

2016

# SisDA

## (Sistema de Decisión Académica)

DISEÑO E IMPLEMENTACIÓN DE HERRAMIENTA INFORMÁTICA A PARTIR DEL ANÁLISIS DE PROYECTOS FINALES DE CARRERA PARA LA TOMA DE DECISIONES EN UTN FRSF



**Llorens, Román Rafael**

Directora: Ing. Esp. Eva Casco

Co-Directora: Dra. Mariel Ale

*El presente informe corresponde al proyecto final de carrera de Ing. en Sistemas de Información del alumno Llorens Román Rafael. El mismo consiste en el desarrollo y la implementación de una herramienta informática, para la toma de decisión académica en la Facultad Regional Santa Fe (FRSF) de la Universidad Tecnológica Nacional (UTN). Tiene como finalidad el análisis de Proyectos Finales de Carrera, su relación con el diseño curricular y la generación de indicadores que permitan la toma de decisión de los actores organizacionales.*

## Índice

Capítulo 1: <b>Introducción</b> .....	4
1.1 Definición del tema.....	4
1.2 Fundamentación.....	5
1.3 Alcance del Proyecto .....	7
1.4 Aportes esperados .....	8
1.5 Sobre el informe.....	8
Capítulo 2: <b>Estudio de Entidades</b> .....	10
2.1 Descripción de entidades involucradas .....	10
2.2 Análisis de relaciones .....	11
2.3 Estudio preliminar de Palabras Claves dentro de PFC .....	12
Capítulo 3: <b>Metodología</b> .....	14
3.1 Modelo de desarrollo de software .....	14
3.2 Características del modelo incremental .....	15
3.3 Desarrollo de incrementos .....	15
3.3.1 Incremento 1 .....	15
3.3.2 Incremento 2 .....	16
3.3.3 Incremento 3 .....	16
3.3.4 Incremento 4 .....	16
Capítulo 4: <b>Solución tecnológica propuesta: SisDA</b> .....	18
4.1 Descripción .....	18
4.2 Diagramas de Casos de Uso.....	19
4.2.1 Módulo 1: Plan de estudios .....	19
4.2.2 Módulo 2: Proyectos Finales .....	20
4.2.3 Módulo 3: Indicadores.....	20
4.3 Arquitectura de sistema .....	21
4.3.1 Modelo Cliente/Servidor .....	21
4.3.2 Arquitectura Lógica.....	22
4.4 Diseño de los módulos .....	23
4.4.1 Módulo 1: Plan de Estudios.....	23
4.4.2 Módulo 2: Proyectos Finales .....	24
4.4.3 Módulo 3: Indicadores.....	24
4.5 Tecnologías y herramientas de desarrollo .....	25
4.5.1 Lenguaje y Tecnologías.....	25
4.5.2 Herramientas de desarrollo .....	27
Capítulo 5: <b>Módulo 1 - Plan de Estudios</b> .....	30

5.1	Descripción.....	30
5.2	Materias .....	30
5.3	Contenidos Mínimos .....	32
5.4	Determinación de Palabras Claves .....	32
<b>Capítulo 6: Módulo 2 - Proyectos Finales .....</b>		<b>33</b>
6.1	Descripción .....	33
6.2	Búsqueda de Palabras Claves .....	33
6.3	Ponderación de Palabras Claves .....	34
6.3.1	Consideraciones previas .....	34
6.3.2	Descripción de TF-IDF.....	36
<b>Capítulo 7: Módulo 3 -Indicadores .....</b>		<b>38</b>
7.1	Descripción .....	38
7.2	Media aritmética para palabras claves agrupadas por materia (MA) .....	38
7.3	Indicadores propuestos .....	40
7.3.1	Indicadores de Proyecto.....	40
7.3.2	Indicadores de Cohorte .....	43
7.3.3	Indicadores Complementarios .....	46
7.4	Valoración de los indicadores propuestos .....	47
<b>Capítulo 8: La herramienta: SisDA .....</b>		<b>49</b>
8.1	Descripción general .....	49
8.2	Módulo 1: Plan de Estudios .....	50
8.3	Módulo 2: Proyectos Finales .....	52
8.4	Módulo 3: Indicadores .....	53
<b>Capítulo 9: Validación de la herramienta .....</b>		<b>57</b>
9.1	Medida de performance .....	57
9.1.1	Cálculo de Precisión y Recall .....	57
9.1.2	Interpretación probabilística .....	57
9.2	Evaluación de la herramienta.....	58
9.2.1	Evaluación de documentos .....	58
9.2.2	Cálculo y análisis de la medida de performance .....	59
<b>Capítulo 10: Conclusiones.....</b>		<b>60</b>
10.1	Conclusiones generales .....	60
10.2	Conclusiones del sistema .....	61
10.3	Impacto .....	61
10.4	Factibilidad .....	62
<b>Anexos.....</b>		<b>64</b>
Anexo 1: Plan de Gestión de Riesgos.....		64

1.1 Aplicación del método TBQ - Taxonomía de Riesgos.....	64
1.2 Evaluación de los riesgos .....	72
1.3 Plan de Contingencia .....	75
Anexo 2: Base de Datos .....	76
2.1 Diagrama de análisis de Base de Datos .....	76
Anexo 3: Casos de Uso.....	77
3.1 Diagrama general de C.U. ....	77
3.2 Especificaciones de C.U. ....	78
Anexo 4: Diagrama de Clases .....	83
4.1 Diagrama de clases Modulo 1: Planes de Estudio.....	84
4.2 Diagrama de clases Modulo 2: Proyectos Finales .....	85
4.3 Diagrama de clases Modulo 3: Indicadores.....	86
Anexo 5: Validación de la herramienta .....	87
5.1 Palabras claves más destacadas .....	87
5.2 Materias más relacionadas.....	88
5.3 Análisis con la herramienta .....	88
<b>Referencias .....</b>	<b>92</b>
<b>Bibliografía.....</b>	<b>94</b>

## Capítulo 1: Introducción

### 1.1 Definición del tema

*“La sociedad actual demanda de las universidades la formación de profesionales idóneos que puedan adaptarse a la naturaleza acelerada de la evolución científica y tecnológica” [1].*

Desde la Universidad se considera que, con la finalidad de que los alumnos alcancen la idoneidad profesional y para facilitar la aproximación al ámbito laboral de los egresados, es necesario analizar qué conocimientos han sido adquiridos por los estudiantes en el trayecto de su carrera de Ingeniería.

Por ello, los Proyectos Finales de Carrera (PFC) son el objetivo de estudio de este proyecto. Ellos surgen como resultado del cursado de la asignatura integradora del último nivel de la carrera de Ingeniería. Durante su desarrollo, los alumnos pueden aplicar todos los conocimientos adquiridos a lo largo de la carrera, por ser el último trabajo integrador necesario para graduarse como ingeniero.

El presente proyecto se realiza en el marco del en el marco del PID 25/O151 “Los proyectos finales de carrera como respuesta a la demanda del medio, su relación con la curricula y el perfil del ingeniero de la Universidad Tecnológica Nacional”.

En el mencionado proyecto de investigación, se generaron distintas instancias en las cuales se realizaron estudios que pretendían analizar la coherencia y cohesión entre los contenidos mínimos del Plan de Estudio de las carreras de Ingeniería de la UTN FRSF, los contenidos de los PFC y las incumbencias según el perfil profesional del ingeniero, establecidas según las Ordenanzas del Consejo Superior Universitario para cada especialidad. Dichos estudios permitieron brindar un aporte teórico-práctico a la gestión de la calidad educativa en la UTN.

Para ello, se llevó a cabo un análisis subjetivo de los PFC de una determinada cohorte, extrayendo de manera manual palabras claves dentro de los mismos, y a éstas se las relacionó de manera arbitraria a las materias según sus contenidos mínimos. Esta actividad daba sustento a la investigación realizada, pero requería de un tiempo considerable ya que el trabajo era manual. En base al procesamiento de información descripto, surge la idea de realizar el presente proyecto.

Se plantea la necesidad de contar con una herramienta que dé agilidad a la tarea mencionada, posibilitando la reducción del tiempo en el acceso a la información referida a los PFC y su relación con las materias del plan de estudios. Además, se requiere de una herramienta flexible que pueda adaptarse y ampliarse según las necesidades organizacionales que a futuro puedan presentarse.

## 1.2 Fundamentación

En el sistema educativo en general, y en el universitario en particular, la evaluación es una necesidad no explicitada ni ejercida suficientemente, pese a que las instituciones de educación formal deberían ser evaluadoras por esencia en su quehacer cotidiano. Lo cierto es que, la evaluación ha funcionado más como una actividad reglamentaria que cumple con la función de certificar, que como una función de comprensión y cualificación del devenir.

Ante tal situación, se plantea la necesidad de que cada institución educativa se autoevalúe en pos de la mejora continua y, para ello, es prioritario analizar el currículum. El motivo de su evaluación radica en que es la unidad funcional de la intervención educativa que permite comprender la acción, asignar valor a sus distintos elementos y gestionar eficazmente la mejora y el cambio.

Es así como, el currículum ya no es un instrumento neutro y científico sino una construcción social, que a través de sus elecciones explícitas o implícitas según creencias e intereses, expresa la participación de todos y los distintos actores de la comunidad educativa.

El currículum es un plan para la acción, un documento que incluye un conjunto de estrategias que persiguen la consecución de los objetivos o fines deseados. Es toda actividad organizada que se prolonga en el tiempo para conseguir unos objetivos, que cuenta con un sistema de gestión y financiación, se dirige a un grupo de individuos y despierta el interés de muchos otros.

La estructura curricular de las carreras de ingeniería de la UTN FRSF, pone en evidencia un adelanto notable en materia curricular pues cuenta con asignaturas que conforman un troco integrador. Además, presenta dos áreas bien delimitadas: la de asignaturas comunes (básicas homogeneizadas y de especialidad) y la de

asignaturas electivas (ciencias sociales, gestión, científico-técnicas). De esta manera, muestra rigidez y flexibilidad curricular al mismo tiempo. Otra característica de dicha estructura es que define objetivos y contenidos mínimos de las asignaturas, así como fija las relaciones de precedencia y correlatividad entre ellas, pretendiendo la inclusión de objetivos formativos ‘transversales’ y ‘horizontales’, los cuales propician la articulación entre los contenidos de las distintas asignaturas, y se orientan hacia el logro de habilidades de pensamiento de orden superior y una elevación de estándares.

Al fijar los contenidos mínimos que guardan relación con las incumbencias y con el perfil profesional de cada especialidad y, al facilitar la orientación a través de las materias electivas, se muestra abierto a los requerimientos de la región a la que pertenece la Unidad Académica, promoviendo así la actualización permanente.

En la fundamentación, intenta superar la tendencia respecto a las racionalidades curriculares, dejando de lado la racionalidad academicista y tecnológica en miras a la racionalidad crítica. Presenta la postura de ampliar el papel de la Universidad, desde el tradicional rol como “fábricas de conocimiento” hacia la tendencia a romper límites institucionales entre empresas y universidades que postula la economía de la innovación, según la cual, la acumulación de conocimiento (proceso complejo de entrelazamiento entre ideas y habilidades) es la base del crecimiento económico.

Este diseño abarca no sólo contenidos programáticos, sino aspectos metodológicos del desarrollo profesional y el trabajo ingenieril. Muestra de ello son los requerimientos que debe cumplir el estudiante como la “Práctica Supervisada” y la asignatura “Proyecto Final de Carrera”, que en su programa requiere la aplicación de los contenidos desarrollados en el transcurso de la carrera por medio de la elaboración y ejecución de un proyecto ingenieril que dé respuesta a la solución de un problema concreto. [2]

El aprendizaje mediante proyectos puede considerarse como uno de los métodos educativos reconocidos como “buenas prácticas” por su uso habitual en las carreras de ingeniería, teniendo en cuenta que la ejecución de los mismos, es una función relevante de la profesión del ingeniero. Es una alternativa pedagógica que le permite al estudiante, aplicar saberes y habilidades adquiridos durante el cursado en una

situación real. El proceso de realización de los Proyectos Finales de Carrera beneficia el desarrollo de competencias con incursión en el mercado laboral.

“...se debe pensar la formación desde el saber-hacer, en forma efectiva y en los diferentes ámbitos del quehacer profesional y social.” [3]

El Proyecto Tuning América Latina define al ingeniero como un profesional con un amplio manejo de las ciencias básicas y las ciencias de la ingeniería, que le permiten desarrollar soluciones de ingeniería y hace mención dentro de las competencias identificadas como muy importantes “la abstracción espacial y las representaciones gráficas” [4], aspectos que son considerados en el diseño curricular de la UTN.

Ahora bien, retomando la importancia de la autoevaluación de las instituciones educativas en vistas a la mejora continua, se considera esencial la evaluación del diseño curricular y si el mismo es efectivamente implementado en el quehacer cotidiano.

Por ello, la herramienta que se propone en el presente proyecto permitirá cotejar si las bondades curriculares analizadas teóricamente son evidenciadas mediante el análisis de los contenidos de los proyectos finales de carrera y su relación con las distintas áreas mencionadas en el diseño.

### 1.3 Alcance del Proyecto

Considerando que en el desarrollo de la asignatura Proyecto Final, que corresponde al tronco integrador del currículo de la Universidad Tecnológica Nacional (UTN), el alumno hace uso de los contenidos desarrollados por las demás asignaturas de los distintos niveles e integra los mismos tanto de forma horizontal como vertical, se desea analizar el contenido de los PFC y la relación que evidencian con el Diseño Curricular al que pertenecen.

La herramienta permitirá agilizar el análisis de los contenidos mínimos utilizados para el desarrollo de los PFC y de esa manera el usuario podrá relacionarlos con las incumbencias propias del perfil, lo cual a su vez posibilitará el fortalecimiento de cada perfil ingenieril, mediante la toma de decisiones académicas.

Los PFC a considerar corresponderán a las carreras de grado dictadas en la UTN Facultad Regional Santa Fe (FRSF), que se encuentren digitalizados en la Biblioteca

de dicha Facultad. Al respecto, la normativa vigente para el desarrollo de los PFC expresa: “El PFC será enviado a la Biblioteca de la Facultad, el que podrá ser facilitado a los usuarios como bibliografía.” [5]

Además, se analizarán los Diseños Curriculares vigentes, teniendo en cuenta los contenidos mínimos y quedando fuera del análisis las materias electivas que pueden variar en el transcurso del tiempo.

#### 1.4 Aportes esperados

La herramienta a realizar pretende contribuir al análisis de PFC en el marco del proyecto de investigación 25/O151. Además, procura aportar una visión más concreta acerca del resultado final en la integración de conocimientos de las distintas áreas y asignaturas. Finalmente, pretende ser un aporte a la mejora continua tanto de los diseños curriculares como de la implementación de los mismos.

Por otro lado, las expectativas del autor radican en poder aplicar, como futuro profesional, los conocimientos adquiridos en el transcurso de la carrera, tanto en lo técnico como en lo humano.

#### 1.5 Sobre el informe

Los siguientes capítulos se centran en el desarrollo del presente proyecto, haciendo hincapié en la estrategia de resolución de la herramienta propuesta como solución.

El informe se complementa con las secciones detalladas a continuación:

**Estudio de Entidades:** contiene las terminologías de las principales entidades del proyecto, la relación entre las mismas y un estudio preliminar para detectar requerimientos.

**Metodología:** se describe la metodología de desarrollo de software empleada.

**Solución Tecnológica propuesta: SisDA:** se desarrolla una breve descripción de la solución propuesta, la arquitectura y diseño de la misma, como así también la descripción de las herramientas y tecnologías aplicadas en el proyecto.

**Plan de Estudios:** define la forma en que se lleva a cabo una de las funcionalidades del sistema, describiendo en profundidad las entidades involucradas y resolución propuesta.

**Proyectos Finales:** describe la resolución de la funcionalidad del sistema, planteando las entidades y métodos utilizados para su desarrollo.

**Generación de Indicadores:** plantea y describe los indicadores propuestos para la herramienta.

**La herramienta: SisDA:** plantea una visión de la herramienta desarrollada, describiendo más precisamente su interfaz.

**Validación de la herramienta:** se desarrolla el análisis abordado para validar la performance de la herramienta.

**Conclusiones:** se exponen conclusiones generales sobre el presente proyecto, sobre el sistema desarrollado, y sobre el impacto y factibilidad de la implementación del mismo.

**Anexos:** se incluyen todos los anexos considerados pertinentes que se desarrollaron en el proyecto.

**Referencias:** contiene todas las referencias bibliográficas citadas a lo largo del informe.

**Bibliografía:** contiene las fuentes bibliográficas consultadas para la realización del proyecto.

## Capítulo 2: Estudio de Entidades

### 2.1 Descripción de entidades involucradas

- **Plan de estudios:** es un plan para la acción dentro de una propuesta formativa, un documento que incluye un conjunto de estrategias que persiguen la consecución de los objetivos o fines deseados. Éste es entendido como una propuesta política-pedagógica, surgida de un contexto histórico-social específico que sintetiza elementos culturales, conocimientos, valores, costumbres, creencias, hábitos; pensada e impulsada por diversos grupos y sectores sociales, de intereses diversos y contradictorios, negociados o impuestos socialmente [6].
- **Materia:** unidad disciplinar perteneciente a un campo de conocimiento específico. La estructura curricular de las carreras de ingeniería cuenta con asignaturas que conforman un tronco integrador. Además, presenta dos áreas bien delimitadas, una de asignaturas comunes (básicas homogeneizadas y de especialidad) y otra de asignaturas electivas (ciencias sociales, gestión, científico-técnicas). Los planes de estudios de ingeniería definen las asignaturas básicas y de especialidad, estableciendo sus respectivos objetivos y contenidos mínimos, pero no así las asignaturas electivas que son establecidas por cada especialidad según la demanda del medio socio-productivo que rodea la Unidad Académica.
- **Contenidos Mínimos:** son conceptos elementales que establecen la base para la enseñanza de una materia, son considerados como la formación indispensable y mínima de la carrera. Los contenidos mínimos establecidos guardan relación con las incumbencias y con el perfil profesional propuesto, flexibilizando y facilitando a través de las materias electivas, la orientación, mostrándose abierto a los requerimientos de la región, posibilitando así la actualización permanente.
- **Proyecto Final de Carrera (PFC):** es el trabajo integrador de la materia Proyecto Final perteneciente al último año de la carrera de ingeniería. El análisis de los proyectos finales es relevante, ya que durante su desarrollo los alumnos pueden

aplicar todos los conocimientos adquiridos en el transcurso de la carrera, por ser el último trabajo integrador necesario para graduarse como ingeniero. Cada PFC integra los conocimientos adquiridos en las distintas asignaturas tanto de manera horizontal (de un mismo nivel) como vertical (de los distintos niveles).

## 2.2 Análisis de relaciones

Los planes de estudios de las carreras de Ingeniería están establecidos por medio de ordenanzas del Consejo Superior Universitario de la UTN. Los mismos cuentan en su estructura con un 'Programa Sintético', el cual establece las materias que integran la curricula de la carrera de ingeniería a la que pertenece, como así también por cada materia, establece su carga horaria, área, objetivos, contenidos mínimos, entre otros. Por otra parte, cabe destacar que la asignatura Proyecto Final de Carrera dentro de su programa contempla la elaboración de un proyecto ingenieril aplicando los contenidos teóricos-prácticos aprendidos a lo largo de la carrera. Además, el PFC es una alternativa pedagógica que le permite al estudiante, aplicar saberes y habilidades adquiridos durante el cursado en una situación real. La realización de los mismos beneficia al desarrollo de competencias con incursión en el mercado laboral.

Tomando por un lado el PFC, y por el otro, el plan de estudio de la carrera a la que pertenece dicho proyecto, se plantea la necesidad de determinar el modo de relacionar los mismos y considerar alguna/s medida/s que permita/n desarrollar un análisis de dichas relaciones.

Como se mencionó anteriormente, el plan de estudios en su programa sintético contiene, entre otras cosas, los contenidos mínimos a desarrollar por las materias de los distintos niveles; y por otro lado, la materia Proyecto Final pretende integrar estos contenidos por medio de un proyecto ingenieril denominado PFC. Por ende, se debería determinar la aplicación de los contenidos establecidos por los planes de estudios dentro del desarrollo del PFC. Para hacerlo, se propone obtener términos que representen al contenido mínimo de una materia, y el mismo pueda ser encontrado dentro del desarrollo de un proyecto. Dichos términos se denominarán **Palabras Claves**.

El concepto 'Palabra Clave' hace referencia a una colección no ordenada de palabras significativas, denominadas también no vacías (es decir, todas las palabras que no son artículos, conjunciones, pronombres, preposiciones, numerales y ciertos verbos y adverbios), extraídas de forma automática por el ordenador, a partir del título, del resumen y, cada vez a menudo, del texto completo de los documentos registrados dentro de un sistema documental dado. [7]

En el presente proyecto, las palabras claves establecen una relación entre los PFC y los contenidos mínimos del Plan de Estudios. Las mismas pueden ser evaluadas en cuanto a su grado de verdad dentro de cada proyecto.

### 2.3 Estudio preliminar de Palabras Claves dentro de PFC

Se realizó un estudio preliminar a partir de 18 PFC pertenecientes a las cohortes 2011 y 2012 de la carrera de Ing. en Sistemas de Información y se detectaron palabras claves dentro de su contenido. Las mismas fueron ordenadas en tablas, asignándoles de manera subjetiva materias del plan de estudios correspondiente. Cabe aclarar que una palabra clave pudo asociarse a más de una materia y, a su vez, una materia pudo vincularse a más de una palabra clave.

En la **Figura 1** se puede observar un listado de las palabras claves utilizadas con mayor frecuencia dentro de los PFC analizados. Dichas palabras pueden relacionarse con materias tales como: Análisis de Sistemas, Diseño de Sistemas, Gestión de Datos, Ingeniería del Software, entre otras.

casos de uso
requerimientos
bases de datos
metodología / métodos
modelo de datos
diagrama de clases
gestión
modelado / modelo
patrones de diseño
arquitectura
documentación
implementación
programación
pruebas
interfaz de usuario
rup
cliente-servidor

**Figura 1.** Listado de P.C. encontradas en PFC.

El estudio de PFC descrito permite realizar diferentes análisis de las relaciones establecidas entre los contenidos mínimos y los proyectos. A modo de ejemplo, podría citarse que el área más involucrada dentro del desarrollo de los PFC es el área integradora de la carrera. La misma, además de establecer sus propios contenidos, integra nivel a nivel los demás contenidos de las asignaturas. Esto posibilita determinar que los estudiantes han podido integrar los contenidos mínimos de las áreas del plan de estudios, respondiendo a la propuesta curricular.

Fue sobre la base de este estudio preliminar que surgió la idea de desarrollar una herramienta informática que facilite la tarea realizada en el marco del proyecto de investigación, que hasta el momento se llevaba a cabo en forma manual.

## Capítulo 3: Metodología

### 3.1 Modelo de desarrollo de software

Para crear la herramienta informática se utilizó un ciclo de desarrollo basado en un modelo incremental, debido a que los requerimientos iniciales fueron razonablemente definidos y los mismos pudieron ser divididos fácilmente en módulos y asignar cada uno a un incremento.

Dicho modelo consiste en un desarrollo inicial de la arquitectura del sistema, seguido de sucesivos incrementos funcionales. Cada incremento tiene su propio ciclo de desarrollo y se construye sobre el anterior, sin cambiar su funcionalidad ni sus interfaces. Una vez finalizado un incremento, no se realizan cambios sobre el mismo, sino sólo corrección de errores.

Considerando la posibilidad de que la herramienta se amplíe a futuro, se desea hacer énfasis en el diseño modular de la misma, buscando favorecer desde la concepción del proyecto la reusabilidad de los módulos que forman parte de la solución informática.

Además, para la trazabilidad y continuidad del proyecto a futuro se utilizaron algunos de los artefactos propuestos por el modelo RUP para generar la documentación necesaria. Dichos artefactos fueron:

- Diagrama de Casos de Uso;
- Diagrama de Clases;
- Modelo Entidad-Relación;
- Especificación de Casos de Usos;
- Pruebas finales de aceptación.

### 3.2 Características del modelo incremental

A continuación, se describen algunos beneficios de este modelo que llevaron a optar por el mismo de acuerdo a las especificaciones del proyecto:

- Es aplicable cuando los requisitos son medianamente bien conocidos, pero no son completamente estáticos y definidos, lo que no es posible con un modelo en cascada donde tal aspecto resulta indispensable.
- Tiene la ventaja de entregar un conjunto de funcionalidades en menor tiempo.
- El usuario, a diferencia de otros modelos clásicos como el modelo en cascada, se involucra más a lo largo del proyecto, con lo cual se reduce la tasa de errores que pueden obtenerse con modelos clásicos.
- Los modelos incrementales admiten incorporar más fácilmente nuevas funcionalidades, respecto de los modelos clásicos. En este sentido, es adecuado para el presente proyecto ya que se busca lograr un sistema extensible.

### 3.3 Desarrollo de incrementos

#### 3.3.1 Incremento 1

##### **Objetivo**

Elicitación de requerimientos, definición de las arquitecturas, investigación y análisis de herramientas de desarrollo, diseño e implementación de Base de Datos.

##### **Actividades a desarrollar**

- a. Diseño de casos de uso (CU) para documentar los requisitos funcionales del sistema, y algunos no funcionales específicos.
- b. Definición y diseño de las arquitecturas.
- c. Diseño e implementación de la Base de Datos.
- d. Diseño de componentes para la delimitación de los módulos a implementar/developar.

- e. Investigación y selección de herramientas de desarrollo.

### 3.3.2 Incremento 2

#### **Objetivo**

Desarrollo del Módulo 1, carga del Diseño Curricular y análisis de Contenidos Mínimos.

#### **Actividades a desarrollar**

- a. Refinamiento de CU para el diseño del Módulo 1.
- b. Diseño de interfaz de usuario.
- c. Diagrama de clases para el desarrollo del módulo a implementar.
- d. Codificación del primer módulo.
- e. Generación e implementación de casos de pruebas para el módulo.

### 3.3.3 Incremento 3

#### **Objetivo**

Desarrollo del Módulo 2, carga de PFC y búsqueda de palabras claves.

#### **Actividades a desarrollar**

- a. Refinamiento de CU para el diseño del Módulo 2.
- b. Diseño de interfaz de usuario.
- c. Diagrama de clases para el desarrollo del módulo a implementar.
- d. Codificación del segundo módulo.
- e. Generación e implementación de casos de pruebas para el módulo.

### 3.3.4 Incremento 4

#### **Objetivo**

Desarrollo del Módulo 3, procesamiento de la información obtenida por los módulos anteriores y generación de indicadores.

**Actividades a desarrollar**

- a. Refinamiento de CU para el diseño del Módulo 3.
- b. Diseño de interfaz de usuario.
- c. Unificar y completar el diagrama de clases en su totalidad.
- d. Codificación del módulo final e integración final del software.
- e. Generación de casos de pruebas unitarios y de integración.

## Capítulo 4: Solución tecnológica propuesta: SisDA

### 4.1 Descripción

Como solución informática, el presente proyecto propone la generación de una herramienta que como funcionalidades principales permita:

- ✓ Obtener palabras claves desde los contenidos mínimos de un plan de estudios;
- ✓ Usar o manipular dichas palabras;
- ✓ Cargar PFC y buscar, dentro de los mismos, palabras claves pertenecientes al plan de estudio correspondiente;
- ✓ Generar indicadores que faciliten el desarrollo de investigaciones relacionadas con el proyecto en el que se enmarcan, y a la vez, orienten la toma de decisiones académicas relacionadas con las currículas y el desarrollo de sus contenidos.

Para dicha solución, se plantea el desarrollo de una herramienta de escritorio, considerando que la misma va a ser utilizada, en primera medida, en el marco del proyecto de investigación y por un número reducido de usuarios.

Se espera además, que la herramienta propuesta pueda dar respuesta a futuros requerimientos y funcionalidades para ser utilizada en un marco académico. El nombre elegido para la misma es **SisDA, Sistema de Decisión Académica**.

A continuación, se exponen los módulos que la conforman, los cuales permiten delimitar sus funcionalidades con mayor especificidad:

- **Módulo 1:** Plan de estudio, para cargar y manipular los planes de estudios, más específicamente sus programas sintéticos (materias, áreas, contenidos mínimos, etc.) e identificar las palabras claves relacionadas.
- **Módulo 2:** Proyectos finales, para cargar datos de los PFC y realizar búsquedas de palabras claves que permitan establecer relaciones con el plan de estudios.
- **Módulo 3:** Indicadores, destinado a la generación de índices e indicadores en base a los datos que se pueden extraer en los primeros dos módulos.

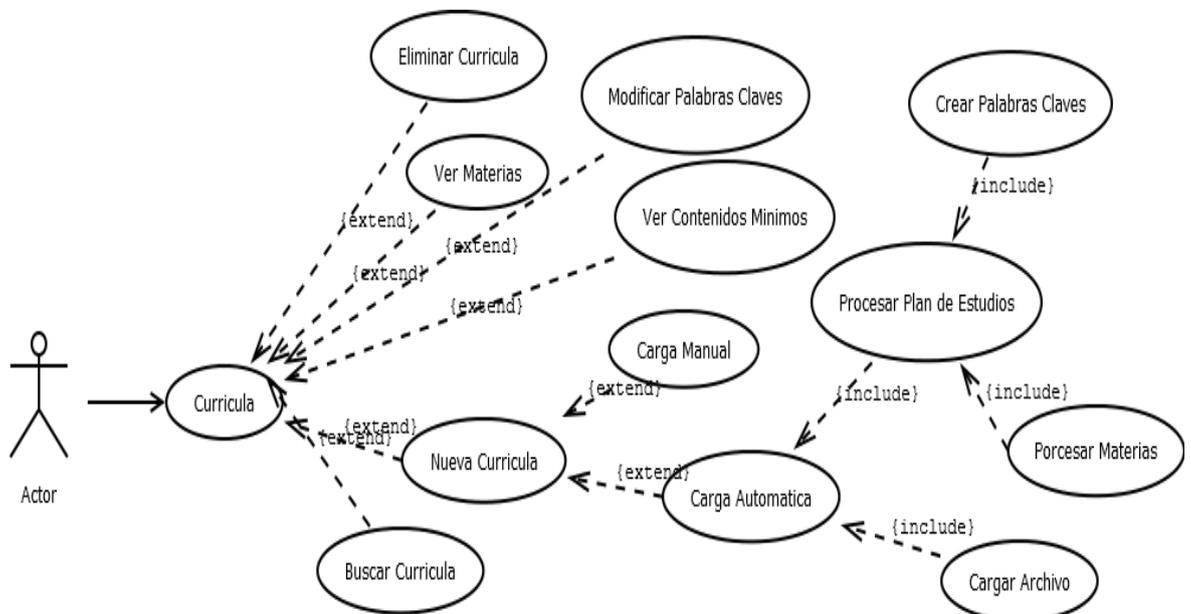
En los capítulos siguientes se detallan con mayor profundidad cada uno de los módulos de la herramienta propuesta como solución.

## 4.2 Diagramas de Casos de Uso

Los diagramas de casos de uso se consideran una buena documentación para especificar la funcionalidad y el comportamiento de un sistema mediante su interacción con los usuarios y/u otros sistemas. Debido a que la utilización de la herramienta se realiza en principio en el marco de un proyecto de investigación, es que se consideró un único actor que es el usuario, cuestión que se deberá rever en el caso de ampliar su implementación hacia otros sectores.

Para entender mejor las funcionalidades se ha dividido el diagrama de casos de uso en tres partes, correspondiente a cada uno de sus módulos. En el Anexo 3 se puede consultar el diagrama completo y las especificaciones de los casos de uso.

### 4.2.1 Módulo 1: Plan de estudios



**Figura 2.** Diagrama de CU Módulo 1.

4.2.2 Módulo 2: Proyectos Finales

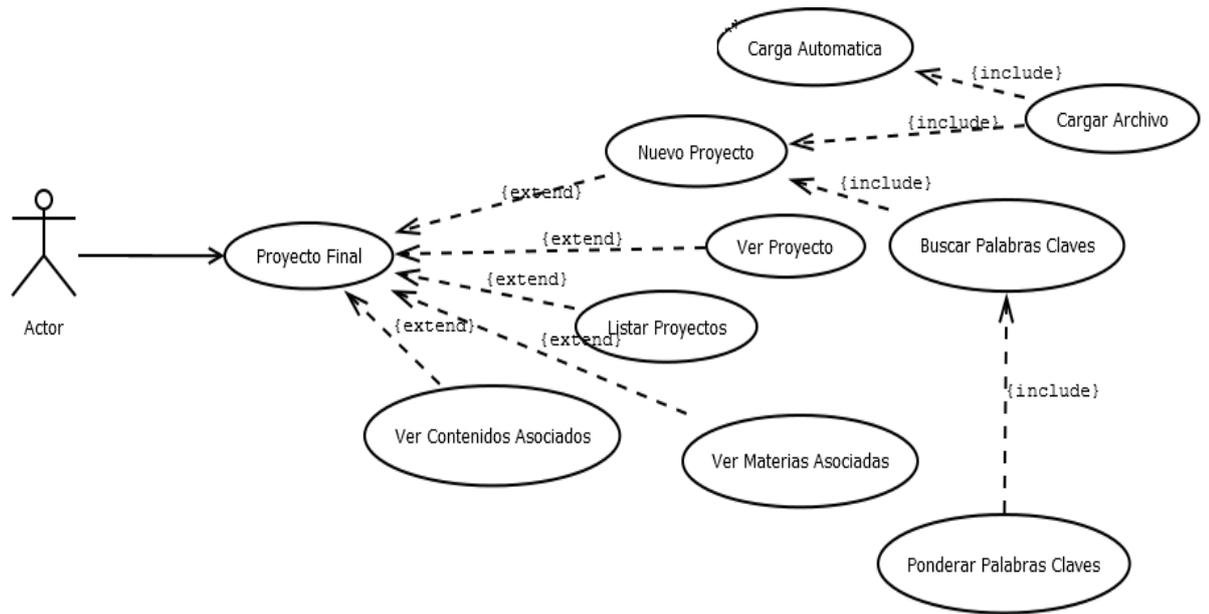


Figura 3. Diagrama de CU Módulo 2.

4.2.3 Módulo 3: Indicadores

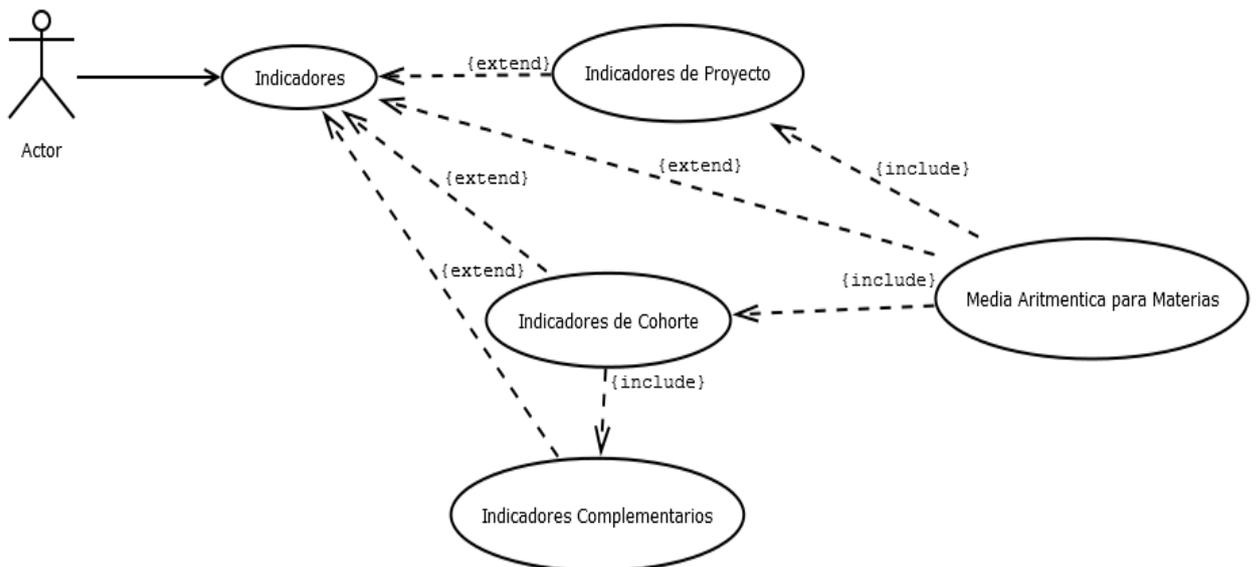


Figura 4. Diagrama de CU Módulo 3.

### 4.3 Arquitectura de sistema

La arquitectura del sistema se basa en un modelo cliente/servidor desde un alto nivel e internamente presenta una arquitectura lógica de tres capas. Esto no sólo satisface las necesidades requeridas para el proyecto, sino que además posibilita que los cambios en el software, se puedan realizar de manera más simple. A continuación, se detallan los modelos mencionados previamente:

#### 4.3.1 Modelo Cliente/Servidor

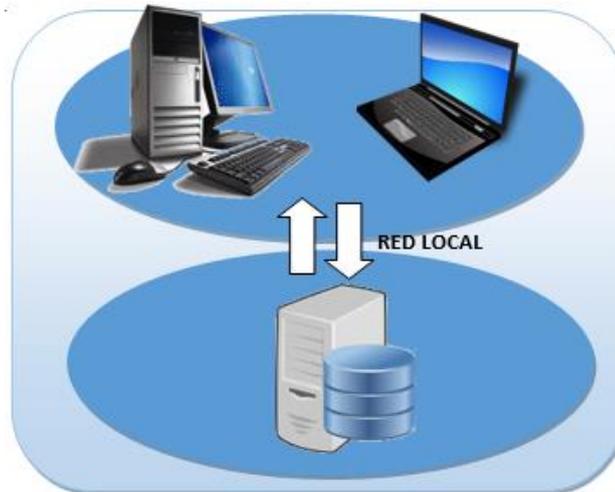
La arquitectura cliente-servidor es un modelo de aplicación distribuida en el que las tareas se reparten entre los proveedores de recursos o servicios, llamados *servidores*, y los demandantes, llamados *clientes*. Si bien la arquitectura planteada posee mayores ventajas cuando el sistema es distribuido, la misma puede aplicarse a sistemas que se ejecutan sobre una sola computadora, como en el caso del presente proyecto.

Para este sistema, la arquitectura consta de dos niveles, por un lado un servidor de base de datos y por el otro, lo que refiere a la interfaz y procesamiento de la información. Este tipo de arquitectura cliente-servidor es una arquitectura tradicional donde el lado cliente es conocido como cliente pesado o cliente grueso, ya que la mayor carga de cómputo está desplazada hacia la computadora que ejecuta dicho programa.

La misma se adapta a las necesidades planteadas por el proyecto y además cuenta con la ventaja de que la interfaz en un cliente pesado no está limitada por las características de un cliente universal. Por tanto, pueden diseñarse interfaces ricas y más fáciles de usar que fue una de las consideraciones para la herramienta.

Por otra parte, se adoptó dicha arquitectura por la escalabilidad que la misma ofrece, debido a que se puede aumentar la capacidad de clientes y servidores por separado. Cualquier elemento puede ser aumentado o mejorado en todo momento, o se pueden añadir nuevos nodos a la red. Para el presente proyecto, esto es importante debido a que se pueden ampliar las estaciones de trabajo (clientes) entre los distintos actores que puedan requerir la herramienta dentro del entorno académico,

así como también compartir la base de datos de la misma alojada en servidores comunes.



**Figura 5.** Arquitectura Cliente/Servidor – Modelo 2 niveles.

#### 4.3.2 Arquitectura Lógica

Desde una vista lógica del sistema se planteó una arquitectura de 3 capas, ya que la ventaja principal de este estilo es que se puede realizar en varios niveles y, en caso de que sobrevenga algún cambio, solo se ataca al nivel requerido sin tener que revisar entre código mezclado. Además, permite distribuir el trabajo por niveles, posibilitando así la abstracción y generando un mejor ordenamiento durante el desarrollo.

También es importante destacar que por medio de esta arquitectura se hace énfasis en una alta cohesión y desacoplamiento, de esta forma se tiene menor dependencia y se especifican los propósitos de cada objeto en el sistema.

Por otra parte, la separación entre la lógica de aplicación y la interfaz de usuario añade una enorme flexibilidad al diseño de la misma. Pueden construirse y desplegarse múltiples interfaces de usuario sin cambiar en absoluto la lógica de aplicación siempre que está presente una interfaz claramente definida a la capa de interfaz. Esta ventaja también se aplica entre la capa de procesos y la de acceso a la base de datos.

Por otra parte, esta separación entre capas se complementa con la arquitectura cliente-servidor, posibilitando que esta última tenga la posibilidad a futuro de desplegarse en más niveles de una manera más sencilla si es necesario.

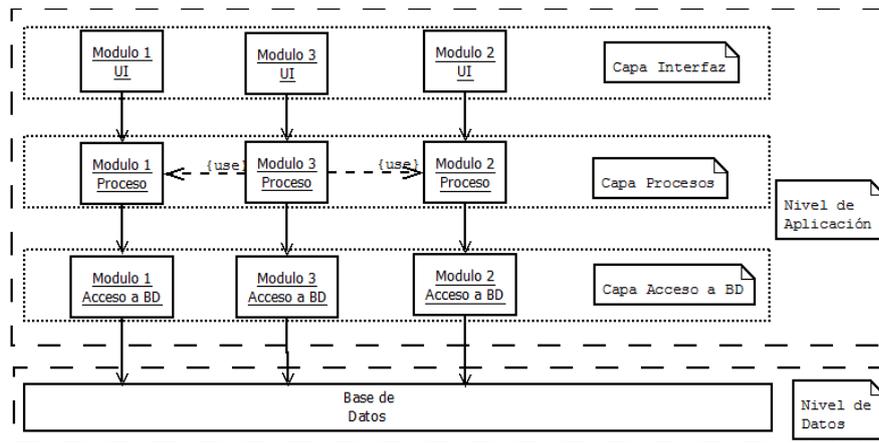


Figura 6. Vista de módulos del sistema con estilo de capas.

#### 4.4 Diseño de los módulos

Para proporcionar una visión del desarrollo de la herramienta se plantea el diseño de los módulos a través de diagramas de clases para entender mejor el dominio de la solución. En el mismo se plantean las principales entidades y sus relaciones, como así también sus atributos, lo que permite detallar el diseño de la misma.

A continuación, se muestran los diagramas divididos por los distintos módulos de la herramienta a fin de tener una mejor perspectiva de la misma. Para obtener mayor información, consultar los anexos 2 y 4 en los que se amplía el modelo de base de datos y los diagramas de clases implementados.

##### 4.4.1 Módulo 1: Plan de Estudios

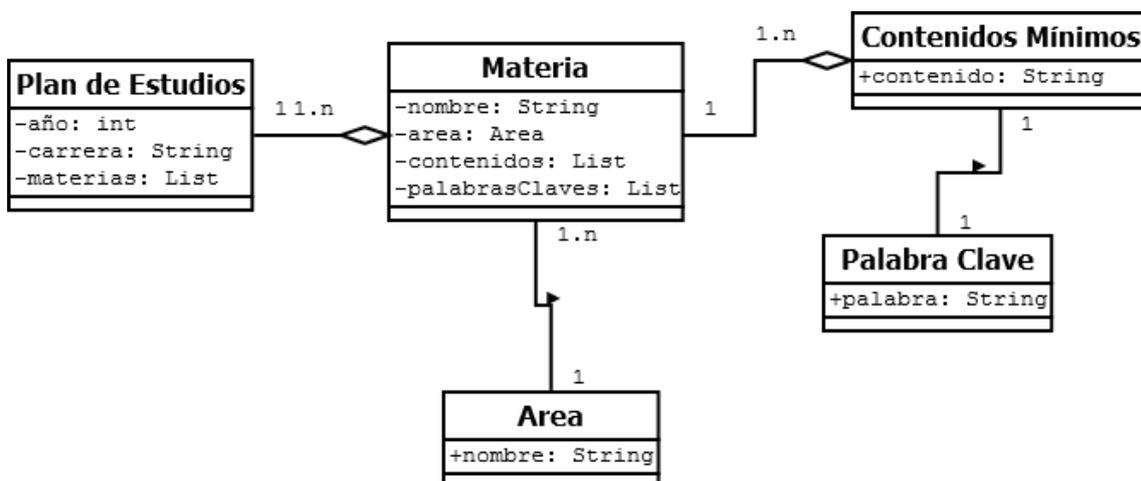


Figura 7. Diagrama de clases Módulo 1.

4.4.2 Módulo 2: Proyectos Finales

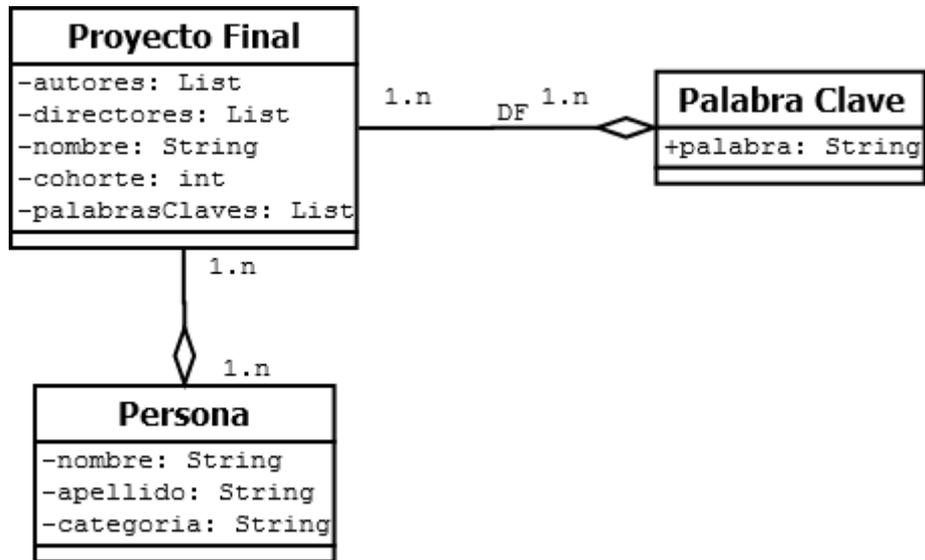


Figura 8. Diagrama de clases Módulo 2.

4.4.3 Módulo 3: Indicadores

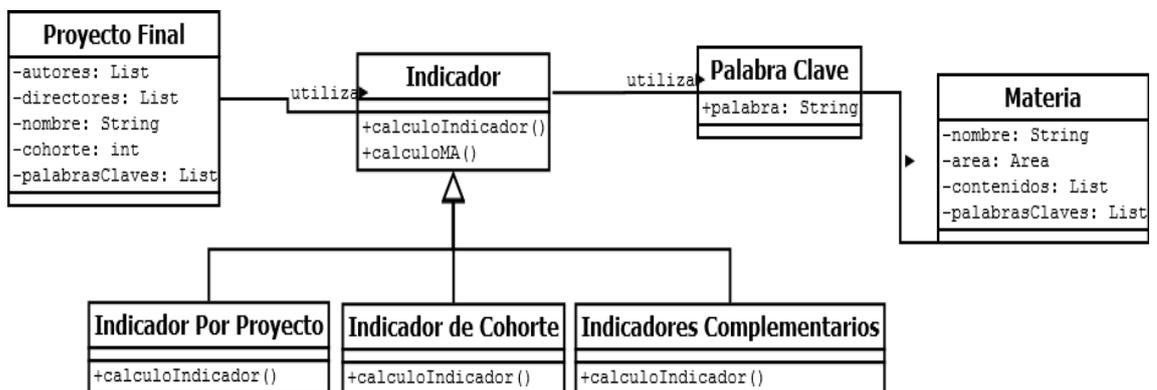


Figura 9. Diagrama de clases Módulo 3.

## 4.5 Tecnologías y herramientas de desarrollo

### 4.5.1 Lenguaje y Tecnologías

En este apartado se presenta una breve descripción del lenguaje de programación y de las tecnologías utilizadas para desarrollar el presente proyecto, detallando las principales características y ventajas de cada una.

- *Lenguaje de Programación: Java*

Java es un lenguaje de programación de propósito general, concurrente, orientado a objetos, que fue diseñado específicamente para tener tan pocas dependencias de implementación como fuera posible. [8]



Está diseñado para permitir el desarrollo de aplicaciones portátiles de elevado rendimiento para el más amplio rango de plataformas informáticas posible. Al poner a disposición de todo el mundo aplicaciones en entornos heterogéneos, las empresas pueden proporcionar más servicios y mejorar la productividad, las comunicaciones y colaboración del usuario final y reducir drásticamente el costo de desarrollo, tanto para aplicaciones de usuario como de empresa.

Este lenguaje se considera el más propicio para el desarrollo del presente proyecto por ser el que mejor servicio presta para elaborar aplicaciones de escritorio, teniendo un entorno amigable con el usuario, una sencilla implementación y ser además del tipo open-source.

- *MySQL*

MySQL es un sistema de gestión de bases de datos relacional desarrollado bajo licencia dual GPL/Licencia comercial por Oracle Corporation y está considerada como la base datos open source más popular del mundo y, junto a Oracle y Microsoft SQL Server, son las más utilizadas sobre todo para entornos de desarrollo web. [9]



Se considera que MySQL es propicia para el presente proyecto ya que posee una buena performance en relación con el lenguaje de desarrollo y además cuenta con licencia GPL, requerida para desarrollar la herramienta propuesta.

- *JavaFX*

JavaFX es una familia de productos y tecnologías de Sun Microsystems, adquirida por Oracle Corporation, para la creación de Aplicaciones de Internet Ricas (Rich Internet Applications, RIA), esto es, aplicaciones web que tienen las características y capacidades de aplicaciones de escritorio, incluyendo aplicaciones multimedia interactivas. Las tecnologías incluidas bajo la denominación JavaFX son JavaFX Script y JavaFX Mobile, aunque hay más productos JavaFX planeados. [10]



Las aplicaciones JavaFX pueden ser ejecutadas en una amplia variedad de dispositivos. JavaFX es compilado a código Java, por lo que las aplicaciones JavaFX pueden ser ejecutadas en computadores con la máquina virtual de Java instalada (JRE), o celulares corriendo Java ME.

Debido a que el proyecto requiere de una interfaz amigable para el usuario y el desarrollo de gráficos que permitan el análisis de resultados, se considera que JavaFX es adecuado para el mismo, ya que ofrece una amplia gama de posibilidades en torno al desarrollo de interfaz, además de ser muy sencilla su utilización.

- *SVN*

Apache Subversion es una tecnología de control de versiones open source basada en un repositorio cuyo funcionamiento se asemeja enormemente al de un sistema de ficheros.



Utiliza el concepto de revisión para guardar los cambios producidos en el repositorio. Entre dos revisiones sólo guarda el conjunto de modificaciones (delta), optimizando así al máximo el uso de espacio en disco. SVN permite al usuario crear, copiar y borrar carpetas con la misma flexibilidad con la que lo haría si estuviese en su disco duro local. Dada su flexibilidad, es conveniente su utilización en relación a la aplicación de buenas prácticas para llevar a cabo una correcta gestión de las versiones del software generado. [11]

Además, es importante destacar que mediante subversion se puede administrar no solo el código fuente sino toda la documentación producida durante el desarrollo del proyecto, lo que hace aún más significativa su utilización.

#### 4.5.2 Herramientas de desarrollo

En relación a las herramientas de desarrollo se decidió optar por tecnologías del tipo open-source y considerar la experiencia en la utilización de las mismas con la que se cuenta.

Entre dichas herramientas se encuentran:

- *Eclipse IDE*

Eclipse es una plataforma de software compuesto por un conjunto de herramientas de programación de código abierto multiplataforma para desarrollar lo que el proyecto llama "Aplicaciones de Cliente Enriquecido", opuesto a las aplicaciones "Cliente-liviano" basadas en navegadores. [12]



Esta plataforma, típicamente ha sido usada para desarrollar entornos de desarrollo integrados (del inglés IDE), pero además Eclipse cuenta con una comunidad de usuarios, extendiendo constantemente las áreas de aplicación cubiertas y brindando respuesta a consultas de desarrolladores. Estas características lo hacen propicio para el desarrollo del presente proyecto ya que además de brindar las herramientas necesarias, cuenta con el soporte adecuado para su utilización.

La versión utilizada para el desarrollo del proyecto es la 4.4, conocida como Eclipse Luna.

- *Java EE*

Java Platform, Enterprise Edition o Java EE es una plataforma de programación—parte de la Plataforma Java—para desarrollar y ejecutar software de aplicaciones en el lenguaje de programación Java. Permite utilizar arquitecturas de N capas distribuidas y se apoya ampliamente en componentes de software modulares ejecutándose sobre un servidor de aplicaciones. [13]



Específicamente para SisDA, se eligió la distribución Java EE versión 8.

- *MySQLWorkbench*

MySQLWorkbench es una herramienta visual de diseño de bases de datos que integra desarrollo de software, administración de bases de datos, diseño de bases de datos, creación y mantenimiento para el sistema de base de datos MySQL. [14]



Esta herramienta se eligió por permitir la administración de bases de datos de MySQL.

- *JavaFX Scene Builder*

JavaFX Scene Builder es un editor para crear archivos FXML, con el mismo se puede construir la interfaz gráfica de una aplicación de escritorio Java de forma más sencilla. JavaFX Scene Builder genera archivos descriptores FXML que podemos cargar en la aplicación evitando la tediosa y no sencilla tarea de construir la interfaz gráfica mediante código. En el caso de SisDA se desarrolló utilizando la versión 2.0 de la presente herramienta.



- *Apache POI*

Apache POI, es un proyecto de Apache Software Foundation, y anteriormente un subproyecto del proyecto Jakarta. Ofrece bibliotecas de java para leer y escribir archivos con formato de Microsoft Office, tales como Word, PowerPoint y Excel. [15]



La elección de esta herramienta se justifica porque permite la lectura de documentos en formato Word, que en el caso del proyecto es el utilizado para los proyectos finales y los planes de estudio.

- *TortoiseSVN*

TortoiseSVN es un cliente gratuito de código abierto para el sistema de control de versiones *Apache Subversion*. Esto significa que TortoiseSVN administra archivos y directorios a lo largo del tiempo. Los archivos se almacenan en un *repositorio* central. El repositorio es prácticamente lo mismo que un



servidor de archivos ordinario, con la excepción de que recuerda todos los cambios que se hayan hecho a sus archivos y directorios. Esto le permite al usuario recuperar versiones antiguas de sus archivos y examinar la historia de cómo y cuándo cambiaron sus datos, y quién hizo el cambio.

Se desarrolla bajo licencia GPL. Lo que significa que es totalmente gratuito para uso, incluso en un entorno comercial, sin ninguna restricción. El código fuente también está disponible gratuitamente, lo que permite desarrollar o extender si así se desea. [16]

Una característica interesante, es que no es una integración de un IDE específico como Visual Studio, Eclipse u otros, sino que se puede utilizar con cualquier desarrollo de herramientas, y con cualquier tipo de archivo.

- *Visual SVN Server*

Visual SVN Server es un servidor de subversion que se instala y administra fácilmente en Windows, el mismo se puede utilizar como repositorio local para el desarrollo del proyecto. Visual SVN Server se distribuye como un solo



paquete de instalación con las últimas versiones de todos los componentes necesarios. El proceso de instalación es extremadamente simple y le permite configurar un servidor Subversion completo y listo para usar en sólo unos pocos clics. [17]

La versión utilizada es VisualSVN Server Standard Edition ya que la misma es una distribución gratuita.

## Capítulo 5: Módulo 1 - Plan de Estudios

### 5.1 Descripción

El desarrollo de este módulo comprende la carga de datos de los diseños curriculares de cada especialidad. Se cargarán las diferentes asignaturas considerando en cada una los siguientes atributos: Nombre; Área a la que pertenece; y Contenidos Mínimos.

Posteriormente, a partir de los contenidos mínimos se identificarán las palabras claves.

Una restricción encontrada para la realización de este módulo fue la lectura de los documentos que contienen los planes de estudios, establecidos como se dijo previamente por ordenanzas del Consejo Superior. Para solventarla se plantearon dos soluciones: por un lado, la carga automática del documento en formato Word, con todo el procesamiento que se describirá en los puntos siguientes; y por otro lado, la carga manual de los planes de estudios que no estuvieran disponibles en formato Word.

Si bien la carga de este primer módulo -ya sea en forma automática o manual- fue diseñada de manera tal que resulte una tarea sencilla, sería conveniente que la misma sea realizada, o a lo sumo supervisada, por personas que tengan conocimiento sobre la carrera y sus contenidos, a fin de contar con, los criterios necesarios para decidir si una determinada palabra clave es representativa de los contenidos de una materia o no. En el caso de Ing. en Sistemas por ejemplo, podrían ser docentes de la carrera, ingenieros en sistemas, etc.

### 5.2 Materias

Dentro de los planes de estudios, existe un apartado que detalla el programa sintético de cada materia, junto a sus contenidos mínimos. Por tanto, una vez cargado el archivo del plan de estudios, se busca de manera automática dentro de éste el apartado del programa sintético y se realiza una iteración por cada materia que el mismo tiene. Para cada iteración se extrae: el nombre de la materia, el área a

la cual pertenece y los contenidos mínimos, dejando de lado otra información que no se consideró pertinente para el análisis.

En la **Figura 10**, se presenta un extracto del plan de estudios de la carrera Ing. en Sistemas de Información, donde se puede observar una materia de especialidad junto a toda la información de la misma.

**Asignatura: ANÁLISIS DE SISTEMAS (INT.)**

**Nº Orden: 12**

**Departamento: Ingeniería en Sistemas de Información**

**Horas/semana: 6**

**Bloque: Tecnologías Aplicada**

**Horas/Año: 192**

**Área: Sistemas de Información**

|

**Objetivos:**

- Elaborar modelos conceptuales de un sistema de información.
- Conocer las distintas etapas del proceso de desarrollo de sistemas de información.
- Modelar las características intrínsecas de los sistemas de información.
- Conocer y aplicar las metodologías, modelos, técnicas y lenguajes de la etapa de análisis.
- Seleccionar adecuadamente los modelos que mejor se adapten para dar soluciones a los problemas de información.
- Conocer y aplicar los elementos que componen la ingeniería de requerimientos.
- Documentar el proceso de análisis de sistemas de información.

**Contenidos Mínimos:**

- Procesos de desarrollo de SI. Metodologías y herramientas de análisis de sistemas.
- Técnicas de relevamiento.
- Identificación, especificación y validación de requerimientos.
- Patrones de Análisis.
- Estudio de Prefactibilidad.
- Modelado de Negocios.
- Documentación Pertinente.

**Figura 10.** Extracto de materia del Plan de Estudios

### 5.3 Contenidos Mínimos

Tal como se mencionó anteriormente, los contenidos mínimos de cada materia serán extraídos del programa sintético de la misma en forma automática, completa y textual por la herramienta. Dichos contenidos mínimos son propiedades de cada materia, por lo cual serán relacionados y almacenados como tal.

Se pueden observar contenidos mínimos de la materia Análisis de Sistemas en la **Figura 10** a manera de ejemplo.

### 5.4 Determinación de Palabras Claves

Las palabras claves serán extraídas de los contenidos mínimos y estarán relacionadas directamente con la materia a la cual pertenecen y no específicamente con el contenido de cada una, ya que el estudio a realizar se focaliza en las materias relacionadas con los PFC.

Las Palabras Claves son constructos elaborados de manera automática por medio del parseo y la identificación de términos específicos, tales como sustantivos o frases sustantivas, dentro de cada contenido mínimo. El parseo del texto se realiza por medio de las conjunciones y signos ortográficos como el punto y la coma.

Si bien las palabras claves se producen en forma automática, se considera que el usuario debe tener la posibilidad de editarlas, modificarlas o eliminarlas en caso que lo crea conveniente; ya que algunas pueden ser comunes a varias materias. Por ejemplo, si la palabra clave fuera “funciones”, la misma resultaría ambigua para el análisis planteado y el usuario podría intervenir.

Tomando la materia Análisis de Sistemas (**ver Figura 10**) algunas palabras claves en base a los contenidos mínimos serían: validación de requerimientos, procesos de desarrollo de SI, técnicas de relevamiento, etc.

## Capítulo 6: Módulo 2 - Proyectos Finales

### 6.1 Descripción

El procesamiento de los PFC consta de la búsqueda de palabras claves dentro de los mismos que se relacionen con los contenidos mínimos de las materias. Una vez halladas las palabras claves, se las pondera para establecer una determinada importancia dentro del proyecto. A partir de esto, es posible generar indicadores relacionados con el desarrollo de los contenidos de las materias dentro de los PFC.

En base a lo descrito previamente, se realizó la búsqueda directa en los PFC de las palabras claves extraídas de los contenidos mínimos, por lo que la relación existente entre el contenido y la palabra clave dentro del PFC es directa. Y una vez obtenidas las palabras claves, se implementó un método de ponderación de las mismas.

### 6.2 Búsqueda de Palabras Claves

Las palabras claves ya obtenidas a partir de los contenidos mínimos, serán buscadas dentro de los textos completos de los proyectos finales. A fin de determinar la relevancia de una palabra dentro del texto, la misma deberá ser ponderada.

Dado que los proyectos no poseen una única estructura definida y las temáticas, áreas en las que se desarrollan, directores, etc. varían en cada uno, las palabras claves se buscarán dentro del contenido de cada PFC, quedando exceptuados del análisis, la Bibliografía y Anexos.

En este módulo, será posible determinar tanto la aparición de las palabras claves dentro del documento, como así también la frecuencia de aparición. Ambos datos serán utilizados para la ponderación de cada palabra clave.

## 6.3 Ponderación de Palabras Claves

### 6.3.1 Consideraciones previas

La ponderación de términos es el proceso que tiene como finalidad conocer la importancia de los mismos para representar un documento y permitir su posterior clasificación. Esto implica que se debe determinar la capacidad de las palabras claves para representar el contenido de un proyecto final.

Dentro de las técnicas a utilizar para ponderar palabras, la más básica es la frecuencia con que aparece dentro del texto, pero existen además, otras métricas estadísticas que complementan a la frecuencia y resulta interesante considerarlas. Dichas técnicas van desde las más estándares como TF-IDF, hasta otras implementadas con algoritmos genéticos. A continuación se detallan algunas de ellas:

- **TF-IDF** (del inglés *Term frequency – Inverse document frequency*), frecuencia de término – frecuencia inversa de documento (es decir, la frecuencia de ocurrencia del término en la colección de documentos). Es una medida numérica que expresa cuán relevante es una palabra para un documento en una colección. [18]
- **Latent Dirichlet Allocation (LDA)** es un modelo generativo que permite que conjuntos de observaciones puedan ser explicados por grupos no observados, los cuales justifican por qué algunas partes de los datos son similares. Por ejemplo, si las observaciones son palabras en documentos, LDA presupone que cada documento es una mezcla de un pequeño número de categorías y la aparición de cada palabra en un documento se debe a una de las categorías a las que el documento pertenece. LDA es un ejemplo de modelo de categorías y fue presentado como un modelo en grafo para descubrir categorías por David Blei, Andrew Ng y Michael Jordan en 2002. [19].

- **Latent semantic analysis (LSA)** es una técnica que permite analizar las relaciones entre un conjunto de documentos y los términos que contienen mediante la producción de un conjunto de conceptos relacionados con los documentos y términos [20].

De acuerdo al análisis requerido para este proyecto, se consideró conveniente aplicar la técnica estándar de TF-IDF, ya que basándose en un conjunto de documentos denominado corpus, permite trabajar sobre un documento en particular y establecer las relaciones que se pretendan. En cambio, las demás técnicas trabajan sobre el conjunto de documentos para realizar una categorización de los mismos.

Por otra parte, es importante mencionar que TF-IDF es una técnica ampliamente aceptada y sencilla de implementar, lo que le da flexibilidad para adaptarse a distintos análisis. Diversos autores recomiendan la utilización de TF-IDF debido a los óptimos resultados observados. Tal es el caso de Cobo Ortega y otros, que utilizan el cálculo TF-IDF para establecer un peso de los términos de la colección, al igual que se hará en este proyecto, y a partir de los resultados logra establecer una representación vectorial de las palabras incluidas en los textos. Además, trabaja con un listado de términos previamente establecidos de cada área disciplinar, que en este caso serían las palabras claves extraídas de los contenidos mínimos de cada materia, para luego obtener un peso respecto de la presencia semántica de cada concepto y reducir así significativamente el número de representaciones en función de rasgos insignificantes, por estar presentes en un excesivo número de documentos de la colección.

Por otro lado, Pérez Iglesias y otros, aplican esta metodología en la elaboración de un ranking de resultados en procesos de recuperación de documentos, donde se trabaja en ponderaciones de términos tanto en la consulta realizada al sistema como también en los contenidos de los documentos que conforman la colección. Esto podría aplicarse al presente trabajo, llegando a elaborar un ranking con los valores ponderados para cada área de la curricula.

### 6.3.2 Descripción de TF-IDF

Para la implementación de TF-IDF es necesario determinar la "Frecuencia de aparición del término" (TF) que se realiza por medio de la búsqueda de la palabra clave dentro del documento. Como muestra la **Fórmula 1**,  $TF(n)$  es la frecuencia de aparición de un término  $n$ , calculado como la suma de las ocurrencias de dicho término en un documento  $D$ .

$$TF_{(n)} = \sum_D (n)$$

*Fórmula 1. Cálculo de TF de un término*

Por otro lado, el valor de dicha frecuencia está ponderado por el valor de la "Frecuencia inversa del documento para un término" (IDF) cuyo cálculo se realiza en base a la aparición de dicha palabra en otros documentos y el total de documentos de la colección, según como se expresa en la **Fórmula 2**.

$$IDF_{(n)} = \log_{10} \frac{N}{DF_{(n)}} + 1$$

*Fórmula 2. Cálculo del IDF de un término*

Siendo  $N$  el número de documentos de la colección,  $DF_{(n)}$  es el número de documentos en los que el término  $n$  aparece dentro de dicha colección, y  $1$  es un factor correctivo.

Dicha métrica se aplica a una colección constituida por todos los proyectos de una determinada carrera analizados con la herramienta. La ponderación IDF está basada en la cantidad de documentos analizados, por lo cual, será más precisa cuanto más documentos sean evaluados.

De esta manera, el valor TF-IDF aumenta proporcionalmente al número de veces que una palabra aparece en el documento, pero es compensada por la frecuencia de la palabra en la colección de documentos, lo que permite tratar el hecho de que algunas palabras son generalmente más comunes que otras.

Por último, la expresión del ponderador TF-IDF resulta de la multiplicación de ambos términos calculados, tal como se muestra en la **Fórmula 3**.

$$TF IDF = TF * IDF$$

*Fórmula 3. Expresión de TF-IDF*

Al finalizar este cálculo, se tendrá para cada palabra clave relacionada con un PFC, la frecuencia de aparición de la misma, junto a su valor de ponderación.

## Capítulo 7: Módulo 3 -Indicadores

### 7.1 Descripción

Una vez elaborados los primeros dos módulos, se llevó a adelante el desarrollo del módulo final, que representa un incremento sustancial de los anteriores. Este identifica las relaciones que existen entre los contenidos mínimos de las materias y las palabras claves más utilizadas en los PFC.

Para realizar dicho análisis se abordaron las bases teóricas de la estadística descriptiva, a fin de obtener las características generales de un conjunto de datos y proceder así a la definición de indicadores.

Al momento de definir los indicadores para este proyecto, se adoptaron los siguientes criterios: en qué medida los PFC están relacionados con una determinada área, qué materias son las más abordadas dentro de los PFC, qué materias dentro de un área son las más evidenciadas, etc.

De esta manera, a partir de los indicadores definidos, será posible evidenciar dentro de los PFC el desarrollo del curriculum propuesto por el plan de estudios, permitiendo analizar el perfil profesional adquirido. Tal conocimiento favorecerá, a su vez, el análisis y la toma de decisiones académicas en torno al curriculum.

### 7.2 Media aritmética para palabras claves agrupadas por materia (MA)

A fin de obtener un valor medible y comparable en el uso de palabras claves relacionadas a una materia en un PFC, se pretende reducir o condensar en pocas cifras el conjunto de observaciones relativas a la aparición de las palabras claves en dichos proyectos. Este proceso de reducción debe continuarse hasta su grado máximo, es decir, hasta substituir todos los valores observados por uno solo. Con ello, se logra una visión más clara de la importancia de la materia en relación al uso de sus contenidos mínimos y, a su vez, se pueden realizar comparaciones con mayor facilidad.

La noción de reducción a un valor único lleva implícita la idea de variación, ya que no tiene sentido promediar cantidades invariables. Este número promedio, que

sustituye al conjunto de observaciones, ha de cumplir con la condición de ser “representativo” del conjunto, para lo cual ha de reflejar la tendencia de observaciones. Dicho promedio se refiere a una medida de tendencia central, que indica cuál es el centro de distribución de datos.

En el caso de los datos obtenidos a través de la herramienta, se trabaja con la frecuencia de las palabras claves y la ponderación de las mismas. Para este tipo de datos, se utiliza como medida de tendencia central la media aritmética para datos agrupados en serie de frecuencia.

La expresión de la media aritmética es la expuesta por la **Fórmula 4**. Donde  $PCa(m)$  es el conjunto de palabras claves pertenecientes a una determinada materia del plan de estudios.

$$MA_{(m,PFC)} = \frac{\sum TF - IDF_{PCa(m)}}{\sum TF_{PCa(m)}}$$

**Fórmula 4.** Cálculo de la media aritmética para palabras claves agrupadas por materia.

Cuando se calcula la media aritmética multiplicando a cada valor por un coeficiente numérico que expresa la importancia de dicho valor (IDF), se dice que la media está “ponderada”.

Cabe destacar que si bien la media aritmética no es la única medida de tendencia central existente, se considera que posee ventajas sobre otras como ser la mediana o el modo, por lo que se optó por su utilización para este proyecto. Entre dichas ventajas se encuentran las siguientes:

- La media aritmética es el punto de equilibrio de la distribución.
- Utiliza toda la información recogida.
- Al estar definida por una expresión algebraica, puede someterse a cálculos matemáticos necesarios para deducir cuestiones importantes.

Esta última ventaja, se considera preponderante para el presente trabajo, ya que la media aritmética es básica para cálculos posteriores.

### 7.3 Indicadores propuestos

*“Cuando puedes medir aquello de lo que estás hablando y expresarlo en números, puede decirse que sabes algo acerca de ello; pero, cuando no puedes medirlo, cuando no puedes expresarlo en números, tu conocimiento es muy deficiente y poco satisfactorio...” [21]*

Para poder evidenciar dentro de los PFC el desarrollo del curriculum propuesto por el plan de estudios, y a su vez analizar el perfil profesional adquirido, se deben identificar aquellos indicadores que sean representativos para dicho análisis. Asimismo, es importante considerar el “paquete” de indicadores, que amplía el análisis y permite conocer u obtener mayores conclusiones sobre lo que se está midiendo.

A fin de establecer dicho “paquete” de indicadores se identifican tres grupos o conjuntos en función de la definición y el propósito que persigue cada indicador. Tales conjuntos se establecen a partir del análisis de un proyecto en particular, un conjunto de proyectos pertenecientes a una misma cohorte o la comparación de áreas comunes a distintas carreras.

#### 7.3.1 Indicadores de Proyecto

<b>Materia por Proyecto</b>	
<b>Definición</b>	A partir de la definición de la media aritmética de palabras claves para medir el uso de contenidos de una materia dentro de un PFC, se define el presente indicador como la medida de utilización de contenidos de una materia específica en relación al total de contenidos utilizados en el PFC.
<b>Propósito</b>	Poder comparar el uso de contenidos mínimos de cada materia dentro de un PFC en relación al uso total de los contenidos del proyecto.
<b>Método de Cálculo</b>	Se calcula como la relación del valor MA para una materia con respecto al total de todos los valores medidos.
<b>Fórmula</b>	$IM_{(m_i,PFC)} = \frac{MA_{(m_i,PFC)}}{\sum MA_{(m,PFC)}}$
	<b>Donde:</b>

$IM_{(m_i,PFC)}$  es el valor de participación de la materia  $m_i$  en un PFC.

$MA_{(m_i,PFC)}$  es la media aritmética de la materia  $m_i$  en el PFC.

$\sum MA_{(m,PFC)}$  es la suma de las medias aritmética de todas las materias en relación al PFC.

#### Interpretación

El valor medido se encuentra comprendido entre cero y uno. Un valor muy cercano al límite inferior representa una escasa participación de la materia dentro del PFC. Sin embargo, si el valor se aproxima a 1 (uno), se puede deducir que la materia ha tenido una alta incidencia en el desarrollo del PFC.

#### Área por Proyecto

#### Definición

Tomando como base la media aritmética para palabras claves agrupadas por materia, se calcula el uso de contenidos mínimos agrupados por área. Las áreas consideradas son las establecidas por los planes de estudios, las cuales agrupan a las materias de acuerdo a áreas de conocimientos.

#### Propósito

Este indicador permite analizar el uso de contenidos mínimos de las distintas áreas, y realizar una comparación entre las mismas.

#### Método de Cálculo

Se calcula como la suma de todas las **MA** de las materias de un área en relación al total de dicha medida para todas las asignaturas del plan curricular en relación a un PFC.

$$IA_{(a,PFC)} = \frac{\sum MA_{(m_i,PFC)}}{\sum MA_{(m,PFC)}}$$

#### Fórmula

Donde:

$IA_{(a,PFC)}$  es el indicador de participación del área  $a$  en un PFC.

	<i>MA<sub>(m<sub>i</sub>,PFC)</sub> es la media aritmética de la materia m<sub>i</sub> perteneciente al área a en el PFC.</i>
<b>Interpretación</b>	Quando el valor es muy cercano a 1 (uno) representa una incidencia considerable del área dentro del proyecto, lo que hace dilucidar una tendencia de dicho proyecto hacia esa área específica de la carrera. No así, si el valor es cercano a cero.
	<b><i>Materia por Área en Proyecto</i></b>
<b>Definición</b>	Con este indicador se pretende considerar la medida de participación de una materia en un PFC en función a las demás materias de su misma área de conocimiento.
<b>Propósito</b>	Poder dilucidar cómo participa una materia en el desarrollo de un proyecto en relación a las demás materias de la misma área. Tal consideración permite analizar los distintos pesos de las materias dentro del área.
<b>Método de Cálculo</b>	El indicador está determinado por la razón entre el valor <b>MA</b> de la materia y el total perteneciente a la misma área de conocimiento.
	$IMxA_{(m_i,a,PFC)} = \frac{MA_{(m_i,PFC)}}{\sum MA_{(m,PFC)}}$
	<b>Donde:</b>
<b>Fórmula</b>	<i>IMxA<sub>(m<sub>i</sub>,a,PFC)</sub> es el indicador de participación de la materia i dentro del área a en un PFC.</i> <i><math>\sum MA_{(m,PFC)}</math> es la suma de las medias aritméticas para las materias pertenecientes al área a en relación al PFC.</i>
<b>Interpretación</b>	El valor representado por el presente indicador determina el peso de la materia dentro del área de conocimiento a la cual pertenece, entonces, un valor alto representa una mayor relevancia de la materia dentro del conjunto.

**Tabla 1.** Indicadores de proyecto.

## 7.3.2 Indicadores de Cohorte

**Porcentaje de Utilización de Materia por Cohorte**

<b>Definición</b>	El presente indicador establece la relación entre los contenidos utilizados de una materia y los contenidos utilizados en los proyectos pertenecientes a una cohorte.
<b>Propósito</b>	Poder analizarla injerencia de los contenidos abordados por una materia en el desarrollo de los PFC de una determinada cohorte.
<b>Método de Cálculo</b>	Se calcula como la suma de todos los valores MA para una materia dentro de una determinada cohorte, sobre el total de todos los valores medidos para la cohorte multiplicados por 100.

**Fórmula**

$$PMC_{(m_i,c)} = \frac{\sum MA_{(m_i,PFC)}}{\sum MA_{(m,PFC)}} * 100$$

Siendo, los **PFC** de la cohorte **c**.

**Donde:**

$PCM_{(m_i,c)}$  es el porcentaje de participación de la materia  $m_i$  dentro de una cohorte  $c$ .

<b>Interpretación</b>	El valor medido se encuentra comprendido entre cero y cien. Un valor muy cercano al límite inferior representa una escasa participación de la materia dentro de los PFC de la cohorte. Sin embargo, si el valor se aproxima al cien por ciento, se puede deducir que la materia ha tenido una alta incidencia en el desarrollo del PFC.
-----------------------	---

**Porcentaje de área por cohorte**

<b>Definición</b>	Usando como base la definición del indicador anterior, se pretende conocer qué contenidos son abordados por los PFC de una cohorte en relación a las áreas del plan curricular.
-------------------	---

**Propósito** Este indicador permite analizar el uso de contenidos mínimos de las distintas áreas y realizar una comparación de las mismas según una cohorte. A partir de dicha medida, será posible analizar el perfil profesional plasmado en los PFC en relación al año de su elaboración.

**Método de Cálculo** Se expresa como la suma total de las medidas adoptadas por las materias de una determinada área para una cohorte específica, en relación con el total de las medidas para esa misma cohorte.

**Fórmula**

$$PAC_{(a,c)} = \frac{\sum(MA_{(m_i,PFC)})}{\sum MA_{(m,PFC)}} * 100$$

**Donde:**

*PAC<sub>(a,c)</sub> es el porcentaje de participación del área a dentro de una cohorte c.*

*m<sub>i</sub> pertenece al área a y los PFC a la cohorte c*

**Interpretación** Esta medida permite conocer el grado de participación de cada área en los proyectos finales de carrera para las distintas cohortes analizadas. Un alto porcentaje indica una tendencia de la cohorte, en aplicar sus conocimientos adquiridos sobre un área de conocimiento específica. Una distribución uniforme entre las distintas áreas representa un uso integral de los contenidos abordados, lo cual implica un perfil profesional más amplio y no específico.

***Número de Proyectos que se encuentran sobre la media de un área***

**Definición** A partir del indicador propuesto para proyectos por área IA, se pretende conocer cuántos proyectos de una cohorte específica se encuentran sobre la media de ese cálculo.

**Propósito** Conocer si el número de proyectos que superan la media en una determinada área es significativo o por el contrario, no

representan una muestra sustancial. Esto permite validar el indicador anteriormente mencionado (**PAC**).

**Método de Cálculo**

Se calcula sumando la cantidad de proyectos que se encuentran sobre la media del indicador **IA** para una cohorte determinada.

**Fórmula**

$$NPA_{(a,c)} = 0;$$

$$\text{Si } (IA_{(a,PFC)} \geq \overline{IA}_{(a,PFC)} \wedge PFC \in c)$$

$$NPA_{(a,c)} = NPA_{(a,c)} + 1;$$

**Donde:**

$NPA_{(a,c)}$  es el número de proyectos que superan la media del indicador **IA** para el área **a** en la cohorte **c**.

$IA_{(a,PFC)}$  es el indicador **IA** para el área **a** y el **PFC** perteneciente a la cohorte **c**.

$\overline{IA}_{(a,PFC)}$  es la media del indicador **IA** para el área **a** de los **PFC** pertenecientes a la cohorte **c**.

**Interpretación**

Si la cantidad es inferior a la media del total de proyectos de la cohorte, el valor hallado para el indicador **PAC** no es representativo, ya que la medida para dicho indicador está desplazada. No así, si el valor es próximo a la media, pudiendo considerarse el indicador dentro del conjunto analizado.

**Tabla 2.** Indicadores de Cohorte.

## 7.3.3 Indicadores Complementarios

**Área Homogénea por Cohorte****Definición**

El plan de estudios de las carreras de ingeniería posee un ciclo básico común formado por las distintas disciplinas de las ciencias básicas. El presente indicador establece el uso de contenidos abordados en los PFC de un área homogénea para una carrera específica, en relación con las restantes carreras de ingeniería.

**Propósito**

Conocer en qué medida cada carrera utiliza dichos contenidos en el desarrollo de sus PFC en relación con las demás.

**Método de Cálculo**

El cálculo se desarrolla por medio de la relación entre el uso de contenidos de un área expresado a través de la media aritmética de las materias pertenecientes al área y el total usado en el conjunto de carreras.

$$AHC_{(a, Ca, c)} = \frac{\sum MA_{(m_i, PFC_i)}}{\sum MA_{(m_i, PFC)}}$$

**Dónde:****Fórmula**

$AHC_{(a, Ca, c)}$  es el indicador de participación del área **a** dentro de una cohorte **c** y una carrera **Ca** en relación al uso de la misma área **a** en las demás carreras para los proyectos de una cohorte **c**.

$m_i$  pertenece al área **a** y los  $PFC_i$  pertenecen a la carrera **Ca** y a la cohorte **c**

**Interpretación**

Si el valor obtenido es cercano a 1 (uno), indica que, en esa carrera, los contenidos abordados por el área común analizada son representativos en comparación a las demás carreras.

**Tasa de Utilización de Área por cohorte**

<b>Definición</b>	La tasa de utilización de área por cohorte, es la variación del uso de los contenidos del área en relación al año anterior.
<b>Propósito</b>	Conocerla variación en el uso de los contenidos pertenecientes a un área de una cohorte a la otra.
<b>Método de Cálculo</b>	La expresión es la diferencia entre el indicador <b>PAC</b> de una cohorte específica y el de la cohorte anterior, por sobre la suma de ambos, multiplicado por 100.
<b>Fórmula</b>	$TUA_{(a,c_i)} = \frac{PAC_{(a,c_i)} - PAC_{(a,c_{i-1})}}{PAC_{(a,c_i)} + PAC_{(a,c_{i-1})}} * 100$ <p><b>Donde:</b>  <i>TUA<sub>(a,c<sub>i</sub>)</sub></i> es la tasa de variación de una cohorte <i>c</i> en correspondencia a un área <i>a</i>, en relación al año anterior.</p>
<b>Interpretación</b>	Expresa la línea de tendencia sobre el uso de los contenidos del área, permitiendo analizar el interés en la misma para el desarrollo de los PFC. Si el resultado es positivo entre una cohorte y otra, indica que la utilización de los contenidos del área se ve incrementada.

**Tabla 3. Indicadores Complementarios****7.4 Valoración de los indicadores propuestos**

A partir del análisis de los indicadores, se busca, entre otras cosas, dejar en evidencia cuales áreas se destacan, como las más utilizadas, según la carrera. A partir de ellas se puede inferir, por ejemplo, en el caso de Ing. en Sistemas de Información, el empleo de metodologías de sistemas y tecnologías para el procesamiento de información, respondiendo a lograr uno de los objetivos previstos curricularmente “un ingeniero creativo capaz de generar cambios”.

La información brindada por los indicadores es necesaria, pues permite analizar si el alumno ha logrado integrar la información proveniente de distintos campos disciplinarios concurrentes en un proyecto común, o se ha desarrollado preferentemente en un área particular. Esta información resulta de insumo para

orientar de la búsqueda laboral del egresado de acuerdo a las incumbencias, para tomar decisiones académicas (equipo docente, evaluación, proceso de enseñanza aprendizaje) orientadas a fortalecer áreas que hasta el momento no han sido consideradas y son requeridas para fortalecer el perfil del egresado, entre otras.

## Capítulo 8: La herramienta: SisDA

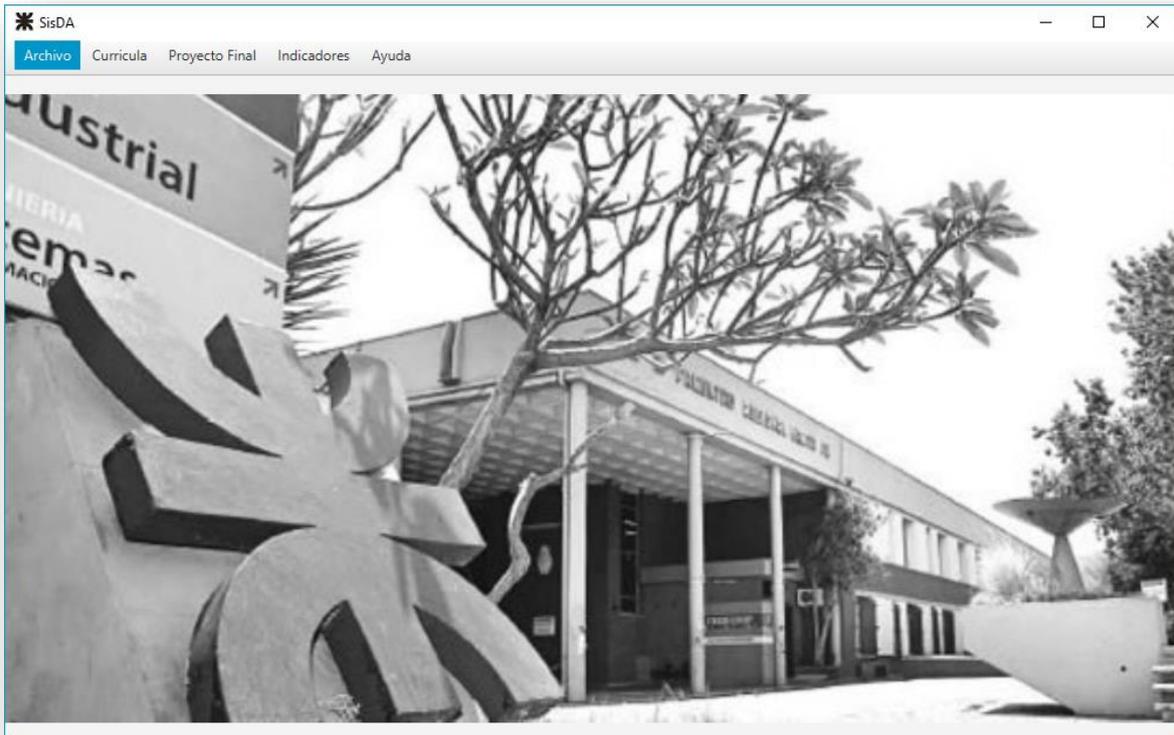
### 8.1 Descripción general

La herramienta se desarrolló mediante los módulos funcionales previamente descritos, tanto en la organización del código fuente como en la interfaz gráfica planteada.

En relación a esta última, se diseñó en base a las siguientes características:

- **Familiaridad:** se optó por una interfaz estándar para herramientas de escritorio, buscando favorecer la familiaridad con el usuario como característica principal.
- **Consistencia:** se mantiene un diseño consistente a lo largo de toda la aplicación, ya que el mismo permite a los usuarios reconocer patrones de uso. Por otro lado evita ambigüedades y confusiones.
- **Flexibilidad:** se brinda una interfaz flexible de acuerdo a las preferencias de interacción, ya sea por medio del uso del teclado, mouse, touchpad, etc.
- **Sencillez:** debido a que en principio su uso es meramente académico, se diseñó una interfaz sencilla de utilizar.

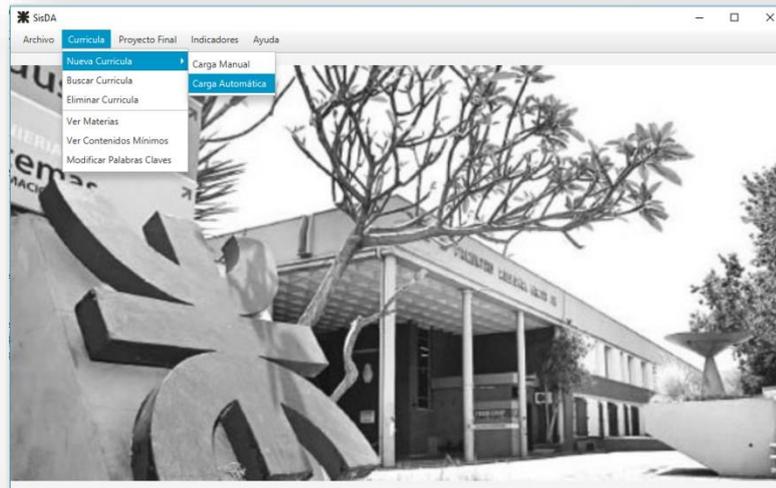
La **Figura 11** muestra una vista de la ventana principal, donde puede verse claramente que dentro del menú de la misma se reflejan los módulos planteados en el presente proyecto.



*Figura 11. Interfaz gráfica de SisDA*

## 8.2 Módulo 1: Plan de Estudios

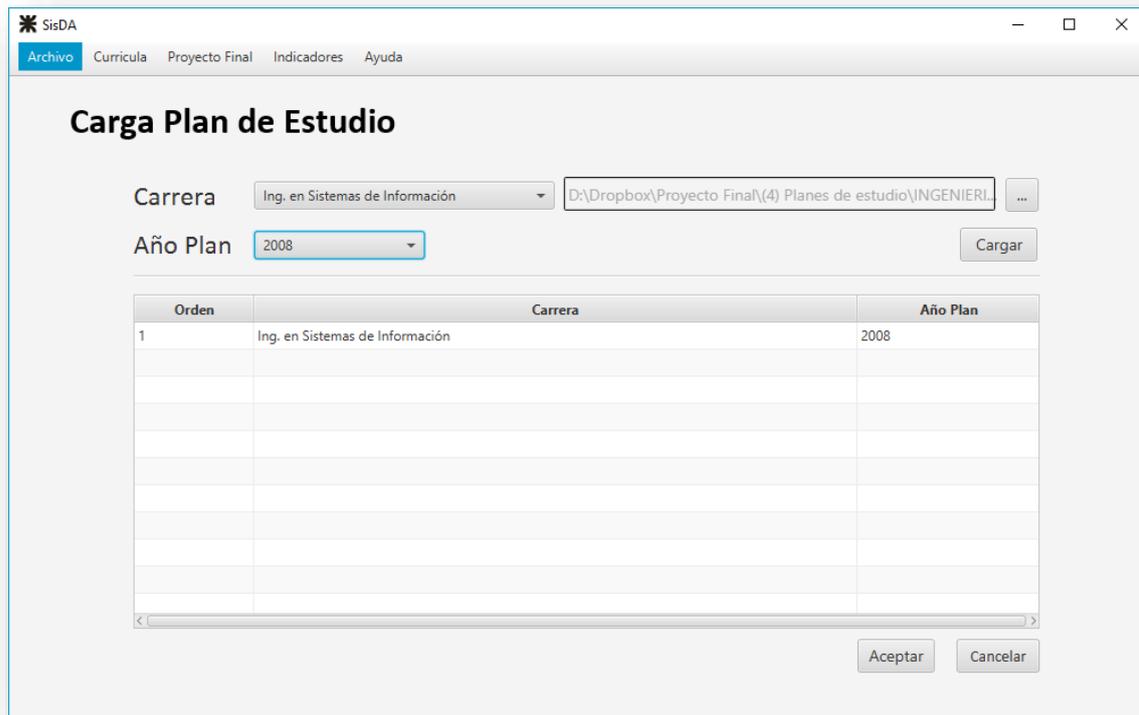
Entre las funcionalidades planteadas dentro del menú del módulo 1 en relación a las curriculas (planes de estudios) se encuentran: la carga tanto automática como manual, la búsqueda y/o eliminación de una curricula, y dentro de una curricula específica, la identificación de las materias y los contenidos mínimos de la misma o la modificación de las palabras claves asociadas. En la **Figura 12** se puede apreciar una vista del menú de 'Curricula' perteneciente a este módulo de la aplicación.



**Figura 12.** Menú de Currícula

Si bien este módulo tiene un alto componente de procesamiento interno, más allá de la parte gráfica, la función principal del menú es la carga automática de la currícula. En dicho módulo puede cargarse el archivo en formato doc desde el equipo donde se está trabajando.

Además, considerando el año del plan de estudios, al seleccionar una carrera, la herramienta sólo deja introducir planes que correspondan a los actualmente en vigencia. Por otra parte, se muestran en la interfaz los planes actualmente disponibles en la base de datos. La **Figura 13** muestra la interfaz gráfica de la carga automática del Plan de Estudios.



**Figura 13.** Interfaz para carga automática de Plan de Estudios

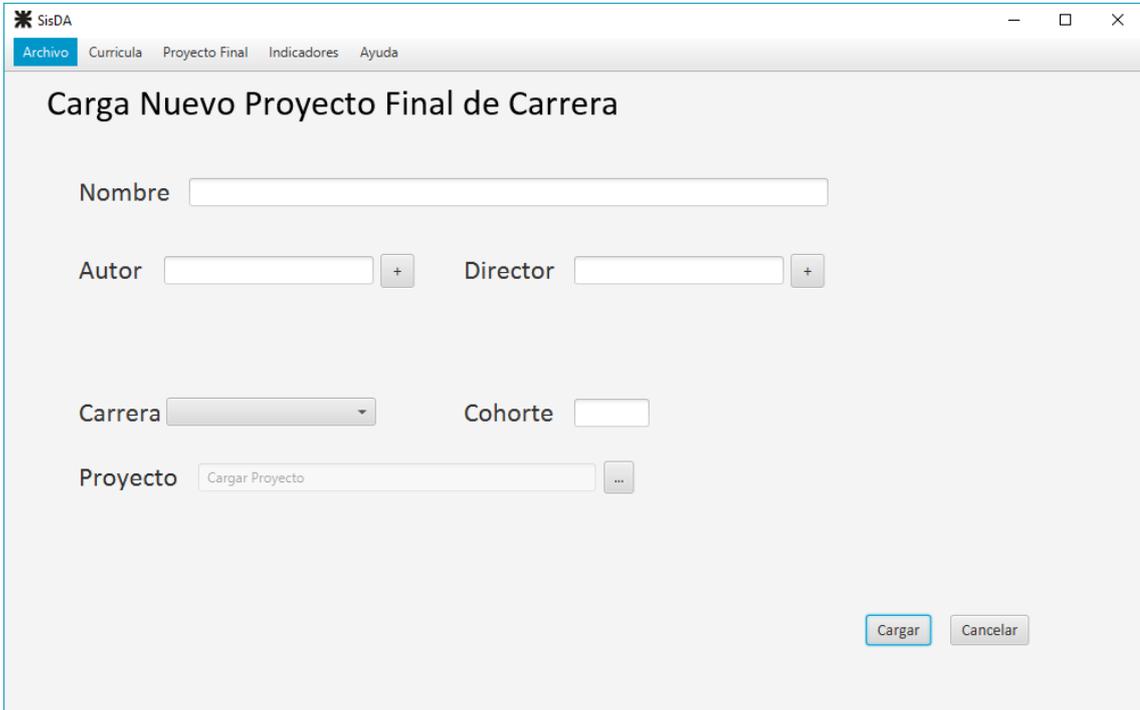
### 8.3 Módulo 2: Proyectos Finales

Dentro del menú de Proyectos Finales se puede realizar la carga de un nuevo proyecto, como así también buscar o listar los disponibles en la base de datos. Por otro lado, para un proyecto que ya haya sido analizado, se pueden observar las materias y los contenidos relacionados. En este caso, al igual que con los planes de estudios, el punto más significativo se encuentra en la posibilidad de carga de los PFC.



**Figura 14.** Menú de Proyecto Final.

Para dicha carga se consideró el nombre, los autores y directores, la cohorte y la carrera a la cual pertenece cada proyecto. En el caso de los autores se permitió ingresar hasta tres personas y en el caso de directores se permitió cargar dos, en el caso que el proyecto posea un director y un codirector. En la **Figura 15** se muestra la interfaz realizada para la carga de PFC.



The screenshot shows a web application window titled 'SisDA' with a menu bar containing 'Archivo', 'Curricula', 'Proyecto Final', 'Indicadores', and 'Ayuda'. The main content area is titled 'Carga Nuevo Proyecto Final de Carrera'. It contains several input fields: a text box for 'Nombre', two text boxes for 'Autor' and 'Director' each with a '+' button, a dropdown menu for 'Carrera', a text box for 'Cohorte', and a text box for 'Proyecto' with a 'Cargar Proyecto' button and a three-dot menu icon. At the bottom right, there are two buttons: 'Cargar' and 'Cancelar'.

**Figura 15.** Interfaz de carga de PFC.

#### 8.4 Módulo 3: Indicadores

Según lo analizado dentro de los requerimientos, se consideró necesario implementar una interfaz que propicie la buena interpretación de los índices e indicadores. Por tal motivo, este módulo plantea una interfaz sencilla de utilizar, permitiendo además la posibilidad de generar gráficos que faciliten la interpretación de los datos cuando corresponda.

En el caso del índice MA, se plantea una ventana dividida en dos, donde en la parte superior se puede realizar la búsqueda de proyectos de una determinada cohorte, para luego realizar la selección del mismo; y en la parte inferior un ranking del índice para todas las materias relacionadas con el proyecto analizado. Dicho ranking se

presenta en formato de tabla para una mejor lectura del mismo. Luego el usuario puede optar por la vista gráfica para tener una visión más concreta sobre la distribución de los datos.

En la **Figura 16** se puede ver la interfaz del índice MA y en la **Figura 17** una gráfica perteneciente a dicho índice.

**Media Aritmetica para Materia en Proyecto Final**

Carrera: Ing. en Sistemas de Información Cohorte: 2011 [Buscar]

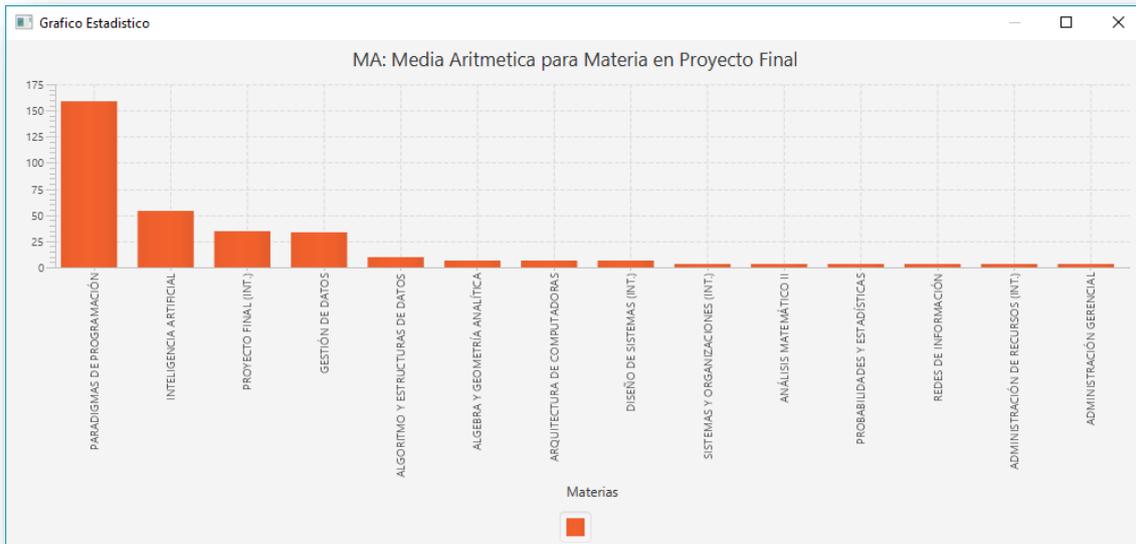
Orden	Proyecto
9	Sistema de Gestión del conocimiento basado en ontologia
11	Juego de Estrategias OnLine
13	Artificial Life Contest

[Analizar]

Orden	Materia	Área	MA
1	PARADIGMAS DE PROGRAMACIÓN	Programación	159
2	INTELIGENCIA ARTIFICIAL	Modelos	54
3	PROYECTO FINAL (INT.)	Sistemas de Información	35
4	GESTIÓN DE DATOS	Programación	33
5	ALGORITMO Y ESTRUCTURAS DE DATOS	Programación	10
6	ALGEBRA Y GEOMETRÍA ANALÍTICA	Matemática	7
7	ARQUITECTURA DE COMPUTADORAS	computación	6

[Ver grafico] [Aceptar]

**Figura 16.** Interfaz gráfica índice MA para un PFC.



**Figura 17.** Gráfica índice MA para un PFC

Además, el menú ofrece la selección de indicadores en base a los planteados en el **Capítulo 6**: de proyecto, de cohorte y complementarios.

Para los indicadores de proyecto, la interfaz es similar al índice MA, con la salvedad que la parte inferior de la pantalla presenta la selección del indicador, el área o materia que se quiera analizar. También permite la opción de establecer el ranking en relación al indicador correspondiente y observar la interpretación del mismo.

En el caso de los indicadores de cohorte o complementarios, la interfaz es similar a la descrita anteriormente, con la salvedad que no presenta la lista de proyectos correspondientes a la cohorte, ya que el análisis no requiere de la selección de uno en particular. Esta interfaz planteada se puede observar en la **Figura 18**.

SisDA

Archivo Curricula Proyecto Final Indicadores Ayuda

### Indicadores por Cohorte

Carrera  Cohorte

Indicador  Materia/Área

---

### Porcentaje de Área por Cohorte (PAC)

Resultado **42 %**

Calculo

$$PAC_{(a,c)} = \frac{\sum(MA_{(m_i, PFC)})}{\sum(MA_{(m, PFC)})} * 100$$

Figura 18. Interfaz del Indicador PAC

## Capítulo 9: Validación de la herramienta

### 9.1 Medida de performance

Para evaluar la funcionalidad de la herramienta implementada, se proponen algunas medidas utilizadas en la recuperación de información, que permiten obtener una métrica respecto a la eficiencia en la evaluación de los documentos.

Precision y Recall son métricas empleadas en la medida del rendimiento de los sistemas de búsqueda y recuperación de información y reconocimiento de patrones.

#### 9.1.1 Cálculo de Precisión y Recall

- **Precision (P):** es la razón entre el número de documentos relevantes recuperados y el número de documentos recuperados. La expresión del cálculo viene dada por la **Fórmula 5**.

$$P = \frac{\text{Documentos recuperados y clasificados correctamente}}{\text{Total de documentos recuperados}}$$

*Fórmula 5. Cálculo de Precision.*

- **Recall (R):** es la razón entre la proporción de documentos relevantes recuperados, y el total de los documentos existentes en la base de datos,  $\gamma$  más allá de que se recuperen o no, definida según la **Fórmula 6**.

$$R = \frac{\text{Documentos encontrados y clasificados correctamente}}{\text{Total de documentos existentes}}$$

*Fórmula 6. Cálculo de Recall.*

#### 9.1.2 Interpretación probabilística

Es posible interpretar Precision y Recall no como relaciones sino como probabilidades.

En este sentido:

- Precision es la probabilidad de que un documento analizado (seleccionado al azar) sea relevante.
- Recall es la probabilidad de que un documento correspondiente a una categoría (seleccionado al azar) sea recuperado correctamente en una búsqueda.

## 9.2 Evaluación de la herramienta

Las métricas consideradas están orientadas a la recuperación de documentos y categorización de los mismos. Sin embargo, la herramienta a evaluar determina un valor de relación entre todos los documentos de una base de datos y las materias del plan de estudios. Por esta razón, la aplicación de las métricas planteadas no puede ser de forma directa.

En este caso, se puede considerar que los documentos recuperados son todos los documentos analizados. Por otro lado, haciendo una búsqueda de las palabras claves dentro de los proyectos y evaluando de manera manual el uso de las mismas, se sabrá si los PFC están bien categorizados por la herramienta en función de la medida de MA asignada a cada materia.

Por último, la métrica adoptada sería la precisión, para conocer la probabilidad de que un documento sea evaluado de manera correcta por la herramienta. En función de lo expuesto, la **P** queda expresada según la **Fórmula 7**.

$$P = \frac{\text{número de documentos analizados adecuadamente}}{\text{número de documentos analizados}}$$

*Fórmula 7. Expresión de P para evaluar la herramienta.*

### 9.2.1 Evaluación de documentos

Tomando como muestra los PFC de las cohortes 2011 y 2012, a partir del estudio previo realizado, se trabajó sobre 18 PFC. Para cada uno de ellos se analizaron, de manera manual, las materias relacionadas a su desarrollo. Por otra parte, se procedió a realizar el cálculo de MA con la herramienta y se analizaron las materias relacionadas que la misma arrojaba. Seguidamente se volcaron los datos en una tabla, a fin de cotejar los resultados abordados. En la **Tabla 4** se presenta una síntesis de la comparación de los datos, en la que se indican las materias más relacionadas con los PFC que se observaron (Materias Observadas) y que fueron más relevantes en el análisis del sistema (Materias Analizadas), como así también el resultado deducido.

PROYECTO	COHORTE	MATERIAS OBSERVADAS	MATERIAS ANALIZADAS	RESULTADO
PROYECTO 1	2011	IA;PP;DS	PP;PYE;AYEDD	INCORRECTO
PROYECTO 2		DS;	AYGA;PP;SYO	INCORRECTO
PROYECTO 3		AS; DS; PP	PP;GD;DS	CORRECTO
PROYECTO 4		IA; PP; GD	PP;IA;GD	CORRECTO
PROYECTO 5		IA;PP;GD	PP;IA;GD	CORRECTO
PROYECTO 6		DS;SG;RC	SYO;PP;GD	INCORRECTO
PROYECTO 7		DS;PP;GD	PP;DS;GD	CORRECTO
PROYECTO 8		DS;IA;GD	IA;PP;GD	CORRECTO
PROYECTO 9		PYE;DS;RC	RC;AYEDD;PYE	CORRECTO
PROYECTO 10	2012	DS; GD; PP	PP; AYEDD; IA	CORRECTO
PROYECTO 11		DS; GD; PP	AYEDD; PP; GD	CORRECTO
PROYECTO 12		GD; RC; PP; DS	SYO; IA; PP	INCORRECTO
PROYECTO 13		DS; AYEDD; GD	PP; GD; AYEDD; DS	CORRECTO
PROYECTO 14		AYEDD; DS; GD	PP; AYEDD; GD	CORRECTO
PROYECTO 15		AYEDD; GD; PP	PP; AR; GD	CORRECTO
PROYECTO 16		PP; DS; GD	PP; GD; SYO	CORRECTO
PROYECTO 17		DS; GD; PP	AYEDD; AC; PP	INCORRECTO
PROYECTO 18		DS; PYE; GD	PP; IA; GD	CORRECTO

**REFERENCIAS:** DISEÑO DE SISTEMAS (DS); INTELIGENCIA ARTIFICIAL (IA); ANALISIS DE SISTEMAS (AS); PARADIGMAS DE PROGRAMACIÓN (PP); ALGORITMOS Y ESTRUCTURAS DE DATOS (AYEDD); GESTION DE DATOS (GD); REDES DE COMUNICACIÓN (RC); PROBABILIDAD Y ESTADISTICAS (PYE); SISTEMAS DE GESTIÓN (SG); ALGEBRA Y GEOMETRIA ANALITICA (AYGA); SISTEMAS Y ORGANIZACIONES (SYO); ADMINISTRACIÓN DE RECURSOS (AR).

**Tabla 4.** Evaluación de PFC

### 9.2.2 Cálculo y análisis de la medida de performance

Luego de haber analizado los PFC de manera manual y utilizando SisDA, se constató que 13 de los 18 documentos fueron bien analizados por la herramienta. Aplicando la **Fórmula 7** para determinar la Precision se obtiene un resultado de 0,722, equivalente a un 72,2% de precisión en el análisis de los PFC. Dicha métrica obtenida se considera aceptable para realizar estudios preliminares de los proyectos. La misma puede ser mejorada trabajando sobre las palabras claves relacionadas con los contenidos de las materias. Es importante recordar que tales palabras claves se obtienen de manera automática por la herramienta al procesar el plan de estudio, y luego pueden ser modificadas o mejoradas por el usuario.

## Capítulo 10: Conclusiones

### 10.1 Conclusiones generales

Todo proyecto denota un desafío, implica metas a alcanzar, las cuales requieren de esfuerzos y conlleva momentos interesantes de aprendizaje y avances como también obstáculos y dificultades.

El presente proyecto no fue la excepción, siendo la pieza final para concretar una etapa y la piedra fundacional para comenzar una nueva; conteniendo además una fuerte carga social por representar la finalización de los estudios en una Universidad Pública.

SisDA no sólo intentó resolver una necesidad, sino que buscó ser un aporte para el desarrollo de las carreras de ingeniería, permitiendo llevar adelante un análisis que promueva la mejora continua de las mismas.

De esta forma, la herramienta favoreció el estudio de los contenidos de las curriculas y su relación con las necesidades que la sociedad demanda al perfil del ingeniero.

Si bien la realización del proyecto tuvo momentos críticos, se pudieron resolver los inconvenientes para concretar los objetivos planteados desde el principio. Esto implicó decisiones que, si bien podían modificar el curso del proyecto, no fueron al azar, sino que estuvieron estipuladas en el plan de gestión de riesgos. Por tal motivo, el plan de contingencia para la gestión de riesgos fue de suma importancia en el desarrollo del proyecto.

Algunos de los riesgos que debieron sortearse tuvieron que ver con la implementación de una herramienta de procesamiento de lenguaje natural y el tiempo estimado para la codificación, además de situaciones personales que prolongaron el tiempo previsto para el desarrollo de la herramienta. Ante tales situaciones se optó por la posibilidad de replanificar y reorganizarse en función de las prioridades del sistema.

Pese a lo antes mencionado, es importante rescatar que el desarrollo del presente proyecto constituyó una instancia más de aprendizaje personal aportando nuevos conocimientos en relación al desarrollo y gestión de proyectos, la temática

abordada, y los recursos y debilidades personales para enfrentar desafíos y momentos críticos.

Para concluir, se destaca el impacto positivo de este proyecto como culminación de la carrera de grado y comienzo de la vida profesional.

## 10.2 Conclusiones del sistema

El sistema desarrollado cumple con los requerimientos necesarios para el proyecto en el que se enmarca, lo cual representa un aporte positivo para el mismo.

Si bien la herramienta propuesta como solución requiere de mayor depuración mediante evaluaciones de distintos proyectos en el marco de la investigación, la misma puede servir como base para desarrollar una herramienta sustancial que ayude a la toma de decisión académica en las carreras de ingeniería.

A su vez, se podría ampliar a futuro hacia el análisis de todo tipo de proyectos institucionales como ser proyectos de investigación o de extensión.

Es importante destacar también, el potencial del sistema para favorecer la gestión académica, debido a que podrían incluirse como variable de análisis los directores o docentes que participan en los proyectos, y los estudiantes que desarrollan los mismos. Esto permitiría realizar una evaluación en el desarrollo de actividades extra curriculares, identificando fortalezas y debilidades académicas de la institución.

Finalmente, podría considerarse la ampliación del sistema hacia otras tecnologías como ser el desarrollo web o integración con otros sistemas de gestión institucional.

## 10.3 Impacto

En primera instancia el presente proyecto impacta de manera satisfactoria en el desarrollo de las investigaciones previas realizadas en el marco del proyecto de investigación, ya que otorga un análisis significativo en la definición de las relaciones de los PFC con los planes de estudio. En tal sentido, amplía y mejora los indicadores estudiados hasta el momento y agiliza el trabajo de campo realizado, dando la posibilidad de extender el alcance de la investigación.

Por otra parte, la herramienta posibilita la evaluación del proceso enseñanza aprendizaje pues toda actividad evaluativa significa una valoración y que, por lo tanto, debe ser comprendida como una aproximación a la realidad. Esta salvedad inicial se realiza como “llamada de atención” para evitar, desde un primer momento, tanto posturas escépticas como dogmáticas en torno a ella. Estas posiciones son las más frecuentes en el ámbito académico y quizás sean tanto una como otra las que más deterioran el sistema educativo.

En este sentido, el sistema posibilita poner de manifiesto la relación entre el producto final realizado por el alumno y la propuesta curricular que como sabemos debe ser flexible e integral, amplia, cognitivamente significativa y socialmente relevante.

Dicha propuesta se encuentra claramente plasmada en los diseños curriculares de los distintos niveles y evidenciados en áreas curriculares. Por esta razón y ante un modelo de enseñanza-aprendizaje como el que se propone, es necesario repensar el proceso de enseñanza y los resultados de aprendizaje plasmados en los proyectos finales.

Por lo expuesto, el desarrollo de proyectos que evalúen el avance académico presenta una mejora continua para las carreras de ingeniería, impactando de manera favorable en el desarrollo de incumbencias que den respuesta a las demandas de la sociedad.

#### 10.4 Factibilidad

La herramienta desarrollada no exige importantes recursos informáticos, ni tiempos de procesamiento. Además, los documentos requeridos para análisis son fácilmente obtenidos por lo que es factible de ser utilizada.

En relación a los documentos de los PFC, se encuentra actualmente en desarrollo el Repositorio Institucional Abierto (RIA), creado por la Ordenanza N° 1480 del Consejo Superior de la UTN. El RIA es un archivo digital cuyo objetivo principal es brindar acceso abierto a producciones científicas y académicas elaboradas por docentes, investigadores y tecnólogos de la Universidad. En este sentido, una vez en funcionamiento el repositorio, los informes de los PFC serán accesibles de manera

on line, posibilitando un mejor acceso, como así también la potencialidad de integrar al sistema el uso del mismo.

Por último, las características brindadas por la interfaz gráfica y la calidad de la información otorgada hacen posible ampliar su uso a diferentes agentes organizacionales.

## Anexos

### Anexo 1: Plan de Gestión de Riesgos

Debido al carácter unipersonal del presente proyecto, se llevo a cabo para la educación de riesgos el uso de la Taxonomía desarrollada por el Software Engineering Institute (SEI), y complementada por la Universidad de las Ciencias Informáticas (UCI) considerando la Gestión de Proyectos. Los eventos hallados servirán para identificar las amenazas y posibilitar la planificación de acciones previas y posteriores a cada evento.

#### 1.1 Aplicación del método TBQ - Taxonomía de Riesgos

Clase: Gestión del Proyecto		
Elemento	Atributo	Consideración
<b>Gestión de la integración del Proyecto</b>	Desarrollo del plan de gestión del proyecto.	El Plan de Proyecto está elaborado.
	Dirección y gestión de la ejecución del proyecto.	La dirección, supervisión y control del proyecto está establecido.
	Supervisión y control del trabajo del proyecto.	
	Realización de un control integrado de cambios.	No se consideran cambios.
	Cierre del proyecto o la fase.	Los cierres de las fases están bien definidos.
<b>Gestión del Alcance del Proyecto</b>	Obtención de los requerimientos del proyecto	Tanto los requerimientos, como el alcance del proyecto están bien establecidos y son controlables.
	Definición del alcance del proyecto.	
	Creación de la estructura de desglose del trabajo del proyecto.	
	Verificación del alcance del proyecto.	
<b>Gestión del Tiempo del Proyecto</b>	Control del alcance del proyecto.	Están bien definidas en el presente Plan de Proyecto.
	Definición de las actividades del proyecto.	
	Establecimiento de la secuencia de actividades del proyecto	

	<p>Estimación de recursos de las actividades del proyecto.</p> <p>Estimación de la duración de las actividades.</p> <p>Desarrollo del cronograma del proyecto.</p> <p>Control del cronograma del proyecto.</p>	<p>Los Recursos establecidos para el proyecto están bajo control.</p> <p>Pueden presentar una amenaza, debido a que el desarrollo del proyecto es unipersonal. No existen actividades a desarrollar en paralelo o equipos de desarrollo extras.</p>
<b>Gestión del Costo del Proyecto</b>	<p>Estimación de costes del proyecto.</p> <p>Determinación del presupuesto del proyecto.</p> <p>Control de costes del proyecto.</p>	No posee erogaciones presupuestarias.
<b>Gestión de la Calidad del Proyecto</b>	<p>Planificación de la calidad del proyecto.</p> <p>Realización del control de la calidad del proyecto.</p> <p>Realización del aseguramiento de la calidad del proyecto.</p>	No está establecido un control de la calidad del proyecto.
<b>Gestión de las Personas del Proyecto</b>	<p>Planificación de los recursos humanos.</p> <p>Adquisición del equipo de trabajo del proyecto.</p> <p>Desarrollo del equipo del proyecto.</p> <p>Gestión del equipo del proyecto.</p>	No existen amenazas en la gestión de RR HH, debido al desarrollo unipersonal del proyecto.
<b>Gestión de las Comunicaciones</b>	<p>Identificación de los involucrados al proyecto.</p> <p>Planificación de las comunicaciones del proyecto.</p> <p>Distribución de la información del proyecto.</p> <p>Gestión de las expectativas de los involucrados.</p> <p>Información del rendimiento/estado del proyecto.</p>	<p>Debido a que el desarrollo es unipersonal no existen amenazas en torno a la comunicación entre los RR HH del proyecto y los involucrados están bien definidos (directores y profesores de cátedra).</p> <p>La información del estado y monitoreo del proyecto están definidos en el presente plan.</p>
<b>Gestión de adquisiciones del Proyecto</b>	Planificación de las adquisiciones del proyecto	

Dirección de las adquisiciones.	No se consideran adquisiciones para el presente proyecto.
Administración de las adquisiciones.	
Cierre de las adquisiciones.	

**Tabla 1 - Anexo 1. Taxonomía de Riesgos para la Clase de Gestión del Proyecto.**

Clase: Ingeniería del Producto		
Elemento	Atributo	Consideración
<b>Negocio</b>	Modelo de dominio.	Al tratarse de un software para la gestión institucional, el modelo de negocio define como clientes a los gestores institucionales. Como no existen intenciones de lucro, el negocio se restringe a la buena utilización del producto y a las decisiones adoptadas a partir del uso del mismo.
	Modelo de negocio.	
	Glosario de términos.	
<b>Requerimientos</b>	Estabilidad	Al tratarse de una herramienta enmarcada en un proyecto de investigación, los requerimientos se encuentran establecidos de manera clara y completa.
	Completitud.	
	Claridad.	Estudios previos validan la necesidad de una herramienta para este tipo de requerimientos.
	Validez.	
	Factibilidad.	Se requiere analizar particularmente la factibilidad en cuanto al análisis de los PFC para poder obtener las palabras claves en los tiempos establecidos.
	Precedentes.	Se investigaron otros precedentes para casos similares de procesamiento de lenguaje natural.
	Escala.	La escala y el modelo del sistema se encuentran establecidos.
Modelo de sistema.		
<b>Diseño</b>	Funcionalidad	No presenta riesgos considerables.
	Dificultad.	
	Interfaces.	Puede llegar a ocasionar alguna dificultad el diseño de gráficos o reportes en la interfaz.

	Rendimiento.	No es riesgoso, ni tampoco implica una performance rigurosa.
	Capacidad de probarse.	No ofrece amenaza al sistema.
	Restricciones de hardware.	No posee restricciones de Hardware.
	Software de terceras partes.	El software se realizara sin contratos a terceros.
	Desarrollo del análisis y diseño	
<b>Gestión de Pruebas</b>	Planificación de las pruebas	Se realizaran pruebas de unidad y de integración con los aportes de algún usuario para el diseño de los mismos. No presenta amenazas al proyecto.
	Diseño de caso de prueba.	
	Desarrollo de listas de chequeo.	
	Tratamiento de evaluaciones y elementos no conformes.	
	Pruebas de unidad.	Se consideran elementos bajo control, y que no implican mayor riesgo considerando el presente proyecto.
	Factibilidad de las pruebas de unidad.	
	Ambiente de integración y pruebas.	
Integración del producto.		
	Integración del sistema.	
<b>Codificación</b>	Factibilidad	Considerando la especificidad del proyecto, es deseable hacer hincapié en la factibilidad de codificación, especialmente en el Modulo 2. Ya que se trabajara con herramientas no tan conocidas, se debe anticipar riesgos en cuanto a las restricciones de implementación.
	Restricciones de implementación	
	Mantenimiento	El mantenimiento del sistema debe ser sencillo, aunque la ventaja se presenta en que el usuario del mismo será el propio desarrollador en una primera instancia.
	Fiabilidad	El sistema debe ser fiable, respecto a la información que brinda. Para ello hay que hacer énfasis en el

		tratamiento de los datos, especialmente en el Modulo 3.
<b>Arquitectura</b>	Implementación y satisfacción de la seguridad.	En principio no se considera necesario medidas de seguridad para el sistema, debido al ámbito de aplicación del mismo.
	Experiencia para implementar los niveles de seguridad.	
	Interfaz de usuario.	Si bien es considerable el desarrollo de una buena interfaz, la arquitectura de la misma no presenta riesgos aparentes.
	Especificaciones.	No presenta inconvenientes.
	Definición de la arquitectura de software.	Se debe prestar atención al desarrollo de la misma, para facilitar la reutilización de módulos.
	Arquitectura de información.	No se consideran riesgos aparentes.

**Tabla 2 – Anexo 1. Taxonomía de Riesgos para la Clase Ingeniería del Producto.**

Clase: Entorno de desarrollo		
Elemento	Atributo	Consideración
<b>Proceso de Desarrollo</b>	Formalidad.	Debido al desarrollo unipersonal, y la metodología planteada de acuerdo al conocimiento de la misma, no presenta riesgo alguno.
	Adecuación.	
	Control del proceso.	
	Familiaridad.	
	Control del producto.	
<b>Sistemas de Desarrollo</b>	Adecuación.	En cuanto a las características del producto, de su implantación en el ámbito de “producción” y su usabilidad, no se consideran riesgos aparentes.
	Usabilidad.	
	Familiaridad.	
	Fiabilidad.	
	Soporte.	
	Entregas.	No están consideradas entregas del mismo.
	Herramientas de desarrollo.	Dada la utilización de herramientas de procesamiento de lenguaje natural, y el desconocimiento hacia las mismas, presenta un riesgo en cuanto a la búsqueda, evaluación y selección. La imposibilidad de hallar una que se adapte a las necesidades del proyecto acarrea un riesgo en la concreción del mismo.
<b>Entorno de Trabajo</b>	Actitud ante la calidad.	En relación a que el proyecto se desarrolla en el marco de un proyecto de investigación y además de que es de carácter unipersonal, no presentan complicaciones asociadas al Entorno de Trabajo.
	Cooperación.	
	Comunicación.	
	Moral.	

**Tabla 3 – Anexo 1. Taxonomía de Riesgos para la Clase Entorno de desarrollo.**

Clase: Restricciones del Programa		
Elemento	Atributo	Consideración
Recursos	Cronograma.	No hay restricciones en cuanto al cronograma, más de las personales.
	Personal Presupuesto	No aplican al proyecto.
	Facilidades	No se considera un elemento en riesgo.
Contratos	Tipo de contrato. Restricciones. Dependencias.	El desarrollo del presente proyecto no implica contratación alguna.
	Cliente.	Los clientes están claramente identificados y se los puede ir consultando a medida que avanza el proyecto, para orientar el desarrollo de las interfaces del mismo, lo que hace suponer la inexistencia de riesgos.
	Integrador	La integración del mismo está bien definida.
Interfaces del programa	Contratistas asociados. Subcontratistas. Contratista principal.	No se utilizarán contratistas en el proyecto.
	Proveedor.	La utilización de software libre en el desarrollo, permite asumir la posibilidad de riesgos.
	Política.	El proyecto se enmarca en una política académica institucional, dentro de un proyecto de investigación. No se aprecian riesgos al respecto.

**Tabla 4 – Anexo 1.** Taxonomía de Riesgos para la Clase de Restricciones del Programa.

## 1.2 Evaluación de los riesgos

Aquellos riesgos que se consideraron relevantes dentro del análisis, se someten a una evaluación cualitativa, teniendo en cuenta las distintas categorías observadas en la

**Tabla 5.**

Categoría	Probabilidad	Impacto
1	Ínfima	Despreciable
2	Baja	Marginal
3	Media	Critico
4	Alta	Catastrófico

**Tabla 5 – Anexo 1. Clasificación de los Riesgos.**

Para caracterizar el grado de exposición del riesgo de manera cualitativa, se recurre a la ecuación:  $ER: P \cdot D$ , siendo P la probabilidad de ocurrencia y D el impacto ocasionado. De acuerdo con los resultados obtenidos, se determina la categoría de nivel de riesgo que se muestran en la **TABLA 6**.

Categoría	ER	Nivel de Riesgo
<b>A</b>	$1 \leq ER < 4$	Aceptable
<b>B</b>	$4 \leq ER < 8$	Poco Riesgoso
<b>C</b>	$8 \leq ER < 12$	Riesgoso
<b>D</b>	$12 \leq ER \leq 16$	Muy Riesgoso

**Tabla 6 – Anexo 1. Categorías de nivel de riesgo.**

En la **Tabla 7** se observa la evaluación realizada.

Ponderación de Riesgos						
Elemento de la Taxonomía	N°	Riesgo identificado	P	D	ER	
<b>Gestión del Proyecto</b> Gestión del tiempo del Proyecto	Estimación de la duración de las actividades.	Imprecisiones en la estimación del tiempo en que se llevaran a cabo las tareas definidas, en especial las referidas a la etapa de desarrollo. Actividades críticas en el desarrollo del modulo 2, en especial en la implementación de una herramienta de procesamiento de lenguaje natural.	4	2	8	
	Desarrollo del cronograma del proyecto.					
	Control del cronograma del proyecto.					
<b>Ingeniería de Producto</b> Requerimientos	Factibilidad	II	No hallar una herramienta que soporte el procesamiento del texto en lenguaje español.	2	3	6
<b>Ingeniería de Producto</b> Diseño	Interfaces	III	Puede no encontrarse un framework adecuado para la realización de reportes gráficos que brinde una buena interfaz de usuario.	2	2	4
<b>Ingeniería de Producto</b> Codificación	Factibilidad	IV	Debido a la falta de dominio en el lenguaje/ implementación del mismo, puede llegar a ocasionar demoras la codificación para los módulos 1 y 2.	3	4	12
	Restricciones de implementación					
	Fiabilidad	V	El tratamiento de los datos para el modulo 3 puede no ser el adecuado, debido a falta de conocimiento, o errores de interpretación del desarrollador, tal aspecto podría ocasionar información que no sea fiable.	2	3	6
<b>Ingeniería de Producto</b> Arquitectura	Definición de la Arquitectura del Software.	VI	Dificultades para llevar a cabo un diseño de arquitectura específico que permita una buena implementación y un posterior reuso.	2	2	4
<b>Entorno de Desarrollo</b>	Herramientas de desarrollo.	VII	La imposibilidad de hallar una herramienta de	2	4	8

Sistema de Desarrollo	procesamiento de lenguaje natural que se adapte a las necesidades del proyecto acarrearía un riesgo en la concreción del mismo.
-----------------------	---

**Tabla 7 – Anexo 1.** Evaluación de los riesgos.

### 1.3 Plan de Contingencia

Para llevar adelante las acciones de contingencia se adoptan los siguientes criterios:

- Los riesgos que arrojaron un grado de exposición menor a 8, que serían del tipo A y B, serán tolerados, no planificándose medidas de contingencia.
- Aquellos que arrojaron un grado de exposición mayor o igual a 8, riesgos del tipo C y D, serán abordados con medidas preventivas apropiadas y con acciones concretas en el caso de hacerse efectivos. De manera tal que le proyecto pueda seguir en curso.

En la **Tabla 8** se ordena el plan de contingencia de manera decreciente según el valor de ER, con el número de riesgo identificado y la clase a la que pertenece el mismo.

Riesgo	Clase	Medidas de Contingencia	
		Preventivas	Correctivas
IV	D	Se realizara una autocapacitación con los medios disponibles, con la ayuda de personas que conocen el tema dentro la Facultad. El tiempo destinado a esto se encuentra contemplado en el cronograma del plan, aunque de ser necesario podrían realizarse horas extras, para el uso exclusivo de la investigación de la herramienta.	Si restan pocos días para culminar la fase de desarrollo, se procederá a replanificar el cronograma referido a dicha fase.
I	C	Utilización de herramientas de soporte como Microsoft Project, apoyándose en técnicas de planificación como Diagrama de Gantt.	Si se avizora el fin de la etapa en el cronograma, y se estima que la misma no llega a concretarse a término, se deberá replanificar la actividad reformulando el cronograma, y verificando las actividades restantes, si no tienen un error de estimación similar.
VII	C	Incrementar la carga horaria destinada a la búsqueda de una herramienta y establecer criterios de prioridad en base a las funcionalidades mínimas de la aplicación que esta herramienta permita codificar.	Reorganizar las funcionalidades en base a las prioridades del sistema, y verificar si es posible la reformulación de los objetivos específicos que serian solventados con esta herramienta, buscando la posibilidad de ser suplantados por algún otro método.

**Tabla 8 – Anexo 1.** Medidas preventivas y correctivas de los riesgos del tipo C y D.

Anexo 2: Base de Datos

2.1 Diagrama de análisis de Base de Datos

Se presenta en la **Figura 1** del presente anexo, el diagrama de análisis de la base de datos obtenido luego de haber recogido los requerimientos del sistema.

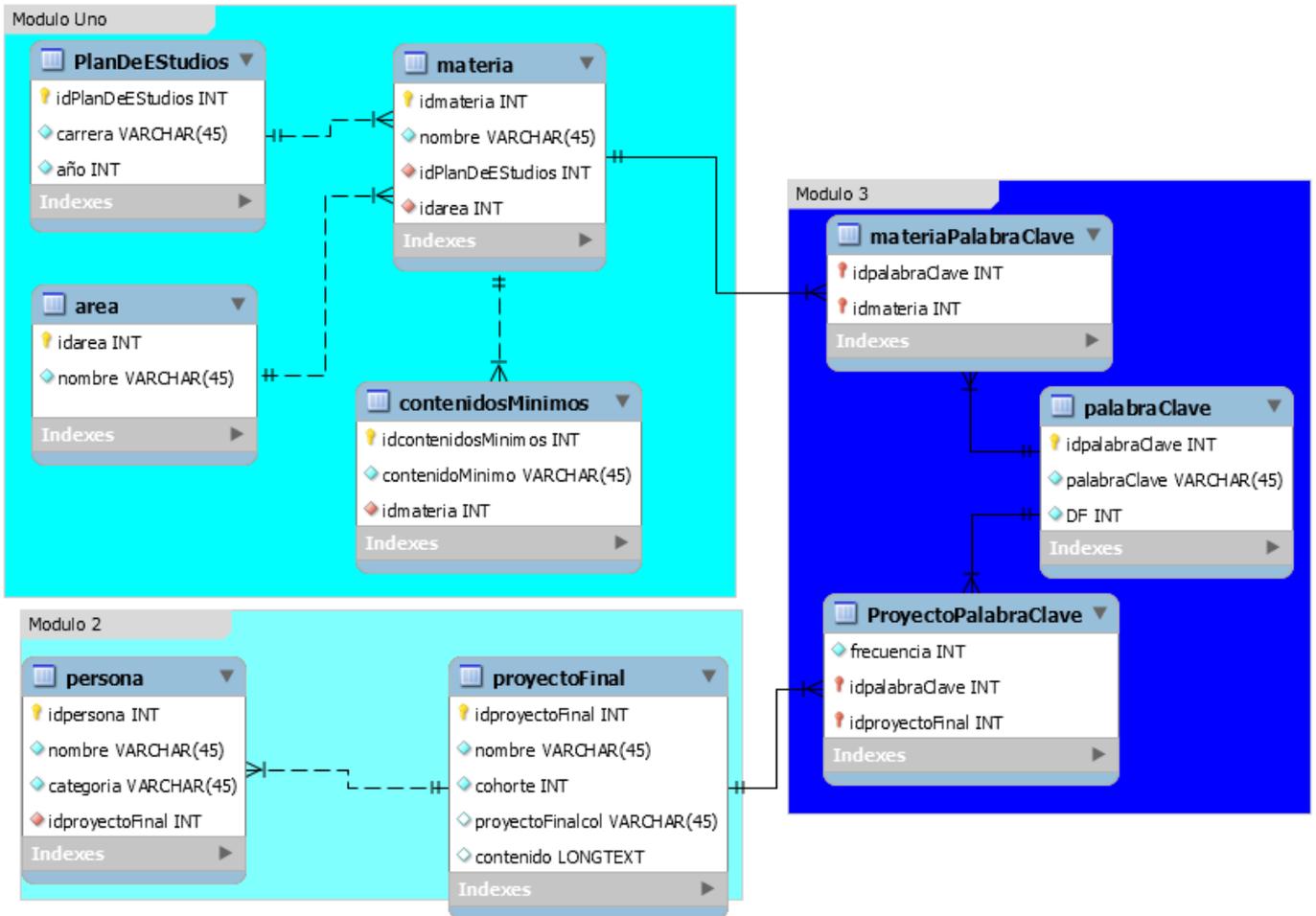


Figura 1 – Anexo 2. Diagrama de Base de datos



## 3.2 Especificaciones de C.U.

A continuación, se presentan las especificaciones de los principales casos de usos desarrollados para el sistema.

CU: Nueva Curricula - Carga Automática	
<b>Complejidad:</b>	Media
<b>Descripción</b>	Mediante este caso de uso el sistema permite cargar una nueva curricula a la base de datos.
<b>Precondiciones</b>	El actor debe contar con un archivo doc perteneciente a la curricula.
<b>Secuencia Normal</b>	<b>Paso: Acción:</b>
	1. El actor selecciono Nueva Curricula - Carga Automática en el Menú Curricula.
	2. Se presenta la pantalla con los campos a completar/cargar y la lista de curriculas actualmente en la base de datos.
	3. El actor completa/carga todos los datos de la pantalla y presiona Cargar
	<ul style="list-style-type: none"> <li>• Carrera (Combo seleccionable)</li> <li>• Archivo (botón para cargar archivo)</li> <li>• Año plan (Numérico)</li> </ul>
	El sistema valida los datos ingresados
	4. El sistema detecta que los datos ingresados no son correctos. El sistema muestra un mensaje referenciando el error sobre el campo erróneo.
	4.a.1 Volver al <b>Paso 3</b> .
	El sistema detecta que los datos ingresados son correctos.
	5. El sistema procesa el plan de estudios (CU: Procesar Plan de Estudios)
5.a El sistema almacena todos los datos en la base de datos	
5.b El sistema muestra la nueva curricula en la lista actual.	
5.c El sistema limpia todos los campos	
5.d	
6. El actor presiona el botón <b>Aceptar</b> .	
7. El sistema vuelve a la pantalla Principal.	
<b>Excepciones</b>	El actor presiona <b>Cancelar</b> . El sistema vuelve a la Pantalla Principal
<b>Pos-Condicion</b>	
<b>Nota</b>	

**Tabla 1-Anexo 3. Caso de Uso Nueva Curricula - Carga Automática.**

CU: Nuevo Proyecto		
<b>Complejidad: Media</b>		
<b>Descripción</b>	Mediante este caso de uso el sistema permite cargar proyecto a la base de datos.	
<b>Precondiciones</b>	<ol style="list-style-type: none"> <li>1. El actor debe contar con un archivo doc perteneciente al informe de un PFC.</li> <li>2. La curricula perteneciente a la carrera del PFC debe encontrarse en la base de datos.</li> </ol>	
<b>Secuencia Normal</b>	<p><b>Paso: Acción:</b></p> <ol style="list-style-type: none"> <li>1. El actor selecciono Nuevo Proyecto - en el Menú Proyecto Final.</li> <li>2. Se presenta la pantalla con todos los campos a completar/cargar.</li> <li>3. El actor completa/carga todos los datos de la pantalla y presiona Cargar <ul style="list-style-type: none"> <li>• Nombre</li> <li>• Autor (Apellido, Nombre)</li> <li>• Director (Apellido, Nombre)</li> <li>• Carrera (Combo seleccionable)</li> <li>• Cohorte (Numérico)</li> <li>• Proyecto (botón para cargar archivo)</li> </ul> </li> <li>4. El sistema valida los datos ingresados <ol style="list-style-type: none"> <li>4.a El sistema detecta que los datos ingresados no son correctos. El sistema muestra un mensaje referenciando el error sobre el campo erróneo. <ol style="list-style-type: none"> <li>4.a.1 Volver al <b>Paso 3</b>.</li> </ol> </li> <li>4.a.2 El sistema detecta que los datos ingresados son correctos.</li> </ol> </li> <li>5. El sistema procesa el archivo del PFC <ol style="list-style-type: none"> <li>5.a El sistema busca cada palabra clave del plan correspondiente dentro del proyecto (CU: Buscar Palabras Claves) <ol style="list-style-type: none"> <li>5.a.1 El sistema almacena todos los datos en la base de datos</li> <li>5.a.2 El sistema limpia todos los campos</li> </ol> </li> <li>5.b El sistema limpia todos los campos</li> </ol> </li> <li>6. El actor presiona el botón <b>Aceptar</b>.</li> <li>7. El sistema vuelve a la pantalla Principal.</li> </ol>	
	<b>Excepciones</b>	El actor presiona <b>Cancelar</b> . El sistema vuelve a la Pantalla Principal
	<b>Pos-Condiciones</b>	
	<b>Nota</b>	En el caso de Autor, se le permitirá cargar hasta tres. En el caso de Director se le permitirá cargar hasta dos.

Tabla 2-Anexo 3. Caso de Uso Nuevo Proyecto.

CU: Media Aritmética para Materias	
<b>Complejidad: Media</b>	
<b>Descripción</b>	Mediante este caso de uso el sistema muestra la media aritmética calculada para cada materia en relación a un PFC.
<b>Precondiciones</b>	El PFC que se desea analizar debe estar cargado en la base de datos.
<b>Secuencia Normal</b>	<b>Paso: Acción:</b>
	1. El actor selecciono Índice de Materia (MA) en el Menú Indicadores.
	2. Se presenta la pantalla con los campos a seleccionar
	3. El actor selecciona los campos correspondientes y presiona <b>Buscar</b>
	<ul style="list-style-type: none"> <li>• Carrera (Combo seleccionable)</li> <li>• Cohorte (Combo seleccionable)</li> </ul>
	4. El sistema muestra una tabla con los proyectos que se encuentran en la base de datos correspondientes a los campos seleccionados
	5. El actor selecciona un proyecto y presiona <b>Analizar</b>
	6. El sistema calcula las MA para cada materia del plan de estudios correspondiente a la carrera seleccionada
	7. Se muestra una tabla en la parte inferior con las materias, el área y el MA correspondiente
	El actor presiona <b>Ver Grafico</b>
	8. 8.a El sistema abre una nueva ventana con un gráfico mostrando la distribución de los datos.
	8.a.1 El actor presiona <b>Aceptar</b>
	8.a.2 El sistema cierra la pantalla con el grafico.
	6. El actor presiona el botón <b>Aceptar</b> .
7. El sistema vuelve a la pantalla Principal.	
<b>Excepciones</b>	
<b>Pos-Condicion</b>	
<b>Nota</b>	

Tabla 3-Anexo 3. Caso de Uso Media Aritmética para Materia..

CU: Indicadores de Cohorte		
<b>Complejidad: Media</b>		
<b>Descripción</b>	Mediante este caso de uso el sistema permite obtener indicadores de cohorte	
<b>Precondiciones</b>	La cohorte que se desea analizar debe contar con PFC en la base de datos	
<b>Secuencia Normal</b>	<b>Paso:</b>	
	<b>Acción:</b>	
	1.	El actor selecciono Indicadores de Cohorte en el Menú Indicadores.
	2.	Se presenta la pantalla con los campos a seleccionar
		El actor selecciona los campos correspondientes y presiona <b>Analizar</b>
	3.	<ul style="list-style-type: none"> <li>• Carrera (Combo seleccionable)</li> <li>• Cohorte (Combo seleccionable)</li> <li>• Indicador (Combo seleccionable)</li> <li>• Materia/Área (Combo seleccionable - De acuerdo al Indicador seleccionado)</li> </ul>
		El sistema verifica que todos los campos fueron seleccionados.
	4.	4.a El sistema detecta que algunos/todos los campos no fueron seleccionados. El sistema muestra un mensaje referenciando el error sobre el campo no seleccionado.
		4.b Volver al <b>Paso 3</b> .
		El sistema corrobora que todos los campos han sido seleccionados
	5.	5.a El sistema muestra en la parte inferior de la pantalla el resultado del Indicador seleccionado
		5.b El sistema muestra la Formula para el calculo del Indicador
		El actor presiona <b>Ver Interpretación</b>
	6.	6.a El sistema abre una nueva ventana con la interpretación del Indicador calculado.
		6.a.1 El actor presiona <b>Aceptar</b>
		6.a.2 El sistema cierra la pantalla con la interpretación
		El actor presiona <b>Realizar Ranking</b>
	7.	7.a El sistema abre una nueva ventana donde muestra una tabla con el ranking del indicador para las demás áreas/materias (según corresponda) de la misma cohorte
		6.a.1 El actor presiona <b>Aceptar</b>
		6.a.2 El sistema cierra la pantalla con el ranking.
6.	El actor presiona el botón <b>Aceptar</b> .	
7.	El sistema vuelve a la pantalla Principal.	
<b>Excepciones</b>		

Pos- Condiciones
Nota

**Tabla 4-Anexo 3. Caso de Uso Indicadores de Cohorte.**

CU: Procesar Plan de Estudios	
<b>Complejidad: Media</b>	
<b>Descripción</b>	Mediante este caso de uso el sistema procesa la curricula
<b>Precondiciones</b>	Se deb� haber cargado un archivo perteneciente a un plan de estudios
<b>Secuencia Normal</b>	<b>Paso: Acci3n:</b>
	1. El sistema detecta el Programa Sintetico dentro del plan de estudios
	2. El sistema extrae cada una de las materias
	3. El sistema extrae para cada materia los contenidos m�nimos
	4. El sistema parsea los contenidos m�nimos para extraer palabras claves.
<b>Excepciones</b>	
<b>Pos-Condiciones</b>	Los datos extra�dos deben ser almacenados en la base de datos
<b>Nota</b>	

**Tabla 5-Anexo 3. Caso de Uso Procesar Plan de Estudios.**

#### Anexo 4: Diagrama de Clases

En los siguientes apartados se muestran los diagramas de clases del sistema. En ellos se pueden observar la estructura del código fuente de la herramienta dividido por módulos. Las clases que no poseen relaciones con otras, es debido a que las mismas solo contienen métodos del tipo static, lo cual implica que dichos métodos pueden ser invocados sin crear ningún objeto de su clase. Estas clases comúnmente son utilizadas para agrupar métodos comunes sin necesidad de crear una instancia de las mismas.

4.1 Diagrama de clases Modulo 1: Planes de Estudio

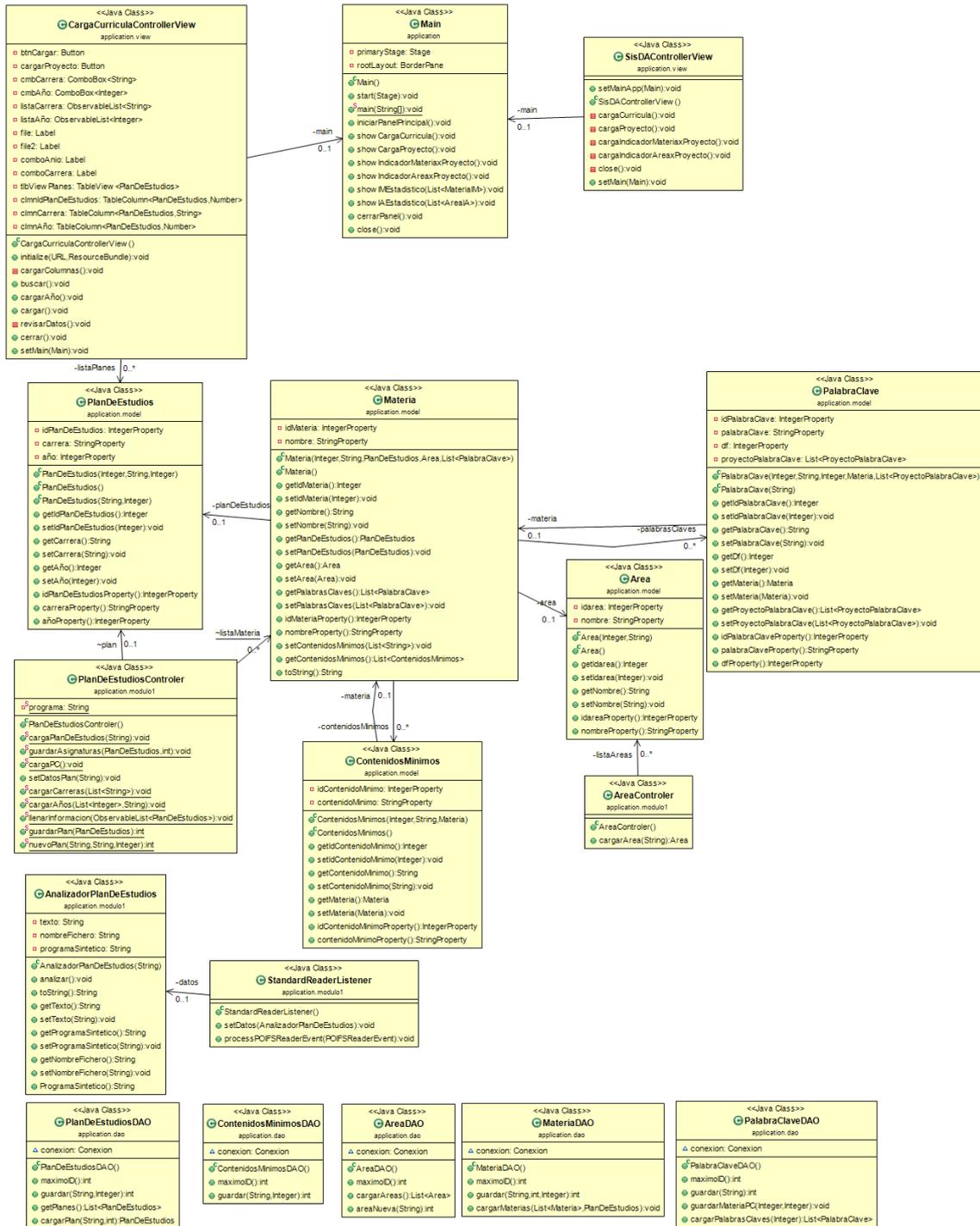


Figura 1-Anexo 4. Diagrama de Clases Modulo 1.

4.2 Diagrama de clases Modulo 2: Proyectos Finales

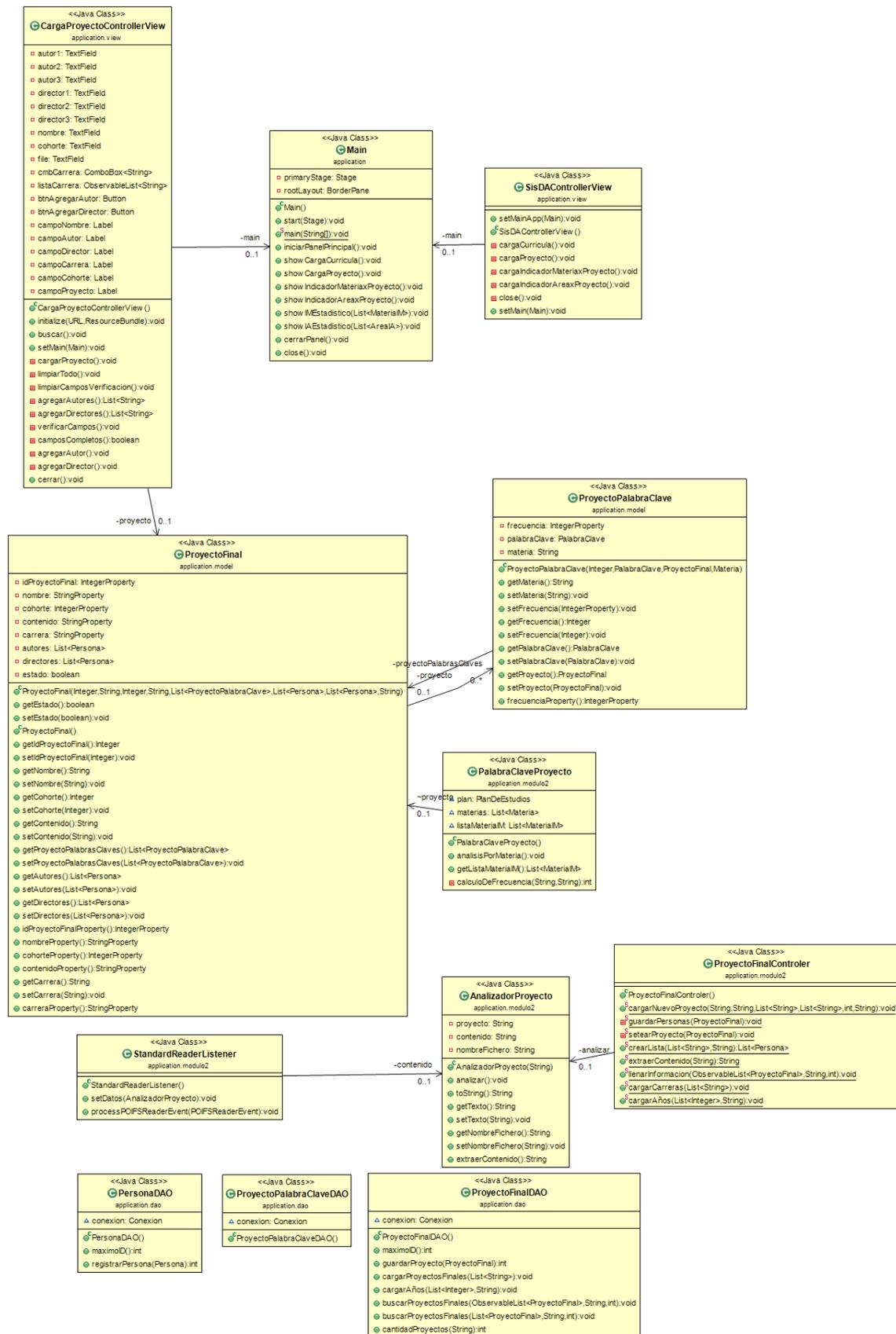


Figura 2-Anexo 4. Diagrama de Clases Modulo 2.

43 Diagrama de clases Modulo 3: Indicadores

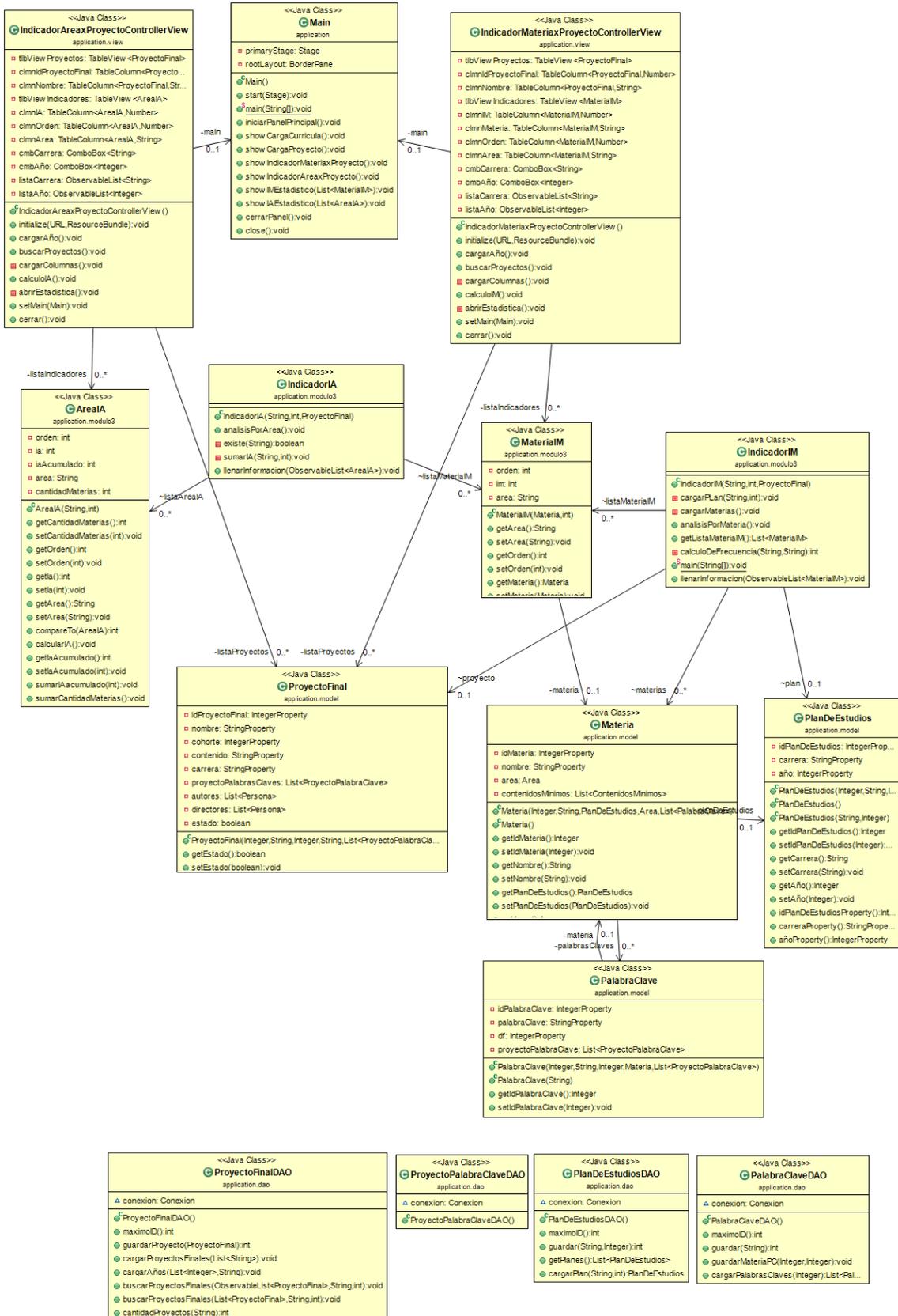


Figura 3-Anexo 4. Diagrama de Clases Modulo 3.

## Anexo 5: Validación de la herramienta

En el presente anexo se muestran los datos más relevantes extraídos a partir del análisis de 18 PFC, los cuales permitieron llevar a adelante la validación de la herramienta.

### 5.1 Palabras claves más destacadas

En la **Tabla 1** del presente anexo, se pueden observar las palabras claves más halladas en el análisis realizado.

Palabras claves	
casos de uso	uml
requerimientos	algoritmos
bases de datos	clases
metodología / métodos	diagrama entidad-relación
modelo de datos	estadística
diagrama de clases	métricas
gestión	minería de datos
modelado / modelo	paquetes
patrones de diseño	protocolo
arquitectura	prototipo
documentación	simulación
implementación	administración
programación	capa
pruebas	casos de prueba
interfaz de usuario	diagrama de actividad
rup	diagrama de componentes
cliente-servidor	diagrama de despliegue
proceso de desarrollo	

**Tabla 1 – Anexo 5.** Palabras Claves halladas.

## 5.2 Materias más relacionadas

En función de las palabras claves halladas se pudo relacionar las mismas con materias, las cuales se encuentran expresadas en la Tabla 2 del presente anexo.

<b>materias</b>
diseño de sistemas
análisis de sistemas
ingeniería de software
gestión de datos
paradigmas de programación
redes de información
algoritmo y estructura de datos
administración gerencial
sistemas de gestión
probabilidad y estadística
proyecto final
simulación
administración de recursos
inteligencia artificial
investigación operativa
comunicaciones

**Tabla 2 – Anexo 5.** Palabras Claves halladas.

## 5.3 Análisis con la herramienta

Del análisis de los PFC realizados con la herramienta se realizaron algunas capturas de pantalla, las cuales se presentan en las siguientes imágenes del anexo.

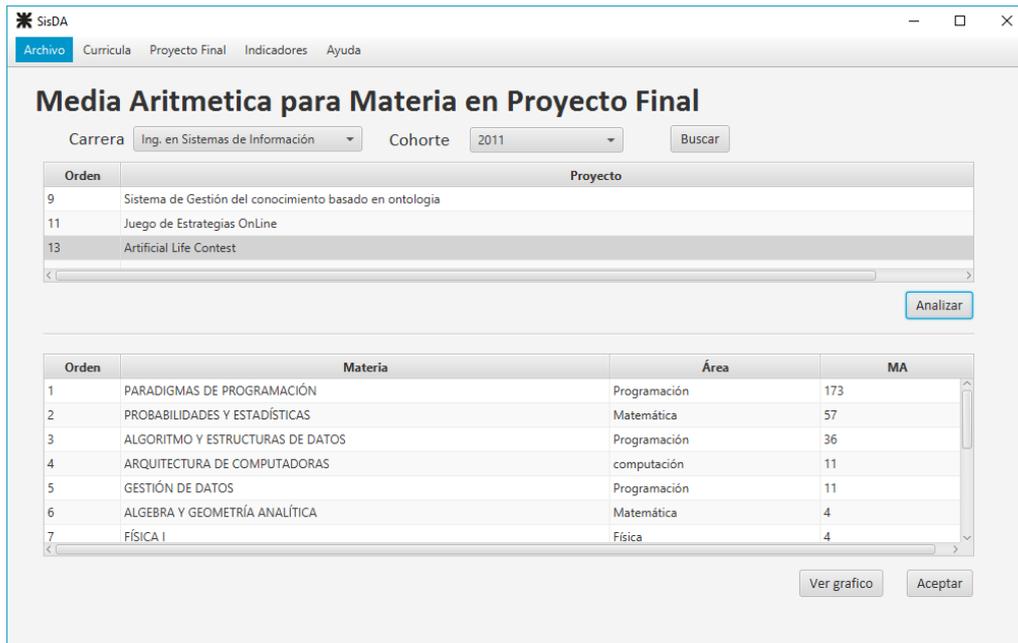


Figura 1 – Anexo 5. Análisis Proyecto 1.

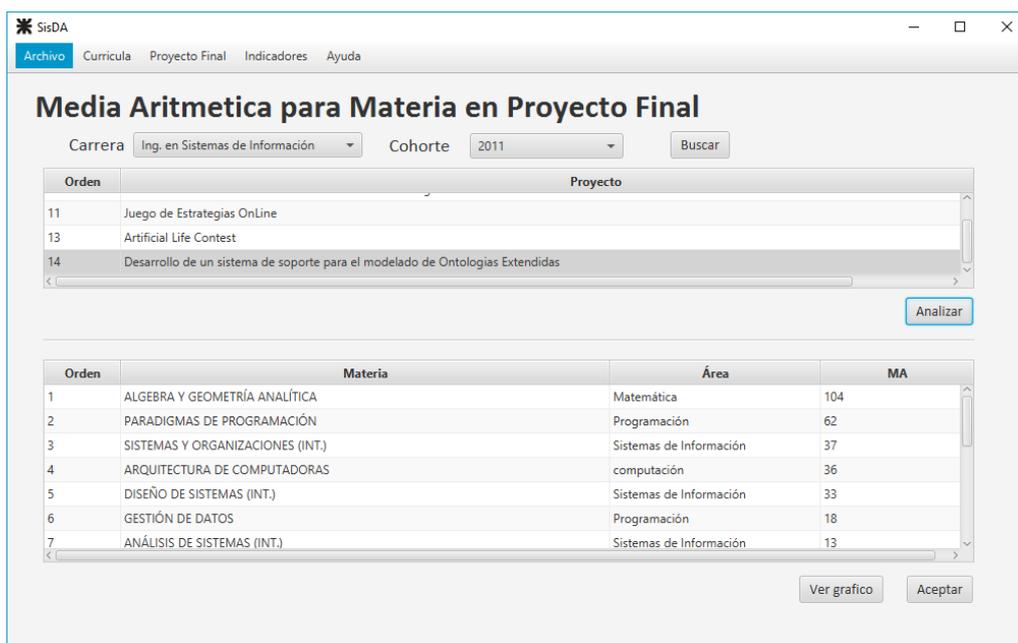


Figura 2 – Anexo 5. Análisis Proyecto 2.

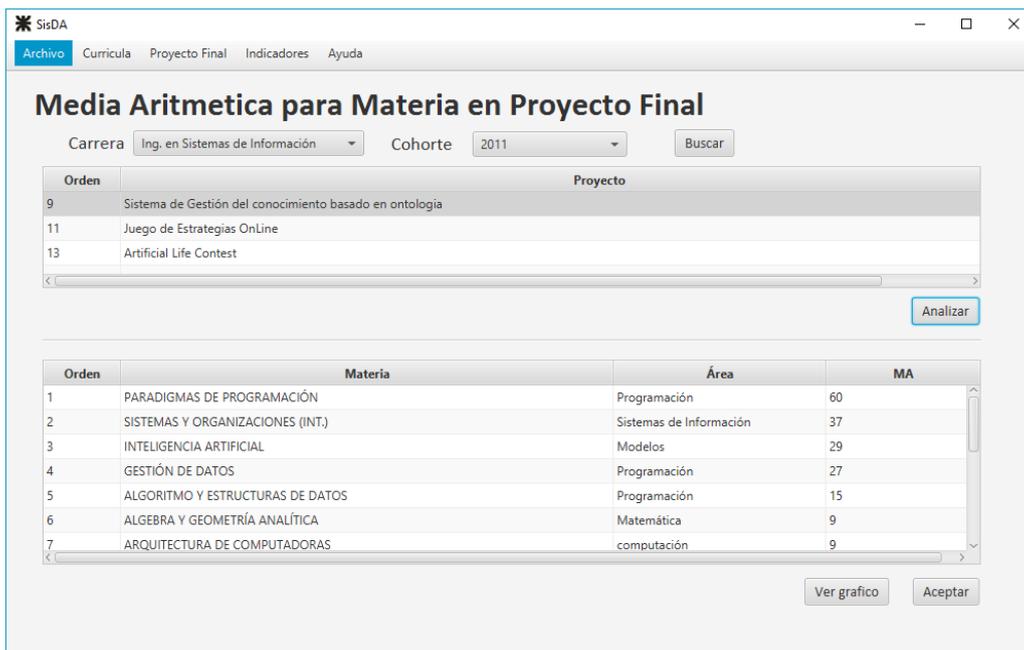


Figura 3 – Anexo 5. Análisis Proyecto 5.

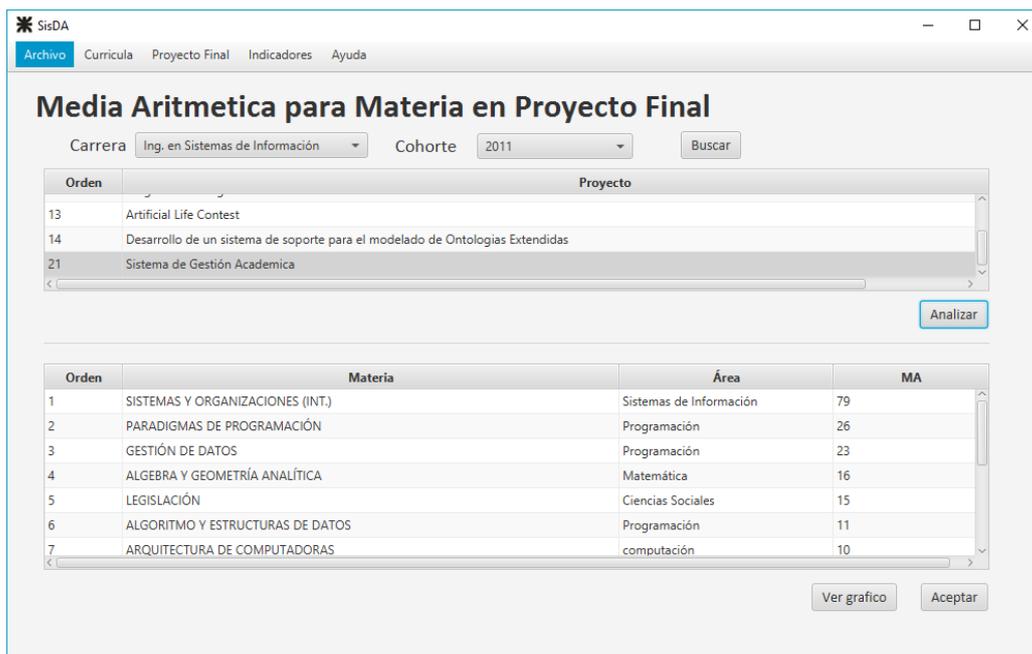


Figura 4 – Anexo 5. Análisis Proyecto 6.

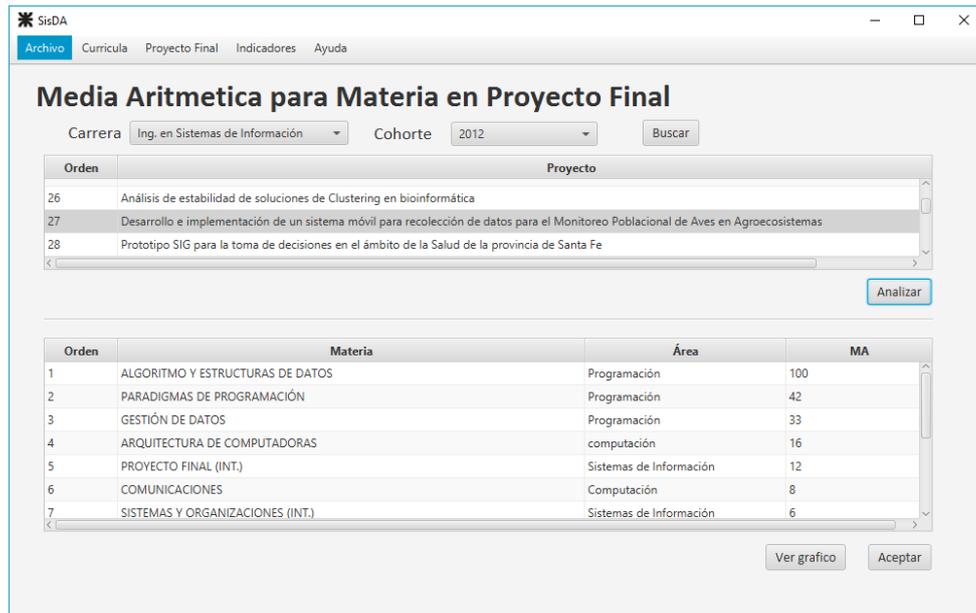


Figura 5 – Anexo 5. Análisis Proyecto 11.

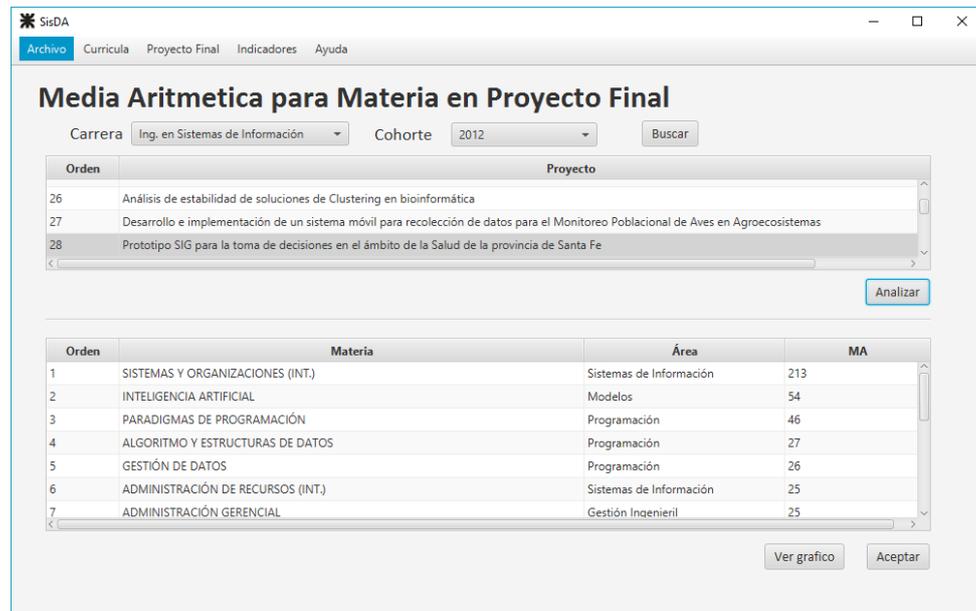


Figura 6 – Anexo 5. Análisis Proyecto 12.

## Referencias

- [1] Marulanda, Agustín; Pires, María; Delgado, José Gregorio. Ingeniería Eléctrica: hacia una propuesta de currículum por competencias; Revista de la Facultad de Ingeniería U.C.V., Vol. 25, N° 3, 2010.
- [2] Ordenanza N° 1030 – Diseño Curricular Ing. Civil; Ordenanza N° 1150 – Diseño Curricular Ing. en Sistemas de Información; Ordenanza N° 1114 – Diseño Curricular Ing. Industrial; Ordenanza N° 1026 – Diseño Curricular Ing. Eléctrica; Ordenanza N° 1027 – Diseño Curricular Ing. Mecánica. Ordenanzas del Consejo Superior Universitario de la Universidad Tecnológica Nacional.
- [3] La competencias sociales y los indicadores en Ingenierías. Conde, Sergio Daniel.
- [4] Reflexiones y perspectivas de la Educación Superior en América Latina. Pablo Beneitone y otros. Proyecto Tuning. 2004 – 2007.
- [5] Res 340/14. Reglamento General de Proyecto Final de Carrera para las especialidades de grado. Consejo Directivo. Facultad Regional Santa Fe. Universidad Tecnológica Nacional.
- [6] Demuth Mercado, Patricia B. (2004). “Modelos Curriculares. Análisis y Reconstrucción”. Comunicaciones Científicas y Tecnológicas. Universidad Nacional del Nordeste, Facultad de Humanidades- Instituto de Ciencias de la Educación. Resistencia, Chaco. Argentina.
- [7] Lenguajes Documentarios. Dolores Vizcaya Alonso. 1997.
- [8] Java [http://es.wikipedia.org/wiki/Java\\_\(lenguaje\\_de\\_programaci%C3%B3n\)](http://es.wikipedia.org/wiki/Java_(lenguaje_de_programaci%C3%B3n))
- [9] MySQL <https://es.wikipedia.org/wiki/MySQL>
- [10] JavaFX <http://es.wikipedia.org/wiki/JavaFX>
- [11] Subversion [https://es.wikipedia.org/wiki/Subversion\\_\(software\)](https://es.wikipedia.org/wiki/Subversion_(software))
- [12] Eclipse [https://es.wikipedia.org/wiki/Eclipse\\_\(software\)](https://es.wikipedia.org/wiki/Eclipse_(software))
- [13] Java EE [https://es.wikipedia.org/wiki/Java\\_EE](https://es.wikipedia.org/wiki/Java_EE)
- [14] MySQL Workbench [http://es.wikipedia.org/wiki/MySQL\\_Workbench](http://es.wikipedia.org/wiki/MySQL_Workbench)
- [15] Apache POI [https://en.wikipedia.org/wiki/Apache\\_POI](https://en.wikipedia.org/wiki/Apache_POI)
- [16] Tortoise [https://tortoisesvn.net/docs/nightly/TortoiseSVN\\_es/](https://tortoisesvn.net/docs/nightly/TortoiseSVN_es/)
- [17] Visual SVN Server <https://www.visualsvn.com/server/>
- [18] TF-IDF <https://es.wikipedia.org/wiki/Tf-idf>
- [19] Latent Dirichlet Allocation [https://es.wikipedia.org/wiki/Latent\\_Dirichlet\\_Allocation](https://es.wikipedia.org/wiki/Latent_Dirichlet_Allocation)

- [20] Latent semantic analysis.[https://en.wikipedia.org/wiki/Latent\\_semantic\\_analysis](https://en.wikipedia.org/wiki/Latent_semantic_analysis).
- [21] Popular Lectures and Addresses, Vol. I. Thomson, W. London. 1891.

## Bibliografía

- *Russell, R. and B. Taylor, Operations Management. Focusing on Quality and Competitiveness. Prentice Hall, Inc. ,2006.*
- *Pressman, R.S. Ingeniería de Software: un Enfoque Práctico, McGraw-Hill, 2010.*
- *Julian R. Salvarrey y Otros. Gerenciamiento de Proyectos con Microsoft Excel y Microsoft Project. OmicrowSystem, Bs As, 2002.*
- *Richard E. Fairley. Managing and Leading Software Projects. IEEE ComputerSociety, 2009.*
- *Ian Sommerville. Software Engineering. Addison – Wesley, 2001.*
- *Clements y otros. Documenting Software Architectures: Views and Beyond. SecondEdition, Addison Wesley, 2011.*
- *Marvin Carr. y otros. Taxonomy-Based Risk Identification. Software Engineering Institute, Carnegie Mellon University. Pittsburgh, Pennsylvania, 1993.*
- *Yampier Medina Tarancón y otros. Una taxonomía para la identificación de riesgos en los proyectos de desarrollo de software de la Universidad de las Ciencias Informáticas. Universidad de las Ciencias Informáticas. La Habana, Cuba, 2012.*
- *Claudio Bracalenti. Gestión de riesgos en los proyectos de software. Monografía. Universidad Nacional de la Plata, 2009.*
- *Term Weighting in Short Documents for Document Categorization, Keyword Extraction and Query Expansion. Mika Timonen. Department of Computer Science Series of Publications A Report A-2013-1.*
- *IMPROVING TEXT CATEGORIZATION BY USING A TOPIC MODEL. WongkotSriurai. Department of Mathematics Statistics and Computer, Faculty of Science, UbonRatchathani University, Thailand*
- *Identifying Keywords and Key Phrases Ashwini Madane. International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012.*
- *Keyword based Text Categorization. Libby Bara. Bar Ilan University, The Department of Computer Science.*

- *Lluís Padró and Evgeny Stanilovsky. FreeLing 3.0: Towards Wider Multilinguality Proceedings of the Language Resources and Evaluation Conference (LREC 2012) ELRA. Istanbul, Turkey. May, 2012.*
- *Lluís Padró. Analizadores Multilingües en FreeLingLinguamatica, vol. 3, n. 2, pg. 13--20. December, 2011.*
- *Automatic keyword prediction using Google similarity distance. Ping-I Chen, Shi-Jen Lin. Department of Information Management, National Central University, Chung-Li 320, Taiwan, ROC.*
- *Estadística básica. Beatriz Spagni, Augusto Dante Bergagna, Gabriela Roldán, Mirta López. Universidad Nacional Del Litoral. 2008.*
- *Aplicación del factor TF-IDF en el análisis semántico de una colección documental. Andrés Vuotto, Celeste Bogetti, Gladys Fernández. Universidad Nacional de Mar del Plata. 2015*
- *Descubrimiento de conocimiento en repositorios documentales mediante técnicas de Minería de Texto y SwarmIntelligence. Cobo Ortega, Rocha Blanco y Alonso Martínez. 2009.*
- *Funciones de Ranking basadas en Lógica Borrosa para IR estructurada. Pérez-Iglesias, Fresno y Pérez-Agüera. 2008.*