



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

La Universidad Católica de Loja

**TITULACIÓN DE INGENIERO EN SISTEMAS INFORMÁTICOS Y
COMPUTACIÓN**

Preprocesador de Planes Académicos de la UTPL

Trabajo de fin de Titulación

Autor:

Calderón Carrión Eyllin Dolores

Director:

Sucunuta España Manuel Eduardo, Ing.

Loja – Ecuador

2012

CERTIFICACIÓN

Ingeniero.

Manuel Eduardo Sucunuta España

DIRECTOR DEL TRABAJO DE FIN DE TITULACIÓN

CERTIFICA:

Que el presente trabajo, denominado: "Preprocesador de Planes Académicos" realizado por el profesional en formación: Eylin Dolores Calderón Carrión; cumple con los requisitos establecidos en las normas generales para la Graduación en la Universidad Técnica Particular de Loja, tanto en el aspecto de forma como de contenido, por lo cual me permito autorizar su presentación para los fines pertinentes.

Loja, 22 de octubre de 2012

f)

CI:

CESIÓN DE DERECHOS

Yo, Eyllin Dolores Calderón Carrión declaro ser autora del presente trabajo y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales.

Adicionalmente declaro conocer y aceptar la disposición del Art. 67 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado que se realicen a través, o con el apoyo financiero, académico o institucional (operativo) de la Universidad”

f)
Autora: Eyllin Dolores Calderón Carrión
Cédula: 1104578206

DEDICATORIA

Con mucho cariño quiero dedicar la presente tesis:

A Dios, por ser la guía en el camino recorrido hasta hoy...

A mi madre Lucía por ser aquella mujer incondicional que ha sabido brindarme su apoyo en las diversas circunstancias de mi vida...

A mi hermana Erika por ser la persona que alegra mis días...

A Juan Carlos que a más de ser el amor de mi vida, es fuerza y la luz que ilumina todos los días de mi vida...

AGRADECIMIENTO

Quiero expresar mis más sinceros agradecimientos al ingeniero Manuel Sucunuta, director de este proyecto de tesis, que sin su con su constante tutoría y guía no hubiera sido posible el desarrollo de este proyecto de tesis.

A mis maestros en especial a quienes conforman la sección departamental de Ingeniería de Software y Gestión de Tecnologías de Información, que me han apoyado de diversas maneras, tanto en el transcurso de mis cinco años de carrera como en la realización de esta tesis.

A mis familiares y amigos que me han brindado su apoyo y se han constituido en un gran aliento para cumplir con mis metas.

Contenido

CERTIFICACIÓN	II
CESIÓN DE DERECHOS	III
DEDICATORIA	IV
AGRADECIMIENTO	V
RESUMEN EJECUTIVO	17
PROBLEMÁTICA	18
OBJETIVOS	20
Sección I:	21
ESTADO DEL ARTE DEL PROCESAMIENTO DE LENGUAJE NATURAL.....	21
1 Capítulo 1	22
Procesamiento de Lenguaje Natural	22
1.1 Introducción	23
1.2 Definición	23
1.3 Áreas de la ciencia relacionadas	24
1.4 Niveles de análisis	25
1.4.1 Nivel Morfológico	25
1.4.2 Nivel Sintáctico	26
1.4.3 Nivel Semántico	27
1.4.4 Nivel Pragmático	28
1.5 Modelos de Procesamiento	28
1.5.1 Modelos Simbólicos	29
1.5.2 Modelos Empíricos o Estadísticos.....	30
1.5.3 Modelos Conexionistas	31
1.5.4 Comparación de los modelos.....	31
1.6 Problemas del Lenguaje Natural	33
2 Capítulo 2	34
Extracción de Información	34
2.1 Introducción	35
2.2 Definición	35
2.3 Fases del Proceso de Descubrimiento de Conocimiento	36
2.3.1 Fase de Selección.....	36
2.3.2 Fase de Pre-procesamiento	37

2.3.3	Fase de Transformación.....	37
2.3.4	Fase de Minería de Datos	37
2.3.5	Fase de Interpretación o Evaluación.....	37
2.4	Herramientas	38
2.5	Problemas.....	40
2.6	Proyectos relacionados	41
2.6.1	Framework para la extracción de información para páginas web semi – estructuradas	45
2.6.2	Sistema de extracción y representación del conocimiento a partir de documentos descriptivos	46
2.6.3	Ontología Multilingüe para una red de conocimiento (Monnet).....	47
Sección II:.....		48
DISEÑO Y CONSTRUCCIÓN DEL PREPROCESADOR DE PLANES ACADÉMICOS.....		48
3	Capítulo 3	49
Descripción del proceso y selección de herramientas		49
3.1	Introducción	50
3.2	Descripción del proceso de construcción.....	50
3.2.1	Actividades de Construcción	50
3.3	Selección de herramientas	53
4	Capítulo 4	62
Construcción del Preprocesador de Planes Académicos de UTPL.....		62
4.1	Introducción	63
4.2	Fase de Inicio	63
4.2.1	Situación Actual.....	63
4.2.2	Descripción de la Problemática	64
4.2.3	Descripción de la solución.....	64
4.2.4	Especificación de Requerimientos.....	65
4.2.5	Especificación de casos de uso	66
4.2.6	Arquitectura de Software.....	66
4.2.7	Esquema de Base de Datos.....	68
4.3	Fase de Preprocesamiento.....	72
4.4	Fase de Comparación.....	73
4.5	Fase de Obtención de Contenidos importantes.....	76

4.6	Fase de Evaluación	78
4.6.1	Análisis de los resultados obtenidos por el sistema.....	78
	Conclusiones, Recomendaciones y Trabajos Futuros	87
	Conclusiones.....	88
	Recomendaciones	89
	Trabajos Futuros	90
	ANEXO A	91
	Etiquetas Eagles.....	91
A.1	Etiquetas Eagles	92
	ANEXO B	94
	Fase de Inicio.....	94
B.1.	Modelo de Procesos de Negocio	95
1.	Introducción	96
1.1.	Objetivo	96
2.	Modelo de Procesos de Negocio Actual	98
3.	Modelo de Procesos de Negocio Propuesto.....	99
B.2.	Situación Actual de la Organización	100
1.	Introducción	101
1.1.	Objetivo	101
1.2.	Alcance	101
1.3.	Definición, Acrónimos y Abreviaturas	101
1.4.	Resumen.....	102
2.	Contexto del Negocio	102
3.	Ideas del Negocio y Estrategias en el contexto del proyecto.....	102
4.	Evaluación de las conclusiones obtenidas del Negocio.....	103
4.1.	Problemas.....	103
4.3.	Nuevas Tecnologías aplicables	104
B.3.	Visión del Sistema.....	105
1.	Introducción	106
1.1.	Objetivo	106
1.2.	Alcance	106
1.3.	Definiciones, Acrónimos y Abreviaturas	107
2.	Posicionamiento del Producto	107

2.1.	Definición del Problema	107
2.2.	Posicionamiento del Producto	108
3.	Resumen de los afectados/involucrados	110
4.	Resumen de Usuarios.....	111
5.	Necesidades de los afectados/usuarios.....	111
5.1.	Necesidades de las secretarias	111
5.2.	Necesidades de los fiscales	111
6.	Resumen del Producto	112
6.1.	Perspectiva del producto	112
6.2.	Resumen de capacidades	113
6.3.	Restricciones.....	114
B.4.	Especificación de Requerimientos	115
1.	Introducción	116
1.1.	Propósito	117
1.2.	Ámbito	117
1.3.	Definiciones, acrónimos y abreviaciones	117
2.	Descripción General	117
2.1.	Asunciones.....	119
3.	Especificación de Requerimientos.....	119
3.1.	Requerimientos Funcionales.....	119
4.7	REQ01 AUTENTICACIÓN INGRESO SISTEMA.....	119
4.8	REQ02 REGISTRAR PLANES ACADÉMICOS EXTERNOS	120
4.9	REQ03 ANALIZAR PLAN ACADÉMICO.....	121
4.10	REQ04 GENERAR ARCHIVO DE TEXTO CON RESULTADO DE ANÁLISIS	122
4.11	REQ05 GENERAR ARCHIVO XML CON CONTENIDOS IMPORTANTES EXTRAÍDOS	123
4.12	REQ06 GENERAR GRÁFICO	123
4.13	REQ07 CREAR NUEVO CORPUS Y LEMATIZACIÓN.....	124
4.14	REQ08 ACTUALIZAR LEMATIZACIÓN EXISTENTE	125
4.15	REQ09 CREAR NUEVAS CARRERAS-UTPL	125
4.16	REQ10 CREAR NUEVAS ASIGNATURAS-UTPL.....	126
4.17	REQ11 GESTIONAR USUARIOS	127
4.18	REQ12 INCREMENTAR VALORES DE LOS CAMPOS DE REGISTRO	

DE PLANES.....	128
3.2. Requerimientos Técnicos.....	131
B.5. Especificación de Casos de Uso.....	132
1. Autenticación en el Sistema.....	132
1.1. Breve descripción	133
1.2. Especificación de Caso de Uso.....	133
2. Registro de Planes Académicos Externos.....	135
2.1. Breve Descripción.....	136
2.2. Especificación de Caso de Uso.....	136
3. Analizar Plan Académico.....	139
3.1. Breve Descripción.....	140
3.2. Especificación de Caso de Uso.....	140
4. Generar archivo de texto con resultados.....	144
4.1. Breve descripción	145
4.2. Especificación de caso de uso.....	145
5. Generar archivo XML de contenidos importantes.....	147
5.1. Breve descripción	148
5.2. Especificación de caso de uso.....	148
6. Generar gráfico de contenidos importantes	150
6.1. Breve descripción	151
6.2. Especificación de Caso de Uso.....	151
7. Crear nuevo corpus y lematización.....	152
7.1. Breve descripción	153
7.2. Especificación de Caso de Uso.....	153
8. Actualizar lematización	155
8.1. Breve descripción	156
8.2. Especificación de Caso de Uso.....	156
9. Crear carreras UTPL.....	158
9.1. Breve descripción	159
9.2. Especificación de Casos de Uso	159
10. Crear asignaturas UTPL.....	161
10.1. Breve descripción.....	162
10.2. Especificación de Caso de Uso	162

11.	Gestionar Usuarios – Crear nuevo usuario	164
11.1.	Breve descripción.....	165
11.2.	Especificación de Caso de Uso	165
12.	Crear asignaturas de UTPL.....	167
12.1.	Breve descripción.....	168
12.2.	Especificación de Caso de Uso	168
B.6.	Arquitectura de Software	169
1.	Introducción	170
1.1.	Propósito	170
1.2.	Alcance	170
1.3.	Definiciones, acrónimos y abreviaturas.....	170
2.	Vista Lógica	171
2.1.	Capa de Presentación	171
2.2.	Capa de Lógica de Negocios	173
2.3.	Capa de Datos	173
3.	Vista de Despliegue.....	174
3.1.	Servidor UTPL – Base de Datos	175
3.2.	Computador de usuario	175
4.	Calidad	175
4.1.	Usabilidad	175
4.2.	Seguridad	175
4.3.	Confiabilidad	176
4.4.	Portabilidad.....	176
4.5.	Mantenibilidad	176
Anexo C.....		177
Manuales.....		177
C.1.	Manual de Usuario	178
1.	Introducción	178
2.	Secretaria	179
2.1.	Inicio	179
2.2.	Registro.....	180
3.	Fiscal.....	183
3.1.	Inicio	183

3.2.	Análisis Específico	184
3.3.	Análisis completo	192
4.	Administrador.....	193
4.1.	Inicio	194
4.2.	Diccionario.....	194
5.	Usuarios	197
6.	Administración de Campos.....	198
C.2.	Manual del Programador	199
1.	Introducción	199
2.	Diagrama de Base de Datos	200
3.	Diagrama de Clases	204
4.	Configuración e Implementación.....	206
4.1.	Base de Datos.....	206
4.2.	Edición de archivos.....	207
5	Referencias Bibliográficas.....	209

Índice de Figuras

Figura. 1-1 Áreas de la ciencia relacionadas con PLN	24
Figura. 1-2 Principales áreas de la ciencia relacionadas con PLN	25
Figura. 1-3 Niveles de Análisis del Lenguaje Natural	25
Figura. 1-4 Regla del Artículo y Sustantivo.....	26
Figura. 1-5 Ambigüedad presente en el Nivel Semántico	27
Figura. 1-6 Problema de sinonimia.....	27
Figura. 1-7 Problema de interpretación de significado.....	28
Figura. 1-8 Clasificación de los Modelos de PLN.....	29
Figura. 2-1 Proceso de Descubrimiento de Conocimiento.....	36
Figura. 2-2 Proceso de Extracción de Información.....	38
Figura. 2-3 Organización estructurada de la información	40
Figura. 2-4 Problema con el tipo de Archivo	41
Figura. 3-1 Proceso de construcción Preprocesador Planes Académicos	52
Figura. 3-2 Modelos de la herramienta Apache OpenNLP	54
Figura. 3-3 Modelos Empleados de Apache Open NLP	58
Figura. 3-4 Ejemplo - Sentence Detector	58
Figura. 3-5 Ejemplo – Tokenizer.....	59
Figura.3-6 Ejemplo - Part-of-Speech Tagger	61
Figura. 4-1 Diagrama General de Casos de Uso	67
Figura. 4-2 Vista Lógica de la Arquitectura General para el Preprocesador de Planes Académicos.....	68
Figura. 4-3 Diagrama de Base de Datos.....	71
Figura. 4-4 Salida - Fase de Preprocesamiento	72
Figura. 4-5 Esquema del Diccionario - Preprocesador de Planes Académicos	74
Figura. 4-6 Salida - Fase de Comparación	76
Figura. 4-7 Organización de Palabras - Frases – Contenidos	76
Figura. 4-8 Salida - Fase de Comparación	77

Figura. 4-9 Interfaz web de Análisis Específico de Planes Académicos Externos	78
Figura. 4-10 Proceso para evaluar un sistema de extracción de información – MUC ..	79
Figura. 4-11 Proceso para evaluar un sistema de extracción de información – MUC...	84
Figura. 4-12 Cobertura de Extracción del Sistema frente a Fundamentos de Programación UTPL – ECTS	84
Figura. 4-13 Precisión de Extracción del Sistema frente a Fundamentos de Base de Datos UTPL – ECTS	85
Figura. 4-14 Precisión de Extracción del Sistema frente a Fundamentos de Programación UTPL – ECTS	86
Figura. B-1 Proceso Actual de Convalidación de asignaturas en UTPL	98
Figura. B-2 Propuesta para Proceso de Convalidación de asignaturas en UTPL	99
Figura. B-3 Arquitectura del Sistema Semi-automático de convalidaciones de Planes Académicos.....	118
Figura. C-1 Autenticación en el Sistema.....	179
Figura. C-2 Pantalla de Inicio – Perfil Secretaria	180
Figura. C-3 Buscar archivo digital del Plan Académico Externo.....	180
Figura. C-4 Buscar archivo en el sistema	181
Figura. C-5 Almacenar Plan Académico Externo en el servidor.....	181
Figura. C-6 Formulario de Registro de Planes Académicos Externos	182
Figura. C-7 Agregar contenidos al Plan Académico Externo	182
Figura. C-8 Alerta – Almacenamiento correcto del plan	183
Figura. C-9 Pantalla de Inicio – Perfil Fiscal	184
Figura. C-10 Buscar Plan Académico Externo – Código de Plan	184
Figura. C-11 Resultado de Búsqueda.....	185
Figura. C-12 Búsqueda avanzada de Planes Académicos Externos	185
Figura. C-13 Resultado de Búsqueda Avanzada.....	186
Figura. C-14 Selección de Plan Académico Interno	187
Figura. C-15 Analizar Plan Académico – Específico	187
Figura. C-16 Resultado de Análisis -Específico	188
Figura. C-17 Alerta – Almacenamiento correcto de contenidos importantes.....	189

Figura. C-18 Archivos de Descarga – Análisis Específico.....	190
Figura. C-19 Descargar archivo de texto de análisis específico.....	190
Figura. C-20 Descargar archivo XML de análisis específico	191
Figura. C-21 Gráfico de Análisis	192
Figura. C-22 Resultado de Análisis Completo	193
Figura. C-23 Descargar archivo XML de análisis completo	193
Figura. C-24 Pantalla de Inicio – Perfil Administrador	194
Figura. C-25 Funcionalidades del Módulo Diccionario	195
Figura. C-26 Búsqueda de Diccionario	195
Figura. C-27 Agregar nuevas carreras y asignaturas de UTPL.....	196
Figura. C-28 Administración de Usuarios	197
Figura. C-29 Administración de Campos	198
Figura. C-30 Diseño de Base de Datos <i>preprocesadorutpl</i> – Manual del Programador	200
Figura. C-31 Diseño de Base de Datos <i>plandocente</i> – Manual del Programador.....	203
Figura. C-32 Diagrama de Clases.....	205
Figura. C-33 Creación de Base de Datos por consola	206
Figura. C-34 Creación de tablas y relaciones.....	207

Índice de Tablas

Tabla 1-1 Analizador morfológico de Rank Xerox	26
Tabla 1-2 Características principales de los Niveles de Análisis de PLN	28
Tabla 1-3 Comparación de los modelos de PLN	32
Tabla 1-4 Problemas del Lenguaje Natural.....	33
Tabla 2-1 Herramientas para Extracción de Información	39
Tabla 2-2 Proyectos relacionados con Extracción de Información	41
Tabla 2-3 Categorización de Proyectos basados en Extracción de Información.	45
Tabla 3-1 Características Apache OpenNLP.....	54
Tabla 3-2 Tipos de Tokenización Apache OpenNLP.....	55
Tabla 4-1 Atributos para evaluación de un sistema de procesamiento de lenguaje natural	79
Tabla 4-2 Métricas de Evaluación	80
Tabla 4-3 Datos de Prueba	81
Tabla A-1 Etiquetas Eagles	93
Tabla B-1 Definición del Problema.....	108
Tabla B-2 Posicionamiento del Producto	109
Tabla B-3 Resumen de afectados	110
Tabla B-4 Resumen de usuarios.....	111
Tabla B-5 Necesidades de las secretarías	111
Tabla B-6 Necesidades de los fiscales	112
Tabla B-7 Perspectiva del producto.....	113
Tabla B-8 Requerimientos Técnicos.....	131
Tabla C-1 Características Técnicas.....	204

RESUMEN EJECUTIVO

La Universidad Técnica Particular de Loja, es una institución educativa de nivel superior del Ecuador; dentro de los procesos de gestión académica se encuentra la convalidación de asignaturas para los estudiantes que han cursado programas de estudio en otras instituciones educativas.

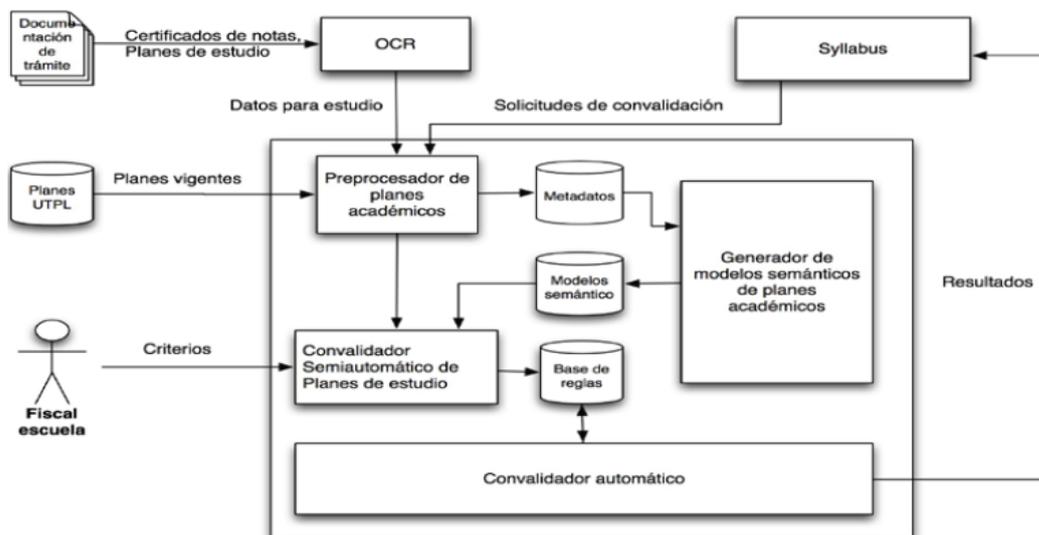
En la actualidad este proceso es realizado de forma manual por docentes en calidad de fiscales de las diferentes titulaciones de la UTPL; ante esta situación, en este proyecto de tesis se realiza la construcción de un Preprocesador de Planes Académicos el cual forma parte del Sistema Semi-automático de convalidaciones de planes académicos, dicho preprocesador es un sistema web semiautomático que realiza la obtención de contenidos importantes de los planes académicos externos comparados con los planes académicos internos. Esta comparación es realizada en base al análisis morfológico de las palabras, a un corpus y lematizaciones obtenidas de los planes académicos internos. Los contenidos extraídos por este sistema serán utilizados por el resto de componentes del Semi-automático de convalidaciones de planes académicos para determinar si una asignatura puede o no ser convalidada con alguna de las asignaturas ofertadas la UTPL.

PROBLEMÁTICA

La Universidad Técnica Particular de Loja, dentro de los procesos de gestión académica que lleva a cabo se encuentra la convalidación de asignaturas para los estudiantes que han cursado programas académicos en otras instituciones educativas.

En la actualidad el proceso de convalidación de asignaturas es realizado de forma manual por docentes de la UTPL, en calidad de fiscales, los cuales verifican que los planes académicos externos cumplen los requerimientos necesarios para ejecutar el proceso anteriormente mencionado.

Para mejorar el desarrollo de este proceso se solicita la creación de un Sistema Semi-automático de convalidaciones de planes académicos, dicho sistema se encuentra dividido en varios componentes uno de ellos el Preprocesador de Planes Académicos, tal como se muestra el diagrama a continuación.



El Preprocesador de Planes Académicos tiene como objetivo realizar la extracción de los contenidos más importantes del plan académico externo (otras entidades educativas) en base a comparaciones con un plan académico interno (UTPL), así como también realizar una comparación general con los contenidos de todos los planes académicos que se encuentran en la UTPL. Esta comparación deberá ser realizada en base al análisis de las palabras que conforman los contenidos del plan académico.

La información obtenida en este componente será utilizada por el resto de componentes, para que finalmente se pueda determinar si la asignatura analizada puede o no ser convalidada con alguna asignatura de la oferta académica que posee la UTPL.

OBJETIVOS

General

- Desarrollar un sistema web semiautomático, que permita extraer los contenidos más importantes de un plan académico

Específicos

- Implementar las funcionalidades de: registro de planes académicos externos, análisis y extracción de contenidos importantes de planes académicos externos en base a un diccionario y representación estructurada de los contenidos obtenidos por el sistema.
- Desarrollar un diccionario compuesto por un corpus y lematizaciones, basado en las asignaturas de la UTPL.
- Construir un módulo de administración de usuarios, formulario de registro y diccionario.

Sección I:

ESTADO DEL ARTE DEL PROCESAMIENTO DE LENGUAJE NATURAL

Capítulo 1

Procesamiento de Lenguaje Natural

1.1 Introducción

En primera parte de esta tesis se comienza con el desarrollo del estado del arte referente al procesamiento de lenguaje natural y la extracción de información, describiendo el marco teórico mediante el cual se va a desarrollar el presente proyecto de tesis.

En este capítulo se realiza el estudio del Lenguaje Natural partiendo del análisis de las diferentes definiciones de esta ciencia, así mismo la relación con las diferentes áreas del conocimiento, de la misma manera la descripción y ejemplificación de los niveles y modelos en los cuales se encuentra estructurado el lenguaje natural, comparación de niveles y modelos entre sí, para finalmente analizar los problemas a los que está expuesto el Lenguaje Natural.

Mediante este estudio se busca tener el marco teórico fundamental que sirva de base para desarrollar el Preprocesador de Planes Académicos de la UTPL

1.2 Definición

El procesamiento de Lenguaje Natural, en inglés (Natural Language Processing), tiene sus inicios después de la segunda guerra mundial, a comienzos de la Guerra Fría (Locke, W.N.; Booth, A.D, 1955) en la década de 1950, iniciándose como una intersección de la lingüística formal y las ciencias de la computación.

En el transcurso del tiempo se ha realizado varias formulaciones para la conceptualización de este término:

"El Procesamiento de Lenguaje Natural (PLN) es el uso de computadoras para entender lenguajes (naturales) humanos tales como español, inglés, francés o japonés. Por entender no se quiere decir que el computador tenga pensamientos, sentimientos y conocimientos humanizados, sino que el computador pueda reconocer y usar información expresada en lenguaje humano" según lo señala Covington. (Covington, Michael, 1994)

Algunos investigadores de esta ciencia mencionan que “el Procesamiento del Lenguaje Natural consiste en el estudio y análisis de los aspectos lingüísticos de un texto a través de programas informáticos” (Leiva, Isidoro; Rodríguez, José, 1996).

Finalmente otros autores definen que un sistema de PLN es aquel que “encapsula un modelo del lenguaje natural en algoritmos apropiados y eficientes, en donde las técnicas de modelado están ampliamente relacionadas con conocimientos de muchos otros campos” (Manaris, Bill; Slator, Brian, 1996)

En base a estas definiciones, de forma personal se puede definir al Procesamiento de Lenguaje Natural como una ciencia que se encarga del reconocimiento, análisis y extracción de información de diferentes fuentes y orígenes de datos definidos en un lenguaje humano, a través del uso de herramientas y sistemas informáticos.

1.3 Áreas de la ciencia relacionadas

Dentro del estudio del procesamiento de lenguaje natural intervienen diferentes áreas del conocimiento como se muestra en la Figura.1-1, estas comprenden: lingüística, ingeniería informática, filosofía, matemática y psicología (Sosa, Eduardo, 1997), las cuales aportan en base a sus especialidades en el desarrollo de este procesamiento.



Figura. 1-1 Áreas de la ciencia relacionadas con PLN

En la Figura. 1-2, se muestran las áreas de la ciencia conjuntamente con sus

aportes, que han ayudado al desarrollo del procesamiento de lenguaje natural (Manaris, Bill; Slator, Brian, 1996).

Ingeniería Informática	<ul style="list-style-type: none">• Brinda métodos para representar modelos, diseño e implementación de algoritmos.
Lingüística	<ul style="list-style-type: none">• Presenta modelos y procesos lingüísticos
Matemática	<ul style="list-style-type: none">• Muestra modelos y métodos formales

Figura. 1-2 Principales áreas de la ciencia relacionadas con PLN

1.4 Niveles de análisis

Para el mejor entendimiento y organización (Covington, Michael, 1994), el lenguaje natural se ha subdividido en varios niveles de análisis Figura. 1-3, los cuales serán descritos a continuación.

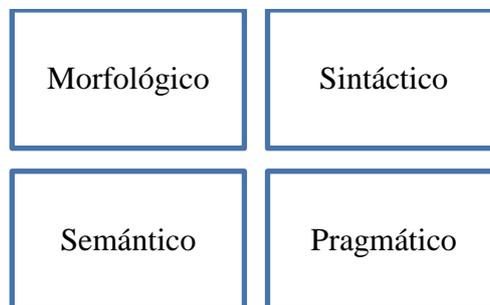


Figura. 1-3 Niveles de Análisis del Lenguaje Natural

1.4.1 Nivel Morfológico

Se refiere a la revisión y detección de la correcta estructura de cada palabra de la instrucción. Además determina la forma, clase o categoría gramatical de cada palabra de una oración (Macario, Hernandez, 2007).

En la Tabla. 1-1, se muestra el analizador morfológico de Rank Xerox (Xerox), en donde se puede observar una lista con todos los posibles valores gramaticales de cada

una de las palabras que aparecen en siguiente instrucción: “Esta nota para mañana” (Leiva, Isidoro; Rodríguez, José, 1996), lo cual es un claro ejemplo del nivel morfológico descrito en este apartado.

Analizador Morfológico	
<Esta>	Pronombre + Demostrativo + Femenino + Singular Determinante + Demostrativo + Femenino + Singular
<nota>	Verbo (notar) + Presente indicativo + 3 ^{ra} persona + Singular Verbo (notar) + Imperativo + 2 ^{da} persona + Singular Sustantivo + Femenino + Singular
<para>	Preposición Verbo (parir) + Presente subjuntivo + 1 ^{ra} persona + Singular Verbo (parir) + Presente subjuntivo + 3 ^{ra} persona + Singular Verbo (parir) + Imperativo + 3 ^{ra} persona + Singular Verbo (parar) + Presente indicativo + 3 ^{ra} persona + Singular Verbo (parar) + Imperativo + 2 ^{da} persona + Singular
<mañana>	Sustantivo + Femenino + Singular Adverbio

Tabla 1-1 Analizador morfológico de Rank Xerox
Tomado de: (Leiva, Isidoro; Rodríguez, José, 1996)

1.4.2 Nivel Sintáctico

Corresponde a la relación que existe entre los diferentes símbolos y signos del lenguaje (Sosa, Eduardo, 1997), así como también la forma en que las palabras se relacionan entre sí (Macario, Hernandez, 2007). En este nivel se analizan todas las combinaciones gramaticalmente aceptadas o las que no lo son.

En la Figura. 1-4, se muestra un ejemplo de las reglas correctas como las incorrectas en cuanto al análisis de la relación de los componentes de una estructura gramatical referente al uso del artículo y del sustantivo.

Artículo (Singular) + Sustantivo (Singular)	=	Correcto
Artículo (Plural) + Sustantivo (Plural)	=	Correcto
Artículo (Singular) + Sustantivo (Plural)	=	Incorrecto
Artículo (Plural) + Sustantivo (Singular)	=	Incorrecto

Figura. 1-4 Regla del Artículo y Sustantivo
Tomado de: (Leiva, Isidoro; Rodríguez, José, 1996)

1.4.3 Nivel Semántico

Este nivel es aquel que estudia el procesamiento y detección del significado que cada elemento tiene por sí mismo dentro de la oración. En este nivel puede producirse ambigüedad debido a que una palabra puede tener uno o varios significados, tal como se muestra en la Figura. 1-5, en donde una palabra posee diferentes significados, lo cual altera el sentido de la oración.

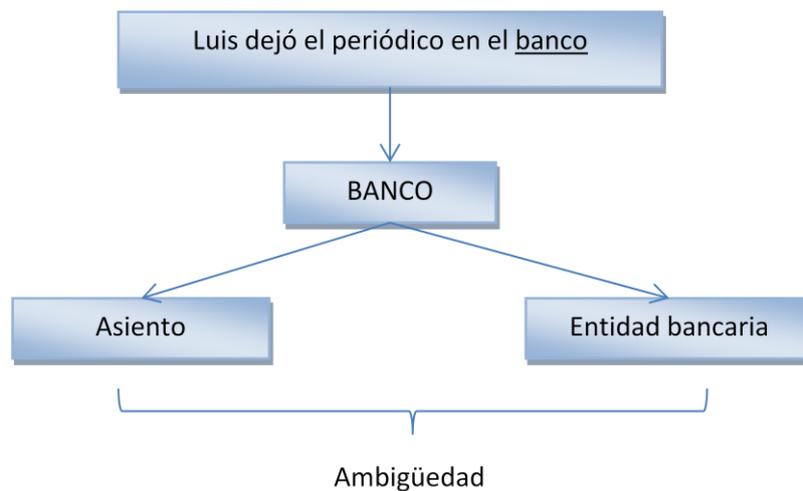


Figura. 1-5 Ambigüedad presente en el Nivel Semántico

En este nivel también se debe tomar en cuenta la variación léxica, la que se refiere a la posibilidad de utilizar distintos términos que muestran un mismo significado, conocido también como sinonimia (Vallez, Mari; Pedraza, Rafael, 2007). Este problema se lo puede observar en la Figura. 1-6, en donde a pesar de que la estructura de la palabra es diferente estas pertenecen a un mismo significado.

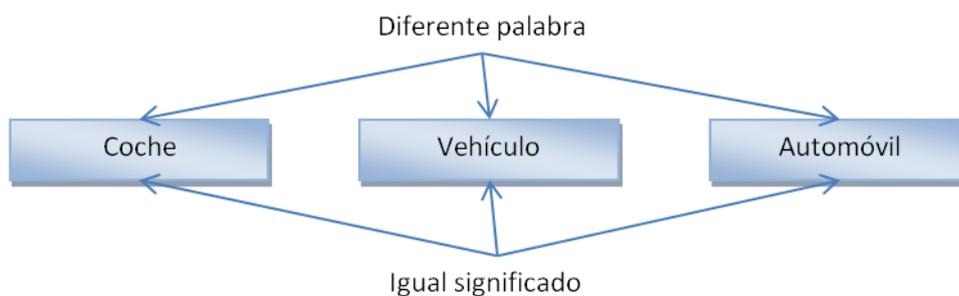


Figura. 1-6 Problema de sinonimia

1.4.4 Nivel Pragmático

En este nivel se interpreta la estructura de la oración para determinar el verdadero significado dentro del contexto específico. En algunas ocasiones el significado de una palabra debe de tomárselo en sentido figurado, un claro ejemplo se muestra en la Figura. 1-7.

El se moría de la risa

Figura. 1-7 Problema de interpretación de significado

Resumiendo en la Tabla. 1.2, se muestran las principales características de los niveles estudiados anteriormente, el contenido de esta tabla está orientada indicar de forma breve el tipo de acciones que realiza y a como realiza estas acciones de cada uno de los niveles antes descritos. (Contreras, Hilda, 2001)

Nivel	Características	
	Qué acciones realiza	Cómo realiza las acciones
Fonológico	Sonidos Hablados	Formar morfemas ¹
Sintáctico	Roles estructurales de palabras	Formar oraciones
Semántico	Significado independiente del contexto	Derivar significado de oraciones
Pragmático	Significado dependiente del contexto	Derivar significado de oraciones relativo al discurso circundante

Tabla 1-2 Características principales de los Niveles de Análisis de PLN

1.5 Modelos de Procesamiento

Dentro del procesamiento de lenguaje natural existen varios modelos, los cuales se han subdividido en base a sus características y similitudes. En la Figura. 1-8, se observa la clasificación de estos modelos.

¹ Morfema: Es el fragmento mínimo de una palabra capaz de por sí mismo expresar significado.



Figura. 1-8 Clasificación de los Modelos de PLN

1.5.1 Modelos Simbólicos

Los modelos simbólicos son objetos matemáticos definidos a partir de un conjunto de expresiones o axiomas y reglas (reglas de derivación). Estas reglas transforman los axiomas en expresiones nuevas, formando de esta manera los teoremas. Además, a partir de estas reglas de derivación, el reconocimiento de oraciones gramaticales se entiende como un tipo de deducción de teoremas (Sandoval, Antonio Moreno, 1998)

Un ejemplo de utilización de símbolos son las computadoras, a través de los programas, ya que estos prueban teoremas lógicos, realizan inferencias, ejecutan juegos como el ajedrez, emiten consejos y traducen textos entre otras actividades.

1.5.1.1 Características

- Se utiliza una metalengua² para hablar de la realidad que consiste en un conjunto de especificaciones para escribir gramáticas (Sandoval, Antonio Moreno, 1998)
- Los símbolos son utilizados como representación del conocimiento, debido a que el pensamiento y el conocimiento racional se basan en la manipulación y ubicación lógica de las ideas pensadas como símbolos (Sandoval, Antonio Moreno, 1998)

² Metalengua, sirve para fijar un conjunto de convenciones que beneficia la comunicación de ideas entre los miembros de la comunidad científica (Sandoval, Antonio Moreno, 1998)

- Autonomía de la sintaxis frente a la semántica, en donde un sistema formal consta de una sintaxis en donde se determinan las condiciones necesarias para la buena formación de una estructura gramatical y una semántica en donde se interpreta el significado de las expresiones de correcta estructura (Sandoval, Antonio Moreno, 1998)

1.5.2 Modelos Empíricos o Estadísticos

El objetivo principal de estos modelos es deducir conocimiento directamente de los datos, buscando de esta forma regularidades significativas de entre los datos.

En este modelo se aplica la estrategia de contar con la mayor cantidad posible de datos para poder establecer una probabilidad lo más cercana posible a la frecuencia relativa³ estable.

Los modelos estadísticos o empíricos, también llamados métodos cuantitativos, proporcionan solución al problema de ambigüedad que poseen los modelos simbólicos, debido a que cuando una oración presenta varias estructuras o interpretaciones posibles para escoger, se elige la más probable en función de las probabilidades de cada opción (Contreras, Hilda, 2001)

1.5.2.1 Características

- Emplea colecciones de muestras de palabras, a lo que se denomina corpus.
- Conteo y análisis de la cantidad y calidad de información que se posee, debido a que a mayor cantidad de información, se podrá obtener un mejor análisis de la misma.

³ Frecuencia Relativa, es el cociente entre la frecuencia absoluta acumulada (es el número de veces que aparece en la muestra el valor de una variable) y el tamaño de la muestra (Álvarez, Juan, 2007)

- Utiliza probabilidades para elegir la interpretación más correcta en base a un conjunto de datos.

1.5.3 Modelos Conexionistas

Los modelos conexionistas están relacionados con los procesos que se llevan a cabo en el cerebro humano, por lo cual estos modelos se dividen en: Modelos inspirados en el cerebro (conexionismo – redes neuronales) y los Modelos inspirados en la vida y la evolución (computación evolutiva – algoritmo genético)

En este modelo existen estructuras formadas por nodos y conexiones en forma de una red, sobre esta red se realiza un reconocimiento de semejanzas⁴ en los patrones de nodos (fonemas, morfemas, oraciones).

1.5.3.1 Características

- Su estructura está compuesta por nodos, conexiones, formando un esquema de red nodal.
- En las redes neuronales el procesamiento se realiza en paralelo, ya que los procesos de reconocimiento de palabras, análisis morfosintáctico y semántica se ejecutan de forma simultánea.

1.5.4 Comparación de los modelos

Una vez analizados cada uno de los modelos de procesamiento de lenguaje natural, se ha creído conveniente realizar una tabla resumida (Tabla. 1-3) con los datos comparativos de las principales características, técnicas, ventajas, desventajas de los tres modelos anteriormente descritos (Contreras, Hilda, 2001).

⁴ Reconocimiento de semejanzas, dos estructuras son similares si excitan los mismos nodos.

	Simbólico	Empírico	Conexionista
Teorías	Teoría de Conjuntos y Lógica Matemática.	Teorías Estadísticas.	Conexionismo, Inteligencia Artificial.
Idea Principal	Conocimiento = Competencia	Conocimiento = Actuación	Conocimiento = Actuación y Conocimiento
Definición	Sistemas formales axiomáticos, formados por reglas y símbolos.	Sistemas que utilizan las probabilidades de ocurrencia de los datos lo cual determina el resultado.	Sistemas que simulan la los procesos llevados en el cerebro humano.
Técnicas	Gramáticas: generativas, de estados finitos, independientes del contexto, Unificación y Rasgos.	N-gramas, Cadenas de Markov, Árbol de Decisión, Gramáticas probabilísticas.	Algoritmos genéticos, Redes Neuronales.
Ventajas	Se encuentran definidos formalmente, facilitando la evaluación de hipótesis.	Ayudan la resolución de problemas de ambigüedad, mediante la probabilidad.	Proporcionan una adaptación directa a nuevos dominios, ejecución en paralelo.
Desventajas	Presencia de ambigüedad.	No se aplica conocimiento sobre el lenguaje	Poco desarrollo conceptual
Aplicación	Sistemas de comprensión, correctores ortográficos, correctores sintácticos, etc.	Reconocimiento de caracteres, etiquetadores, reconocimiento del habla, traducción automática desambiguación léxica y sintáctica, entre otros.	Reconocimiento del habla, simulación y reconstrucción de lenguas, planificación lingüística, etc.

Tabla 1-3 Comparación de los modelos de PLN

1.6 Problemas del Lenguaje Natural

El procesamiento de Lenguaje Natural se ve enfrentado a una serie de problemas, los cuales es necesario conocerlos para poderlos enfrentar (Manaris, Bill; Slator, Brian, 1996). En la Tabla.1-4 se describen los más importantes.

Problemas	Descripción
Inexactitud	Compuesto por: <ul style="list-style-type: none">• Errores ortográficos.• Signos de puntuación incorrectos• Palabras transpuestas.• Oraciones agramaticales⁵
Imprecisión	Corresponde a: <ul style="list-style-type: none">• Términos relativos sin puntos de referencia.• Utilización de términos cualitativos.
Ambigüedad	Diferentes interpretaciones para una sola palabra. La variedad depende del tipo de lingüística a la que pertenezca.
Lingüística	Contempla los diferentes idiomas existentes.

Tabla 1-4 Problemas del Lenguaje Natural

⁵ Oraciones Agramaticales: Oraciones que por su estructura no se ajustan a las reglas de la gramática.

Capítulo 2

Extracción de Información

2.1 Introducción

En este segundo capítulo se finaliza con el estado del arte, mediante la revisión sobre la extracción de información, partiendo del análisis de las diferentes definiciones a este proceso, fases del proceso, descripción de las herramientas de software más utilizadas para la extracción de información, así mismo análisis de los problemas relacionados con este proceso para finalmente concluir con el estudio de los proyectos desarrollados correspondientes a este proceso y que pueden servir de guía en la elaboración del presente proyecto de tesis.

La extracción de información es uno de los principales procesos dentro del Procesamiento de Lenguaje Natural, en el cual se obtienen entidades, eventos y relaciones existentes entre los elementos de un texto o un conjunto de textos, de ahí la importancia de realizar el estudio de este tema, para adquirir de esta manera las bases teóricas necesarias para la construcción del Preprocesador de Planes Académicos de la UTPL.

2.2 Definición

En vista de la necesidad de obtener información de forma rápida se emplea dos técnicas fundamentales como son la recuperación y extracción de información.

La recuperación de la información hace referencia al proceso de obtener documentos ante una solicitud de búsqueda específica. Esta recuperación se basa en el tratamiento del documento como conjunto y no por partes. Las técnicas utilizadas para estos procesos están centradas fundamentalmente en sistemas de indexación⁶ las cuales permiten clasificar los documentos y relacionarlos entre sí.

El proceso de extracción de información se define en cambio como el proceso de obtención de entidades, eventos y relaciones existentes entre los elementos de un texto o un conjunto de textos. (Vallez, Mari; Pedraza, Rafael, 2007). Este proceso implica

⁶ Los sistemas de indexación se realizan de forma manual, automática o semiautomática

separar un bloque de texto de su contexto y luego realizar el su procesamiento.

A diferencia del proceso de recuperación de la información, el proceso de extracción de la información requiere del procesamiento de lenguaje natural, debido a que la extracción realizada debe tener un sentido sintáctico y semántico coherente para el usuario.

2.3 Fases del Proceso de Descubrimiento de Conocimiento

El proceso de extracción de información forma parte de un proceso definido como Descubrimiento de Conocimiento, el cual está compuesto por las fases de: selección, pre-procesamiento, transformación, minería de datos y evaluación o interpretación (Fayyad, Usama, 1996). En la Figura. 2-1 se puede observar las diferentes fases con los procesos realizados en cada una de ellas. Posteriormente se analizará cada una de las fases.

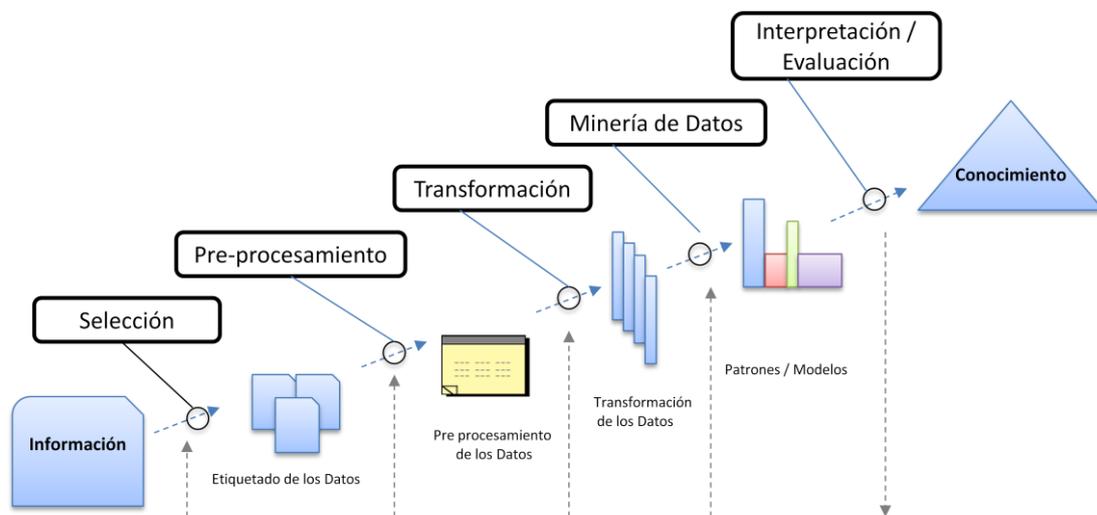


Figura. 2-1 Proceso de Descubrimiento de Conocimiento
Tomado de: (Fayyad, Usama, 1996)

2.3.1 Fase de Selección

En esta fase se seleccionan de algún medio de almacenamiento (repositorios, base de datos, etc.) los contenidos que van a ser parte del proceso de descubrimiento de conocimiento. Una vez elegidos los contenidos se procede a la realización del

reconocimiento y etiquetado de las entidades que son el objetivo dentro del proceso de descubrimiento de conocimiento.

2.3.2 Fase de Pre-procesamiento

En esta fase se realiza el pre-procesamiento de la información, la cual se compone por la ejecución de los niveles descritos en el capítulo uno, en la sección 1.4 Niveles de Análisis. La ejecución de estos niveles puede ayudarse a través de diversas herramientas de software, las cuales se analizarán posteriormente.

2.3.3 Fase de Transformación

Esta fase tiene como objetivo, convertir la información pre-procesada en un formato que sea capaz de ser utilizado para la fase de minería de datos.

2.3.4 Fase de Minería de Datos

En la Fase de Minería de Datos, se realiza el análisis a profundidad de los datos obtenidos en las fases anteriormente estudiadas, con la finalidad de poder identificar información válida, novedosa, útil y entendible que se constituya en patrones o modelos que identifiquen el contenido analizado.

2.3.5 Fase de Interpretación o Evaluación

Una vez obtenidos los patrones o modelos, se debe realizar una interpretación del conocimiento obtenido, así mismo realizar una evaluación del nivel de validez que tiene la información y patrones conseguidos durante el proceso de descubrimiento de conocimiento.

Dentro de la fase de Selección y Pre-procesamiento descrita en la Figura. 9, se realiza la Extracción de Información, la cual posee las siguientes fases: etiquetado, nombrado de la entidad reconocida y análisis sintáctico profundo, las cuales se muestran en la Figura. 2-2, con lo cual se consigue pasar de un texto sin estructura a un texto de

tipo estructurado.

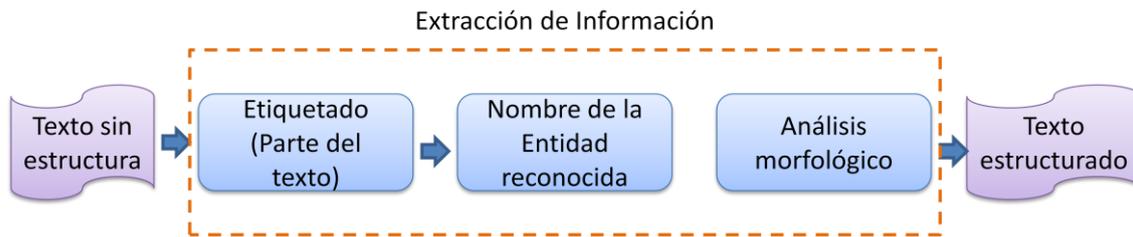


Figura. 2-2 Proceso de Extracción de Información.

- **Etiquetado (Parte del texto).**- En esta fase se procede al reconocimiento de entidades a través del etiquetado de la misma, dentro del texto sin estructura que es parte del proceso de extracción de información.
- **Nombre de la entidad reconocida.**- Aquellas entidades que fueron reconocidas a través del etiquetado en la fase anterior, en esta fase se procede a la asignación de un nombre representativo a la etiqueta que sea correspondiente a la entidad reconocida.
- **Análisis morfológico.**- Como se lo estudió en el capítulo uno en la sección 1.4.1 Nivel Morfológico, este análisis hace referencia al análisis del texto para determinar en base a la estructura de la palabra; la forma, clase o categoría gramatical a la cual pertenece cada palabra de una oración.

2.4 Herramientas

Existen múltiples herramientas de software para ayudar a la realización del proceso de extracción de información. En la Tabla. 2.1, se muestran algunas de las más utilizadas.

Características Herramienta	Características Generales	Atributos de extracción	Idiomas que soporta
Gate (General Architecture for Text	Realiza análisis sintáctico, morfológico,	Palabras etiquetadas según su naturaleza, ordenadas según su	Inglés, alemán, romano e indio.

Engineering)	etiquetado; además posee herramientas de recuperación de información, extracción de información para diferentes idiomas, etc.	aparición en la frase.	
Calais	Realiza análisis sintáctico y morfológico. El producto de la extracción es entregada en formato RDF	Entidades (personas, lugares, productos, etc.), hechos (José trabaja en UTPL) y eventos (Juan fue ascendido como docente)	Inglés, francés, alemán y español.
Apache OpenNLP	Realiza análisis sintáctico y morfológico, en donde realiza: segmentación de la oración, etiquetado de texto, extracción del nombre de las entidades, fragmentación, análisis y resolución de correferencia	Cualquier palabra que se encuentre dentro del idioma en el cual se encuentre configuradas las etiquetas de comparación	Inglés, español, danés, alemán, holandés, portugués
Apache Lucene	Realiza análisis de textos, indexación y búsquedas de textos completos. Es independiente del tipo de archivo a analizar (pdf, html, doc, docx)	Tokens, nombres, organizaciones, Places, Personas.	Inglés, español, italiano, francés, chino, ruso y japonés.
LingPipe	Permite realizar la tokenización de los textos y realizar la extracción de oraciones. Ayuda a verificar la ortografía de las consultas realizadas.	Permite realizar la búsqueda de nombres de personas, organizaciones y ubicaciones.	Todos los lenguajes humanos y algunos no humanos.

Tabla 2-1 Herramientas para Extracción de Información

2.5 Problemas

La complejidad del lenguaje natural puede hacer que el acceso a la información contenida en los textos, la forma y tamaño de los documentos haga más difícil esta tarea, además la información posee diferentes formas entendiéndose por forma a la estructura en la cual se encuentren los datos, donde existe una organización regular y predecible de las entidades y relaciones.

Por ejemplo: Existe una relación entre lugares y empresas, con lo cual una empresa tiene la capacidad de identificar los lugares donde están las empresas, así como también dado un lugar poder conocer que empresas se encuentran en determinado sitio.

Para poder obtener la información de una manera más fácil es necesario convertir los datos no estructurados en datos estructurados, una vez logrado esto es necesario utilizar herramientas de consulta cómo por ejemplo SQL⁷, tal como se muestra en la Figura. 2-3 Organización estructurada de la información

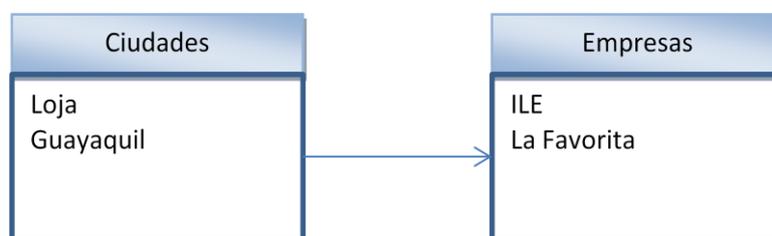


Figura. 2-3 Organización estructurada de la información

Otro de los problemas para realizar la extracción de información, es que el tipo de contenido no se relaciona con el tipo de archivo que se encuentra la misma, por ejemplo un archivo de texto, puede encontrarse en un formato de imagen (.jpg, .gif), en vez de encontrarse en un propio tipo de archivo (.doc, .docx, .txt). La Figura 2-4, ejemplifica este problema, que ocasiona dificultades al momento de extraer el contenido de los archivos con estas características.

⁷ SQL, Structured Query Language o Lenguaje Estructurado de Consulta. Es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones a realizar en estas bases de datos.

Archivo en formato .jpg	Archivo en formato .docx
-------------------------	--------------------------

Figura. 2-4 Problema con el tipo de Archivo

En cuanto a las herramientas de software disponibles para la extracción de información existen varias herramientas que ayudan a realizar este procedimiento, tal como se indica en el apartado 2.4 de este capítulo. Pero a pesar de esto existe dificultad por el idioma para el cual están construidos los analizadores ya que estos son dependientes del idioma para el que fueron programados el cual debe ser el mismo para el contenido que es objeto de la extracción. La mayoría de las herramientas se encuentran adecuadas para el idioma inglés, portugués y alemán, pero existen pocos analizadores desarrollados para el idioma español.

2.6 Proyectos relacionados

Con las necesidades crecientes de la sociedad actual en esta área de la ciencia, se han realizado varias investigaciones y se continúa investigado como el Procesamiento de Lenguaje Natural y la Extracción de Información pueden mejorar las tareas de las personas. En la Tabla. 2-2, se muestran de manera general algunos proyectos relacionados con esta área de la ciencia.

Proyectos relacionados con la Extracción de Información
<ul style="list-style-type: none"> • Inteligencia de Negocios • Análisis de Medios de comunicación. • Clasificación de Documentos • Detección de emociones. • Búsqueda de Patentes. • Resúmenes Automáticos • Análisis de correo electrónico. • Minería de Datos • Extracción de información de literatura científica.

Tabla 2-2 Proyectos relacionados con Extracción de Información
Tomado de: Realizado por la autora

En vista de la gran cantidad de proyectos en los que la Extracción de Información puede ser utilizada, (Sarawagi, 2007) propone una categorización dependiendo de la

orientación que tenga el proyecto, la cual se describe en la Tabla 2-3. Categorización de Proyectos basados en Extracción de Información.

Categoría	Proyecto	Descripción
Aplicaciones Empresariales	Seguimiento de Noticias	Aplicaciones construidas para el seguimiento automático de información en diferentes fuentes de noticias, basándose en la extracción de entidades y relaciones. Como ejemplo de este tipo de proyectos de extracción de información sobre artículos de noticias tenemos la creación automática de noticias multimedia mediante la integración de video y fotografía de entidades y eventos recopilados. ⁸
	Atención al Cliente	Para que la recopilación de información acerca del cliente, sea efectiva y brinde los resultados esperados es necesario que esta se encuentre en bases de datos estructuradas y ontologías empresariales (Sarawagi, 2007). Uno de los proyectos relacionados con esta área se encuentra la extracción de los estados de ánimo de los clientes mediante conversaciones telefónicas.

⁸ Más información (Gupta, 2008): <http://vangelist.wordpress.com/2008/05/08/reuters-spotlight-apis/>

	Limpieza de Datos	Como anteriormente se mencionaba, para que la información almacenada por una empresa se le pueda extraer el máximo provecho, esta debe de encontrarse en un formato estructurado para la fácil consulta y análisis de los mismos. Un ejemplo de esto son las aplicaciones que realizan anuncios clasificados, la cual muestra información previamente estructurada y analizada.
Gestión de Información Personal	Buscadores de Personas	La Gestión de Información Personal mediante el uso de extracción de información hace referencia al almacenamiento estructurado y relacional de los datos informativos de una persona frente a los documentos y proyectos realizados por esta persona. Un ejemplo son las aplicaciones que de forma automática extraen los nombres de una persona que se encuentre en un recurso digital y estos se indexan a una base de datos estructurada con información correspondiente a la persona. ⁹
Aplicaciones Científicas	Buscadores de entidades biológicas.	La extracción de información es utilizada en diferentes áreas de la ciencia como por ejemplo la bioinformática, la cual ha ayudado a potenciar las investigaciones sobre

⁹ Más información <http://google.org/personfinder/global/home.html>

		<p>extracción de información debido a que las entidades que posee la biología como nombres de genes y proteínas es muy diferente de entidades comunes como nombres de personas, empresas, lugares, etc.</p> <p>Un ejemplo de este tipo de aplicación son aquellas que realizan la búsqueda y extracción de entidades biológicas y generar las interacciones que ocurren entre ellas.¹⁰</p>
<p>Aplicaciones Orientadas a la Web.</p>	<p>Bases de Datos de Citas Bibliográficas</p>	<p>En este tipo de proyectos se encuentra los buscadores de artículos científicos como son: Google Scholar¹¹ y CiteSeer¹², los cuales se basan en la extracción de información a diferentes niveles de profundidad partiendo de datos informativos del datos generales del documento (autor, ciudad, fecha de publicación) hasta la extracción de los contenidos importantes del documento, referencias empleadas, etc. Con esta información se puede obtener conocimiento estadístico de cuantas referencias en libros o artículos posee un autor determinado.</p>

¹⁰ Más información <http://www.ncbi.nlm.nih.gov/>

¹¹ Más información <http://scholar.google.com/>

¹² Más información <http://citeseerx.ist.psu.edu/index>

	Colocación de anuncios en páginas web.	También conocida como publicidad contextual, hace referencia a la colocación de anuncios comerciales en páginas web, a través de la utilización de ad-network ¹³ quien realiza la extracción, selección y optimización de información, brindando un beneficio compartido entre el sitio web y el ad-network. Las principales empresas que trabajan de esta manera son Google, Yahoo! y Microsoft. (Broder, Fontoura, Josifovski, & Riedel, 2011)
--	--	---

Tabla 2-3 Categorización de Proyectos basados en Extracción de Información.

Con la finalidad de conocer a mayor detalle proyectos reales basados en Extracción de Información, a continuación se describen tres proyectos que serán de ayuda para el desarrollo de este proyecto de tesis.

2.6.1 Framework para la extracción de información para páginas web semi - estructuradas

Los extractores de información deben ser usados para obtener y estructurar la información del texto de páginas web gratuitas, tales como blogs y páginas de noticias, así como también páginas que sean resultado de alguna búsqueda específica. Este proyecto se centra en la elaboración de un extractor de información para páginas web semi - estructuradas, las cuales se encuentran clasificadas en páginas web basadas en heurísticas y páginas web basadas en reglas.

En base a varias investigaciones realizadas por (Corchado, Bajo, Golinska, Giroux, & Corchuelo, 2012), los paquetes de la arquitectura para extracción de

¹³ **Ad network** hace referencia a la empresa que conecta a los anunciantes a los sitios web que albergan publicidad en sus entornos.

información de páginas web semi-estructuradas definidas en este libro son: Dataset, Annotator, Learners, Tokeniser, Cross Validator, Utilities.

La construcción del framework realizada por (Corchado, Bajo, Golinska, Giroux, & Corchuelo, 2012) ayuda en gran medida a los ingenieros de software a implementar técnicas y comparar resultados del proceso de extracción de información de páginas web semi – estructuradas, con la finalidad de elegir la mejor técnica dependiendo del ámbito de extracción, ahorrando de esta manera costos de desarrollo e implementación.

2.6.2 Sistema de extracción y representación del conocimiento a partir de documentos descriptivos

En el proyecto de tesis elaborado por (López Yebra, 2012), se construye un sistema no-supervisado, a partir de textos que se encuentren en lenguaje natural los cuales describan concretamente el objeto de análisis, con la finalidad de crear una ontología en lenguaje OM¹⁴. En la realización de este proyecto intervinieron los siguientes componentes fundamentales: Corpus, Preprocesador de Texto, Patrones, Conjunto de Reglas.

En esta tesis desarrollada por (López Yebra, 2012) sus autores han podido determinar que los textos que van a ser analizados a través del uso de técnicas de extracción de lenguaje natural deben de ser descriptivos, para asegurar de esta forma que el lenguaje es claro, estructuras de textos sencillas, etc. Así mismo, el pre procesamiento del corpus, según lo indica los autores de este proyecto de tesis, es importante para poder obtener resultados confiables, “ya que se puede contar con mayor información de la que un texto sin pre procesamiento ofrece”.

Además se señala, que en base a las reglas estudiadas utilizando ontologías, las que menor índice de error presentan son: Hiperonimia, Hiponimia y Sinonimia.

Finalmente este proyecto de tesis (Sistema de extracción y representación del

¹⁴ Lenguaje OM, es un lenguaje para el diseño de ontologías con conocimiento y relaciones, las cuales proporcionan más semántica a las operaciones de búsqueda de conocimiento.

conocimiento a partir de documentos descriptivos) tuvo como resultado la búsqueda y extracción de conceptos, identificación de taxonomías, búsqueda y obtención de sinónimos, así como también la relación entre los conceptos encontrados, para culminar de forma exitosa con la representación en el lenguaje de ontologías OM, de lo obtenido en los documentos de texto expresados en lenguaje natural.

2.6.3 Ontología Multilingüe para una red de conocimiento (Monnet)

En base a la demanda que presentan las industrias y las instituciones gubernamentales de acceso multilingüe a la información, varios docentes (Gómez, Aguado-de-Cea, Montiel-Ponsoda, & García, 2010) de la Universidad Politécnica de Madrid, desarrollaron un proyecto denominado Monnet que tiene como objetivo “proporcionar una solución basada en semántica para garantizar el acceso a la información más allá de las barreras lingüísticas”

Para lograr gestionar la información de entrada para este proyecto, es necesario realizar el análisis a nivel semántico, para de esta manera obtener una “mejor consulta, integración y presentación de la información sin importar el lenguaje” en el que se encuentre la información.

La arquitectura utilizada en Monnet está compuesta por cuatro módulos que son: Módulo de Localización de Ontologías, módulo de Extracción de Información basada en Ontologías Multilingües, y módulo de Acceso Multilingüe al conocimiento y módulo de Presentación de la información.

Luego de varias investigaciones se logró conseguir que Monnet localice de forma semi – automática modelos y ontologías, además presente un modelo formal para la representación plurilingüe, conjuntamente con una metodología para obtención de información basándose en técnicas de extracción de la información de diferentes lenguas, formatos y fuentes, para finalmente crear las ontologías en formato RDF u OWL con la información obtenida en varios idiomas y colocarla como un recurso de acceso web.

Sección II:

DISEÑO Y CONSTRUCCIÓN DEL PREPROCESADOR DE PLANES ACADÉMICOS

Capítulo 3
Descripción del proceso y selección de herramientas

3.1 Introducción

En el tercer capítulo de este proyecto de tesis, se describe el proceso a seguir para la construcción del Preprocesador de Planes Académicos de la UTPL teniendo en cuenta los requerimientos planteados para este proyecto.

A continuación se definen, en base a la investigación y al análisis realizado, las mejores técnicas para el procesamiento de lenguaje natural y herramientas de software para la extracción de información, que se adaptan al ámbito de este proyecto y que van a ser utilizadas para el desarrollo de esta tesis.

Se prevé la utilización de una metodología para el desarrollo del Preprocesador de Planes Académicos, en este capítulo se contempla la selección de dicha metodología y en el capítulo siguiente se hará uso de la misma.

3.2 Descripción del proceso de construcción

La entrada principal para el Preprocesador de Planes Académicos de UTPL son los planes académicos tanto internos (UTPL) como externos (otras universidades), por lo tanto es de vital importancia para el proceso de construcción del Preprocesador, conocer la estructura de los planes académicos así como también conocer como se lleva a cabo el proceso de convalidación de asignaturas en la UTPL, para de esta forma construir el software que mejor se adapte a los requerimientos solicitados.

3.2.1 Actividades de Construcción

Para diseñar el Preprocesador de Planes Académicos de la UTPL, se contemplan un conjunto de actividades fundamentadas en la sección 2.3. Fases del Proceso de Descubrimiento de Conocimiento, del capítulo 2 de esta tesis, en donde se contempla el proceso a seguir para realizar el descubrimiento del conocimiento; de la misma forma en la sección 1.4 Niveles de Análisis, del capítulo 1, se define la clasificación del análisis del texto de un lenguaje dependiendo de su tipo y profundidad.

Así mismo se debe tomar en cuenta los temas estudiados en el Capítulo 1, sección 1.5 Modelos de Procesamiento, en donde se analizaron tres modelos que ayudan al procesamiento del lenguaje natural. En base al ámbito, requerimientos y necesidades que posee el desarrollo de este proyecto y al estudio de los Modelos de Procesamiento de Lenguaje Natural, se ha determinado que, el Modelo Empírico o Estadístico es el más idóneo para ayudar en la resolución del problema de este proyecto.

El Modelo Empírico o Estadístico es el más adecuado para este proyecto puesto que es necesario la utilización de muestras de palabras importantes para cada asignatura, con estas palabras se pretende analizar mediante comparaciones consecutivas la información del Plan Académico. Esta información seleccionada permitirá determinar los contenidos más relevantes del plan que está siendo analizado.

Los niveles de análisis del lenguaje natural revisados en el capítulo 1 sección 1.4 también serán utilizados en el desarrollo de este proyecto, es así que en base a los requerimientos que posee el proyecto de construcción del Preprocesador de Planes Académicos, el nivel de Análisis Morfológico es el más adecuado debido a que en este proyecto es necesario conocer y verificar la clase o categoría gramatical de cada palabra de una oración, lo cual servirá para clasificar las palabras en base al tipo de palabra y poder elaborar el diccionario de la herramienta, así como también filtrar el análisis de los planes académicos

Finalmente para construir el software de este proyecto es importante definir un proceso que sirva de guía durante la construcción del mismo, es por esta razón que se ha tomado como base el proceso de descubrimiento de conocimiento y el proceso de extracción de información, tratados en el capítulo 2, sección 2.3 de este documento, para la definición del proceso general de construcción del Preprocesador de Planes Académicos. En la Figura 3-1 se muestra el diagrama de proceso para la construcción de dicho preprocesador.

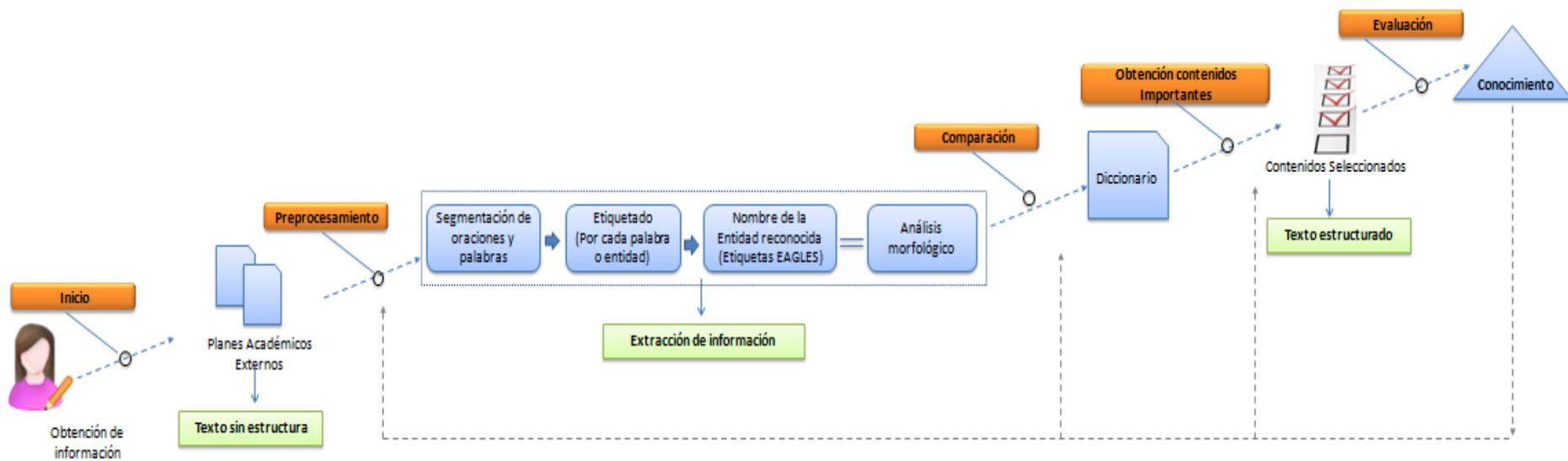


Figura. 3-1 Proceso de construcción Preprocesador Planes Académicos

En el capítulo 4 se detallará las fases y actividades del proceso de construcción del Preprocesador de Planes Académicos que se indican en la Figura 3-1.

3.3 Selección de herramientas

En esta sección se busca seleccionar la mejor herramienta a emplearse durante todo el proceso de desarrollo del Preprocesador de Planes Académicos. Esta selección debe basarse en el análisis y comparación de las diferentes herramientas y descritas en el capítulo 2, lo cual servirá para escoger la más adecuada en base a los requerimientos planteados, para ser utilizadas en la elaboración de este proyecto.

Como se mencionó en la sección 2.4 Herramientas, del capítulo 2, existen múltiples herramientas de software disponibles para ayudar a la realización del proceso de extracción de información. En la Tabla 5, del capítulo anteriormente citado, se realiza la descripción de las utilidades que brindan las herramientas de software más utilizadas para la extracción de información.

Una vez analizado cada una de las herramientas anteriormente mencionadas, se decidió utilizar Apache OpenNLP, una librería open source especializada en las tareas más comunes del procesamiento de lenguaje natural, las razones por las cuales se tomó esta decisión se describen en la Tabla 3-1.

Apache OpenNLP	
Características	Detalle
Interfaz Gráfica	No
Interfaz de Línea de Comandos	Si
Modularidad	Si
Modelo de Desarrollo	Software Libre
Idioma	Inglés, Español, Danés, Alemán, Holandés, Portugués
Lenguaje de Programación	Java
Nivel de Complejidad	Medio
Tareas	Tokens, segmentación de la oración, etiquetado de texto,

	extracción del nombre de las entidades, fragmentación, análisis y resolución de correferencia
Versión	1.5.0
Documentación	Si

Tabla 3-1 Características Apache OpenNLP

Como se puede observar en la Tabla 3-1. Apache OpenNLP es una herramienta gratuita que ayuda al procesamiento del lenguaje natural, de una manera eficaz y eficiente, puesto que su nivel de complejidad es medio, posee gran cantidad de información de ayuda al desarrollador, soporta una variedad de idiomas y principalmente ejecuta las tareas primordiales del procesamiento de lenguaje natural.

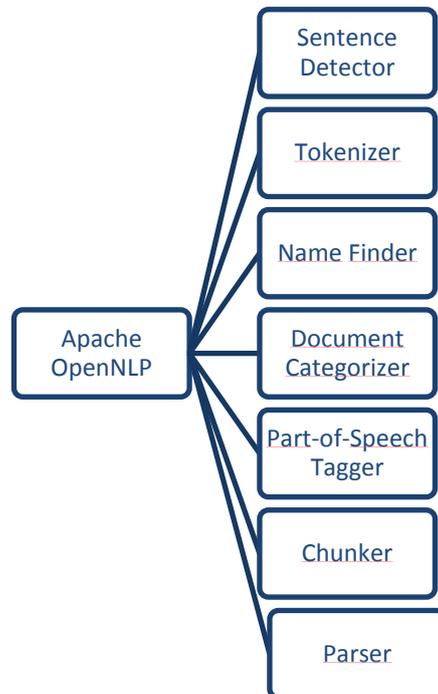


Figura. 3-2 Modelos de la herramienta Apache OpenNLP

Apache OpenNLP está compuesto por un conjunto de modelos para ejecutar las tareas del procesamiento de lenguaje natural. En la Figura. 3-2, se muestran los siete modelos que componen dicha herramienta. (Apache OpenNLP)

Para tener un conocimiento más detallado de la herramienta, a continuación se revisa las funcionalidades de cada uno de los modelos que posee Apache OpenNLP.

- **Sentence Detector:** Es el primer modelo de Apache OpenNLP, ayuda a detectar el final de una oración a través de la identificación de un signo de puntuación al final de la misma y no así teniendo en cuenta el contenido de la sentencia; esta identificación se realiza basándose en que: “Una frase se define como la secuencia más larga de caracteres de espacio en blanco recortado entre dos signos de puntuación, la primera y la última oración hacen la excepción para esta regla. El primer carácter no blanco supone el comienzo de una oración y el último espacio en blanco supone el final de la oración” (Apache OpenNLP)
- **Tokenizer:** Consiste en realizar la tokenización (segmentación del texto) de los atributos requeridos, los cuales pueden ser: palabras, números, signos de puntuación, etc. OpenNLP ofrece tres tipos de implementaciones de tokenización, como se indica en la Tabla 3-2.

TOKENIZER	
Tipo de Tokenización	Descripción
Espacios en Blanco	Las secuencias sin espacios en blanco son identificadas como tokens.
Simple	Las secuencias de la misma clase de carácter son denominadas tokens.
Que se puede aprender	Detecta el límite de los tokens basándose en los modelos de probabilidad.

Tabla 3-2 Tipos de Tokenización Apache OpenNLP

Con Apache OpenNLP, así como también con otros sistemas, está compuesto por dos fases:

- Identificación de los límites de la oración.
- Identificación de los tokens en cada sentencia.
- **Name Finder:** Named Entity Recognition, permite reconocer o detectar nombres de entidades y número en el texto, para poder realizar esto el Name Finder necesita un modelo. Este modelo es dependiente del idioma y la entidad que se forma.

Para encontrar nombres en texto sin formato, el texto debe ser segmentado en tokens y oraciones. Es importante que la tokenización para los datos de entrenamiento y el texto de entrada sean idénticos.

Dentro del Name Finder existe la Evaluation, que consiste en verificar el desempeño de las entidades extraídas. El rendimiento se mide ya sea en un conjunto de datos de prueba o por medio de la validación cruzada.

- **Document Categorizer:** Este modelo se encuentra subdividido en, clasificación y entrenamiento, las cuales se definen a continuación.
 - Clasificación: El categorizador de documentos OpenNLP puede clasificar el texto en categorías predefinidas. Está basado en la máxima entropía de framework. (Apache OpenNLP)
 - Entrenamiento: Los documentos categorizados pueden ser entrenados, como un material didáctico comentado. Los datos deben encontrarse en un formato de entrenamiento definido por el Categorizador de Documentos de OpenNLP, es decir se trata de un documento por línea, el que contiene la categoría y el texto separado por un espacio en blanco.

- **Part-of-Speech Tagger:** Este modelo se encuentra dividido en las siguientes partes (Apache OpenNLP):
 - Etiquetado: Part-of-Speech Tagger, etiqueta las palabras correspondientes basándose en el token por sí mismo o en el contexto del token (Apache OpenNLP). Un token puede tener varios pos etiquetas dependiendo del token y del contexto. Para limitar las posibles etiquetas (tags) para un token, un diccionario de tags puede ser usado para mejorar el desempeño del etiquetador, el cual será descrito posteriormente.

Existe el modelo POS Tagger la cual usa un modelo de probabilidad

para predecir la correcta pos etiqueta que este fuera del conjunto de etiquetas.

- Entrenamiento: El material para realizar el entrenamiento debe poseer las frases con sus respectivos tokens, en donde cada token tiene asignado una etiqueta para facilitar la labor de etiquetado del texto.
- Diccionario de Etiquetas: Es un diccionario de palabras en donde se especifica las etiquetas que pueden ser asignadas a un token. Usar un diccionario de etiquetas tiene las ventajas de que las etiquetas inapropiadas no pueden ser asignadas a los tokens y que al utilizar un algoritmo de búsqueda este es más rápido ya que se considera menos posibilidades debido a que solo contempla las etiquetas que están en el diccionario.
- **Chunker:** Consiste en dividir el texto en partes de palabras que se encuentren sintácticamente correlacionadas, como por ejemplo: grupos de sustantivos, grupos de verbos. Pero en esta división no se especifica su estructura interna, ni su rol en la oración principal.

Este modelo solo está visto para demostración y prueba (Apache OpenNLP).

Se puede obtener valores numéricos para conocer entre otros valores la precisión de los valores obtenidos.

- **Parser:** Este modelo realiza la agrupación en base a la similitud de de las etiquetas asignadas a cada palabra del texto analizado.
Este modelo, al igual que el anterior, solo es usado para realizar pruebas y demostraciones (Apache OpenNLP).

Una vez revisados cada uno de los modelos que conforman Apache OpenNLP y teniendo en cuenta el ámbito y alcance del tema de este proyecto de tesis se ha creído que, la implementación de todos los modelos de la herramienta no es necesario, por cuanto alguno de ellos realizan funciones que se encuentran fuera del alcance del proyecto.

Los modelos a utilizarse en la construcción del Preprocesador de Planes Académicos se muestran en la Figura. 3-3.

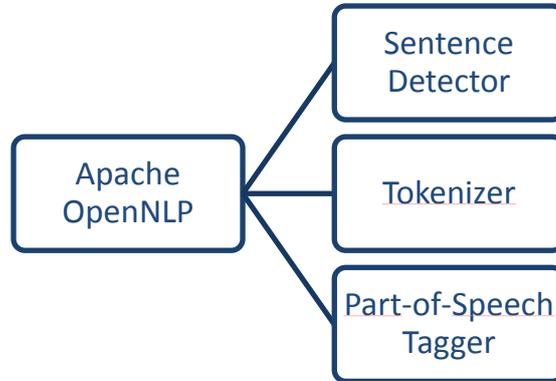


Figura. 3-3 Modelos Empleados de Apache Open NLP

Con la ayuda de los contenidos pertenecientes al plan académico de la asignatura de Arquitectura de Computadores del periodo octubre 2011 – febrero 2012 impartida en la Universidad Técnica Particular de Loja, se muestra el funcionamiento de los módulos utilizados de la herramienta.

- **Sentence Detector:** Como se indicó anteriormente, este módulo permite identificar las oraciones que posee el texto que es objeto de análisis. En la Figura. 3-4, se observa un ejemplo de cómo funciona este módulo de la herramienta.

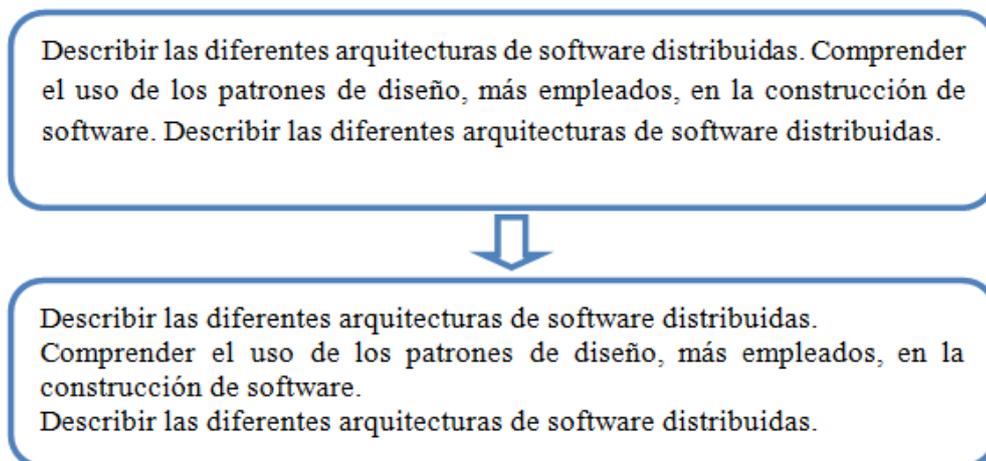


Figura. 3-4 Ejemplo - Sentence Detector

Es importante señalar que para utilizar este módulo se debe tener en cuenta que

para el idioma que fue desarrollado conste en sus reglas gramaticales el punto (.) para indicar el final de una oración, es decir, para analizar un texto en español en base a este modelo, se puede utilizar un modelo Sentence Detector desarrollado para el idioma inglés (y viceversa), sin tener ningún problema ya que en ambos idiomas la regla gramatical del (.) es la misma.

- **Tokenizer:** Consiste en realizar la segmentación del texto en base a un determinado atributo. El atributo seleccionado para este proyecto son los signos de puntuación (punto, coma, punto y coma), debido a que estos no aportan con conocimiento alguno al ámbito del proyecto.

Igual que el modelo anterior, la utilización de este módulo es independiente del idioma para el que fue creado, siempre que en las reglas gramaticales del idioma existan los mismos signos y normas que para el idioma que se está utilizando.

En la Figura. 3-5, se observa un ejemplo de cómo funciona este módulo de la herramienta.

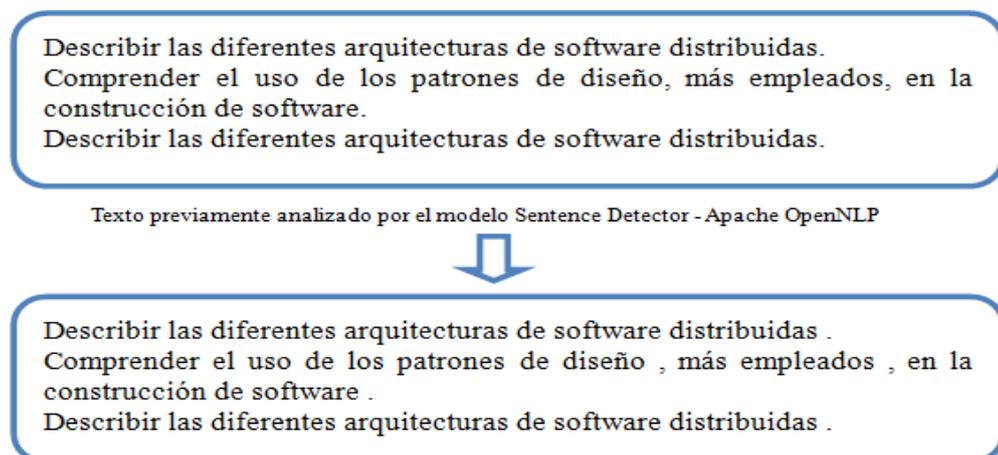


Figura. 3-5 Ejemplo – Tokenizer

Es importante mencionar que para utilizar este modelo, el texto que está siendo utilizado por la herramienta Apache OpenNLP, previamente debe haber sido analizada por el modelo Sentence Detector.

- **Part-of-Speech Tagger:** Como anteriormente se describió, este modelo se

encuentra dividido en tres partes que son: Etiquetado, diccionario de etiquetas y entrenamiento. Para la construcción del Preprocesador de Planes Académicos se ha creído que es necesario únicamente usar la parte de etiquetado y diccionario de etiquetas de este módulo.

A diferencia de los módulos anteriores, este es dependiente del idioma, por lo tanto el idioma del texto a analizar debe estar en el mismo idioma que el módulo que realiza el análisis.

En la página oficial de Apache OpenNLP (Apache OpenNLP), se encuentra desarrollado este modelo para los idiomas: inglés, danés, alemán, holandés, portugués, los cuáles no son posibles de utilizar debido a que el ámbito de este proyecto de tesis está basado en el idioma español.

Luego de realizar pruebas con diversos modelos POS Tagger, disponibles para el español que se encuentran en la red, se determinó utilizar el desarrollado por (Caicedo Carvajal, 2012), el cual es un modelo POS Tagger completo, de fácil adaptación, incluye el diccionario de etiquetas, el cual se encuentra basado en las etiquetas Eagles¹⁵. En el Anexo A. se encuentran cada una de las etiquetas con su respectiva descripción.

Es importante mencionar que para utilizar este modelo, el texto que está siendo utilizado por la herramienta Apache OpenNLP, previamente debe haber sido analizada por el modelo Tokenizer.

En la Figura. 3-6, se observa un ejemplo de cómo funciona este módulo de la herramienta.

¹⁵ Las Etiquetas Eagles fueron creadas para la anotación morfosintáctica de lexicones y corpus para todas las lenguas europeas.

Describir las diferentes arquitecturas de software distribuidas .
Comprender el uso de los patrones de diseño , más empleados , en la construcción de software .
Describir las diferentes arquitecturas de software distribuidas .

Texto previamente analizado por el modelo ~~Tokenizer~~ - Apache OpenNLP



Comprender_VMN el_DA uso_NC de_SP los_DA patrones_NC de_SP
diseño_NC más_RG empleados_VMI en_SP la_DA construcción_NC
de_SP software_NC

Figura.3-6 Ejemplo - Part-of-Speech Tagger

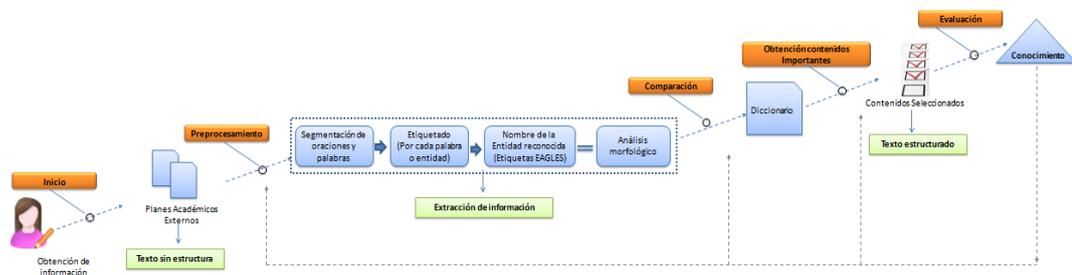
Una vez que el texto haya sido etiquetado, este se encontrará estructurado para ingresar a la fase de selección de contenidos importantes, lo cual se detallará posteriormente.

Capítulo 4
Construcción del Preprocesador de Planes Académicos de UTPL

4.1 Introducción

En los capítulos anteriores se han definido las bases conceptuales para la construcción del Preprocesador de Planes Académicos de la UTPL, por lo cual en este capítulo se presenta la descripción del proceso a seguir para la elaboración del mismo, mostrado en el capítulo 3, apartado 3.2, el cual está compuesto por las siguientes fases: inicio, preprocesamiento, comparación, obtención de contenidos importantes y evaluación.

Este proceso fue construido teniendo como base el proceso de descubrimiento de conocimiento (Fayyad, Usama, 1996) y del proceso de extracción de información, ambos señalados en el capítulo 2.



Proceso para la construcción del Preprocesador de Planes Académicos

4.2 Fase de Inicio

Para realizar un buen desarrollo de un sistema, es necesario conocer claramente la situación actual de la organización y el problema que esta desea dar solución. Con esta información recopilada se podrá capturar de mejor forma los requerimientos y casos de uso necesarios para la construcción del sistema. A continuación se describen cada uno de ellos.

4.2.1 Situación Actual

La Universidad Técnica Particular, es una de entidades educativas más grandes a nivel nacional, la cual acoge a una gran cantidad de estudiantes, en las modalidades de estudios abierta, a distancia, semipresencial y presencial.

La UTPL lleva a cabo una serie de procesos entre los cuales se encuentra la Gestión Académica, la cual conlleva una serie de actividades dirigidas a resolver y mejorar los servicios académicos brindados a la colectividad. Una de estas actividades es la convalidación de asignaturas en base al análisis y comparación manual de los planes académicos externos frente a los planes académicos vigentes dentro de la oferta académica de la universidad.

En la actualidad este proceso es realizado por docentes de cada titulación de la UTPL, en calidad de fiscales, los cuales verifican que los planes académicos externos cumplen los requerimientos necesarios para ejecutar el proceso anteriormente mencionado.

4.2.2 Descripción de la Problemática

El proceso de convalidación de asignaturas es un proceso complejo, debido a que el fiscal encargado de llevar a cabo esta gestión debe manualmente verificar el grado de similitud que presenta el plan académico externo, frente al plan académico de una asignatura de la UTPL igual o semejante y en base a un conjunto de reglas y normas dictaminar si una asignatura puede o no ser convalidada. Por lo tanto el principal problema encontrado es que se trata de un proceso extenso llevado de forma manual y que por sus características hace que este sea culminado en significativas cantidades de tiempo, lo cual ocasiona que el servicio ofrecido por la universidad no sea cien por ciento óptimo en cuanto a tiempo de respuesta se refiere.

4.2.3 Descripción de la solución

La Universidad Técnica Particular de Loja en su continuo avance y desarrollo tecnológico y con el afán de ofrecer un mejor servicio a los estudiantes de la entidad educativa y a la comunidad en general, se ve en la necesidad de construir un sistema informático que sea capaz de llevar a cabo las funciones realizadas por los fiscales de la universidad, pero de una manera más rápida y segura. Este sistema estará conformado por una serie de componentes, siendo uno de los principales el Preprocesador de Planes Académicos en donde se realizará el procesamiento inicial el cual consiste en el análisis y selección de contenidos importantes de los planes académicos que hayan ingresado al

proceso de convalidación.

En el Anexo B – Fase de Inicio, se muestra a detalle la información recopilada en esta fase, la cual se encuentra plasmada en un conjunto de plantillas¹⁶ en las que se muestra los modelos de procesos de negocio, afectados y sus necesidades, oportunidades de mejora para la organización, etc.

4.2.4 Especificación de Requerimientos

Luego del estudio del problema a resolver es necesario identificar las necesidades funcionales de los stakeholders que deberán ser soportadas por el sistema a desarrollar. Estas necesidades fueron recopiladas mediante observación directa, revisión de planes académicos y consultas a expertos.

Los requerimientos para el Preprocesador de Planes Académicos son:

- Autenticación para ingreso al sistema
- Registrar planes académicos externos
- Analizar plan académico
- Generar archivo de texto con resultado de análisis
- Generar archivo XML con contenidos importantes extraídos.
- Generar gráfico
- Crear nuevo corpus y lematización
- Actualizar lematización existente
- Crear nuevas carreras – UTPL
- Crear nuevas asignaturas – UTPL
- Gestionar usuarios
- Incrementar valores de los campos del registro de planes.

En el Anexo B – Fase de Inicio, se desarrolla la especificación de cada uno de los requerimientos antes listados, estos servirán de referencia tanto al desarrollador como al

¹⁶ Las plantillas utilizadas son basadas en la metodología Rational Unified Process

cliente, para validar que el producto de software desarrollado cumpla con las expectativas planteadas.

4.2.5 Especificación de casos de uso

Una vez definidos los requerimientos es importante definir los casos de uso, para definir la secuencia de acciones que va a realizar el usuario sobre el preprocesador de planes académicos. Estos casos de uso nacen a partir de la especificación de requerimientos señalados en el punto anterior.

En la Figura 4-1, se muestra el diagrama general de casos de uso para la construcción de este sistema.

En el Anexo B – Fase de Inicio, se desarrolla la especificación detallada de cada uno de los casos de uso mostrados en la Figura 4-1

4.2.6 Arquitectura de Software

En base a los requerimientos necesarios para la elaboración de este proyecto, se ha creído conveniente implementar una arquitectura donde la carga este dividida en tres capas que son: capa de presentación, capa de lógica de negocios y capa de datos. Cada una de estas capas soportará las transacciones realizadas por los usuarios en el sistema.

En la Figura 4-2 se muestra la Vista Lógica de la Arquitectura General para el Preprocesador de Planes Académicos, con la cual se podrá tener una vista global de cómo se encuentra estructurado el sistema.

En el Anexo B – Fase de Inicio, se muestra la especificación de la arquitectura de software construida para el preprocesador de planes académicos.

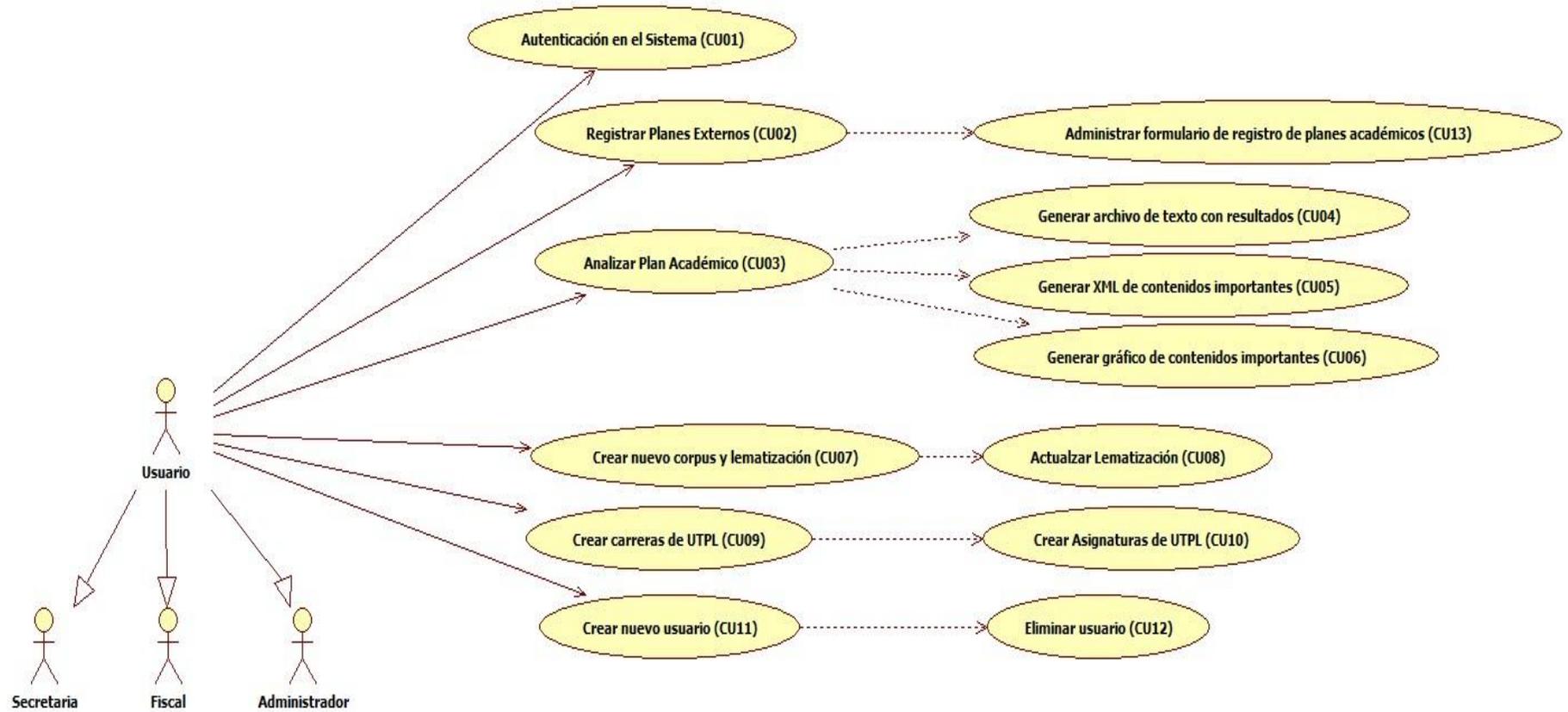


Figura. 4-1 Diagrama General de Casos de Uso

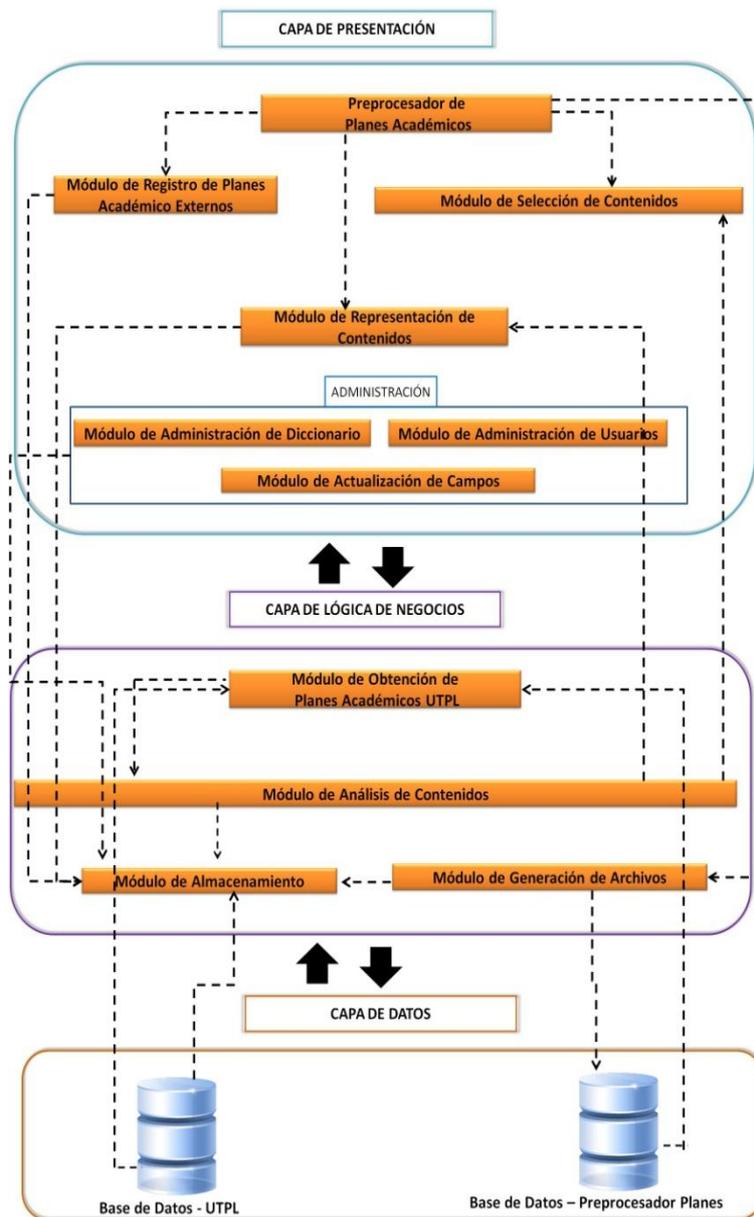


Figura. 4-2 Vista Lógica de la Arquitectura General para el Preprocesador de Planes Académicos

4.2.7 Esquema de Base de Datos

El sistema del Preprocesador de Planes Académicos cuenta con almacenamiento de información en un servidor de base de datos MySQL. Esta base está compuesta por 16 tablas para manejar la información de: planes académicos externos, diccionario, preprocesador y usuarios.

A continuación se lista las tablas utilizadas dentro de la base de datos describiendo su uso dentro de la aplicación:

Registro de Planes Académicos externos

- **reg_carreras:** Almacena el nombre de la carrera a la cual pertenece el plan académico que registra el usuario en el sistema.
- **reg_centrosutpl:** En esta tabla se almacena el nombre del centro de UTPL en donde se registra el plan académico.
- **reg_datosplan:** Esta tabla registra los datos generales del plan académico externo, como son: entidad educativa, carrera, asignatura, centro utpl, periodo, código de plan académico y nombre del archivo digital del plan.
- **reg_entidadeseducativas:** Se guarda en esta tabla el nombre de la entidad educativa a la cual pertenece el plan académico.
- **reg_modalidades:** Se almacena en esta tabla la modalidad de estudios a la cual pertenece el plan académico externo.
- **reg_periodosacademicos:** En esta tabla se guarda el periodo académico al cual pertenece el plan académico que se registra.
- **reg_unidades:** En esta tabla se almacena el contenido de cada una de las unidades que posee el plan académico.

Diccionario

- **corp_asignaturas:** Esta tabla almacena el nombre de la asignatura a la cual pertenece un corpus.
- **corp_carreras:** Esta tabla almacena el nombre de la carrera a la cual pertenece un corpus.
- **corp_corpus:** Se asocia dentro de esta tabla las palabras importantes de una asignatura de una carrera en su forma base.
- **corp_lematizacion:** En esta tabla se guarda las formas más comunes de cada una de las palabras que se encuentran en el corpus.
- **corp_tipocorpus:** Dentro de esta tabla se almacena el tipo de corpus (genérico, específico) al cual pertenece el corpus que se está registrando.

Preprocesador

- **pre_contenidos:** En esta tabla se almacena el identificador del contenido importante obtenido por el sistema y el código del plan al que se encuentra relacionado dicho contenido.
- **pre_frases:** En esta tabla se guarda el identificador de la frase importante obtenida por el sistema y el identificador del contenido al cual se encuentra relacionado la frase.
- **pre_palabras:** Se guarda en esta tabla, la palabra importante extraída por el sistema, el identificador de la frase a la cual está ligada esta palabra y el tipo de palabra (etiquetas Eagles) a la que pertenece.

Usuarios

- **admin_usuarios:** En esta tabla se almacenan los datos informativos sobre los usuarios que utilizan el sistema, así como también datos sobre usuario, contraseña y rol al cual pertenece dicho usuario.

El esquema gráfico de la base de datos para el Preprocesador de la Base de Datos se lo puede observar en la Figura 4-3.

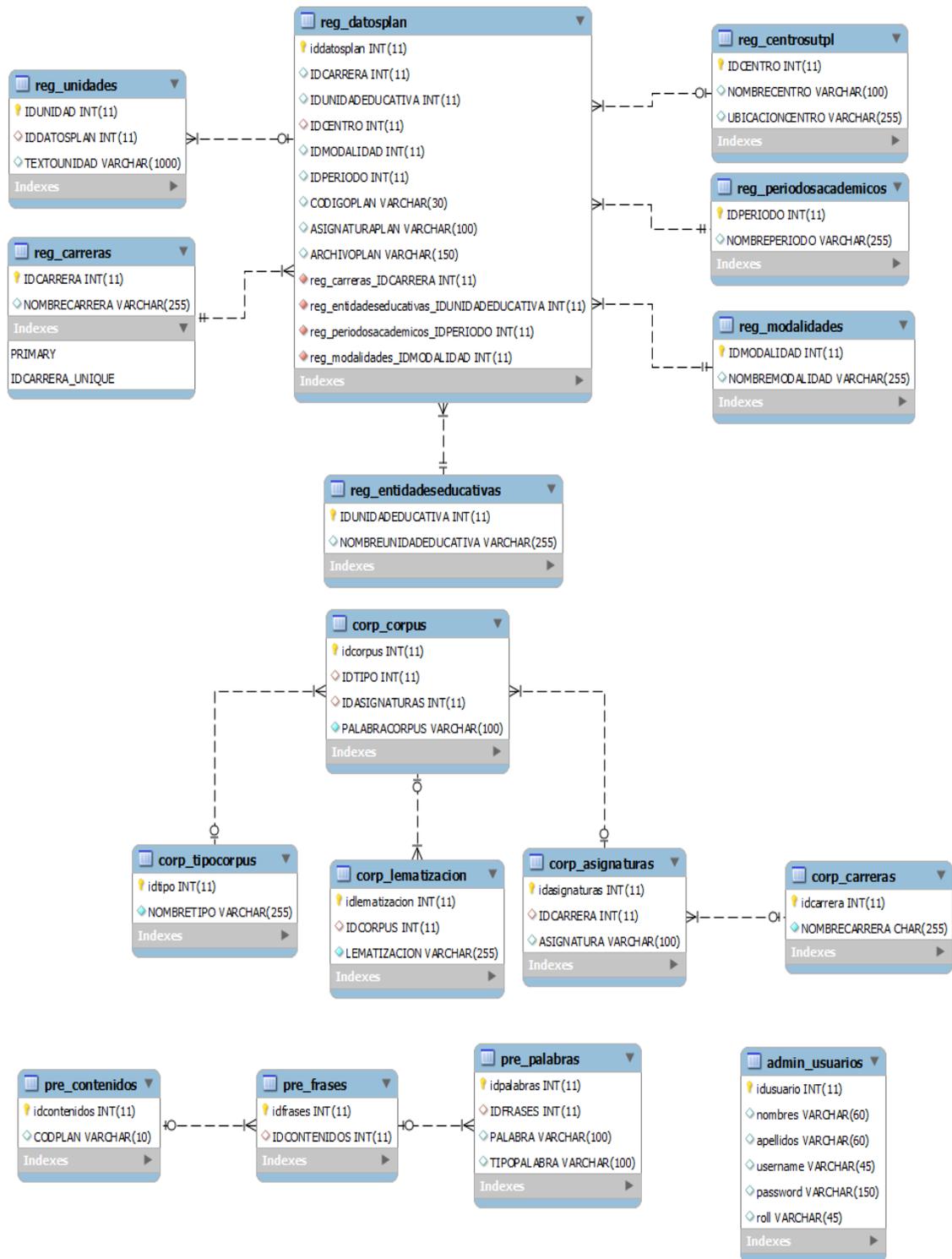


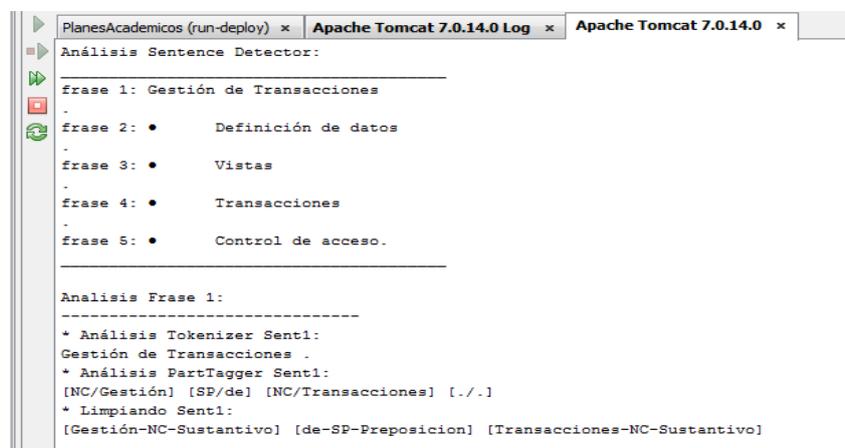
Figura. 4-3 Diagrama de Base de Datos

4.3 Fase de Preprocesamiento

Continuando con la construcción del Preprocesador de Planes Académicos para UTPL, y teniendo el marco de trabajo definido es necesario especificar la forma de implementar los fundamentos teóricos de lenguaje natural y extracción de contenidos analizados en los capítulos uno y dos de este documento.

El Preprocesamiento de un texto sin estructura como lo son los planes académicos externos, consiste en analizar la información de entrada al sistema. Este análisis fue realizado con la ayuda de la librería Apache OpenNLP, descrita en el capítulo 3, sección 3.3. Esta herramienta permitió a través de sus módulos segmentar las oraciones en un conjunto de frases (Módulo Sentence Detector); posteriormente eliminar atributos de estas frases que no aporten significativamente al análisis del texto, como lo son los signos de puntuación (Tokenizer) para finalmente con la utilización de la librería open source indicada anteriormente, etiquetar cada una de las palabras que conforman las frases que fueron segmentadas inicialmente. Para asignar las etiquetas correspondientes a cada palabra, la librería realiza un análisis morfológico de cada una de ellas para reconocer su estructura y así poder asignar la etiqueta idónea. En el Anexo A – Etiquetas Eagles, se encuentran descritas cada una de las etiquetas que utiliza esta librería para asignar a las palabras que analiza, como pueden ser: verbos, sustantivos, adjetivos, etc.

En la Figura 4-4, se observa la salida generada por el sistema desarrollado, para esta fase de Preprocesamiento.



```
PlanesAcademicos (run-deploy) x Apache Tomcat 7.0.14.0 Log x Apache Tomcat 7.0.14.0 x
Análisis Sentence Detector:
-----
frase 1: Gestión de Transacciones
.
frase 2: • Definición de datos
.
frase 3: • Vistas
.
frase 4: • Transacciones
.
frase 5: • Control de acceso.
-----

Análisis Frase 1:
-----
* Análisis Tokenizer Sent1:
Gestión de Transacciones .
* Análisis PartTagger Sent1:
[NC/Gestión] [SP/de] [NC/Transacciones] [./.]
* Limpiando Sent1:
[Gestión-NC-Sustantivo] [de-SP-Preposicion] [Transacciones-NC-Sustantivo]
```

Figura. 4-4 Salida - Fase de Preprocesamiento

4.4 Fase de Comparación

Para desarrollar esta fase fue necesaria la utilización de un concepto fundamental empleado en cualquier sistema de procesamiento de lenguaje natural como es el diccionario o lexicón, el cual almacena un conjunto de palabras basadas en un idioma y gramática común. (Contreras, Hilda, 2001)

A continuación se explican el diccionario empelado, su generación, utilización y lineamientos para futuras expansiones.

La creación de un diccionario para este sistema, tiene como objetivo almacenar las palabras más importantes relacionadas a cada asignatura de una carrera, esta relación uno a uno fue implementa con el fin de evitar posibles ambigüedades entre los términos almacenados.

El diccionario está compuesto por dos partes que son:

- **Corpus:** Está compuesto por un conjunto de palabras en su forma base, como por ejemplo: aprender, conocer, insertar, método, diagrama, etc. Las palabras en el corpus además están clasificadas de la siguiente manera:
 - **Genérico:** En esta clasificación están las palabras que no se encuentran directamente relacionadas con el ámbito de la asignatura a la que pertenece dicho corpus, sino más bien son palabras de propósito general. Ejemplos: aprender, conocer, investigar. Luego de varias pruebas realizadas se pudo determinar que las palabras que se ubican dentro de esta clasificación son verbos.
 - **Específico:** Dentro de esta clasificación se encuentran las palabras que están directamente relacionadas con el ámbito de la asignatura a la que pertenece dicho corpus. Ejemplos:
Asignatura: Programación

Palabras: arreglo, java, diagrama, lista, pseudocódigo.

Luego de varias pruebas realizadas se pudo determinar que las palabras que se ubican dentro de esta clasificación son sustantivos y adjetivos.

- **Lematización:** Está compuesto por las formas más comunes que se derivan de cada palabra que se encuentra en el corpus. Por ejemplo:
Palabra del corpus: Definir
Palabras de lematización: definición, definiciones, definirán.

En la Figura 4-5, se muestra gráficamente como se encuentra la estructura del diccionario para el Preprocesador de Planes Académicos.

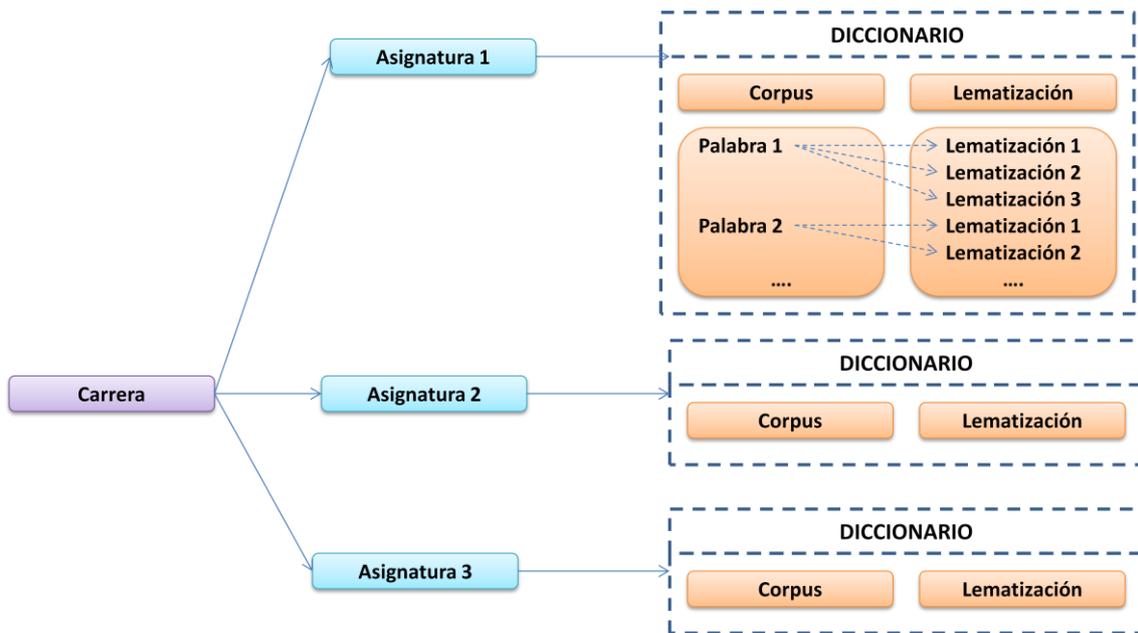


Figura. 4-5 Esquema del Diccionario - Preprocesador de Planes Académicos

Para incrementar la base de palabras del diccionario es importante seguir los lineamientos anteriormente señalados, así mismo es necesario que cada una de las palabras ingresadas en el diccionario sea acorde a las reglas ortográficas y gramaticales del idioma español.

Continuando con el desarrollo de esta fase, se describe como el Preprocesador de Planes Académicos realiza el proceso de comparación.

La salida de la fase de Preprocesamiento, consiste en que todas las palabras que conforman el plan académico se encuentran etiquetadas de acuerdo al tipo de palabra que corresponda. Estas etiquetas desempeñan un papel fundamental en esta fase de comparación, tal como se describe a continuación:

- **Filtrado de Stopwords:** Las stopwords son aquellas palabras que aportan de forma mínima al significado de la oración, como por ejemplo los artículos. Es por esta razón que se ha creído conveniente realizar el filtrado de los mismos a través la verificación de la etiqueta DA, es decir se realiza un análisis de las etiquetas que poseen las palabras, si dicha etiqueta es encontrada, se procede a suprimir dicha palabra temporalmente.
- **Comparación con Corpus Genérico o Corpus Específico:** Para agilizar el proceso de comparación de las palabras que se encuentran en el plan académico externo, frente a las palabras importantes almacenadas en el diccionario, se ha creído conveniente realizar el proceso de comparación utilizando las etiquetas asignadas a cada una de las palabras, es decir, si la palabra a comparar posee cualquiera de las etiquetas que se encuentran dentro de la clasificación de verbos (Anexo A – Etiquetas Eagles), se procederá a buscar la palabra dentro del corpus de tipo genérico y sus respectivas lematizaciones, en caso de poseer cualquier otra etiqueta se buscará la palabra dentro del corpus de tipo específico y sus respectivas lematizaciones.

Cuando la palabra del plan externo que se está analizando se encuentra registrada en el diccionario (corpus o lematización), el sistema le asigna a dicha palabra el atributo de TRUE, en caso de no estarlo, el sistema asigna a dicha palabra el atributo de FALSE, tal como se muestra en la Figura 4-6.

```

PlanesAcademicos (run-deploy) x Apache Tomcat 7.0.14.0 Log x Apache Tomcat 7.0.14.0 x
-----
Analizando Contenido 2:
-----
Palabra='MODELO' ¿Esta En Corpus? =true
Palabra='CONCEPTUAL' ¿Esta En Corpus? =true
Palabra='ENTIDAD' ¿Esta En Corpus? =true
Frase (1) ¿importante?: true
Palabra='Características' ¿Esta En Corpus? =false
Palabra='entidad' ¿Esta En Corpus? =true
Frase (2) ¿importante?: true
Palabra='Atributos' ¿Esta En Corpus? =true
Frase (3) ¿importante?: true
Palabra='Clave' ¿Esta En Corpus? =true
Palabra='primaria' ¿Esta En Corpus? =false
Frase (4) ¿importante?: true
Palabra='Cardinalidad' ¿Esta En Corpus? =false
Palabra='atributos' ¿Esta En Corpus? =true
Frase (5) ¿importante?: true
Palabra='Dominios' ¿Esta En Corpus? =false
Palabra='restricciones' ¿Esta En Corpus? =true
Palabra='atributos.' ¿Esta En Corpus? =false
Frase (6) ¿importante?: false
[MODELO-NC-Sustantivo] [CONCEPTUAL-AQ-Adjetivo] [ENTIDAD-NC-Sustantivo] [Características-NC-Sustantivo] [de-SP-Preposici
Contenido Importante!!!
-----

```

Figura. 4-6 Salida - Fase de Comparación

4.5 Fase de Obtención de Contenidos importantes

Una vez que se han comparado y etiquetado con TRUE o FALSE las palabras del plan académico externo que se encuentran o no en el diccionario de una asignatura de UTPL, con la cual se está comparando el plan académico externo, es importante extraer no solo las palabras importantes, sino que también se deben extraer los contenidos importantes. En la Figura 4-7, se muestra como se encuentra organizada la estructura de los contenidos, frases y palabras.

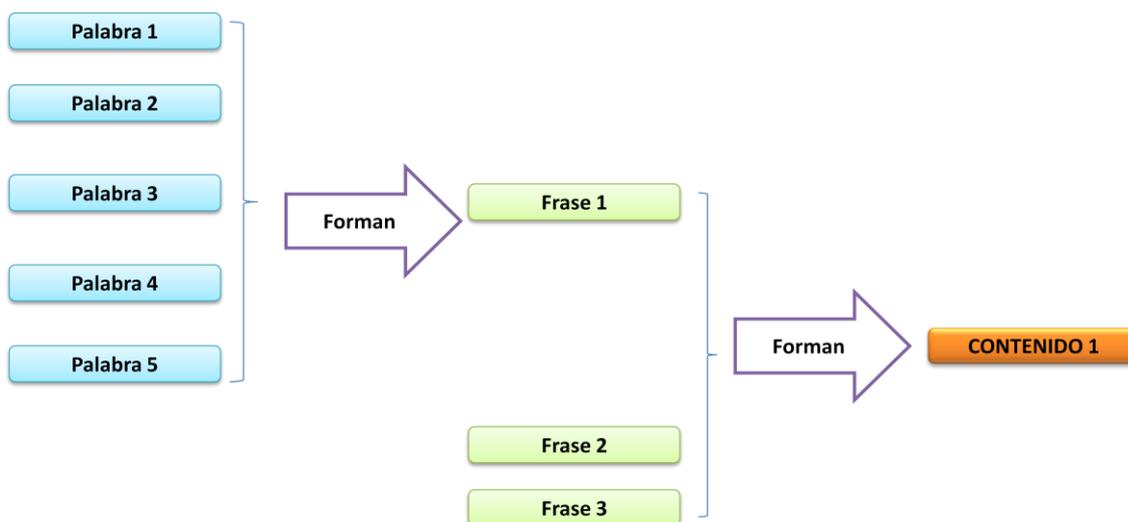


Figura. 4-7 Organización de Palabras - Frases – Contenidos

Para realizar la obtención de contenidos importantes se basó en la estructura

organizativa mostrada en la Figura 4-7, siguiendo la siguiente regla:

- Si el número de palabras que tienen la etiqueta TRUE *es mayor o igual* que las palabras que tienen la etiqueta FALSE, la FRASE ES IMPORTANTE, sino la FRASE NO ES IMPORTANTE.
- Si el número de frases que tienen la etiqueta TRUE *es mayor o igual* que las frases que tienen la etiqueta FALSE, el CONTENIDO ES IMPORTANTE, sino el CONTENIDO NO ES IMPORTANTE.

La explicación de estas reglas se las puede observar a mayor detalle en la Figura 4-8.

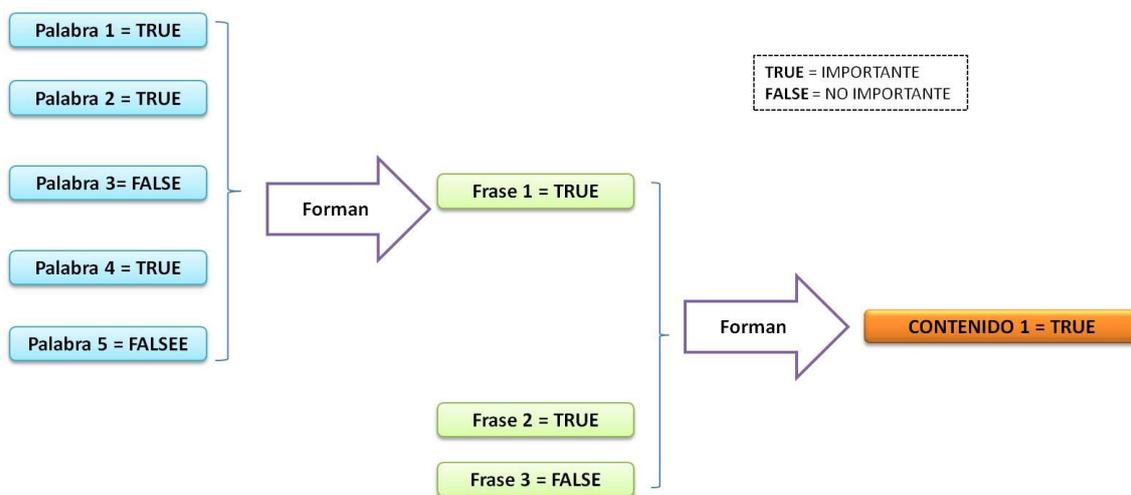


Figura. 4-8 Salida - Fase de Comparación

Los contenidos importantes obtenidos por el sistema son almacenados en la base de datos, para que en caso de que se desee volver a analizar el mismo plan académico ya no se ejecute todo el proceso de análisis sino que únicamente se muestren los contenidos importantes.

En la Figura 4-9, se observa la interfaz web para todo el proceso de análisis de planes académicos externos que se ha venido analizando hasta el momento.

Sistema de Convalidación de Planes Académicos UTPL

Inicio | Análisis Específico | Análisis Completo | Salir | Bienvenid@fiscal1

ANÁLISIS ESPECÍFICO DE PLANES ACADÉMICOS

Buscar Plan Académico

Codigo del Plan:

[Búsqueda Avanzada](#)

DATOS DEL PLAN ACADEMICO

Asignatura: Sistemas de Base de Datos I
 Centro: Loja
 Universidad: Escuela Superior Politécnica del Litoral
 Modalidad: Presencial
 Carrera: Ingeniería en Computación

Analizar con:

Carrera:

Asignatura:

Contenidos Seleccionados

INTRODUCCIÓN. / Tecnología de RDBMS / Evolución de RDBMS / Cliente/Servidor componentes / Glosario de términos usados en bases de datos. /

MODELO CONCEPTUAL ENTIDAD / Características de entidad / Atributos / Clave primaria / Cardinalidad de atributos / Dominios restricciones de atributos. /

MODELO CONCEPTUAL RELACIÓN / Definición de relación / Cardinalidad de relación / Relación de supertipo-subtipo / Relación de dependencia / Relación recursiva / Relación ternaria / Atributos de relaciones. /

Contenido	Quit...
MODELO CONCEPTUAL ENTIDAD / C...	X
MODELO CONCEPTUAL RELACIÓN / ...	X
ÁLGEBRA RELACIONAL / Definición d...	X

Figura. 4-9 Interfaz web de Análisis Específico de Planes Académicos Externos

4.6 Fase de Evaluación

Para finalizar el proceso de construcción del Preprocesador de Planes Académicos es importante evaluar los resultados obtenidos con el sistema por lo cual se ha creído conveniente realizar un Análisis de los resultados obtenidos por el sistema.

4.6.1 Análisis de los resultados obtenidos por el sistema

Al tratarse de un sistema de procesamiento de lenguaje natural, es necesario evaluar los resultados obtenidos por el sistema construido. Es por esta razón que se ha utilizado MUC (Conferencia para la Comprensión de Mensajes - Message Understanding Conference) que además de promover la creación de sistemas de extracción de información, también define métodos para su evaluación.

El sistema de extracción de información propuesto por MUC, cuenta con: un conjunto de documentos de donde se va a extraer la información (planes académicos), conjunto de registros extraídos (diccionario) y el conjunto de registros que el sistema a ser evaluado extrae (contenidos importantes). En la Figura 4-10, se muestra gráficamente el funcionamiento del proceso de evaluación.



Figura. 4-10 Proceso para evaluar un sistema de extracción de información – MUC

Para la utilización de este proceso MUC ha definido un conjunto de atributos necesarios para realizar la evaluación de un sistema de extracción de información, estos atributos se detallan en la Tabla 4-1.

Nombre	Fórmula	Descripción
Número correcto	COR	Ocasiones donde la clave y la respuesta coinciden
Número incorrecto	INC	Ocasiones donde la clave y la repuesta no coinciden
Numero perdido	MIS	Ocasiones donde existe una clave pero no una respuesta
Número falso	SPU	Ocasiones donde existe una respuesta pero no una clave
Número posible	$POS = COR + INC + MIS$	Número de registros en la clave
Número actual	$ACT = COR + INC + SPU$	Número de registros en la respuesta

Tabla 4-1 Atributos para evaluación de un sistema de procesamiento de lenguaje natural

Con los atributos definidos anteriormente, se han asociado un conjunto de métricas para ayudar a evaluar a los sistemas de procesamiento de lenguaje natural. Las métricas seleccionadas para evaluar al Preprocesador de Planes Académicos se muestran en la Tabla 4-2.

MÉTRICAS DE EVALUACIÓN - MUC		
Nombre	Descripción	Fórmula
Cobertura	Es una métrica de integridad ya que mide la cantidad de datos relevantes se extrajo del texto con relación al total disponible.	$REC = COR / POS$
Precisión ¹⁷	Se refiere a la exactitud con la que el sistema extrae datos, es decir, es la cantidad de datos pertinentes en relación con el total puesto en el sistema.	$PRE = COR / ACT$
Sobregeneración	Se refiere a la cantidad de cargas falsas atribuidas en relación con el total de cargas asignadas	$OVG = SPU / ACT$

Tabla 4-2 Métricas de Evaluación

Los datos utilizados para la realización de las pruebas fueron los planes académicos correspondientes a las asignaturas de las siguientes universidades, que se muestran en la Tabla 4-3.

Categoría	Universidad	Asignatura
A	Escuela Superior Politécnica del Litoral	Fundamentos de Programación Sistema de Base de Datos I
B	Universidad de Especialidades Espíritu Santo	Fundamentos de Programación Sistema de Base de Datos I

¹⁷ Precisión también puede ser descrito como la tendencia de un sistema para evitar la asignación de cargas incorrectas, ya que asigna cargas más buenas.

CERRADA ¹⁸	Universidad Autónoma de Quito	Programación I Técnica de Base de Datos I
-----------------------	-------------------------------	--

Tabla 4-3 Datos de Prueba

Luego de la utilización de las métricas de evaluación mostradas en la Tabla 4-2, se obtuvieron los siguientes valores.

Universidad	Escuela Superior Politécnica del Litoral
Carrera	Ingeniería en Computación
Asignatura	Sistemas de Base de Datos I
COR	65
INC	94
MIS	3
SPU	0
POS = COR + INC + MIS	101
ACT = COR + INC + SPU	159
REC = COR / POS	0,643564356
PRE = COR / ACT	0,408805031
OVG = SPU / ACT	0

Universidad	Universidad de Especialidades Espíritu Santo
Carrera	Ingeniería en Sistemas
Asignatura	Sistemas de Base de Datos I
COR	42
INC	64
MIS	6
SPU	0
POS = COR + INC + MIS	101
ACT = COR + INC + SPU	106
REC = COR / POS	0,415841584
PRE = COR / ACT	0,396226415
OVG = SPU / ACT	0

Universidad	Universidad Autónoma de Quito
Carrera	Informática
Asignatura	Técnica de Base de Datos I
COR	53
INC	58

¹⁸ Universidades Cerradas: A esta categoría pertenecen aquellas universidades que luego de un estudio y verificación de sus procesos de gestión académica realizado por el CEAACES – Consejo ecuatoriano de Acreditación y Aseguramiento de la Calidad de la Educación Superior, dictaminó que estas no cumplían con los requisitos mínimos para su funcionamiento, por lo cual se procedió al cierre de las mismas.

MIS	2
SPU	0
POS = COR + INC + MIS	101
ACT = COR + INC + SPU	111
REC = COR / POS	0,324752475
PRE = COR / ACT	0,477477477
OVG = SPU / ACT	0

Universidad	Escuela Superior Politécnica del Litoral
Carrera	Ingeniería en Computación
Asignatura	Fundamentos de Programación
COR	29
INC	69
MIS	2
SPU	0
POS = COR + INC + MIS	87
ACT = COR + INC + SPU	98
REC = COR / POS	0,333333333
PRE = COR / ACT	0,295918367
OVG = SPU / ACT	0

Universidad	Universidad de Especialidades Espíritu Santo
Carrera	Ingeniería en Sistemas
Asignatura	Fundamentos de Programación
COR	26
INC	76
MIS	3
SPU	0
POS = COR + INC + MIS	87
ACT = COR + INC + SPU	102
REC = COR / POS	0,298850575
PRE = COR / ACT	0,254901961
OVG = SPU / ACT	0

Universidad	Universidad Autónoma de Quito
Carrera	Informática
Asignatura	Programación I
COR	16
INC	44
MIS	1
SPU	0

POS = COR + INC + MIS	87
ACT = COR + INC + SPU	60
REC = COR / POS	0,183908046
PRE = COR / ACT	0,266666667
OVG = SPU / ACT	0

- **Resultados de la métrica de evaluación – Cobertura**

En base a los cálculos realizados se puede observar que la mayor parte de datos relevantes extraídos, de las asignaturas de base de datos de las tres universidades señaladas anteriormente comparados con la asignatura de Fundamentos de Base de Datos UTPL-ECTS, corresponden a los contenidos ofertados por la Escuela Superior Politécnica del Litoral (Categoría A) con un valor del 64,4%. Le sigue la Universidad de Especialidades Espíritu Santo (Categoría B) con un 41,6% de valores extraídos. La menor parte de datos relevantes extraídos en base a la comparación tiene un valor de 18,4% que corresponden al plan académico de la Universidad Autónoma de Quito, la cual en la actualidad se encuentra cerrada. Igual situación se observa en la comparación con la asignatura de Fundamentos de Programación UTPL-ECTS, en donde Escuela Superior Politécnica del Litoral (Categoría A) posee un porcentaje de evaluación del 33,33 %, seguida por la Universidad de Especialidades Espíritu Santo (Categoría B) con un porcentaje del 29,88 % y con el menor porcentaje de datos relevantes extraídos la Universidad Autónoma de Quito con un 18,39%.

Estos valores nos ayudan a verificar el nivel eficacia en la cobertura que está realizando el Preprocesador de Planes Académicos de UTPL, ya que como se puede observar en la Figura 4-11 y Figura 4-12, existe una tendencia gradual decreciente, lo cual nos indica que el nivel de cobertura que realiza el Preprocesador de Planes Académicos es aceptable por cuanto al comparar los planes académicos de dos universidades que pertenecen a la misma categoría (UTPL y ESPOL) el nivel de cobertura debe de ser mucho mayor al nivel de cobertura resultante de la comparación de los planes académicos de universidades que pertenecen a distintas categorías (UTPL y UEES) y más aún si estas han sido cerradas (UAQ).

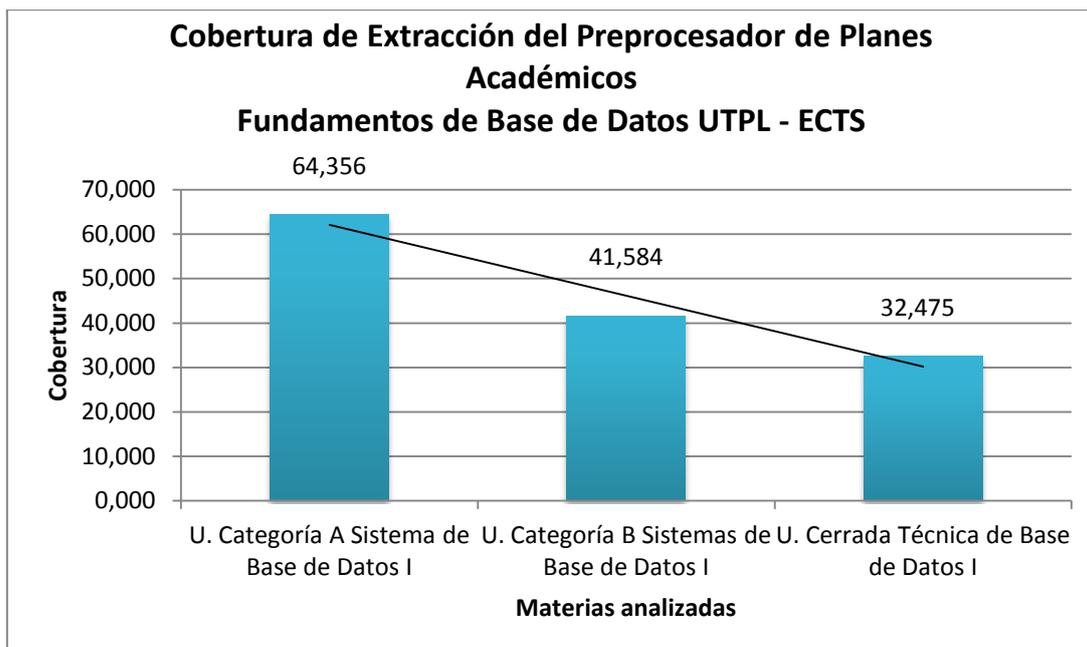


Figura. 4-11 Proceso para evaluar un sistema de extracción de información – MUC

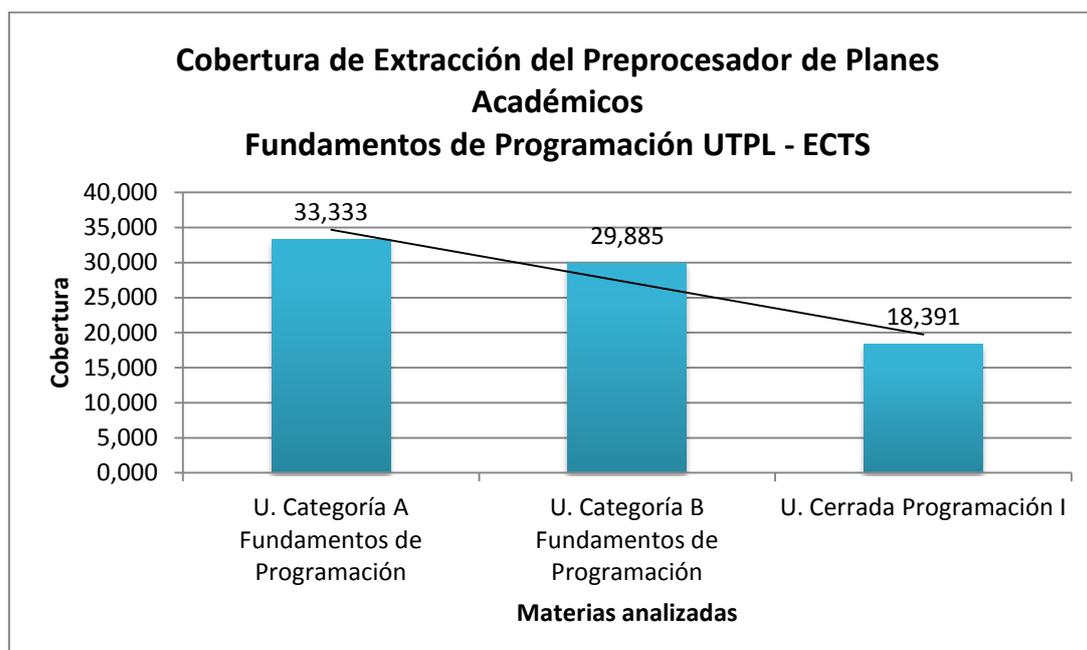


Figura. 4-12 Cobertura de Extracción del Sistema frente a Fundamentos de Programación UTPL – ECTS

- **Resultados de la métrica de evaluación – Precisión**

En base a los cálculos realizados se puede observar que la mayor precisión que tiene el Preprocesador de Planes Académicos para extraer los datos de los planes correspondientes a la asignatura de Fundamentos de Base de Datos UTPL – ECTS frente a las asignaturas de bases de datos de las universidades mencionadas anteriormente, son los ofertados por la Escuela Superior Politécnica del Litoral

(Categoría A) con un valor del 40,8%. Le sigue la Universidad de Especialidades Espiritu Santo (Categoría B) con un 39,62% de precisión. El menor porcentaje de precisión obtenido es para la Universidad Autónoma de Quito con un 37,74%. Igual situación se observa en la comparación con la asignatura de Fundamentos de Programación UTPL-ECTS en donde Escuela Superior Politécnica del Litoral (Categoría A) posee un porcentaje de precisión del 29,59%, seguida por la Universidad de Especialidades Espiritu Santo (Categoría B) con un porcentaje del 25,49 % y con el menor porcentaje de datos relevantes extraídos la Universidad Autónoma de Quito con un 16,66%. En la Figura 4-13 y Figura 4-14 se observan gráficamente los valores indicados.

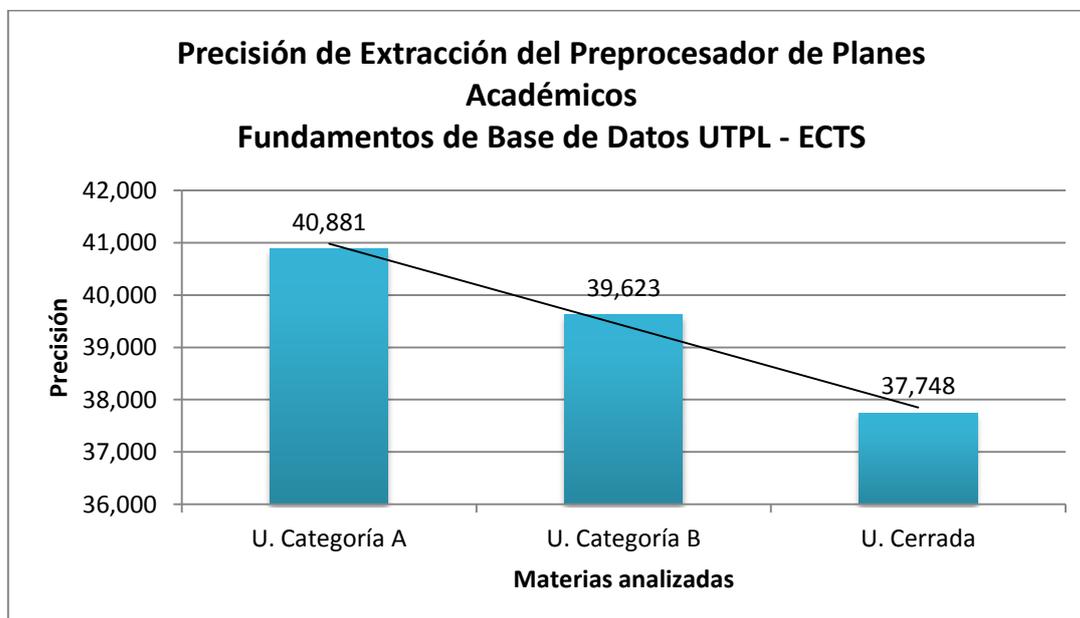


Figura. 4-13 Precisión de Extracción del Sistema frente a Fundamentos de Base de Datos UTPL – ECTS

El nivel de precisión varía debido a que esta métrica se encuentra relacionada de forma inversamente proporcional a la cantidad de palabras extraídas frente al número de respuestas emitidas por el Preprocesador. Teniendo en cuenta lo antes señalado se puede afirmar que el Preprocesador de Planes Académicos tiene un nivel de precisión aceptable, por cuanto al comparar los planes académicos de dos universidades que pertenecen a la misma categoría (UTPL y ESPOL) el nivel de precisión debe ser mucho mayor al nivel de precisión resultante de la comparación de los planes académicos de universidades que pertenecen a distintas categorías (UTPL y UEES) y más aún si estas han sido cerradas (UAQ).

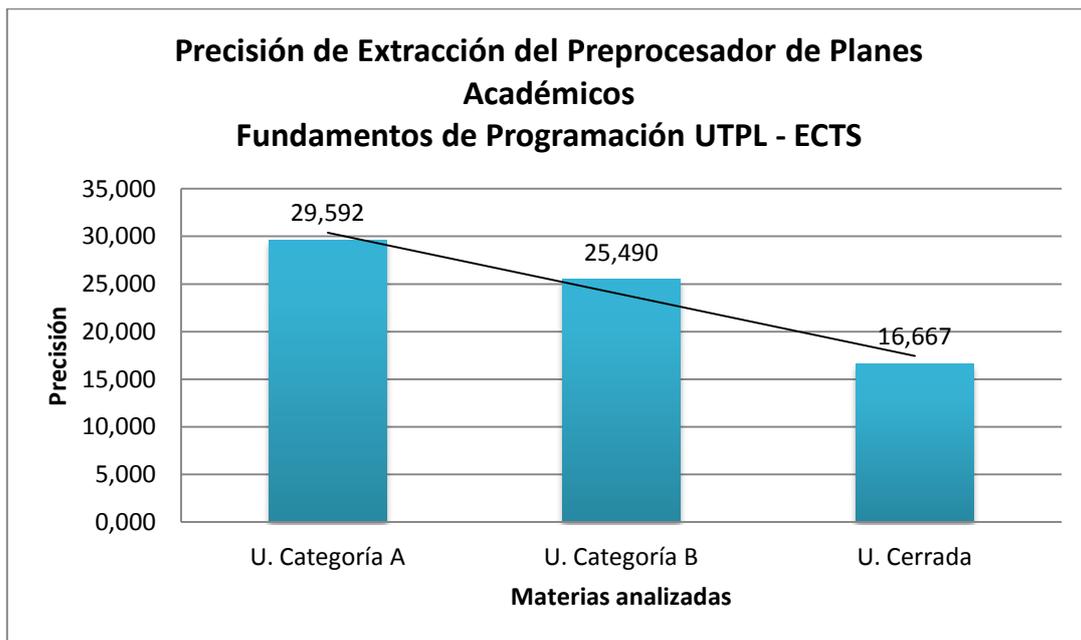


Figura. 4-14 Precisión de Extracción del Sistema frente a Fundamentos de Programación UTPL – ECTS

- **Resultados de la métrica de evaluación – Sobregeneración**

Como se lo mencionó anteriormente la sobregeneración es la cantidad de cargas falsas atribuidas por el sistema en relación con el total asignado. En el caso del Preprocesador de Planes Académicos la sobregeneración es 0, ya que al tratarse de un sistema que realiza un análisis morfológico y sintáctico de los contenidos de los planes académicos, únicamente asigna cargas cuando la palabra que está siendo analizada se encuentra en el diccionario, por lo cual en ninguno de los casos no ocurriría sobregeneración, lo cual es muy importante ya que le da un nivel de credibilidad mayor al sistema.

Conclusiones, Recomendaciones y Trabajos Futuros

Conclusiones

- El plan académico que se utiliza como entrada al proceso, debe tener una estructura que permita reconocer los elementos necesarios para el análisis, ya que al no poseer una estructura estandarizada la diversidad de contenidos y formato hace imposible aplicar de forma adecuada las técnicas de procesamiento. Para tener una estructura adecuada para el procesamiento se ha construido una interfaz que permita registrar los planes académicos.
- Para realizar el análisis a nivel de texto se eligió en base al ámbito del proyecto contar con un diccionario que permita tener un dominio del contexto al cual se va a realizar el preprocesamiento. En el caso concreto de este proyecto tener como base las palabras de mayor relevancia de cada plan académico.
- Dada la naturaleza de los planes académicos y las técnicas disponibles para realizar el análisis de contenidos se debe utilizar técnicas que tengan que ver con la segmentación de palabras, frases y contenidos; en vista de la necesidad de realizar el análisis de dichos contenidos a mayor profundidad.
- Las métricas de cobertura, precisión y sobregeneración permitieron determinar un nivel aceptable al proceso de extracción aplicado a los planes académicos. Estas métricas son proporcionadas por MUC, conferencia en la cual se definen métodos para evaluación de sistemas de extracción de información.

Recomendaciones

- Para optimizar el proceso de registro de planes académicos externos es recomendable digitalizar y definir una misma estructura de organización de contenidos para todos los planes.
- Es importante considerar al momento de seleccionar una librería o herramienta para procesamiento de lenguaje natural, de entre sus funcionalidades, verificar el/los idiomas a los cuales brinda soporte.
- Es recomendable que se realice un análisis ortográfico a los contenidos de los planes académicos que van a ser analizados, para obtener buenos resultados en el proceso de extracción de contenidos importantes.
- Para el proceso de evaluación de resultados es recomendable buscar métodos que se encuentren dedicados exclusivamente a la evaluación de sistemas de extracción de información.

Trabajos Futuros

A partir de la construcción del Preprocesador de Planes Académicos, se han generado algunos proyectos que pueden ser desarrollados en un futuro, en el ámbito del procesamiento de lenguaje natural y la extracción de información, entre los que tenemos:

- Reconocimiento automático de contenidos de los planes académicos externos ingresados a la UTPL.
- Alimentación de diccionarios a través de la web.
- Mejorar el tiempo de respuesta en el preprocesamiento de planes académicos, mediante el uso de paralelización del análisis de los contenidos.
- Utilizar otros métodos y técnicas para la obtención de contenidos importantes como son: ontologías, minería de texto, etc.

ANEXO A
Etiquetas Eagles

A.1 Etiquetas Eagles

Las Etiquetas Eagles, elaboradas por el grupo Eagles, fueron creadas para la anotación morfosintáctica de lexicones y corpus para todas las lenguas europeas. Estas etiquetas son comúnmente utilizadas en los analizadores morfológicos para el castellano, debido a que utiliza un conjunto de etiquetas para representar la información morfológica de las palabras. (Group, 2009)

En la Tabla A-1 se muestran cada una de las Etiquetas Eagles, utilizadas en este proyecto de tesis.

ETIQUETAS EAGLES			
Tipo	Descripción	Tipo	Descripción
<i>ADJECTIVE</i>		PN	Pronoun, Numeral
AO	Adjective, Ordinal	PP	Pronoun, Personal
AQ	Adjective, Qualificator	PR	Pronoun, Relative
<i>CONJUNCTION</i>		PT	Pronoun, Interrogative
CC	Conjunction, Coordinate	PX	Pronoun, Possessive
CS	Conjunction, Subordinative	<i>ADVERB</i>	
<i>DETERMINER</i>		RG	Adverb, General
DA	Determiner, Article	RN	Adverb, Negative
DD	Determiner, Demonstrative	<i>ADPOSITON</i>	
DE	Determiner, Exclamatory	SP	Adposition, Preposition
DI	Determiner, Indefinite	<i>VERB</i>	
DN	Determiner, Numeral	VAG	Verb, Auxiliar, Gerund
DP	Determiner, Possessive	VAI	Verb, Auxiliar, Indicative
DT	Determiner, Interrogative	VAM	Verb, Auxiliar, Imperative
<i>PUNCTUATION</i>		VAN	Verb, Auxiliar, Infinitive
Faa	Punctuation, ;	VAP	Verb, Auxiliar, Participle
Fat	Punctuation, !	VAS	Verb, Auxiliar, Subjunctive
Fc	Punctuation, ,	VMG	Verb, Main, Gerund
Fd	Punctuation, :	VMI	Verb, Main, Indicative
Fe	Punctuation, "	VMM	Verb, Main, Imperative
Fg	Punctuation, -	VMN	Verb, Main, Infinitive
Fh	Punctuation, /	VMP	Verb, Main, Participle
Fia	Punctuation, ¿	VMS	Verb, Main, Subjunctive
Fit	Punctuation, ?	VSG	Verb, Semi-Auxiliar, Gerund
Fp	Punctuation, .	VSI	Verb, Semi-Auxiliar, Indicative
Fpa	Punctuation, (VSM	Verb, Semi-Auxiliar, Imperative
Fpt	Punctuation,)	VSN	Verb, Semi-Auxiliar, Infinitive

Fs	Punctuation, ...	VSP	Verb, Semi-Auxiliar, Participle
Fx	Punctuation, ;	VSS	Verb, Semi-Auxiliar, Subjunctive
Fz	Punctuation, other than those	DATE TIMES	
INTERJECTION		W	Date Times
I	Interjection	UNKNOWN	
NOUN	Sustantivo	X	Unknown
NC	Noun, Common	ABBREVIATION	
NP	Noun, Proper	Y	Abbreviation
PRONOUN		NUMBERS	
P0	Pronoun, Clitic	Z	Figures
PD	Pronoun, Demonstrative	Zm	Currency
PE	Pronoun, Exclamatory	Zp	Percentage
PI	Pronoun, Indefinite		

Tabla A-1 Etiquetas Eagles

ANEXO B
Fase de Inicio

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Modelo de Procesos de Negocio	Fecha: 27/08/2012
PPA-MPN	

B.1. Modelo de Procesos de Negocio

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/08/27	1.0	Elaboración de la Plantilla para conocer el modelo de proceso de gestión académica de convalidación y/u homologación de asignaturas realizadas por la UTPL.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Modelo de Procesos de Negocio	Fecha: 27/08/2012
PPA-MPN	

1. Introducción

La Universidad Técnica Particular de Loja, dentro de los procesos de gestión académica que lleva a cabo se encuentra la convalidación de asignaturas para los estudiantes que han cursado programas académicos en otras instituciones educativas.

Se entiende por convalidación de estudios a la “declaración de equivalencias entre los contenidos temáticos de los programas de cursos, ciclos, módulos, seminarios o asignaturas, impartidas por las distintas unidades académicas de las universidades o escuelas politécnicas, efectuada por el órgano competente, previo análisis comparativo de los contenidos.” (Ciencias Jurídicas UTPL, 2006).

En la actualidad este proceso es realizado por docentes de cada carrera de la UTPL, en calidad de fiscales, los cuales verifican que los planes académicos correspondientes a las asignaturas cumplen los requerimientos necesarios para ejecutar el proceso anteriormente mencionado. Es por esta razón que es necesario conocer el modelo de proceso de convalidación de asignaturas para en base a este modelo realizar la propuesta de mejora del mismo.

1.1. Objetivo

Conocer al detalle la ejecución del proceso de gestión académica de convalidación de asignaturas que lleva a cabo la UTPL.

1.2. Alcance

El alcance de esta plantilla está centrado exclusivamente en el proceso de gestión académica de convalidación de asignaturas que realiza UTPL, así como también la presentación de una propuesta de mejora del mismo hasta la fase de revisión de los planes académicos (preprocesamiento) por parte del fiscal asignado para que lleve el proceso. Cualquier proceso y/o actividad que no haya sido mencionado anteriormente, se lo considera fuera del alcance de esta plantilla.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Modelo de Procesos de Negocio	Fecha: 27/08/2012
PPA-MPN	

1.3. Definiciones, Acrónimos y Abreviaturas

- UTPL: Universidad Técnica Particular de Loja
- PPA: Preprocesador de Planes Académicos
- MPN: Modelo de Procesos de Negocio

1.4. Referencias

- Ciencias Jurídicas UTPL [Blog Internet]. Loja. 2006 – citado [2012-julio-10]. Disponible en: http://www.utpl.edu.ec/juridicas/u_academica.php?pag=3

1.5. Resumen

Como se ha indicado anteriormente en este documento se realizará una investigación y análisis sobre el modelo del proceso actual de gestión de convalidación y homologación de asignaturas con la finalidad de presentar una propuesta de mejora del proceso, enfocándose exclusivamente en realizar la fase de preprocesamiento del plan académico.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Modelo de Procesos de Negocio	Fecha: 27/08/2012
PPA-MPN	

2. Modelo de Procesos de Negocio Actual

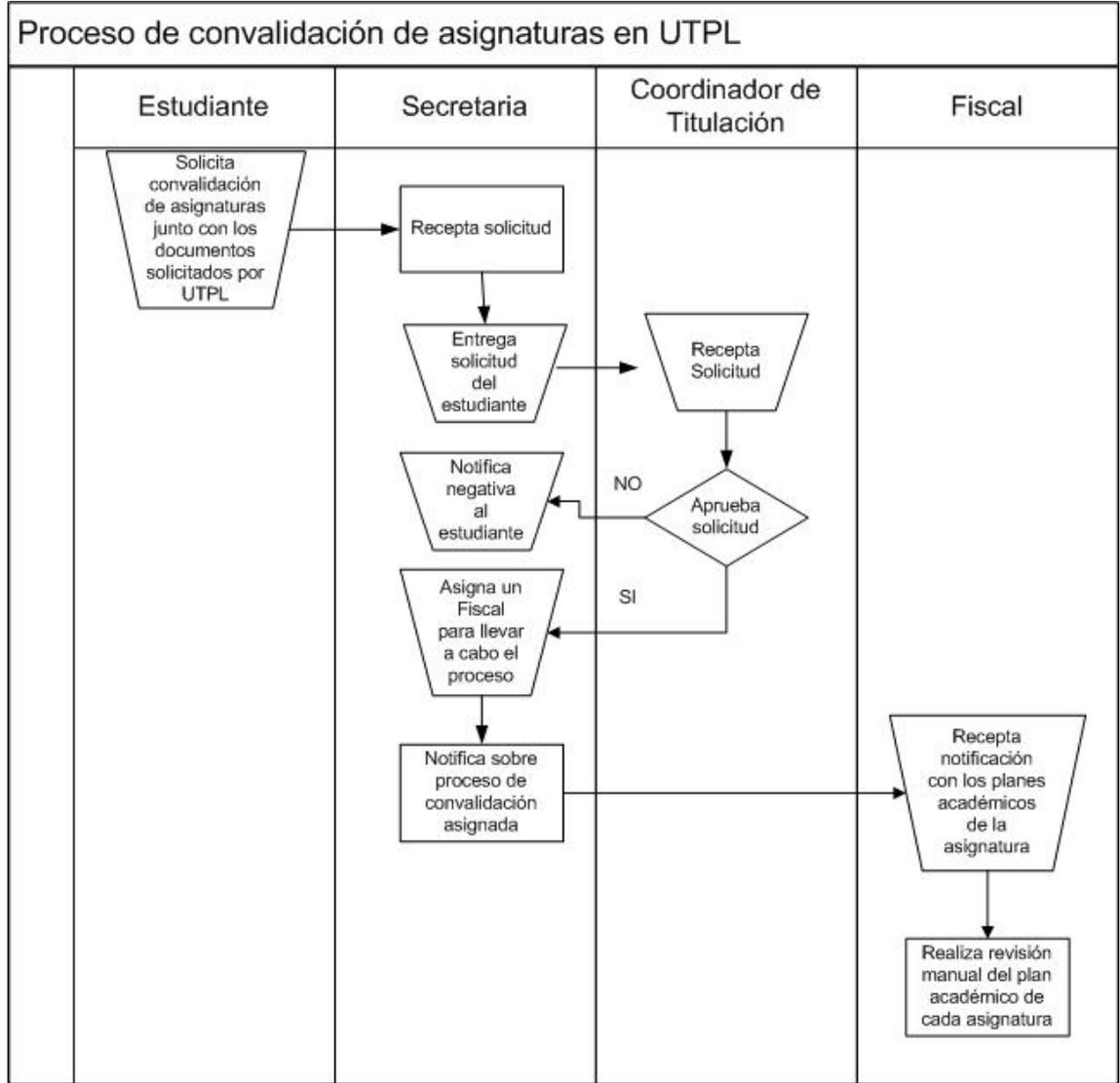


Figura. B-1 Proceso Actual de Convalidación de asignaturas en UTPL

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Modelo de Procesos de Negocio	Fecha: 27/08/2012
PPA-MPN	

3. Modelo de Procesos de Negocio Propuesto

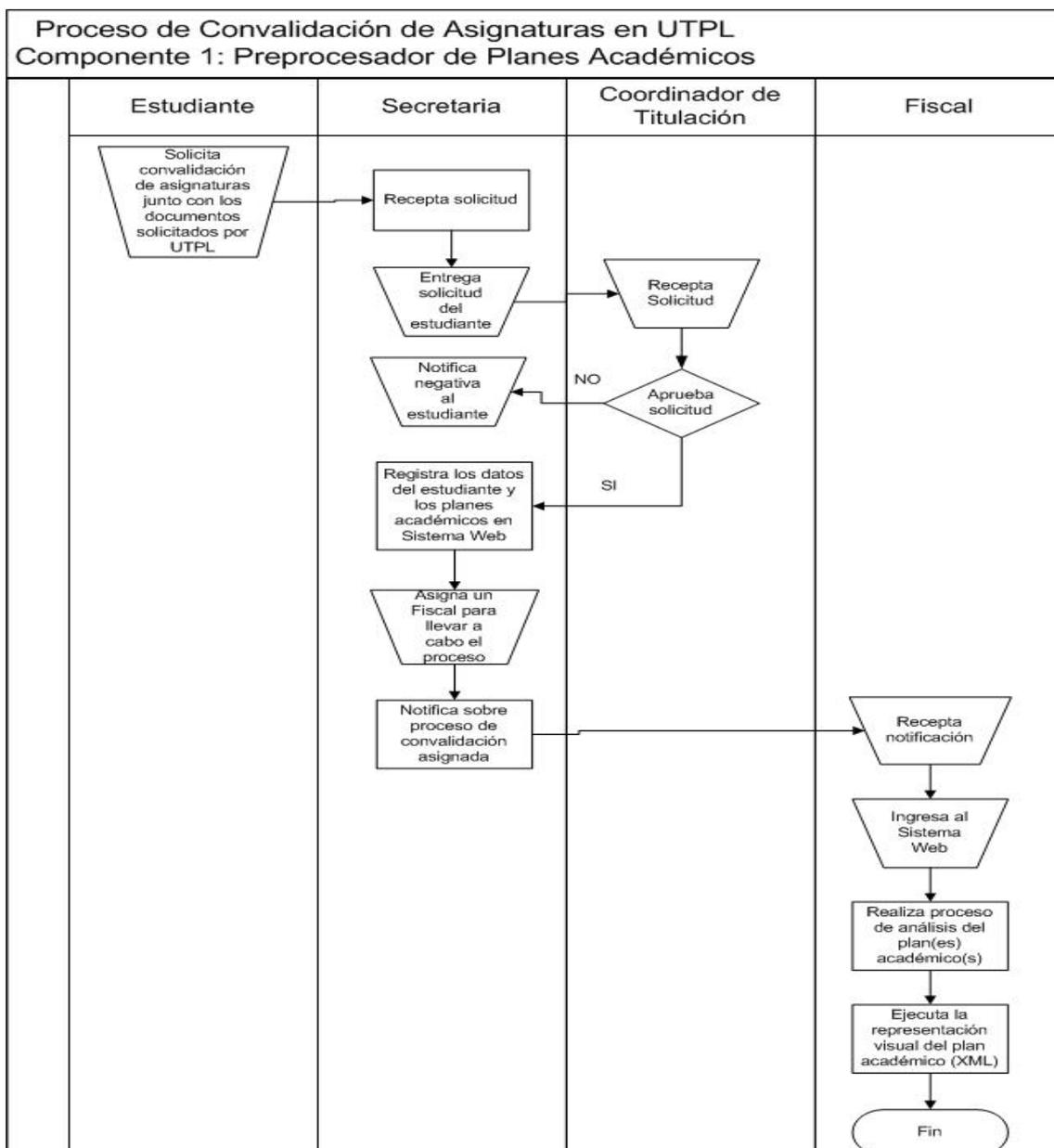


Figura. B-2 Propuesta para Proceso de Convalidación de asignaturas en UTP

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Situación Actual de la Organización	Fecha: 27/08/2012
PPA-SAO	

B.2. Situación Actual de la Organización

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/08/27	1.0	Elaboración de la Plantilla sobre la situación actual de la UTPL con respecto al proceso de gestión académica de convalidación u homologación de asignaturas.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Situación Actual de la Organización	Fecha: 27/08/2012
PPA-SAO	

1. Introducción

La Universidad Técnica Particular de Loja, dentro de los procesos de gestión académica que lleva a cabo se encuentra la convalidación y homologación de asignaturas para los estudiantes que han cursado programas académicos en otras instituciones educativas así como también para aquellos que se encuentran dentro de la universidad, respectivamente.

En la actualidad este proceso es realizado por docentes de la UTPL, en calidad de fiscales, los cuales verifican que los planes académicos correspondientes a las asignaturas cumplen los requerimientos necesarios para ejecutar los procesos anteriormente mencionados.

1.1.Objetivo

Conocer al detalle la ejecución del proceso de gestión académica de convalidación de asignaturas que lleva a cabo la UTPL.

1.2.Alcance

El alcance de esta plantilla está centrado únicamente en el proceso de gestión académica de convalidación de asignaturas que realiza UTPL. Cualquier proceso fuera del antes mencionado se lo considera fuera del alcance de esta plantilla.

1.3.Definición, Acrónimos y Abreviaturas

- UTPL: Universidad Técnica Particular de Loja
- PPA: Preprocesador de Planes Académicos
- SAO: Situación Actual de la Organización.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Situación Actual de la Organización	Fecha: 27/08/2012
PPA-SAO	

1.4. Resumen

Como se ha indicado anteriormente en este documento se mostrará el resultado de la investigación realizada sobre la forma en que UTPL gestiona el proceso académico de convalidación de asignaturas, para lo cual se iniciará conociendo el contexto de manera general al que pertenece UTPL así como las ideas de negocio que poseen y las estrategias que se pueden utilizar en base al contexto del proyecto a desarrollar, para finalmente evaluar los problemas que posee la organización, cuales son las oportunidades de mejora y analizar las herramientas de software que puedan ser útiles en la obtención de mejoras y resolución de problemas.

2. Contexto del Negocio

La Universidad Técnica Particular, es una de las más grandes entidades educativas a nivel nacional, la cual acoge a una gran cantidad de estudiantes, en las modalidades de estudios que son: abierta, a distancia, semipresencial y presencial.

La UTPL lleva a cabo una serie de procesos de gestión académica orientados a resolver y mejorar los servicios que ofrece esta entidad educativa a la colectividad. Uno de estos procesos es la convalidación manual de asignaturas en base a los planes académicos de cada una de las asignaturas que posee la oferta académica vigente de la Universidad.

3. Ideas del Negocio y Estrategias en el contexto del proyecto

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Situación Actual de la Organización	Fecha: 27/08/2012
PPA-SAO	

La Universidad Técnica Particular de Loja, en su continuo avance y desarrollo tecnológico, para ofrecer un mejor servicio a los estudiantes de la entidad educativa, se ve en la necesidad de construir un Sistema Semi-Automático de convalidaciones, para realizar el trabajo de una manera más rápida, eficiente y segura. Este sistema estará conformado por una serie de componentes, siendo uno de los principales el Preprocesador de Planes Académicos en donde se realizará el procesamiento inicial (selección de contenidos importantes) de los planes académicos de las asignaturas a gestionar.

4. Evaluación de las conclusiones obtenidas del Negocio

A continuación se listan los problemas y las oportunidades que se pudo reconocer en la investigación de la situación actual de la Universidad Técnica Particular de Loja.

4.1. Problemas

El proceso de convalidación de asignaturas es un proceso extenso debido a que el fiscal encargado de llevar a cabo esta gestión debe manualmente verificar el grado de similitud que presenta el plan académico de una asignatura frente al plan académico de una asignatura igual o semejante que se encuentre vigente en la oferta académica de la UTPL. Por lo tanto el principal problema encontrado es que se trata de un proceso que es llevado a cabo de forma manual y que por sus características hace que este proceso pueda ser culminado en significativas cantidades de tiempo, lo cual ocasiona que este servicio ofrecido por UTPL no sea óptimo en cuanto a tiempo de respuesta se refiere.

4.2. Oportunidades

Las oportunidades de mejora para el proceso de convalidación de asignaturas de UTPL, es la creación de un sistema informático que sea capaz de llevar a cabo las funciones realizadas por los fiscales de la universidad, con lo cual se podrá mejorar este servicio ofrecido dándole un mayor grado de certeza tanto al estudiante como a la universidad de que el proceso de convalidación fue el

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Situación Actual de la Organización	Fecha: 27/08/2012
PPA-SAO	

correcto, así como también mejorando el tiempo de respuesta de la solicitud de este servicio.

4.3.Nuevas Tecnologías aplicables

La parte tecnológica aplicable a este proyecto es la construcción de un sistema informático automático o semi-automático compuesto por una serie de funcionalidades capaces de llevar a cabo el proceso de convalidaciones de asignaturas para la UTPL.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Visión del Sistema	Fecha: 29/08/2012
PPA-VIS	

B.3. Visión del Sistema

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/08/29	1.0	Elaboración de la Plantilla sobre la visión del producto de software a desarrollar referente al preprocesador de planes académicos de la UTPL.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Visión del Sistema	Fecha: 29/08/2012
PPA-VIS	

1. Introducción

La Universidad Técnica Particular de Loja, dentro de los procesos de gestión académica que lleva a cabo se encuentra la convalidación de asignaturas para los estudiantes que han cursado programas académicos en otras instituciones educativas.

En la actualidad este proceso es realizado de forma manual por docentes de la UTPL, en calidad de fiscales, los cuales verifican que los planes académicos correspondientes a las asignaturas cumplen los requerimientos necesarios para ejecutar los procesos anteriormente mencionados.

Para mejorar el desarrollo de este proceso se solicita la creación de un Sistema Semi-automático de convalidaciones de planes académicos, el cual se encuentra dividido en varios componentes uno de ellos el Preprocesador de planes académicos.

Con la finalidad de tener un entendimiento claro sobre el componente a construir es necesario desarrollar la siguiente plantilla en donde se definirá una visión específica y detallada del mismo, para obtener un producto que satisfaga las necesidades y requerimientos de UTPL.

1.1. Objetivo

- Desarrollar una visión clara sobre el producto deseado, de tal forma que satisfaga las necesidades de UTPL.
- Identificar las personas afectadas/involucradas con el producto de software y sus necesidades.

1.2. Alcance

El alcance de esta plantilla está centrado únicamente al componente denominado Preprocesador de Planes Académicos, el cual es parte de Sistema Semi-

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Visión del Sistema	Fecha: 29/08/2012
PPA-VIS	

automático de convalidaciones de Planes Académicos. Cualquier proceso o componente que no se encuentre dentro del antes mencionado, se lo considera fuera del alcance de esta plantilla.

1.3. Definiciones, Acrónimos y Abreviaturas

- UTPL: Universidad Técnica Particular de Loja
- PPA: Preprocesador de Planes Académicos
- VIS: Visión del Sistema

2. Posicionamiento del Producto

2.1. Definición del Problema

En esta sección se define el problema, los afectados, el impacto y el planteamiento de una solución exitosa. En la Tabla B-1, se observa la definición de estos campos.

El problema de:	El proceso de gestión académica de convalidación de asignaturas llevado por la UTPL, es realizado por un grupo de docentes con rol de fiscales los cuales revisan manualmente la documentación presentada (planes académicos, calificaciones, etc.) y en base a un conjunto de reglas y normas dictaminar si una asignatura puede o no ser convalidada.
Afecta a:	<ul style="list-style-type: none"> • Dirección General Académica • Coordinadores de Titulación • Fiscales de UTPL • Secretarías • Estudiantes de UTPL • Potenciales estudiantes de UTPL
Cuyo impacto es:	<ul style="list-style-type: none"> • Al ser un proceso realizado totalmente de forma manual hace que tener un resultado de este proceso tome largos períodos de tiempo. • La gestión documental para este proceso es

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Visión del Sistema	Fecha: 29/08/2012
PPA-VIS	

	<p>muy extensa.</p> <ul style="list-style-type: none"> • Dominio profundo de las normas y reglas establecidas para el proceso de convalidación de asignaturas.
Una solución exitosa es:	<ul style="list-style-type: none"> • Construir un sistema semi-automático de convalidación de asignaturas el cual está conformado por varios componentes que realizan el proceso manual que se realiza actualmente. • Uno de los componentes antes mencionados consistiría en la construcción de una solución Web supervisada que realice el Preprocesamiento de Planes Académicos, la cual sea extensible, amigable y ágil en donde los usuarios puedan: ingresar los planes académicos externos a la UTPL, realizar el análisis y selección de contenidos importantes de planes académicos tanto internos como externos para que finalmente realice una representación estructurada de los contenidos seleccionados por el producto de software.

Tabla B-1 Definición del Problema

2.2. Posicionamiento del Producto

En esta sección se realiza el posicionamiento del producto, es decir se describe para quien es el producto, que es lo que va a realizar para finalmente mostrar cual es la diferencia con el proceso actual. En la Tabla B-2, se observa la definición de estos campos.

Para:	<ul style="list-style-type: none"> • Dirección General Académica • Coordinadores de Titulación • Fiscales de UTPL • Secretarías
Quién(es):	<p>Dirección General Académica</p> <ul style="list-style-type: none"> • Dependencia responsable general del proceso de convalidación de asignaturas de UTPL. <p>Coordinadores de Titulación</p> <ul style="list-style-type: none"> • Son los responsables del proceso de

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Visión del Sistema	Fecha: 29/08/2012
PPA-VIS	

	<p>convalidación de asignaturas dentro de cada titulación de la universidad</p> <p>Secretarias</p> <ul style="list-style-type: none"> Realizan el proceso de recepción de solicitudes y documentación necesaria para la convalidación de asignaturas. <p>Fiscales UTPL</p> <ul style="list-style-type: none"> Realizan el estudio y análisis de los planes académicos de cada una de las asignaturas para dictaminar al final del proceso si estas poseen las características reglamentarias para ser convalidadas por alguna asignatura perteneciente a la oferta académica de la titulación con la cual está siendo objeto de análisis.
Que	<ul style="list-style-type: none"> Permite acceso online a los Fiscales y a las secretarias de UTPL, para el registro y análisis de los planes académicos. Realiza el registro de los planes académicos externos de la UTPL. Permite ejecutar un análisis morfológico de las palabras que conforman el plan académico, con la finalidad de realizar la extracción de contenidos importantes basándose en un corpus especializado dependiendo de la asignatura. Muestra un diagrama estructurado (XML) con los contenidos importantes extraídos.
A diferencia:	<p>Del procedimiento actual:</p> <ul style="list-style-type: none"> Proceso manual Ocupa largos periodos de tiempo Gestión documental muy extensa. Identificación manual de contenidos importantes.

Tabla B-2 Posicionamiento del Producto

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Visión del Sistema	Fecha: 29/08/2012
PPA-VIS	

3. Resumen de los afectados/involucrados

En la Tabla B-3, se observan los afectados o involucrados en el proyecto, así como también la descripción de estos y sus responsabilidades.

Nombre	Descripción	Responsabilidades
Dirección General Académica	Dependencia encargada de forma general del proceso de convalidación de asignaturas de UTPL.	Es el responsable general de la UTPL del proceso de convalidación y homologación de asignaturas
Coordinadores de Titulación	Quién está encargado de evaluar la solicitud de convalidación de asignaturas y de indiciar el inicio de ejecución de dicho proceso.	Aprobar o reprobar solicitudes de convalidación de asignaturas.
Fiscales de UTPL	Docentes responsables de estudiar, analizar y evaluar cada uno de los planes académicos de cada asignatura.	Analizar los planes de las asignaturas a convalidar, para obtener los contenidos más relevantes y en base a un conjunto de reglas establecidas por UTPL, poder aprobar o rechazar la solicitud de convalidación solicitada por el estudiante.
Secretarias	Persona encargada de receptor las solicitudes de convalidación de asignaturas. Además asigna el fiscal que va a realizar la revisión de los planes académicos.	- Entrega de solicitudes de convalidación a coordinadores de titulación. - Asignación de fiscal
Estudiantes de UTPL	Profesionales en formación los cuales solicitan la homologación de asignaturas.	Presentan solicitud de homologación de asignaturas a la secretaria de la titulación correspondiente de la UTPL.
Potenciales estudiantes de UTPL	Estudiantes de otras entidades educativas de nivel superior, que desean convalidar las asignaturas tomadas en otros centros educativos con las de la malla curricular de la UTPL.	Presentan solicitud de convalidación de asignaturas a la secretaria de la titulación correspondiente de la UTPL.

Tabla B-3 Resumen de afectados

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Visión del Sistema	Fecha: 29/08/2012
PPA-VIS	

4. Resumen de Usuarios

En la Tabla B-4, se observan la descripción de los usuarios que harán uso del sistema.

Nombre	Descripción
Secretarías	Realizan el proceso de recepción de solicitudes y documentación necesaria para la convalidación de asignaturas.
Fiscales	Realizan el estudio y análisis de los planes académicos de cada una de las asignaturas para dictaminar al final del proceso si estas poseen las características reglamentarias para ser convalidadas u homologadas por alguna asignatura perteneciente a la oferta académica de la titulación con la cual está siendo objeto de análisis.

Tabla B-4 Resumen de usuarios

5. Necesidades de los afectados/usuarios

En la Tabla B-5 y Tabla B-6, se describen las necesidades, preocupaciones y soluciones propuestas a las necesidades que poseen los usuarios descritos en la sección anterior.

5.1. Necesidades de las secretarías

Necesidad	Prioridad	Solución Actual	Soluciones Propuestas	Preocupación
Organizar la documentación presentada para la convalidación de asignaturas.	Media	Los estudiantes deben de acercarse a la secretaría de cada titulación con la solicitud de convalidación de los planes académicos de cada una de estas	Proveer de un sistema web de registro de planes académicos externos a la UTPL, en donde una vez q haya sido aceptada la solicitud de convalidación, este plan sea registrado en el sistema	Que el registro de los planes académicos tome bastante cantidad de tiempo

Tabla B-5 Necesidades de las secretarías

5.2. Necesidades de los fiscales

Necesidad	Prioridad	Solución Actual	Soluciones Propuestas	Preocupación
-----------	-----------	-----------------	-----------------------	--------------

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Visión del Sistema	Fecha: 29/08/2012
PPA-VIS	

Automatizar forma de analizar el contenido los planes académicos de las asignaturas	Alta	Los planes académicos son analizados manualmente por los fiscales	Construir un sistema web semi – automático para que analice morfológicamente el contenido del plan académico y pueda así extraer los contenidos más importantes de dicho plan.	Que los contenidos extraídos no sean los más relevantes del plan académico.
---	------	---	--	---

Tabla B-6 Necesidades de los fiscales

6. Resumen del Producto

6.1. Perspectiva del producto

Con la finalidad de tener una visión más clara del producto de software a desarrollar, en la B-7, se describen brevemente los módulos que conforman el Preprocesador de Planes Académicos.

Componente	Módulo	Descripción
Preprocesador de Planes Académicos de UTPL	Login	Autentifica el acceso a los usuarios al sistema y muestra los módulos dependiendo del rol al que pertenezca.
	Registro de Planes	Registra: <ul style="list-style-type: none"> - La información general de los planes académicos externos a la UTPL. - Contenidos del plan. - Archivo digital del plan.
	Análisis de Planes	<ul style="list-style-type: none"> - Realiza la búsqueda del plan académico que se requiere analizar. - Analiza el contenido del plan académico para extraer las palabras importantes. - Si algún contenido no fue tomado como importante por la herramienta, el fiscal, según

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Visión del Sistema	Fecha: 29/08/2012
PPA-VIS	

		<p>su criterio puede seleccionar los contenidos que hayan quedado fuera de la selección.</p> <ul style="list-style-type: none"> - Almacena la información extraída en la base de datos. - Genera un archivo de texto con toda la información del proceso de extracción de contenidos. - Genera un gráfico en base a los contenidos importantes extraídos.
	Administración – Diccionario	<ul style="list-style-type: none"> - Creación de nuevas entradas al diccionario compuesto por un corpus y su respectiva lematización.
	Administración – Usuarios	<ul style="list-style-type: none"> - Crear y eliminar usuarios.
	Administración – Incrementar valores de campos	<ul style="list-style-type: none"> - Definición de centros UTPL, entidades de educación superior y carreras, correspondiente al formulario de registro de planes externos.

Tabla B-7 Perspectiva del producto

6.2. Resumen de capacidades

- **Beneficios del Producto**
 - Permite acceso online a fiscales y a secretarías de UTPL, para el registro y análisis de los planes académicos.
 - Realiza el registro de los planes académicos externos de la UTPL.
 - Realiza la extracción de contenidos importantes basándose en un corpus especializado dependiendo de la asignatura.
 - Permite administrar nuevos usuarios, centros UTPL, entidades de educación superior y carreras, así mismo crear nuevas entradas al diccionario.

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Visión del Sistema	Fecha: 29/08/2012
PPA-VIS	

- Visualización de un diagrama estructurado (XML) con los contenidos importantes extraídos.

- **Características del Producto**

- Facilidad para registrar la información y contenidos de los planes académicos.
- Facilidad para analizar los planes académicos.
- Tiempo de análisis más rápido que el realizado de forma manual.
- Mayor precisión en la extracción de contenidos importante.
- Facilidad para tener una visión amplia de todos los contenidos que posee el plan académico.

6.3. Restricciones

- No incluye la generación de archivos de tipo RDF.
- El sistema no permite la creación de nuevos roles, ni la actualización de información de los usuarios registrados.
- La creación del diccionario (corpus y lematización) es realizado de forma manual.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

B.4. Especificación de Requerimientos

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/08/31	1.0	Este documento define de forma precisa, cuales son los requerimientos para la construcción del preprocesador de planes académicos de la UTPL, el cual está basado en el documento de visión del proyecto en donde se encuentran especificadas las necesidades de los usuarios.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

1. Introducción

En este documento se realiza la definición de los requerimientos solicitados para la construcción del preprocesador el plan académico, basándose en el documento visión realizado en el Anexo B de este documento.

Uno de los principales requerimientos es la creación de una plataforma web supervisada para el preprocesamiento de planes académicos de la UTPL, que servirá para realizar el análisis de los planes académicos que ingresan a la UTPL, para realizar el proceso de convalidación y homologación de asignaturas, con la finalidad de obtener los contenidos más importantes de dicho plan.

El sistema está compuesto por los siguientes módulos:

- Registro de Planes Académicos.
- Análisis de Planes Académicos, que abarca la extracción de contenidos importantes, generación de archivo de texto (con todo el proceso de obtención de contenidos importantes), generación de archivo en formato XML (contenidos importantes obtenidos por el sistema), gráfico (contenidos importantes obtenidos por el sistema).
- Administración que comprende creación de nuevos usuarios, carreras y asignaturas, así como también la creación de corpus y lematizaciones.

Cada uno de estos módulos serán utilizados por diferentes usuarios que están clasificados en tres grupos:

- Secretaria
- Fiscal
- Administrador

Cada grupo tendrá acceso a determinadas partes del sistema dependiendo de los roles asignados.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

1.1. Propósito

El propósito de este documento es capturar todos los requerimientos de software de alto nivel de los distintos usuarios, para la construcción del preprocesador de planes académicos de la UTPL.

1.2. Ámbito

Este producto de software está enfocado al preprocesamiento de planes académicos, entendiendo por preprocesamiento al análisis morfológico de las palabras y la extracción de los contenidos más importantes de dicho plan.

1.3. Definiciones, acrónimos y abreviaciones

- PPA: Preprocesador de Planes Académicos.
- ESR: Especificación de Requerimientos.
- UTPL: Universidad Técnica Particular de Loja
- Autenticación: Es el proceso de verificar la identidad de una persona frente a un producto de software, es decir es un modo de asegurar que los usuarios que están intentando ingresar son quién ellos dicen ser.

2. Descripción General

La Universidad Técnica Particular de Loja, dentro de los procesos de gestión académica que lleva a cabo se encuentra la convalidación y homologación de asignaturas para los estudiantes que han cursado programas académicos en otras instituciones educativas así como también para aquellos que se encuentran dentro de la universidad, respectivamente.

En la actualidad este proceso es realizado por docentes de la UTPL, en calidad de fiscales, los cuales verifican manualmente que los planes académicos correspondientes a las asignaturas cumplen los requerimientos necesarios para ejecutar los procesos anteriormente mencionados.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

Para mejorar el desarrollo de este proceso se solicita la creación de un Sistema Semi-automático de convalidaciones de planes académicos, el cual se encuentra dividido en varios componentes, tal como se muestra en la Figura B-3.

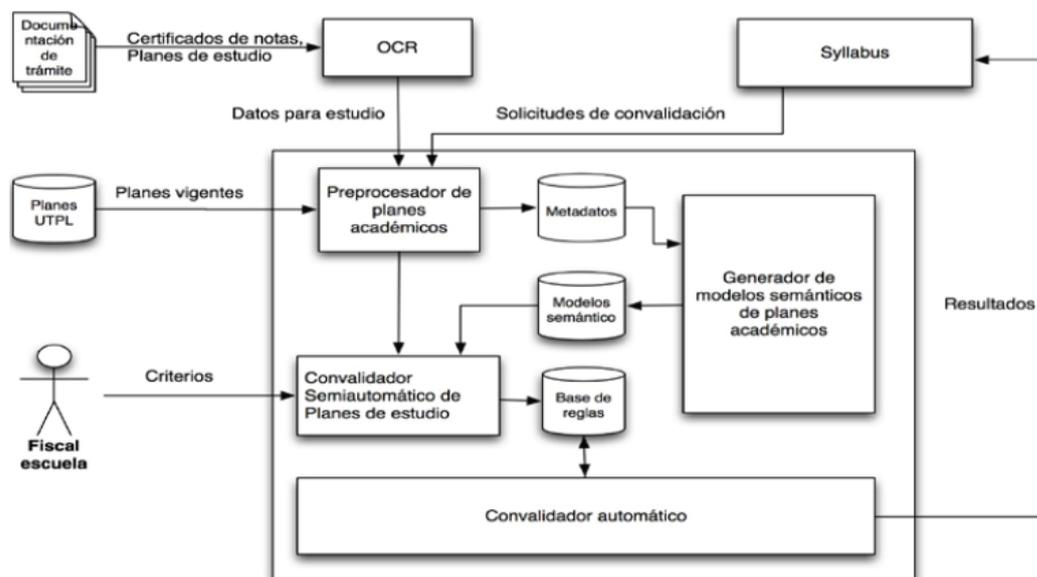


Figura. B-3 Arquitectura del Sistema Semi-automático de convalidaciones de Planes Académicos

El componente para el cual se van a especificar los requerimientos es el Preprocesador de Planes Académicos, el cual es un sistema web semi – automático, que debe ser extensible, amigable y ágil, en donde los usuarios puedan: ingresar los planes académicos externos a la UTPL, realizar el análisis y selección de contenidos importantes de planes académicos tanto internos como externos para que finalmente realice una representación estructurada de los contenidos seleccionados por el producto de software. Esto a diferencia del proceso manual actual en donde la gestión documental es muy extensa y la identificación de contenidos importantes ocupa largos periodos de tiempo, lo cual ocasiona que este servicio al estudiante sea demorado.

A continuación se especifican las funcionalidades, supuestos y dependencias, así como también cada uno de los requerimientos funcionales, no funcionales, de proceso, etc.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

2.1.Asunciones

- El acceso a las características del sistema web se realizará previamente un registro y autenticación de los usuarios
- La extracción de contenidos importantes podrá ser realizado en base al diccionario que posee la aplicación.
- Cada asignatura posee su propio diccionario para evitar ambigüedades.
- Las asignaturas tomadas para realizar el diccionario son las de la UTPL, puesto que para realizar la extracción de contenidos de planes externos e internos, estos deben ser comparados con las materias de la UTPL.

3. Especificación de Requerimientos

3.1.Requerimientos Funcionales

4.7 REQ01 AUTENTICACIÓN INGRESO SISTEMA
<i>Módulo Autenticación</i>
<p><i>Entrada:</i></p> <p>La autenticación para el ingreso de los usuarios al sistema posee los siguientes campos:</p> <ul style="list-style-type: none"> • Usuario (nombre de usuario). • Contraseña (contraseña de ingreso del usuario al sistema).
<p><i>Proceso:</i></p> <p>Ingreso de usuario al sistema.</p> <ul style="list-style-type: none"> • El usuario digita su usuario y contraseña.
<p><i>Salida:</i></p> <p>Autenticación de usuario para ingreso al sistema.</p> <ul style="list-style-type: none"> • Si el usuario ingreso el usuario y contraseña de forma correcta, se procederá a cargar el sistema en base a su perfil (secretaria, fiscal, administrador). • Si el usuario ingresó el usuario y/o contraseña de forma incorrecta, el sistema volverá a cargar la pantalla de autenticación de ingreso al sistema. • Si el usuario deja uno y/o los dos campos vacíos, el sistema volverá a cargar la pantalla de autenticación de ingreso al sistema.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

4.8 REQ02 REGISTRAR PLANES ACADÉMICOS EXTERNOS

Módulo Registro Planes

Entrada:

El registro de los planes académicos se encuentra dividido en tres partes: Subir archivo del plan académico, registrar información general del plan y registrar los contenidos del plan académico.

Para subir el archivo del plan académico, deberá seleccionar el archivo con extensión .pdf, ingresando al botón Seleccionar archivo.

Para registrar la información general del plan académico se utilizará el formulario registro de planes académicos externos, que posee los siguientes campos:

- Centro UTPL (nombre de centro UTPL donde se registra el plan)
- Entidad Educativa (nombre de la entidad educativa a la que pertenece el plan académico)
- Modalidad (tipo de modalidad a la cual pertenece el plan académico)
- Carrera (nombre de la carrera a la que pertenece el plan académico)
- Nombre de la Asignatura (nombre de la asignatura a la que pertenece el plan académico)
- Periodo académico (periodo académico al que pertenece el plan académico)
- Archivo del plan (nombre del archivo del plan académico que fue subido al sistema)

Para registrar los contenidos del plan académico, se utilizará las cajas de texto que se encuentran en el formulario junto al número de unidad de la misma. Para incrementar el número de unidades deberá seleccionar el signo “más” disponible en el formulario y para quitar unidades se seleccionará el signo “menos”.

Proceso:

Subir archivo del plan académico

- La secretaria subirá al sistema el archivo del plan académico (extensión .pdf) que se va a registrar en el sistema.

Registrar información general del plan académico

- La secretaria ingresa los datos del formulario: centro UTPL, entidad educativa, modalidad, carrera, nombre de la asignatura

Registrar los contenidos del plan académico.

- La secretaria ingresa los contenidos de cada una de las unidades que dispone el plan académico externo que se está registrando.

Salida:

Subir archivo del plan académico

- Si la secretaria elige un archivo que no sea de la extensión permitida, se

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

mostrará una alerta indicando que el archivo seleccionado no es permitido.

- Alerta tipo de archivo no permitido
- Si la secretaria no elige un archivo, el sistema cargará nuevamente la misma página.
- Si el archivo subido por la secretaria es el correcto, se muestra la página del formulario de registro del plan académico.

Registrar información general y contenidos del plan académico

- Si la secretaria ingresó todos los campos del formulario de registro de información general así como también al menos un contenido del plan académico, se presentará una alerta de que los datos se han guardado correctamente y regresa a la pantalla de inicio.
- Alerta de que los datos se han guardado correctamente.
- Si la secretaria no ingreso ningún nombre de asignatura y/o ningún contenido en el formulario, se muestra una alerta, en donde se indique que se debe ingresar un nombre de asignatura y por lo menos un contenido para una unidad.
- Alerta de que no se debe dejar en blanco el espacio asignatura y el espacio de contenidos.
- Si al momento de guardar los datos del plan académico, ocurriera un error, se mostrará una alerta de que ocurrió un problema al guardar los datos.
- Alerta, problema al guardar los datos.

4.9 REQ03 ANALIZAR PLAN ACADÉMICO

Módulo Análisis

Entrada:

El proceso de análisis de los planes académicos se encuentra dividido en dos partes: Buscador del plan académico a analizar y ejecución de análisis.

Para buscar un plan académico se deberá utilizar el buscador que posee el sistema a través del empleo del código del plan académico que se desea analizar. Para la ejecución del análisis se deberá tener en cuenta la carrera y la asignatura al que pertenece el plan académico de UTPL, con el cual se desea realizar la comparación y extracción de contenidos importantes.

Proceso

Buscar plan académico.

El fiscal que lleva a cabo el proceso, ingresa el código del plan, en el buscador de planes académicos.

Ejecución de análisis.

El fiscal elige la carrera y asignatura correspondiente al plan académico de UTPL con el cual se requiere realizar la extracción de contenidos importantes.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

Salida

Buscar plan académico.

- Si el código ingresado por el fiscal es el correcto, se muestra la siguiente información del plan académico.
 - Nombre de Asignatura
 - Centro UTPL
 - Universidad
 - Modalidad
 - Carrera.
- Si el código ingresado por el fiscal es incorrecto o no existe, muestra un mensaje indicando lo ocurrido
- Mensaje indicando que el Plan Académico no se ha encontrado y/o el código ingresado es incorrecto.
- Si el fiscal no ingresa ningún código y ejecuta la búsqueda, muestra un mensaje indicando lo ocurrido
- Mensaje indicando que debe ingresar un código de plan académico.

Ejecución de análisis.

- Si el fiscal ha seleccionado la carrera y la asignatura del plan académico a analizar, se listarán los contenidos importantes extraídos del plan, así como también aquellos contenidos que no fueron elegidos por la herramienta de forma separada. El fiscal en base a su criterio tendrá la capacidad para seleccionar contenidos que no hayan sido extraídos por la herramienta y colocarlos en la lista de contenidos seleccionados. El listado final de contenidos importantes se almacenará finalmente en la base de datos, mostrando un mensaje de que los contenidos se han almacenado de forma correcta; en caso de que ocurra algún error al momento de la grabación se mostrará un mensaje alertando de lo sucedido.
- Mensaje, los contenidos importantes se almacenaron correctamente.
- Mensaje, no se pudo almacenar los contenidos en la base de datos.
- Si el fiscal no eligió la carrera y/o asignatura correspondiente al plan académico de UTPL, se mostrará un mensaje indicado q estos campos deben ser seleccionados.
- Mensaje señalando que para ejecutar el análisis se debe seleccionar una carrera y asignatura a la que pertenezca el plan académico.

4.10 REQ04 GENERAR ARCHIVO DE TEXTO CON RESULTADO DE ANÁLISIS

Módulo Análisis

Entrada:

Como entrada para este requerimiento, se tiene la salida del requerimiento 3 (REQ03 ANALIZAR PLAN ACADÉMICO).

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

Para generar el archivo de texto con el resultado de todo el proceso de extracción de contenidos importantes del plan académico, se deberá utilizar el campo correspondiente a la generación de archivo de texto.

Proceso

El fiscal deberá ir al campo referente a la generación del archivo de texto, el cual contendrá todo el proceso de extracción de contenidos importantes.

Salida

Una vez que el fiscal seleccionó el campo que indica sobre la generación del archivo de texto (contiene cada uno de los pasos realizados por la herramienta para obtener los contenidos extraídos), se podrá descargar dicho archivo y almacenarlo en cualquier lugar del computador.

4.11 REQ05 GENERAR ARCHIVO XML CON CONTENIDOS IMPORTANTES EXTRAÍDOS

Módulo Análisis

Entrada

Como entrada para este requerimiento, se tiene la salida del requerimiento 3 (REQ03 ANALIZAR PLAN ACADÉMICO).

Para generar el archivo en formato XML, que contenga los contenidos importantes extraídos, se deberá utilizar el campo correspondiente a la generación de archivo XML.

Proceso

El fiscal deberá ir al campo referente a la generación del archivo XML, para que se ejecute este proceso.

Salida

Una vez que el fiscal seleccionó el campo que indica la generación del archivo XML, se podrá descargar dicho archivo y almacenarlo en cualquier lugar del computador. Además esta estructura será almacenada en la base de datos.

4.12 REQ06 GENERAR GRÁFICO

Módulo Análisis

Entrada

Como entrada para este requerimiento, se tiene la salida del requerimiento 5 (REQ05 GENERAR ARCHIVO XML CON CONTENIDOS IMPORTANTES EXTRAÍDOS)

Para obtener el gráfico de los contenidos importantes extraídos, se deberá utilizar el campo correspondiente a la generación del gráfico.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

Mensaje: no se pudo consultar a la base de datos.
<p>Proceso</p> <p>El fiscal deberá ir al campo referente a la generación del gráfico, para que se ejecute este proceso.</p>
<p>Salida</p> <p>Una vez que el fiscal seleccionó el campo que indica sobre la generación del gráfico, se podrá visualizar el gráfico el cual se origina a través de la consulta a la base de datos de la información almacenada en el requerimiento 5.</p> <p>Si el sistema no pudo consultar a la base de datos la información requerida para realizar el gráfico mostrará una alerta del error.</p>

4.13 REQ07 CREAR NUEVO CORPUS Y LEMATIZACIÓN
Módulo Diccionario
<p>Entrada</p> <p>Para crear un nuevo corpus basado en los planes académicos que tiene UTPL, se utilizará un formulario de registro que posee los siguientes campos:</p> <ul style="list-style-type: none"> • Carrera (Nombre de la carrera a la que pertenece el plan académico) • Asignatura (Nombre de la asignatura a la que pertenece el plan académico) • Tipo corpus (Genérico, Específico) • Palabra de corpus (Palabra que se va a agregar a un corpus) • Primera Lematización (Palabra que se va a agregar a la lematización)
<p>Proceso</p> <p>El administrador ingresa los datos del formulario: carrera, asignatura, tipo corpus, palabra del corpus, lematización.</p>
<p>Salida</p> <ul style="list-style-type: none"> • Si el administrador ingresó la información correspondiente en todos los campos del formulario de registro de un nuevo corpus inclusive la primera Lematización, se presentará una alerta de que los datos se han guardado correctamente. • Alerta de que los datos se han guardado correctamente. • Si el administrador no ingreso el valor de alguno de los campos del formulario, se alertará que debe ingresar información en todos los campos del formulario. • Alerta de que no se debe dejar en blanco los espacios del formulario. • Si al momento de guardar el diccionario, ocurriera un error, se mostrará una alerta de que ocurrió un problema al guardar los datos. • Alerta, problema al guardar los datos. • Si la palabra del nuevo corpus y/o lematización, ya se encuentran

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

almacenadas en el diccionario, alertará indicado que esos datos ya se encuentran almacenados en la base de datos.

- Alertar de que los datos ya existen en la base de datos.

4.14 REQ08 ACTUALIZAR LEMATIZACIÓN EXISTENTE
<i>Módulo Diccionario</i>
<i>Entrada</i> Para actualizar la lematización se utilizará un formulario de registro que posee los siguientes campos: <ul style="list-style-type: none"> • Carrera (Nombre de la carrera a la que pertenece la nueva lematización) • Asignatura (Nombre de la asignatura a la que pertenece la nueva lematización) • Palabra de corpus (Palabra del corpus a la que pertenece la nueva lematización) • Nueva lematización (Nueva palabra para la lematización)
<i>Proceso</i> El administrador ingresa los datos del formulario: carrera, asignatura, palabra del corpus, nueva lematización.
<i>Salida</i> <ul style="list-style-type: none"> • Si el administrador ingresó toda la información en los campos del formulario de registro de una nueva lematización, se presentará una alerta de que los datos se han guardado correctamente. • Alerta de que los datos se han guardado correctamente. • Si el administrador no ingreso el valor de alguno de los campos del formulario, se alertará que debe ingresar información en todos los campos del formulario. • Alerta de que no se debe dejar en blanco los espacios del formulario. • Si al momento de guardar el diccionario, ocurriera un error, se mostrará una alerta de que ocurrió un problema al guardar los datos. • Alerta, problema al guardar los datos. • Si la nueva palabra de la lematización, ya se encuentran almacenada en el diccionario, se alertará indicado que esos datos ya se encuentran almacenados en la base de datos. • Alertar de que los datos ya existen en la base de datos.

4.15 REQ09 CREAR NUEVAS CARRERAS-UTPL
<i>Módulo Diccionario</i>
<i>Entrada</i>

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

<p>Para crear una nueva carrera de la UTPL, se realizará ingresando el siguiente campo: Nombre de Carrera (Nombre de la nueva carrera que se desea agregar).</p>
<p>Proceso El administrador del sistema ingresa el nombre de la carrera a registrar.</p>
<p>Salida</p> <ul style="list-style-type: none"> • Si la nueva carrera se registró correctamente en el sistema, se presentará una alerta de confirmación del registro. • Alerta de confirmación de carrera registrada. • Almacenamiento en la base de datos, del nombre de la carrera registrada en el sistema. • Si el nombre de la carrera que se está registrado ya ha sido registrada, presentará una alerta indicando que esa carrera ya fue registrada. • Alerta de que la carrera ya ha sido registrada anteriormente. • Si la nueva carrera no pudo registrarse por problemas con el servidor de BD o servidor de aplicación, etc., se presentará una alerta de error al almacenar la información • Alerta de error al almacenar la información. • Si no se ingresa ningún valor en el campo de: nombre carrera, se mostrará una alerta indicando que se debe ingresar el nombre de la carrera. • Alerta de que se debe ingresar el nombre de la carrera que se desea ingresar.

<p>4.16 REQ10 CREAR NUEVAS ASIGNATURAS-UTPL</p>
<p>Módulo Diccionario</p>
<p>Entrada Para crear una nueva asignatura de la UTPL, se realizará ingresando el siguiente campo: Nombre de Asignatura (Nombre de la nueva asignatura que se desea agregar).</p>
<p>Proceso El administrador elige de entre las carreras registradas, al cual pertenece la asignatura a crear. El administrador del sistema ingresa el nombre de la asignatura a registrar.</p>
<p>Salida</p> <ul style="list-style-type: none"> • Si el administrador elige la carrera e ingresa el nombre la asignatura a registrar, se presentará una alerta de confirmación del registro. • Alerta de confirmación de asignatura registrada. • Almacenamiento en la base de datos, del nombre de la asignatura registrada en el sistema. • Si el nombre de la asignatura que se está registrando ya ha sido registrada, presentará una alerta indicando que esa asignatura ya fue registrada.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

- Alerta de que la asignatura ya ha sido registrada anteriormente.
- Si la nueva asignatura no pudo registrarse por problemas con el servidor de BD o servidor de aplicación, etc., se presentará una alerta de error al almacenar la información
- Alerta de error al almacenar la información.
- Si no se ingresa ningún valor en el campo de: nombre asignatura, se mostrará una alerta indicando que se debe ingresar el nombre de la asignatura.
- Alerta de que se debe ingresar el nombre de la asignatura que se desea ingresar.
- Si el administrador no elige una carrera a la que pertenezca la asignatura y continúa con el proceso, se mostrará una alerta de que debe seleccionar una carrera.
- Alerta de error, tiene que elegir una carrera a la que pertenezca la asignatura.

4.17 REQ11 GESTIONAR USUARIOS
<i>Módulo Usuarios</i>
<p><i>Entrada:</i></p> <p>La Gestión de usuarios se encuentra dividida en dos partes, que son: crear y eliminar usuarios.</p> <p>Para crear un usuario se realizará utilizando el formulario de información básica que posee los siguientes campos:</p> <ul style="list-style-type: none"> • Nombre (Nombre del usuario). • Apellido (Apellido del usuario). • Usuario (Nombre de usuario). • Contraseña (Contraseña de ingreso del usuario al sistema). • Rol (Que puede ser: Secretaria, Fiscal, Administrador). <p>Para eliminar un usuario se realizará listando todos los usuarios registrados en el sistema.</p>
<p><i>Proceso:</i></p> <p>Registro de información en el formulario.</p> <ul style="list-style-type: none"> • El administrador del sistema registra los datos: nombre, apellido, usuario, contraseña y elige un rol de los definidos. <p>Eliminar un usuario registrado en el sistema.</p> <ul style="list-style-type: none"> • El administrador del sistema elige del listado de usuarios, a cuál de ellos desea eliminar.
<p><i>Salida</i></p> <p>Registro de nuevo usuario.</p> <ul style="list-style-type: none"> • Si la información del nuevo usuario se registró correctamente en el

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

<p>formulario, se presentará una alerta de confirmación del registro.</p> <ul style="list-style-type: none"> • Alerta de confirmación de usuario registrado. • Almacenamiento en la base de datos, de los datos del usuario registrado en el sistema. • Si el nombre de usuario elegido por el nuevo usuario que se está registrado ya está siendo utilizado, presentará una alerta indicando que elija otro nombre de usuario. • Alerta de cambio de nombre de usuario. • Si los datos de registro del nuevo usuario no pudieron registrarse por problemas con el servidor de BD o servidor de aplicación, etc., se presentará una alerta de error al almacenar la información • Alerta de error al almacenar la información. <p>Eliminar usuario registrado en el sistema.</p> <ul style="list-style-type: none"> • Si los datos del usuario fueron eliminados correctamente, se presentará una alerta de confirmación de la eliminación del registro del usuario. • Alerta de confirmación de registro de usuario eliminado. • Si los datos del usuario no fueron eliminados correctamente, se presentará una alerta de error al borrar la información. • Alerta de error al eliminar el registro del usuario.

<h4 style="text-align: center;">4.18 REQ12 INCREMENTAR VALORES DE LOS CAMPOS DE REGISTRO DE PLANES</h4>
<p><i>Módulo Actualización de Campos</i></p>
<p><i>Entrada:</i></p> <p>El formulario de registro de planes académicos se encuentra compuesto por cinco campos, los cuales pueden ser actualizados con más información. Para incrementar los valores de los campos del formulario registro de planes académicos, se deberá ir eligiendo el campo en el cual se vayan a incrementar valores. Estos campos son:</p> <ul style="list-style-type: none"> • Centros UTPL (Nombre del centro de UTPL, en el cual se está realizando el registro del plan académico) • Carrera (Nombre de la carrera a la cual pertenezca el plan académico) • Entidades educativas (Nombre de la entidad educativa a la cual pertenece el plan académico) • Modalidades (Tipo de Modalidad de estudios a la que pertenece el plan académico.) • Periodos académicos (Periodo académico en el cual fue trabajado el plan académico)
<p><i>Proceso</i></p> <p>Incremento de valores de los campos del formulario de registro de planes académicos.</p> <ul style="list-style-type: none"> • Centro UTPL: El administrador ingresa el nombre de la ciudad en la cual

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

se encuentra el centro universitario de la UTPL.

- Entidad Educativa: El administrador ingresa el nombre de la entidad educativa que desea registrar.
- Carrera: El administrador elige previamente la entidad educativa a la cual pertenece la carrera que va registrar. A continuación ingresa el nombre de la carrera que desea registrar.
- Modalidad: El administrador ingresa el nombre de la modalidad de estudios que desea almacenar.
- Periodo académico: El administrador ingresa el periodo académicos que desea almacenar.

Salida

Incremento de valores de los campos del formulario de registro de planes académicos.

- Centro UTPL
 - Si el nombre del centro universitario de UTPL se almacenó correctamente en la base de datos, mostrará un mensaje de confirmación del registro. Si el nombre del centro universitario no pudo ser registrado mostrará un mensaje de error en el almacenamiento. Si el nombre ingresado del centro universitario de UTPL, ya se encuentra anteriormente registrado en la base de datos mostrará un mensaje que el nombre del centro UTPL ya se encuentra registrado en la base de datos. Si el administrador no ingresa información en el campo Centro UTPL y presiona el botón Agregar, mostrará un mensaje que debe de llenar el campo centro UTPL.
 - Mensaje de confirmación del registro del centro UTPL
 - Mensaje de error en almacenamiento del centro UTPL.
 - Mensaje de centro universitario ya registrado en la base de datos.
 - Mensaje que debe llenar el campo centro UTPL.
- Entidad Educativa:
 - Si el nombre de la entidad educativa se almacenó correctamente en la base de datos, mostrará un mensaje de confirmación del registro. Si el nombre de la entidad educativa no pudo ser registrada mostrará un mensaje de error en el almacenamiento. Si el nombre ingresado de la entidad educativa, ya se encuentra anteriormente registrado en la base de datos mostrará un mensaje que la entidad educativa ya se encuentra registrada en la base de datos. Si el administrador no ingresa información en el campo Entidad Educativa y presiona el botón Agregar, mostrará un mensaje que debe de llenar el campo entidad educativa.
 - Mensaje de confirmación del registro de la entidad educativa
 - Mensaje de error en almacenamiento de la entidad educativa
 - Mensaje de que la entidad educativa ya está registrada en la base de datos.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

- Mensaje que debe llenar el campo entidad educativa.
- Carrera:
 - En caso de que el administrador no haya elegido una entidad educativa, para ingresar una nueva carrera, el sistema mostrará un mensaje de error que no se puede registrar una carrera sin elegir una entidad educativa.
 - Si el nombre de la carrera se almacenó correctamente en la base de datos, mostrará un mensaje de confirmación del registro. Si el nombre de la carrera no pudo ser registrada mostrará un mensaje de error en el almacenamiento. Si el nombre ingresado de la carrera, ya se encuentra anteriormente registrado en la base de datos mostrará un mensaje que la entidad educativa ya se encuentra registrada en la base de datos.
 - Si el administrador no ingresa información en el campo carrera y presiona el botón Agregar, mostrará un mensaje que debe de llenar el campo carrera.
 - Mensaje de error que no se puede registrar una carrera sin elegir una entidad educativa.
 - Mensaje de confirmación del registro de la carrera
 - Mensaje de error en almacenamiento de la carrera
 - Mensaje de que carrera ingresada ya está registrada en la base de datos.
 - Mensaje que debe llenar el campo carrera.
- Modalidad:
 - Si el tipo de modalidad se almacenó correctamente en la base de datos, mostrará un mensaje de confirmación del registro. Si el tipo de modalidad ingresada no pudo ser registrado mostrará un mensaje de error en el almacenamiento. Si el tipo de modalidad colocado ya se encuentra anteriormente registrado en la base de datos, mostrará un mensaje que el tipo de modalidad ingresado ya se encuentra registrado en la base de datos. Si el administrador no ingresa información en el campo modalidad y presiona el botón Agregar, mostrará un mensaje que debe de llenar el campo modalidad.
 - Mensaje de confirmación del registro del centro UTPL
 - Mensaje de error en almacenamiento del centro UTPL
 - Mensaje de centro universitario ya registrado en la base de datos.
 - Mensaje que debe llenar el campo modalidad.
- Periodo académico:
 - Si el periodo académico se almacenó correctamente en la base de datos, mostrará un mensaje de confirmación del registro. Si periodo académico ingresado no pudo ser registrado mostrará un mensaje de error en el almacenamiento. Si el periodo académico colocado ya se

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Especificación de Requerimientos	Fecha: 31/08/2012
PPA-ESR	

encuentra anteriormente registrado en la base de datos, mostrará un mensaje que el periodo académico ingresado ya se encuentra registrado en la base de datos. Si el administrador no ingresa información en el campo periodo académico y presiona el botón Agregar, mostrará un mensaje que debe de llenar el campo periodo académico.

- Mensaje de confirmación del registro de periodo académico.
- Mensaje de error en almacenamiento de periodo académico.
- Mensaje de que el periodo académico ingresado ya registrado en la base de datos.
- Mensaje que debe llenar el campo periodo académico.

3.2. Requerimientos Técnicos

Los requerimientos técnicos para el desarrollo de esta aplicación se muestran a continuación, en la Tabla B-8.

Servidor de Aplicaciones	Apache Tomcat - 7.0.27
Sistema Operativo	Windows -7
Lenguaje de programación	Java
Base de datos	MySQL - 5.5.

Tabla B-8 Requerimientos Técnicos

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Autenticación en el Sistema	Fecha: 31/08/2012
PPA-ASI	

B.5. Especificación de Casos de Uso

1. Autenticación en el Sistema

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/08/31	1.0	Este caso de uso describe cada una de las acciones que realiza el sistema en el proceso de autenticación del usuario en el mismo.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Autenticación en el Sistema	Fecha: 31/08/2012
PPA-ASI	

1.1. Breve descripción

Este caso de uso describe claramente cada una de las acciones que realiza el sistema en el proceso de autenticación en el mismo. Además indica cuales son los actores y las pre y post condiciones para el proceso de autenticación funcione correctamente.

1.2. Especificación de Caso de Uso

Número:	CU01	
Nombre:	Autenticación en el Sistema	
Actor(es):	Usuario (Secretaria)	
Descripción:	Su función es dar acceso a los recursos del sistema en base al perfil asignado.	
Precondiciones:	El usuario debe estar registrado en el Sistema.	
Postcondiciones:	El usuario podrá ver las funcionalidades del sistema en base a su perfil asignado (secretaria, fiscal, administrador)	
Flujo Normal:	Actor	Sistema
	1. Solicita acceso a los recursos del Preprocesador de Planes Académicos.	
		2. Muestra plantilla para el ingreso de usuario y contraseña
	3. El usuario ingresa usuario y contraseña.	
		4. Valida los datos ingresados. REF1. El sistema opcionalmente: a. Si los datos ingresados son correctos, da acceso al sistema de preprocesador de planes académicos. b. Si los datos son incorrectos, regresa al paso 2 del flujo normal. FA1, FA2 5. Fin del caso de uso
Flujo Alternativo	FA 1: Datos requeridos no han sido ingresados Esto sucede cuando el usuario no ha ingresado los datos que son obligatorios (usuario y contraseña), en caso de que esto suceda se volverá a cargar la pantalla de autenticación de ingreso al sistema.	

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Autenticación en el Sistema	Fecha: 31/08/2012
PPA-ASI	

	<p>FA 2: Datos incorrectos Sucede cuando el usuario que está en el proceso de autenticación ingresa de forma incorrecta su usuario y/o contraseña, en caso de que suceda el sistema volverá a cargar la pantalla de autenticación de ingreso al sistema.</p>
Prioridad:	Alta
Referencias Cruzadas:	De requerimientos: REQ01
Requerimientos Funcionales	<p>REF1: Validar Datos Los datos ingresados (usuario y contraseña) por el usuario deberán de compararse con los datos que se encuentren almacenados en la base de datos.</p>
Restricciones	Para que el usuario pueda ingresar al sistema este previamente debe haber sido registrado en el sistema por el administrador. No existe la posibilidad que el usuario se pueda registrar por sí mismo en el sistema.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Registro de Planes Académicos Externos	Fecha: 04/09/2012
PPA-RPE	

2. Registro de Planes Académicos Externos

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/04	1.0	Este caso de uso describe exactamente cada una de las acciones que realiza el sistema en el proceso de registro de planes académicos externos a la UTPL.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Registro de Planes Académicos Externos	Fecha: 04/09/2012
PPA-RPE	

2.1. Breve Descripción

Este caso de uso describe cada una de las acciones que realiza el sistema en el proceso de registro de planes académicos externos. Además indica cuales son los actores y las pre y post condiciones para el proceso de registro funcione correctamente.

2.2. Especificación de Caso de Uso

Número:	CU02	
Nombre:	Registro de Planes Académicos Externos	
Actor(es):	Usuario (Secretaria)	
Descripción:	Su función es cargar el archivo correspondiente al plan académico, registrar los datos generales del plan, cada uno de sus contenidos y el archivo correspondiente a dicho plan.	
Precondiciones:	El usuario debe estar autenticado en el Sistema.	
Postcondiciones:	El usuario podrá registrar la información y el archivo digital del plan académico externo.	
Flujo Normal:	Actor	Sistema
	6. Ingresar al módulo de registro de planes académicos externos.	
		7. Muestra formulario para subir al sistema, el archivo digital del plan académico.
	8. El usuario selecciona el archivo digital del plan académico externo a registrar.	
		9. Valida si la extensión del archivo corresponde a la permitida. El sistema opcionalmente: a. Si la extensión del archivo ingresado es la correcta, el sistema da acceso al formulario de registro de los datos generales del plan académico. b. Si la extensión del archivo elegida para subir al sistema es incorrecta, se

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Registro de Planes Académicos Externos	Fecha: 04/09/2012
PPA-RPE	

		regresa al paso 2. REF1 c. Si el usuario no elige un archivo para subir al sistema, este cargará nuevamente la misma página. FA1
		10. Presenta el formulario de registro de información general del plan académico, así como también los contenidos de las diferentes unidades del mismo.
	11. El usuario ingresa los valores solicitados en el formulario	
		12. Valida los datos ingresados en el formulario. El sistema opcionalmente: a. Si los datos ingresados en el formulario son correctos, mostrará una alerta “El plan académico se ha registrado correctamente”. b. Si el usuario no completa toda la información del formulario se alertará sobre la necesidad de que se ingresen los datos del formulario. FA1. c. Si al guardar los datos, existiera un error, se alertará al usuario sobre lo sucedido. FA2. 13. Fin del caso de uso.
Flujo Alternativo	FA3: Datos requeridos no han sido ingresados Sucedo cuando el usuario no ha seleccionado el archivo correspondiente al plan académico externo que se está registrando, en caso de que esto ocurra se volverá a cargar la pantalla para subir el archivo al sistema. Así mismo cuando no ha ingresado toda la información requerida en el formulario de registro de planes externos, se mostrará la siguiente alerta: “Todos los campos de este formulario deben ser completados”.	
	FA4: Error al almacenar los datos	

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Registro de Planes Académicos Externos	Fecha: 04/09/2012
PPA-RPE	

	Cuando la información ingresada en los campos del formulario no se haya podido almacenar por error de conexión con servidor web y/o de base de datos, se mostrará la siguiente alerta: “Error al guardar los datos”.
Prioridad:	Media
Referencias Cruzadas:	De requerimientos: REQ02
Requerimientos Funcionales	REF1: Validar Tipo de Archivo El archivo del plan académico a subir en el sistema deberá verificar que el archivo tenga la extensión .PDF, para poder ser aceptado por el sistema.
Restricciones	Para que el usuario pueda registrar tanto la información general del plan académico externo, así como también los contenidos de cada una de las unidades, es necesario que previamente el usuario suba al sistema el archivo digital del plan académico.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Analizar Plan Académico	Fecha: 04/09/2012
PPA-APA	

3. Analizar Plan Académico

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/04	1.0	Este caso de uso describe exactamente cada una de las acciones que realiza el sistema en el proceso de análisis de planes académicos.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Analizar Plan Académico	Fecha: 04/09/2012
PPA-APA	

3.1. Breve Descripción

Este caso de uso describe claramente cada una de las acciones que realiza el sistema en el proceso de análisis de planes académicos. Además indica cuáles son los actores y las pre y post condiciones para el proceso de análisis funcione correctamente.

3.2. Especificación de Caso de Uso

Número:	CU03	
Nombre:	Análisis de Planes Académicos	
Actor(es):	Usuario (Fiscal)	
Descripción:	Su función es buscar el plan académico a analizar, seleccionar la asignatura con la cual se va a convalidar la asignatura para finalmente obtener los contenidos importantes. El usuario tiene la capacidad para reorganizar los contenidos importantes seleccionados por el sistema con aquellos contenidos que no fueron tomados en cuenta por el sistema.	
Precondiciones:	El usuario debe estar autenticado en el Sistema.	
Postcondiciones:	El usuario podrá tener los contenidos más importantes que se encuentren en el plan académico, en base a la asignatura de convalidación seleccionada.	
Flujo Normal:	Actor	Sistema
	14. Ingresa al módulo de análisis de planes académicos.	
		15. Muestra formulario para buscar el plan académico a analizar.
	16. El usuario ingresa el código del formulario a buscar y presiona el botón Buscar	
		17. Busca el plan académico en base al código ingresado por el usuario. El sistema opcionalmente: d. Si el código ingresado existe, se muestra el nombre de asignatura, centro UTPL de registro, universidad a la que pertenece el plan,

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Analizar Plan Académico	Fecha: 04/09/2012
PPA-APA	

		<p>modalidad de estudios y carrera.REF1</p> <p>e. Si el código ingresado es incorrecto o no existe, mostrará un mensaje indicando el error. FA2</p> <p>f. Si no ingresa un código de plan académico, mostrará un mensaje indicando el error. FA1</p>
		<p>18. El sistema presenta el formulario para seleccionar la carrera y la asignatura de la UTPL con la cual se desea realizar el análisis de convalidación del plan académico.</p>
	<p>19. El usuario selecciona la carrera y la asignatura de la UTPL con la cual se desea realizar el análisis de convalidación del plan académico y presiona el botón Analizar Plan</p>	
		<p>20. El sistema muestra el resultado del proceso de comparación del plan académico buscado (paso 4) con la carrera y asignatura elegida. Este resultado son los contenidos importantes del plan académico elegidos por la herramienta. El sistema opcionalmente:</p> <p>d. Mostrará todos los contenidos que no fueron tomados en cuenta como importantes por la herramienta. REF2</p> <p>e. El sistema ofrece la posibilidad de editar los contenidos importantes y los no importantes, es decir</p>

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Analizar Plan Académico	Fecha: 04/09/2012
PPA-APA	

		que se pueden cambiar contenidos no importantes a importantes y viceversa. FA3
	21. El usuario presiona el botón de confirmar contenidos importantes del plan académico.	
		22. El sistema almacena los contenidos importantes en la base de datos. El sistema opcionalmente: a. Mostrará el mensaje: “Contenidos importantes almacenados correctamente”, si la información se almacenó correctamente. b. Mostrará un mensaje si los contenidos importantes del plan académico no pudieron ser almacenados en la base de datos. FA4 23. Fin del caso de uso.
Flujo Alternativo	FA5: Datos requeridos no han sido ingresados Cuando el usuario no ha ingresado el código de un plan académico a buscar, se mostrará el siguiente mensaje: “Debe ingresar un código de plan académico”	
	FA2: Datos incorrectos o inexistentes Cuando el usuario ingresa un código de plan académico incorrecto o inexistente se mostrará el siguiente mensaje: “Código de plan académico incorrecto o inexistente”.	
	FA3: Error al almacenar los datos Cuando la información ingresada en los campos del formulario no se haya podido almacenar por error de conexión con servidor web y/o de base de datos, se mostrará la siguiente alerta: “Error al guardar los datos”.	
	FA4: Editar contenidos importantes. Los contenidos que no hayan sido elegidos como importantes, el usuario puede elegirlos para que sean almacenados en la base de datos como importantes, de la misma forma los contenidos que hayan sido elegidos como importantes y el usuario en base a su	

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Analizar Plan Académico	Fecha: 04/09/2012
PPA-APA	

	conocimiento cree que no lo son, pueden ser quitados de la lista de contenidos importantes.
Prioridad:	Media
Referencias Cruzadas:	De requerimientos: REQ03
Requerimientos Funcionales	REF1: Validar código de plan académico. El código de plan académico ingresado por el usuario deberá compararse con el código que se encuentre almacenado en la base de datos.
	REF2: Presentar contenidos no importantes para el sistema. El sistema presentará de forma separada los contenidos que no fueron tomados en cuenta como importantes por el sistema.
Restricciones	La búsqueda de los planes académicos únicamente puede ser realizada mediante el código del plan académico.

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Generar Archivo de Texto con Resultados	Fecha: 06/09/2012
PPA-GAR	

4. Generar archivo de texto con resultados

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/06	1.0	Este caso de uso describe como el sistema genera un archivo de texto con los resultados obtenidos paso a paso para la obtención de contenidos importantes.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Generar Archivo de Texto con Resultados	Fecha: 06/09/2012
PPA-GAR	

4.1. Breve descripción

Este caso de uso describe como se realiza el proceso de generación de un archivo de texto, el cual contiene el resultado de todos los pasos realizados por el sistema para obtener los contenidos importantes del plan académico realizado.

4.2. Especificación de caso de uso

Número:	CU04	
Nombre:	Generar archivo de texto con los resultados.	
Actor(es):	Usuario (Fiscal)	
Descripción:	Su función es generar un archivo de texto con cada uno de los pasos realizados por el sistema para presentar los contenidos importantes.	
Precondiciones:	El usuario debe realizar el análisis del plan académico y guardar dicho análisis en la base de datos.	
Postcondiciones:	El usuario podrá ver cada uno de los pasos realizados por el sistema para obtener los contenidos importantes.	
Flujo Normal:	Actor	Sistema
	1. Presiona el botón Generar TXT	
		2. Muestra una ventana de exploración de carpetas. REF1
	3. Selecciona el lugar donde va a almacenar el archivo y coloca un nombre con el cual se va a guardar el archivo.	
		4. Descarga el archivo con el nombre asignado por el usuario y en el lugar especificado por el mismo. REF2 5. Fin del caso de uso
Prioridad:	Baja	
Referencias Cruzadas:	De requerimientos: REQ03, REQ04 De caso de uso: CU03	
Requerimientos Funcionales	REF1: Guardar archivo El usuario tiene la capacidad para seleccionar el lugar en donde se va a almacenar el archivo de texto, así como también el nombre que va a tener dicho archivo.	

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Generar Archivo de Texto con Resultados	Fecha: 06/09/2012
PPA-GAR	

	<p>REF2: Proceso de Extracción de contenidos</p> <p>El archivo generado tendrá cada uno de los pasos realizados por el sistema para obtener los contenidos importantes del plan académico analizado.</p>
Restricciones	<p>El archivo tendrá únicamente el proceso de extracción de los contenidos importantes obtenidos por el sistema. Los contenidos que hayan sido editados por el usuario no constarán en este archivo.</p>

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Generar Archivo XML de contenidos importantes	Fecha: 06/09/2012
PPA-GXML	

5. Generar archivo XML de contenidos importantes

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/06	1.0	Este caso de uso describe como el sistema genera un archivo de texto con los resultados obtenidos paso a paso para la obtención de contenidos importantes.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Generar Archivo XML de contenidos importantes	Fecha: 06/09/2012
PPA-GXML	

5.1. Breve descripción

Este caso de uso describe como se realiza el proceso de generación de un archivo con formato XML, el cual contiene los contenidos importantes del plan académico analizado.

5.2. Especificación de caso de uso

Número:	CU05	
Nombre:	Generar archivo XML de contenidos importantes	
Actor(es):	Usuario (Fiscal)	
Descripción:	Su función es generar un archivo en formato XML con los contenidos importantes correspondientes al plan académico analizado.	
Precondiciones:	El usuario debe realizar el análisis del plan académico y guardar dicho análisis en la base de datos.	
Postcondiciones:	El usuario tendrá un archivo en formato XML con los contenidos importantes del plan académico.	
Flujo Normal:	Actor	Sistema
	1. Presiona el botón Generar XML	
		2. Muestra una ventana de exploración de carpetas. REF1
	3. Selecciona el lugar donde va a almacenar el archivo y coloca un nombre con el cual se va a guardar el archivo.	
		4. Descarga el archivo con el nombre asignado por el usuario y en el lugar especificado por el mismo. REF2 5. Almacena la estructura del archivo XML en la base de datos. FA1 6. Fin del caso de uso
Flujo Alternativo	FA1: Error al almacenar los datos Cuando la estructura del archivo XML no se haya podido almacenar por error de conexión con servidor web y/o de base de datos, se mostrará la siguiente alerta: "Error al guardar los datos".	

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Generar Archivo XML de contenidos importantes	Fecha: 06/09/2012
PPA-GXML	

Prioridad:	Baja
Referencias Cruzadas:	De requerimientos: REQ03, REQ05 De caso de uso: CU03
Requerimientos Funcionales	REF1: Guardar archivo El usuario tiene la capacidad para seleccionar el lugar en donde se va a almacenar el archivo de texto, así como también el nombre que va a tener dicho archivo.
	REF2: Archivo XML Los contenidos importantes serán generados en un formato de archivo estructurado como es XML, el cual tiene la posibilidad de ser descargado por el usuario.
Restricciones	-

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Crear nuevo corpus y lematización	Fecha: 07/09/2012
PPA-NCL	

6. Generar gráfico de contenidos importantes

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/06	1.0	Este caso de uso describe como el sistema realiza la generación de un gráfico en donde se muestran los contenidos importantes extraídos del plan académico analizado.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Crear nuevo corpus y lematización	Fecha: 07/09/2012
PPA-NCL	

6.1. Breve descripción

Este caso de uso describe como se realiza el proceso de generación de un gráfico, el cual contiene los contenidos importantes del plan académico analizado.

6.2. Especificación de Caso de Uso

Número:	CU06	
Nombre:	Generar gráfico de contenidos importantes	
Actor(es):	Usuario (Fiscal)	
Descripción:	Su función es generar un gráfico que muestre visualmente los contenidos importantes extraídos correspondientes al plan académico analizado.	
Precondiciones:	El usuario debe realizar el análisis del plan académico y guardar dicho análisis en la base de datos.	
Postcondiciones:	El usuario podrá visualizar los contenidos importantes extraídos del plan académico.	
Flujo Normal:	Actor	Sistema
	1. Presiona el botón Visualizar gráfico	
		2. Realiza una consulta a la base de datos para poder obtener la información. FA1 3. Muestra un gráfico con los contenidos importantes extraídos.
		4. Fin del caso de uso
Flujo Alternativo	FA1: Error al consultar los datos Cuando los datos necesarios para generar el gráfico no se hayan podido consultar de la base de datos, se mostrará la siguiente alerta: "No se pudo consultar a la base de datos".	
Prioridad:	Baja	
Referencias Cruzadas:	De requerimientos: REQ03, REQ05 De caso de uso: CU03	
Requerimientos Funcionales	REF1: Mostrar gráfico El usuario podrá visualizar un gráfico en donde se encuentren los contenidos importantes extraídos.	
Restricciones	El gráfico se lo podrá visualizar en el sistema, este no podrá ser descargado del mismo.	

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Crear nuevo corpus y lematización	Fecha: 07/09/2012
PPA-NCL	

7. Crear nuevo corpus y lematización

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/07	1.0	Este caso de uso describe exactamente cada una de las acciones que realiza el sistema para crear un nuevo corpus y lematización perteneciente a este corpus.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Crear nuevo corpus y lematización	Fecha: 07/09/2012
PPA-NCL	

7.1. Breve descripción

Este caso de uso describe claramente cada una de las acciones que realiza el sistema en el proceso de creación de un nuevo corpus y lematización para el sistema, correspondiente a una asignatura de una carrera.

7.2. Especificación de Caso de Uso

Número:	CU07	
Nombre:	Crear nuevo corpus y lematización	
Actor(es):	Usuario (Administrador)	
Descripción:	Su función es crear un nuevo corpus para una asignatura de una carrera que no posea un corpus, de la misma manera permite definir la lematización de la nuevo corpus creado.	
Precondiciones:	El usuario debe estar autenticado en el Sistema.	
Postcondiciones:	El usuario tendrá un nuevo corpus con su respectiva lematización para una asignatura de una carrera que no se encuentre aún definida.	
Flujo Normal:	Actor	Sistema
	24. Ingresar al módulo de diccionario.	
		25. Se muestra un formulario para registrar el nuevo corpus.
	26. El usuario ingresa los valores solicitados en el formulario	
		27. Valida los datos ingresados en el formulario. El sistema opcionalmente: f. Si el usuario ingreso información en todos los campos del formulario, el sistema muestra el mensaje “Datos almacenados correctamente” g. Si el usuario no completa toda la información del formulario se alertará sobre la necesidad de que se ingresen los datos del formulario. FA1, REF2

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Crear nuevo corpus y lematización	Fecha: 07/09/2012
PPA-NCL	

		<p>h. Si al guardar los datos, existiera un error, se alertará al usuario sobre lo sucedido. FA2.</p> <p>i. Si la palabra del nuevo corpus y/o lematización, ya se encuentran almacenadas en el diccionario, se alertará al usuario. REF1</p> <p>28. Fin del caso de uso.</p>
Flujo Alternativo	FA6: Datos requeridos no han sido ingresados Cuando el usuario no ha ingresado todos los campos del formulario de registro de un nuevo corpus inclusive la primera lematización, se presentará una alerta con el siguiente mensaje: “Debe ingresar datos en todos los campos del formulario”	
	FA2: Error al almacenar los datos Cuando la información ingresada en los campos del formulario no se haya podido almacenar por error de conexión con servidor web y/o de base de datos, se mostrará la siguiente alerta: “Error al guardar los datos”.	
Prioridad:	Media	
Referencias Cruzadas:	De requerimientos: REQ07	
Requerimientos Funcionales	REF1: Datos Duplicados Si la palabra del nuevo corpus y/o lematización ingresadas en el formulario, ya se encuentran almacenadas en el diccionario, mostrará la siguiente alerta: “Datos duplicados”	
	REF2: Registrar Lematización. El usuario para registrar el nuevo corpus debe de ingresar la primera lematización, debido a que no existe un corpus sin por lo menos una lematización definida.	
Restricciones	Para almacenar el nuevo corpus se debe registrar también la primera lematización.	

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Actualizar lematización	Fecha: 07/09/2012
PPA-ACL	

8. Actualizar lematización

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/07	1.0	Este caso de uso describe exactamente cada una de las acciones que realiza el sistema para actualizar la lematización perteneciente a un corpus existente.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Actualizar lematización	Fecha: 07/09/2012
PPA-ACL	

8.1. Breve descripción

Este caso de uso describe claramente cada una de las acciones que realiza el sistema en el proceso de actualización de una lematización existente, correspondiente a una asignatura de una carrera registrada.

8.2. Especificación de Caso de Uso

Número:	CU08	
Nombre:	Crear nuevo corpus y lematización	
Actor(es):	Usuario (Administrador)	
Descripción:	Su función es crear un nuevo corpus para una asignatura de una carrera que no posea un corpus, de la misma manera permite definir la lematización de la nuevo corpus creado.	
Precondiciones:	El usuario debe estar autenticado en el Sistema. Debe existir un corpus definido para la nueva lematización a registrar.	
Postcondiciones:	El usuario tendrá una nueva lematización correspondiente a un corpus ya registrado.	
Flujo Normal:	Actor	Sistema
	29. Ingresar al módulo de diccionario.	
		30. Se muestra un formulario para registrar la nueva lematización.
	31. El usuario ingresa los valores solicitados en el formulario	
		32. Valida los datos ingresados en el formulario. El sistema opcionalmente: j. Si el usuario ingreso información en todos los campos del formulario, el sistema muestra el mensaje “Datos almacenados correctamente” k. Si el usuario no completa toda la información del formulario se alertará sobre la necesidad de que se ingresen los datos del

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Actualizar lematización	Fecha: 07/09/2012
PPA-ACL	

		<p>formulario. FA1.</p> <ol style="list-style-type: none"> l. Si al guardar los datos, existiera un error, se alertará al usuario sobre lo sucedido. FA2. m. Si la nueva palabra para la lematización, ya se encuentra almacenada en el diccionario, se alertará al usuario. REF1 <p>33. Fin del caso de uso.</p>
Flujo Alternativo	FA7: Datos requeridos no han sido ingresados Cuando el usuario no ha ingresado todos los campos del formulario de registro de un nuevo corpus inclusive la primera lematización, se presentará una alerta con el siguiente mensaje: “Debe ingresar datos en todos los campos del formulario”	
	FA2: Error al almacenar los datos Cuando la información ingresada en los campos del formulario no se haya podido almacenar por error de conexión con servidor web y/o de base de datos, se mostrará la siguiente alerta: “Error al guardar los datos”.	
Prioridad:	Media	
Referencias Cruzadas:	De requerimientos: REQ07, REQ08 De caso de uso: CU07	
Requerimientos Funcionales	REF1: Datos Duplicados Si la nueva palabra de la lematización ingresada en el formulario, ya se encuentra almacenada en el diccionario, mostrará la siguiente alerta: “Datos duplicados”	
Restricciones	Para almacenar la nueva lematización se debe obligatoriamente elegir de entre el conjunto de palabras que conforman el corpus de la carrera y asignatura seleccionada.	

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Crear carreras de UTPL	Fecha: 07/09/2012
PPA-CCU	

9. Crear carreras UTPL

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/07	1.0	Este caso de uso describe exactamente cada una de las acciones que realiza el sistema para crear carreras correspondientes a UTPL.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Crear carreras de UTPL	Fecha: 07/09/2012
PPA-CCU	

9.1. Breve descripción

Este caso de uso describe claramente cada una de las acciones que realiza el sistema en el proceso de actualización de una lematización existente, correspondiente a una asignatura de una carrera registrada.

9.2. Especificación de Casos de Uso

Número:	CU09	
Nombre:	Crear carreras de UTPL	
Actor(es):	Usuario (Administrador)	
Descripción:	Su función es definir en el sistema carreras que pertenezcan a la UTPL pero que aún no hayan sido registradas.	
Precondiciones:	El usuario debe estar autenticado en el sistema.	
Postcondiciones:	El usuario tendrá una nueva carrera registrada en el sistema.	
Flujo Normal:	Actor	Sistema
	34. Ingresar al módulo de diccionario.	
		35. Se muestra un formulario para definir la nueva carrera.
	36. El usuario ingresa el nombre de la carrera a registrar	
		<p>37. Valida los datos ingresados en el formulario.</p> <p>El sistema opcionalmente:</p> <p>n. Si el usuario ingreso el nombre de la carrera, el sistema muestra el mensaje “Carrera almacenada correctamente”.</p> <p>o. Si el usuario no ingresa un nombre a la nueva carrera a registrar, se alertará al usuario sobre lo sucedido. FA1.</p> <p>p. Si al guardar los datos, existiera un error, se alertará al usuario sobre lo sucedido. FA2.</p>

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Crear carreras de UTPL	Fecha: 07/09/2012
PPA-CCU	

		<p>q. Si el nombre para la nueva carrera, ya se encuentra almacenada en el sistema, se alertará al usuario. REF1 38. Fin del caso de uso.</p>
Flujo Alternativo	FA8: Datos requeridos no han sido ingresados Cuando el usuario no ha ingresado un nombre para la nueva carrera, se presentará una alerta con el siguiente mensaje: “Debe ingresar un nombre para la nueva carrera”	
	FA2: Error al almacenar los datos Cuando la información ingresada en los campos del formulario no se haya podido almacenar por error de conexión con servidor web y/o de base de datos, se mostrará la siguiente alerta: “Error al guardar los datos”.	
Prioridad:	Media	
Referencias Cruzadas:	De requerimientos: REQ09	
Requerimientos Funcionales	REF1: Datos Duplicados Si el nombre de la nueva carrera, ya se encuentra almacenada en el sistema, mostrará la siguiente alerta: “Datos duplicados”.	
Restricciones	-	

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Crear asignaturas de UTPL	Fecha: 10/09/2012
PPA-CAU	

10. Crear asignaturas UTPL

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/10	1.0	Este caso de uso describe exactamente cada una de las acciones que realiza el sistema para crear asignaturas correspondientes a carreras pertenecientes a UTPL.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Crear asignaturas de UTPL	Fecha: 10/09/2012
PPA-CAU	

10.1. Breve descripción

Este caso de uso describe claramente cada una de las acciones que realiza el sistema en el proceso de creación de nuevas asignaturas de la UTPL correspondiente a una carrera ya registrada en el sistema.

10.2. Especificación de Caso de Uso

Número:	CU10	
Nombre:	Crear asignaturas de UTPL	
Actor(es):	Usuario (Administrador)	
Descripción:	Su función es definir en el sistema asignaturas que pertenezcan a la UTPL pero que aún no hayan sido registradas.	
Precondiciones:	El usuario debe estar autenticado en el sistema.	
Postcondiciones:	El usuario tendrá una nueva asignatura perteneciente a una carrera registrada en el sistema.	
Flujo Normal:	Actor	Sistema
	39. Ingresar al módulo de diccionario.	
		40. Se muestra un formulario para definir la nueva asignatura.
	41. El usuario selecciona la carrera a la que pertenece la asignatura a registrar e ingresa el nombre de la nueva asignatura.	
		42. Valida los datos ingresados en el formulario. El sistema opcionalmente: r. Si el usuario seleccionó una carrera e ingreso el nombre de la asignatura, el sistema muestra el mensaje "Asignatura almacenada correctamente". s. Si el usuario no seleccionó una carrera y/o ingreso el nombre de la asignatura, se alertará al usuario sobre lo sucedido. FA1.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Crear asignaturas de UTPL	Fecha: 10/09/2012
PPA-CAU	

		<p>t. Si al guardar los datos, existiera un error, se alertará al usuario sobre lo sucedido. FA2.</p> <p>u. Si el nombre para la nueva carrera, ya se encuentra almacenada en el sistema, se alertará al usuario. REF1</p> <p>43. Fin del caso de uso.</p>
Flujo Alternativo	FA9: Datos requeridos no han sido ingresados Cuando el usuario no ha ingresado un nombre para la nueva carrera, se presentará una alerta con el siguiente mensaje: “Debe ingresar un nombre para la nueva carrera”	
	FA2: Error al almacenar los datos Cuando la información ingresada en los campos del formulario no se haya podido almacenar por error de conexión con servidor web y/o de base de datos, se mostrará la siguiente alerta: “Error al guardar los datos”.	
Prioridad:	Media	
Referencias Cruzadas:	De requerimientos: REQ09 De casos de uso: CU09	
Requerimientos Funcionales	REF1: Datos Duplicados Si el nombre de la nueva asignatura, ya se encuentra almacenada dentro de esa carrera en el sistema, se mostrará la siguiente alerta: “Datos duplicados”.	
Restricciones	-	

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Gestionar Usuarios – Crear usuarios	Fecha: 10/09/2012
PPA-CNU	

11. Gestionar Usuarios – Crear nuevo usuario

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/10	1.0	Este caso de uso describe exactamente cada una de las acciones que realiza el sistema para gestionar los usuarios, específicamente la creación de nuevos usuarios.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Gestionar Usuarios – Crear usuarios	Fecha: 10/09/2012
PPA-CNU	

11.1. Breve descripción

Este caso de uso describe claramente cada una de las acciones que realiza el sistema en el proceso de gestión de usuarios, concerniente a la creación de usuarios para el mismo.

11.2. Especificación de Caso de Uso

Número:	CU11	
Nombre:	Gestionar usuarios – Crear nuevo usuario	
Actor(es):	Usuario (Administrador)	
Descripción:	Su función es gestionar la creación de usuarios para que estos puedan utilizar el sistema.	
Precondiciones:	El usuario debe estar autenticado en el sistema.	
Postcondiciones:	El usuario tendrá nuevos usuarios registrados en el sistema.	
Flujo Normal:	Actor	Sistema
	44. Ingresar al módulo de usuarios.	
		45. Muestra un formulario para registrar un nuevo usuario.
	46. El usuario ingresa los valores solicitados en el formulario.	
		47. Valida los datos ingresados en el formulario. El sistema opcionalmente: v. Si los datos ingresados en el formulario son correctos, mostrará una alerta “El plan académico se ha registrado correctamente”. w. Si el usuario no completa toda la información del formulario se alertará sobre la necesidad de que se ingresen todos los datos en el formulario. FA1. x. Si al guardar los datos, existiera un error, se alertará al usuario sobre lo

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Gestionar Usuarios – Crear usuarios	Fecha: 10/09/2012
PPA-CNU	

		<p>sucedido. FA2.</p> <p>y. Si el nombre de usuario, ya se encuentra almacenada en el sistema, se alertará al usuario.</p> <p>REF1</p> <p>48. Fin del caso de uso.</p>
Flujo Alternativo	FA10: Datos requeridos no han sido ingresados Cuando el usuario no ha ingresado datos en todos los campos del formulario, se presentará una alerta con el siguiente mensaje: “Todos los campos de este formulario deben ser llenados”.	
	FA2: Error al almacenar los datos Cuando la información ingresada en los campos del formulario no se haya podido almacenar por error de conexión con servidor web y/o de base de datos, se mostrará la siguiente alerta: “Error al guardar los datos”.	
Prioridad:	Media	
Referencias Cruzadas:	De requerimientos: REQ11	
Requerimientos Funcionales	REF1: Datos Duplicados Si el nombre del nuevo usuario, ya se encuentra almacenado en el sistema, se mostrará la siguiente alerta: “El nombre de usuario ya se encuentra registrado”.	
Restricciones	Esta parte del módulo de usuarios solo permite crear nuevos usuarios y no actualizar información de los mismos.	

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Arquitectura de Software	Fecha: 13/09/2012
PPA-ARS	

12. Crear asignaturas de UTPL

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/10	1.0	Este caso de uso describe exactamente cada una de las acciones que realiza el sistema para gestionar los usuarios, específicamente la eliminación de usuarios.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Arquitectura de Software	Fecha: 13/09/2012
PPA-ARS	

12.1. Breve descripción

Este caso de uso describe claramente cada una de las acciones que realiza el sistema en el proceso de gestión de usuarios, concerniente a la eliminación de usuarios para el mismo.

12.2. Especificación de Caso de Uso

Número:	CU12	
Nombre:	Gestionar usuarios – Eliminar Usuario	
Actor(es):	Usuario (Administrador)	
Descripción:	Su función es gestionar la eliminación de usuarios que utilizan el sistema.	
Precondiciones:	El usuario debe estar autenticado en el sistema.	
Postcondiciones:	El usuario tendrá eliminados los usuarios que haya elegido.	
Flujo Normal:	Actor	Sistema
	49. Ingresar al módulo de usuarios.	
		50. Muestra un listado con todos los usuarios que están registrados en el sistema.
	51. Selecciona el usuario que desea eliminar del sistema.	
		52. Elimina el usuario seleccionado por el usuario y muestra la alerta “Usuario eliminado correctamente”. El sistema opcionalmente: z. Si el usuario seleccionado no fue eliminado alertará al usuario sobre lo sucedido. FA1. 53. Fin del caso de uso.
Flujo Alternativo	FA2: Error al eliminar usuario Cuando el usuario no haya podido eliminar el usuario seleccionado, se mostrará la siguiente alerta: “Error al eliminar usuario”.	
Prioridad:	Media	
Referencias Cruzadas:	De requerimientos: REQ11 De casos de uso: CU11	
Restricciones	-	

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Arquitectura de Software	Fecha: 13/09/2012
PPA-ARS	

B.6. Arquitectura de Software

Historial de Revisión

Fecha	Versión	Descripción	Autor
2012/09/13	1.0	Este documento describe la arquitectura del sistema de Preprocesador de Planes Académicos, en donde la carga está dividida en tres capas: Capa de presentación, capa de lógica de negocios y capa de datos, permitiendo de esta manera tener una vista global de como se encuentra estructurado el sistema.	Eylin Calderón

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Arquitectura de Software	Fecha: 13/09/2012
PPA-ARS	

1. Introducción

Este documento describe la arquitectura del sistema en donde la carga está dividida en tres capas: Capa de presentación, capa de lógica de negocios y capa de datos, permitiendo de esta manera tener una vista global de como se encuentra estructurado el sistema y así poder mostrar las características más importantes del sistema.

La plantilla de este documento se basó en las especificaciones de RUP (Rational Unified Process) para el documento de arquitectura de software.

1.1. Propósito

Este documento de arquitectura de software tiene como propósito brindar una visión clara de la arquitectura del software del preprocesador de planes académicos, mostrando la comunicación entre las capas de la arquitectura, lo cual servirá para tener una mejor comprensión de la misma.

1.2. Alcance

El Preprocesador de Planes Académicos, es una aplicación web que sirve para realizar el análisis de los planes académicos, siendo el objetivo final la extracción de los contenidos más importantes o representativos del plan académico que base a la comparación de un plan académico externo frente a un plan de la UTPL.

1.3. Definiciones, acrónimos y abreviaturas

- PPA: Preprocesador de Planes Académicos.
- ARS: Arquitectura de Software
- UTPL: Universidad Técnica Particular de Loja.
- RUP: Rational Unified Process

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Arquitectura de Software	Fecha: 13/09/2012
PPA-ARS	

- **Capa de presentación:** Es la capa que esta de lado del usuario, ya que es la que le presenta el sistema al usuario y por la cual se recibe la información del sistema.
- **Capa de lógica de negocios:** En esta capa se establecen todas las reglas que deben cumplirse para realizar cada una de las transacciones que realice el sistema.
- **Capa de datos:** En esta capa es donde se almacenan y se acceden a los datos.

2. Vista Lógica

La vista lógica del sistema denominado Preprocesador de Planes Académicos, está compuesto por tres capas: Capa de Presentación, Capa de Lógica de Negocios y Capa de Datos.

En la Figura B-4, se observa un diagrama general de la vista lógica de la arquitectura correspondiente al preprocesador de Planes Académicos de UTPL.

A continuación se realiza una descripción de cada una de las capas de la arquitectura de la aplicación, así como también los módulos que la conforman (la descripción de los módulos se encuentran definidos en el documento de especificación de requerimientos y en el documento de casos de uso), así mismo en la Figura B-5, se observa la vista lógica de la arquitectura del procesador de planes académicos de una forma más detallada.

2.1.Capa de Presentación

Esta capa es la interfaz con el usuario, la cual presenta al usuario el sistema y por la cual este recibe la información transmitida por el usuario. Esta capa está compuesta por una parte de hardware (PC) y una de software en este caso un navegador web.

Esta capa para el preprocesador de planes académicos está compuesta por los siguientes módulos:

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Arquitectura de Software	Fecha: 13/09/2012
PPA-ARS	

- Módulo de Registro de Planes Académicos Externos.
- Módulo de Selección de Contenidos.
- Módulo de Representación de Contenidos.
- Módulo de Administración del Diccionario.
- Módulo de Administración de Usuarios.
- Módulo de Actualización de Campos.

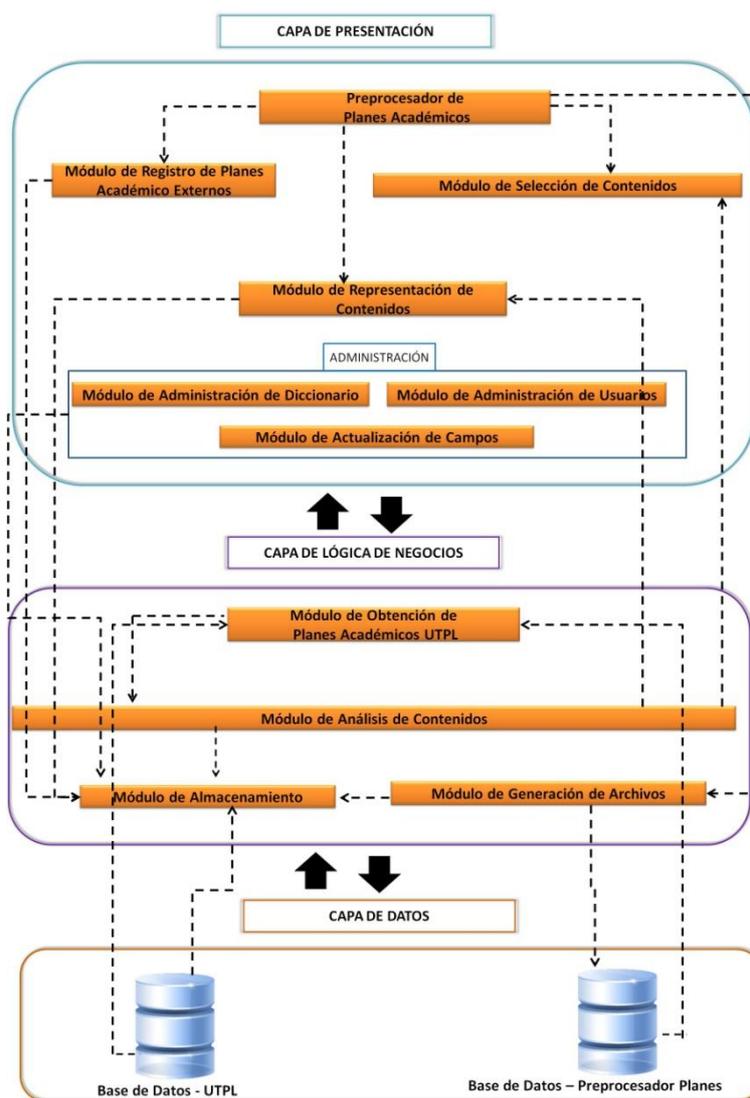


Figura. B-4 Vista Lógica de la Arquitectura General - Preprocesador de Planes Académicos

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Arquitectura de Software	Fecha: 13/09/2012
PPA-ARS	

2.2.Capa de Lógica de Negocios

Esta capa proporciona las funcionalidades requeridas por los usuarios así como también la lógica de negocios de la aplicación, es decir todas las reglas que deben cumplirse para realizar cada una de las transacciones que realice el sistema. Además esta capa es un puente de comunicación entre la capa de presentación y la capa de datos.

Esta capa del preprocesador de planes académicos está compuesta por los siguientes módulos:

- Módulo de Obtención de Planes Académicos.
- Módulo de Análisis de Contenidos.
- Módulo de Almacenamiento.
- Módulo de Generación de Archivos.

2.3.Capa de Datos

Su función es almacenar y devolver los datos a la capa de negocio, esta capa es la única que puede acceder a los mismos. En este caso, esta capa se encuentra formada por un gestor de bases de datos, compuesto por dos bases de datos localizadas en un mismo servidor.

Esta capa del preprocesador de planes académicos está conformada por:

- Base de Datos de Planes Académicos de UTPL
- Base de Datos del Preprocesador de Planes Académicos de UTPL

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Arquitectura de Software	Fecha: 13/09/2012
PPA-ARS	

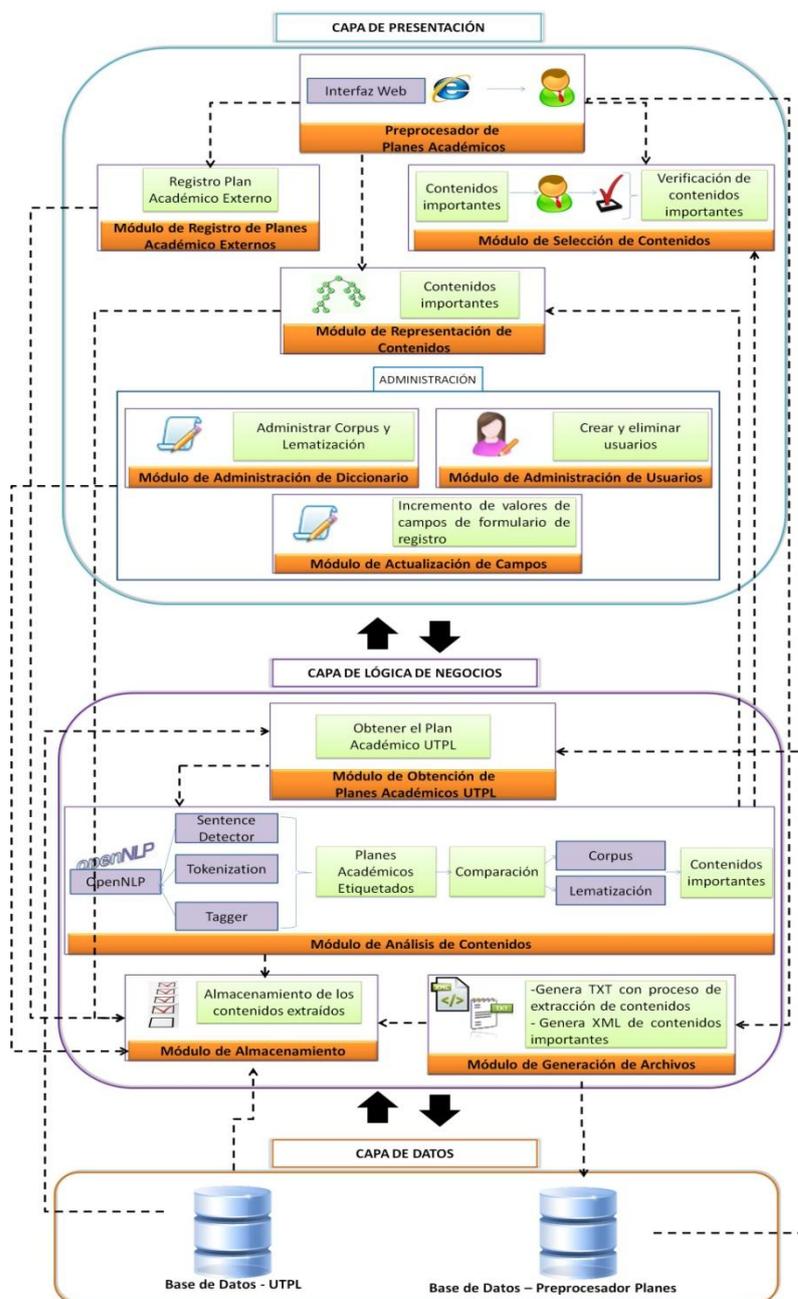


Figura. B-5 Vista Lógica de la Arquitectura Detallada - Preprocesador de Planes Académicos

3. Vista de Despliegue

El diagrama de la vista de despliegue muestra de manera gráfica los nodos que conforman el preprocesador de planes académicos, así como también el medio de comunicación entre ellos. En la Figura B-6, se observa la vista de despliegue de este proyecto de software.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Arquitectura de Software	Fecha: 13/09/2012
PPA-ARS	

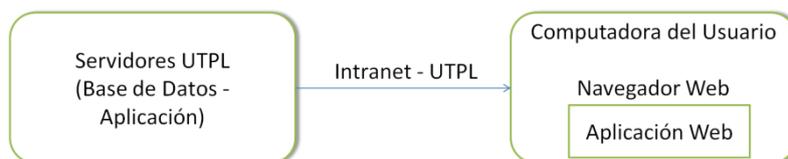


Figura. B-6 Vista de despliegue - Preprocesador de Planes Académicos

3.1. Servidor UTPL – Base de Datos

El servidor de UTPL en donde se encuentra tanto el servidor de base de datos, como el servidor de aplicaciones. Los diferentes usuarios tendrán acceso al servidor de aplicaciones a través de la intranet de la universidad.

3.2. Computador de usuario

Son las computadoras que cada usuario a través del navegador web de la misma accede a aplicación web del Preprocesador de Planes Académicos. Las computadoras se conectan al servidor de aplicaciones vía intranet de la universidad.

4. Calidad

4.1. Usabilidad

El preprocesador de planes académicos, permitirá un manejo intuitivo por parte de los usuarios.

4.2. Seguridad

El sistema web permitirá el acceso a funcionalidades dependiendo del perfil del usuario que ingresa al sistema, validando su ingreso a través de un usuario y clave.

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Arquitectura de Software	Fecha: 13/09/2012
PPA-ARS	

4.3.Confiabilidad

El sistema web del preprocesador de planes académicos, validará la información ingresada por el usuario, por lo cual mostrará mensajes o alertas que expliquen al usuario acerca de los errores cometidos tanto por el usuario como fallos del sistema.

4.4.Portabilidad

Al ser un sistema web el preprocesador de planes académicos, es independiente de la plataforma debido a que únicamente es necesario utilizar un navegador web¹⁹ para poder acceder al sistema web.

4.5.Mantenibilidad

El sistema web será flexible y de fácil mantenimiento, debido a que el sistema se encuentra construido a través de diferentes módulos.

¹⁹ Navegadores Web: Mozilla Firefox , Internet Explorer y Google Chrome

Anexo C
Manuales

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

C.1. Manual de Usuario

1. Introducción

La Universidad Técnica Particular de Loja, dentro de los procesos de gestión académica que lleva a cabo se encuentra la convalidación de asignaturas para los estudiantes que han cursado programas académicos en otras instituciones educativas. Este proceso es realizado de forma manual por diferentes docentes de cada titulación en calidad de fiscales quienes verifican que los planes académicos cumplan con una serie de requisitos para que dicha asignatura sea convalidada.

La UTPL, en sus constantes procesos de mejora plantea la creación de un Sistema Semi-automático de Convalidación de Planes Académicos, el cual está conformado por un conjunto de componentes, uno de ellos el Preprocesador de Planes Académicos el cual tiene como objetivo extraer los contenidos más relevantes de un plan académico externo frente a la comparación con los planes académicos internos.

El Preprocesador de Planes Académicos es un sistema web semi-automático en el cual se han establecido los siguientes perfiles:

- **Secretaria:** Quien realiza el registro de los planes académicos externos en el sistema.
- **Fiscal:** Quien realiza el análisis de los planes académicos en el sistema.
- **Administrador:** Quien realiza la administración de los componentes que conforman el sistema.

A continuación se muestran las actividades que puede realizar cada uno de los perfiles señalados anteriormente.

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

Manual de portal web: PREPROCESADOR DE PLANES ACADÉMICOS

Al acceder al portal web del Preprocesador de Planes Académicos, se presenta la pantalla de autenticación en el sistema, tal como se muestra en la Figura C-1.



Figura. C-1 Autenticación en el Sistema

El usuario ingresa su nombre de usuario y contraseña. Luego de estos datos hayan sido verificados, se mostrará las funcionalidades del mismo en base al perfil correspondiente.

2. Secretaria

El menú que posee el Preprocesador de Planes Académicos para el perfil Secretaria consta de:

- Inicio
- Registro

A continuación se describen cada uno de los componentes del menú.

2.1.Inicio

La presentación inicial para el perfil Secretaria, es la que se muestra en la Figura C-2 a continuación.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	



Figura. C-2 Pantalla de Inicio – Perfil Secretaria

2.2. Registro

En este menú se realiza el registro de los planes académicos externos, para lo cual el usuario con perfil de secretaria deberá realizar los siguientes pasos:

1. Haga clic en el botón Browse (Examinar) para buscar el archivo digital del plan académico externo que se va a registrar, tal como se muestra en la Figura C-3.



Figura. C-3 Buscar archivo digital del Plan Académico Externo

2. Luego aparecerá el explorador de archivos del sistema operativo en el cual este trabajando en este caso Windows. Se deberá buscar dentro de los directorios del sistema, el archivo correspondiente al plan que se va a

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

registrar. Es importante tener en cuenta que el archivo del plan académico debe encontrarse en formato .pdf. En caso de no estarlo no podrá ser subido al sistema y no se podrá continuar con el registro del plan.

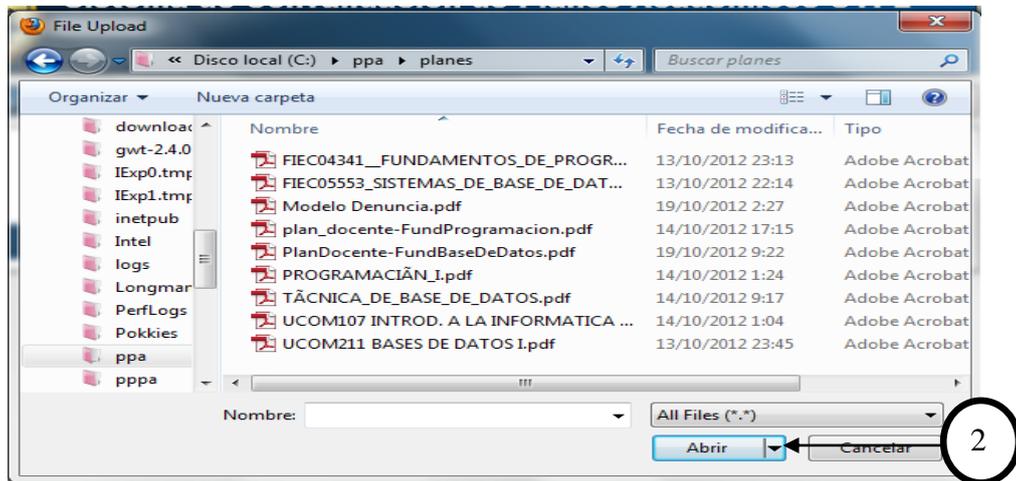


Figura. C-4 Buscar archivo en el sistema

3. Hacer clic en enviar para almacenar el plan académico en el servidor.



Figura. C-5 Almacenar Plan Académico Externo en el servidor

Si el usuario no eligió el archivo del plan académico, o este se encuentra en otra extensión no permitida se mostrará nuevamente la Figura C-3.

Si no existe ningún inconveniente al momento de subir el archivo, se visualizara la Figura C-6.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

4. En la Figura C-6, se observa el formulario de Registro de información general de Planes Académicos Externos, el cual deberá ser llenado en su totalidad.

Figura. C-6 Formulario de Registro de Planes Académicos Externos

5. Para agregar los contenidos del plan, haga clic en el botón más (+), luego de lo cual se observará la Figura C-7.

Figura. C-7 Agregar contenidos al Plan Académico Externo

6. En caso de querer eliminar alguna unidad que haya sido ingresada en el

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

formulario se debe hacer clic en el botón menos (-).

7. Para guardar los datos colocados en el formulario se debe hacer clic en botón Guardar. En caso contrario se debe hacer clic en el botón cancelar.

Si los datos se almacenaron correctamente se podrá observar la Figura C-8.

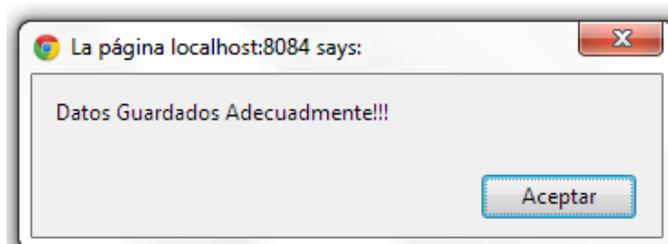


Figura. C-8 Alerta – Almacenamiento correcto del plan

En caso de que los datos no se hayan almacenado correctamente se observará una alerta indicando lo sucedido.

8. Para abandonar del sistema, el usuario deberá hacer clic en el menú Salir.

3. Fiscal

El menú que posee el Preprocesador de Planes Académicos para el perfil Fiscal consta de:

- Inicio
- Análisis Específico
- Análisis Completo

A continuación se describen cada uno de los componentes del menú.

3.1.Inicio

La presentación inicial para el perfil Fiscal, es la que se muestra en la Figura C-9 a continuación.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	



Figura. C-9 Pantalla de Inicio – Perfil Fiscal

3.2. Análisis Específico

En este menú se realiza el análisis de un plan académico externo frente a la comparación con un plan académico interno. Para realizar este análisis el usuario con perfil de fiscal deberá realizar los siguientes pasos:

1. Ingresar el código del plan que se desee analizar y luego haga clic en el botón buscar, tal como se muestra en la Figura C-10.



Figura. C-10 Buscar Plan Académico Externo – Código de Plan

2. Si el código ingresado es correcto se mostrará la siguiente interfaz que se muestra en la Figura C-11.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

The screenshot shows the 'ANÁLISIS ESPECÍFICO DE PLANES ACADÉMICOS' section. At the top, there is a search box titled 'Buscar Plan Academico' with a 'Codigo del Plan:' field and a 'Buscar' button. Below this is a link for 'Búsqueda Avanzada'. The main content area is titled 'DATOS DEL PLAN ACADEMICO' and contains the following information:

- Asignatura: Sistemas de Base de Datos I
- Centro: Loja
- Universidad: Universidad de Especialidades Espíritu Santo
- Modalidad: Presencial
- Carrera: Ingeniería en Sistemas

Below this information, there is an 'Analizar con:' section with a 'Carrera:' dropdown menu set to 'Economía' and an 'Asignatura:' dropdown menu set to 'Macroeconomía'. An 'Analizar Plan' button is located at the bottom of this section. The footer of the page reads 'Copyright © 2012. Todos los derechos reservados.'

Figura. C-11 Resultado de Búsqueda

3. En caso de no conocer el código del plan académico externo, haga clic en el enlace Búsqueda Avanzada, luego de lo cual se muestra el formulario de búsqueda avanzada tal como se muestra en la Figura C-12.

The screenshot shows the 'ANÁLISIS ESPECÍFICO DE PLANES ACADÉMICOS' section. At the top, there is a search box titled 'Buscar Plan Academico' with a 'Nombre de la Asignatura:' field containing the text 'base' and a 'Buscar' button. A circled number '3' with an arrow points to the 'Buscar' button. The footer of the page reads 'Copyright © 2012. Todos los derechos reservados.'

Figura. C-12 Búsqueda avanzada de Planes Académicos Externos

4. Ingresar el nombre o parte del nombre de la asignatura del plan académico que se está buscando y hacer clic en buscar, tal como se puede observar en la Figura C-12. El resultado de la búsqueda se mostrará de forma similar a la Figura C-13.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	



ANÁLISIS ESPECÍFICO DE PLANES ACADÉMICOS

Buscar Plan Académico

Codigo del Plan:

[Búsqueda Avanzada](#)

Nombre del Plan	Unidad Educativa	Carrera	Seleccionar Plan
Sistemas de Base de Datos I	Escuela Superior Politécnica del Litoral	Ingeniería en Computación	Seleccionar
Sistemas de Base de Datos I	Universidad de Especialidades Espíritu Santo	Ingeniería en Sistemas	Seleccionar
Técnica de Base de Datos	Universidad Autónoma de Quito	Informática	Seleccionar
Fundamentos de Base de Datos	Universidad Técnica Particular de Loja	Ingeniería en Sistemas Informáticos y Computación	Seleccionar
BASE DE DATOS AVANZADAS (UTPL-ECTS-CONESUP)	Universidad Técnica Particular de Loja	Sistemas Informáticos y Computación	Seleccionar

Copyright © 2012. Todos los derechos reservados.

Figura. C-13 Resultado de Búsqueda Avanzada

5. Haga clic en el enlace Seleccionar del plan académico que desee analizar. Luego se mostrarán los datos generales del plan seleccionado, igual que la Figura C-11.
6. Para seleccionar con que plan académico interno se va analizar el plan académico externo ya seleccionado, se deberá desplazar los combo box correspondiente tanto a la carrera como a la asignatura, tal como muestra la Figura C-14.
7. Hacer clic en el botón Analizar Plan. Se deberá esperar unos segundos para que el sistema muestre los resultados fruto del proceso del análisis.
8. En la Figura C-16, se observa el resultado del proceso de análisis específico de un plan académico externo frente a un plan académico interno.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	



Figura. C-14 Selección de Plan Académico Interno

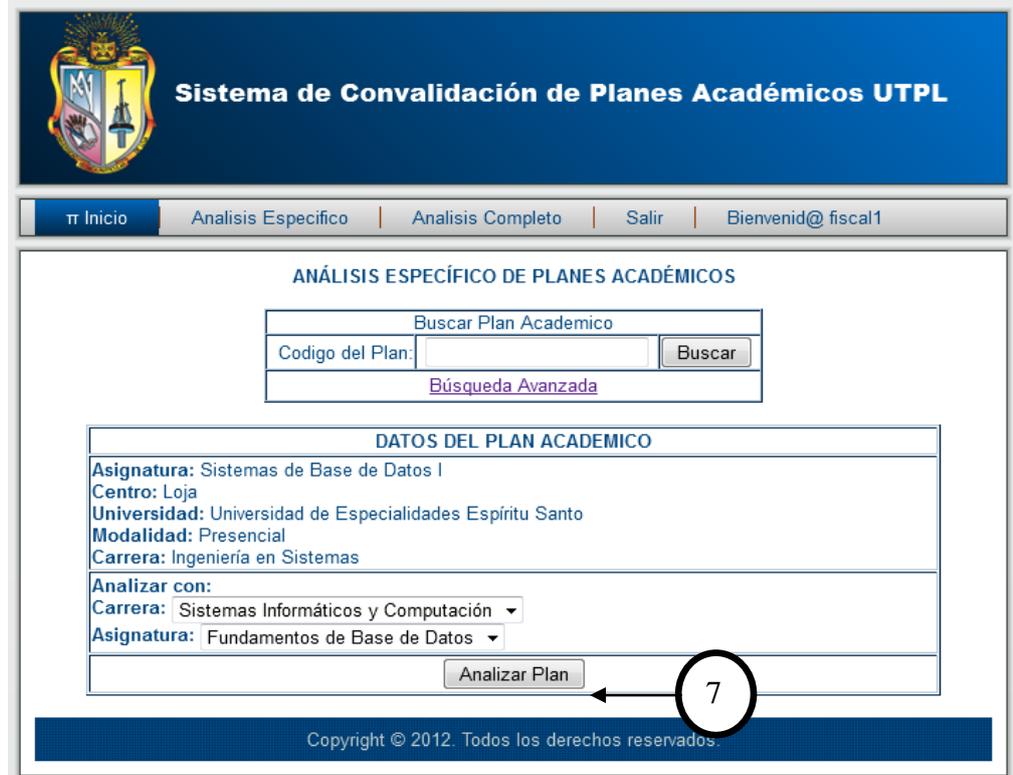


Figura. C-15 Analizar Plan Académico – Específico

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

herramienta. Estos contenidos seleccionados pueden ser editados en base al criterio del usuario fiscal que está haciendo uso del sistema.

10. Si desea quitar un contenido de los seleccionados por la herramienta, haga clic en la equis (x) perteneciente al contenido que desea eliminar.
11. Si desea agregar un contenido que no haya sido seleccionado por la herramienta, haga clic en el contenido que desee agregar y luego arrástrelo a una de las celdas de la tabla perteneciente a los contenidos seleccionados.
12. Para guardar los contenidos importantes del plan académico seleccionado, haga clic en el botón guardar. Si los datos se guardaron correctamente se mostrará la alerta de la Figura C-17. Para continuar haga clic en el botón aceptar.

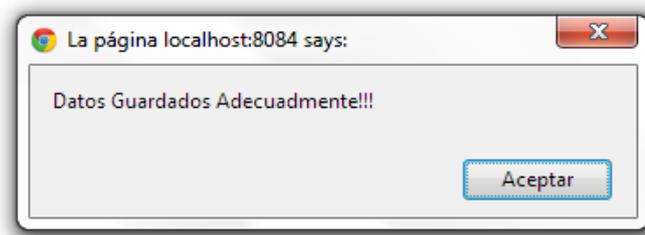


Figura. C-17 Alerta – Almacenamiento correcto de contenidos importantes

En caso de que los datos no se hayan almacenado correctamente se observará una alerta indicando lo sucedido

13. Una vez analizado el plan académico, este presenta tres funcionalidades adicionales que se observan en la Figura C-18 y que se describen a continuación:

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

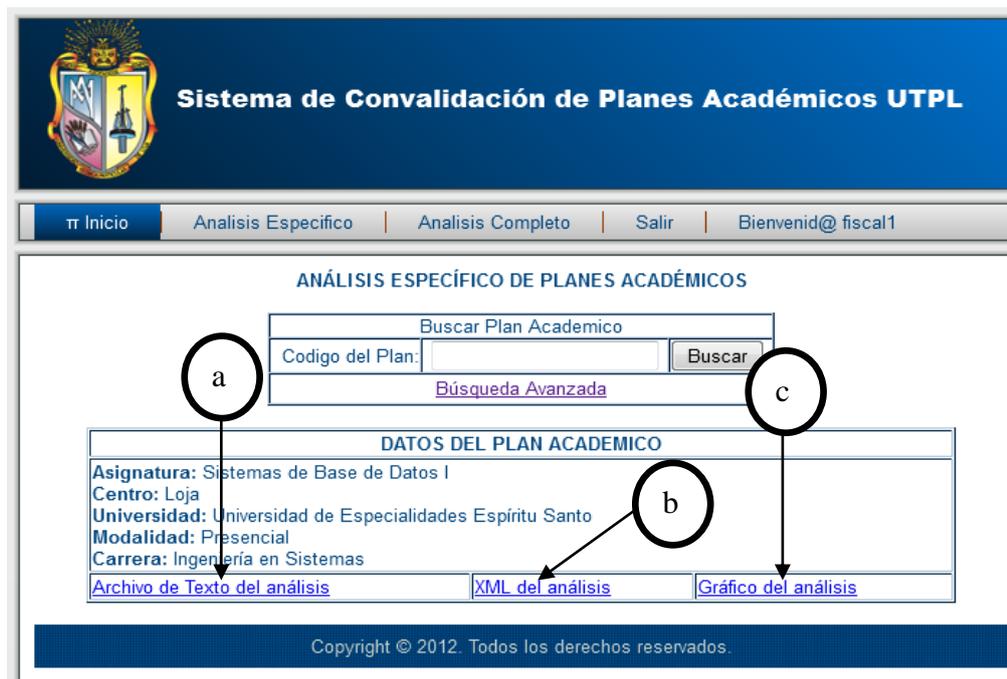


Figura. C-18 Archivos de Descarga – Análisis Específico

- **Descargar archivo de texto del análisis**

Haga clic en el enlace Archivo de texto del análisis. Luego podrá observar la imagen que se muestra en la Figura C-19, en donde se seleccionará la ruta en donde se descargará del sistema el archivo de texto en donde se encuentra el proceso detallado de análisis realizado por el sistema.

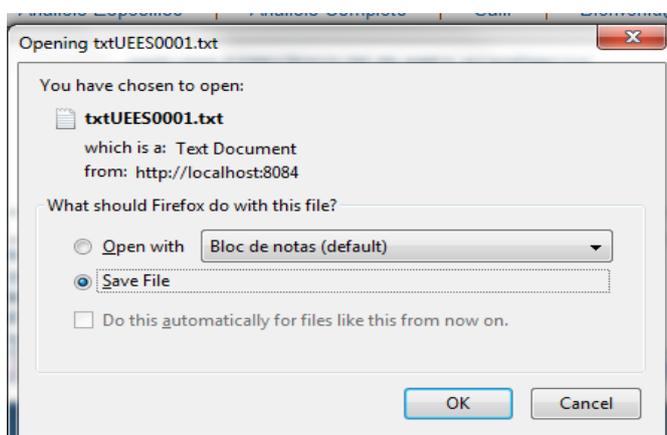


Figura. C-19 Descargar archivo de texto de análisis específico

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

- **Descargar XML de contenidos importantes**

Haga clic en el enlace XML del análisis. Luego podrá observar la imagen que se muestra en la Figura C-20, en donde se seleccionará la ruta en donde se descargará del sistema el archivo de XML en donde se encuentran representados en este formato los contenidos más importantes obtenidos por el análisis específico del sistema.

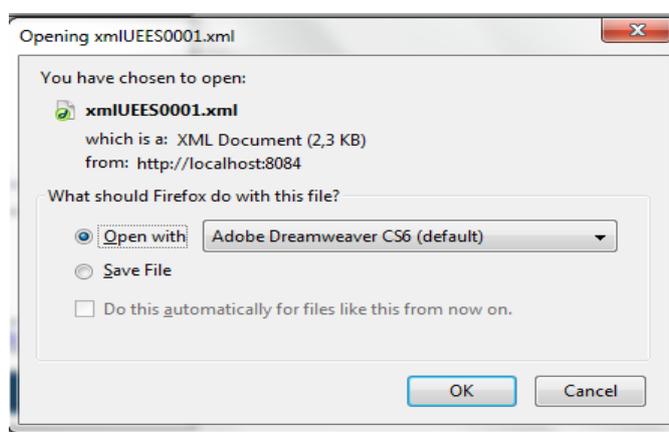


Figura. C-20 Descargar archivo XML de análisis específico

- **Gráfico de análisis**

Haga clic en el enlace, Gráfico del análisis, para observar de forma gráfica los contenidos más importantes extraídos de la comparación de los planes académicos, tal como se muestra en la Figura C-21.

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

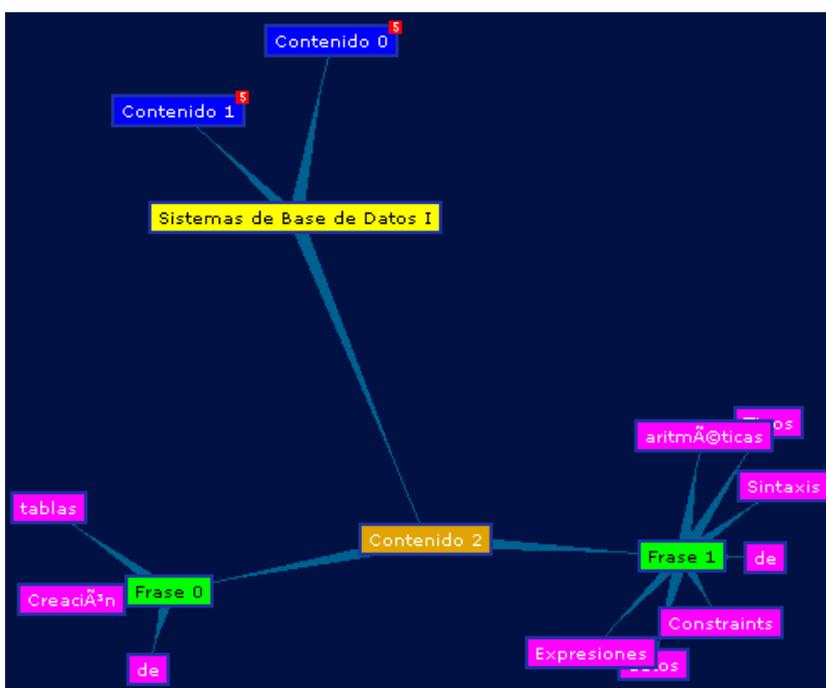


Figura. C-21 Gráfico de Análisis

3.3. Análisis completo

En este menú se realiza el análisis de un plan académico externo en base a la comparación con todos los planes académicos internos que se encuentran registrados en el sistema. Para realizar este análisis el usuario con perfil de fiscal deberá realizar los siguientes pasos:

1. Para realizar la búsqueda del plan académico a analizar, se realizan los pasos del 1 al 7 indicados en el Análisis Específico.
2. En la Figura C-22, se observa el resultado del proceso de análisis completo de un plan académico externo frente a todos los planes académicos internos registrados en el sistema.
3. Haga clic en el enlace XML del análisis. Luego podrá observar la imagen que se muestra en la Figura C-23, en donde se seleccionará la ruta en donde se descargará del sistema el archivo de XML en donde se encuentran representados en este formato los contenidos más importantes obtenidos por

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

el análisis completo del sistema

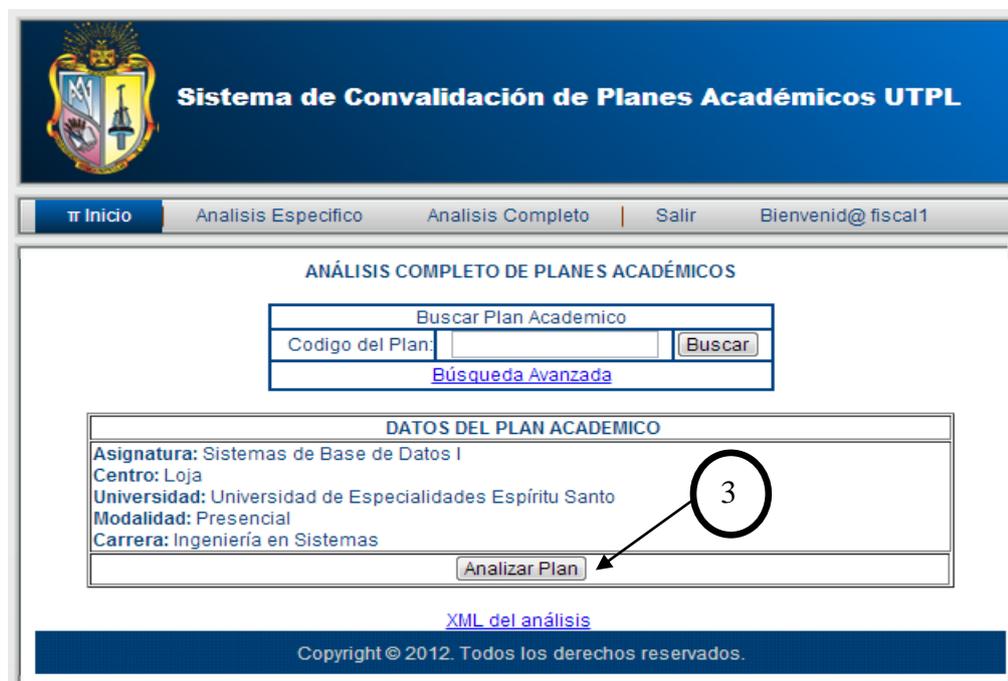


Figura. C-22 Resultado de Análisis Completo

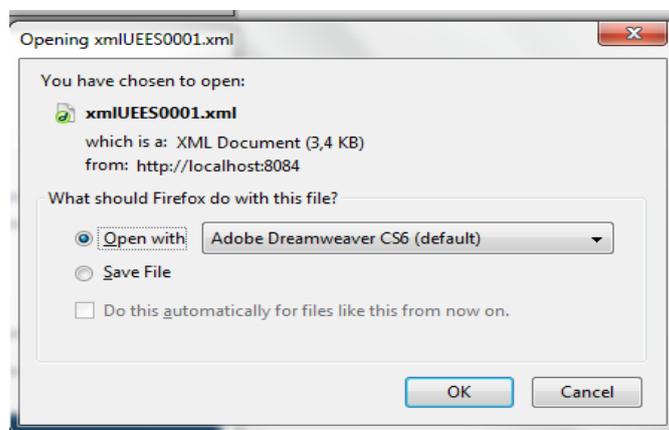


Figura. C-23 Descargar archivo XML de análisis completo

4. Para abandonar del sistema, el usuario deberá hacer clic en el menú Salir.

4. Administrador

El menú que posee el Preprocesador de Planes Académicos para el perfil Administrador consta de:

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

- Inicio
- Diccionario
- Usuarios
- Administración de campos

A continuación se describen cada uno de los componentes del menú.

4.1.Inicio

La presentación inicial para el perfil Administrador, es la que se muestra en la Figura C-24 a continuación.



Figura. C-24 Pantalla de Inicio – Perfil Administrador

4.2.Diccionario

El módulo Diccionario permite realizar las siguientes funcionalidades, tal como se muestra en la Figura C-25.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

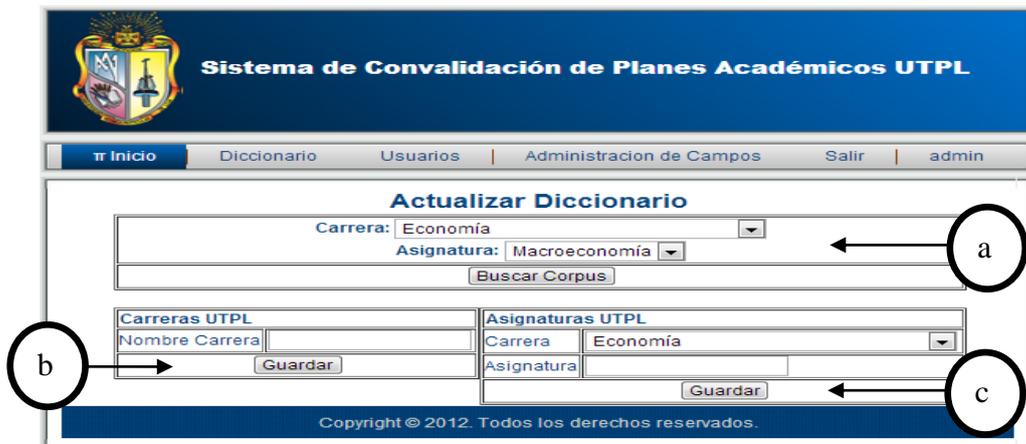


Figura. C-25 Funcionalidades del Módulo Diccionario

- **Actualizar Diccionario**

1. Seleccione la carrera y la asignatura a la cual desea actualizar el diccionario, luego presione el botón Buscar. Inmediatamente se mostrará la Figura C-26.



Figura. C-26 Búsqueda de Diccionario

2. Para agregar una nueva palabra al corpus seleccionado seleccione el Tipo de

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

corpus, sea genérico (palabras que no son propias de la asignatura que se está registrando) o específico (palabras que son propias de la asignatura que se está registrando). Luego ingrese la nueva palabra al corpus en su forma base (si es específico palabras de tipo sustantivo o verbos, si es genérico palabras de tipo verbo). Finalmente ingrese la primera lematización (conjugaciones del verbo, forma plural de los sustantivos, etc.). Finalmente haga clic en el botón guardar.

3. Para agregar una nueva lematización seleccione la palabra del corpus a la cual desea agregar la lematización, ingrese la palabra en el campo lematización. Finalmente haga clic en el botón guardar.

- **Agregar nuevas carreras de UTPL**

4. Para agregar una nueva carrera ingrese el nombre de la misma en el campo correspondiente y presione el botón Guardar, tal como se muestra en la Figura C-27. Esta carrera servirá para registrar nuevos corpus en base a la carrera y asignatura.

- **Agregar nuevas asignaturas de UTPL**

5. Para agregar una nueva asignatura, seleccione la carrera a la cual pertenece la asignatura que se va a ingresar, seguidamente ingrese el nombre de la nueva asignatura, tal como se muestra en la Figura C-27. Esta asignatura servirá para registrar nuevos corpus en base a la carrera y asignatura.

Copyright © 2012. Todos los derechos reservados.

Figura. C-27 Agregar nuevas carreras y asignaturas de UTPL

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

5. Usuarios

Este módulo permite la creación y eliminación de usuarios del sistema.

- **Nuevo Usuario**

Para registrar un nuevo usuario ingrese los datos en el formulario de registro de nuevo usuario, como son: nombres, apellidos, usuario, rol y password. Finalmente presione el botón Enviar. Este formulario se lo puede observar en la Figura C-28.

Sistema de Convalidación de Planes Académicos UTP

Inicio | Diccionario | Usuarios | Administración de Campos | Salir | admin

Administración de Usuarios

NOMBRES	APELLIDOS	USERNAME	ROL	borrar
Eylin Dolores	Calderón Carrión	edcalderon	administrador	X
FISCAL	UNO	fiscal1	fiscal	X
SECRETARIA	UNO	secretaria1	secretaria	X
Administrador	Administrador	admin	administrador	X

Nuevo Usuario

Nombres : jose

Apellidos : salazar

Username : josalar

Rol : Fiscal Administrador Secretaria

Password :

Enviar

Copyright © 2012. Todos los derechos reservados.

Figura. C-28 Administración de Usuarios

- **Eliminar Usuarios**

Para eliminar un usuario haga clic en la equis (x) perteneciente al usuario que se desee borrar de los registros de la base de datos del sistema.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual de Usuario	Fecha: 15/10/2012
PPA-MAU	

6. Administración de Campos



Sistema de Convalidación de Planes Académicos UTPL

π Inicio | Diccionario | Usuarios | Administración de Campos | Salir | admin

Actualización de Campos

Lugar - Centro UTPL	<input type="text"/>	Guardar
Modalidad	<input type="text"/>	Guardar
Periodo Academico	<input type="text"/>	Guardar
Entidad Educativa	<input type="text"/>	Guardar

Nueva Carrera para Planes

Entidad Educativa	Universidad Técnica Particular de Loja	▼
Carrera	<input type="text"/>	
Enviar		

Copyright © 2012. Todos los derechos reservados.

Figura. C-29 Administración de Campos

Este módulo permite realizar la administración de los campos del formulario de registro de planes externos. Estos campos son: Centro UTPL, modalidad, periodo académico, entidad educativa y carrera. Para ingresar un nuevo valor se ingresa este en el lugar correspondiente y luego hacer clic en el botón Guardar. El módulo de administración de campos, se muestra en la Figura C-29.

Preprocesador de Planes Académicos de la UTP	Versión: 1.0
Manual del Programador	Fecha: 15/10/2012
PPA-MAP	

C.2. Manual del Programador

1. Introducción

Con la finalidad de dar a conocer al programador de manera detallada como se encuentra construido el Preprocesador de Planes Académicos, se elabora el siguiente manual, en donde se detallan aspectos técnicos referentes a las bases de datos utilizadas y al diagrama de clases referentes al proyecto.

Este proyecto se encuentra implementado en una arquitectura donde la carga está dividida en tres capas que son: capa de presentación, capa de lógica de negocios y capa de datos, dichas capas se verán reflejadas en el diagrama de clases especificado en este manual.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual del Programador	Fecha: 15/10/2012
PPA-MAP	

2. Diagrama de Base de Datos

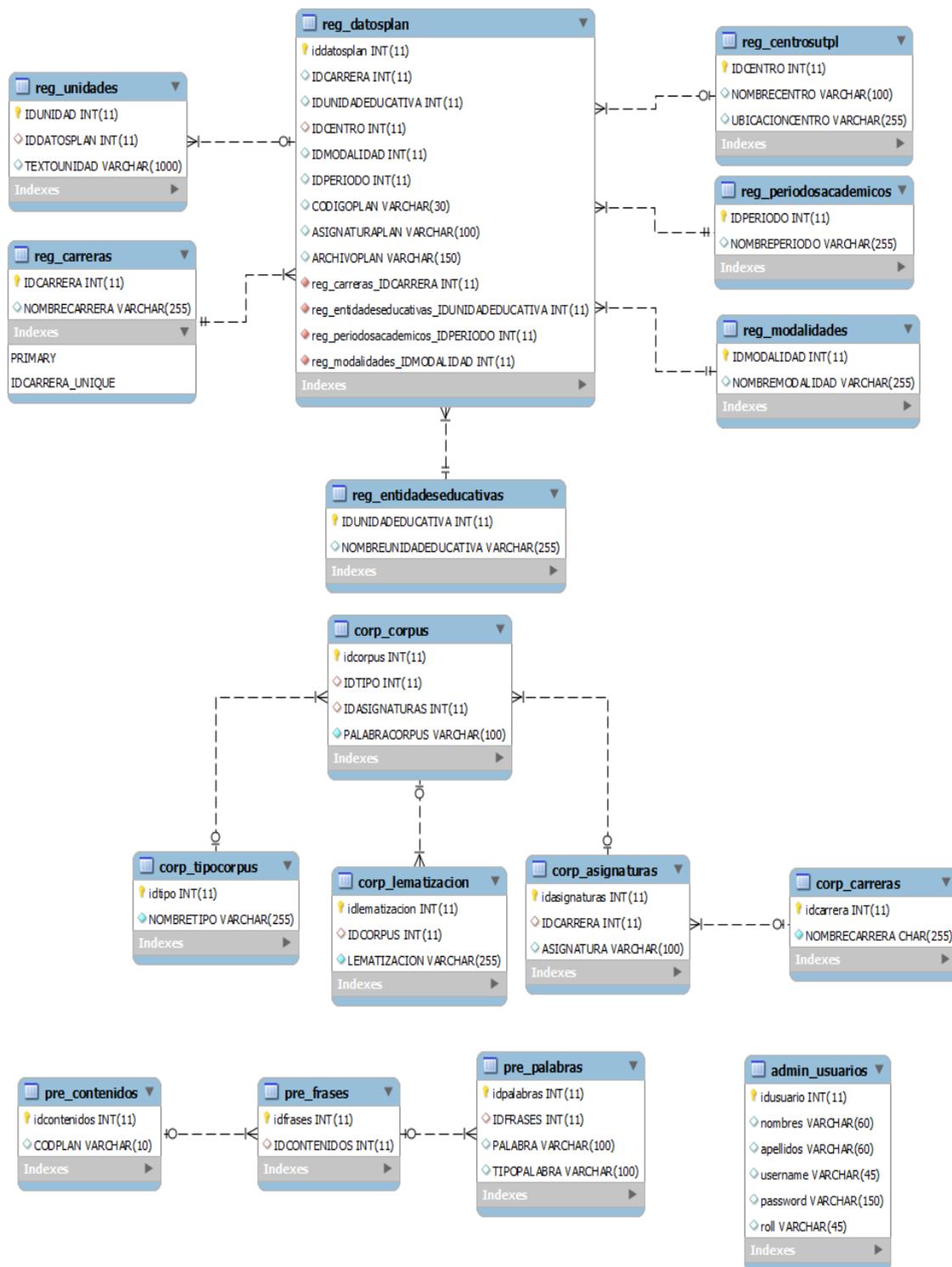


Figura. C-30 Diseño de Base de Datos *preprocesadorutpl* – Manual del Programador

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual del Programador	Fecha: 15/10/2012
PPA-MAP	

La base de datos para el Preprocesador de Planes Académicos, se encuentra almacenada en un servidor MySQL versión 5.5, bajo el nombre de *preprocesadorutpl*. Esta base de datos está compuesta por 16 entidades, las cuales se describen a continuación.

Registro de Planes Académicos externos

- **reg_carreras:** Almacena el nombre de la carrera a la cual pertenece el plan académico que registra el usuario en el sistema.
- **reg_centrosutpl:** En esta tabla se almacena el nombre del centro de UTPL en donde se registra el plan académico.
- **reg_datosplan:** Esta tabla registra los datos generales del plan académico externo, como son: entidad educativa, carrera, asignatura, centro utpl, periodo, código de plan académico y nombre del archivo digital del plan.
- **reg_entidadeseducativas:** Se guarda en esta tabla el nombre de la entidad educativa a la cual pertenece el plan académico.
- **reg_modalidades:** Se almacena en esta tabla la modalidad de estudios a la cual pertenece el plan académico externo.
- **reg_periodosacademicos:** En esta tabla se guarda el periodo académico al cual pertenece el plan académico que se registra.
- **reg_unidades:** En esta tabla se almacena el contenido de cada una de las unidades que posee el plan académico.

Diccionario

- **corp_asignaturas:** Esta tabla almacena el nombre de la asignatura a la cual pertenece un corpus.
- **corp_carreras:** Esta tabla almacena el nombre de la carrera a la cual pertenece un corpus.
- **corp_corpus:** Se asocia dentro de esta tabla las palabras importantes de una asignatura de una carrera en su forma base.
- **corp_lematizacion:** En esta tabla se guarda las formas más comunes

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual del Programador	Fecha: 15/10/2012
PPA-MAP	

de cada una de las palabras que se encuentran en el corpus.

- **corp_tipocorpus:** Dentro de esta tabla se almacena el tipo de corpus (genérico, específico) al cual pertenece el corpus que se está registrando.

Preprocesador

- **pre_contenidos:** En esta tabla se almacena el identificador del contenido importante obtenido por el sistema y el código del plan al que se encuentra relacionado dicho contenido.
- **pre_frases:** En esta tabla se guarda el identificador de la frase importante obtenida por el sistema y el identificador del contenido al cual se encuentra relacionado la frase.
- **pre_palabras:** Se guarda en esta tabla, la palabra importante extraída por el sistema, el identificador de la frase a la cual está ligada esta palabra y el tipo de palabra (etiquetas Eagles) a la que pertenece.

Usuarios

- **admin_usuarios:** En esta tabla se almacenan los datos informativos sobre los usuarios que utilizan el sistema, así como también datos sobre usuario, contraseña y roll al cual pertenece dicho usuario.

El esquema gráfico de la base de datos para el Preprocesador de la Base de Datos se lo puede observar en la Figura C-30.

Además de la base de datos preprocesador utpl, se hace uso de la base de datos de registro de planes académicos de la UTPL, que se encuentra bajo el nombre *plandocente*, la cual se encuentra almacenada en un servidor MySQL versión 5.5. Esta base de datos incluye un conjunto de tablas, dentro de las cuales van a ser utilizadas las tablas que se muestran en la Figura C-31, y que se describen a continuación.

- **plan_registrocontenido:** En esta entidad se almacenan los contenidos del plan académico, métodos de evaluación y

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual del Programador	Fecha: 15/10/2012
PPA-MAP	

acreditación.

- **plan_plan:** En esta tabla se registran los datos generales del plan académico, como son nombre de asignatura, justificación, prerequisites, idioma, temporalidad, periodo, fecha y ciclo.
- **plan_periodo:** Esta entidad es usada para almacenar la descripción del periodo para el cual se ha desarrollado el plan académico.
- **plan_carrera:** Esta tabla es utilizada para almacenar la información correspondiente al nombre de la carrera, modalidad, nivel y unidad académica a la que pertenece el plan académico.
- **plan_asignatura:** Esta entidad es usada para guardar la información perteneciente al nombre de la asignatura, créditos, tipo y malla a la que pertenece el plan académico.

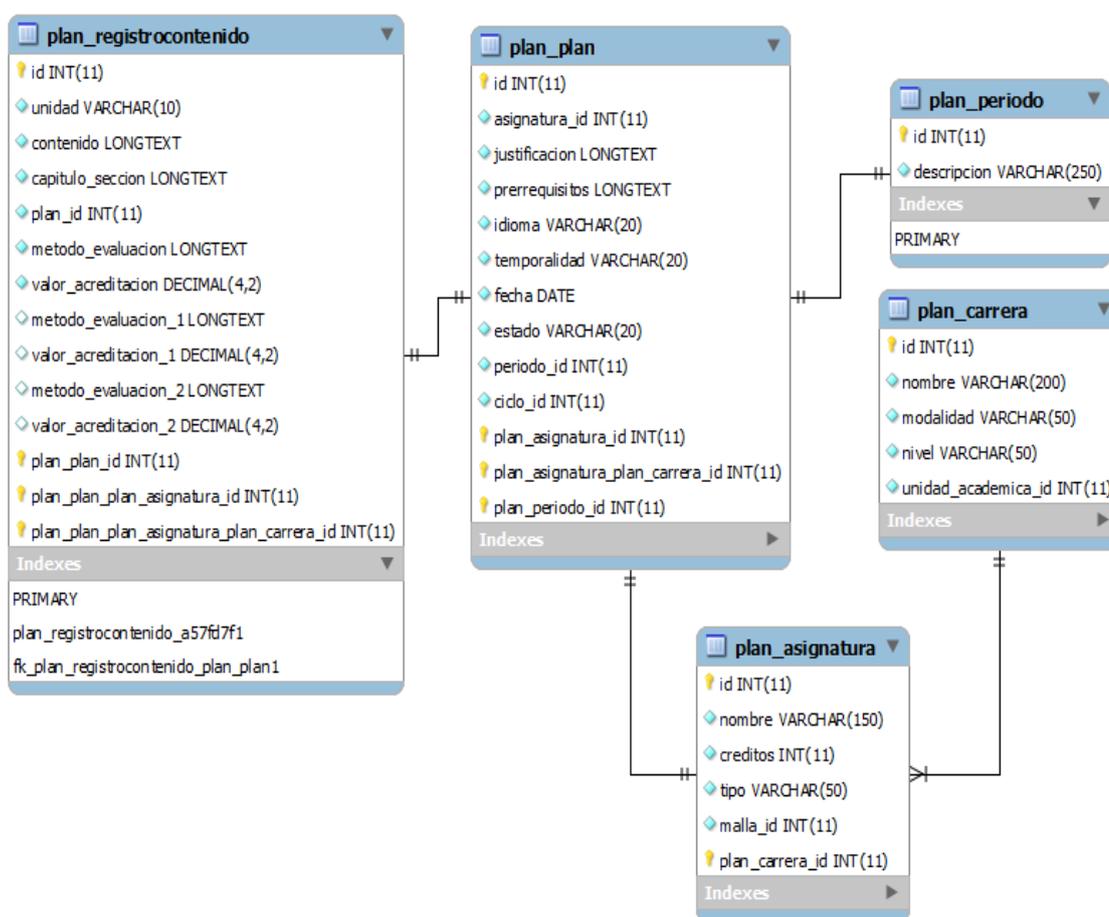


Figura. C-31 Diseño de Base de Datos *plandocente* – Manual del Programador

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual del Programador	Fecha: 15/10/2012
PPA-MAP	

3. Diagrama de Clases

El preprocesador de planes académicos, es un sistema web semiautomático el cual fue desarrollado en base a las siguientes características técnicas que se indican en la Tabla C-1.

Característica	Descripción
Lenguaje de Programación	Java
IDE de Desarrollo	Netbeans 7.0.1
Máquina Virtual de Java	7
Servidor Web	Apache Tomcat 7.0.14.0

Tabla C-1 Características Técnicas

Este preprocesador está compuesto por una serie de clases que han sido organizados en paquetes y estos a su vez se han dividido en las siguientes capas: Librerías, lógica de negocios y presentación. A continuación se describen de manera general los paquetes que conforman este proyecto. El diagrama de clases se lo puede observar en la Figura C-32.

Librerías

- **TouchGraph:** En este paquete se encuentran los archivos utilizados de esta librería para realizar el gráfico de representación de contenidos importantes.
- **DriverMySQL:** En este paquete se ubican los archivos utilizados de esta librería para realizar todos los procesos referentes a la base de datos MySQL.
- **OpenNLP:** En este paquete se localizan los archivos utilizados por esta librería para realizar el análisis (determinar sentencias, separar signos de puntuación y etiquetar palabras) de los planes académicos.
- **Oreilly:** En este paquete se pueden encontrar los archivos utilizados para presentar los contenidos importantes específicos extraídos y que además permiten se modificados dichos contenidos (pasar de contenido importante a no importante y viceversa).

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual del Programador	Fecha: 15/10/2012
PPA-MAP	

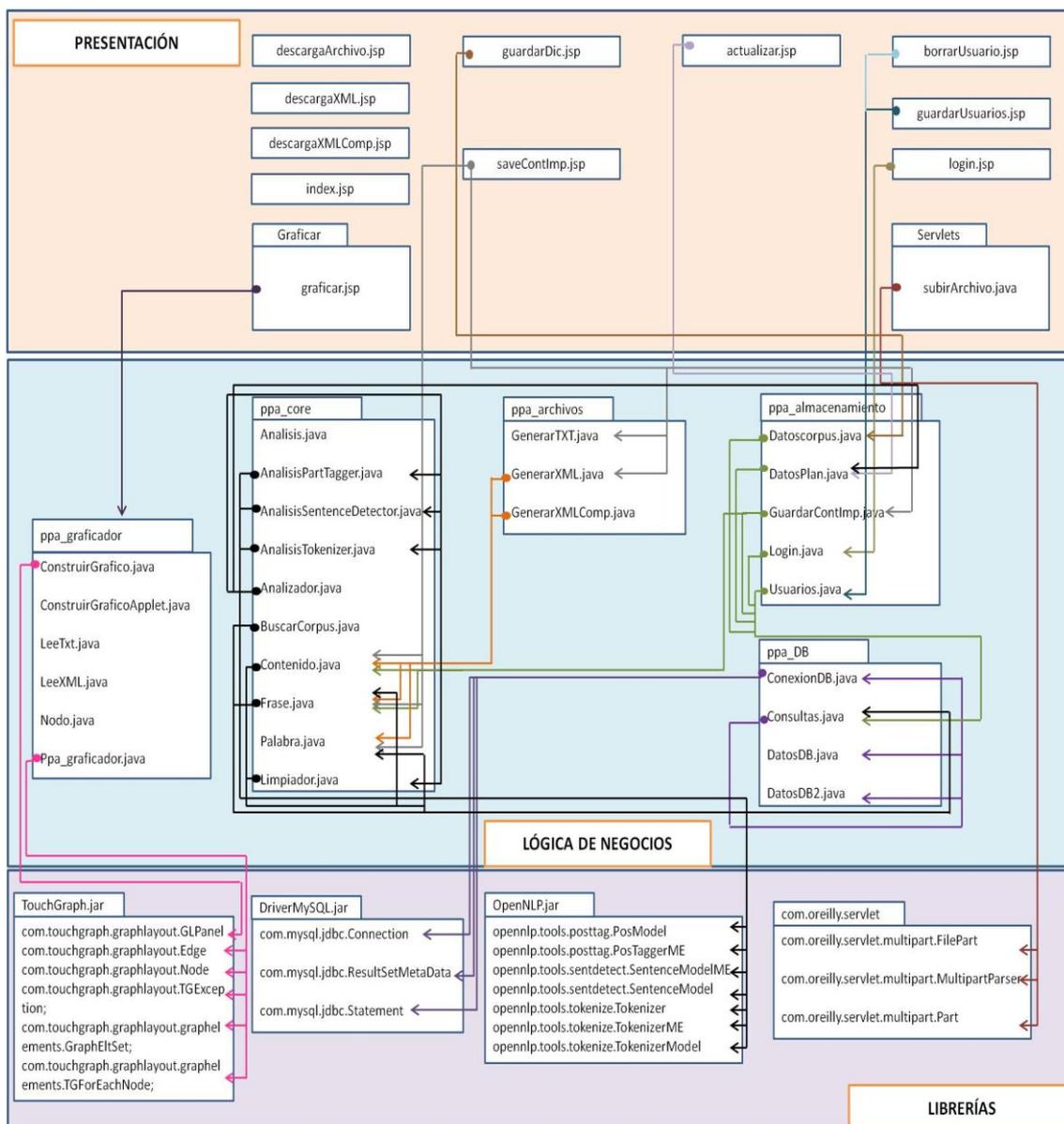


Figura. C-32 Diagrama de Clases

Lógica de Negocios

- **ppa_core:** En este paquete se encuentran las clases construidas para realizar el análisis y extracción de contenidos importantes del plan académico que se está analizando. Este paquete es el más importante en cuanto al tema del proyecto se refiere.
- **ppa_archivos:** En este paquete se ubican las clases elaboradas para generar los archivos (txt y XML), referente al proceso de análisis y extracción de contenidos importantes.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual del Programador	Fecha: 15/10/2012
PPA-MAP	

- **ppa_DB:** En este paquete se localizan las clases construidas para realizar las conexiones a las bases de datos.
- **ppa_almacenamiento:** En este paquete se encuentran las clases construidas para realizar el almacenamiento de los datos en las tablas correspondientes de la base de datos.

Presentación

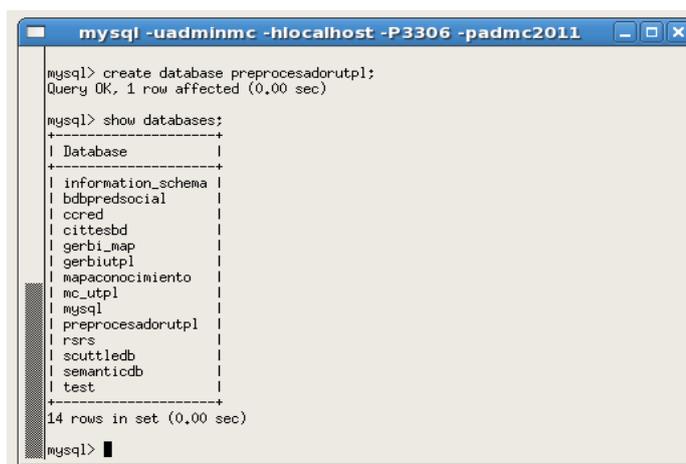
- En este paquete se encuentran cada uno de los archivos que son utilizados para gestionar la presentación de los resultados obtenidos en las capas inferiores.

4. Configuración e Implementación

Una vez revisada la estructura tanto de la base de datos y de la aplicación, a continuación se describe que archivos deben ser configurados al momento de implementar este sistema en un servidor.

4.1. Base de Datos

1. Crear las dos bases de datos con sus respectivos nombres (preprocesadorutpl, y plan docente). Una de las formas de realizar la creación es tal como se muestra en la Figura C-33.



```

mysql -uadminmc -hlocalhost -P3306 -padmc2011
mysql> create database preprocesadorutpl;
Query OK, 1 row affected (0.00 sec)

mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| bdbpredsocial |
| ccred |
| ctttesbd |
| gerbi_map |
| gerbiutpl |
| mapaconocimiento |
| mc_utpl |
| mysql |
| preprocesadorutpl |
| rsrs |
| scuttledb |
| semanticdb |
| test |
+-----+
14 rows in set (0.00 sec)

mysql>

```

Figura. C-33 Creación de Base de Datos por consola

2. Usando los scripts de las bases de datos, implementar la estructura de tablas y relaciones, tal como se indica en la Figura C-34

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual del Programador	Fecha: 15/10/2012
PPA-MAP	



Figura. C-34 Creación de tablas y relaciones

- Una vez ejecutado los dos scripts de la base de datos se podrá observar la estructura completa de las bases de datos.

4.2. Edición de archivos

Finalmente para que el sistema funcione adecuadamente, es necesario realizar la edición de algunas clases que componen el sistema. A continuación se especifica cuáles deben ser editadas para su correcto funcionamiento.

- **DatosDB.java:** En esta clase se debe de modificar la URL en donde se encuentra la base de datos *preprocesadorutpl*, así como también el usuario y la contraseña de la base de datos.
- **DatosDB2.java:** En esta clase se debe de modificar la URL en donde se encuentra la base de datos *plandocente*, así como también el usuario y la contraseña de la base de datos.
- **GenerarTXT.java:** En esta clase se debe establecer la ruta en donde se encuentra el directorio en el cual se van a almacenar los archivos de texto generados por el sistema, producto del análisis específico de los planes académicos.
- **GenerarXML.java:** En esta clase se debe establecer la ruta en donde se encuentra el directorio en el cual se van a almacenar los archivos XML generados por el sistema con los contenidos importantes, producto del análisis específico de los planes académicos.
- **GenerarXMLComp.java:** En esta clase se debe establecer la ruta en donde se encuentra el directorio en el cual se van a almacenar los archivos de XML generados por el sistema producto del análisis completo de los planes académicos.

Preprocesador de Planes Académicos de la UTPL	Versión: 1.0
Manual del Programador	Fecha: 15/10/2012
PPA-MAP	

- **AnalisisPartTagger.java:** En esta clase se debe establecer la ruta en donde se encuentra el directorio de donde se va a extraer el modelo *es-partSpeechTagger.model*.
- **AnalisisSentDetector.java:** En esta clase se debe establecer la ruta en donde se encuentra el directorio de donde se va a extraer el modelo *en-sent.bin*.
- **AnalisisTokenizer.java:** En esta clase se debe establecer la ruta en donde se encuentra el directorio de donde se va a extraer el modelo *en-token.bin*.
- **descargarXML.jsp:** En este archivo se debe establecer la misma ruta que se haya especificado en la clase *GenerarXML.java*, puesto que hace referencia al lugar donde se ubican los archivos XML de contenidos importantes producto del análisis específico de los planes académicos.
- **descargarXMLComp.jsp:** En este archivo se debe establecer la misma ruta que se haya especificado en la clase *GenerarXMLComp.java*, puesto que hace referencia al lugar donde se ubican los archivos XML de contenidos importantes producto del análisis completo de los planes académicos.
- **subirarchivo.java:** En esta clase se debe establecer la ruta en donde se encuentra el directorio en el cual se van a almacenar los planes académicos que se hayan registrado en el sistema.

Una vez realizadas todas las configuraciones se deberá abrir un navegador para hacer uso del Preprocesador de Planes Académicos.

Referencias Bibliográficas

Álvarez, Juan. (30 de Diciembre de 2007). *Sociedad Andaluza de Educación Matemática Thales*. Recuperado el 7 de Febrero de 2011, de <http://thales.cica.es/rd/Recursos/rd97/UnidadesDidacticas/53-1-u-punt14.html>

Apache OpenNLP, C. D. (s.f.). *openNLP*. Recuperado el 12 de Abril de 2012, de <http://opennlp.apache.org/documentation/1.5.2-incubating/manual/opennlp.html>

Benedí, J; Sánchez, j;. (2005). *Estimation of stochastic context-free grammars and their use as language models*. London: Journal Computer Speech and Language.

Bird, Steven; Klein, Ewan; Loper, Edward. (2009). *Natural Language Processing with Python*. EEUU: O'Reilly Media, Inc.

Broder, A., Fontoura, M., Josifovski, V., & Riedel, L. (24 de Agosto de 2011). *fontoura.org*. Recuperado el 14 de Abril de 2012, de <http://fontoura.org/papers/semsyn.pdf>

Caicedo Carvajal, J. M. (2 de Febrero de 2012). *Cavorite.com*. Recuperado el 2 de Abril de 2012, de <http://cavorite.com/labs/nlp/opennlp-models-es/>

Contreras, H. (Marzo de 2001). *Procesamiento del Lenguaje Natural basado en una gramática de estilos para el idioma español*. Recuperado el 10 de Enero de 2012

Contreras, Hilda. (Marzo de 2001). *Procesamiento del Lenguaje Natural basado en una gramática de estilos para el idioma español*. Recuperado el 10 de Enero de 2012

- Corchado, J., Bajo, J., Golinska, P., Giroux, S., & Corchuelo, R. (2012). *Advances in Intelligent and Soft Computing*. Springer.
- Covington, Michael. (1994). *Natural Language Processing for Prolog Programmers*. New Jersey: Prentice Hall, Englewood Cliffs.
- Fayyad, Usama. (1996). *kdnuggets*. Recuperado el 16 de Febrero de 2012, de <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>
- Foundation, T. A. (s.f.). *OpenNLP*. Recuperado el 12 de Abril de 2012, de <http://opennlp.apache.org/>
- Gate. (s.f.). *Gate General Architecture for text engineering*. Recuperado el 12 de Abril de 2012, de <http://gate.ac.uk/>
- Gómez, A., Aguado-de-Cea, G., Montiel-Ponsoda, E., & García, J. (01 de Marzo de 2010). *Monnet*. Recuperado el 18 de Abril de 2012, de <http://www.monnet-project.eu/Monnet/Monnet/English?init=true>
- Group, N. R. (2009). *Universidad Politécnica de Cataluña*. Recuperado el 12 de Abril de 2012, de <http://www.lsi.upc.edu/~nlp/tools/parole-sp.html>
- Gupta, V. (8 de Mayo de 2008). *Vangelist*. Recuperado el 5 de Mayo de 2012, de <http://vangelist.wordpress.com/2008/05/08/reuters-spotlight-apis/>
- IBM. (s.f.). *IBM*. Recuperado el 7 de Julio de 2012, de ftp://public.dhe.ibm.com/software/rational/web/datasheets/RUP_DS.pdf
- Konchady, M. *Building Search Applications Lucene, LingPipe and Gate*. Virginia: Mastru Publishing.
- Leiva, Isidoro; Rodríguez, José. (1996). *Universidad Complutense de Madrid*. Recuperado el 16 de Enero de 2012, de <http://www.ucm.es/BUCM/revistas/byd/11321873/articulos/RGID9696220205A>.

PDF

LingPipe. (s.f.). *LingPipe*. Recuperado el 12 de Abril de 2012, de <http://alias-i.com/lingpipe/>

Locke, W.N.; Booth, A.D. (1955). *Machine Translation of Languages*. Cambridge: Technology Press of MIT and Wiley.

López Yebra, G. (01 de Junio de 2012). *Academia.edu*. Recuperado el 7 de Agosto de 2012, de http://ipn.academia.edu/AdolfoGuzman/Papers/1766542/Sistema_de_extraccion_y_representacion_del_conocimiento_a_partir_de_documentos_descriptivos#outer_page_9

Lucene, A. (s.f.). *Apache Lucene*. Recuperado el 12 de Abril de 2012, de <http://lucene.apache.org/core/>

Macario, Hernandez. (2007). *Generador de los grafos conceptuales a partir de texto en español*. Mexico.

Manaris, Bill; Slator, Brian. (Julio de 1996). *IEEE*. Recuperado el 17 de Enero de 2012, de <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=511965&isnumber=11070>

Moreno, A;. (1998). *Lingüística Computacional. Introducción a los modelos simbólicos*. Madrid: Sintesis.

Nadkarni, P., Ohno-Machado, L., & Chapman, W. (4 de Julio de 2011). *Web of Knowledge*. Recuperado el 9 de Enero de 2012, de <http://jamia.bmj.com/content/18/5/544.full.pdf+html>

Rational, S. (Enero de 2011). *IBM*. Recuperado el 3 de Julio de 2012, de <http://www.ibm.com/developerworks/rational/library/content/03July/1000/1251/1>

251_bestpractices_TP026B.pdf

Reuters, T. (s.f.). *Calais*. Recuperado el 12 de Abril de 2012, de <http://www.opencalais.com/>

Sandoval, Antonio Moreno;. (1998). *Lingüística Computacional*. España: Síntesis.

Sarawagi, S. (2007). *Indian Institute of Technology*. Recuperado el 10 de Marzo de 2012, de <http://osm.cs.byu.edu/CS652s09/papers/Sarawagi.ieSurvey.pdf>

Seidler, K. (s.f.). *Service-oriented Information Extraction*. Recuperado el 9 de Enero de 2012

Sosa, Eduardo. (Enero de 1997). *El profesional de la información - Revista internacional científica y profesional*. Recuperado el 20 de Enero de 2012, de http://www.elprofesionaldelainformacion.com/contenidos/1997/enero/procesamiento_del_lenguaje_natural_revisin_del_estado_actual_bases_tericas_y_aplicaciones_parte_i.html

Technology, T. C. (2012). *CiteSeerX*. Recuperado el 12 de Mayo de 2012, de <http://citeseerx.ist.psu.edu/index>

Vallez, Mari; Pedraza, Rafael. (2007). *El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines*. Recuperado el 9 de Febrero de 2012, de <http://www.upf.edu/hipertextnet/numero-5/pln.html>

Xerox. (s.f.). *Xerox*. Recuperado el 18 de Marzo de 2012, de <http://www.xerox.com/news/news-archive/2011/mex-virtual-lab-030211/esmx.html>