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.

*Aqui vem a revisão de literatura (lembrar de retirar e alterar algumas informações do texto original).*