

SUCUPIRA: a System for Information Extraction of the Lattes Platform to Identify Academic Social Networks

Alexandre Donizeti Alves, Horacio Hideki Yanasse

Laboratório de Computação Aplicada - LAC
Instituto Nacional de Pesquisas Espaciais, INPE
São José dos Campos, Brasil
alexdonizeti@gmail.com, horacio@lac.inpe.br

Nei Yoshihiro Soma

Divisão de Ciência da Computação
Instituto Tecnológico de Aeronáutica, ITA
São José dos Campos, Brasil
soma@ita.br

Abstract—Today, the Lattes Platform is undoubtedly the major source of information on Brazilian researchers. Hence, it has a very high potential for extracting information. This paper describes "SUCUPIRA", a system for information extraction of the Lattes Platform to identify academic social networks.

Keywords: information extraction; academic social network; Lattes Platform

I. INTRODUÇÃO

O crescimento explosivo e a popularidade da Web têm resultado em uma grande quantidade disponível de fontes de informação na Internet. O crescimento acelerado de dados que são tornados públicos em formato digital na Web faz com que seja necessário, igualmente, desenvolver ferramentas capazes de obter automaticamente informações de interesse do usuário. Isto traz grandes desafios na elaboração de metodologias eficazes para pesquisa, acesso e integração de informação [1].

No Brasil, a Plataforma Lattes (PL) é um sistema de informação desenvolvido e implantado pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) para gerenciar informações relacionadas a pesquisadores e instituições do país [2]. A PL é hoje, sem dúvida, a principal fonte de informações sobre os pesquisadores brasileiros e tem um elevado potencial para extração de informação. Este fato foi, inclusive, recentemente reconhecido em artigo publicado na revista Nature [3]. O artigo cita a PL como um exemplo de banco de dados completo e altamente qualificado.

Como exemplos de utilização da PL, poder-se-ia extrair informação para ajudar o Governo Federal a mapear informações sobre áreas do conhecimento específicas e, desta forma, possibilitar um investimento direcionado de acordo com a região em que se encontram ou mantêm vínculo funcional os principais pesquisadores daquelas áreas. Pode também ajudar a responder questões do tipo: o ambiente em que um pesquisador está pode influenciar em sua carreira profissional e acadêmica? O impacto que o trabalho conjunto com algum pesquisador que tem atuação destacada aumenta as chances de sucesso em uma carreira acadêmica?

Descobrir os relacionamentos existentes entre os pesquisadores, entretanto, não é uma tarefa trivial, uma vez que

não existem instrumentos que permitam que isso seja feito de maneira simples e imediata, e sem o auxílio de desenvolvedores experientes.

Um sistema automático poderia ajudar a descobrir esses relacionamentos que, muitas vezes, não são aparentes. Além disso, questões como as apresentadas a seguir poderiam ser igualmente respondidas para grupos de pesquisadores:

- Qual a relação entre os pesquisadores?
- Em que região do país estão? Em que estado? Em qual instituição?
- Quantas publicações em revistas? Em congressos?
- Como é a regularidade e a evolução das publicações?

Esses são alguns dos objetivos do sistema SUCUPIRA, que é um protótipo para extração de informações da PL para identificação de redes sociais acadêmicas.

Este artigo está organizado da seguinte maneira: na Seção II é apresentada uma visão geral da PL, dando ênfase nas ferramentas existentes e em desenvolvimento para extração de informações, além de destacar alguns trabalhos relacionados à extração de informação da PL. Na Seção III é apresentado o sistema SUCUPIRA, destacando sua arquitetura e as principais funcionalidades. Finalmente, na Seção IV são apresentadas as conclusões e trabalhos futuros.

II. A PLATAFORMA LATTES

A PL é mantida pelo Governo Brasileiro e inclui sistemas de informação, bancos de dados e portais Web. O principal componente da plataforma é o sistema Currículo Lattes, que é um sistema de informação curricular.

O sistema Currículo Lattes armazena mais de 2.000.000 de currículos de pesquisadores, docentes, estudantes e profissionais das diversas áreas do conhecimento que atuam em ciência, tecnologia e inovação, principalmente no Brasil.

O *currículo Lattes* ou *CV Lattes* é um documento criado pelo CNPq com o objetivo de padronizar e centralizar as informações pessoais, profissionais e acadêmicas da comunidade científica brasileira. Através do sistema Currículo

Lattes é possível consultar essas informações a qualquer momento e de maneira muito simples via Web. Os dados dos currículos são preenchidos pelo próprio profissional e têm sido utilizados como fonte de informações por agências de fomento no país para avaliação de pesquisadores, projetos, programas de pós-graduação etc. Com isso, este banco de dados é atualizado sistematicamente. Além disso, a própria comunidade científica monitora a qualidade e fidelidade das informações contidas no sistema, uma vez que na disputa por recursos, os currículos de cada profissional são utilizados para comparação. Percebe-se, portanto, que esse sistema tem um elevado potencial para extração de informação.

Apesar do sistema Currículo Lattes ser público e acessível via Web (<http://lattes.cnpq.br/>) há uma restrição de acesso. Atualmente, para consultar um currículo é necessário informar um código alfa-numérico, cuja finalidade é evitar buscas automáticas por *scripts*. A seguir são apresentadas algumas ferramentas que permitem extrair informações desse sistema.

A. Ferramentas para Extração de Informação

Na revisão da literatura efetuada, foram encontradas duas ferramentas que permitem a extração de informações a partir de currículos Lattes: Lattes Extrator e scriptLattes.

Lattes Extrator foi desenvolvida pelo próprio CNPq e é uma das ferramentas que compõem a PL. É acessível via Web (<http://lattesextrator.cnpq.br/lattesextrator/>), sendo que seu acesso é restrito. Atualmente, somente instituições licenciadas podem extrair informações diretamente do banco de dados de currículos Lattes do CNPq e somente informações de seus próprios pesquisadores, docentes, estudantes e colaboradores. As informações extraídas são disponibilizadas em arquivos no formato XML e as instituições podem desenvolver rotinas para a importação dessas informações para as suas próprias bases. As extrações são feitas em lote e podem ser configuradas de acordo com o interesse e as permissões de cada usuário.

scriptLattes é um *script* desenvolvido em Perl para a extração e compilação de produções bibliográficas, produções técnicas, produções artísticas, orientações, participação em bancas examinadoras, participação em comissões julgadoras, eventos, grafos de colaborações e mapa de pesquisa de um conjunto de pesquisadores cadastrados na PL [4].

Na literatura consultada não foi encontrada nenhuma biblioteca que extrai informações de currículos Lattes e/ou identifica relacionamentos entre eles. Por isso, atualmente está em desenvolvimento, pelos autores do presente artigo, uma biblioteca denominada LattesMiner. É uma biblioteca orientada a objetos para extração de informações de currículos Lattes e identificação de redes sociais acadêmicas. A biblioteca é composta por um conjunto de classes escritas em Java que permite que outros desenvolvedores implementem as suas próprias aplicações [5].

A biblioteca LattesMiner é composta por seis componentes: Aquisição de Dados, Descoberta de Dados, Extração de Dados, Estruturação de Dados, Visualização e Análise de Dados. Uma visão geral da arquitetura de componentes da biblioteca LattesMiner é ilustrada na Figura 1. A seguir é apresentada

uma breve descrição da biblioteca LattesMiner, destacando seus principais componentes.

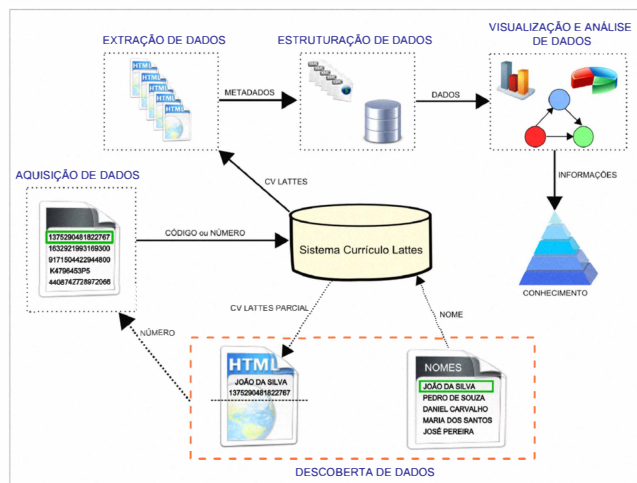


Figura 1. Arquitetura de componentes da biblioteca LattesMiner

O componente “Descoberta de Dados” é usado para descobrir o número (ID) dos pesquisadores. Cada CV Lattes tem uma URL que permite acessá-lo diretamente. Usualmente, apenas o nome do pesquisador está disponível e o sistema Currículo Lattes não permite realizar a busca automática por nome. A URL é composta por um número de 16 dígitos (e.g. <http://lattes.cnpq.br/6187221670775160>). Com esse número (ID), é possível acessar um determinado currículo automaticamente quantas vezes forem necessárias sem restrição e, portanto, pode ser acessada por mecanismos de busca automáticos. A biblioteca LattesMiner permite acessar um CV Lattes sem o problema de restrição de acesso.

Para o componente “Descoberta de Dados” é retornada apenas a parte inicial do CV Lattes para verificar se o nome contido no currículo é igual ao nome procurado. O resultado desse componente pode ser usado como entrada para o componente “Aquisição de Dados”. Este componente é responsável por baixar os currículos Lattes dos pesquisadores a partir do sítio do CNPq e gravá-los como arquivos HTML.

O componente “Extração de Dados” é o principal componente da biblioteca LattesMiner. Este componente é responsável por extrair os dados dos arquivos HTML. Para a extração de dados em um CV Lattes a partir do arquivo HTML foi usada a técnica de extração de informação baseada em expressões regulares. Foi observado que trechos de código no arquivo HTML do CV Lattes têm uma estrutura de repetição, ou seja, têm a mesma formatação HTML [6][7]. Os dados extraídos podem ser armazenados em XML ou em um banco de dados qualquer usando o componente “Estruturação de Dados”.

O componente “Visualização de Dados” é responsável pela identificação e visualização de redes sociais acadêmicas. A identificação dessas redes sociais é feita verificando os relacionamentos entre os pesquisadores. E o componente “Análise de Dados” é responsável pela análise dos dados extraídos e também pela análise dos relacionamentos identificados.

A seguir compara-se a nova ferramenta com as anteriores da literatura.

B. Comparação entre as Ferramentas

Lattes Extrator tem a vantagem de extrair os currículos diretamente do banco de dados da PL e são extraídos como arquivos XML. As ferramentas scriptLattes e LattesMiner extraem os currículos a partir do sistema Currículo Lattes disponível na Web. Nessas ferramentas, os currículos são extraídos como páginas HTML, o que dificulta a extração de informações.

Um dos problemas da scriptLattes é que as páginas são geradas em HTML/JSP, o que obriga o usuário ter um servidor Web instalado e devidamente configurado para executar páginas dinâmicas em Java. Além disso, a ferramenta é restrita ao sistema operacional Linux. A biblioteca LattesMiner é implementada em Java, podendo ser usada em qualquer plataforma e sistema operacional.

A principal vantagem da biblioteca LattesMiner é o fato de ser uma biblioteca que permite aos desenvolvedores programarem suas próprias aplicações, ao contrário das ferramentas Lattes Extrator e scriptLattes. Essas ferramentas realizam busca pelo número (ID) do pesquisador, enquanto a biblioteca LattesMiner também permite a busca pelo nome do pesquisador. Dessa forma, é possível buscar por qualquer nome citado em um CV Lattes, aumentando assim o número de relacionamentos identificáveis.

C. Trabalhos Relacionados

Nos últimos anos, muitos trabalhos foram realizados usando os dados disponíveis na base da PL e por pesquisadores das mais diversas áreas do conhecimento [8][9][10][11]. Alguns desses trabalhos analisaram o perfil dos bolsistas de produtividade em pesquisa em áreas como Saúde Coletiva [12] e Odontologia [13][14]. Essas análises foram feitas usando informações contidas nos currículos Lattes dos pesquisadores dessas áreas específicas.

Em muitos destes trabalhos tanto os dados relativos aos currículos Lattes quanto às informações neles contidas foram obtidas de maneira manual. Além disto, quando os currículos são extraídos no formato XML, usando a ferramenta Lattes Extrator, há restrições, pois somente instituições licenciadas têm acesso e, conforme mencionado anteriormente, somente para os currículos da própria instituição. Análises que envolvam múltiplas instituições e/ou pesquisadores de outras instituições tornam o processo extremamente moroso por não ser feito automaticamente. Mesmo a repetição da análise para o mesmo conjunto de pesquisadores, apenas com variação temporal é também morosa.

Portanto, um sistema que automatize todo esse processo, permitindo refazer o processo quantas vezes forem necessárias é bastante bem vindo.

Na próxima Seção é apresentado o sistema SUCUPIRA, destacando a sua arquitetura e as principais funcionalidades.

III. O SISTEMA SUCUPIRA

A biblioteca LattesMiner faz parte de um projeto maior denominado “Sistema Unificado de Currículos e Programas: Identificação de Redes Acadêmicas - SUCUPIRA”, financiado pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

O projeto SUCUPIRA, cujo acrônimo traz a lembrança do sobrenome do falecido Professor Emérito da Universidade Federal do Rio de Janeiro, Newton Lins Buarque Sucupira e relator do importante Parecer 977/65 sobre a Pós-Graduação, visa ser uma ferramenta computacional automatizada e de domínio público que pode eventualmente auxiliar a obtenção de indicadores de desempenho de docentes, pesquisadores e programas de pós-graduação. A principal ferramenta desse projeto é o sistema SUCUPIRA, que será apresentado a seguir.

O SUCUPIRA é um sistema de extração de informações da PL para identificação de redes sociais acadêmicas. O protótipo do sistema será em breve acessível via Web e de acesso público. No momento, um vídeo está disponível em <http://www.lac.inpe.br/~projetosucupira/cisti2011/demo.html>, demonstrando as principais funcionalidades do protótipo. A Figura 2 ilustra a página inicial do sistema SUCUPIRA que exibe a localização geográfica atual do usuário.

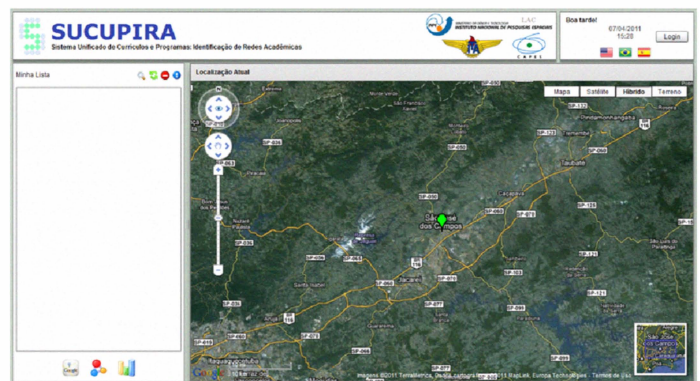


Figura 2. Página inicial do sistema SUCUPIRA

A. Arquitetura

O sistema SUCUPIRA está sendo implementado usando a tecnologia Adobe Flex, que suporta o desenvolvimento de aplicações ricas para Internet e compatíveis com várias plataformas. Para acessar o SUCUPIRA é necessário que o usuário tenha no seu navegador Web o Adobe Flash Player instalado, aplicativo pelo qual são visualizados as aplicações Flex.

No lado servidor está sendo usada a linguagem Java. Para que o Flex consiga se integrar perfeitamente com Java, é necessário ter um *gateway* que possa converter os tipos de dados nativos do Flex (especificamente, do ActionScript) para os tipos de dados nativos do Java e vice-versa. Para isso é usado o AMF (*Action Message Format*), que é um protocolo de especificação aberta, compacto e que trafega os dados em formato binário. Há várias implementações que suportam o AMF. No sistema SUCUPIRA foi utilizado o *gateway* BlazeDS [15] que foi desenvolvido em Java e é de código aberto.

A seguir é apresentada uma descrição da arquitetura do sistema SUCUPIRA, ilustrada pela Figura 3, destacando os principais componentes.

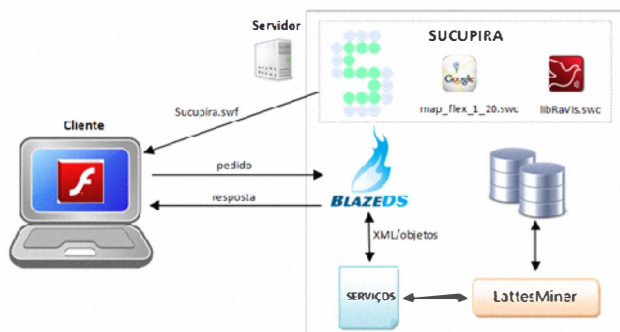


Figura 3. Arquitetura do Sistema SUCUPIRA

Quando um usuário acessa a página do SUCUPIRA, basicamente é retornado o arquivo Flash do sistema (Sucupira.swf). O sistema faz uso da biblioteca do *Google Maps* para Flash [16] para exibir a localização atual do usuário e para exibir a distribuição geográfica dos pesquisadores que fazem parte da lista desse usuário. Também faz uso do componente *RaVis* (*Relational Analysis Visualization*) [17] da biblioteca *BirdEye* [18] para exibir os grafos de relacionamentos entre os pesquisadores.

Em seguida, o sistema fica aguardando as solicitações do usuário. Toda solicitação é atendida pelo gateway *BlazeDS* que, por sua vez, encaminha a solicitação para uma classe escrita em Java que implementa os serviços oferecidos. Cada serviço oferecido é implementado por um método dessa classe, que faz acesso à biblioteca *LattesMiner*, responsável por extrair as informações do CV Lattes dos pesquisadores. As informações extraídas são armazenadas em um banco de dados qualquer que pode ser definido no momento da implantação do sistema. Isso é possível porque a biblioteca *LattesMiner* possui um arquivo texto de propriedades que permite configurar qual gerenciador de banco de dados será usado, podendo ser alterado facilmente a qualquer instante.

A comunicação entre o gateway *BlazeDS* e a classe de serviços é feita através de objetos Java ou através de documentos XML. Por fim, o resultado é enviado ao cliente e as informações apresentadas ao usuário na forma de mapas, grafos de relacionamentos, gráficos ou tabelas. Essas funcionalidades são apresentadas a seguir.

B. Principais Funcionalidades

Uma das principais funcionalidades do sistema SUCUPIRA é gerenciar uma lista de pesquisadores definida por cada usuário do sistema. A ideia é que o usuário adicione na sua lista, chamada no sistema de “Minha Lista”, os pesquisadores que ele deseja comparar e analisar.

Inicialmente, o usuário deve fazer o login no sistema. Caso seja o seu primeiro acesso, é necessário se cadastrar. Para isso, basta informar o login, o nome, a senha e se o usuário tiver uma webcam na sua máquina é possível capturar sua foto.

Após o login, é permitido realizar a busca por pesquisadores, bastando clicar no ícone de lupa do componente

“Minha Lista” do sistema. Feito isso, é exibida a janela de “Busca por Pesquisadores”, ilustrada parcialmente pela Figura 4.



Figura 4. Janela para Busca de Pesquisadores

Na janela de “Busca por Pesquisadores”, o usuário deve informar o nome do pesquisador e clicar no ícone de lupa para realizar a busca. Se o pesquisador for encontrado, são exibidas diversas informações que permitem ao usuário confirmar se é realmente o pesquisador procurado. Caso o seja, o usuário pode clicar no ícone de adicionar para inseri-lo na sua lista. Feito isso, todos os dados desse pesquisador são extraídos e armazenados no banco de dados; o componente “Minha Lista” do usuário é automaticamente atualizado e um novo pesquisador pode ser procurado.

Para demonstrar as principais funcionalidades do sistema SUCUPIRA foram escolhidos, de forma aleatória, seis pesquisadores, todos Bolsistas de Produtividade em Pesquisa do CNPq de nível 1A. Esses pesquisadores foram adicionados ao componente “Minha Lista”, conforme ilustra a Figura 5. Os três primeiros pesquisadores na lista são da área de Medicina e os outros três são da área de Ciência da Computação.



Figura 5. Distribuição Geográfica dos Pesquisadores

A Figura 5 também ilustra a distribuição geográfica dos pesquisadores que fazem parte do componente “Minha Lista” do usuário. Essa funcionalidade pode ser acionada clicando no ícone de mapa. Com isso, é possível visualizar no mapa onde estão trabalhando esses pesquisadores, uma vez que a localização é baseada no endereço profissional informado no CV Lattes de cada pesquisador.

Também é possível visualizar o gráfico de publicações desses pesquisadores, conforme ilustra a Figura 6. Essa funcionalidade é acionada clicando no ícone de gráfico, sendo possível aumentar a área de visualização do gráfico, ocultando o componente “Minha Lista”.

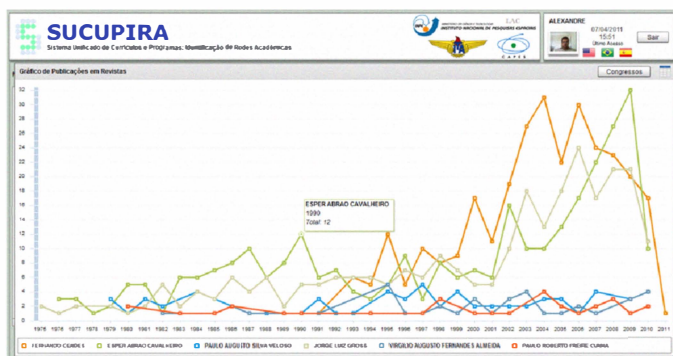


Figura 6. Gráfico de Publicações em Revistas

O gráfico apresentado é referente às publicações dos pesquisadores em revistas. Para visualizar o gráfico de publicações em congressos basta clicar no botão “Congressos”. Para visualizar qualquer um dos gráficos na forma de tabela, basta clicar no ícone e a tabela correspondente é exibida. A Figura 7 ilustra a tabela de publicações em revistas e a Figura 8 ilustra a tabela de publicações em congressos.

Nome	TOTAL	1971	1975	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
ESPER ABRÃO CAVALHEIRO	294	3	3	1	2	5	5	1	9	5	7	8	19	6	4	12	8	7	4	3	2
FERNANDO CENDES	289																1	1	6	5	12
JORGE LUZ GROSS	267	2	1	2	2	1	2	5	2	4	3	6	4	5	2	5	5	6	6	5	5
PAULO AUGUSTO SILVA VELOSO	52				3	1	3	2		4			1	1	1	1	3	1	1	1	4
PAULO ROBERTO FREIRE CUNHA	31					2				1		1	2								
VRIGLIO AUGUSTO FERNANDES ALMEIDA	81							1	1	1	2	1		1		1					6

Figura 7. Tabela de Publicações em Revistas

Nome	TOTAL	1971	1975	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	
ESPER ABRÃO CAVALHEIRO	0																				
FERNANDO CENDES	91																				
JORGE LUZ GROSS	6																				
PAULO AUGUSTO SILVA VELOSO	621	1	1	3	3	1	1	1	6	11	12	13	11	4	6	4	6	2	1	2	1
PAULO ROBERTO FREIRE CUNHA	179					2	2	2	1	5	4	4	1	4	2	8	4	7	3	3	3
VRIGLIO AUGUSTO FERNANDES ALMEIDA	59																				3

Figura 8. Tabela de Publicações em Congressos

Analisando essas tabelas, é possível perceber que os três primeiros pesquisadores, todos da área de Medicina, praticamente não publicam em congressos. Por outro lado, o número de publicações em revistas é significativo. Já os pesquisadores da Ciência da Computação, apesar de também publicarem em revistas, publicam mais em congressos.

A principal funcionalidade do SUCUPIRA é a visualização das redes sociais acadêmicas identificadas entre os pesquisadores. Esta funcionalidade pode ser acionada clicando no ícone . As redes sociais são representadas no sistema na

forma de grafos. Atualmente, apenas um tipo de grafo está disponível. Esse grafo é definido verificando os contatos (links para outros currículos Lattes) contidos no CV Lattes de cada pesquisador. Todo contato contém o número (ID) do pesquisador, o que permite identificar os relacionamentos entre os pesquisadores.

A Figura 9 ilustra o grafo de contatos entre os seis pesquisadores, destacados em vermelho. Esse grafo foi exibido definindo o grau de separação entre os vértices igual a 2. A Figura 10 exibe o mesmo grafo considerando o grau de separação igual a 1. Também é possível trocar o tipo de layout, além de outros controles que são fornecidos pelo componente *RaVis* e são usados no sistema.

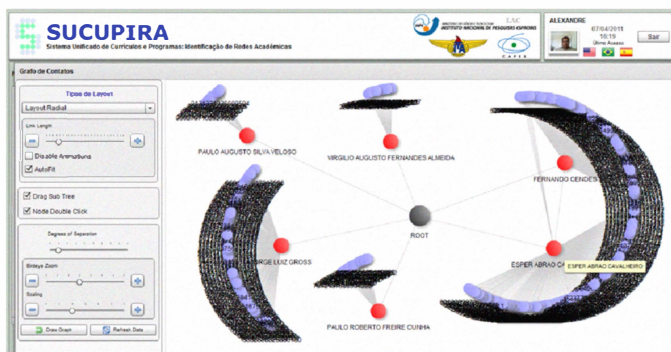


Figura 9. Grafo de Contatos de grau 2 de separação

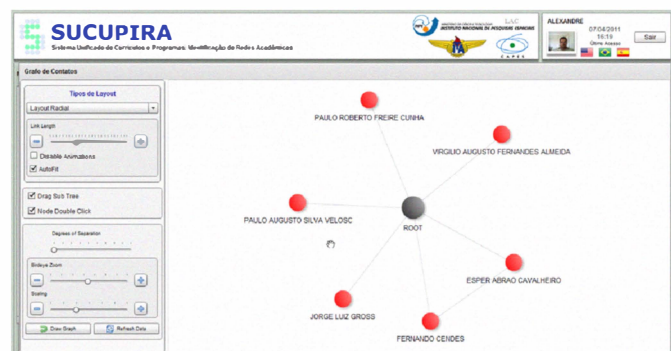


Figura 10. Grafo de Contatos de grau 1 de separação

No grafo da Figura 10 observa-se um relacionamento entre dois pesquisadores, ambos da área de Medicina, sendo que um trabalha na Universidade Federal de São Paulo (UNIFESP) e o outro na Universidade Estadual de Campinas (UNICAMP).

C. Trabalhos Relacionados

Atualmente, agências públicas de fomento como o CNPq e a CAPES se empenham cada vez mais em oferecer ferramentas que forneçam informações sobre as pesquisas no Brasil, desde investimentos financeiros até informações sobre pesquisadores em geral.

Recentemente, o CNPq disponibilizou mais uma ferramenta, denominada Painel Lattes [19], que também compõe a PL. A ferramenta oferece dados estatísticos sobre as pesquisas no Brasil e sua evolução nos últimos anos. As informações são extraídas da base de currículos da PL e apresentadas em forma de painéis, organizados por geografia, sexo e faixa etária, entre outros.

A CAPES também oferece uma ferramenta, denominada GEOCAPES [20], que é uma base de dados estatísticos que consiste em referenciar informações de acordo com sua localização geográfica. As informações são apresentadas na forma de mapas, gráficos e tabelas. Dentre as informações disponibilizadas estão: a concessão de bolsas de pós-graduação; distribuição de discentes, docentes e bolsistas de pós-graduação, entre outras.

Um problema comum nas duas ferramentas é o fato das informações apresentadas serem baseadas em dados coletados em uma determinada data e que podem ter sido modificados. Uma vantagem do SUCUPIRA é que os dados são obtidos diretamente da base de dados da PL durante o uso da ferramenta, permitindo que as análises sejam feitas com dados atualizados.

Existem diversas outras ferramentas que fornecem informações sobre pesquisadores em nível mundial. Algumas delas são: Google Scholar [21], Microsoft Academic Search [22], SciVerse Scopus [23] e Scirus [24]. O sistema SUCUPIRA é limitado no momento a pesquisadores cadastrados na PL e possibilita extrair informações específicas sobre grupos de pesquisadores que, de nosso conhecimento, seriam muito trabalhosas de serem obtidas com as ferramentas existentes.

IV. CONCLUSÕES E TRABALHOS FUTUROS

Este artigo apresentou o sistema SUCUPIRA que permite extrair informações de currículos Lattes para identificação de redes sociais acadêmicas e para análise simples dessas informações. O sistema está em desenvolvimento e o principal objetivo é identificar relacionamentos acadêmicos entre os pesquisadores. Atualmente, apenas um tipo de relacionamento é identificado. Entretanto, com as informações que já são extraídas do CV Lattes dos pesquisadores é possível identificar outros tipos de relacionamentos como, por exemplo, relações de orientado/orientador, participação em bancas etc. Além disso, pode-se adicionar informação à própria representação da rede social, por exemplo, incluindo peso nas arestas representando o número de vezes que um determinado relacionamento acontece. Com isso, é possível visualmente observar o quão intenso tais relacionamentos ocorrem.

AGRADECIMENTOS

Os autores agradecem o apoio financeiro da CAPES e CNPq.

REFERÊNCIAS

- [1] S. Vadrevu, F. Gelgi, and H. Davulcu, "Information Extraction from Web Pages using Presentation Regularities and Domain Knowledge", *World Wide Web*, vol. 10, n. 2, pp. 157-179, 2007.
- [2] Plataforma Lattes, disponível em: <http://lattes.cnpq.br/>, acessado em 08 de fevereiro de 2011.
- [3] J. Lane, "Let's Make Science Metrics more Scientific", *Nature*, vol. 464, n. 7288, pp. 488-489, 2010.
- [4] J. P. Mena-Chalco and R. M. Cesar-Jr., "scriptLattes: an Open-Source Knowledge Extraction System for the Lattes Platform", *Journal of the Brazilian Computer Society*, vol. 15, n. 4, pp. 31-39, 2009.
- [5] A. D. Alves, H. H. Yanasse and N. Y. Soma, "Extração de Informação na Plataforma Lattes para Identificação de Redes Sociais Acadêmicas", In IX Workshop dos Cursos de Computação Aplicada do INPE, São José dos Campos, 2009.
- [6] T. Nanno, S. Saito, and M. Okumura, "Structuring Web Pages on Repetition of Elements", In Second International Workshop on Web Document Analysis. Japão: 2003.
- [7] S. Flesca, G. Manco, E. Masciari, L. Pontieri, and A. Pugliese, "A Exploiting Structural Similarity for Effective Web Information Extraction", *Data & Knowledge Engineering*, Elsevier Science Publishers, vol. 60, n. 1, pp. 222-234, 2007.
- [8] R. Balanciere, A. B. Bovo, V. M. Kern, and R. C. dos S. Pacheco, "Análise de Redes de Colaboração Científica sob as novas Tecnologias de Informação e Comunicação: um Estudo na Plataforma Lattes", *Ciência da Informação*, vol. 34, n. 1, pp. 64-77, 2005.
- [9] A. B. O. Silva, R. F. Matheus, R. F. Parreiras, and T. A. S. Parreiras, "Esuto da Rede de Co-Autoria e da Interdisciplinaridade na Produção Científica através de Métodos de Análise de Redes Sociais: Avaliação do Caso do PPGCI/UFMG", In Encontros Bibli: Revista Eletrônica de Biliotecnomia e Ciência da Informação. Florianópolis, n. 1, pp. 179-194, 2006.
- [10] S. M. R. Vasconcelos, M. M. Sorenson, and J. Leta, "A new input indicator for the assessment of science & technology research?", *Scientometrics*, vol. 80, n. 1, pp. 217-230, 2009.
- [11] D. Arruda, F. Bezerra, V. A. Neris, P. R. de Toro, and J. Wainer, "Brazilian Computer Science Research: Gender and Regional Distributions", *Scientometrics*, vol. 79, n. 3, pp. 651-665, 2009.
- [12] R. B. Barata and M. Goldbaum, "Perfil dos Pesquisadores com Bolsa de Produtividade em Pesquisa do CNPq da Área de Saúde Coletiva", *Cadernos de Saúde Pública*, vol. 19, n. 6, pp. 1863-1976, 2003.
- [13] R. A. Calvacante, D. R. Barbosa, P. R. F. Bonan, M. B. O. Pires, and H. Martelli-Júnior "Perfil dos Pesquisadores da Área de Odontologia do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)", *Revista Brasileira de Epidemiologia*, vol. 11, n. 1, pp. 106-113, 2008.
- [14] A. C. Scarpelli, F. Sardenberg, D. Goursand, S. M. Paiva, and I. A. Pordeus, "Academic Trajectories of Dental Researchers Receiving CNPq's Productivity Grants", *Brazilian Dental Journal*, vol. 19, n. 3, pp. 252-256, 2008.
- [15] BlazeDS – BlazeDS – Adobe Open Source, disponível em: <http://opensource.adobe.com/wiki/display/blazeds/BlazeDS>, acessado em 16 de fevereiro de 2011.
- [16] Google Maps API for Flash – Google Code, disponível em: <http://code.google.com/intl-pt-BR/apis/maps/documentation/flash/>, acessado em 15 de fevereiro de 2011.
- [17] J. Bellone and D. Lang, "Visualizing Relational Data Using Graph Theory", *Flash & Flex Developer's Magazine*, vol. 1, n. 2, pp. 24-27, 2008.
- [18] BirdEye – Information Visualization and Visual Analytics Library, disponível em: <http://code.google.com/p/birdeye/>, acessado em 15 de fevereiro de 2011.
- [19] Painel Lattes (CNPq), disponível em: <http://lattes.cnpq.br/painelLattes/>, acessado em 07 de abril de 2011.
- [20] GEOCAPES Dados Estatísticos (CAPES), disponível em: <http://geocapes.capes.gov.br/geocapesds/#>, acessado em 07 de abril de 2011.
- [21] Google Scholar, disponível em: <http://scholar.google.com>, acessado em 08 de abril de 2011.
- [22] Microsoft Academic Search, disponível em: <http://academic.research.microsoft.com/>, acessado em 08 de abril de 2011.
- [23] SciVerse Scopus, disponível em: <http://www.info.sciverse.com/scopus/>, acessado em 08 de abril de 2011.
- [24] Scirus - for scientific information, disponível em: <http://www.scirus.com/>, acessado em 08 de abril de 2011.