Proposta de métricas de avaliação da qualidade da informação médica para Sistemas de Recomendação baseados no perfil do usuário

Leila Weitzel (Organizador)[1], José Palazzo Moreira de Oliveira (Orientador)[1], Francieli Zanon Boito [1], Henrique Dias Pereira dos Santos [1], Jeferson Campos Nobre [1], João Adolfo Froede Lutz [1], Julio Cesar Santos dos Anjos [1], Marcelo Corrêa Yamashita [1], Márcio Muccillo Sklar [1], Mauricio Volkweis Astiazara [1], Tiago Guimarães Moraes [1]

Resumo: A Web é uma fonte de busca onde as pessoas procuram informações sobre cuidados em saúde. Entretanto, é aberta a vários tipos de publicação e provedores de informação, portanto a qualidade das informações em saúde que são publicadas são altamente variáveis e dinâmicas. Um usuário leigo que busca informação nem sempre possui o conhecimento e educação suficientes para avaliar e validar a informação disponível. Neste relatório aborda-se um sistema de recomendação baseado no perfil do usuário e na qualidade da informação recomendada.

Abstract: The Web is an important source for people who are seeking healthcare information. However, it is open to numerous kinds of publishers and information providers, so the quality of health information published on the Web is highly variant and highly dynamic. A typical healthcare information user may lack sufficient knowledge and training to evaluate the validity and quality of the content of a Web page. In this paper we will explore

[1] Instituto de Informática, UFRGS, Caixa Postal 9999
{lwcsilva,palazzo@inf.ufrgs.br}

the development of a Recommender System that focus on the user profile and information quality.

1 Introdução

Este documento apresenta os resultados provenientes dos estudos realizados no âmbito da disciplina CMP112 no Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande do Sul (UFRGS). A disciplina foi ministrada no primeiro semestre de 2010 pelo Prof. José Palazzo Moreira de Oliveira. O objetivo (e principal atividade) consistiu no desenvolvimento de um projeto de recomendação de páginas web sobre a doença de Alzheimer baseado no perfil do usuário e na qualidade da informação recomendada.

O crescimento atual da informação disponível na web é em parte devido ao aumento da participação de seus usuários na construção de conteúdos. Na sociedade da informação a cada dia agregam-se novas páginas na web de caráter individual, de associações, de grupos, de instituições privadas, governamentais entre outras. Esse conteúdo surge de maneira acelerada e irrestrita acarretando um desenvolvimento não-ordenado e não planejado de páginas da Web. A informação científica e tecnológica, que até a pouco tempo se disseminava apenas em formatos impressos, atualmente está dispersa pela web. Por este mesmo caminho está indo a informação médica, tornando-a numerosa e variada. A evolução das técnicas de gerenciamento do excesso de informação ultrapassa os limites dos sistemas de recuperação de informação. Assim, Sistemas de Recomendação estão sendo estudados e desenvolvidos com o propósito geral de auxiliar no processo social de fornecer sugestões personalizadas de forma automática (total ou parcial) de itens de acordo com o interesse particular de um usuário (ou grupo de usuários).

Com tecnologia disponível e internet mais rápida e de fácil acesso os indivíduos estão buscando na web informações sobre doenças, tratamentos, ente outros, que antes ficavam restritas ao ambiente médico [1].

A Web cria novas oportunidades na melhoria de comunicação e tomada de decisões em saúde. Para os profissionais de saúde pode ser: (i) uma valiosa ferramenta de decisão clínica, (ii) mais um meio de comunicação pelo qual se podem trocar informações com outros profissionais de saúde e (iii) para o aperfeiçoamento continuado. Apesar de seus óbvios benefícios, o aumento da disponibilidade de informação pode resultar em muitos efeitos altamente prejudiciais colocando em risco a saúde dos pacientes [1]. Esses riscos são em parte devido às raras formas sistemáticas e automáticas que avaliem a qualidade da informação obtida [2]. Fica difícil muita vezes tomar decisões acertadas em relação aos cuidados que se deve ter com a saúde, com a presença da desinformação [3].

Os sítios médicos variam desde os altamente acadêmicos, revistas, sítios governamentais, de prestadores de serviços de saúde até as contribuições individuais dos cidadãos, pacientes e parentes de pacientes. Há também um número imensurável de sítios relacionados à indústria, que vão desde as pequenas até as grandes empresas farmacêuticas com uma infinidade de

informação comercial de produtos e serviços. Com tal diversificação de fontes de informação on-line de saúde, não está claro se os usuários sabem diferenciar o propósito, objetivos, intenções que existem entre eles. Desta forma, a preocupação com a qualidade da informação médica na web é um aspecto relevante que deve ser considerado, uma vez que não há monitoramento nem controle sobre o que é publicado [4].

Se por um lado as ferramentas de busca, como o Google ganharam grande popularidade e se tornaram a forma padrão de recuperar informações na Web, por outro, se tratam de ferramentas genéricas, que indexam páginas apenas baseadas em métricas de popularidade tal como o Page-Rank. A evolução das técnicas de gerenciamento do excesso de informação ultrapassa os limites dos sistemas de recuperação de informação.

Os maiores desafios no desenvolvimento de um SR são: a personalização da informação e a qualidade da informação recomendada. Julgar se a informação disponível online é segura e de qualidade pode ser um desafio ainda maior que a própria busca pela informação. Usuários tendem a utilizar ferramentas de busca de uso geral, as quais não possuem formas sistematizadas de avaliação de qualidade de forma automática [2].

No contexto descrito acima, os Sistemas de Recomendação atuais devem trazer para o plano frontal não só a relevância do documento recuperado de acordo com as necessidades dos usuários, mas também a qualidade do documento recuperado. A importância está em deixar os usuários mais aptos a encontrar a "melhor" opção em termos de qualidade [5]..

2 Contextualização do problema

2.1 Dominio do problema

O Mal de Alzheimer (ou Doença de Alzheimer, ou simplesmente Alzheimer) é uma doença degenerativa sem cura cujos efeitos envolvem perda de memória, confusão, irritabilidade, alterações no humor, falhas de linguagem e prejuízo das funções motoras. Como a doença é mais comum entre idosos, e os seus primeiros sintomas são geralmente confundidos com problemas de idade, o seu diagnóstico costuma acontecer anos depois do seu início. Acredita-se que 15 milhões de pessoas no mundo possuem Alzheimer [6]. A doença foi escolhida por ser de difícil e importante o diagnóstico nos estágios iniciais e pela morbidade envolvida.

2.2 Sistemas de Recomendação

Segundo Resnick e Varian [7], Sistemas de Recomendação surgiram a partir do conceito de Filtragem Colaborativa (Collaborative Filtering) que são processos de filtragem de informação ou padrões envolvendo diversos agentes ou fontes de dados. O objetivo dos Sistemas de Recomendação é inferir um conjunto de resultados relevantes ao interesse de um usuário ou um perfil ao qual se adéqüe. Essencialmente, um sistema de recomendação analisa as informações de um usuário em busca de padrões de interesse que possam ser utilizados para a recomendação de conteúdo.

Pode-se dizer que SR possuem dois principais desafios: o primeiro é descobrir informações sobre o interesse do usuário para permitir a construção de um perfil ou adequá-lo a um perfil existente; o segundo desafio é recomendar, a partir do perfil determinado para o usuário, o conteúdo com maior possibilidade de relevância.

Sistemas de Recomendação podem utilizar técnicas explícitas ou implícitas para descobrir informações sobre o usuário. Técnicas explícitas remetem ao caso em que o usuário alimenta o sistema com informações que indiquem seu perfil, como, por exemplo, o preenchimento de formulários ou registro online. Técnicas implícitas são aquelas que visam descobrir os interesses do usuário, geralmente utilizando Mineração de Uso da Web (Web Usage Mining) para analisar, por exemplo, o tempo de permanência de um usuário em um site, a sequência de sites que o usuário visita (clickstream), suas páginas favoritas, entre outros.

A definição de conteúdo para um determinado perfil envolve métodos de Filtragem de Informação (Information Filtering) cujo objetivo é selecionar informações a partir de um perfil de usuário. Em alguns casos, sistemas de filtragem de informação podem ser utilizados em conjunto com técnicas de aprendizagem de máquina para classificação das informações.

2.3 Qualidade da informação em saúde

São raras as formas sistemáticas, disponíveis aos usuários leigos, que avaliam a qualidade da informação em saúde de forma automática [2] . Ao dizer que um usuário é leigo isso significa que este não apresenta conhecimento e formação suficiente para fazer juízo de valor em relação à qualidade ou confiabilidade da informação médica na Web. Do ponto de vista do usuário leigo, existem algumas tentativas para desenvolver aplicações para avaliação da qualidade da informação em saúde de forma automática.

Eysenbach e Kholer [3] propuseram a idéia de filtragem automática através de um software residente no browser do usuário poderia filtrar automaticamente as informações utilizando metadados.

Mais tarde, Eysenbach et al [8] fizeram um levantamento dos critérios de qualidade encontrados na literatura. Verificaram a existência de 26 critérios técnicos de qualidade. Os autores Wang e Liu [2] desenvolveram uma ferramenta baseada nestes 26 critérios técnicos que são divididos em:

(i) autoridade;
(ii) fonte de informação (referências);
(iii) frequência de atualização da informação;
(iv) conteúdo editorial;
(v) propósito e objetivos do sítio;
(vi) interatividade com o usuário;
(vii) patrocinadores e intenção comercial.

Do ponto de vista do profissional de saúde existem algumas iniciativas de se estabelecer um padrão ou certificação de publicação de conteúdos em saúde na Web por empresas e órgãos governamentais. A National Library of Medicine – NLM e a National Institutes of Health propõem diretrizes para avaliação da qualidade através de um tutorial. Este tutorial na

verdade é apenas um guia que destaca alguns pontos que devem ser avaliados pelo usuário quando está buscando informações de saúde na web.

A Health On the Net (HON) Foundation concede selo de certificação de qualidade para sítios que cumprem o seu Código de Conduta (HONCode), de acordo com o modelo de reputação desenvolvido. As métricas utilizadas são:

(i) autoridade - qualificação dos autores, complementaridade - a informação deve ser complementar e não substituir as indicações médicas,

(ii) privacidade - confidencialidade dos dados submetidos ao sítio,

(iii) atribuição - deve citar a fonte da informação disponível,

(iv) justificabilidade - Todas as informações sobre os benefícios ou a realização de qualquer tratamento (médico e/ou cirúrgico), produto comercial ou serviço são considerados como créditos Todas as reclamações devem ser apoiadas por provas científicas (revistas médicas, relatórios ou outros),

(v) transparência - a informação deve ser a mais clara possível, devem disponibilizar informações de apoio e disponibilizar endereços de contato para os visitantes que procuram informações ou apoio,

(vi) apoio financeiro - Suporte para este sítio deve ser identificado claramente, incluindo a identidade das organizações comerciais e não comerciais que tenham contribuído com financiamento, serviços ou materiais para o sítio,

(vii) política de publicidade - Se a publicidade é uma fonte de financiamento esta deverá ser claramente indicada. A publicidade e outros materiais promocionais serão apresentados aos usuários de clara e que facilite a diferenciação entre ela e o material original produzido pela instituição gestora do sítio.

O Health Summit Working Group (HSWG) determina indicadores de qualidade, como por exemplo:

(i) Credibilidade - inclui a fonte, atualização periódica, pertinência/utilidade, e processo de revisão editorial para a informação;

(ii) Conteúdo - acurácia, a hierarquia de evidência, a precisão das fontes, os avisos institucionais e completude;

(iii) Divulgação - inclui informar o usuário o propósito do sítio, bem como qualquer armazenamento de informações relacionadas com o uso do sítio;

(iv) Links - avaliados de acordo com a seleção, a arquitetura, os conteúdos;

(v) Projeto visual - acessibilidade, organização navegabilidade, e capacidade de pesquisa interna;

(vi) Interatividade - inclui mecanismos de feedback e meios para o intercâmbio de informações entre os usuários- fórum de discussão;

(vii) Alerta - esclarecer se a função é a de comercializar produtos e serviços ou é fornecedor de conteúdos de informação primária;

Dentre as medidas que podem afetar a qualidade de um sítio e que devem ser avaliadas incluem:

(i) Período de atualização - se o sítio é sistematicamente atualizado,

(ii) Verificar quem é responsável por manter o sítio (Instituição publica, privada, de comércio, educação etc.),

(iii) Quem patrocina o sítio,
(iv) Objetivo e meta do sítio,
(v) Informações básicas: possui links para outros sítios ou se outros sítios o referenciam.

2.4 Perfil do usuário

Trabalhos na área de personalização da informação podem ser encontrados em Montaner et al [9], Castelano et al [10], Mencar et al [11], Ravindran e Gauch [12], Gauch et al [13], Ziegler et al [14] e Sieg et al [15]. Estas pesquisas utilizam diferentes paradigmas diferentes para categorizar usuários.

Os estudos de perfis de usuários são investigações centradas no sistema, no indivíduo, grupo ou comunidade. Tais investigações objetivam descobrir suas preferências ou hábitos.

Para estudar perfil de usuários são comuns duas técnicas para coletar informações: implícita e explícita. As informações implícitas são coletadas de maneira que o usuário não perceba que o sistema está coletando informações sobre o mesmo. Essas informações são coletadas durante a sua navegação e podem ser: tempo de visita a uma página, movimento do mouse, log de navegação, entre outros. Nas explícitas, as informações são fornecidas intencionalmente, ou seja, nela o usuário se expressa de alguma forma, por exemplo, preenchimento de formulário. Esse tipo de informação é considerada mais confiável por alguns autores, já que o usuário é quem a fornece, mas o custo desse tipo de procedimento é justamente o esforço do usuário que nem sempre está disposto a colaborar [16], [17].

Conceitos que serão adotados para descrever algumas propriedades e relacionamentos do Modulo Usuário são:

Esteriotipização: o primeiro conceito que deve ser destacado é o termo estereótipo - toma-se emprestado da Psicologia Social o conceito de Esteriotipização. Esse termo tradicionalmente tem sido adotado, nesta área, para descrever a classificação ou categorização de estereótipos e apresentá-lo como um portador de traços (atributos) intercambiáveis com outros membros de uma mesma categoria [18]. Estereótipo é um conjunto de características presumidamente partilhadas por todos os membros de uma mesma categoria. Pode envolver praticamente qualquer aspecto distintivo de uma pessoa – idade, raça, sexo, profissão, local de residência. Os estereótipos se referem a suposições sobre a homogeneidade grupal e aos padrões comuns de comportamento dos indivíduos que pertencem a um mesmo grupo.

Pesquisa Etnográfica: Etnografia é também conhecida como pesquisa social, observação participante, pesquisa interpretativa. Compreende o estudo, pela observação direta e por um período de tempo, das formas costumeiras de viver de um grupo particular de pessoas - um grupo de pessoas associadas de alguma maneira, uma unidade social representativa para estudo. A etnografia estuda preponderantemente os padrões mais previsíveis do pensamento e comportamento humanos manifestos em sua rotina diária; estuda ainda os fatos e/ou eventos menos previsíveis ou manifestados particularmente em determinado contexto interativo entre as pessoas ou grupos [19]. A pesquisa etnográfica visa minimizar as alterações mascaradas do comportamento do usuário, entretanto a pesquisa é custosa, exige uma equipe multidisciplinar e depende do esforço do usuário.

Pesquisa Demográfica: a pesquisa demográfica visa conhecer o perfil sócio-cultural do usuário. Pode ser feita através de um pequeno questionário. Escolhe-se um conjunto de perguntas representativas que permitam representar os estereótipos. O questionário tem como vantagem ser simples e de rápida aplicação e como desvantagem o esforço do usuário e a desinformação.

Estilos Cognitivos: O trabalho de Souto et al. [20] e Dias et al. [21] contribuem com a nossa pesquisa na especificação dos estereótipos. Os autores apresentam um estudo sobre modelos cognitivos de aprendizado em educação a distância. Acredita-se que o processo de busca de conceitos e informação na web pode ser comparado às abordagens de ensino-aprendizagem à distância. Modelos cognitivos são representações das facetas do raciocínio humano. De acordo Felder e Soloman [22], estilos de aprendizagem podem ser definidos como as características internas ou as preferências individuais dos aprendizes na forma de receber e/ou processar informações. Tais estilos nem sempre são conscientes e exercem influência marcante nas estratégias utilizadas para aprender. Ou seja, as conceituações de estilos cognitivos de aprendizagem indicam que as experiências às quais os indivíduos são expostos ajudam a determinar suas maneiras privilegiadas de aprender. Tem como principal desvantagem ser custosa, e exigir uma equipe multidisciplinar.

Consulta (query): a consulta formulada pode dar indícios da capacidade linguística do usuário, ou seja, do nível de compreensão de leitura. Pode inclusive auxiliar no mapeamento cultural sobre costumes e regionalismos. Tem como vantagem ser simples e de rápida análise e não depende de esforço adicional do usuário.

Análise comportamental da navegação: é feita por meio de análise de log de navegação. Depende do esforço do usuário e da coleta, e da análise e clusterização dos logs.

Com uma amostra estatisticamente significativa de usuários dispostos a participar da pesquisa, os requisitos acima seriam suficientes para que se obtivesse uma categorização confiável de estereótipos. Entretanto, nesta pesquisa não foi possível contar com o esforço de usuários. Sendo assim, optou-se por utilizar a estratégia de classificar estereótipos de forma sintética. Os estereótipos serão formados por um conjunto de perguntas que associada à query podem fornecer subsídios à classificação desses usuários.

Dentre um rol de perguntas que se pode fazer, foram selecionadas empiricamente as seguintes:

Escolaridade - fundamental, médio, superior ou pós-graduação: No caso da escolaridade ser igual à superior ou pós-graduação será pesquisada qual a área de formação. Este pergunta tem como propósito informar o nível de cultural e de conhecimento. Acredita-se que a escolaridade tem correlação direta com o nível de entendimento do usuário, ou seja, quanto maior o nível de escolaridade maior será seu nível de entendimento. A área de formação permite separar o usuário em dois grupos: leigos e profissionais da área de saúde.

Faixa etária: esta pergunta tem como objetivo verificar o nível de experiências pessoais que esse usuário tem. Acredita-se que quanto mais velho o usuário, maior é seu nível de experiência.

Objetivo da pesquisa: que tem resposta do tipo fechada (o usuário escolhe dentre um rol de respostas). Dependendo da resposta selecionada, por exemplo, ao selecionar: "métodos de profilaxia" têm-se fortes indícios de ser um profissional de saúde.

O perfil é calculado levando-se em consideração o conjunto de respostas de modo ponderado conforme ilustra a equação abaixo.

$$E = \sum_{i=1}^{n} \frac{x_i w_i}{n} \quad (1)$$

Onde:

E classe de estereótipo.

x_i é a variável escolaridade, faixa etária, objetivo e query.

w_i peso associado à variável.

n total de variáveis.

Assim, a partir de dados sintéticos os estereótipos foram classificados em: leigos, profissionais da área de saúde, acompanhante de pacientes e usuários casuais.

3 Estudo de caso

3.1 Arquitetura do sistema

O modelo do sistema possui três áreas de pesquisa distintas, são elas: Perfil de Usuário, Reputação de Páginas e Recomendação de Páginas que compõem os módulos que podem ser visto na Figura 1.

Na área de Perfil de Usuário foram pesquisadas técnicas de como categorizar usuários do sistema e quais as informações necessárias para isso, bem como qual a abordagem para obtenção destas informações. Na Reputação de Páginas, foram pesquisadas métricas de identificar páginas que apresentam qualidade de conteúdo e autoria na área da saúde. E em Recomendação de Páginas, definiu-se a lógica para recomendar páginas em função da adequação do perfil e da qualidade do conteúdo.

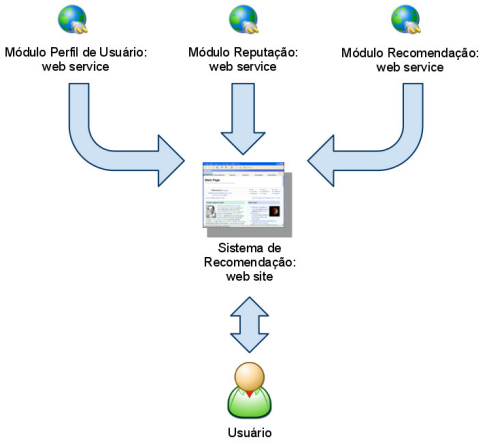

Figura 1. Arquitetura Geral do Sistema

Cada uma destas áreas descritas acima seria encapsulada em módulos independentes, expondo cada uma das suas funcionalidades através de *web services*. Assim, o sistema como um todo seria a integração destes diversos *web services*.

O sistema proposto deve recomendar páginas sobre doença de Alzheimer de acordo com o seu nível de conhecimento. Em especial a um usuário leigo que não tem a formação suficiente para fazer juízo de valor em relação à qualidade ou confiabilidade da informação médica. Trata-se de uma ferramenta de busca, em que o usuário leigo recebe como resultado apenas indicações de páginas de linguagem fácil compreensão, com poucas informações técnicas, enquanto que um especialista da área médica receberia para a mesma busca resultados com teor científico maior.

Neste primeiro protótipo, não haverá armazenamento de dados do perfil de usuários, bem como o seu feedback . A cada consulta, o usuário deverá fornecer informações para que o sistema classifique seu perfil. Os resultados da adequação da informação ao perfil do usuário serão mostrados em ordem de relevância em função da qualidade verificada.

A Figura 2 mostra apenas o caso de uso geral "FAZER UMA BUSCA" que ilustra a situação prevista em que o sistema pode ser utilizado. Neste primeiro protótipo existe apenas um caso de uso previsto, por isso é bem simples.

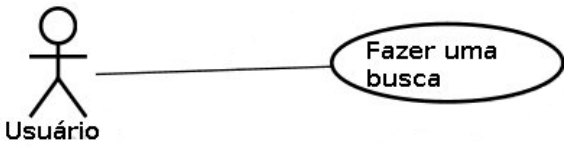

Figura 2. Diagrama de caso de uso Geral do sistema

O diagrama de sequência na Figura 3 especifica como os objetos do sistema interagem para a execução das tarefas. O diagrama dá ênfase à ordenação temporal das comunicações entre as tarefas.

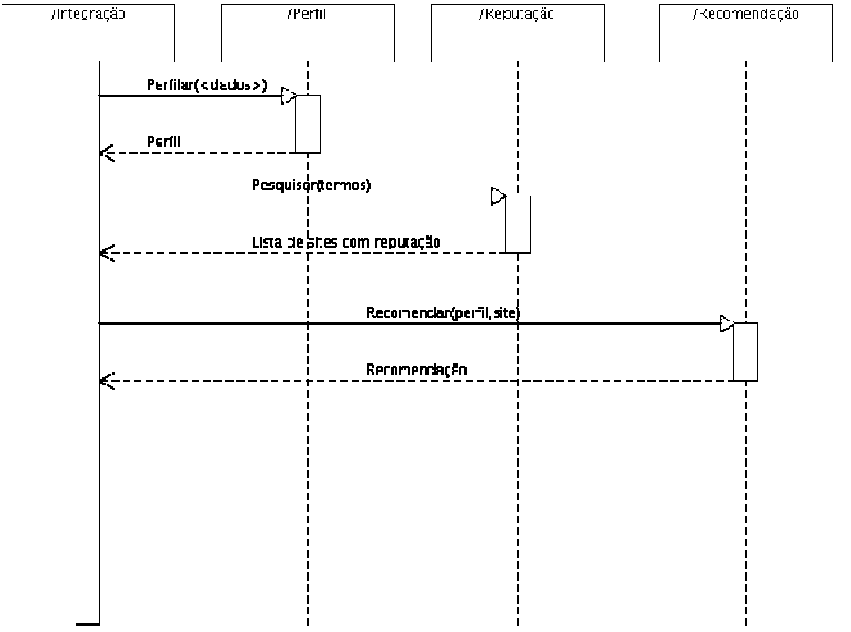

Figura 3. Diagrama de Sequência do sistema

A Tabela 1 apresenta a descrição da sequência de eventos que podem ocorrer no caso de uso "FAZER UMA BUSCA". O passo 3 é opcional, caso o usuário não forneça informações o sistema atuará conforme uma maquina de busca usual.

Tabela 1. Descrição sequência típica de eventos.

Usuário	Sistema
1. Acessa o serviço.	
	2. Mostra a interface com uma janela para realizar a busca e formulários para que o usuário informe as suas informações.
3. Informa as suas informações ao sistema (Esse passo é opcional. Caso o usuário não dê informações o bastante para que o sistema faça a sua recomendação, este funcionará apenas como uma busca simples. 4. Digita no campo apropriado o termo a ser buscado.	
	5. Busca páginas relevantes, determina o perfil de usuário, faz a combinação entre os dois para determinar o que deve ser recomendado. Mostra na tela os resultados para o usuário.
6. Utiliza os links trazidos pelo serviço, acessando páginas que contêm a informação desejada.	

O Diagrama Conceitual de Objetos representa conceitos do mundo real e as relações entre eles. Ele se concentra nas relações e atributos e não nos métodos em si, ajudando a entender a terminologia na área de domínio para qual o sistema está sendo desenvolvido.

O diagrama da Figura 4 mostra a percepção que foi dada sobre as entidades do sistema. Existem perfis e páginas, que devem ser avaliados, gerando, para cada um, uma avaliação dada pelo sistema. Essa avaliação é composta de uma série de atributos, como presença de termos técnicos para páginas e escolaridade para perfis. Os conceitos como avaliação e atributos são expandidos a fim de tornar clara a diferença existente entre avaliar usuários e páginas.

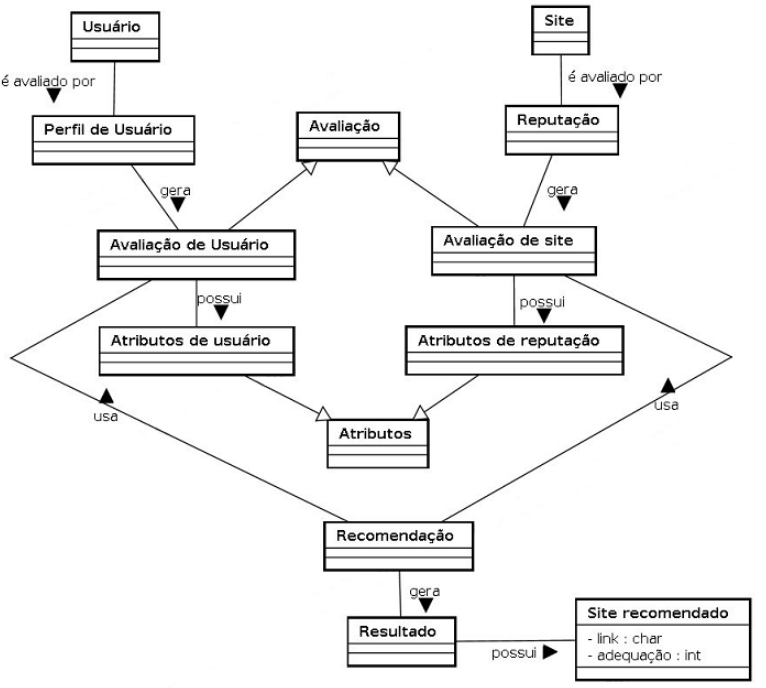

Figura 4. Diagrama conceitual de objetos do sistema

3.2 Módulo Reputação

Vários estudos encontrados na literatura abordam os conceitos, os métodos e as técnicas de recomendação, reputação (qualidade) em documentos na web. As pesquisas de [7], [23], [24], [25], [26], [27] são alguns exemplos e servem de base teórica para este estudo. Além disso, os estudos de [28] e [29] contribuem também para o referencial bibliográfico. Esses estudos fornecem métodos, técnicas e protótipos para extração automática de conteúdo e para a avaliação da qualidade de documentos na web.

Existem algumas iniciativas de se estabelecer um padrão ou certificação de publicação de conteúdos em saúde na Web por empresas e órgãos governamentais. São diretrizes baseadas em princípios básicos de ética. O critério mais citado na literatura é o HON - Health on the Net Foundation, é uma organização fundada em Genebra em 1996 com a missão de auxiliar a identificar web sites confiáveis na área de saúde. Além da HON, duas outras organizações apresentam diretrizes para a avaliação de informações em saúde, a American Medical Association – AMA, e a Health Information Technology Institute – HITI. Entretanto, não existem ainda instrumentos que tenham sido validados metodologicamente. Cabe ao usuário,

em posse desses critérios, considerar quais informações podem ser consideradas de qualidade.

Após avaliar e ponderar sobre as diretrizes estabelecidas pela AMA , HITI e HON o grupo de pesquisa propõem as seguintes métricas:

(I) PRESENÇA DE TERMOS TÉCNICOS: o texto extraído das páginas recuperadas é analisado em termos de seu conteúdo técnico. Isto é, os termos são extraídos e comparados a um dicionário técnico médico. A métrica calcula o percentual de termos técnicos presente, e se este valor for maior que um valor de threshold o texto é considerado técnico e caso contrario não técnico.

(II) PRESENÇA DE IMAGENS: Considera-se que páginas que contam com diversos recursos visuais são mais simples de entender (especialmente para usuários leigos) que as páginas que contam apenas com texto. Destaca-se que as imagens se referem a auxiliar no entendimento do conceito explicado, e não se tratam de imagens de propaganda.

(III) CARÁTER EDUCACIONAL: acredita-se que uma página que esteja vinculada a uma Instituição de ensino apresente maior reputação. Para determinar o valor relativo da métrica, verifica-se, através do domínio exibido na URL, no diretório Colleges and Universities do Google Directory se está página está vinculada. Caso seja encontrada, procura-se a instituição em um ranking de universidades disponível em Ranking Web World Universities. O resultado da métrica é dado então pela divisão da unidade pela posição no ranking. Caso não seja encontrado o registro, atribui-se o valor zero para esta métrica.

(IV) CARÁTER COMERCIAL: acredita-se que páginas que estão vinculadas às instituições comerciais podem apresentam vieses na informação. Para determinar o valor desta métrica utiliza-se a mesma estratégia definida para o caráter educacional, buscando no Google Directory a URL de instituições farmacêuticas ou similares. Caso seja encontrada, é utilizado o valor do PageRank, e caso contrario a métrica terá valor zero.

(V) REFERENCIAL TEÓRICO: acredita-se que páginas que não possuam referencial teórico dos conceitos que são descritos apresentam menor nível de confiabilidade. Esta métrica é do tipo binária onde recebe o valor um (1) caso apresente e zero (0) caso contrário.

(VI) H-INDEX DO AUTOR: havendo autoria na página recuperada, é considerado o H-index do autor do conteúdo.

(VII) PRESENÇA DE PUBLICIDADE: Acredita-se que páginas que contenham propaganda, anúncios e outras comunicações de produtos e serviços vinculados à saúde podem conter informações com viés. Por tanto, páginas que contenham este tipo de comunicação podem apresentar menor nível de confiabilidade na informação.

(VIII) ADVERTÊNCIA DE INCOMPLETUDE: a página deve conter mensagens claras alertando o usuário que a informação disponibilizada não esgota o tema abordado. Devem ser orientados a procurar fontes adicionais de informação e o seu respectivo médico para esclarecer outras duvidas.

(IX) PRESENÇA DO SELO HONCODE: é um selo de qualidade atribuído a sites que garantidamente seguem as suas recomendações. O selo é verificável, portanto, falsificações podem ser facilmente identificadas, sendo uma garantia de qualidade de informação médica.

3.3 Estratégia para avaliação manual

Para fim de validação das métricas propostas, optou-se por fazer a extração e avaliação manual de um conjunto de páginas recuperadas pelo Google na busca pelo termo "Alzheimer". O roteiro de avaliação segue as seguintes heurísticas:

(I) PRESENÇA DE IMAGENS: Conta-se o número de palavras presentes no corpo do texto da página. Assumindo que o algoritmo é capaz de descartar imagens de cabeçalho, fundo e propagandas em barras laterais, entre outros. No entanto, se houver uma imagem de propaganda inserida junto ao texto, não há distinção, é inserida na contagem, independente de ser uma imagem explicativa ou não. O valor da métrica é dado pela divisão da variável: quantidade de imagens presentes no texto pela variável número de palavras, retirar deste conteúdo as palavras sem valor semântico (stopwords).

(II) CARÁTER EDUCACIONAL: Pesquisa-se a URL da instituição fonte (se for, por exemplo, "www.inf.ufrgs.br/pessoa/contexto/disciplina/material.html", retira-se apenas a parte www.ufrgs.br para analise). Verificar no diretório Colleges and Universities do Google Directory, no seu diretório em Inglês acessar em http://www.google.com/Top/Reference/Education/Colleges_and_Universities, se a URL está presente. O resultado poderá ser:

•Se não encontrado, neste caso a métrica tem valor zero. Cabe salientar que o diretório não contém todas as universidades. A UFRGS, por exemplo, não está listada e, portanto, é um viés nesta métrica.

•Se for encontrado deve-se então procurar a posição da mesma no ranking disponível em http://www. webometrics.info. O resultado é dado pela divisão da unidade pela posição neste ranking, por exemplo = 1/800.

(III) CARÁTER COMERCIAL: De forma análoga ao item anterior, procura-se a URL da instituição fonte no diretório farmacêutico do Google Directory (acessar em http://www.google.com/Top/Business/Biotechnology_and_Pharmaceuticals/Pharmaceutical). O resultado deve ser:

•Se não for encontrado, então a métrica recebe o valor de zero

•Se encontrado então utilizar o valor do pagerank da URL

(IV) REFERENCIAL TEÓRICO: O resultado esperado é binário (1) um caso apresente referencial teórico ou (0) zero caso contrário. O algoritmo deve ser capaz de encontrar os termos (palavras) associados como por exemplo, "Fonte:"ou "Retirado de:" etc em notas de rodapé ou seguinte a notação das normas da ABNT para tal fim.

(V) H-INDEX DO AUTOR: recomenda-se o uso do addon para o navegador Firefox e o Google Reader disponível em https://addons.mozilla.org/pt-BR/firefox/addon/45283. O resultado esperado é:

•Se houver uma indicação clara do nome do autor da página ou texto, o H-index é considerado.

•Se não houver indicação, o valor dever ser -1, representando que esse atributo não deve ser considerado.

O conjunto de páginas de teste compreende uma lista de 24 páginas sobre Alzheimer, listada na Tabela 2.

Tabela 2. Lista de páginas avaliadas

Id Página	Endereço da página
1	http://www.nia.nih.gov/Alzheimers/AlzheimersInformation?/Treatment/
2	http://www.brc.cam.ac.uk/research_new/tauopathy.html
3	http://www.pharma.us.novartis.com/diseases-conditions/alzheimer-disease.jsp
4	http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp
5	http://www.labtestsonline.org/understanding/conditions/alzheimers.html
6	http://www.health.com/health/library/topic/0,,hw136623_hw136626,00.html
7	http://www.alzheimer.oxford.on.ca/index.php?menu_id=1697
8	http://www.ucl.ac.uk/news/news-articles/1001/10011401
9	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1414674/
10	http://www.mayoclinic.com/health/alzheimers-disease/DS00161/DSECTION=symptoms
11	http://www.umm.edu/patiented/articles/what_symptoms_of_alzheimers_disease_000002_5.htm
12	http://en.wikipedia.org/wiki/Alzheimer%27s_disease
13	http://dreamsinfo.com/essays/Alzheimer.htm
14	http://www.nature.com/nature/journal/v325/n6106/abs/325733a0.html
15	http://www.reuters.com/article/idUSTRE64O65G20100525
16	http://www.ygoy.com/index.php/alzheimers-disease-treatment-prevention-and-home-remedies-of-alzheimers...
17	http://www.publicaffairs.ubc.ca/2010/02/08/marijuana-ineffective-as-an-alzheimer's-treatment-ubc-vancouver...
18	http://www.medicinenet.com/script/main/art.asp?articlekey=50311
19	http://www.neurokc.com/mani2.aspx?pgID=1078
20	http://alzheimers.org.uk/
21	http://www.alz.washington.edu/
22	http://helpguide.org/
23	http://www.alzheimersprevention.org/
24	http://alzheimers.about.com/

4 Estratégia de avaliação automática

Foi desenvolvido um módulo extrator, com a tecnologia Java, para extração dos termos técnicos de 3 dicionários de dados: (i) *MedicineNet.com*, (ii) *medic8.com* e (iii) *Multilingual Glossary of technical and popular medical terms*. As palavras foram extraídas utilizando a biblioteca *HtmlParser*, que constrói uma estrutura de objetos Java a partir de código HTML, facilitando a localização de conteúdo de interesse. Abaixo, podemos ver nas Figuras 5 e 6 os trechos de código do extrator.

```
82      public void extractWords(){
83
84          for (char c = 97; c < 121; c++){
85
86          String url = "http://users.ugent.be/~rvdstich/eugloss/EN/lijst" + c + ".html";
87
88          try{
89
90              Parser parser = new Parser (url);
91              NodeList list = parser.parse (null);
92
93              PrototypicalNodeFactory factory = new PrototypicalNodeFactory ();
94              factory.registerTag (new BoldTag());
95              parser.setNodeFactory (factory);
96
97              BVisitor bVisitor = new BVisitor();
98
99              list.visitAllNodesWith(bVisitor);
100
101             extratorDAO.insertWordList(bVisitor.getWordList(), 1);
102
103         }
104         catch(ParserException e){
105             e.printStackTrace(System.out);
106         }
107         }
108
109     }
110
```

Figura 5. Trecho do método que extrai palavras de uma URL

```
14 /**
15  *
16  * @author jlutz
17  */
18 public class BulletVisitor extends NodeVisitor{
19
20      private List<String> wordList = new ArrayList<String>();
21
22      @Override
23      public void visitTag (Tag tag)
24      {
25
26          if (tag.getTagName().equals("LI")){
27              if (((Tag) tag.getChildren().elementAt(0)).getAttribute("onclick") == null){
28                  if (((Tag) tag.getParent().getParent()).getAttribute("id") == null){
29                      wordList.add(tag.getChildren().elementAt(0).getChildren().elementAt(0).getText());
30                      System.out.println (tag.getChildren().elementAt(0).getChildren().elementAt(0).getText());
31                  }
32              }
33
34          }
35
36      }
37
38      public List<String> getWordList() {
39          return wordList;
40      }
41
42      public void setWordList(List<String> wordList) {
43          this.wordList = wordList;
44      }
```

Figura 6. Trecho que localiza os termos específicos em determinada página

Após extrair os termos, o extrator deu carga em uma tabela de banco de dados modelado conforme Figura 7.

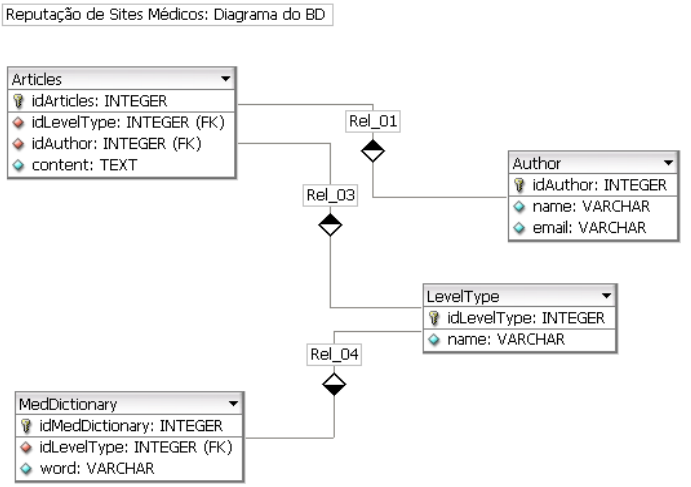

Figura 7. Modelo conceitual das classes do banco de dados gerado

```
54    public void insertWordList(List<String> words, int levelType){
55        Connection con = getConnection();
56
57        String insertString;
58
59        for (String word: words){
60
61            insertString = "INSERT INTO MedDictionary (idLevelType, word) VALUES (?,?) ";
62
63            try {
64                PreparedStatement stmt = con.prepareStatement(insertString);
65                stmt.setInt(1, levelType);
66                stmt.setString(2, word);
67                stmt.executeUpdate();
68                stmt.close();
69
70            } catch(SQLException ex) {
71                System.err.println("SQLException: " + ex.getMessage());
72            }
73
74        }
75        try{
76        con.close();
77        }
78        catch (Exception e){
79
80        }
81    }
```

Figura 8. Trecho de código Java da inserção no banco de dados

Foram extraídos 22 mil termos das páginas avaliadas (Figura 8). O cálculo da relação entre termos técnicos e não técnicos foi feita por meio de própria DML (Linguagem de

Manipulação de Dados) do SQL. A Figura 9 mostra o trecho de código que realiza este cálculo.

```
1 UPDATE Articles SET
2 words = LENGTH( content ) - LENGTH( REPLACE( content, ' ', '' ) ),
3 medWords = ( SELECT COUNT( * ) FROM MedDictionary? WHERE INSTR( content, word ) )
4
5 UPDATE Articles SET percentual = medWords/words
6
7
```

Figura 9. Código SQL que avalia o percentual de termos técnicos

Na Tabela 3 tem-se os resultados da contagem de termos técnicos presentes nas páginas analisadas. Estes valores não se apresentem em ordem crescente e sim no ordenamento do identificador das páginas.

Tabela 3. Lista dos resultados do calculo das métricas

Id página	Percentual de termos técnicos
1	0.253099
2	0.470238
3	0.48
4	0.240618
5	0.351617
6	0.160127
7	0.309353
8	0.342525
9	0.179147
10	0.479245
11	0.294606
12	0.130005
13	0.132862
14	0.280566
15	0.413105
16	0.351792
17	0.275956
18	0.222795
19	0.229557
20	0.360406
21	0.531381
22	0.20063
23	0.222518
24	0.388268

5 Avaliação das métricas

Na Tabela 4 tem-se as métricas calculadas de forma automática (calculo dos termos técnicos) e manual (todas as outras) que serão utilizadas para simulação.

Tabela 4. Tabela dos resultados após o cálculo das métricas

N.Pág.	Possui Imagens	Termos técnicos	Caráter Comercial	Possui Referências	Possui H-index	Possui Publicidade	Possui Honcode
1	0	0,253099	0	0	-1	0	0
2	0,00236	0,470238	0	0	-1	0	0
3	0	0,48	0,8	1	-1	0	0
4	0,00434	0,240618	0	0	-1	0	0
5	0	0,351617	0	0	-1	0	1
6	0,000087	0,160127	0	0	-1	0,00032	0
7	0	0,309353	0	1	-1	0	0
8	0,00143	0,342525	0	0	6	0	0
9	0	0,179147	0	1	14	0	0
10	0	0,497245	0,7	1	-1	0	1
11	0,00063	0,294606	0	1	-1	0	0
12	0,00193	0,130005	0	1	-1	0	0
13	0	0,132862	0	0	2	0	0
14	0	0,280566	0	1	18	0	0
15	0	0,413105	0	1	-1	0	0
16	0	0,351792	0	0	0	0	0
17	0	0,275956	0	0	0	0	0
18	0,00056	0,222795	0,7	1	1	0,00784	1
19	0	0,229557	0	0	0	0,00283	0
20	0,0625	0,360406	0	0	-1	0,03571	0
21	0	0,531281	0	0	-1	0	0
22	0,0015	0,20063	0	1	-1	0	0
23	4	0,222518	0	1	-1	0	0
24	0	0,388268	0	1	-1	0	1

O Sistema de Recomendação desse trabalho utilizou técnicas de aprendizagem de máquina para classificar os sites recebidos e assim retornar apenas os sites que são relevantes ao perfil. O primeiro passo para criar um modelo de classificação foi classificar empiricamente o conjunto de sites listado na Tabela 2 e a partir dele gerar um modelo que permita classificar exemplos futuros. Na Tabelas 5 tem-se os sites já classificados em: sites para leigos e sites para especialistas na área de saúde.

Tabela 5. Classificação manual dos sites

N.Pág.	URL	Perfil
1	http://www.nia.nih.gov/Alzheimers/AlzheimersInformation/GeneralInfo/	Leigo
2	http://www.brc.cam.ac.uk/research_new/tauopathy.html	Especialista
3	http://www.pharma.us.novartis.com/diseases-conditions/alzheimer-disease.jsp	Especialista
4	http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp	Leigo
5	http://www.labtestsonline.org/understanding/conditions/alzheimers.html	Leigo
6	http://www.health.com/health/library/topic/0,,hw136623_hw136626,00.html	Leigo
7	http://www.alzheimer.oxford.on.ca/index.php?menu_id=1697	Leigo
8	http://www.ucl.ac.uk/news/news-articles/1001/10011401	Leigo
9	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1414674/	Especialista
10	http://www.mayoclinic.com/health/alzheimers-disease/DS00161/DSECTION=symptoms	Leigo
11	http://www.umm.edu/patiented/articles/what_symptoms_of_alzheimers_disease_000002_5.htm	Especialista
12	http://en.wikipedia.org/wiki/Alzheimer%27s_disease	Leigo
13	http://dreamsinfo.com/essays/Alzheimer.htm	Leigo
14	http://www.nature.com/nature/journal/v325/n6106/abs/325733a0.html	Especialista
15	http://www.reuters.com/article/idUSTRE64O65G20100525	Leigo
16	http://www.ygoy.com/index.php/alzheimers-disease-treatment-prevention-and-home-remedies-of-alzheimers-disease/	Leigo
17	http://www.publicaffairs.ubc.ca/2010/02/08/marijuana-ineffective-as-an-alzheimer's-treatment-ubc-vancouver-coastal-health-research/	Leigo
18	http://www.medicinenet.com/script/main/art.asp?articlekey=50311	Leigo
19	http://www.neurokc.com/mani2.aspx?pgID=1078	Leigo
20	http://alzheimers.org.uk/	Leigo
21	http://www.alz.washington.edu/	Leigo
22	http://helpguide.org/	Leigo
23	http://www.alzheimersprevention.org/	Leigo
24	http://alzheimers.about.com/	Leigo

O software Weka foi utilizado para a criação e avaliação do modelo classificador e a partir dele foram analisados vários algoritmos de classificação. O algoritmo que melhor apresentou resultados foi o Classificador Bayesiano Ingênuo (*Naive Bayes Classifier*) que classificou 83.33% das instâncias na validação cruzada de 10 sub-conjuntos (*10-fold Cross Validation*). A taxa de acerto pode ser observada na Tabela 6 enquanto a Tabela 7 apresenta a matriz de confusão gerada pelo modelo.

Tabela 6. Performance da classificação da Rede

Naive Bayes Classifier		
Instâncias Classificadas Corretamente	20	83,33%
Instâncias Classificadas Incorretamente	4	16,67%

Tabela 7. Matriz de confusão do modelo

Matriz de Confusão		
Leigo	Especialista	← classificado como
18	1	Leigo
3	2	Especialista

Pode-se observar que apesar do modelo classificar corretamente 83.33% das instâncias a classificação não obteve um alto índice de verdadeiros positivos para o perfil especialista. Isso ocorre pelo baixo número de instâncias utilizadas para a geração do conjunto de treinamento do modelo e pode ser melhorado a partir de um conjunto maior de informações. Na Tabela 8 tem-se os resultados estatísticos dos erros da simulação, percebe-se que a simulação apresentou um erro relativo alto, o que sugere que o conjunto de sites para avaliação não foi expressivo, ou seja, não representou uma amostra representativa da população em estudo.

Tabela 8. Resultados estatísticos da simulação

Resultados da simulação	**Valor**
Erro absoluto	0.1856
Erro Relativo	53.2349 %

Com propósito de validar os resultados obtidos com o Classificador Bayesiano Ingênuo, utilizou-se a técnica de clusterização. Para a formação dos clusters foi utilizada a rede do tipo Auto-organizável (*Self-Organizing Map - SOM*) proposta por Kohonen [30].

Na primeira simulação foram submetidos à rede apenas 3 clusters, com 100 ciclos de repetição, taxa de aprendizado de 0,9 e o valor do parâmetro sigma da função de vizinhança (Gaussiana) de 0,5. Na Figura 10 tem-se os clusters formados nesta primeira simulação. Percebe-se que na amostra existe apenas um cluster bem definido.

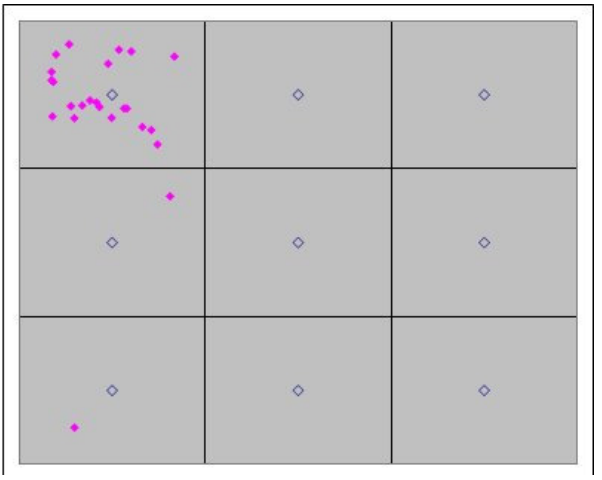

Figura 10. Resultado da simulação com 3 clusters

Na segunda simulação mantiveram-se os parâmetros: 100 ciclos de repetição, taxa de aprendizado de 0,9 e o valor do parâmetro sigma da função de vizinhança (Gaussiana) de 0,5, mas o numero de cluster utilizado foi 4. Mais uma vez foi obtido apenas 1 (um) cluster bem definido.

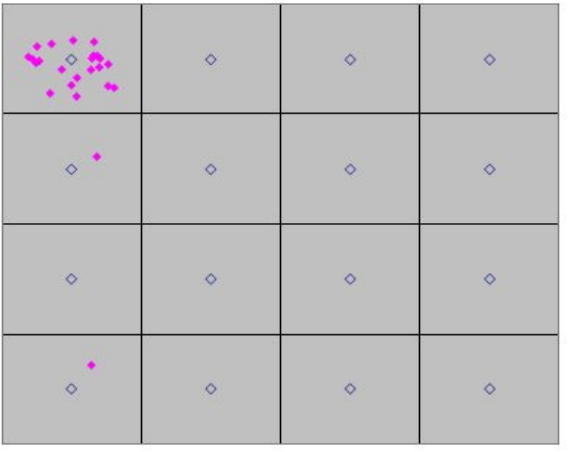

Figura 11. Resultado da simulação com 4 clusters

Baseado nos resultados obtidos no processo de clusterização, fez-se a análise multivariada dos componentes principais. Procura-se assim verificar se cada uma das variáveis pode ser

definida como uma combinação linear dos fatores comuns que irão explicar a parcela da variância de cada variável.

A parcela explicada pelos fatores comuns recebe o nome de comunalidade. Assim, os resultados da análise podem ser visto da Tabela 9. Pelo teste de corte de autovalor tem-se 5 fatores principais que explicam 86,67% da variância encontrada.

Tabela 9. Resultados da análise de fatores

Autovalores

No.	autovalor	% Individual	% Acum.	Scree Plot
1	1,561473	22,31	22,31	|||||
2	1,170902	16,73	39,03	||||
3	1,292554	18,47	57,50	||||
4	1,003177	14,33	71,83	|||
5	**1,038742**	**14,84**	**86,67**	|||
6	0,552622	7,89	94,56	||
7	0,380530	5,44	100,00	||

Na Tabela 10 tem-se o gráfico de barras das comunalidade encontradas e as respectivas variáveis que são explicadas. Por exemplo, o fator 1 pode ser explicado pelas variáveis Caráter Comercial e HONcode resultando em uma das dimensões da amostra. O teste serviu para verificar que as métricas Caráter Comercial e HONcode, Termos técnicos e H-index, Possui-Publicidade e Possui-Referências têm correlação relativamente forte com um determinado fator. Confirmando que existem 5 dimensões, conforme detectado pelo Teste do autovalor (Tabela 9). Concluímos que podemos agrupar as métricas em 5 dimensões da seguinte forma:

Fator 1 representa 22,31 % da variância total composto por 2 variáveis - Caráter Comercial e HONcode, ambas possuindo carga fatorial alta.

Fator 2 representa 16,73% da variância total composto por 2 variáveis - Possui Referências e Possui Publicidade) chamando-se a atenção que a variável Possui Referências tem uma carga fatorial baixa.

Fator 3 representa 18,47% da variância total composto por 2 variáveis - Termos Técnicos e Possui Referências, ambas possuindo cargas fatoriais alta.

Fator 4 representa 14,33% a variância total composto por apenas uma variável (H-index) com alta carga fatorial.

Fator 5, que representa 14,84% a variância total composto pela variável Possui Imagens com alta carga fatorial.

Tabela 10. Gráfico de barras das Comunalidades verificadas em relação aos fatores

Variáveis	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5
Possui Imagens	|	|	|	|	|||||||||||||||
Termos Técnicos	|	||	||||||||||||	|	|
Caráter Comercial	||||||||||||||	|	|	|	|
Possui Referencias	||	|||||	||||||||||	|	|
H_index	|	|	|	|||||||||||||||||	|
Possui Publicidade	|	||||||||||||||||	|	|	|
HONcode	||||||||||||||	|	|	|	|

6 Discussão

A análise de formação de cluster permitiu que avaliássemos as métricas propostas sob o ponto de vista de categorização de perfis de usuários. O resultado mostrou que não foi possível a separação em perfis apenas com o conjunto de métricas propostas. Esses resultados leva-nos a crer que, na amostra existe apenas um grupo e este grupo com fortes indícios de ser o perfil profissional. Além disso, as métricas mostraram-se correlacionadas à qualidade e não ao nível de entendimento do usuário.

A segunda avaliação realizada, a analise de componentes principais, mostrou que algumas variáveis possuem a mesma contribuição (influência) nos fatores. Isto é, pode-se diminuir a dimensão do espaço de variáveis sem se perder generalização.

E por fim, vale ressaltar que ao se recomendar informações em determinados domínios deve-se atentar ao valor dado quando instâncias são classificadas de maneira incorreta. Ou seja, analisar os falso positivos, pois no domínio médico a informação incorreta pode ter consequências desastrosas para saúde em geral.

7 Referências

[1] RISK, A.; DZENOWAGIS, J. Review Of Internet Health Information Quality Initiatives, Journal of Medical Internet Research., v. 3(4), n. E28, 2001. Disponível em: <http://www.jmir.org/2001/4/e28/>. Acesso em: Maio de 2010.

[2] WANG, Y.; LIU, Z. Automatic detecting indicators for quality of health information on the Web. International Journal of Medical Informatics, v. 76, n. 8, p. 575 - 582, 2007

[3] EYSENBACH, G.; KOHLER, C. How do consumers search for and appraise health information on the World Wide Web? British Medical Journal, 324, 573-577, 2002.

[4] ANDERSON, J. G. Consumers of e-health: Patterns of use and barriers. Social Science Computer Review, 22, 242–248, 2004.

[5] O'DONOVAN, J.; SMYTH, B. Trust in recommender systems. In: Proceedings of the 10th international conference on Intelligent user interfaces. Anais... . p.167-174, 2005. San Diego, California, USA: ACM.

[6] AZEVEDO, J. R. D. Doença de Alzheimer: O que há de novo?. Disponível em: <http://www.saudevidaonline.com.br/artigo101.htm>. Acesso em Abril de 2010.

[7] RESNICK P.; VARIAN H. R. "Recommender Systems". Communications of the ACM. Volume 40, issue 3, pages 56 – 58, 1997.

[8] EYSENBACH, G.; DIEPGEN, T. L.; GRAY, J. A. M.; et al. Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. BMJ, v. 317, n. 7171, p. 1496-1502, 1998. Disponível em: <http://www.bmj.com>.

[9] MONTANER, M.; et al., A Taxonomy of Recommender Agents on the Internet. Artificial Intelligence Review. 2003: Kluwer Academic Publishers.

[10] CASTELLANO, G.; FANELLI, A. M.; MENCAR, C.; TORSELLO, M. A. Similarity-Based Fuzzy Clustering for User Profiling. In: Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops. Anais... p.75-78, 2007. IEEE Computer Society .

[11] MENCAR, C.; TORSELLO, M. A.; DELL'AGNELLO, D.; CASTELLANO, G.; CASTIELLO, C. Modeling User Preferences through Adaptive Fuzzy Profiles. In: Proceedings of the 2009 Ninth International Conference on Intelligent Systems Design and Applications. Anais... . p.1031-1036, 2009. IEEE Computer Society.

[12] RAVINDRAN, D.; GAUCH, S. "Exploting hierarchical relationships in conceptual search," in Proceedings of the 13th International Conference on Information and Knowledge Management, ACM CIKM 2004, Washington DC, November 2004

[13] GAUCH, S.; CHAFFEE, J. e PRETSCHNER, A. Ontology-Based User Profiles for Personalized Search, Integrated Series in Information Systems, chapter 24, vol.14, pp 665-694, 2007. Springer US.

[14] ZIEGLER, C.; SIMON, K. e LAUSEN G., "Automatic computation of semantic proximity using taxonomic knowledge," in Proceedings of the 15th ACM International Conference on Information and Knowledge Management, CIKM 2006, Arlington, VA, November 2006, pp. 465–474.

[15] SIEG, A.; MOBASHER, B. e BURKE, R. Learning Ontology-Based User Profiles: A Semantic Approach to Personalized Web Search, IEEE Intelligent Informatics Bulletin, Vol.8 No.1 November 2007, pp 7-18.

[16] MIDDLETON, S. E. Capturing knowledge of user preferences with recommender systems, 2003. Hampshire - United Kingdom: UNIVERSITY OF SOUTHMAPTON. Disponível em: <http://eprints.ecs.soton.ac.uk/7857/1/Thesis-final-lowres.pdf>. Acesso em: 31/5/2010.

[17] CLAYPOOL, M.; LE, P.; WASEDA, M.; BROWN, D. Implicit Interest Indicators. Proceedings of ACM Intelligent User Interfaces Conference (IUI), p. 33-40, 2001. Santa Fé, NM, USA. Disponível em: <http://web.cs.wpi.edu/~claypool/papers/iii/>. Acesso em: 5/4/2010

[18] MOSKOWITZ, G . (2006). Social Cognition. New York: Guilford Press.

[19] WIELEWICKI, V. H. A pesquisa etnográfica como construçàao discursiva. Acta Scientiarum, Maringá, v. 23, n. 1, p. 27–32, 2001.

[20] SOUTO, M.; VERDIN, R.; WAINER, R.; et al . Towards an Adaptive Web Training Environment Based on Cognitive Style of Learning: An Empirical Approach. Adaptive Hypermedia and Adaptive Web-Based Systems, 2006.

[21] DIAS, C. C. L.; KEMCZINSKI, A.; GASPARINI, I. Identificação dos estilos cognitivos de aprendizagem através da interação em um Ambiente EAD. [WEI] - Anais do Workshop sobre Educação em Informática, 2009. Bento Gonçalves- RS.

[22] FELDER, R. M.; SOLOMAN, B. A. Learning Styles and Strategies. . Disponível em: <http://www4.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/styles.htm>. Acesso em: 14/6/2010.

[23] PERUGINI, S. ; M. A. GONÇALVES, et al. Recommender Systems Research: A Connection-Centric Survey. J. Intell. Inf. Syst., v.23, n.2, p.107-143. 2004.

[24] HOFMANN, T. Latent semantic models for collaborative filtering. ACM Trans. Inf. Syst., v.22, n.1, p.89-115. 2004.

[25] KONSTAN, J. A. Introduction to recommender systems: Algorithms and Evaluation. ACM Trans. Inf. Syst., v.22, n.1, p.1-4. 2004.

[26] AGGARWAL, C.C.; WOLF, J.L.; WU, K. e YU, P.S. (1999). Horting Hatches an Egg: A Graph-Theoretic Approach to Collaborative Filtering. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'99) (pp. 201–212). San Diego, CA: ACM Press.

[27] RESNICK, P.; KUWABARA, K. et al. Reputation systems. Commun. ACM, v.43, n.12, p.45-48. 2000.

[28] LICHTNOW, D. et al, Relato e considerações sobre o desenvolvimento de uma ontologia para avaliação de sites da área de saúde, Cadernos de Informática, v.4 n. 01, PP 07-46, 2009.

[29] FLEISCHMANN, A. M. P et al, Relato sobre o desenvolvimento de modelos para obtenção automática do conteúdo de sites sobre saúde, Cadernos de Informática, v. 4 n.01 pp 47-101, 2009.

[30] KOHONEN, T. Self Organizing Maps; Springer Series in Information Sciences Springer: Espoo, Finland, 1994.