

Conhecendo o usuário através de enriquecimento semântico para apoio à aprendizagem informal em saúde

Jonathan Darlan Cunegundes Moreira¹, Francisco Milton Mendes Neto¹, Ricardo Alexsandro de Medeiros Valentim², Alisson Alan Lima da Costa¹, Ênio Lopes Sombra¹, Argemiro Soares de Aliança Neto¹

¹Programa de Pós-Graduação em Ciência da Computação (PPgCC), Departamento de Ciências Exatas e Naturais (DCEN)
Universidade Federal Rural do Semi-Árido (UFERSA) – Mossoró, RN – Brasil

²Laboratório de Inovação Tecnológica em Saúde (LAIS)
Universidade Federal do Rio Grande do Norte (UFRN) – Natal, RN - Brasil
{jdarlan, miltonmendes}@ufersa.edu.br, ricardo.valentim@ufrnet.br,
{alisson, enio}@ufersa.edu.br, argemiront@gmail.com

Abstract. *Informal learning may be useful in healthcare, once recent research shows that one factor which influence to medical treatment's non-adherence by patients with chronic diseases is knowledge lack about their health condition. Digital media favors the spread of knowledge, but the data amount on the web can make hard to individual find the appropriate knowledge. Recommendation systems can provide personalized content to users. However they need to know users well. This paper proposes to determine users' interests about health subjects through their daily actions, using automated techniques of semantic augmentation.*

Resumo. *A aprendizagem informal pode ser útil na área de saúde, uma vez que pesquisas recentes mostram que um dos fatores que influenciam a não adesão ao tratamento médico por portadores de doenças crônicas é a falta de conhecimento acerca de sua condição de saúde. Os meios digitais favorecem a propagação deste conhecimento, porém a quantidade de dados na web pode tornar difícil para o indivíduo encontrar a informação adequada. Sistemas de recomendação podem prover conteúdo personalizado para usuários, necessitando, para isso, conhecê-los. Este trabalho propõe determinar os interesses dos usuários sobre temas de saúde, por meio de suas ações cotidianas, utilizando técnicas automatizadas de enriquecimento semântico.*

1. Introdução

Ao contrário da aprendizagem formal, onde o conteúdo é selecionado e classificado por um professor/facilitador [Green 2001], na informal o conhecimento é adquirido pela experiência diária do usuário e pelo compartilhamento de informações em suas relações sociais [Jiugen et al. 2011]. Uma das áreas onde este tipo de aprendizagem tem grande potencial de aplicação é na área de Saúde. Pesquisas mostram que um dos fatores que

influenciam a não adesão ao tratamento médico por portadores de doenças crônicas é a falta de conhecimento acerca da sua doença e da sua condição de saúde (IBGE, 2009).

Neste cenário, a aprendizagem informal pode auxiliar promovendo a educação sobre temas relacionados à saúde fora do ambiente hospitalar. Isto, aliado ao uso de tecnologias móveis, remete ao conceito de saúde 2.0, que visa estimular o indivíduo a ser responsável pelos cuidados relativos à sua saúde através da utilização de ferramentas da Web 2.0 [Hughes et al. 2008]. Ambientes web possibilitam o compartilhamento de experiências cotidianas e a disseminação do conhecimento. Porém, ao mesmo tempo, fornecem acesso a uma ampla quantidade e diversidade de conteúdos, o que pode acarretar na dispersão dos indivíduos.

Uma forma de contornar esta situação é recomendando conteúdo de forma personalizada, levando em consideração as características particulares de cada usuário. No entanto, para que Sistemas de Recomendação (SR) [Da Silva et al. 2011] possam ser eficientes, eles necessitam conhecer os usuários e estes nem sempre estão aptos ou dispostos a fornecerem informações.

Diante disto, é fundamental o uso de técnicas que, a partir de dados captados das interações diárias do usuário, possibilitem extrair informações relevantes relacionadas à sua saúde pessoal. Contudo, estas não são fáceis de serem recuperadas. Normalmente estão dispersas em diferentes formatos, ferramentas e em ambientes com domínio não definido.

Este artigo se propõe a determinar, por meio de enriquecimento semântico, os interesses do usuário em assuntos relacionados à sua saúde, visando prover aprendizagem informal. Para isso, realiza a contextualização dos Traços Digitais (TD) do usuário com base em ontologias de domínios conhecimento, através de Processamento de Linguagem Natural (PLN). Estas relações permitem que agentes computacionais façam inferências sobre o perfil semântico do usuário possibilitando conhecê-lo sem a necessidade de sua intervenção direta.

As técnicas propostas neste artigo possibilitam a personalização de serviços ao usuário por meio das informações resultantes de suas interações. Estas informações serão utilizadas pelo MobiLEHealth, que é um ambiente de aprendizagem informal no contexto de saúde 2.0 [Mendes Neto et al, 2014]. O sistema de enriquecimento semântico irá processar os dados capturados por este ambiente e auxiliar no processo de recomendação de conteúdos mais adequados ao usuário.

2. Aprendizagem Informal, Saúde 2.0 e Sistemas de Recomendação

A aprendizagem informal é o processo contínuo de aquisição do conhecimento por um indivíduo, sendo este, responsável pelo seu aprendizado e desenvolvimento, que ocorre através de suas experiências cotidianas [Jiugen et al. 2011]. O contexto diário do individuo contribui fortemente para sua aprendizagem, podendo ocorrer através de atividades diárias, das interações sociais, por iniciativa própria, pela reflexão e pelos próprios erros [Castleton 2006].

Segundo Redecker et al. (2010), as tecnologias móveis e as mídias digitais irão ter um forte impacto sobre a aprendizagem informal. Elas representam as experiências do mundo real do usuário por meio do conteúdo gerado por este e as interações com

outros usuários. Isto faz com que sejam ambientes favoráveis à propagação do conhecimento.

Diante disto, o conceito de saúde 2.0 surgiu da necessidade de estimular o cidadão a ser ativo e responsável pela sua saúde e pelos cuidados relacionados a ela [Hughes et al. 2008]. Para isso, promove iniciativas mediadas pelas tecnologias da web 2.0 (portais, fóruns, redes sociais, registros pessoais de saúde, blogs, vídeos, serviços, etc.) para a construção coletiva do conhecimento na saúde [Fernandez-Luque et al. 2010].

Estas iniciativas favorecem a melhoria na qualidade de vida, em uma perspectiva voltada para a atuação preventiva, tais como esclarecimento de dúvidas, acompanhamento de novidades sobre doenças crônicas, facilitação na comunicação com o médico e entre outros. Porém, algumas pessoas ainda não sentem a internet como um instrumento totalmente confiável em razão da procedência e qualidade das informações que são fornecidas. Aliado a isto, a quantidade e a diversidade das informações disponíveis dificultam ainda mais a busca do conhecimento.

Os sistemas de recomendação podem minimizar este problema por meio da recomendação personalizada de conteúdo, considerando as características particulares do usuário e as suas interações cotidianas [Da Silva et al. 2011]. Uma vez que possuam metadados suficientes sobre o conteúdo e o usuário, eles são capazes de analisar as relações existentes e selecionar conteúdos que se adequem às necessidades do usuário.

3. Enriquecimento Semântico

Uma das principais dificuldades dos sistemas de recomendação é a definição correta dos metadados a serem utilizados, que normalmente é feita de forma manual ou semiautomática. Em ambientes de aprendizagem formal, um professor/facilitador seleciona e classifica os conteúdos que serão destinados a um grupo de alunos de um mesmo nível [Green 2001]. Porém, na web os conteúdos estão dispersos sem um domínio de conhecimento definido e há uma grande diversidade de usuários. Estes com interesses e necessidades diferentes e que nem sempre estão aptos ou dispostos a fornecerem informações sobre eles.

Diante disto, surge a necessidade de vincular estes conteúdos a um domínio de conhecimento. Esta contextualização deve acontecer de forma automática para que permita o processamento e entendimento do conteúdo por agentes computacionais. Para isso, podem ser utilizadas técnicas de enriquecimento semântico, que consiste na anexação de conceitos semânticos a partes específicas de um texto, provendo uma estrutura para interpretação automática de seu significado através do uso de ontologias [Zapater e Mendes Neto 2014].

Contudo, para que possam ser aplicadas estas técnicas para mapear perfil de usuário de forma implícita, é necessário conhecer o que ele faz, acessa, compartilha e com quem interage no seu cotidiano. Os TDs são os rastros deixados das interações dos usuários em ambientes virtuais. Eles evoluem como consequência dos ambientes em que estão incorporados e estabelecem uma relação com o mundo real do usuário, representando seus pontos de vistas, interesses, experiências e emoções [Dspotakis et al. 2011]. A análise do TD possibilita estabelecer relações geralmente difíceis de serem

estudadas em ambientes convencionais, possibilitando a medição do comportamento humano coletivo com base em conjuntos de grandes sistemas sociais [Kleinberg 2008].

Para realizar o processamento semântico do perfil do usuário a partir dos TDs, é necessário realizar a identificação e o mapeamento de termos-chave a partir do conteúdo textual [Baldan e Menezes 2012]. Estes termos-chaves são semanticamente associados a conceitos oriundos de uma ontologia, permitindo a compreensão do domínio do TD [Thakker et al. 2012]. Como normalmente os TDs são escritos em linguagem natural, técnicas de PLN são necessárias. Estas técnicas buscam converter ocorrências das linguagens naturais em representações manipuláveis por programas de computador. Para isso analisam os radicais e origem das palavras para determinar e encontrar suas variações, como conjugações de verbos, flexões de adjetivos, graus nos substantivos, etc. Além de isolar frases e períodos para buscar as relações sintáticas e semânticas [Russell e Norvig 2009].

Através do PLN, agentes computacionais podem criar relações semânticas da parte textual de um TD com ontologias de domínio, além de estabelecer relação com os usuários. Isto permite que motores de inferências extraiam informações relativas sobre os usuários.

4. Trabalhos Relacionados

O presente trabalho trata do enriquecimento semântico de perfis de usuários no contexto da saúde, de forma a favorecer a aprendizagem informal. Neste contexto, podemos citar quatro subdivisões da pesquisa usadas para atingir o objetivo proposto: (i) semântica em sistemas de informação na área da saúde, (ii) enriquecimento semântico, (iii) semântica para perfis de usuários e (iv) conhecimento informal em mídias sociais. Todas elas possuem uma ampla área de pesquisa com diversos trabalhos desenvolvidos, dentre os quais podemos destacar alguns.

Nguyen et al. (2009) propõe o desenvolvimento de um modelo PHR (*Personal Health Record*) em forma de ontologia, onde são descritos conceitos tais como, informações demográficas, vacinas, alergias, diagnósticos, procedimentos e medicamentos. Este PHR visa à integração com os sistemas públicos de saúde podendo fornecer informações relativas ao paciente aos profissionais da saúde.

Dung e Kameyama (2007) propõem o enriquecimento semântico pela extração de informações de saúde baseado em ontologias. O sistema varre informações publicadas na internet e, através de inferências provê novos elementos semânticos.

Stan et al. (2008) propuseram um modelo de perfil de usuário usando ontologia. Este modelo permite aos usuários ter uma rede social gerando categorias específicas de pessoas conforme suas similaridades.

Thakker et al. (2012) apresentam uma abordagem para agregar conteúdo de espaços sociais em um navegador de dados enriquecido semanticamente, de modo a facilitar a aprendizagem informal em domínios mal definidos.

Diferentemente dos trabalhos apresentados, esta pesquisa apresenta uma abordagem para determinar os interesses dos usuários, por meio de inferências semânticas, em um ambiente de aprendizagem informal para o contexto da saúde. Para

isso utiliza um conjunto de conceitos e técnicas para gerar relações semânticas entre usuários e domínios de conhecimento.

5. MobiLEHealth

O trabalho proposto é parte do MobiLEHealth, que é um ambiente de aprendizagem informal destinado a pessoas portadoras de doenças crônicas e em tratamento domiciliar. Este promove o conhecimento sobre a doença e, conseqüentemente, uma melhoria na qualidade de vida destas pessoas. Para isso, utiliza a recomendação personalizada de conteúdo a partir do monitoramento dos usuários por meio dos seus dispositivos móveis.

Um modelo abstrato da arquitetura é apresentado na Figura 1, que mostra os seus componentes (Serviços Web, Dispositivo Móvel, Enriquecimento Semântico, Repositório Semântico e Recomendação de Conteúdo), envolvendo três áreas de conhecimento principais: (i) monitoramento contínuo de usuários, (ii) enriquecimento semântico de perfil de usuário e de conteúdos, e (iii) recomendação inteligente de conteúdos.

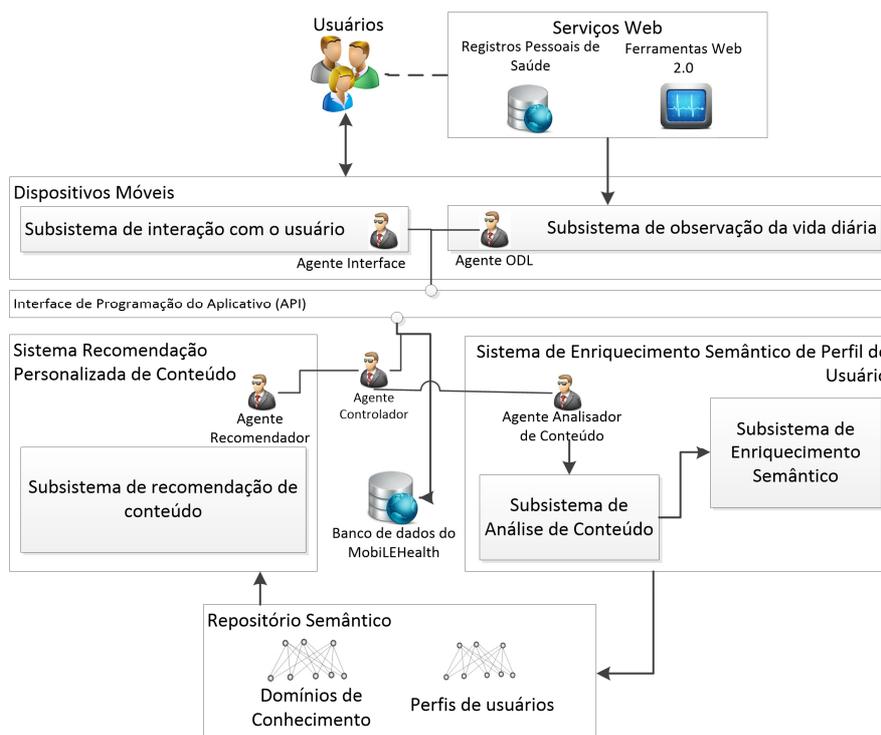


Figura 1 – Modelo da Arquitetura do MobiLEHealth

A arquitetura foi desenhada de forma modular, delegando atividades e funções específicas para cada componente, que são coordenados por um sistema multiagente (SMA) do tipo comunidade de especialistas [Bellifemine et al. 2007]. Cada um dos agentes encontra-se no mesmo nível, sendo cada um deles especialista em determinada tarefa, onde i) o Agente de Interface é responsável pelas ações que exigem interação com o usuário, ii) o Agente ODL (do inglês, *Observations of Daily Living*) trata do monitoramento e a captura de informações contextuais do usuário, iii) o Agente Analisador de Conteúdo é responsável pelo processamento do enriquecimento

semântico, iv) o Agente recomendador é responsável pela seleção dos conteúdos adequados a serem recomendados baseado no repositório semântico e no perfil do usuário, e o v) Agente Controlador (ou Agente DF, do inglês *Directory Facilitator*) fornece uma arquitetura do tipo “quadro-negro” mediando a comunicação entre os agentes.

6. Sistema de Enriquecimento Semântico de Perfil do Usuário

Para que o MobiLEHealth recomende conteúdo relevante ao usuário de forma mais precisa, é necessário que exista um mecanismo que possibilite avaliar os interesses do usuário de forma implícita. O sistema de enriquecimento semântico de perfil do usuário visa suprir esta necessidade.

O sistema proposto apresenta uma arquitetura (Figura 2) modular fornecendo interfaces que o integra ao MobiLEHealth de forma coesa, isolando o componente e garantindo a consistência do seu funcionamento. Ele tem a função de enriquecer semanticamente o perfil de usuário com os dados coletados pelo (i) subsistema de interação com usuário, (ii) subsistema de observação da vida diária, e (iii) com base nos domínios de conhecimentos registrados no repositório semântico.

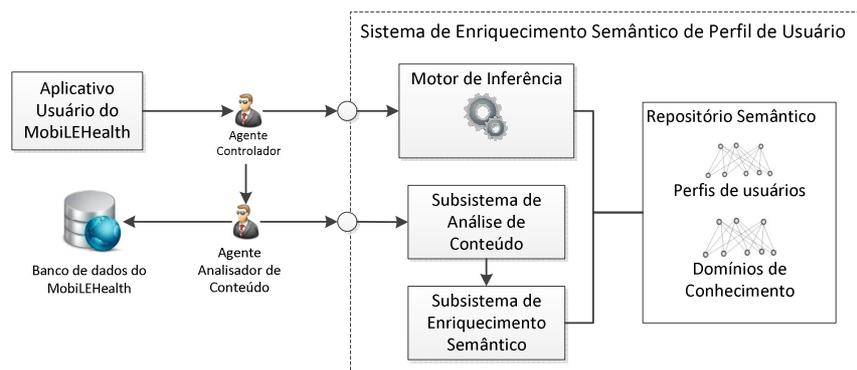


Figura 2 – Arquitetura do Sistema de Enriquecimento Semântico

O processo pode ser iniciado pelo agente controlador, por solicitação do aplicativo usuário para um recurso específico. Ou pelo agente analisador de conteúdo, que varre a base de dados do MobiLEHealth procurando recursos que ainda não foram semanticamente enriquecidos. O agente analisador de conteúdo realiza um pré-processamento do recurso para verificar sua integridade e se é compatível com as funcionalidades do sistema antes de ser enviado para o processamento semântico.

O subsistema de análise de conteúdo realiza dois tipos de tarefas conforme o recurso a ser processado. Na primeira, inicialmente analisa elementos textuais por meio de PLN eliminando conteúdos desnecessários, como, por exemplo, registros de linguagens de marcação. Posteriormente analisa morfológica e sintaticamente o restante do conteúdo. A segunda tarefa é realizada quando o recurso for um dado bruto oriundo da base do MobiLEHealth. Neste caso o processamento é feito levando em consideração não só o valor do recurso, como também a que ele se refere. O resultado do processamento deste subsistema é uma estrutura de dados a ser utilizada pelo enriquecedor semântico.

Esta estrutura é enviada ao subsistema de enriquecimento semântico que, com base nestes dados, realiza uma busca por termos semelhantes nas ontologias de domínio. As correspondências encontradas são armazenadas em estruturas de dados específicas nas ontologias de perfil. O conteúdo é processado com todos os domínios do sistema, podendo gerar relações semânticas com um ou vários domínios.

O Repositório Semântico fornece ontologias de domínio, que representam a modelagem dos conceitos relacionados às doenças crônicas que compõem o sistema, e as ontologias de perfil de usuário. A independência entre a ontologia de domínio e as marcações semânticas existentes na ontologia de perfil do usuário é outro fator importante da arquitetura. Isto garante a portabilidade das ontologias e possibilita estender o sistema a novos domínios relacionados à saúde.

O Motor de Inferência é responsável por responder as solicitações oriundas do MobiLEHealth, que são coordenadas pelo Agente Controlador. Estas solicitações podem ser (i) para um usuário, onde o sistema deverá determinar quais os interesses relacionados à sua saúde ou (ii) para um conteúdo, definindo a que domínio de conhecimento se refere.

Estas informações fornecerão suporte ao módulo de recomendação personalizada de conteúdo do MobiLEHealth, que as utilizará para auxiliar na seleção de conteúdos relevantes aos interesses do usuário.

7. Resultados

A validação da pesquisa foi realizada a partir de dados simulados com objetivo de avaliar o desempenho do sistema em um ambiente controlado. Foram utilizadas como domínios da aplicação as doenças crônicas Diabetes e a Esclerose Lateral Amiotrófica (ELA). Também foram criadas ontologias de domínio para cada uma delas e selecionados manualmente, pela internet, 100 (cem) conteúdos para cada um dos assuntos: diabetes, ELA e diversos.

Com base nos domínios e nos assuntos dos conteúdos, foram traçados quatro perfis de usuários considerando o seu interesse: diabetes, ELA, diabetes e ELA, nenhum. Para cada perfil foram simulados 50 (cinquenta) usuários e para cada usuário 30 (trinta) acessos a conteúdos, totalizando 200 (duzentos) perfis de usuários e 6.000 (seis mil) acessos. Para simular os perfis, os acessos do usuário foram distribuídos conforme apresentado na Figura 3.

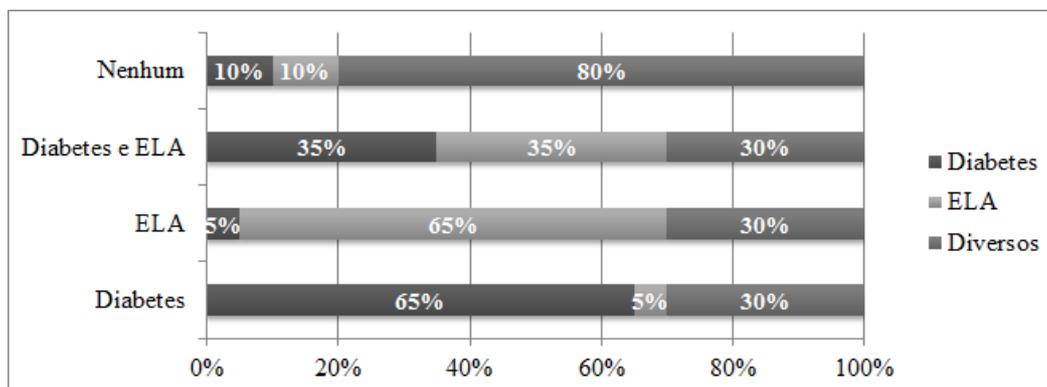


Figura 3 – Distribuição de recursos ao usuário por Perfil x Assunto

Para verificar a eficácia do sistema de enriquecimento, foram realizadas validações do índice de interesse do usuário. Após alguns testes foi considerado que os índices iguais ou acima de 30% representam que o usuário possui algum interesse no domínio. Abaixo disso não possui. Para este artigo foi avaliado apenas a existência ou não do interesse do usuário no domínio de conhecimento. Não foi considerado o grau do interesse do usuário, que seria a variação do índice.

Os resultados foram validados com base nas métricas, verdadeiro positivo (VP) que representa se o índice corresponde ao recurso simulado e verdadeiro negativo (VN) que representa se o índice não corresponde ao recurso simulado. Por exemplo, se um conteúdo é relacionado à diabetes e tiver relação com domínio de diabetes é considerado como VP, porém se este mesmo conteúdo tiver relação com o domínio da ELA é considerada como VN. O índice de eficácia é avaliado pela representatividade do VP em relação ao total (VN / VP+VN).

Foi analisado o índice de cada usuário para cada um dos domínios, contabilizando a quantidade de VP e VN. Obteve-se como resultado final uma média de 86,54% de acerto ao determinar qual o interesse do usuário. O detalhamento dos resultados para cada um dos perfis pode ser observado na Figura 4.

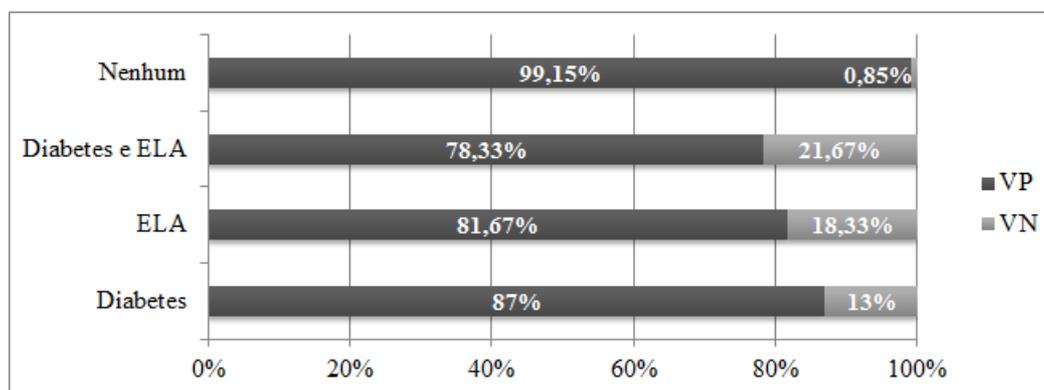


Figura 4 – Resultados por perfil de usuário.

As validações feitas em um ambiente controlado mostraram uma elevada capacidade do sistema em determinar interesses do usuário a partir de traços digitais oriundos de domínios não definidos. Esta relação é importante quando consideramos ambientes de aprendizagem informal, que exige uma necessidade de relacionar milhares de conteúdos oriundos de diversas origens a uma conceptualização formal, possibilitando o direcionamento de conteúdo relevante ao usuário.

8. Considerações Finais e Trabalhos Futuros

Neste artigo foi apresentado um sistema para enriquecimento semântico do perfil de usuário para determinar os interesses do usuário como apoio a um ambiente de aprendizagem informal no contexto da saúde, o MobiLEHealth.

Os resultados demonstraram que através de um perfil semântico do usuário, gerado a partir de seus traços digitais, é possível determinar os interesses deste relacionados à sua saúde. Isto fornece aos sistemas de recomendação a capacidade de recomendar conteúdo mais relevante ao usuário considerando o seu contexto diário, percebendo suas mudanças comportamentais e se adaptando a elas.

Como trabalhos futuros, pretende-se (i) validar a eficácia da variação dos índices dos usuários, (ii) integrar ao MobiLEHealth e verificar o uso pelo sistema de recomendação e (iii) realizar um estudo de caso real com usuários portadores de doenças crônicas. O objetivo é comprovar que os índices fornecidos pelo sistema proposto possibilitam um bom desempenho na recomendação personalizada de conteúdos e, conseqüentemente, na aprendizagem sobre saúde pessoal de forma flexível e ubíqua.

Referências

- Baldan, M.A., and Menezes, C.S. (2012). “Um Ambiente para Construção de Perfis a Partir de Textos Pessoais”. Anais do Simpósio Brasileiro de Informática na Educação 23.
- Bellifemine, F.L., Caire, G. e Greenwood, D. (2007). “Developing Multi-Agent Systems with JADE”. Hoboken, NJ: Wiley.
- Castleton, G., Gerber, R. e Pillay, H. (2006). “Improving Workplace Learning: Emerging International Perspectives”. New York: Nova Science Pub Inc.
- Da Silva, L.C.N., Mendes Neto, F.M., e Jácome Júnior, L. (2011). MobiLE: Um ambiente Multiagente de Aprendizagem Móvel para Apoiar a Recomendação Sensível ao Contexto de Objetos de Aprendizagem. Anais do XXII Simpósio Brasileiro de Informática na Educação.
- Dspotakis, D., Thakker, D., Lau, L. e Dimitrova V. (2011). “Capturing the semantics of individual viewpoints on social signals in interpersonal communication,” in Journal of Web Semantics, Special Issue on Personal and Social Semantic Web.
- Dung, T.Q. e Kameyama, W. (2007). “A Proposal of Ontology-based Health Care Information Extraction System: VnHIES”. In 2007 IEEE International Conference on Research, Innovation and Vision for the Future, pp. 1–7.
- Fernandez-Luque, L., Karlsen, R., Krogstad, T., Burkow, T.M. e Vognild, L.K. (2010). “Personalized health applications in the Web 2.0: The emergence of a new approach”. In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1053–1056.
- Green, L. (2002). “How Popular Musicians Learn: A Way Ahead for Music Education” Aldershot, Hants ; Burlington, VT: Ashgate Pub Ltd.
- Hughes, B., Joshi, I. e Wareham, J. (2008). “Health 2.0 and Medicine 2.0: Tensions and Controversies in the Field”. Journal of Medical Internet Research 10, e23.
- Jiugen, Y., Ruonan, X. e Xiaoqiang, H. (2011). “Constructing informal learning mode based on social software”. In 2011 6th International Conference on Computer Science Education (ICCSE), pp. 1227–1230.
- Kleinberg, J. (2008). “The Convergence of Social and Technological Networks”. Commun. ACM 51, 66–72.
- Mendes Neto, F.M., Costa, A.A.L., Sombra, E.L., Moreira, J.D.C., Valentim, R.A.M., Samper, J.J., Nascimento, R.P.C. e Flores, C.D. (2014). “Content’s Personalized Recommendation for Implementing Ubiquitous Learning in Health 2.0”, Latin

- America Transactions, IEEE (Revista IEEE America Latina), (Accepted for Publication).
- Nguyen, M.T., Fuhrer, P. e Pasquier-Rocha, J. (2009). “Enhancing E-health Information Systems with Agent Technology”. *Int. J. Telemedicine Appl.* 2009, 1:1–1:13.
- Redecker, C., Ala-Mutka, K. e Punie, Y. (2008) “Policy Brief: Learning 2.0 – The Impact of Social Media on Learning in Europe”. *JCR Technical Notes*, JCR56958. Disponível em: <<http://ftp.jrc.es/EURdoc/JRC56958.pdf>>. Acesso em: 12 dez. 2013.
- Russell, S. e Norvig, P. (2009). “Artificial Intelligence: A Modern Approach”. Third Edition. Upper Saddle River: Prentice Hall.
- Stan, J., Egyed-Zsigmond, E., Joly, A. e Maret, P. (2008). “A User Profile Ontology For Situation-Aware Social Networking”. 3rd Workshop on Artificial Intelligence Techniques for Ambient Intelligence.
- Thakker, D., Despotakis, D., Dimitrova, V., Lau, L. e Brna, P. (2012). “Taming Digital Traces for Informal Learning: A Semantic-driven Approach”. In *Proceedings of the 7th European Conference on Technology Enhanced Learning*, Berlin, Heidelberg: Springer-Verlag, pp. 348–362.
- Zapater, J.J.S. e Mendes Neto, F.M. (2014). “Uso de tecnologías semánticas en diferentes dominios de aplicación: Entorno educativo y sistemas de información de tráfico vial”. Saarbrücken: Editorial Académica Española.