

O Uso de Agentes Ontológicos para Identificação de Plantas com o OntoPlant

Marco Antonio S. Silva¹, Mayara Alexandre², Pablo F. Matos³, Avaldo O. S. Filho⁴

¹Discente Superior em Sistemas de Informação, ²Discente Superior em Engenharia Elétrica,
³Docente de Informática, ⁴Docente de Biologia

¹²³Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA)
Av. Amazonas, 3150, Zabelê – 45.075-265 – Vitória da Conquista – BA – Brasil

⁴Universidade Estadual do Sudoeste da Bahia (UESB)
Estrada do Bem Querer, km 4, Candeias – 45.083-900 – Vitória da Conquista – BA – Brasil

marcoantoniosilva.tec@gmail.com, mayara_25@hotmail.com,
pablofmatos@ifba.edu.br, avaldosf@gmail.com

Abstract. *This paper describes the development of the OntoPlant software that allows the identification of forest species with ontological agents. The computational program uses taxonomic keys to classify these species, analyzing the main characteristics provided by the user. The software is a means that facilitates and streamlines the process of identification of the species collected in the Vitória da Conquista city and region. 47 species in 15 families and 88 characteristics were registered in the whole. The software offers a database of images, a glossary and a dialog box that allows the user, when selecting the feature, to understand in a simple and accessible way what each classification represents.*

Resumo: *Este artigo descreve o desenvolvimento do software OntoPlant que permite identificar as espécies de plantas florestais com agentes ontológicos. O programa computacional utiliza de chaves taxonômicas para fazer a classificação dessas espécies, analisando as principais características fornecidas pelo usuário. O software é um meio que facilita e agiliza o processo de identificação das espécies coletadas no município de Vitória da Conquista e região. Ao todo foram cadastradas 47 espécies em 15 famílias e 88 características. O software oferece um banco de imagens, um glossário e uma caixa de diálogo que permite o usuário, no momento da seleção da característica, compreender de forma simples e acessível o que representa cada classificação.*

1. Introdução

O processo de evolução das máquinas ocorre desde a Revolução Industrial até nos dias de hoje. Depois de participar da criação do World Wide Web (WWW), Tim Bernes-Lee avança para algo mais sofisticado e atual, nomeando como Web Semântica. Segundo Berners-Lee, Hendler e Lassila (2001), a Web Semântica representa para a humanidade algo como um “conjunto revolucionário de tecnologias” que foi desenvolvido com o objetivo de dar a máquina a capacidade de refinar buscas, fazer compras verificando ao mesmo tempo se aquele site é seguro para tal finalidade, além de tornar a máquina capaz de “compreender o sentido” das informações e dados a partir da descrição dos domínios

onde estão inseridos no grande banco dados interligados que a Web Semântica se tornou após o advento dessa tecnologia.

A Web Semântica é formada por agentes inteligentes conhecidos como agentes ontológicos, ou Ontologia, que fazem o processo de “organização e interligação” das informações contidas no sistema, fazendo com que todos os domínios possam se relacionar. Ao observar essa aplicação, a área da Inteligência Artificial buscou aplicar aos robôs, com o objetivo de criar uma área conhecida como “*Machine Learning*” que possibilita dar a máquina uma capacidade inerente ao homem, que é identificar as coisas através da semântica das mesmas.

Após a criação da Web Semântica, sua aplicação conseguiu atingir outros recursos, tecnologias e funcionalidades, como é o caso das linguagens: *Resource Description Framework* (RDF), *eXtensible Markup Language* (XML), *SPARQL Protocol and RDF Query Language* (SPARQL) e *Ontology Web Language* (OWL), expandindo ainda mais suas possíveis aplicações [Segundo, Coneglian e Oliveira Lucas 2017].

Ao utilizar os recursos oferecidos pela Web Semântica é possível formar um sistema mais intuitivo, além de auxiliar o homem nas tarefas que o sistema se prepõe a solucionar [World Wide Web Consortium 2011]. Desde então, observou-se essas aplicações o que possibilitou a verificação de uma relação entre a justificativa inicial do projeto e as soluções que a Web Semântica pode oferecer. Por conseguinte, utilizou-se agentes ontológicos, uma das extensões da Web Semântica, para desenvolver um sistema capaz de facilitar a atividade de identificação de espécies florestais. Com esses recursos foi possível construir um sistema que possibilitasse o usuário fazer triagem/identificação de plantas por espécies sem perder a hierarquia das chaves taxonômicas que é base fundamental para obter um sistema preciso e assertivo na sua resposta.

O restante deste artigo está estruturado da seguinte forma: Na Seção 2, são discutidos os trabalhos correlatos, na Seção 3, são apresentados os temas relacionados a este trabalho, na Seção 4, são mostrados os recursos do *software* OntoPlant, e na Seção 5, são apresentadas as Considerações Finais.

2. Trabalhos Correlatos

Os trabalhos encontrados na literatura têm como objetivo a identificação de plantas, seja por meio de fotos como é o caso do PlantSnap (2010), do PlantNet (2010), do LeafSnap (2010) e do Garden Answers (2013); seja por meio de chaves taxonômicas criadas pelo próprio usuário como é o caso do Project Xper Botanica (2010); ou ainda através do modo de selecionar características fornecidas pelo usuário como faz o Lucid (2010) ou finalmente codificando descrições taxonômicas que é o caso do Free Delta (2010).

O *software* OntoPlant utiliza de agentes ontológicos para a identificação das espécies florestais. Através do uso da ontologia no sistema, foi possível inserir imagens, características, descrição das características e um glossário para descrição das palavras acadêmicas. Este sistema de identificação permite que as informações sejam colocadas em formato de árvore não alterando a hierarquia das informações.

3. Fundamentação Teórica

O termo taxonomia está relacionado ao desenvolvimento de uma estrutura constituída por nomes ou informações que auxiliam a descrever um domínio [Vital e Café 2011]. A organização em forma taxonômica apresenta uma ordem de informações hierárquica, ou

também pode ser uma forma de organizar através da classificação das informações da mais genérica as mais específicas.

A taxonomia é o estudo das relações filogenéticas entre as plantas. Taxa é o termo usado para qualquer nível hierárquico de grupos de plantas relacionadas, como espécies, gêneros, família e ordem. Cada táxon recebe um nome segundo as regras de nomenclatura. Espécie é definida como um grupo de indivíduos que pode cruzar e produzir descendentes férteis, ou ainda um grupo de indivíduos morfologicamente semelhantes e que provavelmente podem cruzar entre si. Gênero são espécies próximas entre si. O nome completo de cada espécie inclui o nome do gênero e da espécie e os dois são escritos em itálico. Família são gêneros próximos. A família é o nível de classificação mais usado pelos botânicos no campo. Um botânico quando está tentando identificar as espécies ele começa pela família através de uma série de características.

Para fazer a triagem das espécies observando suas características, o profissional da área de botânica faz o uso de chaves taxonômicas, ou chaves dicotômicas que é um sistema que utiliza as características morfológicas das famílias para diferir uma espécie da outra. Na Figura 1 é mostrado a forma de classificação por meio do uso das chaves taxonômicas.

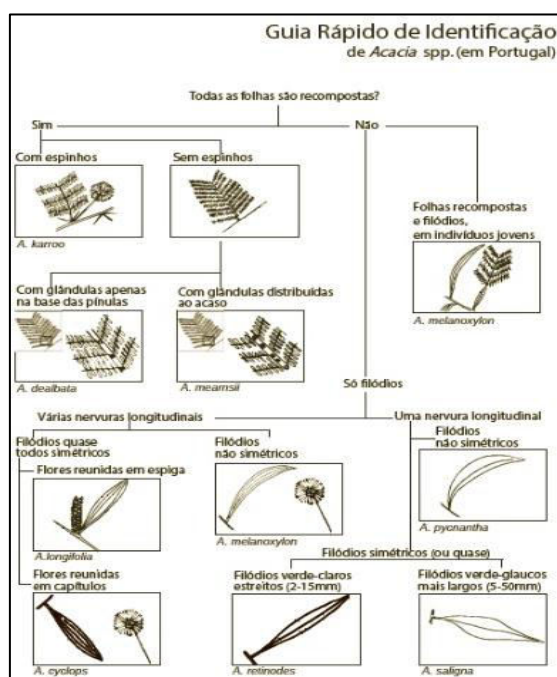


Figura 1. Chave taxonômica para identificação de espécies florestais.

Na área da Web Semântica, o estudo da taxonomia possui um conceito semelhante, que é a organização das informações em hierarquias. Ainda dentro da Web Semântica, existe uma aplicação conhecida como Ontologia, que apresenta estruturas de classes, organizadas de forma taxonômica, apresenta também informação das propriedades dessas classes, a possibilidade de relacionar essas propriedades, e instâncias ou características que estão relacionadas com essas classes. Sendo assim, o conhecimento da Ontologia é organizado em formato de teia. Dessa forma, entende-se Ontologia como uma extensão da taxonomia e apesar da relação entre os conceitos, para a finalidade de classificação de espécies florestais, o uso do mecanismo taxonômico é o mais conhecido [Vital e Café 2011].

Dessa forma, a partir da chave taxonômica é possível fazer a associação do mesmo com a Ontologia. Após analisar a relação entre essas organizações é possível verificar que a Taxonomia se organiza de maneira hierárquica, de forma que as classes e subclasses dos domínios não podem se relacionar entre si. Na Ontologia, as classes e subclasses podem ter alguma informação em comum formando uma espécie de teia de informações. Entretanto, nada impede que o sistema continue obedecendo uma hierarquia predefinida, característica essa que possibilita a modelagem das informações por meio da Ontologia. Na Figura 2 é possível visualizar a diferença entre Taxonomia e Ontologia.

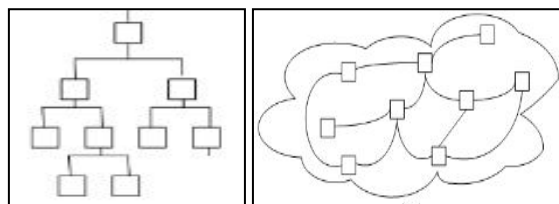


Figura 2. Organização da Taxonomia e da Ontologia.

Esse trabalho de organização de informações em forma hierárquica também é aplicado a área da botânica, com o intuito de organizar informações classificatórias de forma que o usuário dessas chaves possa, ao seguir a sequência, chegar a instância final. Assim, se o usuário observar uma característica presente na planta, ele pode optar pela característica apresentada na chave, e descartar a outra. Ao fazer esse trabalho de observação da característica e escolha da mesma na chave, o usuário tem uma boa chance de acertar a espécie estudada.

Dessa forma, através da identificação morfológica, a chave proporciona duas ou mais alternativas nas etapas e a cada escolha de uma dessas características geram-se características na etapa seguinte. Entretanto, fazer uma relação das características de uma espécie com espécies que possuem características correlacionadas é um papel da Ontologia, uma vez que esta área utiliza a organização de uma informação em forma de teia.

Esse processo de identificação, normalmente é realizado em herbários, onde são armazenadas as espécies de diversas localidades. De acordo com os dados do *Index Herbariorum* [Thiers s.d], estão catalogados cerca de 3.000 herbários no mundo todo, dos quais, 150 estão localizados no Brasil onde 125 desses estão em intercâmbio para troca de dados científicos. A Universidade Estadual do Sudoeste da Bahia (UESB), *campus* Vitória da Conquista, possui o herbário chamado HUESBVC [Filho 2017], cuja finalidade básica está na manutenção da coleção das amostras do material botânico (amostras de plantas) coletadas em todo o país, mas com um foco na região do Sudoeste e Oeste da Bahia. Todo esse material contabiliza um acervo com cerca de 8.500 espécimes (exsicatas).

4. OntoPlant

O OntoPlant é um *software* que utiliza agentes ontológicos, no contexto da Web Semântica, para facilitar a identificação de espécies florestais. As principais características do *software* são: (i) fornece detalhes e fotos de uma determinada planta, família ou característica presentes em uma Ontologia; (ii) permite que o usuário associe imagens a elementos da Ontologia, além de possibilitar que o usuário faça consultas de

termos técnicos em um glossário; (iii) e permite que o usuário amplie a base de dados da Ontologia.

Na Figura 3 é mostrado a interface inicial do OntoPlant. Na ontologia foi possível modelar os domínios ou famílias, com suas respectivas subclasses que foram utilizadas como opção das características das famílias até alcançar a espécie final que são as instâncias (Figura 3 - Área 1). O *software* também tem a opção de visualização de imagens relacionadas as características selecionadas (Figura 3 - Área 2), além de uma área do programa destinada a descrição das características selecionadas pelo usuário à medida que ele avança nas chaves (Figura 3 - Área 3).

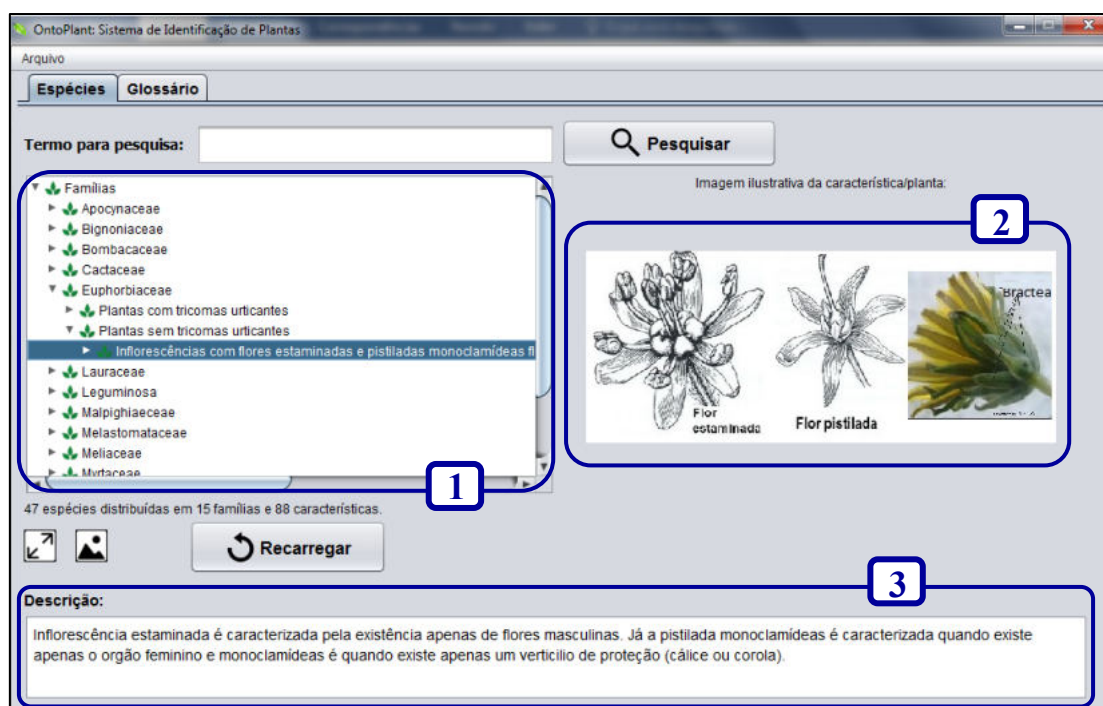


Figura 3. Interface inicial do OntoPlant.

Outra opção do sistema é a ferramenta de busca por termos conforme destacado na Figura 4. O usuário pode digitar a palavra referente a característica ou as espécies desejadas para o sistema realizar a busca na Ontologia.



Figura 4. Ferramenta de busca na ontologia.

O sistema ainda apresenta um mecanismo de expandir e retrair a Ontologia, inserir imagens através do banco de imagens do computador e a opção de recarregar a pesquisa e a ontologia, além de apresentar um sistema de contagem das espécies, famílias e características cadastradas no sistema (Figura 5).

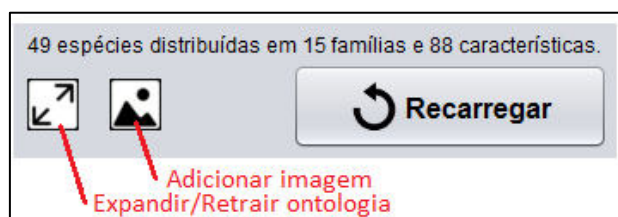


Figura 5. Opções oferecidas pelo programa.

Ademais, é possível realizar uma pesquisa direta de um termo no glossário apenas clicando sobre um item e escolhendo a opção de pesquisar um determinado termo, como mostrado na Figura 6.

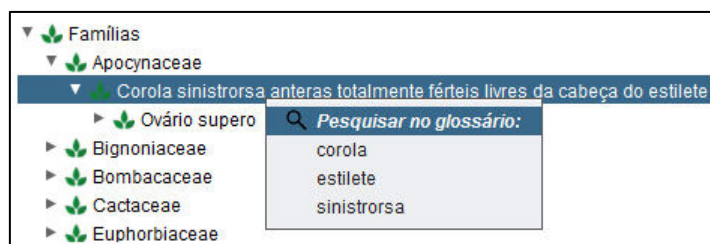


Figura 6. Pesquisa direta no glossário.

Na aba Glossário é mostrado os termos e seus respectivos significados (Figura 7).

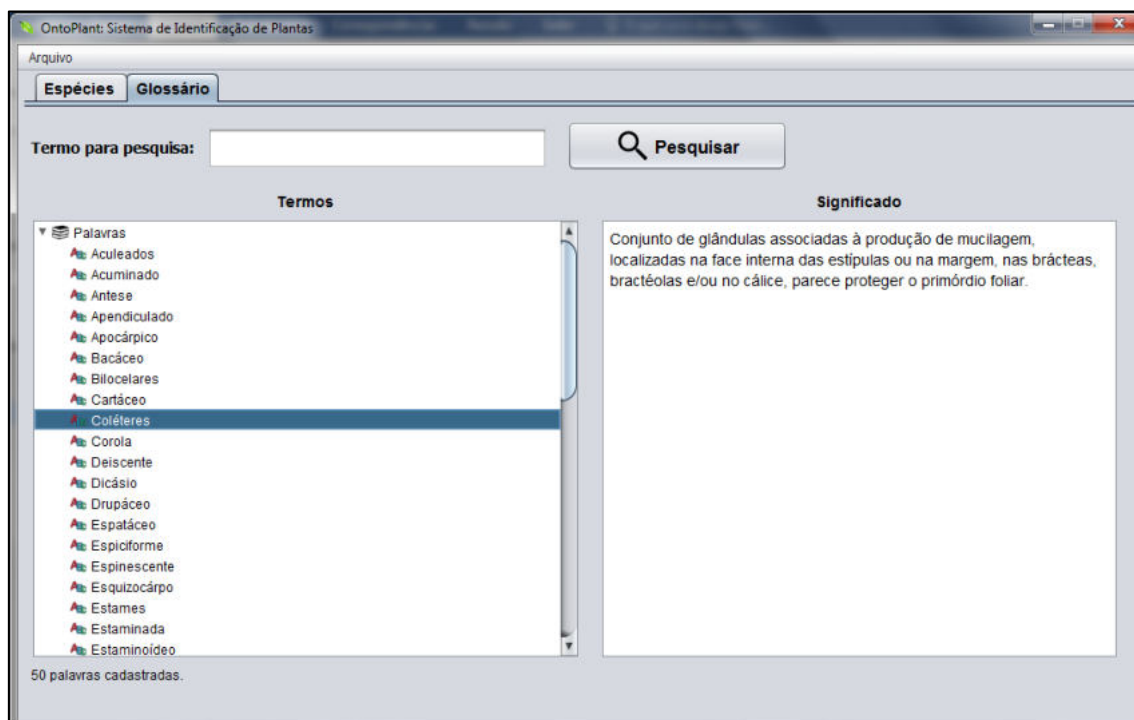


Figura 7. Glossário do OntoPlant.

A ontologia do OntoPlant foi modelada a partir do editor de ontologias denominado Protégé (2018), o qual modela ontologias nas linguagens RDF, RDFS, OWL e OWL2. Na Figura 8 é mostrado um exemplo de chave taxonômica de uma família modelada no Protégé.

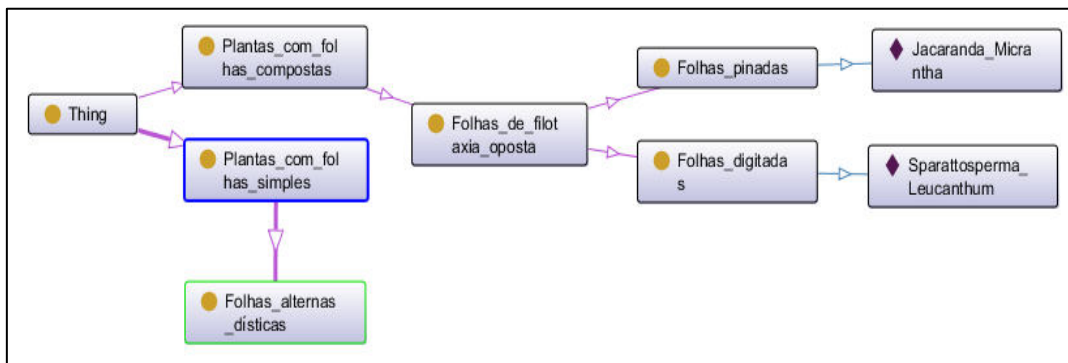


Figura 8. Exemplo de chave taxonômica de uma família modelada no Protégé.

O *framework* Jena (2018) foi escolhido para construir o *software* OntoPlant na linguagem de programação Java a partir da ontologia modelada no Protégé. Dessa forma, o Jena garante que a hierarquia das chaves taxonômicas seja mantida por meio da ordem das classes e subclasses existentes na ontologia.

Ao todo foram cadastradas 47 espécies em 15 famílias e 88 características. Na Figura 9 podem ser visualizadas as 15 famílias no OntoPlant, modeladas no editor de ontologias Protégé. Para o usuário inserir mais famílias ou mais palavras no glossário nesta versão do OntoPlant, é necessário ter o Protégé (2018) instalado e realizar as inserções ou exclusões por dele.



Figura 9. Famílias modeladas no editor de ontologias Protégé no programa OntoPlant.

O software, assim como seu código fonte e o manual do usuário, encontra-se disponível para download por meio do seguinte link: <http://www.pablofmatos.pro.br/tools/17-OntoPlant>.

5. Considerações finais

Com a presente proposta, foi possível verificar uma aplicação para o uso da ontologia, além dos que já são comumente visto em publicações como é o caso da Web Semântica, do mecanismo de busca por preço mais barato ou ainda sugestão de combinação de acessórios de uma loja de roupas para os usuários, por exemplo. A partir do OntoPlant também foi possível validar que o uso da ontologia para chaves taxonômicas é um método

eficiente para armazenar e recuperar as informações, uma vez que a estrutura hierárquica é mantida de forma integral tanto nas classes/subclasses quanto nas espécies.

O uso das ontologias no projeto também tornou o programa mais leve, o que poderia ser uma dificuldade no momento da opção de efetuar o download do programa. Verificou-se que não foi necessário excluir nenhuma das opções para que o programa ficasse leve. Além de ser um mecanismo rápido de busca. Por fim, para trabalhos futuros o programa pode ser expandido para uma aplicação mobile e web além de aumentar o acervo já existente atualmente.

Agradecimentos

Agradecemos o apoio financeiro concedido pelo IFBA por meio do Programa Institucional de Iniciação em Desenvolvimento Tecnológico e Inovação (PIBITI) – Edital Nº 04/2017/PRPGI.

Referências

- Berners-Lee, T., Hendler, J. e Lassila, O. (2001) **The Semantic Web**. Scientific American.
- Filho, A. S. (2017) **Entrevista Concedida a Mayara Alexandre**. Vitória da Conquista. 12 jun. 2017.
- Free Delta (2010), **An Open Source System for Processing Taxonomic Descriptions**, <http://freedelta.sourceforge.net/>.
- Gardem Answers (2013) **Plant Identification**, <http://www.gardenanswers.com/>.
- Jena (2018) **Apache Jena: A free and open source Java framework for building Semantic Web and Linked Data applications**, <https://jena.apache.org/>.
- LeafSnap (2010) **An Electronic Field Guide**, <http://leafsnap.com/>.
- Lucid (2010) **Identification Keys**, <http://www.lucidcentral.com/>.
- PantSnap (2010) **A Plant Expert at your Fingertips**, <https://www.plantsnap.com/>.
- PlantNet (2010) **Identify, Explore and Share your Observations of Wild Plant**. <https://identify.plantnet-project.org/>.
- Project Xper Botanica (2010) **Identification de la Flore de France**, https://www.keytonature.eu/wiki/Projet_XPer_Botanica/.
- Protégé (2018) **Protégé: A free, open-source ontology editor and framework for building intelligent systems**, <https://protege.stanford.edu/>.
- Thiers, B. (s.d.) **Index Herbariorum: A Global Directory of Public Herbaria and Associated Staff**. New York Botanical Garden's Virtual Herbarium, <http://sweetgum.nybg.org/science/ih/>.
- Segundo, J. E. S., Coneglian, C. S., e Oliveira Lucas, E. R. (2017) Conceitos e Tecnologias da Web Semântica no Contexto da Colaboração Acadêmico-Científica: um Estudo da Plataforma Vivo. **TransInformação**, 29(3), 297-309.
- Vital, L. P., e Café, L. M. A. (2011) Ontologia e Taxonomia: Diferenças, **Perspectivas em Ciência da Informação**, v.16, n.2, 115-130.
- Word Wide Web Consortium (2011) **Web Semântica**, <http://www.w3c.br/Padroes/WebSemantica>.