FASE 1 (esto pertenece al actual proyecto, para comparar con la nueva propuesta)

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 cienciométrico 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 cienciométrico 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.

Falta completar último año, último objetivo (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.

FASE 2

**SECCION 2: Denominación del PID: A DEFINIR**

**SECCION 3: Resumen técnico del PID:**

El proyecto será una continuación natural del PID SIUTNCO0007848, “Análisis cienciométrico de la producción en investigación científica y tecnológica en la Red de Ingeniería en Informática”, al cual ampliará y complementará.

Se prevé la incorporación a la base de datos de un set de datos compuesto por trabajos de estudiantes para ampliar la información de la misma y poder definir métricas específicas e indicadores de producción asociados a los estudiantes, además con esta nueva información, poder realizar análisis comparativos entre las áreas de interés demostradas por los estudiantes con respecto a los investigadores formados.

Aprovechando la disposición del texto completo de los artículos bajo análisis, se aplicarán distintas técnicas de minería de textos con respecto al contenido, a fines de obtener una categorización de los artículos que complemente las palabras clave ya existentes. Dicha categorización servirá para nomenclar los artículos de acuerdo a estándares internacionales tales como los soportados por Scopus y/o Web of Science.

Con la herramienta de consulta y análisis de métricas desarrollada en el proyecto anterior, se prevé mejorarla y ampliarla con la incorporación de una API (interfaz de programación de aplicaciones, por sus siglas en inglés) que posibilite que otras entidades y organizaciones hagan uso del repositorio de datos académicos.

Por último, se desea ampliar el sistema de gestión de congresos y conferencias implementado, para que agregue nuevas funcionalidades de gestión de un congreso.

**SECCION 5: Proyecto – Palabras clave**

cienciometría – producción científica – métricas – big data – data/text mining – key process indicators – open data

**SECCION 6: Fechas de realización**

Inicio: 01/01/2023 Fin: 31/12/2024 Duración: 24 meses.

**SECCION 10: Personal Científico Tecnológico**

* Muñoz, Roberto Miguel: Director
* (??): Co-Director
* Guzman, Analía: Investigador de Apoyo
* Casatti, Martin Gustavo: Investigador de Apoyo
* ???

**SECCION 11: Datos de la investigación**

**Estado actual de conocimiento 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].

La cienciometría busca caracterizar la investigación científica desde el punto de vista de la producción, sin establecer a priori criterios con respecto a la validez o importancia relativa de los contenidos presentes en los artefactos de investigación [2].

Si bien existen en la actualidad formatos bien establecidos y ampliamente aceptados con respecto a la estructuración de artefactos de investigación científica, en forma de papers, journals, abstracts, etc.[3] no menos cierto es que una gran parte del conocimiento sigue estando en forma de texto libre y formatos no estructurados, de los cuales la extracción de información estructurada reviste múltiples dificultades [4].

Por otra parte, las bases de datos documentales, si bien son herramientas adecuadas para la registración y catálogo de documentos, no están eficientemente diseñadas para la realización de tareas analíticas que incluyan la obtención de conocimiento valioso a través del análisis de textos.

Vemos entonces una necesaria complementariedad en los dos enfoques cienciométricos principales. Por un lado el análisis cuantitativo de la producción científica, en la forma de métricas de producción, de relación, de citas y referencias y por otro un análisis cualitativo más enfocado en el contenido de los documentos, como es la detección de líneas de investigación, relaciones no formales, terminología y ontologías emergentes, etc.

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, dando una posible respuesta a la problemática cuantitativa de la cienciometría [5].

Por otra parte, las técnicas de text mining, machine learning y reconocimiento de patrones en texto libre brindan un conjunto de técnicas valiosas para lidiar con el problema de la extracción de información cualitativa de documentos científicos [6].

Durante los años 2019 y 2020, principalmente, la pandemia de COVID-19 puso de manifiesto la necesidad de realizar trabajos de investigación en un entorno de movilidad limitada y con importantes restricciones para el desarrollo de actividades presenciales. La necesidad de contar con acceso a importantes recursos de investigación en un tiempo extremadamente corto con vistas a la elaboración de tratamientos, protocolos y eventualmente vacunas, puso de manifiesto que el acceso a información científica de calidad, de manera rápida y confiable, es una necesidad acuciante y no debe ser menospreciada en el planteo de cualquier proyecto de investigación [7][8].

La existencia de bases de datos públicas o de acceso libre demostró ser una excelente noticia para investigadores de todo el mundo que, no solo obtuvieron información importante para sus investigaciones y protocolos sino que encontraron repositorios abiertos en los cuales publicar hallazgos, tanto preliminares como definitivos, acelerando de esta manera los procesos de investigación, validación y verificación.

Muchos de estos proyectos hicieron uso de interfaces de acceso, protocolos de comunicación abiertos y formatos de intercambio de datos que ya se encontraban en uso, solo que no aplicados a este tipo de problemática.

Tecnologías como REST, JSON, OpenAPI, posibilitaron el acceso a la información y la consecuente reducción en los tiempos de investigación y en el esfuerzo humano asociado a la misma [9][10].

Es crucial que cualquier nuevo proyecto de acceso a información científica prevea la posibilidad de interconexión con otros repositorios así como que implemente mecanismos de consulta abiertos para su uso por parte de la comunidad académica en general y de los investigadores en particular.

[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] 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.

[3] Wu, J. (2011). Improving the writing of research papers: IMRAD and beyond. *Landscape Ecology*, *26*(10), 1345-1349.

[4] Boyd, J. (2015). *Automatic Metadata Extraction-The High Energy Physics Use Case* (Doctoral dissertation, Ecole Polytechnique, Lausanne).

[5] Leonelli, S. (2020). Scientific research and big data.

[6] Kononova, O., He, T., Huo, H., Trewartha, A., Olivetti, E. A., & Ceder, G. (2021). Opportunities and challenges of text mining in materials research. *Iscience*, *24*(3), 102155.

[7] Sohan, M. F. (2020). So you need datasets for your COVID-19 detection research using machine learning?. *arXiv preprint arXiv:2008.05906*.

[8] Zuo, X., Chen, Y., Ohno-Machado, L., & Xu, H. (2021). How do we share data in COVID-19 research? A systematic review of COVID-19 datasets in PubMed Central Articles. *Briefings in Bioinformatics*, *22*(2), 800-811.

[9] Viljoen, M., Dutka, Ł., Kryza, B., & Chen, Y. (2016). Towards European open science commons: the EGI open data platform and the EGI dataHub. *Procedia Computer Science*, *97*, 148-152.

[10] Ferrari, T., Scardaci, D., & Andreozzi, S. (2018). The Open Science Commons for the European Research Area. *Earth Observation Open Science and Innovation. ISSI Scientific Report Series*, *15*, 43-68.

**Grado de avance**

Respecto al Congreso Nacional de Ingeniería en Informática / Sistemas de Información (CoNaIISI), 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 seguido administrando 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.

Actualmente se han procesado todos los documentos de investigadores, correspondientes a los años 2013 a 2021, obteniendo un total de:

* 7882 Artículos (entre articulos de Conaiisi y artículos referenciados)
* 16807 Autores (entre autores de Conaiisi y autores referenciados)
* 1046 keywords detectadas
* 1059 Instituciones (organismos y dependencias incluídas)

Los Ing. Esp. Analía Guzmán y Martín Casatti han participado, durante los últimos 8 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.

Es importante mencionar que se desarrolló, como parte del proyecto precedente (PID SIUTNCO0007848, “Análisis cienciométrico de la producción en investigación científica y tecnológica en la Red de Ingeniería en Informática”) un software de gestión de congresos que busca reemplazar el antiguo sistema OCS que se utiliza en la gestión de congresos CoNaIISI, en dicho software participaron adicionalmente 6 alumnos que realizaron su tesis de grado en el ámbito de la cátedra proyecto final. Este software implementa importantes mejoras con respecto a OCS y está pensado para alimentar automáticamente la base de datos cienciométrica a medida que los autores vayan enviando trabajos en su versión final o Camera Ready. Si bien al mencionado software le resta implementar algunas funcionalidades para abarcar todas las operaciones de gestión de un congreso, la funcionalidad existente es un excelente punto de partida para avanzar con una segunda etapa donde se puede agregar más funcionalidades.

ESTO DEPENDE DE SI LOS INTEGRANTES SE MANTIENEN

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 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 graduado reciente, 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 (CoNaIISI 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 (CoNaIISI) 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 cienciométrico.
* 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 es spaCy pero también se pueden evaluar algunas alternativas, como NLTK, textaCy, polyglot o practNLPTools.
* Desarrollo de herramientas automatizadas para la generación de archivos intermedios en formato TIE utilizando el software GROBID de extracción de información de documentos científicos.
* Desarrollo de herramientas automatizadas que toman los archivos TIE intermedios los pre-procesan y realizan la inserción en la base de datos cienciométrica (OrientDB, orientada a grafos) para su procesamiento en lo que respecta a obtención de métricas y patrones.
* Planteo de métricas estandarizadas para la caracterización de la información cienciométrica.
* Análisis de librerías de presentación gráfica de grafos, para su utilización en conjunto con las rutinas analíticas.

**Objetivos de investigación**

**Objetivo general:** Describir y caracterizar la producción en investigación por parte de estudiantes, en contraposición con la producción en investigación por parte de docentes formados.

**Objetivos particulares:**

1. Incorporar un set de datos compuesto por los trabajos de estudiantes presentados en las diferentes ediciones de CoNaIISI, para ampliar la base de conocimiento y poder definir métricas específicas e indicadores de producción.
2. Realizar análisis comparativos entre las áreas de interés demostradas por los estudiantes con respecto a investigadores formados.
3. Aplicar técnicas de minería de textos con respecto al contenido de los artículos, a fines de obtener una categorización de los mismos que complemente las palabras claves ya existentes.
4. Mejorar y ampliar la herramienta de consulta y análisis de métricas desarrollada, de tal forma de dotarla de una API (interfaz de programación de aplicaciones, por sus siglas en inglés) que posibilite que otras entidades y organizaciones hagan uso del repositorio de datos académicos.
5. Ampliar el sistema de gestión de congresos y conferencias implementado, para que agregue funcionalidades de gestión de un congreso.

**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 dividirá en las siguientes etapas, estando las tareas principales enumeradas en cada una de ellas.

***Investigación básica:***

Investigar los modelos y herramientas de minería de textos existentes, analizando sus características y comparándolas con las necesidades del proyecto.

Investigar las ontologías del conocimiento científico en vistas a plantear un método de mapeo entre los términos utilizados en un documento con respecto a las categorías científicas estandarizadas a nivel mundial.

***Modelado***

Modelar los cambios necesarios en la base cienciométrica para aceptar la incorporación de trabajos de investigación de estudiantes sin confundirlos con los trabajos de investigadores formados.

Determinar qué datos específicos de estudiantes se pueden obtener de los trabajos de investigación presentados e incorporarlos a las estructuras para categorizar mejor la población bajo estudio.

Determinar las necesidades operativas y de conectividad que debe cubrir una API, para cumplir con los estándares de ciencia abierta y de Open Data.

Determinar las herramientas y tecnologías más adecuadas para la implementación de las API de consulta a la base de datos cienciométrica.

Diseñar los puntos de acceso (endpoints) y las estructuras de datos a intercambiar por dicha API.

***Desarrollo***:

Desarrollar los scripts para extracción masiva de información en formato estándar TIE a partir de los documentos PDF enviados a las ediciones del congreso CoNaIISI.

Implementar los cambios en las estructuras de la base de datos cienciométrica y verificar la consistencia de los nuevos atributos de nodos y arcos.

Desarrollar los algoritmos de extracción de datos, utilizando archivos TIE como fuente, e insertar dicha información en la base de datos cienciométrica.

Codificar las consultas estándar en formato Graph Query para su ejecución sobre la base de datos.

Desarrollar la funcionalidad de back-end que implemente la API de consulta de datos, sobre la base de información cienciométrica.

Desarrollar las rutinas de visualización gráfica de resultados, en base a las nuevas métricas y patrones definidos.

***Análisis e informes cienciométricos***

Analizar la información registrada en la base de datos.

Caracterizar las principales líneas de investigación, a nivel general y discriminadas por institución, la evolución, con respecto a los trabajos de investigación de estudiantes.

Realizar análisis comparativos de la investigación por parte de estudiantes con respecto a investigación por parte de investigadores formados.

Analizar los distintos patrones que exhiben los datos y las posibles causas de la existencia de los mismos.

Analizar los resultados obtenidos al aplicar técnicas de minería de textos sobre el texto completo de los trabajos de investigación.

Determinar la categorización de artículos y líneas de investigación a partir de los datos obtenidos por medio de técnicas de text mining.

**NOTAS**:

En el proyecto se plantea la utilización de información 100% pública para el análisis cienciométrico. 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 a la RIISIC la posibilidad de proponer el presente proyecto, las características y el uso de los datos que emergen de las ediciones de CoNaIISI.

**SECCION 12: Contribuciones del Proyecto**

**Contribuciones al avance científico, tecnológico, transferencia al medio**

Las contribuciones que los análisis a realizar, utilizando métodos cienciométricos, 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, específicamente en lo que respecta al involucramiento temprano de estudiantes en el ámbito de la investigación.
* Estudiar las temáticas preferidas por estudiantes para realizar sus primeras actividades de investigación.
* 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 cienciométrico 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.

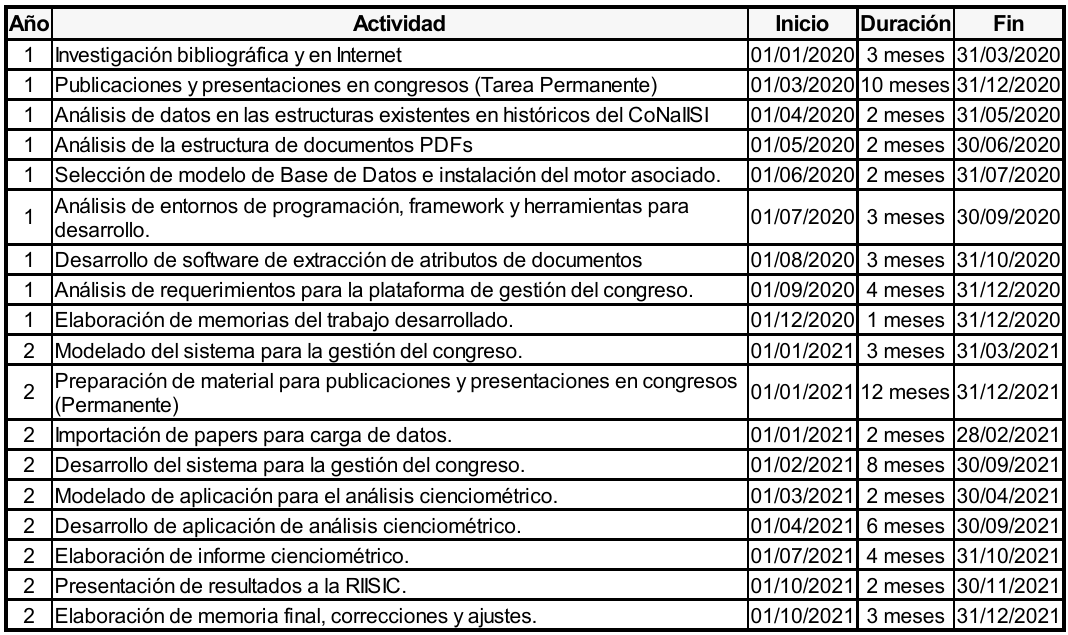
**Contribuciones a la formación de Recursos Humanos**

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.
* 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.

**SECCION 13: Cronograma de actividades**

**NO DEFINIDO. ESPEREMOS A VER SI EL RESTO ESTAMOS DE ACUERDO Y AHI PLANIFICAMOS LAS ACTIVIDADES**



FASE 2

* Ampliar objetivo 5: ampliar herramienta analítica para definir las métricas por parte del usuario, poner con que tecnología??
* Incorporar los trabajos de los estudiantes para analizar las áreas de interés, la cantidad de producción de los mismos, etc.
* Definir métricas especificas para estudiantes, tales como…
* Definir métricas generales para investigadores y alumnos, tales como…
* Aplicar técnicas de minería de texto en los artículos para ampliar y complementar la lista de keyword y permitir la categorización de los artículos según normas internacionales (scopus)

Objetivo general: Ampliar la caracterización de 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.

La caracterización relacionada a:

Comparaciones de productividad institucional, clasificaciones de investigación institucional, estableciendo estándares de productividad y permanencia, evaluando la influencia de los artículos académicos más importantes y desarrollando

Los perfiles de los principales autores e instituciones en términos del rendimiento de su investigación. OTROS que sean mas adecuados politicamente​