

Pesquisa em Ontologias no Grupo de Engenharia do Conhecimento (GRECO)

Programa de Pós-graduação em Informática - UFRJ

Maria Luiza M. Campos¹, Maria Luiza A. Campos³, Maria Cláudia Cavalcanti⁴,
Linair Maria Campos², João Carlos Pereira da Silva¹
Vanessa Braganholo¹, Alberto Dávila⁵

¹Dep. Ciência da Computação/IM-UFRJ, ²Núcleo de Computação Eletrônica-UFRJ

³Dep. Ciência da Informação/UFF, ⁴Dep. de Engenharia de Sistemas/IME-RJ

⁵Laboratório de Biologia Computacional e Sistemas/Instituto Oswaldo Cruz-RJ

Resumo

O Grupo de Engenharia do Conhecimento (GRECO) surgiu da fusão de outros grupos de pesquisa: DatAware e CHORD, o primeiro com foco em integração de informações e o segundo com foco em suporte por computador ao trabalho cooperativo e gestão do conhecimento. Nos últimos anos, o grupo tem concentrado suas pesquisas nos diversos aspectos da Engenharia do Conhecimento, onde ontologias e outros mecanismos de organização e representação do conhecimento ocupam papel central. Nas investigações conduzidas nesta temática, o grupo tem mantido estreita colaboração com pesquisadores de outras instituições e áreas do conhecimento, em especial com pesquisadores da Ciência da Informação e Bioinformática, trabalhando diversos aspectos da construção e re-uso de ontologias na recuperação e integração de recursos em aplicações nas áreas de genômica comparativa, biologia de sistemas, prontuários médicos, sistemas de controle e apoio a emergências, web semântica, dentre outras.

Palavras-chave: construção de ontologias, re-uso de ontologias, integração de informações, bioinformática, sistemas de controle e apoio a emergências,

1. Introdução

Recentemente, ontologias passaram a receber atenção especial por parte da comunidade de computação, que vem aumentando sua utilização em mecanismos de classificação e recuperação mais sofisticados, e também como representações de domínios processáveis capazes de dar suporte a estratégias de integração e interoperabilidade, a exemplo das utilizadas em iniciativas da web semântica. No entanto, apesar da difusão de linguagens e ferramentas para a representação e construção de ontologias, as metodologias propostas resultam de pouca utilidade, pois em geral, ainda não contemplam diretrizes satisfatórias para identificação dos conceitos e relacionamentos entre estes, nem tampouco para a criação de definições sistemáticas associadas a esses conceitos. Em consequência, muitas das ontologias e esquemas terminológicos já em utilização por algumas comunidades, apresentam limitações quanto a sua expressividade e adequação às necessidades de seus usuários.

Domínios de aplicação como a Bioinformática e Sistemas de Controle e Apoio a Emergências apresentam diversos requisitos em termos de organização/representação/gestão do conhecimento e integração de recursos. Por isso mesmo, têm sido alvo de grandes investimentos e esforço por parte de pesquisadores da Ciência da Computação. No entanto, caracterizam-se também por serem aplicações de natureza tipicamente multidisciplinar, seja por requererem conhecimentos específicos das áreas temáticas em questão, seja pela diversidade de aspectos críticos a tratar, envolvendo as interações

humanas, a interpretação das informações trocadas, as demandas cognitivas por se lidar com grandes volumes de informações distribuídas e heterogêneas, apenas para citar algumas.

O grupo GRECO, desde suas origens, reconheceu a necessidade de parcerias para atuar nestas temáticas de maneira abrangente, capazes de trazer resultados reais e um retorno efetivo à sociedade no médio prazo. Neste sentido, vem trabalhando em estreita colaboração não só com outros colegas da Computação, mas também com pesquisadores da Ciência da Informação, utilizando bases teóricas e estratégias definidas nesta área e na Teoria da Terminologia, para a melhoria de abordagens e ferramentas de construção e re-uso de ontologias. Em Bioinformática, desde 2001, mantém parceria em diversos projetos de genômica comparativa em *Tripanosomatídeos* e, mais recentemente, em biologia de sistemas. Em Sistemas de Emergência, já há mais de 12 anos vem desenvolvendo trabalhos conjuntos com pesquisadores de Engenharia Cognitiva e de Resiliência, em projetos para o Corpo de Bombeiros e Defesa Civil do Rio de Janeiro, dentre outros, com parcerias com universidades chilenas e européias.

Neste artigo, apresentamos de forma resumida as principais pesquisas conduzidas pelo grupo envolvendo ontologias, um histórico destas parcerias e as perspectivas de trabalho no tema.

2. Trabalhos em Andamento, Perfil e Experiência do Grupo

Os fundamentos das técnicas de organização do conhecimento já vêm sendo experimentados há décadas na construção de tesauros e de terminologias no âmbito da Ciência da Informação, mas são ainda pouco conhecidos e utilizados na Ciência da Computação. Assim, nosso grupo vem ampliando os laços já estabelecidos com pesquisadores daquela área, especialmente com a Profa. Maria Luiza de Almeida Campos, da Ciência da Informação da UFF, buscando compreender e aplicar sua base teórica em representações de domínio necessárias no suporte aos mecanismos de integração e recuperação de recursos típicos, por exemplo, das aplicações distribuídas e da web semântica. Desta forma, no desenvolvimento de nossa pesquisa, a linha teórico/metodológica vem sendo baseada no aporte teórico da Ontologia Formal de Guarino (GUARINO, 1998), da Teoria do Conceito de Dahlberg (DAHLBERG, 1978), da Teoria Geral da Terminologia (WUESTER, 1981) e da Teoria da Classificação Facetada (RANGANATHAN, 1967; CAMPOS, 1995), em complemento às metodologias existentes para construção de Ontologias.

Nesta linha, diversos trabalhos foram desenvolvidos, sejam aqueles de cunho metodológico (MARINO, 2001; CAMPOS, 2004; CAMPOS, PIRES, CAMPOS, 2004; CAMPOS et al., 2009) sejam outros visando explorar o uso das ontologias como suporte ao acesso aos dados, ao processo de anotação, à gerência de dados de proveniência, dentre outros (MENDES, 2005; CAMPOS, CAMPOS, SOUZA, 2003; CAMPOS, CAMPOS, CAMPOS, 2005; CAMPOS, 2007^a; CAMPOS et. al. 2007^b; CRUZ, CAMPOS, MATTOSO, 2009).

A Profa. Vanessa Braganholo tem investigado o uso de padrões relacionados a ontologias e representação do conhecimento como RDF (BRAGANHOLLO; HEUSER, 2001) e XML (BRAGANHOLLO; DAVIDSON; HEUSER, 2006), e atuado nas pesquisas envolvendo o uso de ontologias na melhoria de técnicas de recuperação de informação (ELIAS, CAMPOS, BRAGANHOLLO, 2009).

Para que as melhorias descritas sejam viáveis do ponto de vista da processabilidade das ontologias resultantes, é fundamental que as extensões incorporadas sejam mapeadas e verificadas, (quanto à complexidade introduzida) para formalismos da Lógica Descritiva e Programação em Lógica - sob os quais se apóiam as linguagens de representação de ontologias como a OWL- já que estas constituem a base dos mecanismos de inferência que dão suporte à manipulação do conhecimento explicitado nas ontologias. Neste sentido, o grupo conta também em sua equipe com pesquisador especialista em Inteligência Artificial, o Prof. João Carlos Pereira da Silva, valendo-se de sua experiência nas diversas

técnicas e representações já utilizadas há muitas décadas, agora aplicadas a investigações na área de ontologias e bioinformática (LUPO, CERDERAL, 2006; NUNES, 2006; SILVA, 2008; MENEZES, 2009).

A colaboração da Profa. Maria Cláudia Cavalcanti do IME/RJ com o grupo se dá por sua atuação de muitos anos junto aos projetos de Bioinformática em parceria com o IOC, particularmente o laboratório coordenado pelo Prof. Alberto Dávila, desde o término de sua tese de Doutorado na UFRJ (CAVALCANTI et. al., 2002; CAVALCANTI et. al., 2003; CAVALCANTI, CAMPOS, MATTOSO, 2003). Teve continuidade com a dissertação de mestrado, sob sua orientação (BELLOZE, 2007), voltada para o mapeamento do processo de anotação baseado em ontologias, no sentido de apoiar a descoberta de problemas no uso de ontologias, reportá-las e facilitar com isso a evolução das mesmas. Além da dissertação, este trabalho gerou um protótipo, que foi desenvolvido para o sistema de anotação Stingray (BELLOZE et al., 2007a; BELLOZE et al., 2007b).

3. Equipe envolvida

A equipe envolvida com Ontologias junto ao Grupo GRECO encontra-se listada abaixo:

- PESQUISADORES: Maria Luiza Machado Campos (DCC/IM-UFRJ), Vanessa Braganholo (DCC/IM-UFRJ), João Carlos Pereira da Silva (DCC-IM/UFRJ), Maria Luiza Almeida Campos (Ciência da Informação – UFF), Maria Cláudia Cavalcanti (IME/UFRJ), Alberto Dávila (Biologia Computacional e de Sistemas/Instituto Oswaldo Cruz/RJ)
- ALUNOS (DOUTORADO): Kelle Belloze (doutorado BCS/IOC-RJ), Linair Maria Campos (doutorado PPCI-UFF), Sérgio Manuel Serra da Cruz (doutorado COPPE-UFRJ)
- ALUNOS (MESTRADO PPGI-UFRJ): Vivian Santos Silva, Débora Constantino, Renato Brum Alves, Eliêmia Araújo, André Bechara Elias, Miguel Gabriel Prazeres, João Victor Freitas, Maria Célia Tavares.
- ALUNOS (GRADUAÇÃO C.Computação UFRJ): Bianca de Oliveira Pereira, Renan de Oliveira Lobo, Ana Carolina Assaife, Fernanda Bruno, Ruben Albuquerque, Marcus Leopoldino.

O grupo teve diversos projetos aprovados com financiamento de órgãos de fomento (FINEP, FAPERJ, CNPq) onde o tema Ontologias é explorado seguindo as linhas indicadas anteriormente.

4. Direções de pesquisa

A equipe multidisciplinar que investiga o tema de ontologias aplicadas à Bioinformática tem trabalhado em colaboração durante os últimos 10 anos de maneira bastante produtiva, em diferentes aspectos do uso de ontologias na análise e anotação funcional de genomas de tripanosomatídeos e, mais recentemente, na construção de um sistema que usa a abordagem de “biologia de sistemas” para a identificação e análise de alvos para diagnósticos e fármacos nos genomas dos protozoários patogênicos (*T. cruzi*, *T. brucei*, *L. major*, *P. falciparum* e *E. histolytica*), com financiamentos CNPq e FAPERJ.

Para o futuro próximo, o grupo continuará sua pesquisa aplicada, investigando um recorte/extensão das ontologias da OBO para o ramo dos Tripanosomatídeos, visando atender às demandas dos projetos do grupo. No trabalho de BELLOZE (2004), verificou-se que na pesquisa genômica realizada com o *Trypanosoma rangeli*, das 375 seqüências não redundantes estudadas, apenas 17% do total das seqüências tiveram anotação utilizando termos da GO. Evidenciou-se também a possibilidade de uso de outras ontologias da OBO, que, pela dificuldade de manipulação e visualização de mais de uma ontologia no processo de anotação, não são ainda utilizadas. Assim, pesquisas visando dar suporte ao uso integrado de mais de uma ontologia, através de estratégias de alinhamento que possam tirar vantagem dos aportes teóricos da Ciência da Informação são prioridade e já se encontram em andamento.

De forma complementar, vem-se também investigando o povoamento e melhoria de ontologias a partir de acervos documentais e de definições associadas aos termos, utilizando-se de técnicas de recuperação de informação e mineração de texto.

No domínio de Emergências, a pesquisa do grupo em ontologias é trabalho recente, mas tem por meta o desenvolvimento de ontologias para suporte aos Sistemas de Controle e Apoio. Iniciativa neste sentido já se encontra em andamento, em parceria com a Universidade Politécnica de Valência, baseando-se em resultados anteriores de trabalhos do grupo sobre coleta de conhecimento de situações de emergência, tratamento de informações de contexto, modelagem de processo de apoio, dentre outros (OCHOA et. al., 2006; DINIZ et.al., 2005; SOUZA et.al. 2009).

Bibliografia

- BELLOZE, K. T. *Uma Extensão do Processo de Anotação Genômica para Ampliar o Uso e a Evolução Colaborativa de Ontologias no Domínio da Biologia Molecular. Dissertação (Mestrado em Sistemas e Computação) - Instituto Militar de Engenharia, 2007.*
- BELLOZE, K. T. ; DÁVILA, A. ; ARAÚJO, R. M. ; CAVALCANTI, M. C. . *Ampliando o Uso Colaborativo de Ontologias em Processos de Anotação Genômica. In: e-Science Workshops - SBBD/SBES. João Pessoa : v. 1. p. 71-80, 2007.*
- BELLOZE, K.T. ; ALBRECHT, F.; DÁVILA, A. ; ARAÚJO, R. M.; CAVALCANTI, M.C. *Capturing Ontology-based Annotation History. In: International Workshop on Genomic Databases (IWGD'07), Angra dos Reis - RJ., 2007.*
- BRAGANHOLO, V. ; DAVIDSON, Susan B. ; HEUSER, Carlos A. . *PATAXO: a framework to allow updates through XML views. ACM Transactions on Database Systems, v. 31, p. 839-886, 2006..*
- BRAGANHOLO, V.; HEUSER, C. *XML Schema, RDF(S) e UML: umacomparação. In: IDEAS 2001 - 4th Iberoamerican Workshop onRequirements Engineering and Software Environments, Santo Domingo,Heredia, Costa Rica, 2001. Proceedings... pages 78-90.*
- CAMPOS, L.M. *Uma Abordagem Sistemática de Descrição de Serviços Web para fins de Descoberta Dinâmica. Dissertação (Mestrado em Informática) – UFRJ, 2004.*
- CAMPOS, L.M.; PIRES, P.F.; CAMPOS, M.L.M. *OntoGuide-4WS: Uma Abordagem Sistemática para Descrição de Serviços Web, WebMedia/LA-Web, 2004, Ribeirão Preto, São Paulo, Brasil.*
- CAMPOS, M.L.A., CAMPOS, M.L.M., DÁVILA, A.M.R., GOMES, H.E., CAMPOS, L.M., LIRA, L. “Aspectos metodológicos no reuso de ontologias: um estudo a partir das anotações genômicas no domínio dos tripanosomatídeos”. *RECIIS – Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, v.3, n.1, p.64-75. 2009.
- CAMPOS, M.L.A. *A Organização de Unidades do Conhecimento em Hiperdocumentos: o modelo conceitual como um espaço comunicacional para a realização da autoria. 2001a. Tese de Doutorado – UFRJ, Escola de Comunicações, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro. Disponível em: <http://genesis.nce.ufrj.br/dataware>. Último acesso em 01 dez 2003.*
- CAMPOS, M.L.A.; CAMPOS, M.L.M. *Princípios para a Modelização de Domínios de Conhecimento: estudo comparativo entre abordagens da Ciência da Informação, Ciência da Computação e Teoria da Terminologia. In: IV Coloquio Internacional de Ciencias de la Documentación y VI Congreso del Capitulo Espanol de ISKO, 2003, Salamanca., 2003. p. 135-144.*
- CAMPOS, M. L. A.; CAMPOS, M. L. M.; GOMES, H. E.; CAMPOS, L. M.; MARTINS, A. E.; SALES, L. F.. *Estudo comparativo de Softwares de Construção de Tesauros. Perspectiva em Ciência da Informação*, v.11,n.1. Janeiro/Abril 2006.
- CAMPOS, M.L.M.; CAMPOS, M.L.A.; SOUZA, R.F. *Organização de unidades de conhecimento em hiperdocumentos: o modelo conceitual como espaço educacional para a realização da autoria. Ciência da Informação*, Rio de Janeiro, v. 32, n. 2, p. 7-16, 2003.
- CAMPOS, M. L. M. ; CAMPOS, M. L. A. ; CAMPOS, L. M. *Web semântica e a gestão de conteúdos informacionais. In: Carlos H. Marcondes; Hélio Kuramoto; Lídia Brandão Toutain. (Org.). Bibliotecas digitais: saberes e práticas. 1 ed. Brasília: , 2005, v. 1, p. 55-75.*
- CAVALCANTI, M. C., CAMPOS, M. L. M.; RÖSSLE, S.; BISCH, P. M; BAIÃO, F.; TARGINO, R.; PIRES, P. F.; *Structural Genomic Workflows Supported by Web Services. In: DEXA '03, WORKSHOP ON BIOLOGICAL DATA MANAGEMENT, 2003, Praga. Proceedings of the Workshop on Biological Data Management- DEXA'03. República Tcheca: IEEE CS Press, v. 1, pp.45-49, 2003.*
- CAVALCANTI, M. C.; CAMPOS, M. L. M.; MATTOSO, M.L.Q. *Managing Scientific Resources with Web Services. In: CAISE '03, WORKSHOP ON E-BUSINESS AND SEMANTIC WEB, Klagenfurt/Velden. Proceedings of the Workshop on*

- E-business and Semantic Web*. 2003.
- CAVALCANTI, M. C.; MATTOSO, M.L.Q.; CAMPOS, M. L. M.; LLIRBAT, F.; SIMON, E. An Architecture for Managing Distributed Scientific Resources. In: 14 TH INTERNATIONAL CONFERENCE IN SCIENTIFIC AND STATISTICAL DATABASE MANAGEMENT, Edimburgo. *Proceedings of the 14 th International Conference in Scientific and Statistical Database Management*. 2002.
- CRUZ, S. M. S., CAMPOS, M. L. M., MATTOSO, M. Towards a Taxonomy of Provenance in Scientific Workflow Management Systems In: 7th IEEE International Conference on Web Services (ICWS 2009) - Third International Workshop on Scientific Workflows (SWF 2009), 2009, Los Angeles.
- DAHLBERG, I. A referent-oriented analytical concept theory of interconcept. *International Classification*, v. 5, n. 3, pp.142-150, 1978.
- DÁVILA, A. M R ; CAVALCANTI, M. C. ; CAMPOS, M. L. M. ; MATTOSO, M. L. Q. ; MENDES, P. N. ProtozoaDB: dynamic visualization and exploration of protozoan genomes (the 2008 Annual Database Issue). *Nucleic Acids Research*, v. 36, p. 1, 2007.
- DINIZ,V.B., BORGES, M.R.S., GOMES, J.O., CANÓS, J.H. "Knowledge Management Support for Collaborative Emergency Response," *Proc. of the 9th International Conference on Computer Supported Cooperative Work in Design*, Coventry, UK, May 2005, pp. 1188-1193.
- ELIAS, A. B., CAMPOS, M. L. M., BRAGANHOLO, V. Applying Biomedical Ontologies on Semantic Query Expansion In: *International Conference on Biomedical Ontology (ICBO 2009)*, 2009, Buffalo EUA.
- GO, The Gene Ontology Consortium, 2006. Disponível em: <http://www.geneontology.org/>. Último acesso em 05 mai 2006.
- GUARINO, N. Formal ontology and information systems. In: FOIS '98, 1, 1998a, Trento, Italy. *Proceedings*. Amsterdam: IOS Press; Tokyo: Omsa, pp. 3-15, 1998.
- GUERREIRO, L. ; SOUZA, S.S. ; WAGNER, G. ; SOUZA, E. A.; MENDES, P. N. ; CAMPOS, L. M.; BARROS, L. A.; PIRES, P. F. ; CAMPOS, M. L. M. ; GRISARD, E. C.; DÁVILA, A.M. R . Exploring the Genome of Trypanosoma Vivax through GSS and in Silico Comparative Analysis. *Omics A Journal Of Integrative Biology*, New York, v. 9, n. 1, pp. 116-128, 2005.
- LUPO, P. P. ; CERDEIRAL, C. T. . JPE: Um Ambiente Integrado de Desenvolvimento de Sistemas de Representação de Conhecimento e Raciocínio" Trabalho de Conclusão de Curso.(Bacharelado em Ciência da Computação) – UFRJ, 2006. Prof. João Carlos Pereira da Silva.
- MARINO, M. T. Integração de Informações em Ambientes Científicos na Web: uma Abordagem baseada na Arquitetura RDF. Dissertação (Mestrado em Informática) – UFRJ, 2001. Orientador: Maria Luíza M. Campos.
- MENDES, P.N. Uma Abordagem para Construção e Uso de Ontologias no Suporte a Integração e Análise de Dados Biológicos. Dissertação (Mestrado em Informática) – UFRJ, 2005. Orientador: Maria Luíza M. Campos.
- MENEZES, G.M.C. Técnicas de inteligência artificial aplicadas ao problema das redes de regulação biológicas. Trabalho de Conclusão de Curso.(Bacharelado em Ciência da Computação) – UFRJ, 2009. Orientador: Joao Carlos P. da Silva.
- NUNES, B. N. . Implementação de um Interpretador Prolog, Trabalho de Conclusão de Curso. (Bacharelado em Ciência da Computação), UFRJ, 2006. Orientador: Prof. João Carlos P. da Silva.
- OCHOA, S.F., PINO, J.A., NEYEM, A., BORGES, M.R.S. Using Context to Support Group Decision Making in Disasters Affecting Urban Areas. In: IFIP WG8.3 International Conference on Creativity and Innovation in Decision Making and Decision Support. Londres: v. 2, p. 546-561, 2006.
- RANGANATHAN, S. R. Prolegomena to library classification. Bombay: Asia Publ. House, 1967.
- SILVA, V.S. Alinhamento de ontologias através do algoritmo de alinhamento local de caminhos. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – UFRJ, 2008. Orientador: Joao Carlos P. da Silva.
- SILVA, Alexandre ; RUBERG, Gabriela ; BAIÃO, Fernanda ; BRAGANHOLO, V. ; MATTOSO, Marta Lima de Queirós . Efficiently processing XML queries over fragmented repositories with PartiX. In: *Database Technologies for Handling XML Information on the Web (DATAx)*, 2006, Munique, Alemanha. *Current Trends in Database Technology EDBT 2006*. Berlin / Heidelberg : Springer, 2006. v. 4254. p. 150-163.
- SMITH, B.; ASHBURBER M.; ROSSE, C.; BARD, J.; BUG, W.; CEUSTERS, W.; GOLDBERGG, L. J.; EILBECK, K.; IRELAND, A.; MUNGALL, C. J. Et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, v. 25, n. 11, p. 1251-5, nov. 2007.
- SOUZA, A.P ; FREITAS, J.V.V.B. ; CUNHA, A. M. ; CAMPOS, M. L. M. . Um Método de Modelagem de Processos para a Investigação de Problemas em Sistemas de Informação: Aplicação em uma Central de Regulação Médica. In: V Simpósio Brasileiro de Sistemas de Informação, Brasília, 2009.
- WAGNER, G.; SORIANO, K. ; JUCÁ, H. ; BELLOZE, K. T. ; TSCHOEKE, D. A. ; GERONIMO, G. A. ; ALBRECHT, F. ; MATTOSO, M.L.Q.; CAVALCANTI, M. C. ; CAMPOS, M. L.M.; GRISARD, E.; DÁVILA, A. M. R. STINGRAY: System for Integrated Genomic Resources and Analysis. In: *International Workshop on Genomic Databases (IWGD'07)*, Angra dos Reis – RJ. 2007.