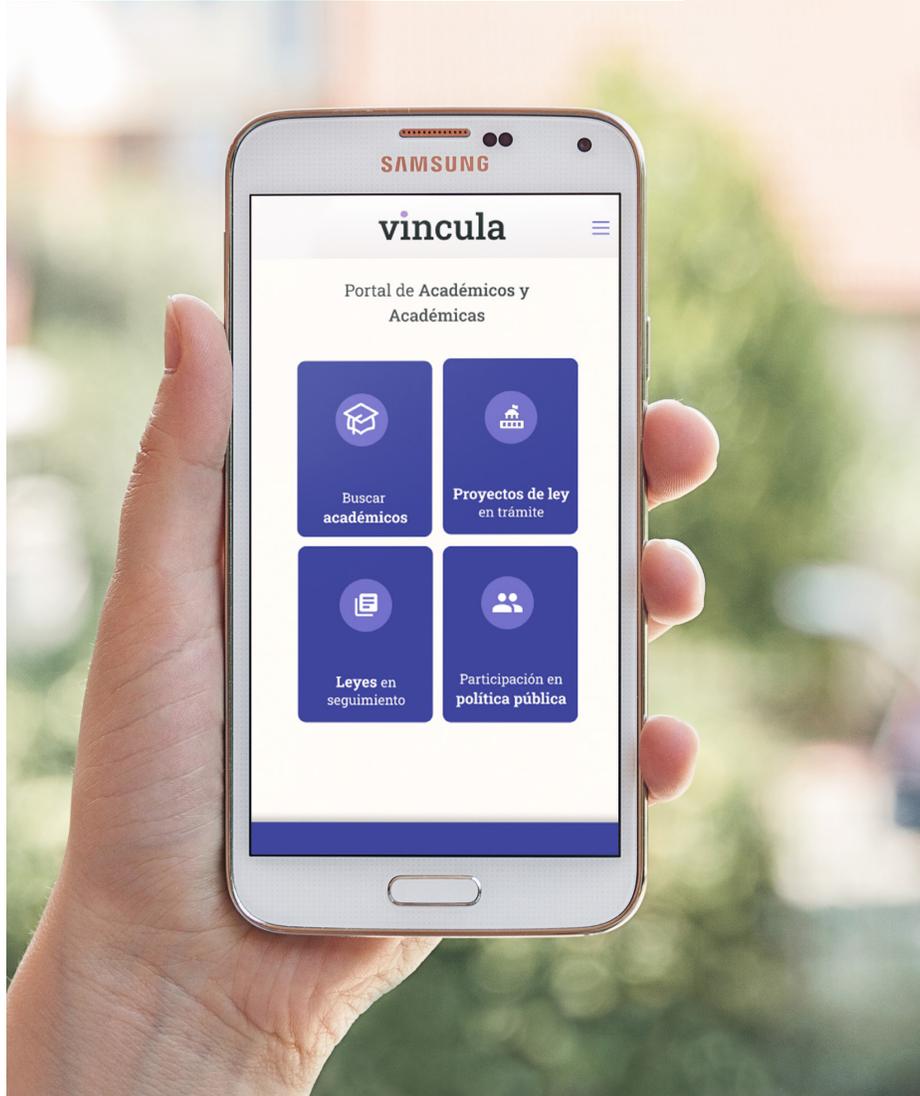




NLP al servicio de conectar el conocimiento científico con la creación de las leyes en Chile



JAZMINE MALDONADO

Magíster en Ciencias mención Computación por la Universidad de Chile. Actualmente es Directora de la Unidad de Innovación del Instituto Milenio Fundamentos de los Datos (Innovación IMFD) y profesora experta del Departamento de Ciencias de la Computación de la Universidad de Chile.

✉ jazmine@imfd.cl



RESUMEN. El Instituto Milenio Fundamentos de los Datos (IMFD), a través de su Dirección de Innovación se ha planteado, entre algunos de sus objetivos, apoyar a empresas e instituciones en el desarrollo de nuevas tecnologías y productos basados en el uso y el procesamiento de datos. Uno de los proyectos de innovación al servicio de la sociedad de los que el Instituto ha sido parte es la Plataforma de Encuentro Vincula. Este producto tecnológico se desarrolló como parte de la colaboración entre el IMFD y la red de colaboración Vincula, liderada por Francisca Reyes, académica de la Universidad Católica, y Andrea Rodríguez, académica de la Universidad de Concepción, y tiene como objetivo principal promover el intercambio de conocimientos entre las universidades chilenas y el Congreso.

Esta plataforma se integra con las universidades del país para obtener información validada y utiliza técnicas avanzadas de procesamiento del lenguaje para facilitar la creación de perfiles, potenciar la búsqueda de expertos y expertas, y detectar similitudes entre la investigación y los proyectos de ley. Además, presenta la información del avance de la tramitación de los proyectos de ley de forma intuitiva, permitiendo a los usuarios suscribirse para recibir actualizaciones sobre, por ejemplo, el cambio de estado de los proyectos de ley de su interés. Luego de una marcha blanca de seis meses, la Plataforma Vincula comenzó a operar masivamente en 2023, alcanzando, en abril, un total de 412 usuarios provenientes de seis universidades y del Congreso.

Introducción

La innovación es fundamental para el progreso y el desarrollo en todos los ámbitos de la sociedad: desde la tecnología y la economía, hasta la política y la cultura. En un mundo en constante evolución, innovar es esencial para mantenerse a la vanguardia y adaptarse a los cambios en el entorno.

Es por esto que el Instituto Milenio Fundamentos de los Datos, a través de su Dirección de Innovación (Innovación IMFD) se ha planteado, entre algunos de sus objetivos, apoyar a empresas e instituciones en el desarrollo de nuevas tecnologías y productos basados en el uso y el procesamiento de datos.

El equipo de Innovación IMFD, compuesto por científicos/as de datos, ingenieros/as de datos, desarrolladores de software y diseñadores/as de experiencia de usuarios, desde 2019 ha apoyado a diferentes instituciones públicas y privadas a enfrentar sus desafíos de gestión de datos, uso eficiente y estratégico, y en el desarrollo de productos innovadores.

En este artículo se presenta el proyecto Vincula (www.vincula.cl), de cuyo desarrollo hemos sido parte. La iniciativa es fundada y liderada por la Universidad Católica y la Universidad de Concepción desde el año 2021. Vincula es un espacio de encuentro entre productores de conocimiento y tomadores de decisión en el espacio legislativo, que busca promover una cultura de la evidencia que fortalezca el rol de la evidencia y el conocimiento de investigación en todas las etapas del proceso de formación de la ley. La dirección de Innovación IMFD se suma a este proyecto para desarrollar la Plataforma de Encuentro Vincula, una aplicación web que facilita el intercambio de conocimiento entre miembros del mundo académico y miembros del mundo legislativo. La Plataforma utiliza técnicas de procesamiento de lenguaje

para automatizar el proceso de conectar a miembros de ambos mundos de forma rápida y efectiva.

Después del periodo de marcha blanca de seis meses, en la que realizamos pruebas con usuarios y correcciones de usabilidad, la Plataforma Vincula ya está operando con 412 usuarios provenientes de seis universidades y del Congreso. Esperamos que, durante el segundo semestre del 2023, la Plataforma Vincula cuente con nuevas funcionalidades que serán fruto de la investigación y desarrollo que continuamos haciendo y se haya completado la integración de siete universidades más.

Plataforma de Encuentro Vincula

Promover el intercambio de conocimientos entre las universidades chilenas y el Congreso, es el objetivo de la Plataforma de Encuentro Vincula. Este es un producto tecnológico desarrollado por el equipo de Innovación IMFD, como parte de la colaboración entre el IMFD y la red de colaboración Vincula. Liderada por la Pontificia Universidad Católica y la Universidad de Concepción, Vincula es una innovación social, que tiene como finalidad que la mejor y más diversa evidencia y conocimiento que provenga de la ciencia y la investigación, sea utilizada en el proceso de análisis y creación de leyes en nuestro país.

La solución que ofrece Vincula elimina las barreras de acceso para quienes quieren aportar con conocimiento, al mismo tiempo que ofrece un programa de formación para lograr que investigadoras e investigadores logren una participación efectiva. Por último, aspira a reducir las brechas de género y centralismo existentes al visibilizar a las personas que hacen investigación en universidades de todo Chile.

El proceso de creación comenzó mucho antes de que el equipo de desa-

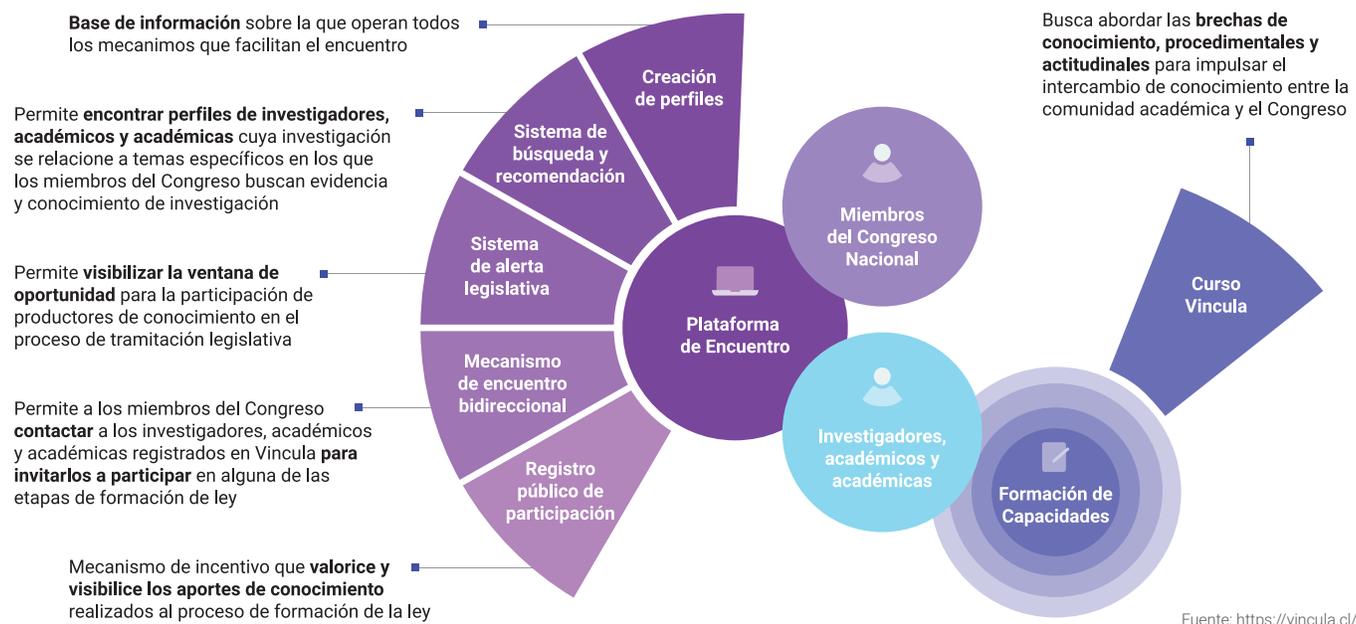


Figura 1. Principales componentes de la plataforma.

rollo de software escribiera el primer código. En 2020 el equipo directivo de Vincula, junto con la Escuela de Diseño de la Pontificia Universidad Católica, realizó una actividad de levantamiento de necesidades y cocreación que permitió identificar los principales “puntos de dolor” del público objetivo de Vincula. En esa actividad participaron representantes del mundo académico de diferentes universidades y representantes del sector legislativo, quienes —gracias al relato de sus experiencias— permitieron identificar las necesidades que Vincula resolvería.

A partir de ese levantamiento inicial se definieron los componentes principales de Vincula (ver Figura 1). Además, se priorizaron algunos aspectos como la simplicidad del registro, el uso de términos simples para referirse al estado del proceso de tramitación de un proyecto de ley, la capacidad de visibilizar los aportes realizados y facilitar el contacto bidireccional entre miembros del mundo académico y el mundo legislativo de una manera segura y efectiva.

Como resultado, la plataforma se integra con las universidades del país automatizando el registro de usuarios y utiliza técnicas avanzadas de procesamiento del lenguaje para facilitar la creación de perfiles, y potenciar la búsqueda de expertos y expertas. Además, presenta la información del avance de la tramitación de los proyectos de ley de forma intuitiva, permitiendo a los usuarios suscribirse para recibir actualizaciones, por ejemplo, cambio de estado de los proyectos de ley de su interés.

Procesamiento de lenguaje para vincular a la universidad y el Congreso

Las técnicas de procesamiento de lenguaje natural (natural language processing o NLP) han avanzado a pasos agigantados en los últimos años. El equipo de Innovación IMFD vio en este proyecto la oportunidad de aplicar técnicas de

análisis de texto, para agilizar algunos procesos que ocurren dentro de la plataforma. Además, con el salto que observamos en el año 2023 con el modelo de lenguaje GPT 3.5, se está desarrollando una actualización que promete mejorar la usabilidad y la efectividad de algunos de los algoritmos.

Creación automática de perfiles

Como se mencionó previamente, la plataforma intenta simplificar el registro de académicos y académicas, al mismo tiempo que busca generar perfiles verificados de personas afiliadas a un centro académico que genera conocimientos. Para esto, la plataforma se integra con las universidades y centros de investigación, quienes validan la afiliación de los usuarios a su institución y proveen información actualizada sobre publicaciones académicas y proyectos en los que han participado.

La creación de perfiles se compone de dos etapas, la primera consiste en generar perfiles de expertos y expertas



anónimos y se ejecuta al momento de integrar a la universidad o centro por primera vez con Vincula, y luego se actualiza una vez al mes. Para esto, el sistema usa únicamente información pública obtenida de publicaciones científicas, es decir, el título, *abstract* y *keywords*, y un ID generado sólo para este propósito y que representa a la persona.

La segunda etapa consiste en la desanonimización del perfil y se ejecuta al momento que el académico o académica se registra en la plataforma y autoriza el tratamiento de datos para los fines de Vincula. Es en esta etapa donde validamos con la universidad respectiva si el usuario que se acaba de registrar pertenece a la institución, solicitamos algunos datos adicionales para autocompletar su perfil y asociamos el perfil experto que se creó previamente de forma anónima con el usuario que se acaba de registrar.

La primera versión del algoritmo generador de perfiles utiliza Sentence-BERT¹ para seleccionar las frases clave más representativas del trabajo del autor a partir de los títulos, *abstracts* y *keywords* de las publicaciones científicas (ver Figura 2). Sentence-BERT es un modelo *transformer* multilinguaje preentrenado, capaz de mapear oraciones y párrafos a vectores densos en un espacio de 384 dimensiones y que ha sido probado para tareas como *clustering* y búsqueda semántica. Sentence-BERT genera *embeddings* para oraciones usando redes BERT siamesas que reducen el tiempo para encontrar pares semánticamente similares de 65 horas a alrededor de cinco segundos, manteniendo la exactitud de BERT.

En resumen, el algoritmo genera todas las frases clave posibles compuestas por una, dos o tres palabras a partir de las publicaciones de un autor (frases candidatas). Luego, utilizando Sentence-BERT, genera *embeddings* para

Innovar es esencial para mantenerse a la vanguardia y adaptarse a los cambios en el entorno.



Figura 2. Frases representativas de un usuario de la academia perteneciente a la Universidad de Concepción.

las publicaciones y para todas las frases candidatas generadas, para posteriormente seleccionar las frases más representativas utilizando la medida de similitud coseno. Esto permite determinar cuáles frases candidatas son las más similares al documento original y este índice de similitud se utiliza como una métrica de “relevancia” del concepto dentro del trabajo de cada autor.

Durante la generación de las frases representativas para cada autor, hay un paso para resumir el conjunto de frases clave y para ello se utilizan técnicas de detección de comunidades para detec-



Figura 3. Sistema de búsqueda de expertos y expertas.

tar frases representativas que semánticamente son muy similares entre sí, y así seleccionar únicamente las que representan conceptos diferentes para la creación del perfil.

Siguiendo este proceso, cada académico o académica que tiene publicaciones y que pertenece a una universidad o centro de investigación integrado en Vincula cuenta con un perfil anónimo en la plataforma. Este perfil anónimo representa la existencia de un experto o experta asociado a una universidad específica que ha trabajado en un conjunto de temáticas.

¹ <https://arxiv.org/abs/1908.10084>.



Sistema de búsqueda de expertos y expertas

La aplicación permite a todos los usuarios buscar expertos o expertas ingresando una frase de búsqueda (ver Figura 3). Cada vez que se realiza una búsqueda se calcula un puntaje para cada perfil académico. El puntaje se calcula considerando la frase de búsqueda, las frases clave que conforman el perfil y la relevancia de cada frase en las publicaciones del autor o autora tal como indica la siguiente fórmula.

$$\text{Puntaje} = \text{similitud}(\text{frase búsqueda}, \text{frase clave}_i) \times \text{relevancia}_i$$

En otras palabras, para cada frase clave de un perfil académico, se calcula la similitud entre la frase de búsqueda y una frase clave, y luego la similitud se pondera por la relevancia de cada frase dentro del total de publicaciones del autor o autora. Finalmente, todos estos valores se suman y se obtiene el puntaje del académico o de la académica para esa búsqueda.

Para realizar este proceso se utiliza el mismo modelo *transformers* multilinguaje utilizado para la creación de perfiles y la medida de similitud coseno.

El sistema de búsqueda opera sobre todos los perfiles académicos creados, ya sean perfiles correspondientes a usuarios registrados o no. Por lo tanto, al mostrar los resultados en la plataforma, se utiliza un código de color y simbología específica para presentar los perfiles que aún son anónimos.

Seguimiento de proyectos de ley y sistema de alerta legislativa

En la plataforma los usuarios también pueden revisar intuitivamente los proyectos de ley que se encuentran en tramitación. Para permitir esta funcionalidad, la plataforma se conecta a los sistemas del Senado de Chile, los que proveen información actualizada sobre el estado de los diferentes proyectos de ley en tramitación,

así como también, información actualizada de quienes integran las comisiones que tramitan cada proyecto y sus integrantes.

En relación a los proyectos de ley la plataforma Vincula ofrece diferentes funcionalidades. La principal es que los usuarios también pueden suscribirse a los proyectos de ley de su interés y recibir actualizaciones por correo cuando estos cambian de estado y avanzan en su tramitación. Esto facilita que se mantengan informados del proceso y puedan contribuir de mejor manera, en caso de querer hacerlo.

Además de esto, los académicos y académicas pueden aportar documentación o informar sobre su participación en comisiones relacionadas a la ley, antecedentes que quedan disponibles para otros usuarios de la aplicación, tanto en la página del proyecto de ley, como en el perfil del académico(a). También podrán ver quiénes son los integrantes de la comisión que tramita cada proyecto y contactar a los parlamentarios o sus asesores vía correo electrónico.

Siguientes pasos

Con el lanzamiento de GPT 3.5, el equipo de investigación y desarrollo de Vincula ya está trabajando en mejorar algunas de las funcionalidades utilizando este nuevo modelo de lenguaje.

A nivel de prototipo ha resultado ser muy útil para generar taxonomías a partir de los títulos y *abstracts*. Es decir, usando este modelo es posible instruir a la máquina para generar un tipo de respuesta determinado, dado un tipo de *input* específico. En este caso, la entrada son títulos y *abstracts*, y el resultado es una clasificación jerárquica de las áreas de conocimiento representativas. Esto permitiría de manera preliminar mejorar la creación de perfiles de expertos y expertas, ya que —en vez de presentar un perfil compuesto por frases clave generadas exhaustivamente y luego priorizadas por relevancia— sería posible generar un árbol de áreas de conocimiento ordena-

das por relevancia que, en términos de usabilidad y precisión, facilitaría mucho la interpretación a los usuarios.

Por otro lado, siguiendo esta misma línea, esperamos poder utilizar esta técnica para analizar los documentos públicos de los proyectos de ley en tramitación haciendo un ejercicio similar, es decir, generar una taxonomía de áreas de conocimiento científico a partir de los documentos que describen el proyecto. Esto permitirá incorporar una nueva funcionalidad, que consiste en poder recomendar proyectos de ley a los académicos y académicas registrados, detectando automáticamente cuándo un nuevo proyecto de ley está relacionado con el trabajo que un experto o experta ha venido desarrollando.

El desafío de acceder a la información

Para funcionar, Vincula se integra con las universidades, recibiendo información clave relacionada con la productividad académica y verificando la afiliación de quienes se registran en la plataforma.

La información provista por las universidades es la que utiliza el algoritmo para crear los perfiles de los expertos y expertas, y la que permite hacer la conexión entre proyectos de ley y las personas que han estudiado temáticas afines.

Esto significó un esfuerzo conjunto por parte de los equipos de tecnología y vicerrectorías de investigación de las universidades, unidades que —además de conocer y comprometerse con la visión del proyecto—, trabajaron internamente para tener la información necesaria disponible y preparar las APIs que permiten la integración con la plataforma Vincula.

“Las universidades almacenan esta información, porque es muy relevante para su gestión en términos de investigación y productividad académica. Sin embargo, no siempre tienen todos los campos que el sistema necesita. Por ejemplo, no



siempre están completos los *abstracts* o resúmenes de las publicaciones, y esto es muy importante para que el sistema pueda “entender” de qué se trata una publicación y los temas que aborda”, explica Roberto Vargas, desarrollador de software en Innovación IMFD, quien es el encargado de la integración de Vincula con las universidades.

Por otro lado, la plataforma Vincula utiliza información de proyectos de ley que es pública y que cualquier persona puede conocer a través de la página del Senado. Para esto, el equipo de Tecnología del Senado contribuyó de forma muy eficiente y efectiva en el despliegue de las APIs que la Plataforma Vincula utiliza para obtener los datos y mostrar intuitivamente los proyectos de ley y su avance en la tramitación. Además, la plataforma muestra en cada momento quiénes son los parlamentarios/as que conforman las comisiones que están trabajando en el proyecto, permitiendo el contacto a través de sus correos públicos o a través de sus asesores parlamentarios que están registrados en Vincula.

Uso de datos y privacidad

Desde su diseño, la plataforma Vincula tiene dentro de sus objetivos respetar cabalmente los derechos de los usuarios titulares sobre sus datos personales y promover los principios fundamentales de la protección de datos personales, que consisten en el principio de licitud, lealtad y transparencia, el principio de limitación de la finalidad, el principio de minimización de datos, el principio de exactitud, el principio de limitación del plazo de conservación y el principio de integridad y confidencialidad.

Es por esto que la aplicación Vincula solicita a las universidades información en su mayoría de carácter público que consiste en metadatos de publicaciones

científicas y proyectos, tales como, títulos, *abstracts* o descripción, *keywords*, año y DOI. Esta información se provee de forma anónima.

Luego, únicamente después de la autorización explícita de los usuarios al momento de registrarse en la aplicación, Vincula también solicita información personal. Los datos personales que Vincula solicita son RUT, nombre, apellidos, género, año de nacimiento, cargo en la institución y ORCID. El RUT y el correo electrónico que utiliza el usuario al registrarse nunca serán públicos y otros datos, como el género y el año de nacimiento, podrán ser omitidos del perfil a voluntad del académico o académica.

Como se explicó anteriormente, todas las acciones internas de la aplicación que tienen relación con la generación de perfiles expertos o la recomendación de proyectos de ley, sólo consideran la información de carácter público, como metadatos de publicaciones y proyectos y documentos públicos de proyectos de ley.

Por último, cada usuario titular tiene el derecho a acceder a la información respecto de cómo fueron recolectados, tratados y almacenados sus datos, y con qué propósito, el derecho a rectificar la información inexacta o incompleta, el derecho a suprimir todos o algunos de sus datos, el derecho a oponerse momentáneamente a que sus datos sean utilizados y el derecho a la portabilidad de sus datos, pudiendo recibir una copia de sus datos personales.

Agradecimientos

Los desafíos de desarrollar un producto tecnológico como la Plataforma de Encuentro Vincula son múltiples y el equipo de Innovación IMFD contribuyó a resolver muchos de ellos. Sólo para desarrollar el software era necesaria una visión amplia que combinara el desarrollo de software, el conocimiento en ciencia de

Vincula es un espacio de encuentro entre productores de conocimiento y tomadores de decisión en el espacio legislativo.

datos, las regulaciones para el tratamiento de datos, la automatización de procesos y también conocimientos generales sobre la gestión de datos universitarios o el proceso legislativo chileno. El equipo de Innovación IMFD que trabajó en este proyecto está compuesto por Camilo Garrido (científico de datos), Elena Villalón (analista de datos), Etienne Toussaint (Ph.D., IMFD), Gabriela Contreras (desarrolladora *front-end*), Jazmine Maldonado (directora de Innovación IMFD), Juan José Alegría (científico de datos), Julio Hurtado (diseñador UX), Marcelo Arenas (investigador y director del IMFD), Patricio Isbej (desarrollador *full-stack*), Roberto Vargas (desarrollador *full-stack*), Rodrigo Oportot (practicante, Ingeniería en Computación, U. Chile) y Vicente Contreras (practicante, Ciencia Política, UC).

Finalmente, todo esto fue posible gracias a un liderazgo efectivo y altamente comprometido por parte de la Directora de Vincula, Francisca Reyes, académica del Instituto para el Desarrollo Sustentable de la Pontificia Universidad Católica y directora de la iniciativa Votaciones Ambientales en el Congreso Nacional, y de la la codirectora del proyecto, Andrea Rodríguez, vicerrectora de Investigación y Desarrollo de la Universidad de Concepción, quienes trabajaron arduamente para dar forma a la idea e iterarla hasta lograr el resultado final, crear grandes equipos colaborativos y multidisciplinarios, sumar a personas comprometidas con el proyecto y acercar a representantes del Senado, de la Cámara de Diputados, de la Biblioteca del Congreso Nacional y de las vicerrectorías académicas de todas las universidades que hoy son parte de Vincula. ■