponibilizada torna difícil o seu acesso, e muito do que é de nosso interesse particular é perdido, visto que nossa atenção pode estar voltada para outras tarefas cotidianas.

Vários serviços na internet oferecem a agregação de conteúdo através de feeds de forma gratuita, e isso nos possibilita potencializar o acesso àquilo que é de nosso interesse. Porém, muitas vezes, estas ferramentas focam apenas nesta funcionalidade, apenas agregando o conteúdo, não se preocupando em oferecer novas possibilidades de descoberta, e também não levando em consideração a forma como o usuário interage com este conteúdo.

A ferramenta que este projeto se preocupa em discutir e desenvolver, permite não apenas o simples agrupamento de inúmeras bases de informações, mas se preocupa imensamente na forma como os usuários interagem com o conteúdo previamente cadastrado por eles, tomando as suas avaliações como fonte de dados para uma nova experiência ao oferecer recomendações que levam em consideração o cruzamento destes dados, e que parte da ideia de similaridade entre os usuários para decidir o que pode ou não ser relevante.

Para a realização deste projeto, com o intuito de desenvolver uma ferramenta ao final dos estudos, foi proposta uma pesquisa qualitativa e outra quantitativa, em duas etapas:

* **1ª etapa: pesquisa qualitativa:** foi aplicado um questionário de 20 questões (já somado as sub-questões) em um público específico de 20 entrevistados. A pesquisa visa entender o comportamento de blogueiros e consumidores de notícias na internet em relação a sua forma de ler, publicar e participar colaborativamente dos processos pertinentes a esse campo.
* **2ª etapa: pesquisa quantitativa:** foi feita uma pesquisa analisando as características e funcionalidades de alguns sites específicos. A pesquisa tem como alvo XX sites, entre eles: blogs, portais de notícias, indexadores de links, agregadores de conteúdo e sites de recomendação de notícias. Esta pesquisa analisou inúmeras funcionalidades pertinentes a estes sites, e serviu para definir a arquitetura e design da informação, design da navegação e o design visual da ferramenta, bem como definir propostas melhoradas das funcionalidades já existentes e oferecer novas baseadas no escopo deste projeto.

Uma ferramenta que toma como base a Web 2.0 só se faz possível com os estudos que levem em consideração o usuário. O usuário de qualquer aplicação web é tido como o centro de maior importância, pois ele é tanto o detentor da informação quanto o principal agente formador de crítica, e também o principal interessado nas relações homem x máquina.

“A prática de criar experiências de usuário que são envolventes e eficientes é chamada de *User-centered design* (Design centrado no usuário)” (ANGULO, 2010). O conceito prevê que a cada etapa do projeto o usuário seja levado em consideração. Garrett aborda mais afundo esse assunto em *The Elements of User Experience* (2002), e propõe um diagrama que descreve as etapas de um projeto centrado no usuário. Esta é a metodologia de projeto utilizada para o desenvolvimento da ferramenta.

**2 – WEB 2.0**

**2.1 - CONCEITO E PROPRIEDADES**

Em março de 2000, um grande furor tomou conta da Nasdaq, bolsa de valores voltada para ações da área de tecnologia, e onde as empresas chamadas “ponto com” receberam grande valorização na época. Foi nesse momento que algo se tornou insustentável, provocando o tão famoso “estouro da bolha” (DISCOVERY BRASIL, 2012). O estouro foi uma queda vertiginosa das ações supervalorizadas das empresas de tecnologia que tinham seus serviços baseados na internet, e que chegavam a atingir mais que o dobro de seu capital. Com uma especulação totalmente diferenciada, achando encontrar nas empresas ponto com um modelo diferente de investimento, que não seguia a receita das empresas clássicas, o que se viu foi um momento marcado pelo falecimento de inúmeras empresas que não suportaram o estouro, e que não estavam preparadas para esse momento tão particular da história da economia.

O fato do estouro da bolha ter ocorrido não significou a queda dos modelos de negócio baseados na internet, e sim dos modelos de economia que se imaginava serem totalmente diferentes. Mesmo com inúmeras empresas ponto com fechando as portas em 2000 e nos anos que se seguiram, histórias de sucesso mostraram sua força e as diferenças que determinaram a quebra de um e o sucesso de outro. Vendo dessa forma, ao contrário do que se imaginava com o estouro, a internet ganhou cada vez mais importância, marcando uma virada que talvez desse sentido a uma convocação do tipo “Web 2.0” (O’REILLY, 2006).

O termo Web 2.0 surgiu numa conferência de *brainstorming* em outubro de 2004, entre a O’Reilly e a MediaLive Internacional. O conceito foi largamente discutido nesse momento, onde se reparou que as empresas sobreviventes do estouro da bolha tinham algo em comum, e que acabaram por determinar o seu sucesso ou pelo menos o seu crescimento regular.

Em “O que é Web 2.0”, Tim O’Reilly (2006, p. 2) formula a idéia do que seria a Web 2.0 em comparação com o modelo de web 1.0.

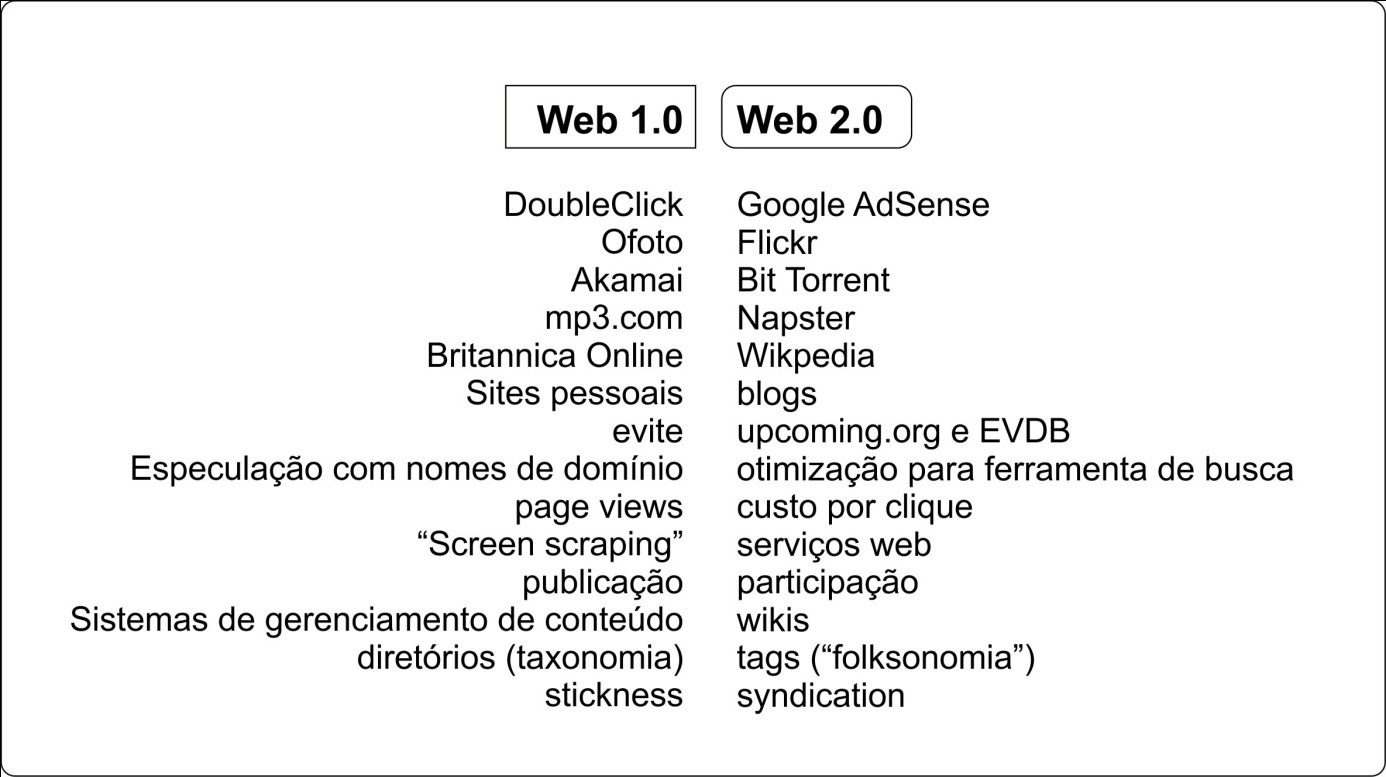


Figura 1 – Exemplo discutido na conferência Web 2.0 (O’REILLY, 2006, p. 2)

O uso do termo web 2.0, desde então, passou a ser citado de forma crescente na internet, e em muitos casos, apenas como uma forma de marketing das empresas que se diziam pertencentes a ele. Mas a raiz do conceito é muito mais que se apropriar de um modelo de web, é ter uma postura diferenciada de outros modelos, e em muitos casos, uma evolução.

Na figura um, são emparelhados alguns pontos que exemplificam a evolução dos modelos de sites e serviços. Muitos dos sites pertencentes à web 2.0 aplicam parte dos princípios pertencentes a ela, fazendo entender este conceito como um centro gravitacional, com práticas determinadas, onde algumas se aplicam a determinados sites, e outros não.

É entendido como noções centrais da Web 2.0, os seguintes pontos:

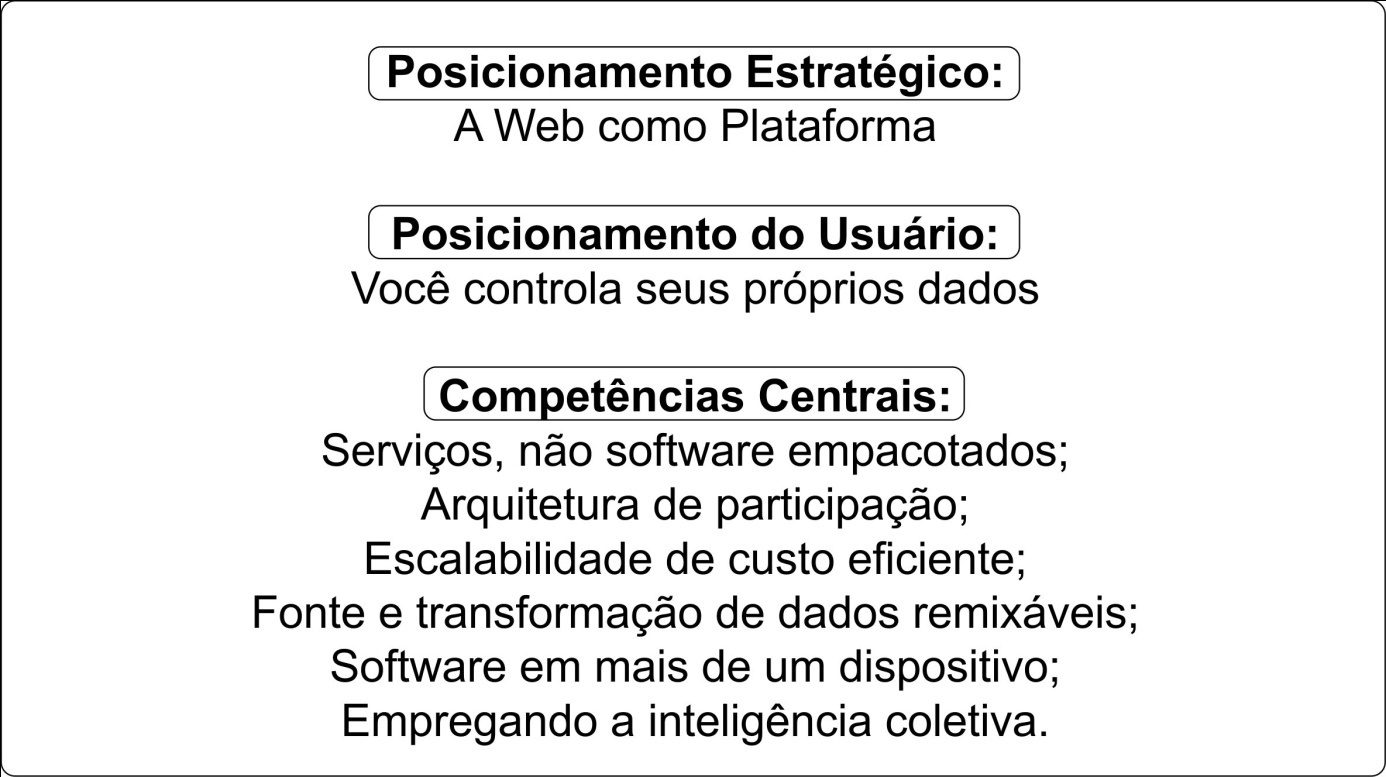


Figura 2 – Noções centrais da Web 2.0 (O’REILLY, 2006, p. 3)

A Figura 2 ilustra as ideias como centro da Web 2.0. A partir daí, vários sites se propõem a utilizar esses princípios, usando apenas parte deles.

Podemos citar como exemplos pioneiros da Web 2.0, e que utilizam parte destas ideias, os seguintes sites e serviços:

* **Flickr e Del.icio.us:** Utilização de *Tagging*, não taxonomia.
* **GMail, Google Maps e AJAX:** Experiência rica do usuário.
* **PageRank, Reputação do eBay e Avaliações da Amazon:** O usuário como colaborador.
* **Google AdSense:** Auto-serviço, possibilitando um novo modelo de negócio.
* **BitTorrent:** Descentralização radical.
* **Wikipedia:** Confiança radical.

Todos os exemplos acima se aproveitam do conceito de maior valor dado por O’Reilly, em sua conceituação da web 2.0:

Web 2.0 é a revolução de negócios na indústria de computadores causada pela mudança para uma internet como plataforma, e uma tentativa de entender as regras para o sucesso nessa nova plataforma. A regra principal é desenvolver aplicações que aproveitem os efeitos da rede para se tornarem melhores quanto mais são usados pelas pessoas, aproveitando-se da inteligência coletiva. (O’REILLY RADAR, 2006).

Seguindo estes conceitos, a web 2.0 vê a internet como plataforma, e os softwares, como serviços que rodam nela. Quanto mais os aplicativos são usados, mais é enriquecido, tornando o conteúdo construído através da inteligência coletiva.

Segundo Marcello Póvoa, escritor da Webinsider (PÓVOA, 2006), podemos listar uma síntese dos principais padrões que são considerados parte desse grupo de tendências da segunda geração da web.

* **A Web como plataforma:** os sites deixam de ser estáticos e passam a ser entendidos como softwares que rodam diretamente no navegador. Para isso, se utiliza de novas linguagens de programação e bibliotecas de recursos para tornar o conteúdo dinâmico, e não apenas texto simples inserido em páginas estáticas, com pouca interação do usuário. Na internet, passa a se oferecer serviços, diferentes de softwares “empacotados” (O’Reilly, 2006, p. 2). Tomamos como exemplo rápido o Google, que inicialmente ofereceu seus serviços diretamente na internet, sem a necessidade de pagamento para o seu uso e também não se via obrigado a fazer constantes lançamentos, apenas propondo melhorias em tempo real nas suas aplicações web.
* **Beta eterno**: tradicionalmente, softwares de *desktop* são empacotados em versões seqüenciais (versão 1.0, versão 2.0, e assim por diante), e vários sites também recebem esse tratamento. O conceito de beta eterno aborda o lançamento de versões de outro modo, já que a possibilidade de *feedback* dos usuários é mais constante e dinâmica. Os sites sempre necessitam de atualizações, muitas vezes pequenas, que envolvem partes específicas de um serviço que está alocado na rede. “O Beta Eterno vem do princípio de que o desenvolvimento de um site nunca termina. É uma evolução constante e sem fim” (SEIXAS, 2010).

Segundo Seixas (2010), o termo deve ser usado com cuidado, pois juntamente com a frase “evolução constante”, o recado do termo é claro: “Ao usar esse sistema, você poderá encontrar erros, *bugs* e instabilidade.” Sendo assim, a crítica comum no marketing ao se usar o termo vem justamente nesse sentido, propondo ao usuário usar um serviço que pode conter bugs e instabilidade, e já que se considera uma versão em beta eterno, os problemas também serão eternizados, propagando uma imagem negativa do site.

* **Redes Sociais:** o conceito sempre existiu, apesar do termo ter ganhado destaque mais recente. Mas com as evoluções tecnológicas que possibilitaram uma maior interação entre usuários, as redes sociais ganharam mais evidência, oferecendo um número sem fim de propostas que oferecem diversas formas de interatividade. Só no Brasil, segundo um infográfico da Agência Click, as redes sociais agregam mais de 55 milhões de usuários (CLICK, 2010).
* **Conteúdo flexível:** o conteúdo dos sites passa a ser muito mais dinâmico, utilizando banco de dados nos servidores web para guardar todos os tipos de informação. Com essa possibilidade, o conteúdo pode ser requisitado de acordo com o interesse do usuário, uma vez que ele se encontra em um banco próprio. Os usuários passam de simples consumidores para geradores de conteúdo, participando efetivamente da produção mundial de informação, e principalmente, passa a classificá-la, para posterior pesquisa. A internet leva a colaboração ao seu nível mais alto.
* **Tags:** essa talvez seja a mais importante tendência da segunda geração da internet. Baseado nos conceitos de folksonomia (ver tópico mais a frente), a classificação se dá de forma inteligente, levando em consideração termos comuns incidentes nos conteúdos publicados.

**2.2 – PRÓXIMO INTEL INSIDE: A IMPORTÂNCIA DOS DADOS**

A Wikipedia (2010) define banco de dados como sendo “[...] um conjunto de registros dispostos em estrutura regular que possibilita a reorganização dos mesmos e produção de informação”.

Todo aplicativo que oferece importantes serviços na internet dispõe de poderosos bancos de dados. Saber gerenciar estes dados é de extrema importância para o sucesso de serviços oferecidos por empresas como Google, Amazon.com, Yahoo, Wikipédia, entre outros. A Amazon.com, em particular, teve grande sucesso em criar um sistema de dados atrelados aos seus produtos, enriquecendo a experiência do usuário ao visitar o seu site, tendo disponível uma grande base de informação sobre publicações que não estão disponíveis em nenhum outro lugar.

Com a supervalorização das informações contidas nos bancos de dados, que são alimentados constantemente por inúmeros usuários todos os dias, se destacam empresas que conseguem utilizá-las da melhor forma possível. Saber captar os dados e transferi-los para uma interface interativa e intuitiva, levando em consideração a atualização dinâmica desses dados, é também transformar a informação em fonte de riqueza de um serviço.

A informação contida em sites passa de direitos reservados a “alguns direitos reservados”, dando ao usuário o direito de remixá-los (O’REILLY, 2006, p. 23). Por outro lado, é possível ver em diversos sistemas disponíveis na internet, como em maps.google.com, os avisos de *copyright* (direitos reservados) que apontam para os detentores dos direitos dos dados contidos nesta aplicação. No exemplo do GoogleMaps, os dados são licenciados por várias empresas, cabendo a Google comprar os direitos de utilizá-los, disponibilizando-os em aplicativos que oferecem diversas possibilidades de utilização, inclusive agregados a outros tipos de serviços com temáticas diferenciadas, que utilizam mapas como forma de sistematização. O termo Intel Inside deriva exatamente de um caso onde a empresa NavTeq imitou o logotipo “Intel Inside” em alguns de seus produtos. Isso mostra que os dados é que são a parte essencial desses aplicativos, que são de código fonte aberto e que possibilitam a sua distribuição.

Hoje os bancos de dados mais importantes, são, em grande parte, proprietários (de posse particular de uma empresa), mesmo abrindo um leque de possibilidades de utilização. Assim como ocorre com os softwares proprietários, que cada vez perde mais espaço para os softwares de código fonte aberto, é possível imaginar e até mesmo prever, que os dados proprietários abram espaço para uma espécie de “dados abertos”, sendo de posse indeterminada. Um ótimo exemplo dessa tendência é o Wikipédia, espécie de enciclopédia colaborativa *online* livre. Este exemplo defende a universalidade da informação, com os dados pertencendo a toda comunidade inserida na web.

**2.3 – INTELIGÊNCIA COLETIVA**

É sabido que na história do homem, a escassez de informação sempre esteve presente. Mesmo com o que sabemos dos tempos passados, através de livros e outras fontes de informação, não é suficiente pra contar nem se quer parte de tudo que já se produziu de conhecimento. Nos dias atuais, viajar num tapete mágico pela imensa biblioteca de Alexandria para consultar todos os livros da humanidade ou até mesmo o desejo de um financista visualizar, numa bola de cristal, tudo sobre as bolsas de valores de todo o mundo, depende apenas de alguns caracteres inseridos em um pequeno computador pessoal com acesso a internet (LEMOS; PALACIOS, 2001, p. 5). Ver uma exposição, se deslumbrar com imagens de alguma praia paradisíaca ou acompanhar ao vivo uma palestra do outro lado do mundo, é questão apenas de alguns códigos. A maior dificuldade acaba sendo a filtragem, pois com a imensa quantidade de informação que se faz disponível, o trabalho acaba sendo separar o bom do ruim.

A Internet é livre, qualquer um pode inserir na rede a informação que bem entender, mesmo que não seja legal ou ético. A produção do saber se torna coletivo e não apenas por aqueles que se dizem aptos a produzirem conhecimento.

Com a utilização das redes abertas na internet, é possível viabilizar um novo tipo de pensamento sustentado por conexões sociais. Esse é o enfoque da discussão promovida por Pierre Lévy, ao debater o conceito de Inteligência Coletiva.

**2.3.1 – O QUE É INTELIGÊNCIA COLETIVA?**

É uma inteligência distribuída por toda a parte, incessantemente valorizada, coordenada em tempo real, que resulta em mobilização efetiva das competências. Acrescentemos à nossa definição este complemento indispensável: a base e o objetivo da inteligência coletiva são o reconhecimento e o enriquecimento mútuo das pessoas, e não o culto de comunidades fetichizadas ou hipostasiadas (LÉVY, 1998, p. 28).

É a partir daí que podemos definir que o conhecimento passa a ser descentralizado, desviando o poder de um para a democratização da informação e valorização da participação de cada indivíduo, resultando num reconhecimento e enriquecimento cultural de todos.

Cavalcanti e Nepomuceno (2007, p. 39) identificam três formas de gerar Inteligência Coletiva: inteligência coletiva inconsciente, consciente e plena.

* **Inteligência Coletiva inconsciente:** é a contribuição feita pelo usuário sem se dar conta de que está fazendo, ao simplesmente visitar um site, digitar alguma informação em um formulário ou questionário, avaliar um produto em uma loja virtual, clicar em um determinado *link*. Esses registros são passíveis de armazenamento em um banco de dados por parte do servidor, e é possível que os proprietários de um desses sites visitados use as informações obtidas de alguma forma.
* **Inteligência Coletiva consciente:** acontece quando um usuário tem consciência de sua contribuição em algum tipo de ação efetuada na rede. Existem várias comunidades onde os participantes contribuem com dados que poderão ser usados para proveito próprio ou em prol de outros usuários. Podemos citar como exemplo prático as listas de discussão, onde existe um movimento para discutir assuntos com a finalidade de obter um parecer comum. Existem inúmeros casos onde ocorre este tipo de inteligência coletiva, como nos *wikis*, que produzem verbetes que é produzido por mais de um usuário e que é de acesso comum a todos, em aplicativos de código fonte aberto, onde é possível ser modificado ou atualizado por alguém que disponibiliza livremente na rede.
* **Inteligência Coletiva plena:** é obtido ao unir os dois tipos de inteligência coletiva em um mesmo ambiente.

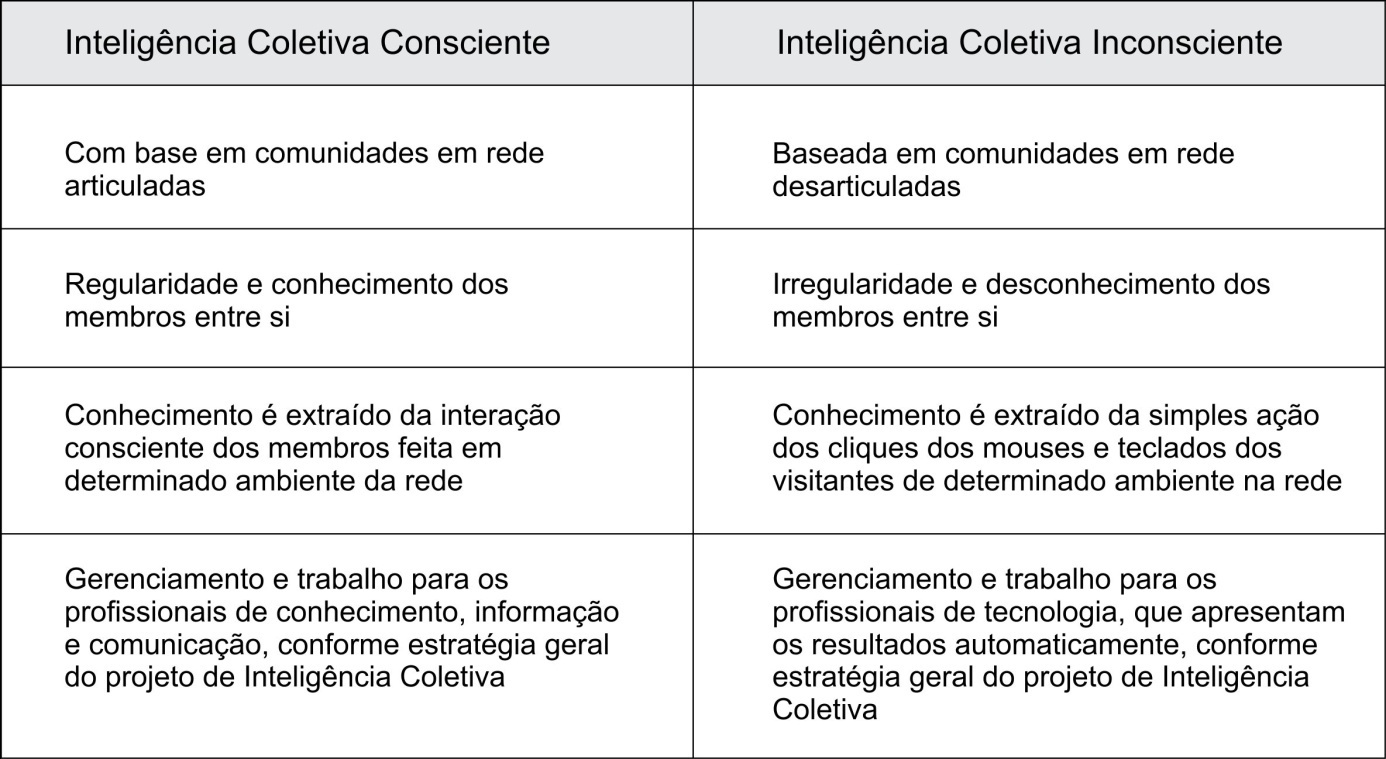


Figura 3 – Características da Inteligência Coletiva Consciente e Inconsciente. (CAVALCANTI; NEPOMUCENO, 2007, p. 40).

**2.3.2 - TIRANDO PARTIDO DA INTELIGÊNCIA COLETIVA**

Uma lição que vários casos de sucesso parecem ter aprendido, é tomar partido da inteligência coletiva para enriquecer os seus aplicativos. Os efeitos na rede resultantes das contribuições dos usuários pode ser a chave da supremacia de mercado na web 2.0 (O’REILLY, 2005).

* **Yahoo!:** Desde antes até mesmo nos dias de hoje, a Yahoo! sabe tirar proveito da contribuição de seus usuários, seja agregando links de sites em um diretório específico ou promovendo discussões a cerca de questões comuns a todos usuários da rede.
* **Google:** com o PageRank, a Google reinventou o modo de se fazer buscas na internet. Ela levou em consideração as incidências dos links na rede para resultar em melhores buscas. A relevância de uma busca está diretamente ligada a quantidade de apontamentos de um mesmo link, que é feita de pelos próprios usuários.
* **eBay e Mercado Livre:** são lojas virtuais onde os donos do site não possuem sequer um produto em estoque. Todas as ofertas e compras são realizadas pelos usuários, bem como as avaliações dos produtos.
* **Amazon:** é tido como um exemplo de grande sucesso de uso da colaboração dos usuários. Para muitos pode parecer que a Amazon procede da mesma forma que seus inúmeros concorrentes, mas ela utiliza uma ciência muito diferenciada em seu sistema. Os comentários dos compradores, as informações adicionais produzidas por eles para cada produto e suas avaliações, são fatores determinantes na relevância das buscas feitas no site.
* **Wikipedia:** a premissa de que qualquer usuário pode adicionar ou editar um verbete é para muitos uma proposta arriscada. Mas a Wikipedia foi a fundo no que diz respeito a confiança radical em seus usuários, e é um grande sucesso na produção de conteúdo colaborativo.

**Flickr, Del.icio.us e Youtube**: tomaram partido da folksonomia, uma nova forma de classificação, para categorizarem o seu conteúdo. Os usuários escolhem livremente as palavras chave que posteriormente produzirão resultados mais relevantes nas buscas.

**2.4 - BLOGS**

Qualquer usuário munido de uma ferramenta com acesso a internet pode inserir inúmeros tipos de informação na rede, mesmo sem estar ciente disso, como podemos ver acontecer com a inteligência coletiva inconsciente. Visitar um site gera um rastro captado por inúmeras aplicações que podem utilizar essa simples informação de alguma forma, como registrar a visita em um sistema de análise de tráfego (*CAVALCANTI; NEPOMUCENO, 2007, p. 36*). A todo tempo a internet recebe esses e muitos outros tipos de informação, inundando a rede com um número exorbitante de dados. Mas a Internet só faz sentido se conter informação passível de absorção por parte dos usuários, como textos, imagens, sons, vídeos, etc. Esses dados, por um bom tempo, só eram inseridos por um número pequeno de indivíduos, por ser necessário um conhecimento prévio das linguagens que faziam necessário para a criação de um site.

Sites pessoais simples já foram bem populares a mais de uma década atrás, possibilitando a qualquer um com conhecimento básico em HTML - linguagem para descrever a estrutura das páginas web (W3C, 2010) - publicar um site na rede, tornando-o disponível em todo mundo (KOMESU, 2004). Mas com o surgimento dos blogs, a tarefa de tornar público um conteúdo específico por um usuário com conhecimentos básicos de informática, se tornou muito mais simples e posteriormente, uma febre.

**2.4.1 – A ASCENSÃO DOS BLOGS**

A palavra blog é uma contração do termo *weblog*, originalmente criado por Jorn Barger em dezembro de 1997, para dar nome ao seu site Robot Wisdom (BAKER, 2010). Provavelmente o primeiro uso público da palavra blog, segundo Baker, foi no site criado pela Pyra Labs em 22 de junho de 1999, o Blogger, e que ainda hoje é uma famosa ferramenta de criação de blogs.

O conceito geral que podemos atribuir à idéia de blog é a de uma plataforma onde sua estrutura permite, de forma dinâmica, a atualização do conteúdo (chamados de artigo ou *post*), e que são ordenados de forma cronológica inversa, podendo ser de autoria de um ou vários autores. Geralmente um blog pode ser criado através de ferramentas na internet, que possibilitam a pessoas sem conhecimentos de qualquer tipo de linguagem de programação, publicar conteúdo na internet em um espaço que possa chamar de seu. Segundo Walker e Bruns (2003, 2005) o blog é visto como um formato, e não como uma plataforma, bastando apenas uma página ter atualizações freqüentes em ordem cronológica inversa para se considerar um blog, independendo da ferramenta usada para isso. Mas para o estudo que se pretende este trabalho, aceitaremos a idéia de blog como plataforma, bem como defende Komesu (2004).

**2.4.2 – DOS SITES PESSOAIS AOS BLOGS**

Partindo da idéia de que para se ter um site pessoal era necessário um mínimo conhecimento de HTML para a sua criação, possuir uma página própria não era muito comum. No início da história da web, ainda na década de 1990, os sites eram estáticos e o conteúdo era renovado com pouca freqüência. Na era da Web 2.0 os sites passaram a ser dinâmicos, com o seu conteúdo armazenado em bancos de dados, permitindo a atualização mais rápida e freqüente. “Ao invés de serem lugares para se visitar, as páginas da web passaram a se tornar plataformas de interação, espaços abertos os quais permitem que qualquer um possa não só consumir como também produzir conteúdo” (ZAGO, 2009, p. 2). Os blogs então ganham forte notoriedade neste contexto, por se tratar de uma plataforma onde qualquer um possa ter um espaço próprio para escrever o que quiser e publicá-lo, sem nem mesmo possuir conhecimento específico em linguagens de programação.

Mas os blogs existem mesmo antes do termo Web 2.0 ser cunhado, porém, novas ferramentas e tecnologias se tornaram disponíveis, como o RSS (*Really Simple Syndication*) e o *permalink* (links permanentes), que possibilitaram aos blogs ganharem ainda mais força, ao transformar o seu conteúdo mais passível de difusão pela rede.

No mesmo momento do surgimento dos primeiros blogs, o termo blogosfera também ganhou força. O termo é normalmente usado para definir a rede de blogs ou blogueiros (usuários que publicam em blogs) que estão interconectados e que formam uma comunidade com uma cultura própria, e que caracterizam um fenômeno social (ZAGO, 2009, p. 8).

**2.4.3 – CARACTERÍSTICAS DOS BLOGS**

Todos os blogs assumem características comuns, que possibilitam a sua identificação. São em sua maioria criados por ferramentas publicadoras de blogs, como no caso do Wordpress e Blogger. Essas ferramentas possuem várias funcionalidades em comum:

* **RSS:** abreviatura de *Really Simple Syndication*, apesar de ser uma tecnologia usual em outros tipos de sites, é mais comum em blogs, e permitem agregar conteúdo que pode ser acessado sem a necessidade de se visitar a página do blog, através de sites ou programas agregadores (ZAGO, 2009, p. 5).
* **Blogroll:** são as listas de blogs favoritos ou indicados pelo autor do blog, que geralmente são posicionados na lateral do site (idem, 2009, p. 5).
* **Permalink:** são ligações permanentes representadas através de uma URL (endereço de uma página específica na internet), que permite o acesso imediato a determinado conteúdo de um blog (O’REILLY, 2006, p. 13).
* **Comentários:** apesar da possibilidade dos blogs não disponibilizarem os comentários, esta funcionalidade se torna uma das mais importantes nos blogs. Elas permitem a interação de um leitor com o autor do conteúdo, promovendo uma discussão que pode resultar numa rica troca de informações (ZAGO, 2009, p. 5).

Com a grande popularização dos blogs, inúmeras variações do formato puderam ser percebidas gerando uma “crise global” do conceito (idem, 2009, p. 6). Essas variações nem sempre obedecem as características iniciais de um blog, como a ausência de comentários, somado a novas funcionalidades que se tornaram possíveis com o surgimento de novas tecnologias, formatos novos surgiram no cenário atual da web. Sendo assim, são variações dos blogs os *tumblelogs*, *fotologs*, *videologs*, *audioblogs*, *moblogs*, *tecnoblogs*, *warblogs*, *splogs*, *klogs*, *blawgs*, *celeblogs*, *microblogs*, entre outras. Essas variações podem ser definidas pelo tipo de conteúdo, como os especialistas em publicação de imagens e vídeos, pela temática, como no caso dos blogs sobre tecnologia e celebridades, e também por características de funcionalidade, como nos casos onde a principal diferença é a brevidade do conteúdo e a forma como é publicada.

**2.4.5 - A SABEDORIA DAS MASSAS**

A blogosfera utiliza da inteligência coletiva para exercer um grande papel na Web 2.0. Ela é responsável por um resultado desproporcional nos mecanismos de busca, primeiramente pelo fato dos blogueiros fazerem parte de uma comunidade com produção frequente de conteúdo e em segundo lugar, “o fato da comunidade de blogs ser tão auto referida – com blogueiros focalizando a atenção em outros blogueiros – aumenta sua visibilidade e poder” (O’REILLY, 2006).

O’Reilly (2006) ainda enfoca o fato da mídia convencional promover uma discussão enfurecida contra os blogs, criticando a informação da comunidade blogueira como algo sem real validade. Mas o que está em evidência é a real discussão sobre um modelo de negócio, onde a era da Web 2.0 promove todos nós como parte da mídia e que define o que realmente é importante.

**3 – PORTAIS DE NOTÍCIA**

Na década de 1990 os meios de comunicação viram nos portais uma plataforma para convergir mídias, que acabou sendo explorado durante as primeiras gerações do jornalismo na web. Os portais de notícia nascem deste contexto, e recebem grande visibilidade ao transportar para a internet parte dos meios convencionais de veiculação da informação, tendo como principal aposta a disponibilização gratuita e o fácil acesso por meio dos leitores (SECCON, 2009).

A evolução do jornalismo que é praticado nas redes telemáticas pode ser categorizada em cinco superfícies, segundo Mielniczuk (2003, p. 44).

Essa separação em cinco superfícies se divide em jornalismo eletrônico, onde se desfruta de equipamentos eletrônicos para sua produção, o jornalismo multimídia, onde existe a conversão dos dados para o meio digital, o ciberjornalismo, onde se usa o computador tanto para realizar a pesquisa jornalística quanto para a sua produção, o jornalismo on-line, que se refere a instantaneidade do fluxo de informação, com uma conexão em tempo real, e por último o webjornalismo, onde o conteúdo é produzido unicamente para a web.

Todas essas superfícies, propostas por Mielniczuk (2003, p. 44), são resultados de uma reformulação do jornalismo após a popularização da web. Em um primeiro momento, o que se pode chamar de primeira geração do webjornalismo, não se via a possibilidade de exploração dos recursos que a web oferecia, podendo constatar a despreocupação com diversas empresas apenas copiando seu conteúdo do impresso para sites, na íntegra. Apenas em uma segunda fase é que se viu a experimentação e exploração de vários recursos, como links e e-mails, onde o conteúdo de uma notícia poderia ser linkado para referências externas, ou até mesmo para textos dentro do próprio conteúdo, imagens e vídeos, e o contato do jornalista com o leitor, onde poderiam haver sugestões, críticas, análises ou simples conversas entre as duas partes.

No que se pode chamar de terceira geração do jornalismo (Idem, 2003, p. 44), o webjornalismo deveria “[...] produzir conteúdos originais em formato multimídia, constituindo sistemas descentralizados próprios, capazes de incorporar as contribuições dos usuários, para apuração, produção e circulação de conteúdos (MACHADO; BORGES; MIRANDA, 2003, p.131).

Como visto anteriormente, o surgimento dos blogs possibilitou a produção e publicação de conteúdo por parte de usuários comuns, sem a necessidade de se tornar um jornalista para isso, ou como era feito anteriormente, a criação de um site pessoal, que necessitava de mínimos conhecimentos técnicos para esta tarefa.

Existe muita discussão no meio acadêmico sobre as informações vindas de blogs ou de sites que não se caracterizam como jornalístico. Porém, isso não impediu que diversos portais de notícias se apropriassem dos blogs para difundir diversos tipos de conteúdo, com o intuito de estreitar as distâncias entre jornalistas e leitores. Os dois lados exploram os diversos recursos oferecidos pelos blogs e em muitos momentos não se enxerga diferença entre as origens de seu conteúdo, sendo eles vindos de grandes agências de notícias ou simplesmente de entusiastas da internet que publicam seu conteúdo de forma profissional.

**3.1- RSS em portais de notícias**

O RSS, tecnologia característica da Web 2.0, muitas vezes é assumida como funcionalidade pertencente apenas aos blogs. Porém, este recurso pode ser visto na web em diversos tipos de sites, onde a intenção é disponibilizar conteúdo de forma ordenada e cronológica através de uma linguagem específica, podendo ser consumida por ferramentas que excluem a necessidade de se visitar o site em questão.

Os portais de notícias, grandes produtores de conteúdo, viram neste recurso uma forma de oferecer aos seus leitores toda sua base de conteúdo através de ferramentas de terceiros, podendo assim atingir uma parcela maior de usuários.

**4 – CLASSIFICAÇÃO COLABORATIVA**

Desde quando nos conhecemos como seres capacitados a pensar, classificamos tudo o que nos interessa, seja pra uma posterior pesquisa ou simplesmente pelo fato de tornar as coisas mais organizadas. Ao separar documentos no nosso ambiente de trabalho por características em comum, ao arrumarmos objetos pessoais de forma a agrupar em porções comuns, ou até mesmo no momento em que separamos as contas pagas das não pagas, estamos classificando objetos e informações, para tornar mais fácil o modo como relacionamos com as coisas (RUFINO, 2010).

Na sociedade contemporânea, com o advento dos sistemas informatizados, acostumamos a classificar as coisas ou informações de formas diferentes. Com o computador, dados podem ser recuperados com mais facilidade, evitando muitas vezes o erro humano, contribuindo para uma disposição da informação de forma mais organizada.

Classificar é dividir em grupos ou classes, segundo as diferenças e semelhanças. É dispor os conceitos, segundo suas semelhanças e diferenças, em certo número de grupos metodicamente distribuídos (PIEDADE, 1983, p. 16).

Tomamos como exemplo o CDD (Classificação Decimal de Dewey) e a CDU (Classificação Decimal Universal), ambas são sistemas de classificação utilizadas em diversas áreas para organizar documentações. Mas esses sistemas, apesar de facilitar a classificação para o que se pretende, não promovem de fato a interação (RUFINO, 2010).

Se a colaboração é a característica primordial da Web 2.0, a classificação colaborativa é a nova ideia de classificação de conteúdos.

Uma nova concepção, que passa agora a ser descentralizada, e na qual o sujeito torna-se um ser ativo e participante sobre a criação, seleção e troca de conteúdo postado em um determinado site por meio de plataformas abertas. Nesses ambientes, os arquivos ficam disponíveis on-line, podendo ser acessados em qualquer lugar e momento, ou seja, não existe a necessidade de gravar em um determinado computador os registros de uma produção ou as alterações feitas na estrutura de um texto. As alterações são realizadas automaticamente na própria Web, pelo próprio usuário e em tempo hábil (SILVA; BLATTMANN, 2007, p. 198).

Em 2004 se torna popular na internet um novo conceito de categorização de conteúdo, uma maneira de classificar que leva em consideração o linguajar dos pertencentes a comunidade que o utilizam: Folksonomia.

**4.1 – FOLKSONOMIA**

Cunhado em 2004 pelo arquiteto de informação Thomas Vander Wal (WIKIPEDIA, online), o termo folksonomia ganha visibilidade ao aplicar de uma forma diferente os conceitos de classificação de conteúdo. Segundo Catarino e Baptista (2007), podemos definir a folksonomia como “[...] o resultado da atribuição livre e pessoal de etiquetas de informações dos recursos na Web, em um ambiente social, compartilhado e aberto a outros, pelos próprios usuários da informação, visando a sua recuperação”. Ainda nesta mesma definição, destacam-se três fatores essenciais:

1. É o resultado de uma indexação livre, feita pelo próprio usuário do recurso;
2. Objetiva a recuperação posterior da informação;
3. É desenvolvida num ambiente aberto que possibilita o compartilhamento e, até, em alguns casos, a sua construção conjunta.

Tomando como partido esse conceito do termo, a folksonomia é utilizada para classificar o conteúdo de terminada comunidade na internet de forma livre e compartilhada para a sua posterior recuperação.

**4.1.1 – USANDO A FOLKSONOMIA**

O uso das *tags* (palavras-chave que caracterizam um assunto ou categoria, seja de uma imagem, texto, vídeo, som, etc) é o resultado da adoção da folksonomia, principalmente quando essas *tags* são escolhidas livremente pelo usuário. Vários sites adotam a folksonomia como forma de classificar o seu conteúdo, e a maneira como isso é feito é através das *tags*. Nas figuras abaixo, alguns exemplos de sites populares que adotam a folksonomia:

  
Figura 4 – Twitter: lista de Trending Topics à esquerda, e no centro, resultados da busca pela *tag* “Vila Cruzeiro” (<http://twitter.com>).

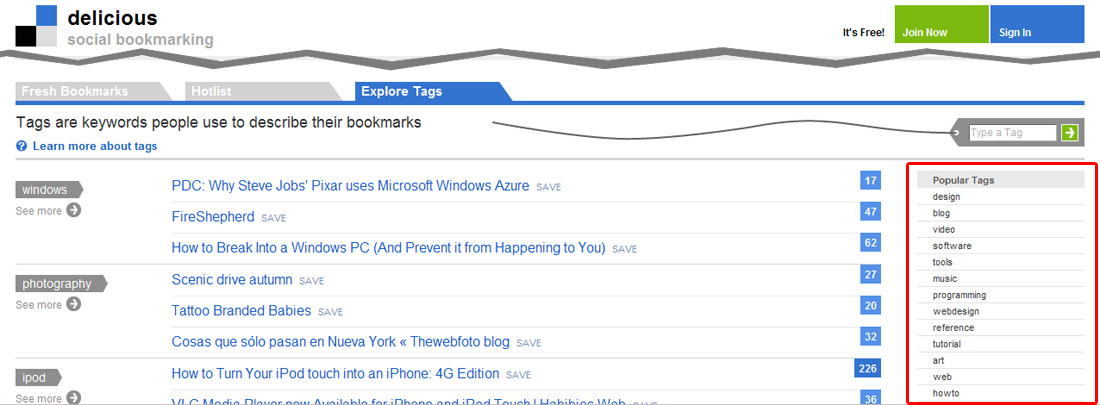


Figura 5 – Delicious: lista lateral com as *tags* mais populares (http://del.icio.us).

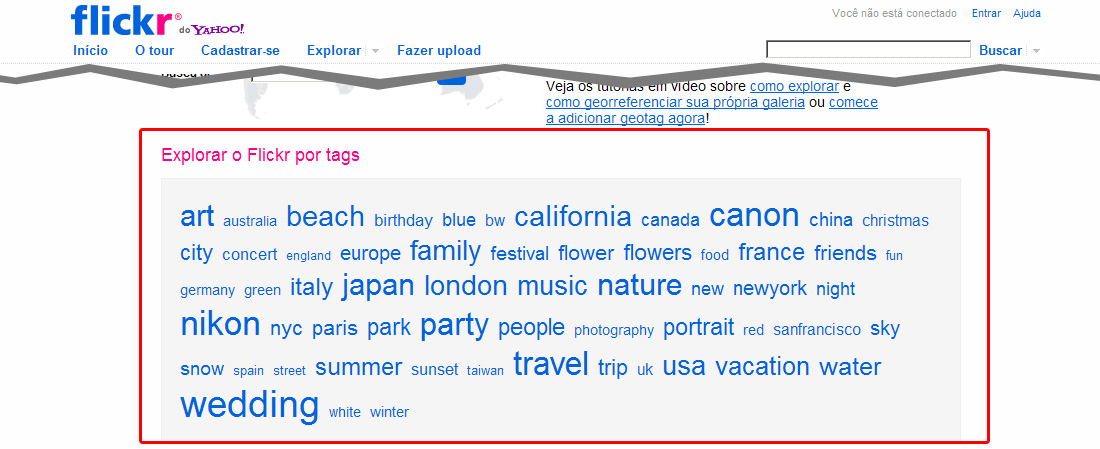


Figura 6 – Flickr: *tags* representadas de acordo com sua relevância (http://flickr.com).



Figura 7 – Youtube: *possibilidade de cadastrar tags* (http://www.youtube.com).

No Youtube, famoso site de vídeos, é possível agregar palavras-chave no momento em que é inserido um arquivo de vídeo. Essas *tags* servirão para complementar a busca feita pelos usuários, onde as palavras escolhidas para categorizar o conteúdo farão parte da classificação inerente ao vídeo.

Já no Flickr, site de armazenamento de imagens, a pesquisa feita pelo usuário para encontrar uma imagem dentre centenas de milhares é facilitada ao atribuir uma tag a imagem. Essa funcionalidade permite classificar todas as imagens por assuntos em comum, como por exemplo, todas as fotos que possuam um beija-flor. As pesquisar pela palavra, o sistema varrerá todas as imagens que possuam a palavra-chave beija-flor agregada a imagem e retornará para o usuário um resultado mais otimizado.

No Twitter, ao usuário escrever uma mensagem (*tweet*), é possível inserir *tags* para posterior classificação. Essa tag é chamada de *hashtag*, e para inseri-la, basta anteceder a palavra com o símbolo #. Não existe uma regra definida para a criação dessas palavras, cabendo ao usuário utilizá-la da forma que achar conveniente. Porém, isso servirá para uma classificação de todas as postagens que contenham a mesma *hashtag*. É possível também seguir uma *hashtag*, assim como acontece com os contatos da ferramenta, que são chamados de seguidores. Ao segui-la, todas as postagens que a contenham, serão apresentadas na linha do tempo de forma cronológica inversa, na página inicial do serviço. Um assunto que é muito discutido e contém uma mesma *hashtag*, poderá ser apresentado na forma de um link, na lateral da página do serviço. Essa listagem é chamada *Trend Topics*, e lista os assuntos mais discutidos que contenha a mesma palavra como padrão classificador.

Todas as formas acima de se classificar um conteúdo levam em consideração a colaboração, os usuários são os responsáveis por determinar a que palavra o conteúdo está atrelado, e nisso resulta um conteúdo que não só é construído por todos, mas também dita a importância de cada informação classificando-a da forma que preferir.

**4.2 – FILTRAGEM COLABORATIVA**

A forma mais simples de se obter uma indicação sobre algo novo, assim como em um filme, livro, música, site, é obter a opinião de amigos e colegas. Além de simples, espera-se obter um resultado satisfatório, já que mantemos amigos em grande parte por interesses em comum. E também é sabido que muitas dessas opiniões recebem um peso diferente por você, por ter em mente o bom gosto de uns e nem tanto de outros. Muitas dessas opiniões podem ser valiosas no momento de decidir consumir algo novo, mas no momento em que várias sugestões vão ficando disponíveis, se torna mais difícil a sua escolha, visto que nem sempre seus amigos estão a par de todas as opções (SEGARAN, 2008, p. 7).

Um grande aliado no momento de obter tais recomendações é o desenvolvimento de técnicas conhecidas como filtragens colaborativas. Segundo Cabré (2011), filtragem colaborativa “é um método de fazer predições automáticas (filtragem) sobre interesses de um usuário, coletando informações sobre o gosto de muitos usuários (colaboradores)”. Isto pressupõe que usuários que concordaram em algo no passado, tendem a concordar novamente no futuro.

O termo filtragem colaborativa foi usado pela primeira vez por David Goldberg no Centro de Pesquisas de Palo Alto da Xerox em 1992 num artigo intitulado Using collaborative filtering to weave na information tapestry. Ele projetou um sistema chamado Tapestry (tapeçaria) que permitia que pessoas fizessem anotações em documentos como interessantes ou desinteressantes e usou essa informação para filtrar esses documentos para outras pessoas [...] (SEGARAM, p. 8, 2007).

Encontramos bons exemplos de sistemas que usam a filtragem colaborativa na internet. Esses sites se aproveitam de tais técnicas para criar recomendações, e para isso necessita das preferências de cada usuário, para só assim poder determinar as similaridades entre eles.

Para a realização de tais recomendações, no uso da filtragem colaborativa, utilizam-se algoritmos capazes de recolher as preferências de um grande grupo de usuários, para logo em seguida criar subconjuntos menores com gostos semelhantes. A partir de então, listam-se o que imagina ser de interesse do usuário. A realização de tal tarefa é possível através da utilização de diversos algoritmos que usam abordagens diferentes. Segaram (2007) cita duas formas bem conhecidas, onde se obtém, através de um calculo matemático, uma escala de similaridade: Escala de Distância Euclidiana e a Escala de Correlação de Pearson.

**Escala de Distância Euclidiana:**

Utiliza os itens em comum que pessoas avaliaram ou categorizaram, e os utiliza como eixos num gráfico. Ela encontra a distância entre dois pontos no espaço multidimensional. Como exemplos, os pontos, que representam a avaliação de cada item para cada usuário, são representados da seguinte forma:

**A** = Adriano, sendo **Ax** a avaliação de Adriano para o item **x**, e **Ay** a avaliação para o item **y**. Da mesma forma, toma-se **B** = Bianca, sendo **Bx** a avaliação de Bianca para o item **x**, e **By** a avaliação para o item **y**.

Temos então a seguinte fórmula para o cálculo da distância euclidiana, num plano bidimensional:

C:\wamp\www\TCC-Dehbora\imagens\formula-euclidiana.jpg

A distância entre Adriano e Bianca é encontrada pegando a diferença entre os itens x e y, elevando-os ao quadrado e somando-as, e depois extraindo a raiz quadrada desta soma. O resultado pode ser utilizado para definir a similaridade entre usuários, permitindo o seu uso em sistemas de recomendação.

**Escala de Correlação de Pearson**

Esta forma é considerada mais sofisticada ao determinar a similaridade entre usuários, pois obtém melhores resultados em situações onde os dados não estão bem normalizados, como no momento onde os usuários avaliam mais negativamente um item em comum.

O Coeficiente de Correlação de Pearson também é conhecido como Coeficiente de Correlação do Momento Produto. Considera-se este como o primeiro método de correlação, estudado por Francis Galton e seu aluno Karl Pearson, em 1897 (SCHULTZ e SCHULTZ, 1992).

Para exemplificar o seu uso, imaginemos uma situação onde Adriano avalia cinco itens, e Bianca avalia também estes mesmos cinco itens. O coeficiente então calcula as somas e as somas dos quadrados das notas nas duas avaliações, e calcula a soma de seus produtos. Finalmente, ele usa esses valores para calcular o Coeficiente de Correlação de Pearson. O resultado será um valor entre -1 e 1. Um valor de 1 significa uma correlação perfeita e positiva entre os dois usuários, enquanto que um valor -1 equivale a uma correlação perfeita e negativa entre os dois usuários. Temos então, de maneira mais complexa, a definição da fórmula:

C:\wamp\www\TCC-Dehbora\imagens\formula-pearson.png

Diferente da métrica da distância, exemplificada na Escala de Distância Euclidiana, a fórmula para o cálculo do Coeficiente de Correlação de Pearson não se mostra muito intuitiva, mas ela efetivamente indica o quanto as variáveis mudam em conjunto, divididas pelo produto de quanto elas variam individualmente (SEGARAN, 2008, p. 7).

Os simples resultados podem determinar o grau de semelhança entre os usuários submetidos ao cálculo do coeficiente, porém, várias linhas de pensamento tendem a especificar limites diferentes para o entendimento destes resultados. Para o desenvolvimento do produto final deste projeto, estabelece, de forma geral, o entendimento dos seguintes resultados:

P = 0 a 0.30 indica fraca correlação.

P = 0.30 a 0.70 indica moderada correlação.

P = 0.70 a 1 indica forte correlação.

Este trabalho não espera esgotar as informações referentes a tipos de coeficientes para o cálculo de correlação, mas os apresenta a fim de esclarecer que o seu uso foi de grande importância para o desenvolvimento da ferramenta proposta neste projeto. Levando em consideração estes esclarecimentos, a escolha para utilizar o Coeficiente de Correlação de Pearson na aplicação final é tomada de forma despreocupada, pois o resultado final, mesmo com sua utilização em caráter crucial, pode prevalecer mesmo com a escolha de outros tipos de algoritmos que calculem correlações entre pessoas.

**5 – DESENVOLVIMENTO**

Este projeto se organiza em três etapas de desenvolvimento, sendo a primeira a aplicação e analise de entrevistas, de onde se preocupa em obter os hábitos dos usuários, buscando comportamentos em comum através das respostas conseguidas e também uma pesquisa de funcionalidades comuns em serviços na internet que se utilizam da classificação colaborativa. A segunda etapa elabora uma proposta de criação de uma ferramenta que tenha em vista os dados das entrevistas e pesquisa, e a revisão de todo o conteúdo literário estudado. Na terceira e última etapa, é descrito o processo de construção desta ferramenta, tomando como base a metodologia de projeto proposta por Jesse James Garrett, em *The Elements of User Experience* (2002).

**5.1 - PESQUISAS**

Para este projeto foi aplicado uma pesquisa com as seguintes características:

* **Participantes:** participaram da pesquisa inicial 20 blogueiros que publicam ou lêem assuntos relacionados à tecnologia em geral. Pode-se prever que este público utilize mais ferramentas disponibilizadas em serviços na web, sendo de grande valia para os estudos deste projeto.
* **Instrumentos:** um questionário disponibilizado na internet pelo GoogleDocs (docs.google.com), com 20 questões sobre o comportamento destes entrevistados diante dos conteúdos disponíveis na web e a forma pessoal de como é feita a participação nesses espaços.
* **Procedimentos:** para definir os participantes da pesquisa, foi feita uma busca em sites específicos sobre tecnologia, onde se pretendeu escolher um pré-perfil de entrevistado, a fim de aperfeiçoar os resultados dos dados do questionário. Todos possuem blogs e leem postagens na internet sobre tecnologia. Após essa escolha, os participantes foram contatados por e-mail e informados sobre este projeto acadêmico. Em um segundo momento, detectado o interesse do blogueiro em participar da pesquisa, um novo e-mail foi enviado, com o link do questionário. Após atingir a quantidade de 20 formulários preenchidos, a pesquisa foi encerrada. Os dados foram tabulados e a partir daí foi feita a análise dos dados.

**5.1.1 – RESULTADOS INICIAIS**

O questionário aplicado na 1ª etapa de pesquisa visou obter dados para entender um comportamento comum entre blogueiros no que diz respeito à participação efetiva em blogs e portais em geral, para poder fazer implementações tanto de funcionalidades como de conteúdo na ferramenta que se propõe desenvolver, bem como atender as necessidades desses potenciais usuários.

No total foram 20 perguntas, sendo elas de escolha única ou preenchimento livre, podendo ser dividido em quatro tipos de dados, mesmo que para o entrevistado isto não tenha ficado claro. São eles:

1. **Dados pessoais:** data de nascimento, sexo, localidade, escolaridade, tempo de acesso de internet e onde mais acessa.
2. **Publicação de conteúdo:** sobre a utilização de alguma ferramenta de publicação de conteúdo, sobre assuntos que costuma escrever e sua frequência, a forma de divulgação, tipos de funcionalidades indispensáveis na hora de se escrever e publicar conteúdo.
3. **Conteúdo específico de tecnologia:** se lê ou não conteúdo na web sobre tecnologia, em que sites costuma encontrar este tipo de informação, se tem o hábito de comentar nestes sites.
4. **Avaliação do conteúdo:** sobre o hábito de avaliar o que se lê e a forma como se dá essa avaliação.

Abaixo, um resumo geral dos entrevistados:

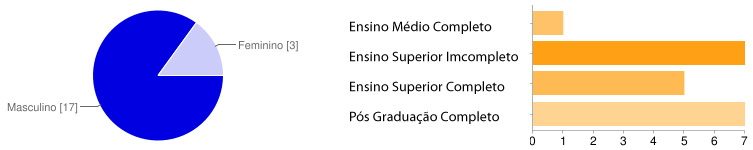


Gráfico 1 – Sexo e escolaridade dos entrevistados

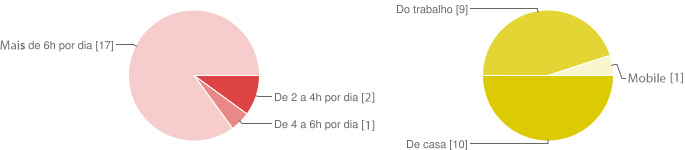


Gráfico 2 – Freqüência e localidade de acesso a internet dos entrevistados

É possível constatar que os entrevistados passam muito tempo na internet, já que o fazem, em sua maioria, por mais de seis horas diariamente, seja acessando do trabalho ou de casa. Isso pode indicar, mesmo que sem afirmar, que todos são familiarizados com a internet.

Todos os entrevistados publicam conteúdo na internet, sejam em blogs, microblogs, grupos de e-mail, listas de discussão ou em redes sociais. Ao publicarem, 12 disseram divulgar o seu conteúdo na rede social de *microblogging* Twitter. Este tipo de atitude ajuda a entender o processo de compartilhamento de uma notícia, e consequentemente suas formas de disponibilização. Na ferramenta proposta por esse trabalho, entender este processo pode ser muito útil.

Além de publicar, todos também lêem outras postagens sobre tecnologia. De Todos eles, 17 disseram avaliar o conteúdo de alguma forma, favoritando no navegador, retuitando, deixando um comentário, compartilhando com amigos ou utilizando funcionalidades existentes em alguns serviços, como ícones de gostei e dando uma certa quantidade de estrelinhas douradas.

Quando perguntado se ao término da leitura de algum conteúdo ele avaliasse de alguma forma, e de que forma, um entrevistado respondeu: “Costumo dar uma certa quantidade de estrelinhas, quando este recurso está disponível. Esta é uma forma rápida de se avaliar um conteúdo sem precisar escrever qualquer coisa”. Outro entrevistado, para a mesma pergunta, respondeu: “ícones de *gostei*”.

Este comportamento indica a necessidade de se poder categorizar ou avaliar o conteúdo de alguma forma. Para uma ferramenta que tome partido das preferências dos usuários, em algum momento será necessário recolher as suas preferências. Nesta pesquisa, ficou clara a ação direta das pessoas neste processo, utilizando algo que talvez se soubesse o seu real motivo, que seria justamente o de classificação.

Ao ler um conteúdo na internet, a maioria dos entrevistados disse não ter o costume de fazer comentários nestas postagens, salvo em raros casos. Para a pergunta ***Você costuma deixar algum tipo de comentário ao ler essas publicações? Explique sua resposta*** um entrevistado respondeu: “Não. Geralmente leio pelo Google Reader e nem vou ao blog”. Um outro respondeu: “Nunca deixo comentários”.

Por este ser o público que se espera ao usar a ferramenta deste projeto, o comportamento de não se fazer comentários pode ser um direcionamento, para uma funcionalidade bem conhecida, porém que pode ser descartada.

Nos anexos deste trabalho se encontra o questionário completo aplicado para os entrevistados, bem como todas as respostas.

Com essa análise dos resultados da pesquisa, é possível determinar alguns itens chaves para a definição dos objetivos e escopo que terá a ferramenta de agregação e recomendação de conteúdo, já que os dados deram um parecer geral do comportamento dos entrevistados. São eles:

1. Os entrevistados passam grande tempo na internet, e possivelmente usam dispositivos móveis para acessar.
2. Descartar a possibilidade de comentários para as notícias que são lidas, já que se demonstrou não ser um hábito comum dos entrevistados.
3. Possibilitar algum tipo de avaliação.
4. Permitir a leitura das notícias na própria ferramenta, mesmo que ainda ofereça um link que direcione o usuário para a notícia em sua fonte real.

**5.1.1 – ANÁLISE DE FUNCIONALIDADES EXISTENTES EM SERVIÇOS NA WEB**

O questionário serviu para direcionar este projeto quanto aos hábitos dos possíveis usuários que fariam uso da ferramenta, porém é necessário, neste momento, entender as funcionalidades já existentes na web, principalmente em sistemas de recomendação. Neste caso, foram coletados alguns sites com ideias semelhantes a este projeto, e que promovam de alguma forma a interação do usuário com o seu conteúdo. Estes sites tem em comum a possibilidade de se avaliar, categorizar ou classificar o conteúdo apresentado por eles.

Foram analisados três sites com grande volume de usuários e já consagrados na internet:

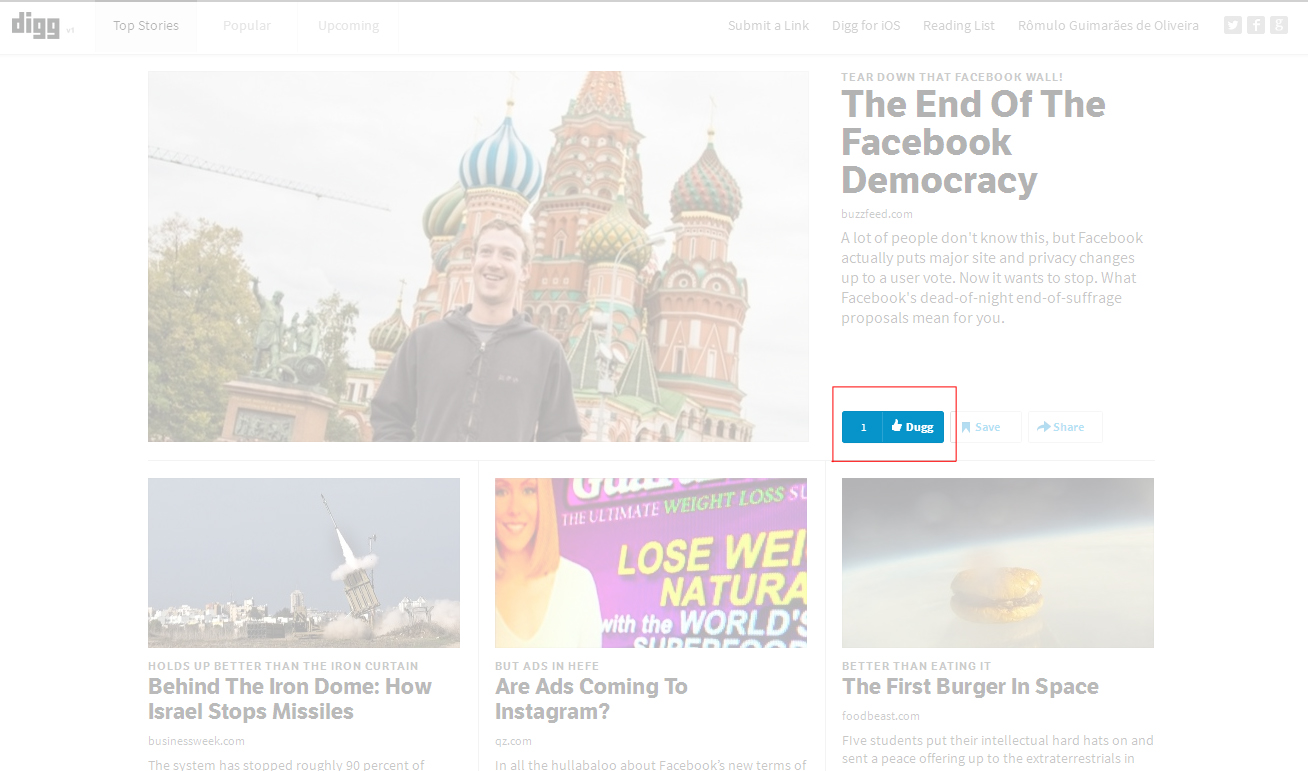
1. **Digg**. Reúne resumos de notícias, vídeos, fotos e outros tipos de conteúdo enviados pelos próprios usuários e avaliado por eles. Site em: www.digg.com.
2. **Google Reader**. Um agregador de feeds, que reúne várias fontes de notícias em um único lugar, possibilitando a leitura de conteúdo sem precisar visitar a fonte. Site em: reader.google.com.
3. **Amazon**. Site de vendas de todos os tipos de produtos e que possui um grande banco de dados de informações sobre seus produtos, que são submetidos pelos próprios usuários. Site em: [www.amazon.com](http://www.amazon.com).

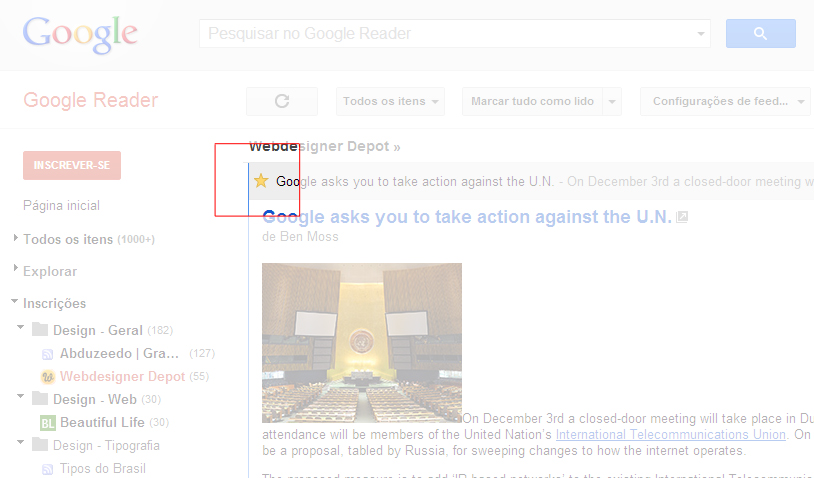
Todos os três exemplos acima utilizam pelo menos uma forma de se avaliar o seu conteúdo ou produto.

 Figura 8 - Avaliação através de estrelas douradas. www.amazon.com

Freqüência e localidade de acesso a internet dos entrevistados No exemplo acima, do site da Amazon, o meio de se avaliar um produto é através de estrelas douradas. Também encontrado em diversos outros sites, este medidor serve para indicar a qualidade ou relevância da informação, onde um menor número de estrelas significa um produto ou conteúdo de baixa qualidade ou simplesmente não muito recomendado pelos usuários, e um número alto de estrelas apontando para algo altamente recomendado e de boa qualidade. No caso da Amazon em específico, um bom uso deste recurso é para a comparação entre dois produtos que se queira comprar, porém ainda se encontra com dúvidas para a escolha final.

No exemplo da imagem abaixo, o site Digg disponibiliza um botão, com o ícone de um sinal de joia seguido do texto digg. No seu contexto, ele serve para o usuário determinar se gosta ou não de uma notícia, e o número a esquerda exibe a quantidade de usuários do site que também usaram este botão para este mesmo conteúdo. Notícias mais populares tendem a ter um número mais elevado nesta funcionalidade, porém isto não indica necessariamente a sua relevância para o usuário que não avaliou ainda. O Digg também lista as notícias mais populares levando em consideração a quantidade de vezes que os usuários clicaram no botão digg. Ele determina essa lista como sendo os artigos mais populares.

 Figura 9 - Botão digg. Possibilita que o usuário determine o seu interesse. www.digg.com

 Figura 10 - Estrela dourada. Possibilita ao usuário determinar a importância da informação. reader.google.com

Na ferramenta Google Reader, uma forma de classificar as notícias agregadas pelos feeds cadastrados é atribuir uma estrela dourada para um determinado item. Presente também no Google Gmail, esta funcionalidade serve principalmente para lembrar o usuário que em algum momento ele atribuiu certa importância para aquela notícia, sendo possível, posteriormente, filtrar apenas os itens indicados com a estrela dourada.

Um recurso comum nos dois últimos casos é o fato de existir um botão ou link direcionando o usuário para a notícia em seu local de origem. Esta possibilidade serve para lembrar o usuário que o conteúdo presente em sua ferramenta não é de própria autoria, sendo agregado de uma fonte externa, e que se for de sua escolha, pode ler a notícia no site de onde a derivou.

Enquanto o Google oferece a possibilidade de se cadastrar um feed, que seria a fonte de notícias de algum site, o Digg oferece a possibilidade de cadastrar apenas u link, que se refere a uma página específica. No primeiro caso, o que acontece é a agregação de conteúdos de diversos sites, onde cada fonte provê os seus conteúdos com atualização por parte das fontes, sendo listados os itens de cada site em separado. Essas informações são de cada usuário, não sendo compartilhados, pelo menos inicialmente, com outras pessoas. No Digg o cadastro é de um item em específico, onde você insere o link e este mesmo é compartilhado com todos os usuários do site, podendo assim receber avaliações.

Todos os três exemplos analisados se preocupam em classificar o seu conteúdo de alguma forma, para que seja possível uma filtragem posterior ou analise de um item mais relevante para determinado usuário. A característica de avaliação se faz necessária na ferramenta proposta por este projeto por capturar as preferências dos usuários e possibilitar o seu uso posterior em recomendação de conteúdo. Apesar de ser uma forma simples de se obter a preferência de usuários, pode ser uma poderosa informação, já que esta ação foi feita de forma direta e espontânea.

**5.2 – DESENVOLVIMENTO DA FERRAMENTA - DEHBORA**

Com as pesquisas e analises realizadas, agora se propõe a criação de uma nova ferramenta, disponibilizada em um website que forneça funcionalidades mínimas que possam testar os conceitos abordados neste projeto. O site se chama Dehbora e está disponível em www.dehbora.com.br.

Para a definição de todas as etapas de desenvolvimento, este projeto se utiliza dos conceitos abordados por Jesse James Garret em *The Elements of User Experience* (Os Elementos da Experiência do Usuário) (2002). Para Garret, a experiência do usuário não diz respeito ao funcionamento interno de um produto ou serviço e sim sobre como ele funciona do lado de fora, quando uma pessoa entra em contato com ele. Este projeto se preocupa quase que unicamente neste quesito, não entrando em detalhes sobre o funcionamento interno do site. Talvez o único momento onde a preocupação não é de fato com a interface do usuário é na forma como é realizado o cálculo de semelhança entre os usuários, onde se procura utilizar algoritmo de filtragem coletiva e que não será apresentado de fato para o usuário final, sendo apenas parte do processamento interno do sistema que faz funcionar o site.

**Os Cinco Planos**

Para Garret a experiência do usuário ao visitar um determinado site é praticamente a mesma. O resultado final é resultado de um conjunto de decisões, de como o site se parece, como ele se comporta e o que ele permite que o usuário faça. Todas essas decisões são construídas umas em cima das outras, e o seu entendimento para as tomadas de decisão podem então ser definidas em camadas, cuja proposta vai da parte mais abstrata possível até a parte mais concreta.

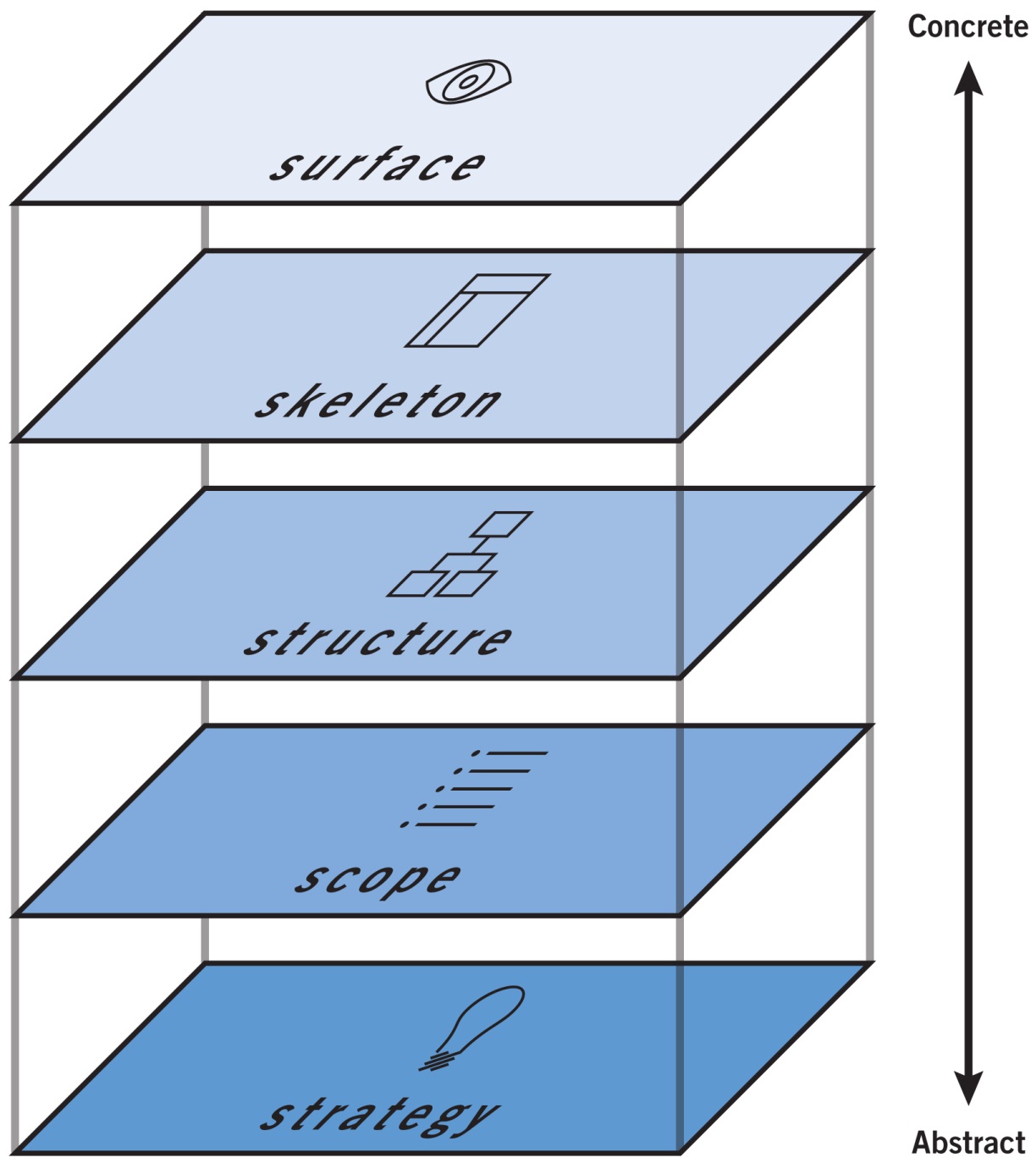


Figura 11 - The Five Planes (Os cinco planos) (GARRET, 2002, p. 22)

1. **Superfície**: neste plano são apresentadas uma série de páginas da web, que podem conter textos, imagens, áreas clicáveis que direcionam o usuário para mais páginas ou que realize determinadas funções.
2. **Esqueleto**: logo abaixo do plano de superfície, o esqueleto é a organização de tudo que é visível no plano superior, projetado para aperfeiçoar a disposição de todos os elementos na página para efeito máximo de eficiência.
3. **Estrutura**: o esqueleto poderia ser considerado a parte mais concreta do que é a estrutura. Estrutura tenta definir o caminho que determinado usuário pode percorrer para acessar determinada função, que pode estar contida em uma categoria ou página específica, e também determina como essas funções se encaixam.
4. **Escopo**: define os recursos e funções que estarão incluídos no site, determinando uma espécie de limite para o desenvolvimento da aplicação.
5. **Estratégia**: o escopo é fundamentalmente definido pela estratégia. Esta estratégia, por ser o mais abstrato dos planos, define o modelo de negócio, a relação entre os usuários e os objetivos do site.

Apesar da ordem apresentada partir da Superfície para a Estratégia, Garret (2002) deixa claro que a construção deve partir de cima para baixo. No plano mais abstrato, onde se define a estratégia, não é necessário e nem mesmo aconselhável pensar na forma visual final do site. Bem como quando se chega no plano da superfície, onde a preocupação é apenas visual.

Os cinco planos podem não ser suficientes para cobrir a grande capacidade da web em permitir diversos tipos do seu uso. Com a Web 2.0, já discutida neste projeto, surgiram novas formas de se usar a web, e o que era apenas um local para se exibir informações diversas, passou a oferecer aplicações com diversas funções. Desta forma, os cinco planos básicos proposto por Garret (2002) podem ser divididos ao meio e sub-categorizados, com o intuito de cobrir um número bem maior situações, aplicando terminologias diferentes para tipos diferentes de requisições.

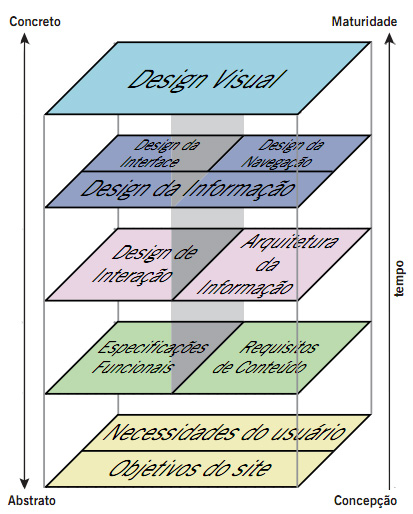


Figura 12 - Diagrama dos elementos da experiência do usuário. Tradução para o Português por Livia Labate (GARRETT, 2002, p. 29)

O lado direito direciona o desenvolvimento para a web como um sistema de hipertexto, orientado a informação. O da esquerda tenta cobrir a web como uma interface de software, onde se orienta à tarefa.

Como o Dehbora pretende oferecer um sistema que permita realizar tarefas e que não se orienta apenas como um sistema de hipertexto, então o lado esquerdo do diagrama é que será utilizado para o desenvolvimento da ferramenta.

Estes cinco planos, partindo de baixo para cima, definem as etapas de desenvolvimento do site Dehbora, sendo melhor detalhadas nos tópicos seguintes.

**5.2.1 – ESTRATÉGIA**

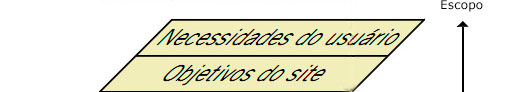


Figura 13 - Plano de Estratégia.

Para definir a estratégia do site Dehbora, é preciso definir os objetivos do projeto e entender as necessidades dos usuários. Na aplicação do questionário foi possível entender muito do comportamento de usuários familiarizados com a internet e que potencialmente podem ser o perfil de quem usaria esta ferramenta. A partir disso já é possível entender a parte das necessidades do usuário, e aliar aos propósitos da criação da ferramenta proposta por este projeto. Sendo assim, algumas perguntas são feitas, e suas respostas direcionam todo o campo de estratégia:

**Objetivos do Site**: criar uma ferramenta de agregação de notícias e que ofereça recomendações personalizadas para os usuários. O site deve permitir que o usuário possa escolher a fonte de notícias preferida e que o seu conteúdo possa ser visualizado em um único local. Também deverá sugerir novas leituras de notícias, realizando recomendações baseadas na filtragem colaborativa entre todos os usuários da ferramenta.

**Necessidades do usuário**: com a analise dos dados da pesquisa realizada neste trabalho, percebeu-se os seguintes pontos que determinam as necessidades do usuário:

1. Possibilitar a avaliação de uma notícia/conteúdo;
2. Não é necessária a funcionalidade de comentários para as notícias;

Tendo isto em vista, o site Dehbora deverá oferecer recursos suficientes para que o usuário cadastre fontes de notícias, possa lê-las no próprio site, permitir a sua avaliação, excluir a possibilidade de comentários e listar recomendações de novas notícias que foram lidas e avaliadas por outros usuários.

A estratégia determina o início do desenvolvimento, porém não exclui a possibilidade de alterações no decorrer das outras superfícies do diagrama.

**5.2.2 – ESCOPO**



Figura 14 - Plano de Escopo.

Com a estratégia definida, o próximo ponto é definir o escopo, que segue como segundo plano no diagrama.

A definição do escopo cria um conjunto de requisitos que determina o que será construído no projeto. Segundo Garret (2002) o escopo também ajuda a saber o que você não está construindo. Partimos da questão abstrata “Por que estamos fazendo este produto?” que lidamos no plano de estratégia e vamos construir sobre ela uma nova pergunta: “O que nós vamos fazer?”.

Como falado anteriormente, o diagrama poderá ser divido ao meio, e é neste plano que esta divisão fica clara. Este trabalho se encarrega apenas de tratar da parte referente a construção de um site como um objeto orientado a tarefas, e por isso apenas o lado esquerdo de todo o diagrama deverá ser levado em consideração.

O escopo então é definido através das especificações funcionais. Essas especificações descrevem o que o site deverá fazer e como se comportará ao ser usado. Com a estratégia desenvolvida, então listamos todos os requisitos presentes neste projeto:

1. **Cadastro gratuito de usuários para uso completo da ferramenta/site**: o usuário que entrar no endereço [www.dehbora.com.br](http://www.dehbora.com.br) receberá informações suficientes que permitam o cadastro gratuito na ferramenta. Esse cadastro deverá ser rápido e simples, solicitando apenas informações básicas como nome, e-mail e senha. O acesso completo a ferramenta só será permitido para usuários cadastrados;
2. **Cadastro de feeds (RSS ou ATOM)**: o site Dehbora não conterá notícias proprietárias, apenas permitirá ao usuário o cadastro de fontes de notícias de sua preferência. O cadastro será realizado com a inserção de um link válido de um feed no formato RSS ou ATOM, possibilitando também a sua atribuição a um nome de livre escolha do usuário. Esta nomeação é uma forma de permitir a categorização por parte do usuário, e pode ser considerado, mesmo que em pequeno grau, como uma forma de folksonomia;
3. **Gerenciar os feeds cadastrados:** assim como o usuário possui a liberdade de criação de fontes de notícias, ele também poderá excluí-las se assim achar necessário.
4. **Listar as notícias por feed cadastrado**: com um determinado feed cadastrado, o site deverá ser capaz de exibir a listagem de notícias referentes àquela fonte de notícias específica. Essa listagem é o ponto de partida para a leitura de conteúdo que o próprio usuário escolheu obter;
5. **Leitura da notícia sem sair do site (porém, exibindo a página completa da notícia)**: o Dehbora deverá ser capaz de exibir a página da notícia na íntegra, sem sair do endereço [www.dehbora.com.br](http://www.dehbora.com.br). Esta funcionalidade é necessária para que possibilite a sua avaliação de forma simples e rápida, pois ao ler a notícia, o Dehbora deverá exibir a forma de avaliação neste mesmo contexto;
6. **Oferecer o recurso de avaliação de cada notícia**: no momento da leitura da notícia, o Dehbora deverá exibir também a funcionalidade de avaliação. Esta avaliação se dará através da computação de estrelas douradas. Quanto mais estrelas um usuário der para uma determinada notícia, mais próxima ela estará das preferências do usuário;
7. **Oferecer um link que direcione o usuário para a página original da notícia**: como o Dehbora irá exibir a notícia na íntegra e com a funcionalidade de avaliação, então deverá também permitir que o usuário se redirecione para a página original do conteúdo;
8. **Listar recomendações de notícias baseado na correlação entre os usuários**: o Dehbora deverá utilizar o algoritmo de correlação de Pearson para calcular a correlação entre os usuários. Com o grau de similaridade calculado entre todos os usuários, a ferramenta será capaz de exibir uma lista de recomendações de notícias ainda não avaliadas para o usuário;
9. **Permitir o uso do site em dispositivos móveis**: o Dehbora deverá se comportar de forma satisfatória ao ser acessado através de dispositivos móveis;
10. **Compatibilidade da ferramenta com navegadores padrões**: o site [www.dehbora.com.br](http://www.dehbora.com.br) deverá ser compatível com os principais navegadores do mercado.

**5.2.3 – ESTRUTURA**



Figura 15 - Plano da Estrutura.

Para Garret (2002), a estrutura é onde se inicia o processo concreto de desenvolvimento do site, apesar de ainda envolver assuntos em parte conceituais.

O reino da estrutura é o terceiro dos cinco planos, e de forma adequada, é o ponto em que nossas preocupações mudam de questões mais abstratas de estratégia e escopo para os fatores concretos que vão determinar o que é realmente a experiência. Mas a linha entre o abstrato e o concreto pode ser borrada, embora muito do que decidir aqui vai ter uma influência perceptível e tangível sobre o produto final, mas as decisões ainda envolvem assuntos em grande parte conceituais. (GARRET, 2002, p. 80).

No desenvolvimento de um software, a estruturação da experiência do usuário é chamada de design de interação. Design de interação está preocupado em descrever os possíveis comportamentos do usuário e definir como o sistema irá acomodar e responder a esse comportamento (Idem, 2002, p. 85). Como o site Dehbora procura se definir como uma ferramenta pertencente à Web 2.0, o modelo de estruturação é totalmente baseado nesta ideia.

Para descrever a estrutura de um site com detalhes específicos da ferramenta, podemos fazer uso do chamado diagrama de arquitetura (Idem, 2002). O diagrama não precisa documentar cada link de cada página do site, ele serve mais para estabelecer as relações conceituais, onde cada função específica se relaciona com outra, e como etapas de uma sequência de interação se encaixam.

Para o Dehbora, levando em consideração os itens propostos no plano de escopo, temos o seguinte diagrama de arquitetura:

DIAGRAMA

**5.2.4 – ESQUELETO**



Figura x - Plano do Esqueleto.

**5.2.5 – SUPERFÍCIE**



Figura x - Plano da Superfície.

**6 – DISCUSSÃO**

Discussão

**7 – CONSIDERAÇÕES FINAIS**

Considerações finais

**8 – ANEXOS**

Anexos