



Universidad Tecnológica Nacional
Rectorado
Secretaría de Ciencia, Tecnología y
Posgrado

SISTEMA DE INFORMACION DE CIENCIA Y
TECNOLOGIA (SICyT)

FORMULARIO PARA PROYECTOS DE INVESTIGACIÓN Y DESARROLLO

Código del Proyecto: SIUTNCO0007848

1. Unidad Científico-Tecnológica

- FR Córdoba - DEPARTAMENTO DE INGENIERIA EN SISTEMAS DE INFORMACIÓN
- FR Córdoba - CIDS - CENTRO DE INVESTIGACIÓN, DESARROLLO Y TRANSFERENCIA DE SISTEMAS DE INFORMACIÓN

2. Denominación del PID

Análisis cuantitativo de la producción en investigación científica y tecnológica en la Red de Ingeniería en Informática / Sistemas de Información de CONFEDI

3. Resumen Técnico del PID

El proyecto caracterizará la producción científica y tecnológica desarrollada por las universidades miembro de la Red de Ingeniería en Informática / Sistemas de Información de CONFEDI (RIISIC), por medio de la elaboración de una metodología de análisis cuantitativo a partir de la documentación producida por los investigadores, becarios y centros de investigación de las casas de estudios que componen la red. Se realizará un análisis de la información contenida en la documentación y se desarrollarán técnicas y herramientas para la extracción de la misma y el almacenamiento en un medio que permita el correcto análisis cuantitativo. Parte de este desarrollo incluye una plataforma web para la gestión administrativa y académica del Congreso Nacional de Ingeniería en Informática / Sistemas de Información (CoNallSI), obteniendo así la información de análisis en tiempo real. El insumo para el análisis son los artículos enviados por los autores. Está considerado el modelado, diseño y desarrollo de una herramienta de análisis para obtener indicadores, métricas y patrones, en base a la información almacenada en la base de datos cuantitativa, que permitan la visualización simple y efectiva de los datos registrados y que oficie de mecanismo de consulta general para elaborar informes y análisis.

4. Programa

Sistemas de Información e Informática

5. Proyecto

Tipo de Proyecto: UTN (PID UTN) SIN INCORPORACION EN PROGRAMA INCENTIVOS

Tipo de Actividad: Investigación Aplicada

Campos de Aplicación:

Rubro	Descrip. Actividad	Otra (especificada)
INDUSTRIAL (Producción y tecnología)	Otros - Industrial (Especificar)	Software para investigación

Disciplinas Científicas:

Rubro	Disciplina Científica	Otras Disciplinas Científicas
CIENCIAS DE LA COMPUTACIÓN E INFORMÁTICA	Ingeniería en Sistemas de Información	-

Palabras Clave

cuantitativa - producción científica - métricas - big data - data mining

6. Fechas de realización

Inicio	Fin	Duración	Fecha de Homologación
01/01/2020	31/12/2021	24 meses	-

7. Aprobación/ Acreditación / Homologación / Reconocimiento (para ser completado por la SCTyP - Rectorado)**7.1 Aprobación / Acreditación / Reconocimiento (para ser completado por la FR cuando se posea N° Resolución)**
N° de Resolución de aprobación de la FR:**7.2 Homologación (para ser completado por la SCTyP - Rectorado)**

Código SCTyP: SIUTNCO0007848

Disposición SCTyP:

Código Ministerio:

8. Estado (para ser completado por la SCTyP - Rectorado)

EN TRÁMITE: EVALUACIÓN CONSEJO DE PROGRAMAS

9. Avals (presentación obligatoria de avals)**10. Personal Científico Tecnológico que participa en el PID**

Apellido y Nombre	Cargo	Hs/Sem	Fecha Alta	Fecha Baja	Otros Cargos
CUEVAS, JUAN CARLOS	CO-DIRECTOR	12	01/01/2020	31/12/2021	-
MUÑOZ, ROBERTO MIGUEL	DIRECTOR	15	01/01/2020	31/12/2021	-
MALDONADO, CALIXTO	INVESTIGADOR FORMADO	10	01/01/2020	31/12/2021	-
GUZMAN, ANALÍA	INVESTIGADOR DE APOYO	10	01/01/2020	31/12/2021	-
CASATTI, MARTÍN GUSTAVO	INVESTIGADOR DE APOYO	10	01/01/2020	31/12/2021	-
DAMIANO, LUIS ESTEBAN	INVESTIGADOR DE APOYO	10	01/01/2020	31/12/2021	-
BENITO, FEDERICO	BECARIO BINID	10	01/01/2020	31/12/2021	-
ACHÁVAL VINUESA, JULIO GUILLERMO	BECARIO ALUMNO UTN-SAE	10	01/01/2020	31/12/2021	-
TOLEDO ALONSO, RODRIGO	BECARIO ALUMNO UTN-SAE	10	01/01/2020	31/12/2021	-
WISMER, AXEL FRANCIS	BECARIO ALUMNO UTN-SCYT	10	01/01/2020	31/12/2021	-

11. Datos de la investigación**Estado actual de concimiento del tema**

"Lo que no se define no se puede medir. Lo que no se mide, no se puede mejorar. Lo que no se mejora, se degrada siempre". Esta cita pertenece a William Thomson Kelvin (Lord Kelvin), físico y matemático británico (1824 – 1907). La cienciometría (scientometrics) como disciplina se denomina al "estudio de los aspectos cuantitativos de la comunicación científica, las prácticas de investigación y desarrollo, así como las políticas de ciencia y tecnología"[1]. Con la aparición de big data se logra, a la vez, que grandes conjuntos de datos puedan ahora ser tratados y utilizados con fines analíticos y de predicción, lo que hasta hace poco era imposible de procesar mediante el uso de herramientas tradicionales debido a su tamaño y complejidad. La combinación de big data, DataMining, bibliometría y redes sociales, ha dado lugar a lo que se ha denominado como science mapping [2] cuyo objetivo es visualizar la estructura y relaciones de la ciencia en general, y de las áreas científicas, disciplinas, líneas de investigación, autorías, etc., en particular.

Ningún estudio científico de importancia es completamente independiente y autocontenido. Todos los trabajos de investigación se basan parcialmente en hallazgos o trabajos previos, los cuales se sostienen para producir un avance propio y de esta manera corren hacia adelante los límites del conocimiento.

El reconocimiento y las referencias, sobre las que un investigador o equipo ha trabajado para construir el conocimiento propio, se representan por medio de citas.

Una cita es una referencia a un trabajo, ya sea propio o de otros autores o grupos, que se ha tomado total o parcialmente para construir el propio trabajo de investigación.

En cierta medida el tipo y cantidad de citas que un trabajo de investigación registra es un indicador, con mayor o menor grado de fidelidad, del impacto que dicho trabajo produce en la comunidad científica. Un trabajo muy citado es la base para diversos papers científico y líneas de investigación, mientras que un trabajo raramente citado no produce un efecto importante en la comunidad.

Otro análisis importante que se puede realizar por medio del análisis de citas es el de los patrones de colaboración entre instituciones o entre investigadores. Las redes de coautoría son una importante clase de redes sociales. El análisis de estas redes revela las características de las comunidades académicas que ayudan en el entendimiento de los trabajos científicos colaborativos y en la identificación de los investigadores prominentes [3].

Es importante destacar que el análisis de redes de co-autoría puede ayudar a caracterizar una cierta población de investigadores, determinando cuales son las temáticas de interés y la producción total y relativa por área temática. Si se considera que las relaciones entre los investigadores se puede representar como un grafo no dirigido existe una métrica que es de especial importancia a la hora de caracterizar dicho grafo. Dicha métrica es el grado de centralidad.

La centralidad puede indicar, en un grafo que represente co-autoría de trabajos de investigación, aquellos autores que tienen un grado alto de colaboración con otros autores. Por otra parte, una métrica asociada, la centralidad de

intermediación, puede indicar aquellos autores que ofician de nexo entre otros investigadores o grupos de investigación.

Por otra parte, se pueden utilizar las redes de co-autoría para evaluar la interrelación entre las instituciones académicas, el sector gubernamental y el sector productivo [7]. De esta manera es posible identificar el impacto generado por esa interrelación y se pueden diseñar políticas tendientes a reforzar esos vínculos para generar resultados beneficiosos a mediano y largo plazo.

Si bien los indicadores cuantitativos absolutos son útiles, también existe un amplio campo de aplicación para el análisis entre indicadores, o indicadores relativos.

Dentro de los estudios comparativos llevados a cabo por métodos cuantitativos tenemos las comparativas entre las publicaciones realizadas por distintos países, las publicaciones de acuerdo al idioma en el que se publica y el ratio entre el impacto producido por una institución comparado con el promedio de impacto mundial dentro de la misma área de interés.

Un estudio de extrema importancia, sobre todo en economías en donde la investigación científica y tecnológica no es prioritaria en las políticas de gobierno, es el análisis que establece relaciones entre la producción científica y tecnológica y el ámbito de los investigadores sobre todo desde el punto de vista socioeconómico [1] [4].

Desde este punto de vista también se puede analizar el surgimiento o decadencia de ciertas líneas de investigación.

Si bien la información numérica y tabulada reviste un gran importancia, una de las líneas de investigación que más auge tiene en los últimos años es la denominada "Knowledge Mapping" o mapas de conocimiento. Se conoce como mapa de conocimiento al conjunto de técnicas, métodos y herramientas, destinadas a representar de forma intuitiva y clara, las relaciones existentes entre los componentes de la investigación científica [5], siendo el "knowledge mapping" un subconjunto de una disciplina más amplia y general conocida como "science mapping".

"El objetivo central del "science mapping" es el de revelar la estructura y dinámica del conocimiento científico...", tal como menciona el trabajo de Chen, Dubin y Schultz [9]. Los análisis cuantitativos sirven como punto de partida para contar con información cuantitativa, heterogénea y relacionada para poder plantear análisis de mapping [8]. Existen tres ramas de estudio que comprenden la utilización de la información como fuente para realizar análisis cuantitativos. Dichas ramas se diferencian desde el punto de vista de su objeto de estudio, las variables que maneja cada uno de ellos, los métodos que utiliza y los objetivos de dichos análisis.

Si bien es indiscutible que las tres ramas tienen un importante grado de solapamiento y que a menudo se unifican bajo el término más general de "cienciometría", es conveniente analizar las particularidades de cada una, desde el punto de vista de los trabajos de McGrath [6]:

Tipología	Bibliometría	Cienciometría	Informetría
Objeto de estudio	Libros, documentos, revistas, artículos, autores y usuarios.	Disciplinas, materias, campos y esferas.	Palabras, documentos y bases de datos.
Variables	Números en circulación, citas, frecuencia de aparición las palabras, longitud de las oraciones, etc.	Aspectos que diferencian a las disciplinas y a las subdisciplinas. Revistas autores, trabajos, forma en que se comunican los científicos.	Difieren de la cienciometría en los propósitos de las variables, por ejemplo, medir la recuperación, la relevancia, el recordatorio, etc.
Métodos	Clasificación, frecuencia, distribución.	Análisis de documento y correspondencia.	Modelo rector-espacio, modelos booleanos de recuperación, modelos probabilísticos, lenguaje del procesamiento, enfoques basados en el conocimiento, tesauros.
Objetivos	Asignar recursos, dinero, tiempo, etc.	Identificar esferas de interés; dónde se encuentran las materias; comprender cómo y con qué frecuencia se comunican los científicos.	Aumentar la eficiencia de la recuperación

Hay que mencionar que a la hora de optimizar el uso de recursos escasos, como pueden ser los lugares físicos para laboratorios, las plazas en las plantas de investigadores permanentes, los recursos económicos, de materiales o de insumos, es fundamental contar con un diagnóstico adecuado de la situación actual del ámbito en donde esos recursos se van a utilizar.

En este tema en particular el análisis cuantitativo puede servir para tener una imagen actualizada y confiable de la infraestructura de I+D+i a los fines de asistir en la toma de decisiones.

[1] RENDÓN, ESMERALDA CERVANTES AND VICTORIANO GARZA-ALMANZA (2015). "La cienciometría como herramienta para analizar el impacto de la investigación científica en una región." Cultura Científica y Tecnológica 48.

[2] M.J. COBO, A.G. LÓPEZ-HERRERA, E. HERRERA-VIDEIRA AND F. HERRERA (2011). "Science mapping software tools: Review, analysis, and cooperative study among tools". Journal of the American Society for Information Science and Technology, vol. 62, nº 7, pp. 1382-1402.

[3] UMADEVI, V. (2013). Case study—centrality measure analysis on co-authorship network. Journal of Global Research in Computer Science, 4(1), 67-70.

[4] LEYDESDORFF, LOET, AND ISMAEL RAFOLS (2011). "Local emergence and global diffusion of research technologies: An exploration of patterns of network formation." Journal of the American Society for Information Science and Technology 62.5: 846-860.

[5] EBENER, STEVE, et al. (2006). "Knowledge mapping as a technique to support knowledge translation." Bulletin of the World Health Organization 84: 636-642.

[6] MCGRATH, W. (1989). What bibliometricians, scientometricians and informetricians study; a typology for definition and classification; topics for discussion. In International Conference on Bibliometrics, Scientometrics and Informetrics (Vol. 2). Ontario: The University of Western Ontario.

- [7] SÁNCHEZ, M. A., SCHMIDT, M. A., & OBIOL, L. C. (2018). Redes de coautorías y patrones de colaboración institucional: análisis de un caso en el campo de la ingeniería.
- [8] ROBLES, E., & VINCK, D. (2010). Mapeo de la evolución de la producción de ciencias y tecnologías emergentes: comparación de estrategias lexicales sobre las nanociencias y nanotecnologías. In ESOCITE 2010.
- [9] CHEN, C., DUBIN, R., & SCHULTZ, T. (2015). Science mapping. In Encyclopedia of Information Science and Technology, Third Edition (pp. 4171-4184). IGI Global.
- [10] Plataforma OCS. <https://pkp.sfu.ca/ocs/>. Accedida el 26/05/2019.

Grado de Avance

Respecto al Congreso Nacional de Ingeniería en Informática / Sistemas de Información (CoNallSI), que organiza la Red de Ingeniería en Informática / Sistemas de Información de CONFEDI (RIISIC) desde el año 2013, el Departamento de Ingeniería en Sistemas de Información de UTN-FRC organizó la primera edición y a partir de allí ha administrado las demás ediciones. Cabe aclarar que en la red participan más de 30 Facultades del país que implementan la terminal y que el Congreso es anual, cambiando la sede en cada edición y actualmente se trabaja en la organización de la 7ma. edición para noviembre de 2019.

Es decir que en el equipo de investigación cuenta con la experiencia de haber organizado el CoNallSI y además posee los datos de todas las ediciones, por lo que dichos archivos históricos contienen todos los artículos y trabajos de investigación enviados a los CoNallSI en todas sus sedes. Esta información está disponible para la realización de todas las tareas de análisis y modelado de datos, así como las pruebas necesarias durante el desarrollo del software de extracción de atributos, que deberán ser importadas a la base de datos, con el proceso necesario, y dando el formato correspondiente para la gestión desde la plataforma propuesta en el proyecto.

En CoNallSI se utiliza la plataforma libre Open Conference Systems (OCS) [10] y personal docente de UTN-FRC administró el sitio de OCS, instalado en un servidor propio de la Facultad. Por ello se conocen concretamente las fortalezas del software y las funcionalidades necesarias para administrar un congreso como CoNallSI.

Existe un conjunto de actividades desarrolladas por los docentes y estudiantes investigadores, en el marco de sus respectivos proyectos en ejecución, que brindan una base importante para el desarrollo del presente trabajo, así como también se cuenta con acceso a información de gran valor para la concreción del presente.

Los Ing. Esp. Analía Guzmán y Martín Casatti han participado, durante los últimos 4 años, en proyectos que utilizan bases de datos de grafos como mecanismo de almacenamiento de información, siendo éste además el título del trabajo final integrador del Ing. Casatti (Bases de datos de grafos como mecanismo de almacenamiento de información heterogénea) con el que obtuvo el título de Especialista en Ingeniería de Sistemas.

Otro integrante del equipo de investigación, el Ing. Calixto Maldonado, cuenta con importante experiencia en el ámbito de Bases de Datos y específicamente en Big Data. Actualmente es el Adjunto a cargo de la asignatura "Big Data: Arquitectura y Estrategias para el Análisis de Datos Masivos", electiva de 5to año de Ingeniería en Sistemas de Información, de la UTN-FRC. También es Jefe de Trabajos Prácticos Ordinario en Gestión de Datos, asignatura curricular de 3er año de Ingeniería en Sistemas de Información, de la UTN-FRC. Además Maldonado ha participado en proyectos asociados a las bases de datos, es certificado en Oracle y se desempeña profesionalmente como Consultor Informático.

Algo semejante sucede con el Lic. Luis Damiano que se desempeña como docente en asignaturas del área Programación y con acceso a bases de datos relacionales, además cuenta con continuidad en la participación en proyectos de investigación donde el tema central son las bases de datos y en la faz profesional es desarrollador en ámbitos del Gobierno de la Provincia de Córdoba.

El codirector propuesto, Lic. Juan Carlos Cuevas, es docente y coordinador de las asignaturas electivas del 5to año de la carrera de Ingeniería en Sistemas de Información: Gestión de la Calidad y Auditoría de SI/TI. En la actividad profesional fue Auditor en la Empresa Provincial de Energía de Córdoba, consultor en temas de Calidad y Auditoría. En materia de Investigación ha dirigido dos proyectos vinculados a las temáticas de riesgo en mantenimiento de software y vulnerabilidades en aplicaciones web y ha sido co-director en un proyecto de Bases de Datos. Por lo expuesto el principal aporte a realizar está vinculada a la medición como paso imprescindible para conocer el dominio, para su ulterior mejora continua, como así también en los aspectos relacionados a la formalización de la gestión del conocimiento que se genere a partir del presente proyecto.

Una mención especial merece la participación del Ing. Esp. Roberto Muñoz, quien coordina la cátedra de Gestión de Datos y actualmente es Director de la Carrera de Ingeniería en Sistemas de Información de UTN-FRC. Al ser director es el representante, por UTN-FRC, ante la red RIISIC y se desempeñó como coordinador de la misma durante el año 2017. Esto provee un vínculo directo con la red y las universidades miembro, a los contenidos generados por sus equipos de investigación y publicados en el congreso.

Respecto a los estudiantes becarios han sido seleccionados en función de la relación de ellos con los temas a investigar:

- Julio Achaval: es un estudiante que ha participado del proyecto "Modelo de Análisis para la derivación de requerimientos funcionales a partir de la implementación de Patrones en la construcción del Modelo Conceptual", desde el año 2018. Ha presentado un trabajo en WICC 2019 y fue colaborador de los eventos organizados en UTN-FRC: 46 JAIIO / XLIII CLEI y en las VI JEIN.
- Federico Benito: es un estudiante que sólo adeuda 2 asignaturas para recibirse como Ingeniero en Sistemas de Información, ha sido becario en el Laboratorio de Ingeniería en Sistemas durante 2017 y luego accedió a ser Auxiliar de Segunda en dicho laboratorio. Posee experiencia como becario en Investigación, habiendo participado del proyecto: Automatización de Corrección de Exámenes, mediante Grafos Dirigidos, del Laboratorio de Investigación de Software de la UTN-FRC. También ha presentado artículos en el 4to. Congreso Nacional de Ingeniería en Informática/ Sistemas de Información (CoNallSI 2016).
- En el caso de Axel Wismer y Rodrigo Toledo Alonso son estudiantes con alto rendimiento académico y que desean iniciarse en la investigación. Ambos poseen experiencia laboral y específicamente en modelado y desarrollo de software.

Las particularidades de UTN FRC también proveen un valor muy importante para la ejecución de este proyecto, debido a que se cuenta con un vínculo directo con las otras doce (12) Facultades Regionales de UTN que poseen ofrecen la carrera de Ingeniería en Sistemas de Información.

Otro aspecto destacable es que el Departamento de Ingeniería en Sistemas de Información de UTN FRC cuenta con una larga y fructífera trayectoria en la organización y administración de Congresos y Jornadas académicas, por ejemplo:

- V Jornadas Nacionales de Metodologías Ágiles en 2018.
- Jornadas de Enseñanza de la Ingeniería (JEIN) en 2018.
- Jornadas Argentinas de Informática (JAIIO) en 2011 y 2017.
- Jornada del PMI Capítulo Córdoba en 2016, 2017 y 2018.
- Olimpiadas Informáticas Argentinas (OIA) de Programación en 2016.
- Oracle Developer Tour Argentina en 2017.
- Jornada Nacional de Comercio Electrónico y Community Management (CoECoM) en 2017, 2018 y 2019.
- 1er Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNallSI) en 2013.

El equipo ha desarrollado algunas actividades previas que aportarán avances importantes para el desarrollo del presente proyecto:

- Evaluación y prueba de librerías de captura de texto a partir de archivos PDF. Dichas librerías, como por ejemplo pyPDF, pyPDF2 o pdfwr, sirven como punto de partida para la elaboración del software de extracción y captura de información textual obtenida a partir de los documentos PDF de trabajos de investigación y artículos del congreso que se consideren para el análisis cuantitativo.
- Análisis de herramienta de procesamiento de textos en lenguaje natural, la cual podría tomar como entrada la información proveniente de los archivos PDF de los trabajos de investigación y reconocer ciertas estructuras importantes, como ser los abstract, la lista de autores, las palabras clave y las citas bibliográficas. La herramienta analizada hasta el momento es spaCy pero también se pueden evaluar algunas alternativas, como NLTK, textaCy, polyglot o prattNLPTools. Esto se considera necesario porque, en algunas muestras, se han detectado documentos que no cumplen con los formatos establecidos por la organización, pero que sí cuentan con información relevante la cual deberá ser extraída con técnicas de minería de texto.

Objetivos de la investigación

Objetivo general: Caracterizar la producción e impacto de las líneas de investigación que llevan adelante las universidades pertenecientes a la red RIISIC por medio de una herramienta desarrollada ad-hoc.

Objetivos particulares:

1. Modelar la representación de la información asociada a las tareas de I+D que permita caracterizar todos los trabajos producidos por las universidades pertenecientes a la red RIISIC.
2. Estructurar un sistema de almacenamiento cuantitativo que permita la registración y consulta de la información registrada.
3. Procesar los documentos de los congresos, extrayendo de los mismos los atributos para el análisis cuantitativo e infiriendo lo que no se obtiene directamente de los textos.
4. Diseñar, desarrollar e implementar un sistema de gestión de congresos y conferencias que registre automáticamente la información en el sistema de almacenamiento, a partir de los papers y artículos presentados, sin necesidad de interacción adicional o carga manual de datos.
5. Desarrollar una herramienta de consulta y análisis que permita obtener métricas asociadas a la producción científica y que posibilite analizar los diversos patrones que pudieran surgir a partir del análisis de citas, autores y líneas de investigación.

Descripción de la metodología

Se utilizará el método científico en base a un enfoque experimental, con una duración total de 24 meses, el cual a los fines de la correcta ejecución se analizó la metodología a seguir, estando las tareas principales enumeradas en cada una de ellas.

Para cumplimentar el **objetivo particular 1:**

“Modelar la representación de la información asociada a las tareas de I+D que permita caracterizar todos los trabajos producidos por las universidades pertenecientes a la red RIISIC.”

Para modelar es necesario estudiar las características que debe reunir la información asociada a las tareas de I+D de forma tal que posibiliten el análisis cuantitativo.

A partir de esa caracterización es posible desarrollar un modelo de representación para el almacenamiento de la información, de forma tal que se mantenga la mayor cantidad posible de atributos existentes en los paper y trabajos de investigación que se van a utilizar para el análisis.

Para cumplimentar el **objetivo particular 2:**

“Estructurar un sistema de almacenamiento cuantitativo que permita la registración y consulta de la información registrada.”

Esto hace necesario analizar las posibilidades de persistencia de datos según los modelos de datos existentes. Actualmente hay diversas alternativas de persistencia de la información y diversas herramientas existentes en el mercado, abarcando al modelo Relacional pero sin circunscribirse para así considerar modelos como NoSQL,

Clave/Valor, GraphDB, etc. De las alternativas disponibles se seleccionará la mejor opción y se procederá a su instalación y configuración en un entorno adecuado para comenzar con las pruebas de concepto.

Una vez generado el modelo será conveniente contrastarlo con diversos tipos de papers de investigación, de diversos congresos y variadas universidades, para verificar que el mismo sea aplicable a la mayor cantidad de escenarios posibles y brindar flexibilidad, sin quedar limitados por la información de CoNallSI. Aquí pueden surgir ajustes y adecuaciones en función de los datos a almacenar y del modelo elegido.

Se desarrollará una librería que procese automáticamente los papers enviados por medio de la aplicación, a fines de extraer toda la información necesaria para el análisis cuantitativo, la cual se almacenará directamente en la base de datos correspondiente, minimizando la operatoria manual y la interacción con un operador.

Para cumplimentar el **objetivo particular 3:**

"Procesar los documentos de los congresos, extrayendo de los mismos los atributos para el análisis cuantitativo e infiriendo lo que no se obtiene directamente de los textos".

Desarrollar una aplicación de importación masiva que permita cargar, en la base de datos cuantitativa, textos extraídos de la estructura de los documentos e informes de investigación, para contar con una base de datos completamente funcional.

Será necesario aplicar técnicas de minería de textos para detectar atributos no visibles directamente o no se infieren del formato de los documentos.

Para cumplimentar el **objetivo particular 4:**

"Diseñar, desarrollar e implementar un sistema de gestión de congresos y conferencias que registre automáticamente la información en el sistema de almacenamiento, a partir de los papers y artículos presentados, sin necesidad de interacción adicional o carga manual de datos."

Relevar requerimientos de usuarios y de administradores del sitio de CoNallSI, partiendo la experiencia con OCS, con el objetivo de mejorar la experiencia de usuario y la calidad de los datos cargados.

Diseñar la arquitectura del sistema de gestión de congresos de forma tal que sirva a la doble función de uniformar la gestión de los trabajos de investigación (en cuanto a envío, evaluación, organización y publicación) y que sirva de base como plataforma para la carga automática de la base de datos, sobre la que se realizarán los análisis.

Desarrollar una librería que procese automáticamente los papers enviados por medio de la aplicación a fines de extraer toda la información necesaria para el análisis cuantitativo, la cual se almacenará directamente en la base de datos correspondiente, minimizando la operatoria manual, facilitando la tarea del usuario, reduciendo la tasa de error de la carga manual y mejorando la interacción con el usuario.

Desarrollar e implementar el sistema de gestión de congresos, según los lineamientos surgidos de los requerimientos.

Para cumplimentar el **objetivo específico 5:**

"Desarrollar una herramienta de consulta y análisis que permita obtener métricas asociadas a la producción científica y que posibilite analizar los diversos patrones que pudieran surgir a partir del análisis de citas, autores y líneas de investigación".

Desarrollar una herramienta de análisis que acceda a la base de datos y permita consultar, navegar y visualizar la información de I+D, proveyendo una interfaz intuitiva y simple, pero que no restrinja la potencia de las consultas que se pueden realizar, para lo cual se diseñará un entorno de consultas no preprogramadas.

Brindar una herramienta que permita analizar la información registrada y caracterizar las principales líneas de investigación, a nivel general y discriminadas por institución, la evolución en el tiempo de los principales temas, el impacto de los distintos autores y grupos de investigación, las líneas de investigación que están mostrando interés o desinterés, etc.

Detectar y visualizar patrones (referencias entre autores, relaciones de autores con líneas de investigación, concentración de investigaciones o temáticas en las instituciones, etc.) que se extraigan de los datos, para posterior análisis de causales de la existencia de los mismos en el seno de RIISIC, de las Universidades o de los propios autores.

NOTAS:

En el proyecto se plantea la utilización de información 100% pública para el análisis cuantitativo. Todos los documentos sobre los que se va a trabajar están alojados en repositorios institucionales abiertos o han sido publicados en memorias de eventos de carácter público o indexadas por repositorios de acceso abierto.

En ningún caso se utilizará información de carácter propietario, privado o que requiera algún tipo de suscripción o pago.

Además se informó en Plenario de RIISIC la posibilidad de proponer el presente proyecto, las características y el uso de los datos que emergen de las ediciones de CoNallSI.

12. Contribuciones del Proyecto

Contribuciones al avance científico, tecnológico, transferencia al medio

Las contribuciones que los análisis a realizar, utilizando métodos cuantitativos, consideramos que permiten alcanzar, entre otros, los siguientes beneficios:

- Identificar el avance de la ciencia en un ámbito dado.
- Detectar la caducidad de algunos campos científicos.

- Analizar el cambio cronológico con respecto a algún ámbito del saber.
- Analizar la productividad de autores, instituciones y centros de investigación.
- Estudiar las redes de colaboración entre científicos y los resultados de dicha colaboración.
- Brindar una herramienta eficaz para un análisis cuantitativo tendiente a optimizar la distribución de recursos para la investigación.

Se espera que los resultados del presente trabajo posibiliten tener una imagen certera y actualizada de la producción científica de las entidades que componen la RIISIC.

Por otra parte, para los investigadores, un análisis cuantitativo puede ser utilizado como herramienta potenciadora dentro de cada uno de sus proyectos, proponiendo líneas de investigación, investigadores para colaboración y/o medios de difusión de los resultados.

Contribuciones a la formación de Recursos Humanos

Los recursos humanos involucrados en el presente proyecto puede ver su formación fortalecida de acuerdo a los siguientes criterios:

- Los docentes investigadores explorarán los usos de una metodología que les permitirá ampliar sus posibilidades, en cuanto a líneas de investigación de interés, tanto local como regional o internacional, y accederán a nombres y trabajos de posibles colaboradores en sus respectivos campos de estudio.
- El Esp. Ing. Martín Casatti realizará su tesis de maestría en el marco del proyecto, la cual permitirá finalizar los estudios de posgrado en la Maestría en Ingeniería en Sistemas de Información.
- A los estudiantes de grado que aún cursan asignaturas y no poseen experiencias en investigación, que en este proyecto son dos y uno que ya posee experiencia, podrán realizar sus primeras experiencias en investigación, a la vez que aplicarán los conocimientos adquiridos durante el cursado en el desarrollo de las herramientas informáticas que se describen en los objetivos e incluso proponer la temática en sus Prácticas Supervisadas, contando con los docentes integrantes del PID como Docentes Tutores.

13. Cronograma de Actividades

Año	Actividad	Inicio	Duración	Fin
1	Investigación bibliográfica y en Internet	01/01/2020	3 meses	31/03/2020
1	Publicaciones y presentaciones en congresos (Tarea Permanente)	01/03/2020	10 meses	31/12/2020
1	Análisis de datos en las estructuras existentes en históricos del CoNallSI	01/04/2020	2 meses	31/05/2020
1	Análisis de la estructura de documentos PDFs	01/05/2020	2 meses	30/06/2020
1	Selección de modelo de Base de Datos e instalación del motor asociado.	01/06/2020	2 meses	31/07/2020
1	Análisis de entornos de programación, framework y herramientas para desarrollo.	01/07/2020	3 meses	30/09/2020
1	Desarrollo de software de extracción de atributos de documentos	01/08/2020	3 meses	31/10/2020
1	Análisis de requerimientos para la plataforma de gestión del congreso.	01/09/2020	4 meses	31/12/2020
1	Elaboración de memorias del trabajo desarrollado.	01/12/2020	1 meses	31/12/2020
2	Modelado del sistema para la gestión del congreso.	01/01/2021	3 meses	31/03/2021
2	Preparación de material para publicaciones y presentaciones en congresos (Permanente)	01/01/2021	12 meses	31/12/2021
2	Importación de papers para carga de datos.	01/01/2021	2 meses	28/02/2021
2	Desarrollo del sistema para la gestión del congreso.	01/02/2021	8 meses	30/09/2021
2	Modelado de aplicación para el análisis cuantitativo.	01/03/2021	2 meses	30/04/2021
2	Desarrollo de aplicación de análisis cuantitativo.	01/04/2021	6 meses	30/09/2021
2	Elaboración de informe cuantitativo.	01/07/2021	4 meses	31/10/2021
2	Presentación de resultados a la RIISIC.	01/10/2021	2 meses	30/11/2021
2	Elaboración de memoria final, correcciones y ajustes.	01/10/2021	3 meses	31/12/2021

14. Conexión del grupo de Trabajo con otros grupos de investigación en los últimos cinco años

Grupo Vinc.	Apellido	Nombre	Cargo	Institución	Ciudad	Objetivos	Descripción
-	-	-	-	-	-	-	-

15. Presupuesto

Total Estimado del Proyecto: \$ 1396370,64

15.1. Recursos Humanos - Inciso 1 e Inciso 5

Primer Año

Becarios Inciso 5	Cantidad	Pesos	Origen del financiamiento
1. Becario Alumno Fac.Reg.	0	\$ 0,00	-
2. Becario Alumno UTN-SAE	2	\$ 22533,84	Facultad Regional
3. Becario Alumno UTN-SCTyP	1	\$ 18288,00	UTN- SCTyP

4. Becario BINID	1	\$ 42291,00	UTN- SCTyP	-
5. Becario Posgrado-Doctoral en el país	0	\$ 0,00	-	-
6. Becario Posgrado Doctoral en el extranjero	0	\$ 0,00	-	-
7. Becario Posgrado - Especialización	0	\$ 0,00	-	-
8. Becario Posgrado - Maestría en el país	0	\$ 0,00	-	-
9. Becario Posgrado - Maestría en el extranjero	0	\$ 0,00	-	-

Docentes Investigadores y Otros - Inciso 1	Cantidad	Pesos
1.Administrativo	0	\$ 0,00
2.CoDirector	1	\$ 127178,74
3.Director	1	\$ 134615,00
4.Investigador de apoyo	3	\$ 191100,00
5.Investigador Formado	1	\$ 127178,74
6.Investigador Tesista	0	\$ 0,00
7.Otras	0	\$ 0,00
8.Técnico de Apoyo	0	\$ 0,00

Totales	Inciso 5	Inciso 1	Total
Primer Año	\$ 83112,84	\$ 580072,48	\$ 663185,32

Segundo Año

Becarios Inciso 5	Cantidad	Pesos	Origen del financiamiento	
1. Becario Alumno Fac.Reg.	0	\$ 0,00	-	-
2. Becario Alumno UTN-SAE	2	\$ 22533,84	Facultad Regional	-
3. Becario Alumno UTN-SCTyP	1	\$ 18288,00	UTN- SCTyP	-
4. Becario BINID	1	\$ 42291,00	UTN- SCTyP	-
5. Becario Posgrado-Doctoral en el país	0	\$ 0,00	-	-
6. Becario Posgrado Doctoral en el extranjero	0	\$ 0,00	-	-
7. Becario Posgrado - Especialización	0	\$ 0,00	-	-
8. Becario Posgrado - Maestría en el país	0	\$ 0,00	-	-
9. Becario Posgrado - Maestría en el extranjero	0	\$ 0,00	-	-

Docentes Investigadores y Otros - Inciso 1	Cantidad	Pesos
1.Administrativo	0	\$ 0,00
2.CoDirector	1	\$ 127178,74
3.Director	1	\$ 134615,00
4.Investigador de apoyo	3	\$ 191100,00
5.Investigador Formado	1	\$ 127178,74
6.Investigador Tesista	0	\$ 0,00
7.Otras	0	\$ 0,00
8.Técnico de Apoyo	0	\$ 0,00

Totales	Inciso 5	Inciso 1	Total
Segundo Año	\$ 83112,84	\$ 580072,48	\$ 663185,32

TOTAL GENERAL	Inciso 5	Inciso 1	Total General
Todo el Proyecto	\$ 166225,68	\$ 1160144,96	\$ 1326370,64

15.2 Bienes de consumo - Inciso 2

Año del Proyecto	Financiación Anual	Solicitado a
1	\$ 3.000,00	UTN - SCTyP
2	\$ 2.000,00	UTN - SCTyP
Total en Bienes de Consumo		\$ 5.000,00

15.3 Servicios no personales - Inciso 3

Año	Descripción	Monto	Solicitado a
1	Viáticos y movilidad Nacional e Incripción a Congresos Nacionales	\$ 16.510,00	UTN - SCTyP
2	Viáticos y movilidad Nacional e Incripción a Congresos Nacionales	\$ 23.500,00	UTN - SCTyP
Total en Servicios no personales		\$ 40.010,00	

15.4 Equipos - Inciso 4.3 - Disponible y/o necesario

Año	Disp/Nec	Origen	Descripción	Modelo	Otras Espec.	Cantidad	Monto Unitario	Solicitado a
1	Disponible	UTNFRC-CIT 003229	Notebook Compaq Presario 21	I5	4Gb RAM - 1TB disco	1,00	\$ 0,00	UTN - SCTyP
1	Disponible	UTN INV 25118	Disco Externo Seagate	1TB	-	1,00	\$ 0,00	UTN - SCTyP
1	Disponible	UTNFRC-CIT 3167	Disco Externo WD	1TB	-	1,00	\$ 0,00	UTN - SCTyP
1	Necesario	VENEX Computación	PC DE ESCRITORIO	Intel Celeron G3900	SSD 120GB - 4GB DDR4 + MONITOR 19"	1,00	\$ 15.490,00	UTN - SCTyP
2	Necesario	VENEX Computación	Memoria RAM	DDR4 SODIMM 16GB 2400MHz		2,00	\$ 4.750,00	UTN - SCTyP
Total en Equipos						\$ 24.990,00		

15.5 Bibliografía de colección - Inciso 4.5 - Disponible y/o necesario

Año	Disp/Nec	Origen	Descripción	Modelo	Otras Espec.	Cantidad	Monto Unitario	Solicitado a
Total en Bibliografía							\$ 0,00	

15.6 Software - Disponible y/o necesario

Año	Disp/Nec	Origen	Descripción	Modelo	Otras Espec.	Cantidad	Monto Unitario	Solicitado a
-	-	-	-	-	-	-	-	-
Total en Software							\$ 0,00	

16. Co-Financiamiento

Año	RR.HH.	Bienes de Consumo	Equipamiento	Servicios no personales	Bibliografía	Software	Total
1	\$663.185,32	\$3.000,00	\$15.490,00	\$16.510,00	\$0,00	\$0,00	\$698.185,32
2	\$663.185,32	\$2.000,00	\$9.500,00	\$23.500,00	\$0,00	\$0,00	\$698.185,32
Total del Proyecto	\$1.326.370,64	\$5.000,00	\$24.990,00	\$40.010,00	\$0,00	\$0,00	\$1.396.370,64

Financiamiento de la Universidad

Universidad Tecnológica Nacional - SCyT	\$ 191.158,00
Facultad Regional	\$ 1.205.212,64

Financiamiento de Terceros

Organismos públicos nacionales (CONICET, Agencia, INTI, CONEA, etc.)	\$ 0,00
Organismos / Empresas Internacionales / Extranjeros	\$ 0,00
Entidades privadas nacionales (Empresas, Fundaciones, etc.)	\$ 0,00
Otros	\$ 0,00
Total	\$ 1.396.370,64

Avales de aprobación, Financiamiento y Otros

	Orden	Nombre de archivo	Tamaño
Descargar	1	AvalDepto.pdf	167821
Descargar	2	avalesscycdecanoMuñoz.pdf	846793
Descargar	3	ResolDirectivoNº931-19.pdf	3939539

Currículums (Currículums de los integrantes cargados en el sistema)