

SURPRISE.ME - SISTEMA MOTIVACIONAL DE RECOMENDAÇÕES COLABORATIVAS

Cléber Moraes Reis

Faculdades Integradas de Taquara – Faccat – Taquara – RS – Brasil
7cleber@gmail.com

Francisco Assis Moreira do Nascimento

Professor Orientador
Faculdades Integradas de Taquara – Faccat – Taquara – RS – Brasil
assis@faccat.br

Resumo

Diversos estudos apontam que o uso em demasia de aparelhos eletrônicos e internet, faz com que estes usuários tenham uma maior propensão a apresentar sintomas de depressão, mal-estar e comportamento antissocial. Este trabalho trata do desenvolvimento de uma ferramenta inserida nesse contexto, mas que se propõe justamente a combater estes problemas. A ferramenta consiste de um aplicativo que indica atividades aos usuário baseando-se no humor deste e nas experiências e avaliações de outros usuários. Mediante o uso de técnicas de filtragem colaborativa, os dados coletados a partir de avaliações de diversos grupos de usuários são classificados e quantificados de forma que possam servir como base para ajudar outros usuários, que apresentam características semelhantes a esses grupos, a amenizar os referidos problemas.

Palavras-chave: sistemas de recomendação, filtragem colaborativa, atividades.

SURPRISE.ME – A MOTIVATIONAL COLLABORATIVE RECOMMENDATION SYSTEM

Abstract

Several studies suggest that people that make excessive use of electronic devices and Internet have a great tendency to present symptoms of depression, and antisocial behavior. This paper presents the development of a tool oriented to cope with these problems. The tool consists of an application that recommends activities for the user based on his mood and experiences as well as based on reviews from another users. By means of collaborative filtering techniques, the data collected from assessments of different groups of users are classified and quantified in order to be used as a basis to help others users, who have similar characteristics to those groups, to alleviate their problems.

Keywords: *recommender systems, collaborative filtering, activities.*

1 INTRODUÇÃO

Nos dias de hoje é possível perceber uma grande quantidade de sistemas de recomendação nas mais diversas áreas do comércio eletrônico (RONEN, 2013). O sistema proposto no presente trabalho se baseia também neste conceito, em que o usuário recebe indicações de produtos e serviços conforme vai delineando seu perfil dentro desses sistemas. Entretanto, o produto a ser indicado não terá finalidade comercial e sim emocional. Todo e qualquer conteúdo que sugira algum estado de humor e que possa surpreender os mais variados usuários do sistema será o produto em questão para o sistema de recomendação desenvolvido neste trabalho.

Há ainda outros diferenciais: não há qualquer administrador, pois tudo que é indicado, criado e avaliado é de origem dos próprios usuários. Além disso, a técnica de filtragem e recomendação é uma versão condensada e compacta de algoritmos de grande popularidade, não trazendo resultados tão precisos mas ainda assim relevantes, o que facilita a implementação em sistemas de pequeno e médio porte.

Baseado no contexto da web 3.0 e design responsivo (ZERVAS, 2014), foi desenvolvido um sistema de recomendações motivacionais que trata a filtragem e apresentação de informações relacionadas a diversas atividades voltadas para desconectar o usuário do mundo virtual. O sistema desenvolvido combina funcionalidades colaborativas que devem surpreender o usuário sempre que este estiver disposto a novas experiências e avalia estas experiências a fim de criar um fluxo de opiniões que classifique o conteúdo consultado, e possa influenciar positivamente na escolha de outros usuários. Secundariamente a arquitetura do sistema desenvolvido pode ser reutilizada por outros desenvolvedores e assim facilitar a criação de sistemas de recomendações colaborativas para as mais diversas áreas.

O presente artigo é organizado da seguinte maneira: na Seção 2, é apresentado o conceito dos sistemas de recomendação com foco nos algoritmos mais usados em técnicas de filtragem colaborativa; na Seção 3 são discutidos trabalhos correlatos; na Seção 4 são expostas a arquitetura do sistema e os métodos de coleta de dados adotados; na Seção 5, o processo de desenvolvimento é descrito; na Seção 6 os resultados obtidos são reportados; na Seção 7 são feitas as considerações finais e finalmente na Seção 8 são listadas as referências bibliográficas.

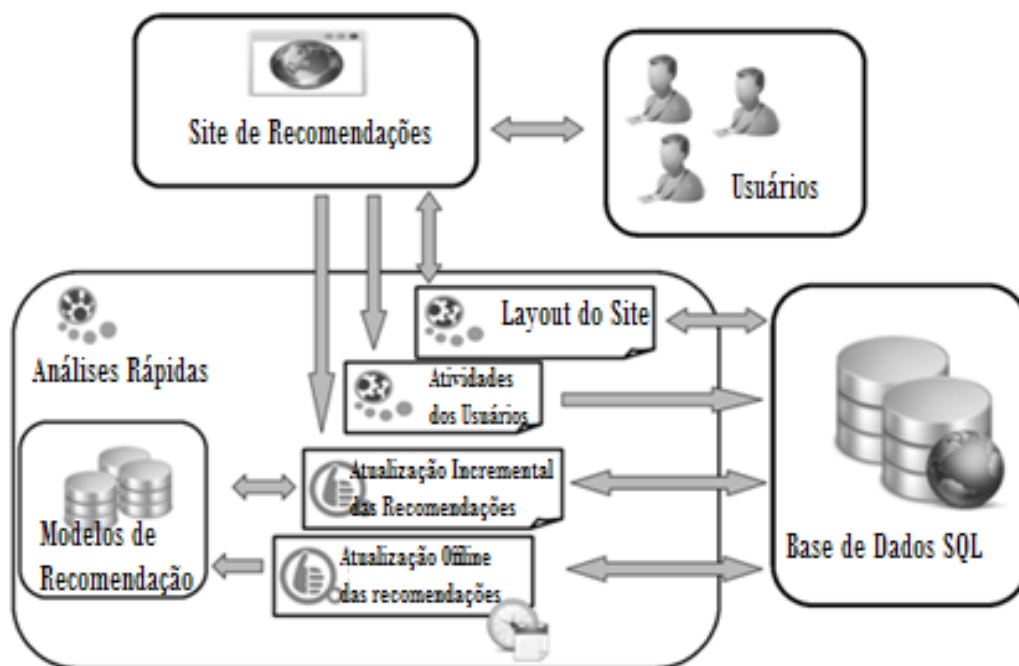
2 REFERENCIAL TEÓRICO

2.1 Sistemas de recomendação

Segundo Ricci, Rocach e Shapira (2011), os sistemas de recomendação surgiram como uma tentativa de otimizar com eficiência e eficácia o processo de exibição, recomendação e direcionamento de conteúdo para possíveis interessados. No cenário corporativo, que é o foco desta forma de sistematização, os principais objetivos são a fidelidade e conseqüentemente o aumento da lucratividade das empresas. Conforme o plano de negócio de cada empresa, podem ser implementadas diferentes estratégias de recomendação, que podem ser, como ilustrado na Figura 1, baseadas em conteúdo, colaboração ou em métodos híbridos.

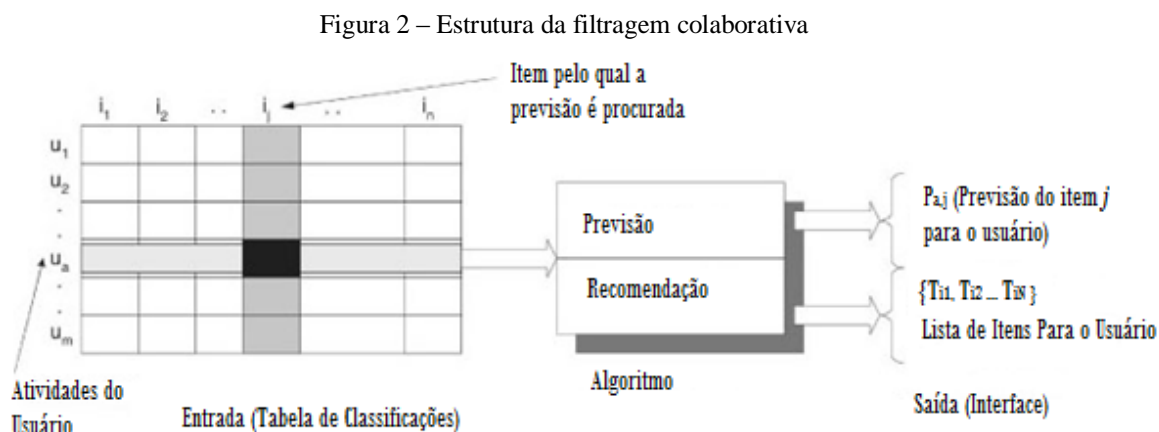
Os sistemas que utilizam o conceito de Recomendações Baseadas em Conteúdo indicam itens ao usuário semelhantes a itens preferidos pelo mesmo no passado. Quando a estratégia utilizada é a de Filtragem Colaborativa, o usuário receberá recomendações de itens que pessoas com gostos similares aos dele preferiram no passado. Em sistemas que utilizam métodos híbridos as recomendações são baseadas tanto em estratégias de Recomendação Baseadas em Conteúdo quanto em estratégias baseadas em Filtragem Colaborativa (BOBADILLA, 2013).

Figura 1 – Estrutura dos Sites de Recomendação



2.2 Filtragem Colaborativa

Segundo Su e Khoshgoftaar (2009), o princípio básico de um sistema baseado em filtros colaborativos é que ele deve se utilizar de uma base de dados com experiências dos usuários. Estas experiências são representadas pela avaliação e/ou interação do usuário em relação a itens. A partir disso, o sistema tende a adivinhar itens adicionais de interesse para seus usuários. Em geral, como mostrado na Figura 2, na base de dados consultada, há usuários e itens. Para cada usuário, há uma lista de itens com uma classificação do usuário para cada item. Essas classificações, geralmente, são indicadores explícitos de aprovação e reprovação em uma escala de 1 a 5, mas também podem ser indicações implícitas de interação, como compra de itens, visualizações e *clicks*.



Fonte - Su e Khoshgoftaar (2009)

Para a implementação deste tipo de sistemas, a previsão de indicações pode ser feita através de métodos de aprendizado de máquinas, teorias de aproximação e vários tipos de heurísticas. Dentre essas diversas técnicas é possível destacar dois algoritmos que são extensamente utilizados. O primeiro é um algoritmo baseado em vizinhanças de K (K-Nearest-Neighbors – KNN) (LUO *et al.*, 2013) e o outro é o algoritmo de filtragem colaborativa baseada em tendências (CACHEDA, 2011).

2.2.1 Algoritmo Baseado em Vizinhanças de K(KNN)

Segundo Luo *et al.* (2013), esse algoritmo busca encontrar semelhanças entre itens, o conjunto dos k itens mais semelhantes a um item recebe o nome de vizinhança desse item. A partir da vizinhança, prevê-se uma avaliação para um usuário da matriz usuário-item ainda não avaliado por este usuário alvo. Busca-se estabelecer uma vizinhança formada pelos K itens mais parecidos com o usuário alvo. Se $S(i)^k$ for a vizinhança definida pelos k itens mais similares ao item i, a nota que o usuário u atribuiria ao item i, $\hat{r}_{u,i}$ pode ser estimada pela equação (1):

$$\hat{r}_{u,i} = \frac{\sum_{j \in S(i)^k} sim(i,j)r_{u,j}}{\sum_{j \in S(i)^k} sim(i,j)} \quad (1)$$

Nesta equação (1), $sim(i,j)$ é o grau de semelhança entre os itens i e j e $r_{u,j}$, a nota que o usuário u atribuiu ao item j. Para a aplicá-la, a similaridade entre dois itens i e j pode também ser determinada pela correlação de Pearson, conforme a equação (2), onde U é o conjunto de todos os usuários que avaliaram os itens i e j, \bar{r}_i é o valor médio das avaliações recebidas por i e \bar{r}_j é o valor médio das avaliações recebidas por j (LUO *et al.*, 2013).

$$sim(i,j) = \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{u,j} - \bar{r}_j)^2}} \quad (2)$$

2.2.2 Algoritmo Baseado em Tendências

Para Cacheda (2011), usuários parecidos podem avaliar itens de maneiras diferentes, reservando notas boas ou ruins somente para itens realmente bons ou ruins, respectivamente. Por isso, este algoritmo fundamenta suas indicações nas diferenças entre usuários ou itens, ao invés das similaridades.

A equação (3) resulta na tendência de um usuário τ_u , através da diferença média entre as suas avaliações e a média do item, sendo I_u o conjunto de itens avaliado pelo usuário u, $r_{u,i}$ a nota que o usuário u atribuiu ao item i e \bar{r}_i é a média de notas que o item i recebeu. Da

mesma forma, a equação (4) demonstra o cálculo de tendência de um item τ_i (CACHEDA, 2011).

$$\tau_u = \frac{\sum_{i \in I_u} (r_{u,i} - \bar{r}_i)}{|I_u|} \quad \tau_i = \frac{\sum_{u \in U_i} (r_{u,i} - \bar{r}_u)}{|U_i|} \quad (4)$$

Para a predição de um item são consideradas, além das tendências, as médias do usuário e do item. Se tanto o item quanto o usuário tiverem tendências positivas, a previsão será superior a ambas médias, conforme a equação (5). Em caso contrário, o cálculo será igualmente contrário, tal qual a equação (6) (CACHEDA, 2011).

$$\hat{r}_{u,i} = \max(\bar{r}_u + \tau_i, \bar{r}_i + \tau_u) \quad (5)$$

$$\hat{r}_{u,i} = \min(\bar{r}_u + \tau_i, \bar{r}_i + \tau_u) \quad (6)$$

Possuindo usuário e item, tendências e médias contrárias, a previsão deve ficar entre os dois, como mostra a equação (7). β varia entre 0 e 1 e regula a contribuição da média do item e do usuário na previsão. Quando apenas a média é contrária, o cálculo é feito tal qual a equação (8) (CACHEDA, 2011).

$$\hat{r}_{u,i} = \min(\max(\bar{r}_u, (\bar{r}_i + \tau_u)\beta + (\bar{r}_u + \tau_i)(1 - \beta)), \bar{r}_i) \quad (7)$$

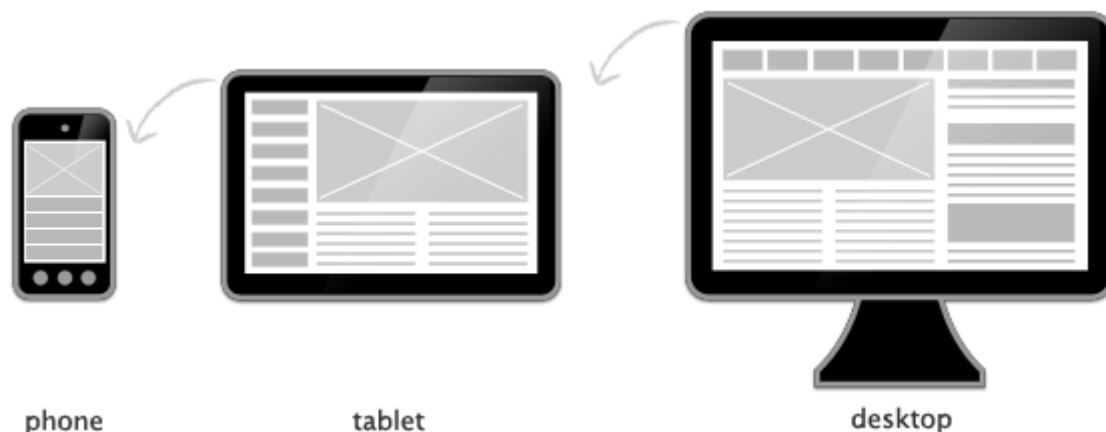
$$\hat{r}_{u,i} = \bar{r}_i\beta + \bar{r}_u(1 - \beta) \quad (8)$$

2.3 Design Responsivo

Segundo Zervas (2014), com o surgimento de novas plataformas de acesso à internet surgiram também diversos problemas de adaptabilidade dos sites existentes com relação às suas visualizações. A solução seria fugir do design estático, ou seja, desenvolver o layout de forma que ele possa ‘responder’ conforme o dispositivo que o usuário está utilizando. É tão somente uma questão de adequar a formatação do conteúdo para cada tamanho e o tipo de tela. Para esta finalidade é possível utilizar a ferramenta Media Queries (SILVA, 2011).

Media Queries ou consulta de mídia, são códigos específicos de redirecionamento do visitante do sistema em HTML5 para um código CSS3 apropriado ao dispositivo que ele estiver utilizando, conforme ilustrado pela Figura 3. Nestes códigos auxiliares, é possível definir o tipo, tamanho, orientação, resolução e proporção de cada dispositivo (ZERVAS, 2014).

Figura 3 – Design Responsivo



Fonte - Zervas (2014)

A figura 3 demonstra a versatilidade e adaptabilidade dos conteúdos, quando orientados através de Media Queries. É possível perceber que os elementos das páginas são redimensionados de acordo com o dispositivo utilizado pelo usuário, resultando em uma melhor visualização e acessibilidade.

3. TRABALHOS CORRELATOS

Reategui (2006) diz que desde o surgimento dos sistemas de recomendação, o interesse de pesquisas na área vem aumentando em decorrência da vasta aplicação destes sistemas como parte crucial de sistemas de e-commerce e sites de empresas como Amazon (AMAZON.COM INC., 2014), Google (GOOGLE CORPORATION, 2014) e eBay (EBAY INC., 2014).

Last.fm é um site que usa o seu sistema de recomendação como sendo o seu produto principal. Ainda que, a base de arrecadação do site sejam as mensalidades, o serviço de descobertas de músicas que faz recomendações personalizadas com base nas músicas que o usuário ouve é o atrativo que a empresa usa para conseguir mais clientes. O algoritmo deste site é baseado, principalmente, nas tags (etiquetas que definem o estilo e outras características

das músicas) e artistas parecidos com os que o usuário costuma ouvir (HENNING; REICHELT, 2008).

Um outro site, que também se utiliza de sistemas de recomendação como uma de suas principais ferramentas, é o Netflix. Usando metadados (dados sobre outros dados), os analistas do Netflix conseguem encontrar as semelhanças entre filmes e programas de TV, como o ano de produção, diretor, gênero, avaliações, entre outras coisas. Além disso, fatores como o horário do acesso, o dia da semana e comportamentos específicos dos usuários também influenciam na seleção do algoritmo.

Inserindo o sistema desenvolvido neste trabalho no mesmo contexto de ambos os sites citados, pode-se perceber inicialmente uma grande diferença em um dos principais fatores de um sistema de recomendação: a relação usuário-item. Tanto Last.fm quanto Netflix se utilizam de gostos pessoais e características do usuário para prever as próximas músicas ou filmes com suas respectivas características. O sistema deste estudo trata usuários e indicações de forma semelhante. Ambos possuem as suas classificações em uma mesma escala. E o algoritmo selecionará o item que tem classificação semelhante ao usuário.

Não seria apropriada uma comparação entre as duas abordagens, uma vez que, o produto a ser indicado neste trabalho não possui uma definição formal e por isso também não dispõe de outras características a serem avaliadas, ou seja, não há nenhuma exigência do sistema em relação ao conteúdo a ser inserido, sendo assim o usuário livre para sugerir o que achar apropriado. Esse fato facilita a implementação, mas dificulta em uma maior precisão pelo fato de que quanto mais características tenham os produtos, mais parâmetros podem ser utilizados para um refinamento das indicações.

4 DESCRIÇÃO DO SISTEMA DE RECOMENDAÇÃO

4.1 Requisitos Funcionais do Sistema

O sistema é composto de cinco funcionalidades principais:

- a) o cadastro de usuários, que é onde o usuário responderá as questões básicas (sexo, idade, estado de humor), que o classificarão em um perfil de usuário, além de incluir os dados de login;

- b) o cadastro de recomendações, intitulado ‘Surpreenda’, onde o usuário cadastrado poderá incluir no sistema qualquer conteúdo (textos, indicações culturais, experiências), para que estes possam receber uma classificação posteriormente;
- c) a seção avaliação de recomendações, intitulado ‘Avalie’, na qual o usuário poderá classificar os conteúdos que foram incluídos na seção ‘Surpreenda’;
- d) a seção solicitação de recomendações, intitulada ‘Experimente’, onde o sistema se valerá das classificações do usuário e dos conteúdos cadastrados para indicar os conteúdos mais apropriados, essas indicações devem ser avaliadas pelo usuário; e
- e) a seção histórico, intitulada ‘Relembre’, na qual o usuário poderá rever todas as suas avaliações.

A Figura 4 mostra, da direita para a esquerda as telas das seções 'Experimente', 'Surpreenda' e 'Avalie'.

Figura 4 –Design do Sistema para dispositivos mobile.



Fonte – o Autor

No canto superior esquerdo da figura 4 está a seção experimente, onde o usuário tem a sua disposição 5 níveis de humor identificados por ‘emoticons’. O usuário deve escolhe um deles, o que melhor lhe represente no momento, e isso definirá a classificação de humor do

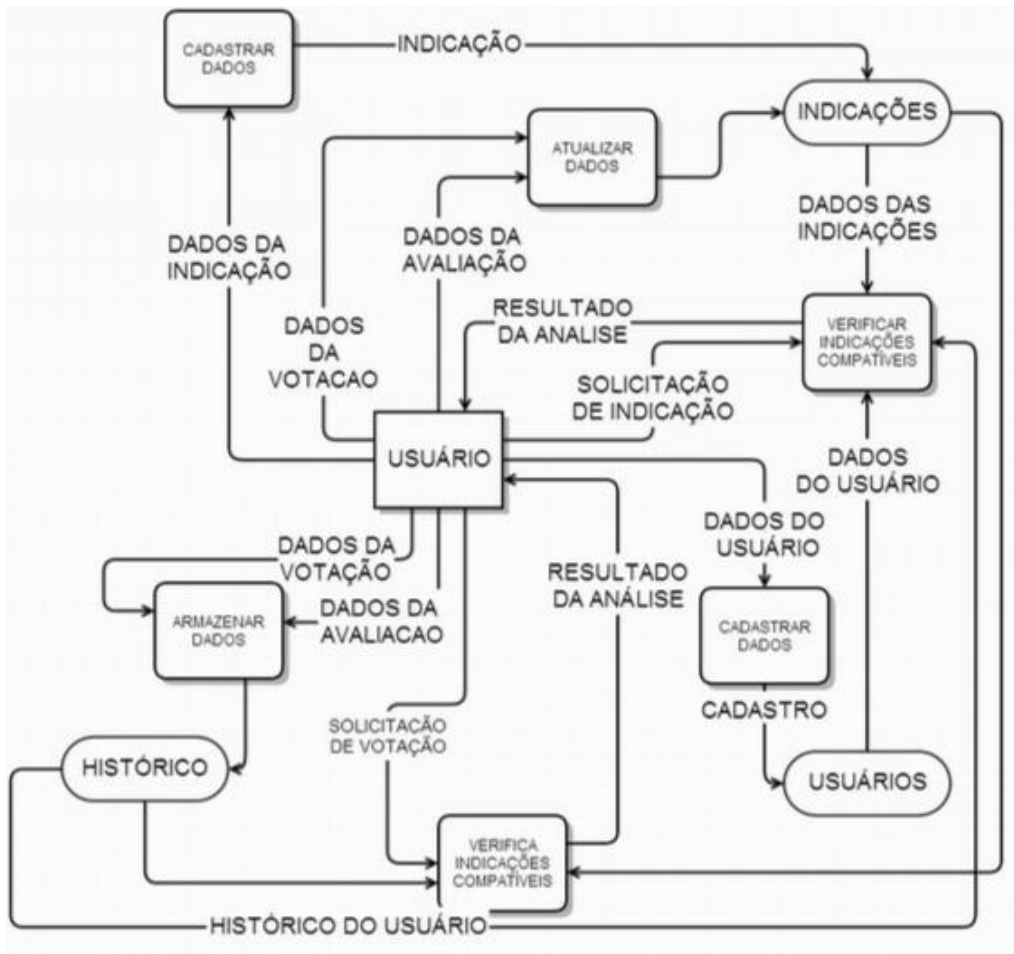
usuário e automaticamente ativará o algoritmo de seleção de indicações, que selecionará os conteúdos que estejam o mais próximo possível desta classificação definida pelo usuário.

Na tela exibida no canto inferior esquerdo da imagem 4 está o conteúdo que foi selecionado pelo sistema, o número de visualizações deste conteúdo, os ratings de usuário e conteúdo, a compatibilidade entre estes, os percentuais de cada avaliação entre os usuários e duas informações que não estão sendo utilizadas, neste momento, no sistema, mas já foram incluídas para futuras atualizações e aperfeiçoamentos: categoria e perfil (ambas as informações ainda não possuem definição sobre suas características).

Na tela central da figura 4 está a seção ‘Surpreenda’, onde os usuários devem inserir os conteúdos (e em futuras atualizações, deverá ser informado ainda a categoria) a serem indicados para outros usuários na seção ‘Experimente’, entretanto, para que este conteúdo de fato se torne indicável, ele deve passar por uma avaliação de 5% dos usuários. Avaliação esta que tem a sua tela exibida na direita da figura 4. O nível de humor mais votado será a classificação de humor (rating) do conteúdo.

No diagrama DFD mostrado na Figura 5, é possível notar que o usuário se situa no centro do sistema. O usuário fornece os dados de suas indicações e as suas avaliações à tabela “Indicações” e secundariamente à tabela “Histórico”. Fornece também os seus próprios dados pessoais à tabela “Usuários”. A função ‘Verificar indicações compatíveis’ se vale destas três tabelas para retornar a indicação que melhor poderá satisfazer o usuário. Na tabela “Usuário”, essa função observa o rating do usuário, na tabela de “Indicações”, ele seleciona o rating de indicação mais aproximado dentre os que não estão na tabela “Histórico”.

Figura 5 – Diagrama DFD do Sistema



Fonte - o Autor

4.2 Coleta de Dados

Os dados a serem coletados pelo sistema são: dados pessoais e estado de humor do usuário, recomendações do usuário para outros usuários e avaliações do usuário a cerca de recomendações. Nas seções seguintes serão detalhados esses processos de coleta.

O usuário, ainda no cadastro, além de informar seus dados, deve indicar um estado de humor entre os níveis de 1 a 5, onde 1 indica um maior nível de infelicidade ou mau humor e 5 indica felicidade e/ou bom humor. Esta primeira escolha é o que definirá a classificação do humor em seu primeiro acesso ao sistema. Nos acessos seguintes, esse usuário será sempre questionado em qual estado de humor ele se encontra e a resposta que ele dará em cada acesso contribuirá para a formação do estado de humor médio do usuário, denominado “rating do usuário”.

Todo usuário do sistema, opcionalmente, pode indicar, na seção ‘Surpreenda’, todo e qualquer conteúdo que sugira algum estado de humor ou melhora de estado e que possa surpreender ou simplesmente entreter os demais usuários do sistema. Estes conteúdos não serão de imediato submetidos ao sistema de indicações, pois eles ainda não possuem o seu rating. Por isso, essas indicações são submetidas à seção ‘Avalie’, onde ficarão disponíveis para avaliação opcional de qualquer usuário, até que se chegue ao limite de cinco por cento do total de usuários, com a finalidade de estabelecer o estado de humor médio deste conteúdo, que é denominado “rating de indicação”. Essa avaliação varia entre 0 e 5, onde o 0 significa que o conteúdo deve ser excluído do sistema, 1 é a recomendação para usuários com extrema infelicidade ou mau humor e 5 é proporcionalmente o inverso de 1. Notas entre 2 e 4 representam humores intermediários.

Para fins de exemplificação, há o caso hipotético do acesso do usuário *u1* que insere na seção ‘Surpreenda’ um conteúdo *c1*, que equivale a uma letra de música, pronta para ser tocada e recomenda que as pessoas que sabem tocar, o façam. Esse conteúdo ainda não possui rating, e assim ainda não é sabido para quais níveis de humor ele pode ser indicado, por isso ele é disponibilizado na seção ‘Avalie’ para que 15 quaisquer usuários (5% do total de usuários, assumindo que sejam 300 no momento da inserção do conteúdo) possam definir uma nota de 0 a 5 para o conteúdo. Onde 0 representa a exclusão da indicação, 1 e 5 representam estados extremos de humor e 2, 3 e 4 indicam estados intermediários. Após as 15 avaliações *c1* recebeu uma média de avaliações de 2.50, e assim esse será o rating da indicação, que será exibido sempre que um usuário com rating aproximado acessar a sessão ‘Experimente’.

Esses registros são armazenados em uma conta online no Sistema Gerenciador de Banco de Dados MySQL hospedado no servidor das Faculdade Integradas de Taquara (fit.faccat.br). A estrutura desenvolvida para esse armazenamento é composta de 3 tabelas: Usuário, Indicações e Histórico. A tabela Usuário armazena os seguintes dados do cadastrado: nome, sexo, ano de nascimento, email, senha para acesso, rating e número identificador único do usuário. A tabela indicações contém: conteúdo da indicação, notas de 0 a 5 dadas por cinco por cento dos usuários na seção ‘Avalie’, notas de 1 a 4 dadas pelos usuários que receberam o conteúdo como indicação, rating da indicação, categoria e número identificador único do conteúdo.

4.3 Cálculo de Ratings

De posse dos ratings de usuário e ratings das indicações é possível se utilizar do conceito de filtragem baseado em vizinhanças. Entretanto, o vizinho a ser indicado não será baseado nas escolhas feitas pelo usuário no passado e sim pela proximidade do seu rating com o rating da indicação. Essa indicação será avaliada pelo usuário entre 1 e 4. Se o usuário gostar do conteúdo, ele deve dar uma nota 3 (curtir) ou 4 (favorito). Estas duas notas farão com que o rating daquele conteúdo contribuía para o rating do usuário, a fim de que haja uma aproximação entre usuário e conteúdo. Isso faz com que os próximos vizinhos indicados sejam mais aproximados daquele que foi apreciado. Se o usuário não gostar do conteúdo, ele deve dar uma nota 2 (não gosto) ou 1 (ignorar). Neste caso o rating do usuário se manterá com o mesmo valor.

Para ilustrar este processo, considere o seguinte exemplo. O usuário *u1* cadastrou-se no sistema e informou como seu estado de humor o nível 2. O sistema escolherá para este usuário todos os conteúdos que, na seção 'Avalie', foram avaliados como nível 2 ou mais próximos deste nível, conforme mostrado na Tabela 1). Após o processamento da solicitação, o sistema encontra para o usuário o conteúdo *c1*, que possui rating de 2.5. O *u1* entende que aquele conteúdo eleva ou pode elevar o seu estado de humor ou ajudar-lhe de alguma forma, então ele concede ao conteúdo a nota 3 (curtir). Nesse momento o rating do usuário é alterado de forma que se aproxime do rating do conteúdo, ficando agora com 2.25 de rating de humor. Novamente, o *u1* solicita uma nova indicação, e dessa vez, baseado no novo rating, o sistema exibirá um conteúdo *c2* com rating 2.55, ao invés de um *c3* com rating 1.90, que seria indicado caso o usuário não tivesse apreciado *c1* (de acordo com o mostrado na Tabela 2). Neste exemplo, *u1* teve uma alteração de 12,5% no seu rating, após aprovar um conteúdo. Evidentemente, é uma mudança muito radical, portanto a quantidade de indicações disponíveis no sistema é um fator relevante para o sucesso do sistema. Quanto mais indicações, mais vizinhos próximos e menor a necessidade de recorrer a vizinhos distantes, vizinhos estes que acabam por deixar uma discrepância muito grande nos ratings dos usuários.

Tabela 1 – Primeiro Acesso de *u1*

Usuários		Conteúdos			
Usuários	Ratings	c1	c2	c3	c4
		2,5	2,55	1,9	1,6
<i>u1</i>	2	X			
<i>u2</i>	3				
<i>u3</i>	2,5				
<i>un</i>	4				

Onde X representa a indicação apresentada.

Fonte - Autor

Tabela 2 – Segundo Acesso de *u1*

Usuários		Conteúdos			
Usuários	Ratings	c1	c2	c3	c4
		2,5	2,55	1,9	1,6
<i>u1</i>	2	I	X		
<i>u2</i>	3				
<i>u3</i>	2,5				
<i>un</i>	4				

Onde X representa a indicação apresentada e I representa que a recomendação já não pode mais ser selecionada, pois já está no histórico do usuário.

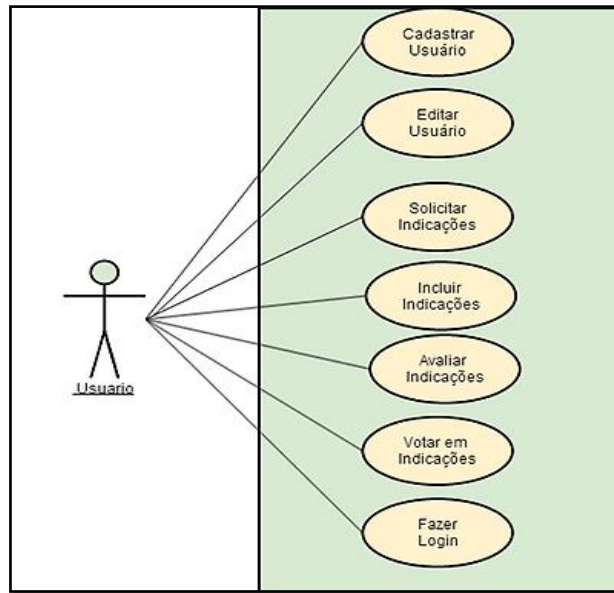
Fonte - Autor

5 DESENVOLVIMENTO

O modelo de desenvolvimento utilizado neste sistema foi o modelo clássico ou cascata, iniciando pela fase de análise de requisitos e seguindo sequencialmente pelas etapas de projeto, implementação, testes (validação), integração e finalmente a manutenção de software. Para análise e projeto foi utilizado o software DFD Editor ¹ conforme mostrado no diagrama de casos de uso na Figura 6. Este diagrama expressa com perfeição o fato de que apenas o usuário tem permissões dentro do sistema.

¹ DFD Editor é uma ferramenta de desenho independente que ajuda o usuário a desenhar, editar e validar os seus Diagramas de Fluxo de Dados de forma eficiente. Os usuários podem desenhar DFDs multi-nível e verificar seu saldo de fluxo.

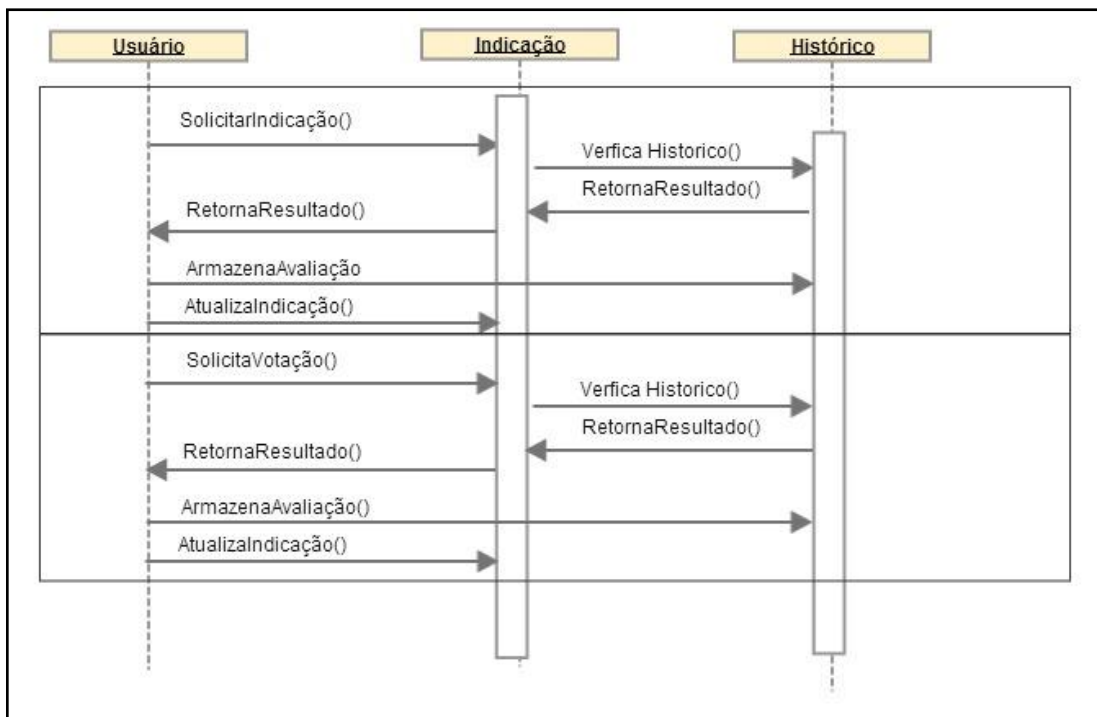
Figura 6 – Diagrama de Casos de Uso



Fonte: Autor

A Figura 7 mostra um diagrama de sequência que modela o cenário da seleção de indicações.

Figura 7 – Diagrama de Sequencia da Seleção de Indicações



Fonte – Autor

A Figura 7 apresenta a modelagem, na forma de uma diagrama de sequencia editado com o software DFD editor, para a funcionalidade implementada no código da figura 8, mostrada a seguir.

A codificação do sistema foi feita através do software Adobe Dreamweaver². A interface envolve as linguagens HTML5³, CSS3⁴ e Javascript⁵ com Framework jQuery⁶. As indicações, avaliações e demais conteúdos dinâmicos foram desenvolvidos com as linguagens PHP5⁷ e SQL⁸. A Figura 8 a seguir mostra um trecho de código na linguagem de programação PHP da aplicação desenvolvida, que inclui a execução de comandos na linguagem SQL para bancos de dados .

Figura 8 – Seleções de Indicações com PHP5 e SQL

```
$sql = "SELECT TEXTO, ID_FRASE, OK, NOK, FAVORITOS, UP, RATING, OFICIAL
FROM FRASE
WHERE OFICIAL=1
AND RATING>'$RATINGUSER'-'$DIFERENCA'
AND ID_FRASE NOT IN
(SELECT ID_FRASE FROM HISTORICO WHERE ID_USUARIO='$IDUSUARIO')
ORDER BY RATING ASC LIMIT 1 ";
$recursoExec = mysql_query($sql, $conect);
while($escrever=mysql_fetch_array($recursoExec)) {
echo $escrever['ID_FRASE'] .
    $escrever['TEXTO'] ;
```

Fonte - Autor

² *Adobe Dreamweaver*, é um software de desenvolvimento voltado para a web com notável suporte para várias tecnologias web, tais como XHTML, CSS, *Javascript*, *Ajax*, PHP, ASP, ASP.NET, JSP, ColdFusion e outras linguagens Server-side (BONATTI, 2013).

³ *Hypertext Markup Language*, 5 é uma linguagem para estruturação e apresentação de conteúdo para a World Wide Web. Esta versão traz consigo importantes mudanças no mundo da Web, com novas funcionalidades de semântica e acessibilidade (CROWTHER, 2013).

⁴ *Cascading Style Sheets 3* é uma linguagem de estilo onde se define efeitos de transição, imagem, e outros dentro de uma página HTML (CROWTHER, 2013).

⁵ *JavaScript* é atualmente a principal linguagem para programação *client-side* em navegadores web.

⁶ *jQuery* é uma biblioteca JavaScript cross-browser desenvolvida para simplificar os scripts *client-side* que interagem com o HTML (CROWTHER, 2013).

⁷ *Hypertext Preprocessor* é uma linguagem interpretada livre, usada para o desenvolvimento de aplicações presentes e atuantes no lado do servidor, capazes de gerar conteúdo dinâmico em páginas HTML (CROWTHER, 2013).

⁸ Linguagem de Consulta Estruturada é a linguagem de pesquisa declarativa padrão para banco de dados relacional (base de dados relacional) (CROWTHER, 2013).

Na figura 8 é possível visualizar o uso das linguagens PHP5 e SQL em uma seleção pela indicação mais próxima do rating do usuário. A consulta em questão é feita diretamente no banco de dados MySQL e é retornado em uma estrutura padrão desenvolvida em HTML5 e CSS3.

6. RESULTADOS OBTIDOS

Para avaliar a efetividade e eficácia do sistema, foram incluídas, manualmente na base de dados do sistema, 250 indicações com as mais variadas classificações de rating, estes conteúdos foram selecionados a partir de buscas realizadas no site Google com as seguintes palavras chave: “me deixa feliz”, “me deixa melhor” e “me faz sentir bem”. As classificações a respeito do conteúdo, que em geral serão feitas pelos usuários, neste primeiro momento, foi definido pelo contexto em que tal conteúdo foi encontrado.

Foram incluídos, também manualmente, 100 usuários fictícios igualmente diversificados. Cada usuário possui rating igual a 0.05 acima do antecessor, partindo de 1.00 até 5.00. Esses usuários são representados por 6 usuários reais. Cada um dos três pares de usuários reais foi enquadrado, de acordo com seu comportamento e humor rotineiros, em uma das 3 zonas em que foram divididas as recomendações: ratings baixos (entre 1 e 2.5), ratings médios (entre 2.6 e 3.9) e ratings altos (entre 4 e 5).

Cada usuário real, designado para avaliar a zona de ratings baixos e médios é responsável por 17 usuários fictícios destas mesmas categorias. Os usuários designados para a zona de ratings altos são responsáveis por 16 usuários fictícios desta mesma categoria. As avaliações dos usuários em relação às recomendações apresentadas foram divididas em aprovadas (notas 3 e 4) e reprovadas (1 e 2). Os percentuais obtidos pelas avaliações vão sendo calculados e mostrados automaticamente, logo abaixo da própria indicação, conforme mostrado na Figura 9.

Figura 9 – Cálculo de Aprovações e Reprovações.



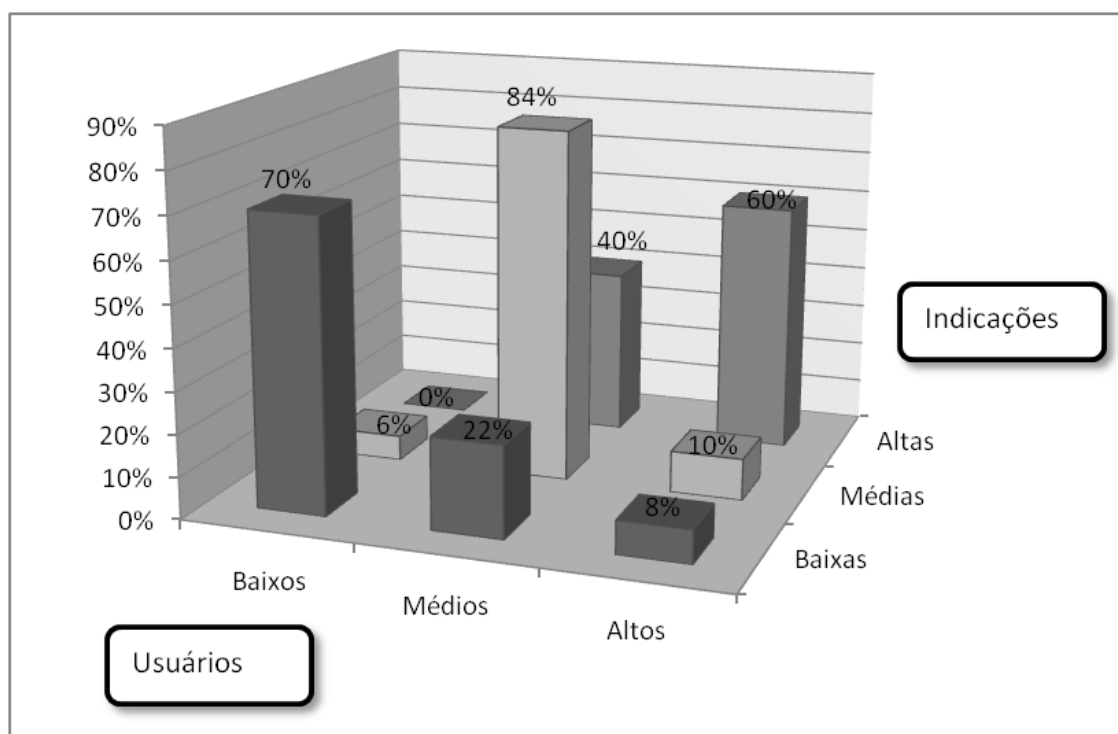
Fonte - Autor

Os usuários que se declararam com baixo nível de humor aprovaram na grande maioria os conteúdos que, em muitos casos, se referiam a livros de auto-ajuda, filmes de comédia e músicas e/ou álbuns alegres. Isso fez com que, aos poucos, o rating destes usuários fosse subindo para que novos vizinhos daquelas indicações aprovadas fossem exibidas. Entretanto, ao circular entres indicações de rating médio, os usuários aprovaram menos da metade da indicações, que se referiam, em grande parte, a livros de gêneros amenos, música, álbuns e filmes de romance e drama, fazendo com que o rating decrescesse, voltando a zona anterior.

Já os usuários que partiram de ratings médios, proporcionalmente, aprovaram e reprovaram indicações, fazendo com que ora o rating subisse e ora baixasse, e com que 84% se mantivesse nesta zona de indicações e os outros 16% acabaram por elevar ou reduzir o seu ratings às outras zonas.

Por fim, os usuários com ratings altos, aprovaram 60% das indicações que recebiam (em grande parte se tratavam de filmes de ação, livros de aventura e músicas de grande intensidade), mas ainda assim, 40% de reprovações não foi suficiente para que o rating provocasse uma troca de zona, fazendo com que 100% dos usuários se mantivessem com rating acima de 4, conforme mostrado no gráfico da Figura 10.

Figura 10 - Relação de Aprovação entre usuários e indicações.

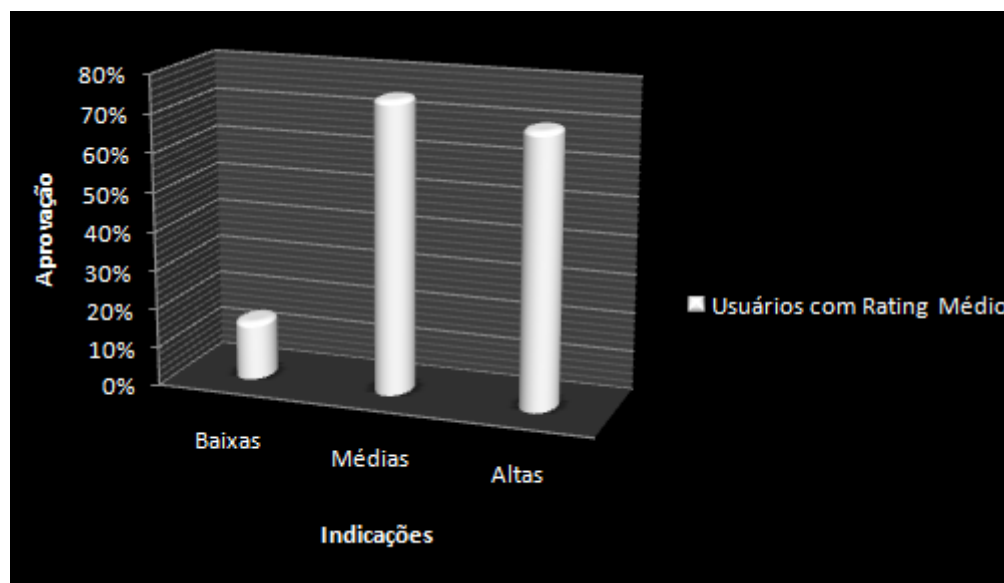


Fonte - Autor

Para de fato comprovar que existe influência da indicação sobre o usuário, os testes sobre a base de dados foram refeitos e, desta vez, o algoritmo de seleção das recomendações foi substituído por uma função randômica, objetivando a aleatoriedade das indicações, ou seja, não eram verificados os ratings das indicações. Os usuários com ratings baixos recebiam indicações dos mais variados ratings, assim como os outros dois grupos (médios e altos). A finalidade seria verificar se não há tendências de aprovações entre os grupos, comprovando assim a importância do rating de indicação.

Na primeira verificação, os grupos de usuários com ratings baixos (entre 1 e 2,5) foram submetidos a 250 indicações diversificadas. Devendo aprová-las com nota 3 (curtir) ou 4 ('favoritar') ou reprová-las com nota 2 (reprovar) ou 1 (ignorar). Do total de indicações com ratings baixos, 60% foram aprovadas: 45% do total de ratings médios e apenas 35% de todas as indicações com ratings altos apresentados, conforme ilustrado pelo gráfico da Figura 11.

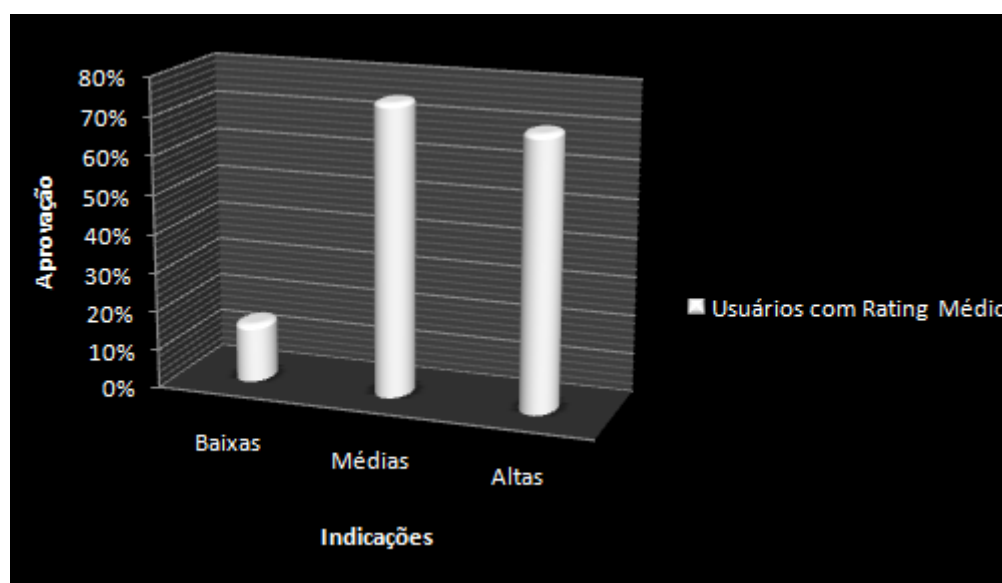
Figura 11 – Avaliação de Recomendações Aleatórias por Usuários com Ratings Baixos



Fonte - Autor

Na segunda verificação, os usuários com ratings médios foram submetidos às mesmas indicações, em condições totalmente aleatórias, e novamente, as indicações preferidas foram aquelas que realmente pertencem a mesma classificação do usuário. No entanto, as aprovações para indicações de ratings baixos foram expressivamente rejeitadas e os de ratings altos foram igualmente aprovados, chegando a, respectivamente, 15%, 75% e 70% de aprovações, conforme mostrado no gráfico da Figura 12.

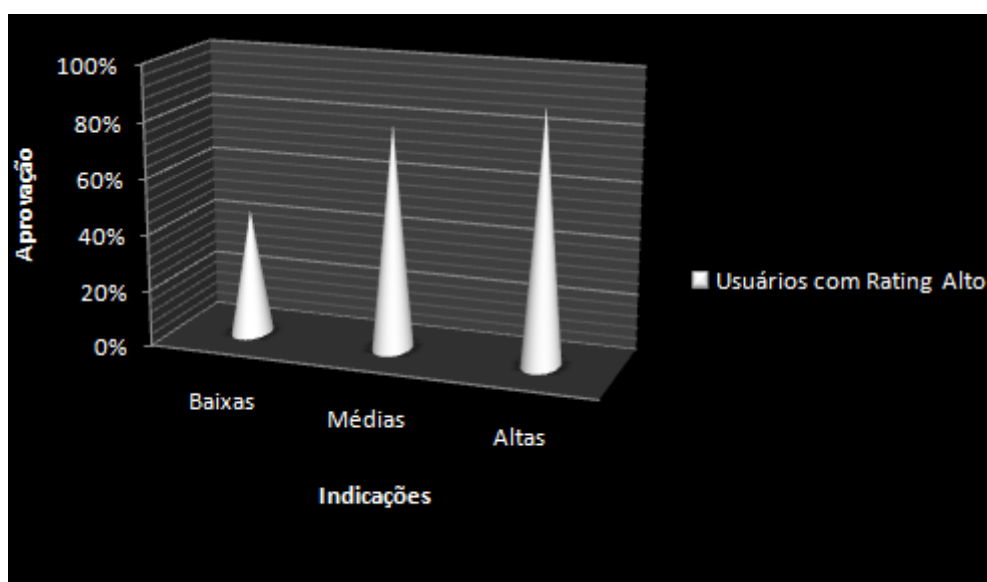
Figura 12 – Avaliação de Recomendações Aleatórias por Usuários com Ratings Médios



Fonte - Autor

Finalmente, a última classe de usuários, aqueles que possuem rating entre 4 e 5, avaliaram todas as recomendações do sistema. A expectativa de que houvessem altos índices de aprovação acabou sendo confirmada, chegando a se ter 45% de aprovações em recomendações para ratings baixos, 80% para ratings médios e 90% para ratings altos.

Figura 13 – Avaliação de Recomendações Aleatórias por Usuários com Rating Alto



Fonte - Autor

É possível concluir que, a partir dos gráficos apresentados, existe de fato uma tendência maior na aprovação de indicações que é proporcional ao rating do usuário. Entretanto, mesmo com essa tendência, é notável, no comparativo dos gráficos, que há uma inclinação maior do usuário em aprovar cada indicação quando essa foi de fato selecionada pelo sistema.

7. CONCLUSÃO

Este trabalho teve como finalidade apresentar uma pequena introdução aos conceitos dos sistemas de recomendações colaborativas e filtragem colaborativa, assim como, mostrar na prática o desenvolvimento de um sistema deste gênero, em ambiente web e com design responsivo, baseado em técnicas e algoritmos de predição largamente difundidos entre os desenvolvedores da área.

Pôde-se perceber, através dos resultados obtidos, que o aplicativo atende de forma satisfatória a uma parcela significativa dos potenciais usuários deste sistema. No entanto, para que os números futuros continuem condizentes com os já obtidos, será de fundamental importância uma classificação adequada por parte dos usuários. Todo o sucesso do sistema depende do interesse e da preocupação do usuário em associar da melhor forma possível a recomendação ao seu provável público alvo.

A principal contribuição deste trabalho não se encontra no desenvolvimento do algoritmo de análise da base de dados, uma vez que, para se alcançar os resultados esperados, em um sistema que trabalha com sentimentos humanos, não se pode esperar precisão extrema das escolhas-, ou seja, o sistema atende as expectativas dos usuários do ponto de vista de que não se pode definir o que é um sentimento exato e portanto a indicação exata para o sentimento em questão também pode apresentar incoerências. Porém, isso faz com que o usuário seja ainda mais surpreendido por essa discrepância entre o que pode ser exato e o que pode não fazer nenhum sentido. Finalmente, o que pode e deve ser tratado em trabalhos futuros é o conceito por trás do sistema e a aproximação do usuário com a finalidade dos dados que este sugere através do uso de algum mecanismo de motivação, por exemplo.

Outro fator de sucesso do sistema desenvolvido é o fato de que, durante todos os testes, não foi necessária a presença e/ou intervenção de qualquer tipo de administrador, as inserções, avaliações e recomendações fluíram naturalmente entre as bases de dados, atendendo uma das principais pretensões do sistema quando da sua projeção inicial. Ainda que isso possa sugerir uma eventual propagação de conteúdos que fogem o objetivo do sistema, o fato de o usuário possuir total liberdade para expressar suas ideias é um componente de interesse a mais para os públicos-alvo.

6 REFERÊNCIAS

- BOBADILLA, Jesús et al. Recommender systems survey. **Knowledge-Based Systems**, v. 46, p. 109-132, 2013.
- BONATTI, Denilson. **Desenvolvimento de Sites Dinâmicos com Dreamweaver CC**. Brasport, 2013.
- CELMA, Oscar. **Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space**. Springer, 2010.
- CROWTHER, Rob. **Hello! HTML5 & CSS3: A user-friendly reference guide**. Manning, 2013.
- REATEGUI, Eliseo Berni; CAZELLA, Sílvio César; OSÓRIO, Fernando Santos. Personalização de Páginas Web através dos Sistemas de Recomendação. **Tópico em Sistemas Interativos e Colaborativos**. São Carlos, 2006.

RICCI, Francesco; ROKACH, Lior; SHAPIRA, Bracha. **Introduction to recommender systems handbook**. Springer US, 2011.

SILVA, Maurício Samy. **CSS3: desenvolva aplicações web profissionais com uso dos poderosos recursos de estilização das CSS3**. Novatec Editora, 2011.

AMAZON.COM INC. **Informações corportativas**. 2014. Disponível em: <http://www.amazon.com.br/gp/feature.html/ref=gw_m_b_corporate/179-5179379-3173364?ie=UTF8&docId=5721910011>. Acesso em: 20 out.2014.

CACHEDA, Fidel et al. Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. **ACM Transactions on the Web (TWEB)**, v. 5, n. 1, p. 2, 2011.

EBAY INC. **Who We Are**. 2014. Disponível em: <http://www.ebayinc.com/who_we_are/one_company>. Acesso em: 19 out.2014.

GOOGLE CORPORATION. **Sobre o Google**. 2014. Disponível em: <<https://www.google.com.br/intl/pt-BR/about/>>. Acesso em: 18 out.2014.

LUO, Li; GAO, Xiao Ling; WU, De Jun. The Research of Cloud Platform Based on KNN Collaborative Filtering Algorithm in Fruits Marketing System. **Applied Mechanics and Materials**, v. 373, p. 1826-1829, 2013.

RONEN, Royi et al. Selecting content-based features for collaborative filtering recommenders. In: **Proceedings of the 7th ACM conference on Recommender systems**. ACM, 2013. p. 407-410.

SU, Xiaoyuan; KHOSHGOFTAAR, Taghi M. A survey of collaborative filtering techniques. **Advances in artificial intelligence**, v. 2009, p. 4, 2009.

ZERVAS, Panagiotis et al. A Responsive Design Approach for Supporting Mobile Access to Virtual and Remote Laboratories. In: **14th IEEE International Conference on Advanced Learning Technologies**. 2014.

AMAZON.COM INC. **Informações corportativas**. 2014. Disponível em: <http://www.amazon.com.br/gp/feature.html/ref=gw_m_b_corporate/179-5179379-3173364?ie=UTF8&docId=5721910011>. Acesso em: 20 out.2014.