

Aplicaciones Informáticas para el Estudio de Biodiversidad de Poliquetos Espiönidos en los Golfos Nordpatagónicos

Gustavo Samec^{1,2,3}, María Emilia Diez^{1,2,5}, Marcos Zárate^{1,2,4}, Carlos Buckle^{1,2}, Joaquín Lima^{1,2}, Rodrigo Jaramillo^{1,2,3}, Alejandro Sánchez^{1,2,6}, Renato Mazzanti^{1,2,3}

¹ Departamento de Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB), Puerto Madryn.

² LINVI, Laboratorio de Investigación en Informática, Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB)

³ Unidad de Gestión de la Información, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (CCT CONICET-CENPAT)

⁴ Centro para el Estudio de Sistemas Marinos, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (CESIMAR) (CCT CONICET-CENPAT)

⁵ Laboratorio de Parasitología (LAPA), Instituto de Biología de Organismos Marinos, Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (IBIOMAR) (CCT CONICET-CENPAT)

⁶ Departamento de Informática, Universidad Nacional de San Luis (UNSL), San Luis.
{gsamec, emiliadiez, zarate, renato}@cenpat-conicet.gob.ar, cbuckle@unpata.edu.ar

RESUMEN

En la actualidad, la información biológica es de tal magnitud que se hace necesario el uso de herramientas informáticas para optimizar la labor del científico. En este sentido, una base de datos regional constituye una buena herramienta para la validación y la precisión de datos con estándares de control. En las costas patagónicas uno de los principales problemas para la marisqueería es el causado por los poliquetos espiönidos polidorinos. Una correcta identificación de estas especies es clave para atender los problemas de infestación. El objetivo general es estudiar teorías y técnicas que permitan el desarrollo de herramientas que asistan a biólogos en la validación de las especies involucradas en sus estudios. Se propone realizar un proceso de ingeniería de software para atender el caso particular de espiönidos de los golfos nordpatagónicos mediante la incorporación de información de distribución, herramientas avanzadas de consulta y semántica para la interoperabilidad con repositorios de especies de referencia mundial que garanticen la calidad de los datos. El proyecto está integrado por especialistas de

Ciencias de la Computación y Biológicas, conformando un grupo de investigación interdisciplinario en informática para la biodiversidad dentro del Laboratorio de Investigación en Informática (LINVI) de la Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB).

Palabras clave: Base de Datos, Taxonomía, Poliquetos, Spionidae, Web Semántica.

CONTEXTO

Este proyecto fue realizado por especialistas de Ciencias de la Computación y de las Ciencias Biológicas, que conforman un grupo de investigación interdisciplinario en informática para la biodiversidad dentro del LINVI de la UNPSJB. En los últimos años Orensanz J.M., Mazzanti R., Samec G. y Lima J. han desarrollado la base de datos Southwest Atlantic Benthic Invertebrates¹ (SWATL) con la finalidad de centralizar la información sobre invertebrados bentónicos de la región con sus datos asociados para

¹ <http://sistema.cenpat-conicet.gob.ar:8081/sgrb/sgrb/system/Menu.jsf>

hacer interpretaciones y realizar aplicaciones prácticas en el campo de la conservación. Sobre esta base se pretende el desarrollo de herramientas informáticas que faciliten el estudio de la taxonomía y ecología de estas especies. Para poder llevar a cabo el proyecto será necesario especificar, diseñar y desarrollar un sistema de bases de datos que considere la semántica y la interoperabilidad con repositorios de referencia mundial de especies, como por ejemplo Global Biodiversity Information Facility² (GBIF) y World Register of Marine Species³ (WoRMS). En este caso en particular se trabajará con uno de los grupos más abundantes y diversos de la comunidad bentónica como es la clase Polychaeta. Este proyecto apunta a consolidar un sistema de información de registros biológicos como un repositorio de registros sobre especies de poliquetos espionidos para la región patagónica, con características de base de datos abierta e interoperable, que por su implementación genérica pueda incorporar a futuro otras especies de otros grupos.

1. INTRODUCCIÓN

Existen iniciativas a nivel global como GBIF, Especies2000⁴ o WoRMS que gestionan listas taxonómicas de especies cuyo objetivo es la gestión de índices uniformes y la validación por especialistas taxónomos. Como antecedente directamente relacionado con el proyecto se cuenta con la base de datos NONATObase⁵ que trabaja con registros de especies de clase Polychaeta del Atlántico Sudoccidental (y áreas cercanas al antártico), desde 5°N a 80°S y cuenta con un comité

científico externo de asesoramiento. La fuente de datos primarios que utilizan se basa en la exploración y recopilación de información desde distintos sitios Web como Web of Knowledge⁶, Scopus⁷, Scielo⁸ y desde librerías y/o repositorios de distintas instituciones de investigación. Dentro del contexto nacional el Ministerio de Ciencia, Tecnología e Innovación Productiva (MinCyT) creó el Sistema Nacional de Datos Biológicos⁹ (SNDB) sobre una plataforma desarrollada por GBIF basada en el estándar *Darwin Core* [1]. Además creó el Sistema Nacional de Datos del Mar (SNDM) con el fin de visualizar la información de los centros nacionales productores de datos marinos de Argentina de tipo satelitales, geológicos, batimétricos, físico-químicos, geofísicos y biológicos. Sumado a esto en mayo de 2014 se lanza oficialmente la iniciativa Pampa Azul¹⁰, y se propone utilizar la plataforma del SNDM para el manejo del conocimiento generado en los puntos focales de la iniciativa (Golfo San Jorge, Islas Sándwich del Sur, Agujero Azul y Talud, Banco Burwood/Namuncurá). La gestión de la biodiversidad requiere el uso de información biológica heterogénea de múltiples fuentes. La indexación, la agregación y la búsqueda de dicha información se basan en los nombres y el conocimiento taxonómico de los organismos. Sin embargo, las taxonomías cambian con el tiempo debido a nuevos hallazgos científicos. Además, los nombres de las especies y su significado cambian en el tiempo, diferentes autoridades usan diferentes nombres científicos para un mismo taxón en diferentes momentos, y varios nombres vernáculos están en uso en diferentes idiomas.

² <https://www.gbif.org/>

³ <http://www.marinespecies.org/>

⁴ <https://sp2000.org/>

⁵ <http://nonatobase.ufsc.br/>

⁶ <https://www.webofknowledge.com/>

⁷ <https://www.scopus.com>

⁸ <http://www.scielo.org>

⁹ <http://datos.sndb.mincyt.gob.ar/>

¹⁰ <http://www.pampazul.gob.ar/>

Esto dificulta la integración de datos y la recuperación de información si no se dispone de información biológica detallada. Entre los trabajos más importantes podemos mencionar [2,3,4]. De acuerdo con [5] el surgimiento de la Web Semántica [6] ha permitido obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información. En lo que refiere a los datos científicos, los datos abiertos enlazados [7] se presentan como el camino estándar hacia la integración de datos heterogéneos de diferentes dominios. Este proyecto apunta a desarrollar un sistema de información que contenga datos de diversidad de poliquetos (Annelida: Polychaeta) de la familia Spionidae sobre la base de datos SWATL. La finalidad de este sistema es aportar herramientas para la gestión taxonómica y de la información asociada a cada registro.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El presente proyecto plantea el desarrollo de herramientas que asistan a biólogos y otros actores interesados a la validación de las especies involucradas en sus estudios de poliquetos, mediante la taxonomía y propone especificar y diseñar un sistema de bases de datos, con el agregado de tecnologías semánticas y estándares de *World Wide Web Consortium* (W3C) como RDF¹¹, SPARQL¹², etc. y el estándar GeoSPARQL¹³ del *Open Geospatial Consortium* (OGC) para la

interoperabilidad con registros de referencia mundial de especies.

Por otro lado, el proyecto de investigación *Infraestructura de Acceso a Datos Primarios con aporte de semántica en Repositorios Digitales* [8] tiene varios objetivos generales. Uno de ellos es el de desarrollar conocimiento especializado en el área de Interoperabilidad Semántica de la Información. En este sentido, se estudian, entre otras, el desarrollo de agentes de información cuyo ambiente de trabajo es la Web, además de técnicas de representación de conocimiento, razonamiento automático y modelado ontológico.

Estas líneas de investigación confluyen en el estudio de formalismos y tecnologías para cubrir las necesidades emergentes de compartir, actualizar e integrar el conocimiento de sistemas computacionales pre-existentes como lo es SWATL.

3. RESULTADOS ESPERADOS

En este proyecto se estudiarán teorías y se desarrollarán modelos y aplicaciones de software para asistir a biólogos que trabajan en biodiversidad, contribuyendo a la validación de las especies involucradas y aportando calidad a sus registros. En este caso particular se trabajará con un grupo de organismos muy abundante y diverso de la comunidad bentónica como son los poliquetos espiónidos de los golfos nordpatagónicos.

Objetivos particulares:

1. Analizar el estado del arte respecto de la interoperabilidad con repositorios o bases de datos de especies de referencia mundial.
2. Analizar y rediseñar el modelo de datos de la base de datos SWATL para la implementación de un sistema de información.

¹¹ <https://www.w3.org/RDF/>

¹² <https://www.w3.org/TR/sparql11-query/>

¹³ <https://www.opengeospatial.org/standards>

3. Facilitar mediante herramientas informáticas la obtención de sinonimias, las fuentes y referencias para un taxón, como así también la distribución geográfica y sus atributos.
4. Estudiar y aplicar tecnologías de la Web Semántica para la publicación de datos abiertos enlazados, interoperabilidad semántica y explotación de información.
5. Consolidar un grupo de investigación interdisciplinario en informática para la biodiversidad dentro del grupo LINVI de la UNPSJB.

En este sentido, el proyecto permitirá el desarrollo de un sistema de información taxonómica/biológica, que constituirá una herramienta para facilitar y mejorar la calidad de los estudios taxonómicos y ecológicos, en un continuo flujo de retroalimentación de la información [9].

La relevancia de la propuesta se identifica a diferentes niveles:

1. Tecnológico: La herramienta a desarrollar utiliza tecnologías semánticas para la integración con diferentes repositorios internacionales. Esto será de gran relevancia no sólo para investigadores taxónomos o ecólogos, sino también para aquellos organismos estatales o instituciones que trabajan en temáticas de conservación de la biodiversidad, de manejo de recursos pesqueros entre los principales.
2. Socioeconómico: La infestación de polidorinos afecta a moluscos de interés comercial en los golfos nordpatagónicos. Esto es relevante y pertinente en tanto es necesario contar con información sobre el estado de salud de los moluscos como satisfacer estándares de calidad del producto.
3. Política científica: Contar con un sistema que gestione este tipo de información facilitará la investigación de métodos combinados de control biológico y prácticas de manejo para una explotación más eficiente.

4. FORMACIÓN DE RECURSOS HUMANOS

En este proyecto participan integrantes de formación docente y académica en las áreas de Ingeniería de Software, Bases de Datos, Inteligencia Artificial y Biología. Cinco de los docentes son del Departamento de Informática y uno del Departamento de Matemáticas de la Facultad de Ingeniería de la UNPSJB Sede Puerto Madryn. La Doctora en biología M. Emilia Diez es especialista en taxonomía y ecología de poliquetos, dos de los autores están realizando carreras de doctorado y otros dos se encuentran desarrollando carreras de especialización y maestrías. Uno de los autores de este proyecto está inscripto en el Doctorado en Ciencias de la Computación en la Universidad Nacional del Sur y cuenta con beca interna doctoral del CONICET hasta el 2020 y otro está finalizando su doctorado en Ciencias de la Computación en la Universidad Nacional de San Luis. También forman parte del proyecto un graduado de la carrera de Licenciatura en Informática y tres alumnos del ciclo superior. Su participación tiene el objeto de introducirlos a la tarea científica y permitirles incorporar conocimientos sobre temas no desarrollados en la currícula de la carrera. En el caso de los alumnos que están próximos a graduarse, este proyecto tiene la intención de guiarlos en el desarrollo de sus tesis en esta rama de la disciplina. Otro aporte para la formación académica radica en la posibilidad de que los alumnos puedan realizar Instancias Supervisadas de Formación en la Práctica Profesional en el marco de este proyecto y en pasantías y/o terrenos de capacitación en el área de Biología. Si bien el proyecto de investigación es nuevo, continúa con las líneas de investigación iniciadas en otros proyectos anteriores. El objetivo es combinar los esfuerzos realizados en forma individual en un proyecto común.

5. BIBLIOGRAFÍA

1. Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, et al. Darwin Core: an evolving community-developed biodiversity data standard. *PLoS One*. 2012;7: e29715.
2. Tuominen J, Laurenne N, Hyvönen E. Biological Names and Taxonomies on the Semantic Web – Managing the Change in Scientific Conception [Internet]. *The Semantic Web: Research and Applications*. 2011. pp. 255–269. doi:10.1007/978-3-642-21064-8_18
3. Michel F, Faron-Zucker C, Terceire S, Olivier G. TAXREF-LD: A Reference Thesaurus for Biodiversity on the Web of Linked Data [Internet]. *Proceedings of TDWG*. 2017. p. e20232. doi:10.3897/tdwgproceedings.1.20232
4. Plitzner P, Henning T, Müller A, Güntsch A, Karam N, Kilian N. Bottom-up taxon characterisations with shared knowledge: describing specimens in a semantic context. *S4BioDiv@ ISWC*. 2017. Available: <https://pdfs.semanticscholar.org/7f02/4d5917825e07370ac3f19df4ae4429fdf98a.pdf>
5. Koper R. Use of the Semantic Web to Solve Some Basic Problems in Education: Increase flexible, distributed lifelong learning; decrease teacher's workload [Internet]. *Journal of Interactive Media in Education*. 2004. p. 5. doi:10.5334/2004-6-koper
6. Berners-Lee T, Hendler J, Lassila O, Others. *The semantic web*. Sci Am. New York, NY, USA; 2001;284: 28–37.
7. Heath T, Bizer C. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers; 2011;1: 1–136.
8. Mazzanti R, Zàrate M, Samec G, Buckle C. Infraestructura de acceso a datos primarios con aporte de semántica en repositorios digitales. *XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste)*. 2018. Available: <http://sedici.unlp.edu.ar/handle/10915/67259?show=full>
9. Escalante T, Llorente J, Espinosa D, Soberón J. Bases de datos y sistemas de información: aplicaciones en biogeografía. *Rev Acad Colombiana Cienc Exact Fis Natur*. 2000;24: 325–341.