

Asistencia a la identificación de especies botánicas del NE del Chubut a través de aplicaciones basadas en ontologías

Renato Mazzanti^{1,2}, Gustavo Daniel Samec^{1,2}, Alejandro Bisigato³, Alejandro Sanchez^{1,4}

¹LINVI, Dpto. de Informática, Fac. de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco
+54-0280-4472885-int. 117 - Puerto Madryn, Argentina

²Unidad de Gestión de la Información – UGI CENPAT-CONICET
+54-0280-4450401-int. 1260 - Puerto Madryn, Argentina

³Instituto Patagónico para el Estudio de los Ecosistemas Continentales – IPEEC-CONICET
+54-0280-4450401-int. 1321 - Puerto Madryn, Argentina

⁴Departamento de Informática, Universidad Nacional de San Luis
+54 266 4520300 – San Luis, Argentina

e-mail: {renato, gsamec, bisigato}@cenpat-conicet.gob.ar, aljsanchez@gmail.com

Resumen

La determinación de especies botánicas dista de ser trivial. El enfoque mayormente aplicado consiste en utilizar claves dicotómicas que a través de una serie preestablecida de observaciones, ayudan al experto a determinar la identidad de una planta. Estas claves se materializan en voluminosas publicaciones que incluyen un extenso conjunto de especies, las que raramente coexisten en un sitio, y requieren reconocer caracteres sutiles, o el uso de instrumentos de laboratorio. El proceso es complejo, y resulta inviable cuando se debe completar en campo.

Esta línea de investigación, desarrollo e innovación apunta a definir y desarrollar aplicaciones basadas en ontologías para asistir en la determinación de especies a investigadores botánicos del NE de la provincia del Chubut. Estas aplicaciones deberán funcionar, tanto en laboratorio como en campo, sin seguir una secuencia preestablecida, ajustándose a las características reconocidas por el usuario.

Palabras clave: Bioinformática, Ontología, Morfología de Plantas, Botánica

Contexto

Esta línea de investigación se encuentra en su etapa inicial y será llevada adelante en el contexto del Laboratorio en Investigación en Informática (LINVI) del Departamento de Informática de la Universidad Patagónica San Juan Bosco (UNPSJB), en su sede de Puerto Madryn. Se relaciona con las líneas de I-D-I del LINVI, BigData y Ontologías, e involucra al Profesor de Ecología de Comunidades de la UNPSJB el Dr. Alejandro Bisigato.

Introducción

La identificación de especies de plantas suele ser un desafío incluso para los expertos. La situación es aún más desafiante cuando el proceso debe completarse en campo. Aunque existen

numerosas claves dicotómicas, éstas suelen abarcar grandes regiones e incluir cientos o miles de especies. Ello hace que las claves hagan referencia a sutiles caracteres para diferenciar especies emparentadas, las que pueden estar especificadas utilizando conceptos (o lenguaje) que resulte ajeno a un botánico menos experto por su especificidad o taxonomía subyacente, o que difícilmente puedan observarse en campo sin la ayuda de material óptico. Un ejemplo de estas claves es la Flora Patagónica (Correa 1969-1999), la que incluye varios tomos y nunca terminó de publicarse de modo impreso. Ello ha originado la publicación de numerosas guías locales, las que frecuentemente incluyen sólo algunos grupos de plantas y muchas veces, incluso carecen de claves para la determinación (Bertiller y Beeskow S/F, Green y Ferreyra 2011, Kröpfl y Vilasuso 2012, Quintana 2014).

En los últimos años se han desarrollado varias iniciativas en el dominio de la biología donde se utilizan distintos enfoques para acercar comunidades que realizan investigaciones en diferentes disciplinas tales como la genética, biología molecular, biología celular, taxonomía, botánica y genómica. Estos enfoques incluyen, por ejemplo, la utilización de vocabulario controlado, como en Plant Ontology Database¹, bases de datos online, como el caso de Flora Argentina², generada y mantenida por la autoridad taxonómica de la botánica para Argentina, o lenguajes y herramientas específicos, como es el caso de DELTA³ (DEscription Language for TAXonomy)

¹ <http://www.plantontology.org/>

² <http://www.floraargentina.edu.ar/>

³ <http://delta-intkey.com/>

(Dallwitz 1980), cuyas herramientas interactivas permiten seguir, a diferencia de las claves dicotómicas, la trayectoria que mejor se adapte a la información que dispone el usuario.

El uso de ontologías para la representación formal y compartida del conocimiento (Borst 1997, Gruber 1993) emerge también como una opción para atender a problemas como el enunciado (Fang y colegas 2004). Los lenguajes y tecnologías que soportan este enfoque poseen fundamentos sólidos en las lógicas descriptivas (Baader y colegas 2003). Por ejemplo, un razonador responderá la especie/subespecie de una instancia una vez se describan todas las características necesarias para tal clasificación.

En general, la conceptualización se formaliza en una red de ontologías relacionadas de distintos tipos. Una posible clasificación de las ontologías distingue entre ontologías de alto nivel, de tarea, de dominio, y de aplicación (Guarino 1998). Las ontologías de un dominio son esenciales en estos enfoques. Las entidades en estas se relacionarán con conceptos definidos en ontologías de alto nivel, como ser, por ejemplo, la ontología de unidades estándar de medida.

Por otro lado, las ontologías de aplicación refinan las de dominio para responder a preguntas/problemas específicos dentro del dominio, dando soporte al funcionamiento de una aplicación. Las relaciones en esta red de ontologías se establecen con tecnologías de correspondencia entre ontologías (Euzenat 2013).

Las ontologías están lejos de ser desarrolladas de manera anárquica. Existe un conjunto de metodologías para esto, incluyendo (Grüniger y Fox 1995, Uschold y King 1996, Bernaras y colegas 1996, Gómez-Pérez y colegas 1996, Staab y colegas 2001, Corcho y colegas 2003, Tautz y Wangenheim 1998). Muchas de

estas metodologías consideran la participación de expertos de dominio (no familiarizados con lenguajes de especificación de ontologías) en el proceso de desarrollo.

Líneas de Investigación, Desarrollo e Innovación

Esta línea de I-D-I procura definir y desarrollar prototipos de aplicaciones basadas en ontologías para asistir en la determinación de especie en el NE de la provincia de Chubut. El enfoque se inspira en la necesidad de brindar funcionalidad complementando la provista por DELTA, pero explotando la expresividad y herramientas disponibles para ontologías especificadas en lógicas descriptivas (o también llamadas de la web semántica). Los principales ejes de la línea son:

- Participación de expertos del dominio de la región como usuarios y fuente de conocimiento.
- Selección y ajuste de metodología de desarrollo de ontologías, y software asociado.
- Utilización herramientas para el desarrollo de ontologías especificadas en lenguajes basados en lógicas descriptivas.
-

Objetivos y resultados esperados

Con el objetivo general de asistir en la determinación de especie en el NE de la provincia de Chubut, a través de aplicaciones basadas en ontologías, se desarrollarán una serie de actividades que se esperan que permitan:

- Obtener una aproximación explícita e informal a la conceptualización para

responder a las preguntas del experto, a partir de las fuentes de conocimiento disponibles que describen la flora de la provincia, y descripciones de los expertos de la región.

- Desarrollar una red de ontologías a partir de la (re)utilización de ontologías existentes.
- Desarrollar aplicaciones basadas en ontologías que permitan al investigador identificar especies botánicas de la provincia en diversos escenarios de trabajo.

Formación de Recursos Humanos

Está previsto en esta línea de I-D-I continuar con la formación de integrantes del LINVI, incluyendo el desarrollo de dos tesis de Maestría en Ciencias de la Computación de la Universidad Nacional del Sur, y dos tesinas de grado de la Licenciatura en Informática de la UNPSJB.

Referencias

Atkinson B., Cooper L, Moore L, Preece J, Lingutla NTV, Todorovic S, Walls RL, Stockey R, Rothwell G, Smith B, Gandolfo MA, Stevenson DW, Jaiswal P. (2013) Plant Ontology, a controlled and structured plant vocabulary for all botanical disciplines. Botany Conference. New Orleans, LA

BERTILLER, MÓNICA y BEESKOW, ANA MARÍA. S/F. Las flores de la Patagonia. Alrededores de Puerto Madryn y Península Valdés. I. Arbustos y

Subarbustos. Área Científica de Zonas Áridas. Centro Nacional

Patagónico. CONICET. Puerto Madryn. Chubut. Argentina

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point?. *Data & knowledge engineering*, 46(1), 41-64.

Correa, M. N., 1969. Flora Patagónica. Parte II. Typhaceae a Orchidaceae. Colección Científica del INTA. Buenos Aires. 219 pp.

Correa. M. N., 1971. Flora Patagónica. Parte VII. Compositae. Colección Científica del INTA. Buenos Aires. 451 pp.

Correa, M. N., 1984. Flora Patagónica. Parte IVa. Dicotiledóneas Dialipétaleas (Salicaceae a Cruciferae). Colección Científica del INTA. Buenos Aires. 559 pp.

Correa. M. N., 1988. Flora Patagónica. Parte IVb. Dicotiledóneas Dialipétaleas (Droseraceae a Leguminosae). Colección Científica del INTA. Buenos Aires. 309 pp

Correa. M. N., 1988. Flora Patagónica. Parte V. Dicotiledóneas Dialipétaleas (Oxalidaceae a Cornaceae). Colección Científica del INTA. Buenos Aires. 381 pp.

Correa. M. N., 1999. Flora Patagónica. Parte VI. Dicotyledones. Gamopétaleas (Ericaceae a Calyceraceae). Colección Científica del INTA. Buenos Aires. 536 pp.

Dallwitz, M.J.. A general system for coding taxonomic descriptions. 1980.

Fang Gu, Cun-Gen Cao, Yue-Fei Sui, and Wen Tian. 2004. Domain-specific ontology of botany. *J. Comput. Sci. Technol.* 19, 2 (March 2004), 238-248.

DOI=<http://dx.doi.org/10.1007/BF02944802>

Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider (Eds.). 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, New York, NY, USA. Borst, Willem Nico (1997). "Construction of Engineering Ontologies for Knowledge Sharing and Reuse". Tesis doct. Enschede.

Gómez-Pérez, A., Fernández, M., & Vicente, A. D. (1996). Towards a method to conceptualize domain ontologies.

Green L y Ferreyra M. 2011. Flores de la estepa patagónica. Vázquez Mazzini Editores, Buenos Aires.

Gruber, Thomas R. (1993). "A Translation Approach to Portable Ontology Specifications". En: *Knowl. Acquis.* 5.2, págs. 199-220. ISSN : 1042-8143.

Grüniger, M., & Fox, M. S. (1995). *Methodology for the Design and Evaluation of Ontologies*.

Guarino, N. (1998). *Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy*. 1st. Amsterdam, The Netherlands, The Netherlands: IOS Press. ISBN : 9051993994.

Jérôme Euzenat and Pavel Shvaiko. 2013. *Ontology Matching (2nd ed.)*. Springer Publishing Company, Incorporated

Kröpfl Ai & Villasuso Nm. 2012 *Guía para el Reconocimiento de Especies de los Pastizales del Monte Oriental de Patagonia*. San Carlos de Bariloche, INTA.

L. Watson and M. J. Dallwitz (1998 onwards). *DELTA Sample Data: Descriptions, Illustrations, Identification*,

and Information Retrieval. Version: 21st September 2000.

Quintana, DR 2014 Plantas de la Patagonia Árida. Nativas y exóticas del noroeste del Chubut. Fondo editorial de la Secretaría de Cultura de la Prov. del Chubut

Schreiber, A. T., Wielinga, B. J., Akkermans, J. M., & Van de Velde, W. (1994). CommonKADS: A comprehensive methodology for KBS development. Deliverable DM1 (Vol. 1). 2a KADSII.

Staab, S., Studer, R., Schnurr, H. P., & Sure, Y. (2001). Knowledge processes and ontologies. *IEEE Intelligent systems*, (1), 26-34.

Šimún, M., Andrejko, A., & Bieliková, M. (2007, September). Ontology-based models for personalized e-learning environment. In *Proc. of 5th Int. Conf. on Emerging e-learning Technologies and Applications (ICETA'07)*.

Tautz, C., Wangenheim, G., & Refseno, C. (1998). A representation formalism for software engineering ontologies. *Fraunhofer IESE-Report No, 15*, 1-151.

Uschold, M., & King, M. (1995). Towards a methodology for building ontologies (pp. 15-30). Edinburgh: *Artificial Intelligence Applications Institute, University of Edinburgh*.