

Gestão do Conhecimento Taxonômico Aplicado na Conservação da Flora Brasileira

Luís Alexandre. E. Silva^{1,2}, Jonice de Oliveira Sampaio¹, Eduardo Dalcin²,
Geraldo Zimbrão¹, Jano M. de Souza¹

¹COPPE/UFRJ, Universidade Federal do Rio de Janeiro, Brasil
Caixa Postal 68.511 – CEP 21945-970 Rio de Janeiro/RJ, Brasil

²Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Diretoria de Pesquisa Científica, Rua Pacheco Leão 915, cep. 22460-030, Rio de Janeiro, RJ, Brasil.
{estevao, jonice, zimbrao, jano}@cos.ufrj.br, {edalcin@jbrj.gov.br}

Abstract. *Information systems for biodiversity management and monitoring are effective tools for nature conservation. However, the taxonomic data are difficult to handle, considering the rate of new discoveries and constant updates of the names by the specialists. Thus, the definition of a methodology to manage and share knowledge of these researchers is needed. In this study, we present an approach to the use of collaboration technologies, ontologies and data mining, enabling the management of scientific knowledge in plant conservation.*

Resumo. *Os sistemas de informação para gestão e monitoramento da biodiversidade são ferramentas efetivas para a conservação da natureza. Porém, os dados taxonômicos são de difícil manipulação, tendo em vista que novas espécies de plantas são descobertas, necessitando de identificação e as constantes atualizações dos nomes por especialistas. Assim, a definição de uma metodologia para gerenciar e compartilhar o conhecimento destes pesquisadores se faz necessária. Neste trabalho, é realizado um estudo do uso das tecnologias de colaboração, ontologias e mineração de dados possibilitando a gestão do conhecimento científico na conservação da flora.*

1. Introdução

A Botânica é um importante domínio científico sendo seu conteúdo disponibilizado em muitos tipos de aplicações. Nesses sistemas, parte dos dados utilizados são obtidos por meio de taxonomistas. Esses profissionais realizam o estudo de famílias de plantas, identificado-as; naturalmente acumulam, com o tempo muito conhecimento, porém encontram dificuldades para o compartilhamento deste conhecimento. Assim, a motivação para este trabalho e a de promover a divulgação do conhecimento taxonômico para ações de conservação da flora.

2. Objetivo

Este trabalho tem por objetivo analisar o uso da gestão do conhecimento científico no auxílio as atividades relacionadas à conservação da flora. Apresentando a como mais um campo de aplicação da gestão de conhecimento. Assim, a proposta apresentada vem

a ser mais uma opção de aplicação aos taxonomistas participantes do projeto do Centro Nacional de Conservação da Flora no Instituto de Pesquisas Jardim Botânico do Rio de Janeiro¹. Para alcançar este objetivo será proposta uma arquitetura para o desenvolvimento de um sistema colaborativo para apoiar às tomadas de decisões.

4. A Gestão do conhecimento

Após a análise de alguns sistemas de informação para a biodiversidade, foi considerado o uso da gestão do conhecimento [Comas e Poch 2008] para esta aplicação. Especificamente, foram estudados os Sistemas de Gestão do Conhecimento Científicos. O uso da tecnologia como suporte à gestão do conhecimento facilita no registro, organização, transformação e busca do conhecimento tácito. Tecnologias utilizadas:

4.1. Mineração de dados

A mineração de dados objetiva a extração ou mineração de conhecimento a partir de grande quantidade de dados. Sua aplicação em bases de dados botânicas vem sendo cada vez mais usada [Hochachka et al 2007], pode auxiliar na descoberta de padrões e predições de espécies vulneráveis e descoberta de relacionamentos entre dados auxiliando na descoberta de conhecimento de forma integrada [Kuo et al 2007] [Zaková et al 2008] [Bernstein et al 2005].

4.2. Ontologias

Ontologias são utilizadas para criar terminologias únicas, possibilitando o compartilhamento e a reutilização do conhecimento ou ainda na definição dos conceitos e relacionamentos em um domínio particular [Guarino 1998]. Iniciativas tem sido desenvolvidas para a integração de sistemas e fontes de informações por meio de ontologias [Pohle 2002] [Brisson 2005]. Uma abordagem é o uso de um ambiente de criação e desenvolvimento de ontologias de domínio de modo colaborativo [Sampaio 2003]. Neste sentido podemos observar iniciativas na Botânica [Gu et al 2004] [Tochtermann e Maurer 2000]. Na prática a integração de ontologias com modelos conceituais pode ser feita por meio de linguagens de representação de conhecimento, como a UML. Como permitem exportar o modelo em XML, este pode ser traduzido em linguagem ontológica e, posteriormente importadas para ferramentas como o Protegé² segundo a OWL³ [Gasevic 2004].

5. Arquitetura proposta

Para a implementação do sistema é apresentada uma proposta de componentes para sua arquitetura (Figura 1). O fluxo de informações da arquitetura é feito da seguinte forma:

- A ontologia é extraída no banco de dados e pela iteração com taxonomistas;
- A base de conhecimento, é obtida em um passo inicial por meio de entrevistas com os usuários ou pesquisas em fontes de dados internas ou externas e na literatura;

1 <http://www.jbrj.gov.br>

2 <http://protege.stanford.edu/>

3 <http://www.w3.org/TR/owl-features/>

- Um sistema colaborativo permite que os especialistas possam interagir e trocar conhecimento, agilizando o processo;
- Com a definição de uma ontologia e a base de conhecimento armazenada, conjuntos de dados podem ser definidos como um pré-processamento da mineração de dados [Pohle 2003] [Pan, Jun-Yi e Mu-Xin 2006] [Brisson, Collard e Pasquier 2005];
- Aplicação da mineração para a descoberta de padrões. Na procura pelo aumento do grau de confiança nos resultados, um módulo de avaliação da qualidade dos dados foi adicionado a arquitetura. Pode-se ainda usar algoritmos de limpeza de dados;
- Um pós-processamento no qual os padrões descobertos podem ser interpretados ou filtrados de acordo com o conhecimento prévio armazenado no banco de conhecimento ou pela consulta aos taxonomistas. Assim, pode-se filtrar os padrões obtidos como relevantes ou não para as pesquisas relativas com a conservação;

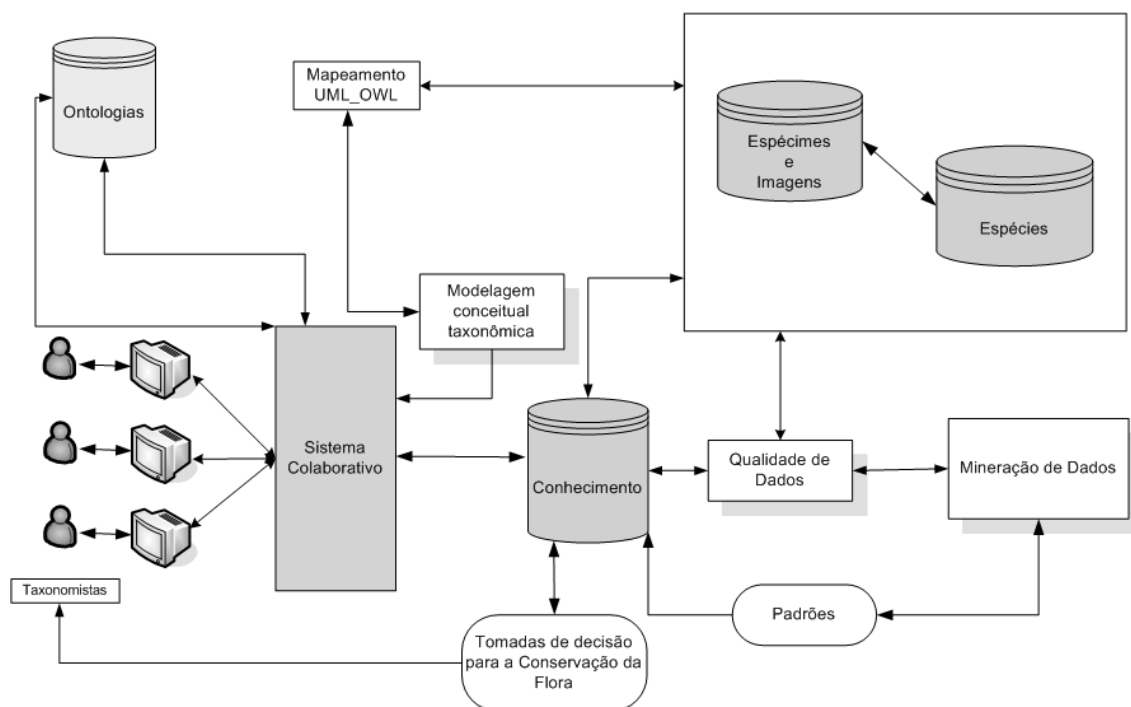


Figura 1 – Solução proposta para o CNCFlora

- Para que isto possa ser realizado, estruturas hierárquicas e generalizações/especializações podem ser definidas para os conceitos ontológicos na construção de regras gerais. Essas regras podem reduzir o volume de padrões extraídos e permitir a construção de um módulo para interpretar os resultados obtidos pelos.

5. Conclusão

Neste trabalho, foram estudadas diversas técnicas de tratamento de dados que podem ser consideradas componentes para um sistema de gestão do conhecimento no contexto do processo de análise de dados, visando incentivar ações de conservação da flora. Além de apresentar a Botânica com mais um campo de aplicação para as tecnologias de mineração de dados, qualidade de dados, trabalho colaborativo e ontologias. De modo a permitir o uso das tecnologias, foi apresentada uma proposta de uma arquitetura para o

controle da gestão do conhecimento. Concluímos que um melhor aproveitamento do conhecimento dos taxonomistas pode ser alcançado tanto pelo compartilhamento quanto pela motivação promovida com o uso de um ambiente colaborativo. Desta forma os órgãos ambientais e instituições de pesquisas poderão se beneficiar do enorme potencial da gestão do conhecimento no uso conjunto destas diversas ferramentas quando aplicadas na gestão de seus sistemas de informações ambientais.

Referências

- Bernstein, A., Provost, F., Hill, S. (2005) "Toward Intelligent Assistance for a Data Mining Process: An Ontology-Based Approach for Cost-Sensitive Classification" <http://cat.inist.fr/?aModele=afficheN&cpsidt=16598278>, março;
- Brisson, L., Collard, M., Pasquier, N. (2005) "Improving Knowledge Discovery Process Using Ontologies", https://portail.telecom-bretagne.eu/publi/public/download.jsp?id_publication=4263, setembro;
- Comas, J., Poch, M., (2008) "A Framework for the Development of Intelligent Environmental Decision Support Systems", http://www.iemss.org/iemss2008/uploads/Main/S04-04-Comas_et_al-IEMSS2008.pdf, março;
- Gasevic, D. (2004) "UMLtoOWL: Converter from UML to OWL", <http://www.sfu.ca/~dgasevic/projects/UMLtoOWL/>, setembro;
- Gu, F., Cao, C., Sui, Y. e Tian, W. (2004) "Domain-Specific Ontology of Botany Other attributes", <http://www.springerlink.com/index/P61N85NXX368J7R1.pdf>, março;
- Guarino, N. (1998) "Formal Ontology in Information Systems", Proceedings of the 1st International Conference, IOS Press Trento, Italy;
- Hochachka, W. M., Caruana, R., Fink, D., et al. (2007) "Data-Mining Discovery of Pattern and Process in Ecological Systems", Journal of Wildlife Management. <http://www.bioone.org/perlserv/?request=get-bstract&doi=10.2193/2006-503>, março;
- Kuo, Y., Lonie, A., Sonenberg, L. e Paizis, K. (2007) "Domain ontology driven data mining," *Proceedings of the 2007 international workshop on Domain driven data mining*, <http://portal.acm.org/citation.cfm?doid=1288552.1288554>, março;
- Pan, D., Jun-Yi, S., Mu-Xin, Z. (2006) "Incorporating domain knowledge into data mining process: An ontology based framework", <http://www.springerlink.com/content/b257566w24111875/>, setembro;
- Pohle, D.C. (2002) "Integrating and Updating Domain Knowledge with Knowledge Discovery," Knowledge Creation Diffusion Utilization, http://www.hhl.de/fileadmin/LS/micro/Download/Pohle_2003_IntegratingAndUpdatingA.pdf, março;
- Pohle, C. (2003) "Integrating and Updating Domain Knowledge with Data Mining", <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-76/pohle.pdf>;
- Sampaio, J. O. (2003) "Epistheme: Um Ambiente de Gestão de Conhecimento Científico", www.cos.ufrj.br/~jonice/tese_msc/TeseMScJoniceOliveira.zip, setembro;
- Zaková, M., Kremen, P., Zelezný, F. e Lavrac, N. (2008) "Using Ontological Reasoning and Planning for Data Mining Workflow Composition", <http://http://ida.felk.cvut.cz/cgi-bin/docarc/public.pl/document/129/4.pdf>, março.