



Facultad 2

Trabajo de diploma para optar por el Título de Ingeniero en Ciencias Informáticas

Título: Herramienta informática con minería de texto para la
caracterización psicosocial de estudiantes y grupos universitarios

Autor: Carlos Alberto Francés Benítez

Tutores: MSc. Reynel Fals de Pedro
Dr.C. Úrsula Puentes Puentes

La Habana, Cuba

24 de junio, 2014

DATOS DE CONTACTO

Tutores

MSc. Reynel Fals de Pedro

Graduado de ingeniero en Ciencias informáticas de la Universidad de Ciencias Informáticas en el año 2006, profesor asistente, máster en neurociencias. Se ha desempeñado como jefe de proyecto y jefe de asignatura. Ha impartido las asignaturas de máquinas computadoras I y II, teleinformática I y arquitectura de computadoras; y postgrado de introducción al ensamblador para GNU-Linux. Ha participado en el proyecto de desarrollo de software CSI. Actualmente es jefe de la asignatura inteligencia artificial II.

Correo electrónico: rfals@uci.cu

Dra.C. Ursula Puentes Puentes

Graduada en Pedagogía y psicología En el Instituto Pedagógico Superior Estatal V.I.Lenin Moscú en 1978. Especialista en Psicología en 1987, ISP. E.J. Varona y Master en Psicología Educativa en 1998 en la Universidad de La Habana. Es Profesor Titular. Se ha desempeñó durante 41 años a la formación de profesores en la Universidad de Ciencias Pedagógicas de Pinar del Río, (antiguamente. ISP) y hace dos años que trabaja en esta universidad. Siempre ha impartido docencia en áreas de la Psicología y la Pedagogía. Lo mismo en pre grado como en postgrado y maestrías. Ha sido miembro de varios proyectos entre los que se encuentran: La profesionalidad del profesor, la orientación familiar, el Modelo de Integración (Formación Producción- Investigación) en la UCI (MIFPI) y actualmente en CALINFOR. La calidad de la formación permanente del ingeniero sustentada en el vínculo universidad industria y uso de las TIC. Presta servicio en el Departamento de Ciencias Sociales de la facultad 2. Es profesora de Formación Pedagógica y trabaja en el Centro de Innovación y Calidad Educativa (CICE). Es Jefe de colectivo de la asignatura que imparte a nivel de facultad.

Correo electrónico: ursulapp@uci.cu

Agradecimientos

Agradezco a:

Mi novia por sus consejos, ayuda y apoyo incondicional para la realización de la investigación.

Mis padres por sus sabias palabras y contribuciones para la realización de la investigación.

Mis tutores, por la ayuda, colaboración y recomendaciones brindadas para la realización de esta investigación.

A la MSc. Annia Arencibia por su oportuna y cuidadosa revisión de la redacción y presentación de la tesis.

Al Dr.C. Febe A. Ciudad por sus recomendaciones para la realización de la investigación.

Mis compañeros de estudio por compartir sus experiencias en la búsqueda de soluciones para la realización de la investigación.

Los que de una manera u otra han contribuido a la realización de este trabajo.

A todos ¡muchas gracias!

Dedicatoria

A mi novia, que me hace sentir el hombre más feliz y afortunado con su amor, entrega, paciencia y apoyo incondicional.

A mis padres, por la confianza y el apoyo que siempre han depositado en mí y su perseverancia porque siempre triunfe.

A mi hermana, por haberme apoyado siempre bajo cualquier circunstancia.

A mi sobrina, por ser la fuente de alegría de la familia.

En la Facultad 2 de la Universidad de las Ciencias Informáticas se realiza el proceso de caracterización psicosocial en cada curso escolar. Este proceso es fundamental para obtener la información necesaria de cada estudiante, para la toma de decisiones en determinadas cuestiones, como la exploración de la personalidad y la caracterización integral del estudiante al llegar al quinto año de la carrera, para su ubicación laboral. En este proceso se realiza un análisis de los datos obtenidos a través de técnicas psicológicas que se les aplican a los estudiantes. Es realizado de forma manual por profesores guías y especialistas, pero en la mayoría de los casos son muy jóvenes y no tienen la experiencia ni los conocimientos necesarios para llevar a cabo esta tarea, lo que provoca que este análisis no se realice con el rigor que requiere.

El objetivo de este trabajo es desarrollar una solución informática para realizar el análisis de datos en el proceso de caracterización psicosocial en la Facultad 2. Son definidos los principales conceptos y requisitos identificados a partir de las necesidades del problema, son estudiados los lenguajes, herramientas y tecnologías necesarias para desarrollar la solución informática. Como resultado se obtiene la Herramienta informática con minería de texto para la caracterización psicosocial que permite automatizar el análisis de los datos de los estudiantes encuestados en la facultad con la técnica psicológica de los Diez Deseos para la obtención de información necesaria en la toma de decisiones referidas a mejorar la formación del estudiante universitario.

Palabras clave: análisis de datos, caracterización psicosocial de estudiantes, minería de texto, obtención de información, técnicas psicológicas.

Índice

Introducción	1
Capítulo 1. Fundamentación teórica	7
1.1 Introducción.....	7
1.2 Conceptos relacionados con el dominio del problema	7
1.2.1 Ámbito psicológico y social	7
1.2.2 Ámbito informático	12
1.3 Sistemas homólogos existentes	22
1.3.1 Sistemas existentes a nivel internacional	22
1.3.2 Sistemas existentes a nivel nacional	26
1.3.3 Sistemas existentes en la UCI	27
1.4 Análisis de los sistemas homólogos	30
1.5 Estudio de las técnicas, herramientas, metodologías y tecnologías empleadas en la investigación 30	
1.5.1 Metodología para el desarrollo.....	30
1.5.2 Lenguajes de programación.....	31
1.5.3 Lenguaje de modelado	36
1.5.4 Herramienta CASE	37
1.5.5 Sistema gestor de base de datos.....	39
1.6 Conclusiones del capítulo	41

Índice de contenido

Capítulo 2. Características de la Herramienta.....	42
2.1 Introducción.....	42
2.2 Propuesta de solución.....	42
2.2.1 Técnica seleccionada para la exploración de los principales motivos del estudiante.....	43
2.3 Descripción del análisis de los datos.....	44
2.4 Ontología.....	48
2.5 Modelo de Dominio.....	49
2.5.1 Definición y descripción de los conceptos del Modelo del Dominio:.....	50
2.6 Especificación de los requerimientos.....	51
2.6.1 Obtención de los requerimientos.....	52
2.6.2 Requisitos funcionales.....	52
2.6.3 Requisitos no funcionales.....	54
2.6.4 Definición de los casos de uso.....	56
2.7 Conclusiones del capítulo.....	58
Capítulo 3. Análisis y diseño del sistema.....	59
3.1 Introducción.....	59
3.2 Arquitectura de la Herramienta.....	59
3.2.1 Cliente-servidor en tres capas.....	60
3.3 Patrón arquitectónico.....	61

Índice de contenido

3.3.1	Modelo Vista Controlador	62
3.4	Patrón de diseño.....	63
3.4.1	Patrones Generales de Software para Asignar Responsabilidades (GRASP).....	63
3.5	Extracción de los datos del Sistema Lime Survey.....	64
3.6	Conclusiones del capítulo	65
Capítulo 4. Implementación y validación		66
4.1	Introducción	66
4.2	Modelo de datos	66
4.3	Diagrama de despliegue	68
4.4	Estrategias de codificación	69
4.4.1	Estándar de codificación.....	69
4.5	Validación de requisitos	71
4.5.1	Criterio de expertos	72
4.5.2	Proceso de validación.....	74
4.6	Conclusiones del capítulo	75
Beneficios		76
Conclusiones generales.....		77
Recomendaciones		78
Referencias bibliográficas		79

Índice de contenido

Anexo 1	84
Anexo 2	93
Anexo 3	95

***“La importancia de la tecnología no está en sí misma,
sino en su aplicación como herramienta
para hacer concreta una filosofía educativa”.***

Collis, B. (1997).

Introducción

Hoy en día la calidad en la formación de los estudiantes en las universidades es un objetivo primario a lograr. Para llegar de la manera más idónea a los estudiantes y captar toda su atención, es muy necesario realizar un estudio psicosocial de los mismos. Estudiar la personalidad de cada cual o de un grupo de individuos, permite conocer cómo los pensamientos, sentimientos y comportamientos de estos son influidos por la presencia real, imaginada o implícita de otras personas, ayuda a comprender cómo es el comportamiento en grupos. También abarca lo que son las actitudes de cada cual ante su forma de reaccionar o pensar en el medio social, lo cual es importante para realizar los planes y programas de clases, la forma en que se va a transmitir el contenido y las formas en las que se podrá evaluar para obtener los mejores resultados educativos y formativos en la institución. El proceso de caracterización permite dar una evaluación psicológica y social para conformar el aval del estudiante al culminar cada curso de la carrera y ser insertado en el ámbito laboral y profesional.

Según el área de resultado clave número 1 (ARC1), de los Objetivos de Trabajo para el año 2013 y hasta el 2016 del Ministerio de Educación Superior (MES), “Los proyectos educativos constituyen un importante instrumento institucional de dirección del trabajo educativo, en la base que consta de la integración de tres dimensiones: curricular, extensión universitaria y actividad sociopolítica. Su prioridad fundamental es ubicar al estudiante en el centro del trabajo educativo. De aquí la necesidad de que se refleje, con una determinada fundamentación científica, una caracterización del grupo de estudiantes. Esta caracterización debe plasmar las diferencias que existen en los grupos de un mismo año de una carrera, así como entre los diferentes años de las distintas carreras y de cada estudiante.” (Ministerio de Educación Superior, 2012; pág. 53)

Para la Universidad de las Ciencias Informáticas (UCI) este proceso es fundamental. Desde el año 2008 la institución utiliza el sistema de encuestas en línea Lime Survey, que posteriormente será explicado, para el proceso de caracterización, modificado por el Centro de Estudios de Innovación y Calidad Educativa (CICE), que ofrece a los profesores, profesores guías de las brigadas y profesores principales, una variada información sobre diferentes aspectos de la vida de los estudiantes, que son de utilidad para la elaboración de las Estrategias Educativas y Proyectos Educativos. También pueden servir para la ubicación en los proyectos productivos. En la actualidad la UCI cuenta además, con otros sistemas que apoyan este proceso de caracterización en algunas de sus Facultades. En la Facultad 2 se realizó el **Sistema automatizado**

para el proceso de caracterización de los estudiantes, con el objetivo de permitir la gestión de la información para el proceso de caracterización de los estudiantes (Basso Mesa, y otros, 2009); en la antigua Facultad 7 se desarrolló la **Herramienta de apoyo al proceso de caracterización psicológica para la entrada a proyectos informáticos**, aplicación capaz de realizar una gestión dinámica de los test psicológicos, que da como resultado una caracterización psicológica para los estudiantes (Valdivia Mena, y otros, 2010); también se cuenta con el sistema **Registro de información para la caracterización de los estudiantes de la Universidad de las Ciencias Informáticas**, con vista a ser integrado al Sistema de Gestión Universitaria, para un mejor desarrollo del proceso de registro de la información, para la caracterización de los estudiantes de la UCI (Reyes Cueto, y otros, 2012); en la Facultad 1 se desarrolló la **Solución informática para la gestión del proceso de caracterización estudiantil y generación de reportes del Sistema de Caracterización Integral**, cuyo fin es caracterizar a los estudiantes de la Universidad de las Ciencias Informáticas durante su carrera, facilitando la toma de decisiones para su formación profesional y proporcionando una evaluación integral del estudiante, para las entidades empleadoras a través del módulo Caracterización desarrollado (Martínez Rivera, y otros, 2012).

Existen varios instrumentos válidos para inferir rasgos psicológicos, como: los diez deseos, completamiento de frases, composición, inventario de problemas, adivina quién es, test socio métrico y expectativas y propósitos. La Facultad 2 de la Universidad emplea algunos de estos instrumentos en su proceso de caracterización de estudiantes, el análisis de los datos que se obtienen de estas encuestas y la aplicación de las mismas son realizados de forma manual, trayendo como consecuencia la pérdida de los documentos que contienen estos datos de los estudiantes, como el nombre, grupo y los deseos, en el caso de la técnica psicológica de los Diez Deseos, durante la recolección de los mismos por parte de los profesores, quedando el estudiante sin evaluación o teniendo que volver a realizar la encuesta; también como parte de esta consecuencia se deriva el lento proceso de análisis de los datos obtenidos para emitir una evaluación psicosocial, debido al gran volumen de documentos y datos que contienen, pues por cada estudiante de un grupo, existe un documento en formato duro con los datos encuestados o varias tablas de Excel, que el profesor debe analizar uno por uno manualmente; otra consecuencia es el posible deterioro de estos documentos, debido a los lugares donde son almacenados y al tiempo que permanecen guardados bajo condiciones que no les permite perdurar en ocasiones, como la humedad; el poco orden que se establece con estos documentos en ocasiones, mezclándose los pertenecientes a un grupo de estudiantes con los de otro; y que además este análisis es llevado a cabo por los profesores guías, que en su mayoría son muy

jóvenes, carecen de experiencia y formación psicológica y pedagógica, para poder realizar este análisis de los datos obtenidos con el rigor y la seriedad requeridos.

Teniendo en cuenta la problemática anteriormente planteada se formula la siguiente interrogante como **problema a resolver**: ¿Cómo mejorar el análisis de datos en el proceso de caracterización psicosocial de los estudiantes y grupos universitarios en la Facultad 2 de la Universidad de las Ciencias Informáticas (UCI)?

Es definido como **objeto de estudio**: el proceso de caracterización psicosocial. El **campo de acción**: herramientas informáticas aplicadas al análisis de datos en el proceso de caracterización psicosocial de los estudiantes y grupos universitarios en la Facultad 2 de la Universidad de las Ciencias Informáticas usando la minería de texto.

El presente trabajo tiene como **objetivo general**: desarrollar una herramienta informática usando la Minería de Texto, que permita realizar el análisis de datos en el proceso de caracterización psicosocial, de los estudiantes y grupos universitarios en la Facultad 2 de la Universidad de las Ciencias Informáticas.

Se trazaron las siguientes **tareas de investigación** para dar cumplimiento al objetivo anterior:

1. Determinación de los fundamentos teóricos acerca del uso de la minería de texto, para el análisis de datos en el proceso de caracterización psicosocial de la personalidad, a nivel internacional y nacional.
2. Definición de la arquitectura a emplear en el desarrollo de la herramienta informática, para estructurar la forma en que realizará el análisis de los datos, con el fin de obtener la información que se necesita para el proceso de caracterización.
3. Elaboración de la ontología, que contendrá las categorías de evaluación psicosocial, en las que se clasificarán cada uno de los conceptos determinados luego del preprocesamiento de los datos para generar el vector de frecuencias relativas por cada uno de los estudiantes y grupos universitarios.
4. Validación de la herramienta informática a través de un prototipo funcional y del criterio de expertos, para medir la exactitud de la información emitida por la misma luego del análisis de los datos.

En la investigación se adoptaron los **métodos científicos** siguientes:

Métodos del nivel teórico

Permitieron abordar el problema científico, la interpretación de los datos empíricos y para la construcción y desarrollo de la teoría científica:

Histórico-lógico: para la fundamentación de los antecedentes acerca de sistemas que abordan asuntos semejantes al que plantea la situación problemática.

Análisis y síntesis: para el análisis detallado de bibliografías asociadas al objeto de estudio y la síntesis de las mismas en los aspectos a tener en cuenta en la propuesta.

Inductivo-deductivo: para arribar a conclusiones sobre el proceso de caracterización psicosocial de los estudiantes, así como grupos universitarios; precisar las técnicas y herramientas a emplear para elaborar la solución haciendo uso de la minería de texto.

Métodos del nivel empírico

Permitieron recoger los datos necesarios para verificar la situación problemática planteada:

Análisis documental: permitió revisar toda la bibliografía relacionada con el objeto de estudio con el fin de comprender su esencia y particularidades; así como para el estudio de los referentes teóricos que permiten desarrollar la solución.

Entrevistas: realizadas a especialistas y directivos de la facultad, calificados y que llevan un prolongado tiempo trabajando sobre el tema asociado al objeto de estudio, para la búsqueda de información relacionada con los métodos empleados en la facultad en el proceso de caracterización de los estudiantes y grupos universitarios, así como la forma en que fueron realizados los mismos. Estas entrevistas se pueden encontrar en el [Anexo 2](#).

Métodos matemático- estadísticos

Análisis de clasificación múltiple: permitió clasificar a los estudiantes en estudio formando grupos o conglomerados (clúster) de elementos, de modo tal que los estudiantes dentro de cada conglomerado presenten cierto grado de homogeneidad en base a los valores adoptados sobre un conjunto de variables. (de la Puente Viedma, 1995)

Criterio de expertos: permitió la utilización sistemática del juicio intuitivo de un grupo de expertos para obtener un consenso de opiniones informadas sobre la posible aplicabilidad de una propuesta determinada, las cuales podrán estar fundamentadas, tanto en un análisis estrictamente lógico como en su experiencia intuitiva.

Método Delphi: permitió garantizar la selección de los especialistas para el criterio de expertos, a través de los métodos que este propone, con el fin de obtener criterios confiables para la validación de la herramienta informática.

Población y muestra:

Se define como población a los estudiantes de la Facultad 2 del segundo año de la carrera, presentándose como muestra la brigada 2201 de la propia facultad. Es una muestra seleccionada de manera intencional porque los rasgos se están midiendo de manera individual y global, si se incluyen estudiantes de otros grupos la medición individual servirá pero la global carecerá de sentido, puesto que no puede referirse a un grupo en específico.

Unidad de análisis: la esfera motivacional, presente en los datos colectados de los estudiantes, la cual será identificada haciendo uso de la minería de texto.

Aportes prácticos esperados del trabajo:

Con este trabajo se obtendrá como resultado una herramienta capaz de coleccionar datos de un estudiante o de un grupo de estudiantes universitarios, analizarlos y emitir un esbozo sobre la esfera motivacional aplicando minería de texto; la cual reduce los problemas de agilidad en el análisis de los datos obtenidos, pérdida de los documentos con estos datos durante su recolección y el posible deterioro de los mismos por su estado en formato sólido; aporta ventajas en el proceso docente educativo, permite caracterizar al estudiante y a un grupo universitario en cada curso de la carrera sirviendo de referencia para ubicar al estudiante en la práctica profesional al momento de insertarlo en el ámbito laboral. También favorecerá el trabajo de los Profesores Guía y del Colectivo de Año, aún más, en esta Universidad cuyos profesores en su mayoría son muy jóvenes y con pobre formación psicológica y pedagógica.

El presente trabajo está compuesto por 4 capítulos, estructurados de la siguiente manera:

Capítulo 1. Fundamentación teórica: se presenta el marco conceptual del trabajo y los principales conceptos relacionados con el objeto de estudio y el uso dado por diferentes autores. Se realiza un análisis de otras soluciones existentes que abordan el mismo tema a nivel internacional, nacional y de la Universidad, así como de las técnicas, metodologías y tecnologías utilizadas en estas soluciones y las que fueron adoptadas para el desarrollo de la presente solución.

Capítulo 2. Características del sistema: se evidencian los detalles del tipo de técnica relacionada con el objeto de estudio que se emplea, se establece el modelo de dominio a través del cual se determinan los conceptos y definiciones fundamentales tratados en el problema, se especifican los requisitos de la herramienta, funcionales y no funcionales, y se hace la definición de los casos de uso.

Capítulo 3. Análisis y diseño del sistema: se hace el análisis de la arquitectura y de las técnicas de desarrollo definidas, se seleccionan los patrones de diseños a emplear para este desarrollo y se describe la forma en que se relacionan las clases para la extracción de información de los datos obtenidos del Sistema Lime Survey.

Capítulo 4. Implementación y validación: se hace una descripción del modelo de datos realizado en la herramienta desarrollada, se define y describe el diagrama despliegue, se explican las estrategias de codificación, estándares y estilos a utilizar. Se describe la forma en que será validada la herramienta para comprobar su correcto funcionamiento.

Capítulo 1. Fundamentación teórica

Capítulo 1. Fundamentación teórica

1.1 Introducción

En el presente capítulo se abordan los elementos teóricos que sirven de base para la fundamentación del marco teórico de la investigación. Se realiza un estudio de los principales conceptos asociados al uso de la minería de texto y sus técnicas para la extracción de información a partir del análisis de datos, a la caracterización psicosocial, las técnicas y métodos de la psicología para realizar la misma. También se analizan los sistemas existentes en que se apoya el proceso de la caracterización de estudiantes y grupos universitarios, así como las técnicas, metodologías, tecnologías y herramientas, en cuanto a las características, ventajas y desventajas que poseen para el desarrollo de la investigación en el ámbito internacional, nacional y de la Universidad.

1.2 Conceptos relacionados con el dominio del problema

1.2.1 Ámbito psicológico y social

Caracterización

La caracterización, según (Hurtado, y otros, 2009) es un tipo de descripción cualitativa que puede recurrir a datos o a lo cuantitativo con el fin de profundizar el conocimiento sobre algo. Para cualificar ese algo previamente se deben identificar y organizar los datos; y a partir de ellos, describir (caracterizar) de una forma estructurada; y posteriormente, establecer su significado (sistematizar de forma crítica). Desde una perspectiva investigativa (Sánchez Upegui, 2010) establece que la caracterización es una fase descriptiva con fines de identificación, de los componentes, acontecimientos (cronología e hitos), actores, procesos y contexto de una experiencia, un hecho o un proceso.

La actividad de caracterizar parte de un trabajo de indagación documental del pasado y del presente de un fenómeno en cuestión, y donde no deben prevalecer las interpretaciones, pues el objetivo de la actividad es esencialmente descriptivo.

Según publicación realizada por (Fernández Fernández, y otros, 2006) en la Universidad de Camagüey, desde un punto de vista psicopedagógico la caracterización constituye un proceso donde se definen las

Capítulo 1. Fundamentación teórica

particularidades individuales de los alumnos; la valoración de su entorno escolar, familiar y comunitario y de las formas de relación que se establecen entre el estudiante y su medio. Para la determinación de esas particularidades el docente en la actualidad parte del diagnóstico integral, que incluye el conocimiento previo del entorno familiar y comunitario, aspectos biológicos, psicológicos y pedagógicos, así como el conocimiento que tiene el estudiante de cada materia que cursa. Los rasgos distintivos de la misma son:

- Personalizada: se valora a un sujeto en su individualidad.
- Dinámica: el sujeto y su entorno son susceptibles a constantes cambios y transformaciones.
- Colaborativa: no es del maestro en particular sino del colectivo docente e involucra además a la familia y todos los factores que interviene en el desarrollo integral del escolar.
- Preventiva: no solo se diagnostican aquellas insuficiencias del escolar, sino los factores de riesgo que pueden provocar su aparición.
- Desarrolladora: permite conocer en qué nivel se encuentra el alumno y con qué recursos cuenta para seguir desarrollándose.
- Universal: no se realiza solo a los alumnos con dificultades sino a todo el universo.
- Explicativa: no solo se describe el problema sino que se va a la causa que lo origina.
- Objetiva: refleja la realidad de manera científica, evalúa al estudiante en cuanto a su aceptación y respeto como ser humano que se es, la toma de decisiones, autorregulación, seguridad en cuanto a sus potencialidades y valores.

Por lo que para caracterizar a un alumno se necesita suficiente información sobre su entorno escolar, familiar y comunitario. Esta es obtenida por el docente a partir de los elementos utilizados en el diagnóstico integral:

- Expediente acumulativo del escolar.
- Métodos científicos aplicados como: la observación, la entrevista, pruebas pedagógicas y psicológicas, entre otras.
- Historia clínica.

En el proceso docente educativo de la UCI esta actividad es de vital importancia pues permite conocer el recorrido del estudiante durante la carrera universitaria y su desenvolvimiento en la misma, lo cual actúa como un currículum estudiantil el cual será tomado en cuenta para su ubicación laboral al culminar sus estudios en la Universidad.

Capítulo 1. Fundamentación teórica

Personalidad

En Cuba, diferentes investigadores (F. González, 1989; L. Domínguez, 1992; L. Pérez Martín y otros, 1991; L. Pérez, 2003; etc.) a partir de las investigaciones realizadas durante varios años, han planteado una concepción sistémica de la personalidad, en la cual con un enfoque materialista dialéctico del estudio de su estructura como una integración de contenidos y funciones que se configuran en cada sujeto y desempeñan un papel esencial en la regulación del comportamiento del hombre. (Colectivo de Autores, 2004)

“La personalidad es una configuración que expresa una organización estable y sistemática de los contenidos y funciones psicológicas que caracterizan la expresión integral del sujeto en sus funciones reguladoras y autorreguladoras del comportamiento, constituye por lo tanto, una integración sistemática de los contenidos y funciones de la psiquis, que se expresa de manera individualizada en el comportamiento de cada sujeto.” (Colectivo de Autores, La personalidad: su diagnóstico y su desarrollo, 2004; pág. 47)

Es la personalidad la que regula el comportamiento del individuo manifestando su función reguladora y autorreguladora y constituye el nivel superior regulador de la actividad del sujeto lo cual se evidencia en el carácter activo y consciente de esa regulación. (Colectivo de Autores, 2004)

Los autores (Collazo Delgado, y otros, 1992) plantean que la exploración de la personalidad para su caracterización ha sido estudiada por los siguientes principios, que desde el enfoque histórico cultural responden a la concepción filosófica del materialismo dialéctico:

- **Carácter socio histórico de la Personalidad:** El determinismo socio histórico es expresión dialéctica de la relación individuo – sociedad. La concepción asumida por el enfoque histórico cultural define la esencia humana como social, como resultado de interrelaciones que orgánicamente caracterizan la vida del ser humano en sociedad. Este determinismo no es mecánico e inmediato; se trata de un proceso en donde el ser humano es un elemento inseparable del sistema de interrelaciones y determinantes sociales del momento que le tocó vivir en función de las acciones individuales y colectivas (aun cuando lo trasciendan) que lo involucran.
- **Concepción sistémica de la personalidad y su desarrollo:** Lo psíquico y la personalidad como su nivel más alto de desarrollo, deben comprenderse en la multiplicidad de las relaciones externas e internas en las cuales existe. La personalidad es un sistema que integra lo psíquico a un nivel superior de

Capítulo 1. Fundamentación teórica

regulación. En ella lo afectivo y lo cognoscitivo están representados en una síntesis reguladora que condiciona la actuación del sujeto.

- La actividad y la comunicación como principios metodológicos en la comprensión y exploración de la personalidad: La actividad y la comunicación como formas de relación humana con la realidad, constituyen vías y expresión del desarrollo de la personalidad.
- La participación de la conciencia en la actuación de la personalidad: La autoconciencia participa en los procesos y formaciones que integran la personalidad y se expresa en momentos activos de integración de lo psíquico mediante las elaboraciones y reflexiones conscientes del sujeto en relación con su actuación. Aunque lo consciente no agota de ningún modo lo psíquico en la personalidad, lo que distingue la acción transformadora del sujeto como personalidad es la participación de la autoconciencia.

La definición brindada por el (Colectivo de Autores, 2004) en su cita bibliográfica será la que se tomará como base teórica para el desarrollo de la herramienta, aunque fueron mostradas otras definiciones dadas por otros autores para tener un dominio más amplio sobre este concepto.

Motivación

Según (Colectivo de autores, 1995) el carácter activo de la personalidad se aprecia en el hecho de que ella se forma y se desarrolla en la actividad, y a su vez regula su actividad. Esta actividad del hombre es provocada por algo, es sostenida por algo con una cierta energía o intensidad en una determinada dirección. La idea general anterior es traducida bajo el amplio término de motivación. Los elementos de base de esta motivación humana, los constituyentes básicos sobre los cuales se conforma todo el proceso de desarrollo del proceso motivacional son encontrados en las necesidades y los motivos.

Necesidad

“Podemos entender la necesidad como un estado de carencia del individuo que lleva a su activación con vistas a su satisfacción, con dependencia de las condiciones de su existencia.” (Colectivo de autores, 1995, Psicología para Educadores; p.97). Estos establecen que se debe hacer la distinción de la necesidad como lo que dirige y regula la actividad concreta de un individuo en un medio objetivo, como resultado del encuentro de la necesidad con aquel objeto capaz de satisfacerla, es que la necesidad se convierte en algo capaz de orientar y regular la actividad . (Sic)

Capítulo 1. Fundamentación teórica

Según (Leontiev, 1979) el encuentro de la necesidad con el objeto es un hecho de objetivación de la necesidad, de completamiento de la misma con un contenido, que es sacado del mundo circundante y esto dirige a la misma a un nivel psicológico propiamente dicho.

Motivo

De esta manera “Podemos entonces definir el motivo como aquel objeto que responde a una u otra necesidad y que, reflejado bajo una forma u otra por el sujeto, conduce su actividad.” (Colectivo de autores, 1995, Psicología para Educadores; p. 100).

La necesidad no conoce su objeto hasta que es satisfecha por primera vez, tiene que descubrirlo. El resultado de este descubrimiento es quien le da a la necesidad su carácter objetal y este objeto percibido (representado, pensado), deviene en motivo (Colectivo de autores, 1995). (Sic)

Según (Colectivo de autores, 1995) los motivos se clasifican atendiendo a distintos criterios:

- Por su contenido: cognoscitivos, laborales, artísticos.
- Por su forma de manifestación: intereses, convicciones, aspiraciones, ideales, intenciones, autovaloraciones.
- Por su nivel de conciencia: consientes o inconscientes.
- Por su polaridad: positivos o negativos.
- Por su estabilidad: estables y situacionales.
- Por su generalidad o amplitud: generales o amplios, particulares o estrechos.
- Por su estructura jerárquica: rectores (dominantes), secundarios (subordinados).

Grupo

“Grupo es un conjunto de personas que interactúan entre si durante un tiempo determinado en la realización de una actividad común para alcanzar un objetivo grupal.” (González Soca, y otros, 2002. Nociones de sociología, psicología y pedagogía; pág. 116). De forma semejante, otros autores definen que “Un grupo es un conjunto de personas que interactúan directamente entre sí (cara a cara), durante un tiempo relativamente estable, para alcanzar determinadas metas, mediante la realización de una tarea.” (Bermúdez, y otros, 2002, La facilitación de la dinámica grupal en educación; pág.1).

Capítulo 1. Fundamentación teórica

Desde un punto de vista psicopedagógico “El grupo escolar no constituye una sumatoria de estudiantes, sino un órgano vivo, con identidad propia, que se conforma en las interacciones y la comunicación, generando normas, funciones, metas y objetivos comunes, códigos compartidos y una especial dinámica, que condiciona de forma notable los caminos que tomará el proceso en cada caso particular. Por tanto, el grupo ha de ser tomado en consideración en todos los momentos del proceso, cuando el profesor estructura, como mediador, los sistemas de actividad y comunicación que fluyen entre los estudiantes, apoyándose en el conocimiento de sus potencialidades para conducir a los/las aprendices al logro de los objetivos esperados.” (Castellanos, 2001, Hacia una concepción del aprendizaje desarrollador; pág.63).

En la investigación se asume la definición dada por (González Soca, y otros, 2002).

1.2.2 Ámbito informático

Inteligencia Artificial

El concepto de Inteligencia Artificial es definido por diversos autores de disimiles formas, no existe un concepto que defina de manera general esta disciplina, a continuación se muestran diversos términos definidos por algunos autores:

La Inteligencia Artificial (IA) estudia cómo lograr que las máquinas realicen tareas que, por el momento, son realizadas mejor por los seres humanos. (Rich, y otros, 1994)

Algunas definiciones de Inteligencia Artificial, organizadas en cuatro categorías (Russel, y otros, 2010):

- Pensando humanamente

El nuevo esfuerzo excitante para hacer que los ordenadores piensen... máquinas con mentes, en el sentido pleno y literal. (Haugeland, 1985).

La automatización de las actividades que asociamos con el pensamiento humano, actividades tales como la toma de decisiones, resolución de problemas, aprendizaje. (Bellman, 1978).

- Pensando racionalmente

El estudio de las facultades mentales a través del uso de modelos computacionales. (Charniak, y otros, 1985).

Capítulo 1. Fundamentación teórica

El estudio de los cálculos que hacen posible percibir, razonar y actuar. (Winston, 1992).

- Actuando humanamente

El arte de crear máquinas que realicen funciones que requieren de inteligencia cuando es realizada por la gente. (Kurzweil, 1990).

El estudio de cómo hacer que las computadoras hagan las cosas en las que, por el momento, la gente es mejor. (Rich, y otros, 1991).

- Actuando racionalmente

Inteligencia Computacional es el estudio del diseño de agentes inteligentes. (Poole, y otros, 1998).

IA... tiene que ver con el comportamiento inteligente en artefactos. (Nilsson, 1998)

Analizando las definiciones anteriores se toma como conclusión que la inteligencia artificial es una disciplina que estudia cómo hacer que los ordenadores realicen determinadas actividades y tareas involucradas con la toma de decisiones, la resolución de problemas y el aprendizaje tal y como los seres humanos somos capaces de hacer y en muchas ocasiones de una manera más rápida y eficiente. Es el estudio de cómo transpolar las facultades mentales de los humanos al ordenador.

Minería de texto

La Minería de Texto o Text Mining es una de las principales herramientas que se utilizan en el proceso de descubrimiento del conocimiento o KDD (del inglés Knowledge Discovery in Data Bases), ha sido configurada por las disciplinas de la Inteligencia Artificial y la Computación Lingüística, según (Nasukawa, y otros, 2001) esta se refiere al examen de una colección de documentos y al descubrimiento de información no contenida en ningún documento individual de la colección; se trata de obtener información sin haber partido de algo. Por su parte (Bressán, 2003) indica que la minería de texto consiste en la búsqueda de regularidades o patrones que se encuentran en un texto, a partir de técnicas de aprendizaje automático; por tanto, se considera como una de las muchas ramas de la lingüística computacional. Como proceso se encarga del descubrimiento del conocimiento que no existe en el texto, pero que surge al relacionar el contenido de varios textos y se divide en varias etapas.

Capítulo 1. Fundamentación teórica

Pero desde el punto de vista del dominio de aplicación (Olivenza, 2007) determina, que la minería de texto es una herramienta capaz de abarcar una amplia gama de dominios, desde aquellos de la recuperación y extracción de información, presentación, resumen de multidocumentos, minería de datos aplicada a textos, etcétera. Es un término que en dependencia del autor que lo emplee se restringen o se amplían sus esferas de aplicación.

(Bordón, y otros, 2004) determinan que las aplicaciones de la minería de textos se utilizan principalmente para:

- Extraer información relevante de un documento.
- Agregar y comparar información automáticamente.
- Clasificar y organizar documentos según su contenido.
- Organizar depósitos para búsqueda y recuperación.
- Clasificar textos e indizarlos en el Web.

Descripción de las principales técnicas utilizadas en la minería de textos por (Olivenza, 2007):

Tabla 1. Principales técnicas utilizadas en la minería de texto.

Técnica	Descripción		Observaciones
	Etapas	Descripción de la etapa	
Técnicas clásicas	Pre-procesamiento	Es el proceso mediante el cual los textos se transforman en algún tipo de representación estructurada que facilite su análisis.	Como se puede observar, todas las etapas están muy interrelacionadas, así entonces, la primera etapa condiciona el descubrimiento de los patrones que la minería de texto puede realizar.
	Representación	La representación depende de la técnica de pre-procesamiento utilizada y determinará el algoritmo de descubrimiento a utilizar.	
	Descubrimiento	Son algoritmos que, a partir de una representación estructurada de la información, son capaces de descubrir regularidades en los textos.	
	Un grafo conceptual es un grafo bipartito que tiene dos tipos de nodos, conceptos y relaciones conceptuales. Los grafos se comparan utilizando conocimiento del dominio como diccionarios de sinónimos y jerarquías de conceptos. Se realiza una operación de intersección entre dos grafos		Las técnicas de grafos conceptuales aportan mayor

Capítulo 1. Fundamentación teórica

Grafos conceptuales	para dar un resumen de ambos y a dicho resumen se le valora con una puntuación que indica el grado de similitud entre ambos textos. La agrupación de dos o más grafos permite descubrir la estructura oculta de la colección de textos. Para agrupar los grafos, se pueden utilizar técnicas de agrupamiento como las estrategias colaborativas y el agrupamiento en k medias ¹ .	semántica. Todas estas técnicas se basan en suministrar a los algoritmos, un conjunto de ejemplos a partir de los cuales se generan las agrupaciones.
Programación lógica inductiva	Esta técnica permite introducir conocimiento apriori del dominio en forma de definiciones mediante predicados relacionados. Requiere no sólo de un conjunto de entrenamiento con ejemplos sino también de las relaciones descubiertas por el diseñador y basadas en las cláusulas de Horn ² .	La ventaja que posee es su capacidad de representación basada en una lógica de segundo orden, que permite generalizar conceptos y descubrir definiciones de conceptos de forma automática. Normalmente se utiliza Prolog para programar las herramientas.
Programación genética	Es un método de generación automática de programas para computadoras con inspiración evolutiva; sobre la base de programas muy simples, mediante el cruce de unos con otros y procesos de mutación aleatoria, se generan programas más y más aptos para la realización de la tarea que se le asigna. La aptitud de los programas se mide de forma numérica mediante una función denominada de fitness.	Esta técnica ha tenido problemas con las definiciones recursivas, debido a que la programación

¹ K medias, es un método de agrupamiento heurístico con número de clases conocido (K). El algoritmo está basado en la minimización de la distancia interna (la suma de las distancias de los patrones asignados a un agrupamiento al centroide de dicho agrupamiento). Este algoritmo minimiza la suma de las distancias al cuadrado de cada patrón al centroide de su agrupamiento.

² Una cláusula de Horn, es una regla de inferencia lógica con una serie de premisas (cero, una o más), y un único consecuente. Las cláusulas de Horn son las instrucciones básicas del lenguaje de programación Prolog, de paradigma declarativo.

Capítulo 1. Fundamentación teórica

	Algunas extensiones de la programación genética permiten describir nuevas primitivas a partir de las primitivas inicialmente descritas. Básicamente, la idea consiste en introducir como primitivas las relaciones expresadas como cláusulas de Horn y utilizar un sistema que permita utilizar las denominadas ADFs ³ , que no son más que evoluciones paralelas de otras primitivas que se pueden utilizar en la definición principal, para generar definiciones muy resumidas de los conceptos. La función de fitness ⁴ deberá medir el número de ejemplos que se cubren con la definición de cada uno de los individuos generados en cada generación, que se pondera con la longitud de la definición para guiar al algoritmo a soluciones sencillas y el descubrimiento de conceptos intermedios.	genética tiene problemas de eficiencia con primitivas recursivas. Si se conoce que la naturaleza de la definición a encontrar es recursiva, probablemente sean mejor solución las técnicas anteriormente descritas.
--	--	---

En la investigación se emplearon las técnicas clásicas de la Minería de Texto, sobre la base del Sistema de arquitectura para un avanzado sistema de minería de texto con base de conocimientos que se explica a continuación.

Sistema de arquitectura para un sistema de minería de texto avanzado con una base de conocimientos previos

(Feldman, y otros, 2007) dividen este sistema en 4 áreas principales: Ver Figura 1.

- **Tareas de Preprocesamiento:** incluye todas aquellas rutinas, procesos y métodos requeridos para preparar los datos para las operaciones básicas de descubrimiento de conocimiento en un sistema de minería de texto. Estas tareas se centran típicamente sobre las actividades de preprocesamiento

³ ADF (definición automática de funciones) es una versión avanzada de la PG (Programación Genética) que intenta aprovechar las regularidades de los dominios a resolver aprendiendo simultáneamente subrutinas que codifiquen dichas regularidades. Dicha versión, permite reutilizar una subrutina varias veces dentro de un mismo individuo.

⁴ La función fitness (en inglés fitness function), es una función de ajuste que evalúa la capacidad de cada programa de llevar a cabo la tarea de optimizar una población de programas, utilizada por la técnica de aprendizaje automático, programación genética.

Capítulo 1. Fundamentación teórica

y categorización de las fuentes de datos. Generalmente convierten la información de cada fuente de datos en un formato canónico antes de aplicar varios tipos de métodos de extracción de rasgos contra estos documentos para crear una colección de documentos completamente representada por conceptos.

- **Operaciones básicas de minería:** son el corazón de un sistema de minería de texto e incluye el descubrimiento de patrones, análisis de las tendencias, y algoritmos de descubrimiento de conocimiento incremental. Entre los patrones comúnmente usados para el descubrimiento de información en datos textuales son distribuciones (y proporciones), frecuente y cerca de colecciones frecuentes de conceptos, y asociaciones. Estas pueden concernir ellas mismas con comparaciones entre - y la identificación de niveles de interés – algunos de estos patrones. Sistemas de minería de textos avanzados u orientados a dominios, o ambos, pueden también ampliar la calidad de sus diversas operaciones apalancando las fuentes de base de conocimientos. A estas operaciones básicas de minería en un sistema de minería de texto se les ha referido también, colectivamente, como procesos de destilación del conocimiento.
- **Componentes de la capa de presentación:** incluye GUI (Interfaz gráfica de usuario, por sus siglas en inglés) y funcionalidad de búsqueda de patrones, como también acceso a los lenguajes de consulta. Herramientas de visualización y editores de consultas para el cliente y optimizadores también entran dentro de esta categoría de arquitectura. Estos pueden incluir herramientas basadas en caracteres o gráficas para crear o modificar grupos de conceptos como también para crear acotados perfiles para conceptos o patrones específicos.
- **Técnicas de refinamiento:** dentro de lo más simple, incluyen métodos que filtran información redundante y agrupan cerradamente datos relacionados pero pueden crecer, en un sistema de minería de texto dado, para representar una completa, una suite global de supresión, ordenamiento, poda, generalización, y agrupamiento dirigido a la optimización del descubrimiento.

Capítulo 1. Fundamentación teórica

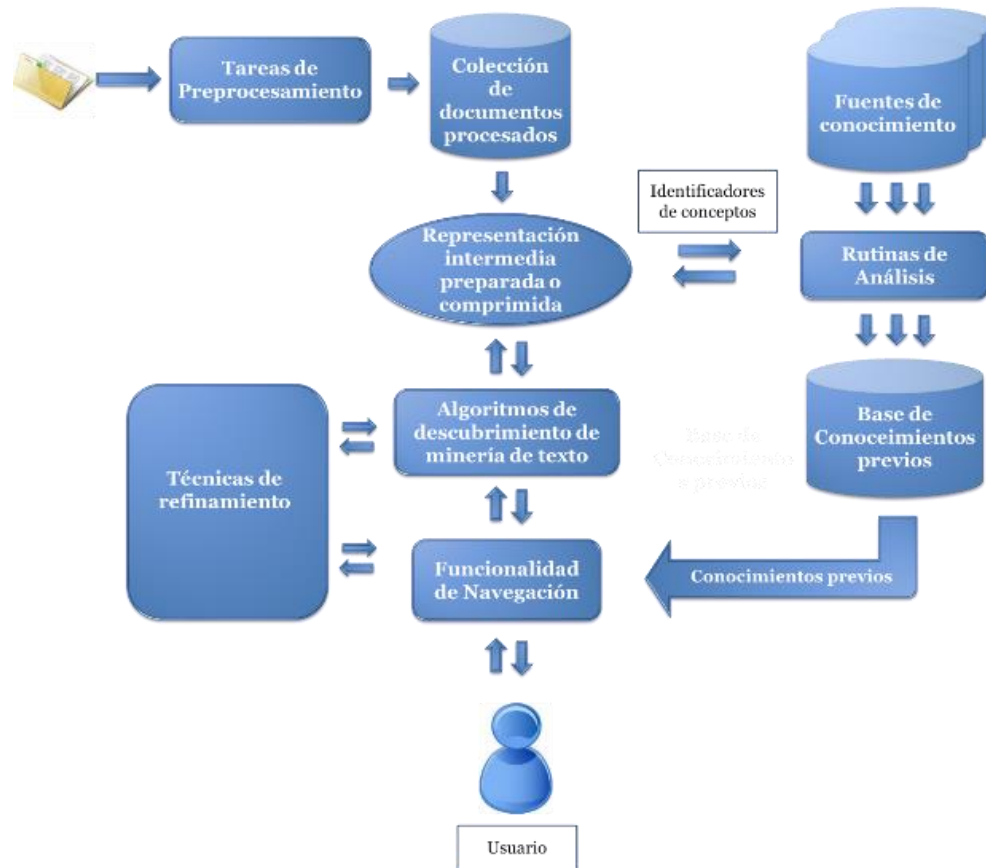


Figura 1. Sistema de arquitectura para un avanzado sistema de minería de texto con base de conocimientos. (Elaboración propia)

Procesamiento de lenguaje natural (PLN)

El PLN es una rama de las ciencias de la computación, relacionada con la inteligencia artificial y la lingüística. Confecciona sistemas computacionales para la comunicación eficiente entre personas y máquinas a través de lenguajes naturales. Permite esbozar mecanismos para que los programas ejecuten o simulen la comunicación. Involucra aspectos cognitivos, de memoria y de comprensión del lenguaje. Plantea (Sosa, 1997) que el PLN se concibe como el reconocimiento y utilización de la información expresada en lenguaje humano a través del uso de sistemas informáticos. Las aplicaciones del PLN se encuentran en la extracción de información a través de la minería de textos y la recuperación de información.

Capítulo 1. Fundamentación teórica

Extracción de información

Según (Alarcón, 2009) la extracción de información se encarga de desarrollar sistemas para la búsqueda y selección de datos específicos sobre eventos, entidades o relaciones a partir de un conjunto de documentos, esta será la labor que realizará la herramienta como resultado del análisis de los datos.

Recuperación de información

La recuperación de la información declara Alarcón en la misma bibliografía, que se basa en la elaboración de sistemas para la búsqueda y selección de documentos que cumplan ciertos criterios señalados por un usuario. Con el uso de la minería de texto se diseñan herramientas con el fin de extraer de manera automática aquellos documentos que puedan resultar interesantes para el usuario a partir de una consulta realizada por éste, ejemplo de ello es la labor que realizan los buscadores de Internet.

Etiquetado de partes de la oración

Las partes de la oración (POS, por sus siglas en inglés: Part Of Speech): clases de palabras, clases morfológicas, categorías gramaticales y etiquetas léxicas son muy importantes para el PLN por la información que proporcionan de una palabra y las palabras adyacentes. Las partes de la oración se utilizan para la recuperación de información y para extraer verbos o sustantivos de diferentes textos. La localización automática de las partes de la oración es útil en el análisis sintáctico, en algoritmos de desambiguación, en análisis sintácticos superficiales de palabras, para encontrar nombres, datos, fechas en textos y para otras entidades de la extracción de información. Finalmente, los corpus que han sido marcados con partes de la oración se usan para la investigación de la lengua.

Lematización

La lematización se puede entender como el proceso informático de reducir una palabra a su forma más simple, llamada lema. Según (Gómez, 2005) es un neologismo que se aplica al proceso de eliminación automática de partes no esenciales de los términos (sufijos, prefijos) para reducirlos a su parte esencial (lema) y facilitar la eficacia de la indización y la consiguiente recuperación. Ejemplo de ello es cuando queremos buscar una palabra en un diccionario, pero sabemos que no todas las palabras están contenidas en el, por ejemplo, si la palabra es un verbo, solo el infinitivo aparecerá representando a toda la familia

Capítulo 1. Fundamentación teórica

conjugada, este proceso de transformar una palabra al término por el que lo encontraríamos (lema) en ese diccionario se llama lematización.

Corpus

El corpus se define en su forma más simple como cualquier recopilación de muestras de lenguaje. Desde la filología (Torruella, y otros, 1999) entienden que un corpus es una recopilación de textos bien organizada. Pero para (Sierra, 2008) un corpus es un conjunto de datos reales y aceptables, debidamente ordenado, codificado y organizado, de diferentes textos recopilados, pertenecientes a un código lingüístico determinado, oral o escrito y el cual debe contar con las características de variedad y representatividad de la muestra de lenguaje que estén representando.

Análisis de Datos

Según (Rouse, 2012) el Análisis de Datos (Data Analysis, o DA) es la ciencia que examina datos en bruto con el propósito de sacar conclusiones sobre la información, este es usado en las ciencias para verificar o reprobando modelos o teorías existentes, se centra en la inferencia, el proceso de derivar una conclusión basándose solamente en lo que conoce el investigador. Por esta razón es empleado este proceso por la herramienta para simular el procedimiento y las operaciones que hace un especialista al trabajar con los datos en el área de la psicología para realizar una caracterización psicosocial.

Ontología

Según (Bosch, 2012) la ontología es un instrumento de organización y representación del conocimiento el cual permite hacer explícitas las reglas implícitas de una parte de la realidad. Su presentación formalizada permite que estas declaraciones explícitas sean independientes del sistema que las utiliza y que, a su vez, pueda reutilizarse por otros sistemas. Argumenta (González, 2006) que estos instrumentos como repertorios del conocimiento compartido, retoman aspectos teóricos que tienen sus bases en las raíces de la profesión de información:

- Son sistemas que establecen clases generales como sinónimo de representación más abstracta del dominio y subclases más específicas que pueden utilizarse en función del grado de profundidad de la descripción.

Capítulo 1. Fundamentación teórica

- Definen los términos básicos y en función de ellos, se esbozan las relaciones que se pueden establecer en ese dominio y las reglas para su combinación.
- Contemplan las restricciones (slots) para delimitar las cualidades de los conceptos y las facetas como expresión de los valores que puede adoptar un slot,
- Establecen las instancias como objetos de la clase.

Características de las ontologías según González:

- Los términos en los cuales se basan las ontologías se encuentran más cercanos al lenguaje natural, en tanto se defiende la utilización de adjetivos, adverbios, verbos, prefijos.
- El sustantivo no siempre tiene un lugar primordial.
- Permite expresar los conceptos como se conciben por los usuarios, en toda su complejidad de interconexiones concretas y en función de los distintos sistemas de los que forman parte.
- Evidencian las contradicciones reales que existen entre los conceptos de la realidad, y portan un desarrollo semántico más profundo para las relaciones de clase/subclase y las relaciones cruzadas.
- Posibilitan el trabajo con sistemas heterogéneos.
- Contemplan un mayor número de relaciones diversas entre los conceptos, que proceden del esquema conceptual con disponible en el dominio formalizado.

Por su parte (Foo, y otros, 2002) establecen que las diferencias fundamentales entre una ontología y un vocabulario de representación convencional están situadas en el nivel de abstracción, en las relaciones entre los conceptos, en la capacidad de ser comprensibles para las máquinas y, principalmente, en la expresividad que pueden proporcionar.

(González, 2006) plantea que las ontologías, fungen como herramientas tecnológicas que:

- Facilitan e intervienen en la selección y adquisición de la información.
- Enriquecen el procesamiento de la información, a partir de una descripción semántica más profunda.
- Propician un almacenamiento que sienta nuevos espacios para la búsqueda, recuperación y diseminación efectiva de la información.

Por estas razones, la ontología es empleada por la herramienta para clasificar la información obtenida teniendo en cuenta cada una de las diez categorías que establece el proceso de caracterización psicosocial de los estudiantes y grupos universitarios en la Facultad 2 de la UCI.

Capítulo 1. Fundamentación teórica

1.3 Sistemas homólogos existentes

1.3.1 Sistemas existentes a nivel internacional

Armstrong:

El estudio de los antecedentes permitió conocer que a nivel internacional existe el Sistema Integral de Evaluaciones ArmStrong. Solución integral desarrollada en México que favorece la autogestión de todo el proceso de reclutamiento, evaluación y selección de los mejores candidatos tanto para procesos de incorporación como desarrollo, capacitación y desincorporación, reduciendo tiempos de respuesta y mejorando la efectividad. Apoya cualquier actividad de evaluación brindando información del perfil personal de los evaluados así como habilidades técnicas o de gestión. Obtiene la información, la almacena, analiza, sintetiza y la presenta ejecutivamente en una radiografía laboral de cada empleado, existente o potencial. (Armstrong, 2014)

Este software hace más eficiente la operación del departamento de Recursos Humanos, reflejado en la minimización de costos y tiempos en cada una de las etapas inherentes al proceso de reclutamiento, selección y evaluación continua del personal, y por consecuencia en la maximización de las utilidades. Incluye la automatización de evaluaciones cognoscitivas y la aplicación de baterías psicométricas avaladas, así como el proceso completo de entrevistas digitalizadas (general, de presión, profunda y de salida), administración de competencias, generación de perfiles de puesto y evaluaciones de desempeño, además de contar con retroalimentación mediante un buzón de sugerencias electrónico y la creación del expediente electrónico del empleado. Cuenta con un módulo de Evaluaciones Psicométricas en el que se destacan cuatro tipos de evaluaciones que permiten obtener experiencias acerca del funcionamiento de estos sistemas:

- 16 Personality Factors ©: instrumento diseñado para medir las tendencias de personalidad de un individuo, en 16 factores primarios y 4 secundarios que muestran los principales rasgos de personalidad del evaluado. Es importante conocer el perfil de las personas para poder predecir sus actitudes, su conducta y sus reacciones ante diferentes estímulos.
- Estudio de Valores Allport ©: podemos definir los valores como algo de suma importancia para la persona, que representan fines buenos o deseados y que, por tanto, son perseguidos en forma

Capítulo 1. Fundamentación teórica

continua y constante, pues representan bienestar o le satisfacen un interés. Los valores pueden ser considerados incentivos que disparan la acción, o motivadores que mueven nuestras decisiones. El reporte ArmSTRONG® muestra claramente la escala de valores del evaluado.

- Escala de Intereses Vocacionales - Kuder ©: los intereses son impulsores que mueven a tomar decisiones en situaciones determinadas y son excelentes motivadores. Cuando las personas realizan actividades en las que están interesadas se muestran más atentas, con buena disposición y ánimo. Conocer los intereses que motivan a las personas es una herramienta muy importante que permite conocerlas mejor y orientarlas debidamente en el desempeño de una carrera, profesión u oficio. El resultado de esta evaluación permite conocer cómo están conformados los intereses del sujeto según diez áreas ocupacionales.
- Escala de Preferencias Personales - Kuder ©: muchas investigaciones han revelado la importancia del éxito profesional vinculado con las preferencias del individuo. La motivación y la satisfacción que el individuo encuentra al realizar determinadas actividades actúan como impulsores de un trabajo más productivo. El reporte ArmSTRONG® nos permite examinar en percentiles los resultados de cada uno de las puntuaciones para poder determinar diferencias del evaluado.

LimeSurvey:

LimeSurvey es un sistema de encuestas en línea, que pertenece a la comunidad de software libre pues ha sido liberado bajo licencia GPL⁵ v2 o posterior, significa que es gratis, por lo que cualquiera puede realizar modificaciones, así como copiar y redistribuir el trabajo o cualquier versión derivada.

Es un sistema de scripts en PHP⁶, que interactúa con MySQL⁷ con el fin de publicar, desarrollar y recoger los resultados de las encuestas. Permite la creación de encuestas que pueden ser publicadas en una página web. Centraliza y recoge los cuestionarios efectuados por las visitas, para analizar los resultados. Está

⁵ GPL, Licencia Pública General de GNU o más conocida por su nombre en inglés GNU General Public License (o simplemente sus siglas del inglés GNU GPL) es la licencia más ampliamente usada en el mundo del software y garantiza a los usuarios finales (personas, organizaciones, compañías) la libertad de usar, estudiar, compartir (copiar) y modificar el software.

⁶ PHP, es un lenguaje de programación de uso general de código del lado del servidor originalmente diseñado para el desarrollo web de contenido dinámico.

⁷ MySQL, es un sistema de gestión de bases de datos relacional, multihilo y multiusuario con más de seis millones de instalaciones.

Capítulo 1. Fundamentación teórica

compuesto por un panel completo de opciones que incluye el envío de correos electrónicos, gestión de la fecha de inicio y fin del cuestionario, la impresión del cuestionario en línea y el análisis de sus resultados en gráficos. Visualiza las imágenes de las encuestas en diferentes estilos gráficos. Su número de preguntas y cuestionarios no están restringidos. Cuenta con un sistema de plantillas que permite variar el punto de vista de los resultados de sus encuestas y permite el acceso a los análisis estadísticos básicos de los resultados de las encuestas. (LimeSurvey, 2014)

LimeSurvey tiene como ventajas:

- Ofrecer una interfaz amplia y amigable de administración, lo que proporciona que pueda ser utilizado fácilmente por otros. Permite realizar un ilimitado número de encuestas simultáneas y que varias personas accedan a una misma encuesta.
- Contiene funciones mejoradas de importación y exportación a texto como: CSV⁸, PDF⁹, SPSS¹⁰, XML¹¹ y formato MS Excel¹². Ofrece la posibilidad de integrar imágenes y videos en las encuestas.
- Brinda la posibilidad de establecer condiciones para las preguntas dependiendo de respuestas anteriores.
- Conjuntos de respuestas reutilizables y editables.
- Gestión de usuarios y cuotas.

⁸ CSV, (del inglés comma-separated values) son un tipo de documento en formato abierto sencillo para representar datos en forma de tabla, en las que las columnas se separan por comas (o punto y coma en donde la coma es el separador decimal: España, Francia, Italia...) y las filas por saltos de línea.

⁹ PDF, (sigla del inglés portable document format, formato de documento portátil) es un formato de almacenamiento de documentos digitales independiente de plataformas de software o hardware.

¹⁰ SPSS, es un programa estadístico informático muy usado en las ciencias sociales y las empresas de investigación de mercado. Originalmente SPSS fue creado como el acrónimo de Statistical Package for the Social Sciences aunque también se ha referido como "Statistical Product and Service Solutions". Sin embargo, en la actualidad la parte SPSS del nombre completo del software (IBM SPSS) no es acrónimo de nada.

¹¹ XML, siglas en inglés de eXtensible Markup Language ('lenguaje de marcas extensible'), es un lenguaje de marcas desarrollado por el World Wide Web Consortium (W3C) utilizado para almacenar datos en forma legible.

¹² MS Excel, Microsoft Excel es una aplicación distribuida por Microsoft Office para hojas de cálculo. Este programa es desarrollado y distribuido por Microsoft, y es utilizado normalmente en tareas financieras y contables.

Capítulo 1. Fundamentación teórica

- Contiene más de 20 tipos de preguntas.
- Encuestas anónimas y no anónimas.
- Encuestas multi-idioma.

LimeSurvey cuenta con varios filtros para la visualización de sus respuestas, entre los cuales se encuentran:

- Filtrado de respuestas por identificador: toma en cuenta las respuestas que cumplan con el identificador sean menores que, mayor que o igual que lo dispuesto en los cuadros de texto.
- Filtrado de respuestas por fecha: toma las respuestas que cumplan con la fecha de condición impuesta.
- Selección de preguntas a presentar y filtrado por respuestas: permite mediante cuadros de selección incluir preguntas en el reporte de resultados. Brinda la posibilidad de marcar una o varias de las opciones posibles de respuesta de una pregunta, con el objetivo de mostrar solamente los resultados que cumplan con la condición establecida, permitiendo así una evaluación cruzada de preguntas. En el caso de las preguntas de texto, muestra un cuadro donde introducir palabras relevantes para el filtrado de respuestas que las contengan.
- Resumen de todos los campos disponibles: incluye todas las preguntas en el reporte de resultados, hayan sido marcados o no los cuadros de inclusión.
- Filtrar respuestas incompletas: toma en cuenta solamente las respuestas de las encuestas que hayan sido terminadas y enviadas completamente, aunque posee la opción de incluir aquellas respuestas que algunas de las personas encuestadas guardaron pero cuya encuesta no fue completada.

Debido a su amplia gama de funcionalidades es empleado en la UCI para la realización de encuestas a través del sitio web, <https://encuesta.uci.cu>, también la Universidad brinda este tipo de servicios a diversas instituciones a nivel nacional logrando una alta demanda en su utilización y por las bondades que brinda; es empleado por la herramienta desarrollada en la investigación para la obtención de los datos que analizará posteriormente en función de la extracción de información.

Estos sistemas tienen como elementos comunes, que permiten realizar encuestas en línea para la obtención de datos de interés para realizar una caracterización del individuo. Pero ninguno de los dos puede ser

Capítulo 1. Fundamentación teórica

tomado como solución pues no permiten realizar un análisis de estos datos de manera automatizada usando minería de texto para emitir esta caracterización por lo que no cumplen con las necesidades de la investigación.

1.3.2 Sistemas existentes a nivel nacional

Paquete tecnológico para la evaluación psicológica en salud ocupacional:

“El diseño de este paquete tecnológico (PT) representa un proceso de integración a partir de instrumentos desarrollados en el Instituto Nacional de Salud de los Trabajadores (INSAT) con el objetivo de conformar procedimientos que permitan la ejecución de estudios a personas y a grupos poblacionales laborales en las condiciones de laboratorio y de terreno. El PT constituye una alternativa de solución, ya que permite la exploración de variables tanto de las condiciones, como de los riesgos y efectos que se presentan en el ambiente psicosocial laboral.

Tiene como objetivo facilitar el acceso a pruebas psicológicas de autores diversos en Salud ocupacional. El PT cuenta con veinticuatro herramientas construidas y/o validadas en el INSAT, clasificadas en: cuestionarios de síntomas; cuestionarios de factores de riesgo psicosocial y de riesgos psicosociales emergentes; cuestionarios sobre factores psicosociales protectores; autopercepción de la capacidad de trabajo; instrumentos para evaluación de funciones cognitivas; y procedimientos para análisis del trabajo, exigencias y efectos percibidos.

La aplicación del PT logra el acceso a datos de alto valor, ya que brinda: 1) la identificación de alteraciones clínicas y subclínicas potencialmente asociadas a condiciones y riesgos del ambiente laboral, 2) El establecimiento de niveles de frecuencia de alteraciones que fundamentan recomendaciones acerca del manejo de los factores de riesgo psicosociales, 3) la creación de líneas de base para el establecimiento de sistemas de vigilancia epidemiológica sobre riesgos laborales, y 4) una contribución a la caracterización de la naturaleza, evolución y particularidades de los efectos de los riesgos psicosociales.” (Grupo de autores, Paquete tecnológico para la evaluación psicológica en salud ocupacional, Revista Cubana de Salud y Trabajo, 2012; pág. 62-74).

Las herramientas incluidas en este paquete tecnológico se clasifican en:

Capítulo 1. Fundamentación teórica

- Cuestionarios de síntomas
- Cuestionarios de factores de riesgo psicosocial
- Cuestionarios de riesgos psicosociales emergentes
- Cuestionarios sobre factores psicosociales protectores
- Cuestionario de autopercepción de la capacidad de trabajo
- Instrumentos para evaluación de funciones cognitivas
- Procedimientos para análisis del trabajo, exigencias y efectos percibidos

AsDi:

Asistente para el Diagnóstico (AsDi), herramienta informática desarrollada por la Empresa Desoft atendiendo a los requerimientos del trabajo de la Dirección de Diagnóstico de FORDES¹³. Es una aplicación que permite el diseño de técnicas psicológicas, el diseño de normas para evaluar dichas técnicas y ofrece reportes individuales de evaluación. Es utilizada por vez primera en el Programa de Caracterización Psicológica de Recién Graduados de Nivel Superior del año 2013, donde se realizan evaluaciones socio psicológicas a grandes muestras de jóvenes egresados de las Universidades que se incorporan al Sistema del Ministerio de Comunicaciones. (Cabrera Ardanás, y otros, 2013)

El estudio de estos sistemas homólogos permite conocer las técnicas y herramientas empleadas para el desarrollo del mismo, lo cual ha servido de guía para el desarrollo del software, pero estos sistemas no se pueden adoptar como solución porque no cumplen con las expectativas de la problemática al estar enmarcada la situación en estudiantes universitarios con características propias de la Universidad y sistema docente educativo.

1.3.3 Sistemas existentes en la UCI

Sistema automatizado para el proceso de caracterización de los estudiantes:

¹³ Centro Coordinador para la Formación y el Desarrollo del Capital Humano.

Capítulo 1. Fundamentación teórica

Es un sistema capaz de permitir la gestión de la información en el proceso de caracterización de los estudiantes de la Facultad 2 de la UCI, con el fin de facilitar este proceso haciéndolo de manera automatizada. Emplea como metodología el Proceso Unificado de Desarrollo (RUP, por sus siglas sus inglés), es una metodología de desarrollo que proporciona un enfoque disciplinado para la asignación de tareas y responsabilidades dentro de una organización de desarrollo. La notación IDEF¹⁴ como técnica de modelado de sistemas para lograr una mayor colaboración del cliente.

Como tecnologías y herramientas para su desarrollo emplea las siguientes: herramientas CASE¹⁵ Visual Paradigm y Office Visio 2003, herramienta de desarrollo el Zend Studio for Eclipse, como servidor Web Apache, el marco de trabajo CodeIgniter, el PostgreSQL SQL como Sistema Gestor de Bases de Datos (SGBD), los Servicios Web que son un conjunto de aplicaciones o de tecnologías con capacidad para interoperar en la Web y como lenguaje de programación PHP5, se utiliza además HTML¹⁶, CSS¹⁷ y JavaScript principalmente en el diseño ya que posibilita la inclusión de ambos en su declaración de código.

Herramienta de apoyo al proceso de caracterización psicológica para la entrada a proyectos informáticos:

Es una aplicación desarrollada en la antigua Facultad 7 de la Universidad para la gestión dinámica de los test psicológicos que dan como resultado una caracterización psicológica de mayor calidad, facilitando la aplicación de los test, agilizando el proceso de obtención de resultados y garantizando la seguridad de la información contenida. Para el desarrollo de la herramienta se emplean: el lenguaje de programación PHP, el marco de trabajo CodeIgniter que utiliza la arquitectura Modelo-Vista-Controlador (MVC), el Sistema

¹⁴ IDEF (del inglés Integration DEFinition) es una familia de lenguajes de modelado en el campo de la Ingeniería de sistemas y la Ingeniería de software. Cubren una amplia gama de usos, desde el modelado funcional, simulación, análisis orientado a objetos hasta el diseño y adquisición de conocimientos.

¹⁵ Las herramientas CASE (Computer Aided Software Engineering, Ingeniería de Software Asistida por Computadora) son diversas aplicaciones informáticas destinadas a aumentar la productividad en el desarrollo de software reduciendo el costo de las mismas en términos de tiempo y de dinero.

¹⁶ HTML, siglas de HyperText Markup Language («lenguaje de marcas de hipertexto»), hace referencia al lenguaje de marcado para la elaboración de páginas web.

¹⁷ CSS, Hojas de Estilo en Cascada (Cascading Style Sheets) es el lenguaje de hojas de estilo utilizado para describir el aspecto y el formato de un documento escrito en un lenguaje de marcas.

Capítulo 1. Fundamentación teórica

Gestor de Bases de Datos MySQL y la metodología de Desarrollo el Proceso Unificado de Modelado (RUP). Además, se hace uso del Lenguaje Unificado de Modelado (UML) y Visual Paradigm para UML 2.0 para la creación de los artefactos que se generan durante el ciclo de vida del software.

Registro de información para la caracterización de los estudiantes de la Universidad de las Ciencias Informáticas:

Es un sistema informático que permite el registro de información para la caracterización de los estudiantes de la Universidad de las Ciencias Informáticas de manera organizada, realiza el registro de información desde que el estudiante ingrese a la Universidad hasta que concluye sus estudios superiores, el registro de información en cualquier momento del curso académico, consulta la trayectoria estudiantil facilitando a los profesores llevar el seguimiento de los estudiantes para la toma de decisiones y permite suministrar información al Módulo de Caracterización del sistema CICE con el objetivo de realizarles a los estudiantes el proceso de caracterización. La investigación fue desarrollada utilizando las tecnologías y herramientas siguientes: como lenguaje de programación Hypertext Pre-Processor 5 (PHP5), Gestión Universitaria Universo Digital Gestión Documental (GUUD) como marco de trabajo y PostgreSQL como gestor de base de datos, siguiendo el proceso de desarrollo ágil con segundo nivel de CMMI¹⁸.

Solución informática para la gestión del proceso de caracterización estudiantil y generación de reportes del Sistema de Caracterización Integral:

Es un módulo para la gestión del proceso de caracterización estudiantil y generación de reportes del Sistema de Caracterización Integral. Permite la caracterización integral del estudiante en cada curso de la carrera universitaria, para el estudio del comportamiento individual del estudiante, que permita la orientación hacia una formación eficaz, entregar la caracterización integral del graduado al empleador, la cual reúne la trayectoria del estudiante durante la carrera y consultar la información referente al proceso de caracterización estudiantil, para el seguimiento, evaluación y toma de decisiones por parte de los directivos docentes. Fue desarrollado utilizando los lenguajes, tecnologías y herramientas establecidas por el Centro

¹⁸ CMMI, Integración de modelos de madurez de capacidades o Capability maturity model integration (CMMI) es un modelo para la mejora y evaluación de procesos para el desarrollo, mantenimiento y operación de sistemas de software.

Capítulo 1. Fundamentación teórica

de Informatización Universitaria (CENIA) y guiado por el proceso de desarrollo con enfoque ágil basado en el nivel 2 de CMMI.

1.4 Análisis de los sistemas homólogos

El análisis de los sistemas mencionados anteriormente permite realizar un estudio de las herramientas, técnicas, metodologías y tecnologías empleadas en la UCI para la solución de estos problemas similares. Siendo una guía a tener en cuenta para el desarrollo de la aplicación web que se necesita desarrollar y permiten incorporar ciertos requisitos funcionales para el desarrollo de la herramienta, a pesar de que no se pueden utilizar como solución, porque no están relacionados estrechamente con el campo de acción de la investigación, no realizan extracción de la información empleando la minería de texto para el análisis de los datos obtenidos, accediendo luego a la caracterización del estudiante y del grupo universitario, son sistemas para la gestión de la información ya recogida previamente a través de técnicas para la caracterización y evaluación de los estudiantes en la Universidad, por lo que no se encuentran vinculados directamente con la investigación.

1.5 Estudio de las técnicas, herramientas, metodologías y tecnologías empleadas en la investigación

1.5.1 Metodología para el desarrollo

1.5.1.1 Programación Extrema (XP)

Para hacer más énfasis en la adaptabilidad que en la previsibilidad se escoge XP como metodología para el desarrollo de la herramienta. Según (Robles Martínez, y otros, 2002) la programación extrema es una metodología de desarrollo ligera basada en una serie de valores y una docena de prácticas de buenas maneras que propician un aumento en la productividad a la hora de generar software. Resumiendo (Beck, 1999) plantea que la programación extrema es una forma ligera, eficiente, flexible, predecible, científica y divertida de generar software. El desarrollo empleando esta metodología debe estar guiado por los valores de:

1. Comunicación,
2. Simplicidad,
3. Realimentación y

Capítulo 1. Fundamentación teórica

4. Coraje.

De la misma forma se guía por los siguientes principios que se apoyan en los valores anteriores:

1. Realimentación veloz,
2. Modificaciones incrementales,
3. Trabajo de calidad y
4. Asunción de simplicidad.

Estos suponen un puente entre las prácticas a tener en cuenta en el desarrollo, enunciadas a continuación, y los valores establecidos, ellas son:

1. El juego de la planificación (the planning game)
2. Pequeñas entregas (small releases)
3. Metáfora (metaphor)
4. Diseño simple (simple design)
5. Pruebas (testing)
6. Refactorización (refactoring)
7. Programación por parejas (pair programming)
8. Propiedad colectiva (collective ownership)
9. Integración continua (continuous integration)
10. 40 horas semanales (40-hour week)
11. Cliente en casa (on-site customer)
12. Estándares de codificación (coding standards)

1.5.2 Lenguajes de programación

Para el desarrollo de la Herramienta era necesario un lenguaje de programación que permitiera trabajar con texto, que hiciera uso de las expresiones regulares pues hace mucho más fácil el trabajo con el mismo, que fuera gratis y que fuera multiplataforma. Son muchos los lenguajes que cumplen con estos requisitos,

Capítulo 1. Fundamentación teórica

ejemplo de ellos son PHP, Ruby¹⁹, Python²⁰, Java²¹ y Perl²². Este último por sus características y ventajas fue el que mejor satisfizo las necesidades de la investigación. También se hacía necesario emplear un lenguaje de programación para el desarrollo de la interfaz gráfica con la que trabajaría el usuario (GUI), para ello de los tantos existentes que pueden ser usados se seleccionó PHP, puesto que el Sistema LimeSurvey está desarrollado empleando este lenguaje, de esta manera se mantiene la compatibilidad y se aprovechan las experiencias y prácticas del equipo de desarrollo de ese sistema.

1.5.2.1 Perl

Perl (del inglés Practical Extraction and Report Language) fue creado por Larry Wall en 1987. Es un lenguaje de programación de distribución gratuita. Su licencia es dual: Artistic License y GPL. Toma características de C, Bourne Shell (sh), AWK²³, Sed y Lisp. Es multi-paradigma: imperativo, funcional, orientado a objetos y reflexivo. Se pensó originalmente para la manipulación de texto. Sin embargo, en la actualidad, se utiliza en múltiples tareas, sobretodo en administración de sistemas y hasta hace poco menos de una década (antes de la existencia de PHP), en el desarrollo de aplicaciones web.

Aunque Perl es un lenguaje interpretado, es muy rápido comparado con otros lenguajes como Burne Shell. Plantea (Montes, 2013) que esto se debe a cómo se ejecuta un programa, que se podría dividir en:

1. Tiempo de compilación:

Donde se analiza (parsing) el texto del programa en un árbol sintáctico. Este árbol es optimizado simplificando las expresiones constantes, propagando el contexto y optimizando partes sueltas del código.

¹⁹ Ruby, es un lenguaje de programación interpretado, reflexivo y orientado a objetos, creado por el programador japonés Yukihiro "Matz" Matsumoto, quien comenzó a trabajar en Ruby en 1993, y lo presentó públicamente en 1995.

²⁰ Python, es un lenguaje de programación interpretado cuya filosofía hace hincapié en una sintaxis muy limpia y que favorezca un código legible.

²¹ Java, es un lenguaje de programación de propósito general, concurrente, orientado a objetos y basado en clases que fue diseñado específicamente para tener tan pocas dependencias de implementación como fuera posible.

²² Perl, es un lenguaje de programación diseñado por Larry Wall en 1987. Adoptado por su destreza en el procesado de texto y no tener ninguna de las limitaciones de los otros lenguajes de script.

²³ AWK, es un lenguaje de programación diseñado para procesar datos basados en texto, ya sean ficheros o flujos de datos.

Capítulo 1. Fundamentación teórica

2. Tiempo de ejecución:

Se ejecuta el programa siguiendo el árbol generado en la compilación.

Estas características le confieren las siguientes ventajas:

- Es un eficiente, completo y fácil de usar lenguaje de programación.
- Está disponible para gran cantidad de sistemas operativos, es un lenguaje accesible a cualquier usuario y además es gratis. Se puede descargar sin algún costo desde su sitio web oficial en www.perl.org para casi cualquier sistema operativo.
- Cuenta con un poderoso sistema de procesamiento de texto haciendo uso de una amplia gama de expresiones regulares para el trabajo con los textos.
- Su implementación es potente y fácil de invocar.

1.5.2.1.1 Módulos para Perl

Un módulo es un conjunto de funciones que proveen una cierta funcionalidad (Lizama, 2008), en este juicio el lenguaje Perl y su capacidad para el procesamiento. Por lo cual se emplearon diversos módulos del mismo para el desarrollo de la herramienta. Entre ellos:

- Perl::Lingua::TreeTagger

Lingua es un módulo que permite invocar al TreeTagger desde el código de Perl para el procesamiento del texto. Es empleado en la herramienta para el uso de este etiquetador POS (Xantos, 2012).

- Perl::File::Type

Se encarga de detectar el tipo de algún archivo de acuerdo con la información contenida, ya sea en la cabecera de este o en la estructura de los datos que contiene. Se emplea en la herramienta para detectar el tipo de archivo de entrada y tratarlo según sea el caso. (CPAN, 2013)

1.5.2.1.1.1 Procesador de texto para el lenguaje

Etiquetador POS

Es una herramienta que consiste en identificar la categoría gramatical (ya sea verbo, adjetivo, sustantivo, etc.) de cada palabra y asignarle una etiqueta que será utilizada posteriormente en los sistemas inteligentes de extracción de información. Para el etiquetado con partes de la oración se utilizan varios algoritmos, desde

Capítulo 1. Fundamentación teórica

las reglas elaboradas a mano, los métodos probabilísticos (etiquetado HMM y etiquetado de máxima entropía), hasta el etiquetado basado en la transformación y en la memoria para la etiquetación de partes de la oración (Jurafsky, 2007). De los muchos algoritmos que existen se ha aplicado en la investigación el TreeTagger.

- Algoritmo TreeTagger

TreeTagger es un etiquetador POS probabilístico, robusto y portable, trabaja de manera similar a un etiquetador de trigramas con algunas diferencias (Schmid, 1995). Para funcionar requiere de un lexicón de palabras completas, un lexicón en forma de árbol de sufijos y una categoría por defecto. El etiquetador estima la transición de probabilidades generando un árbol binario de decisión a partir del corpus de entrenamiento. El árbol de decisión automáticamente determina el tamaño apropiado del contexto que se utiliza para definir las probabilidades de transición. Ver Figura 2.

Del corpus de entrenamiento etiquetado se calculan las probabilidades de las categorías de acuerdo a los trigramas, para esto se normaliza procurando que a los trigramas raros o mal formados se les asigne un valor diferente de cero, pero muy bajo y se normalizan las probabilidades.

Un trigrama se determina siguiendo la senda marcada por el árbol hasta que se alcanza la última hoja. Cada hoja contiene un vector de probabilidades de categorías. Posteriormente, se recorta el árbol de trigramas siguiendo una medida de cantidad de información, para eliminar hojas y nodos con poco contenido de información. Ver Figura 3.

Capítulo 1. Fundamentación teórica

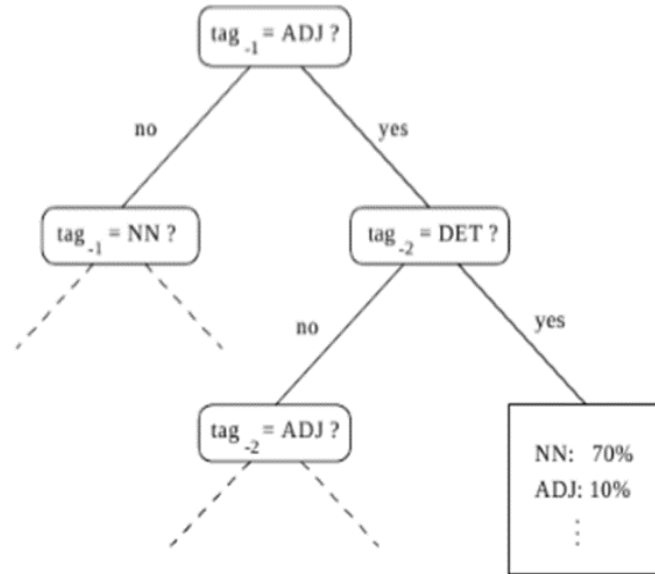


Figura 2. Árbol de decisión de TreeTagger (Schmid, 1994; pág.3)

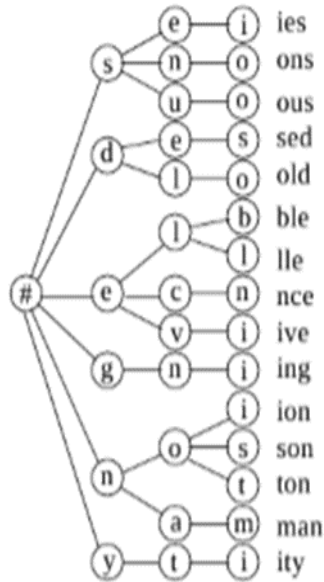


Figura 3. Árbol de sufijos de longitud tres (Schmidt, 1994; pág.5)

Capítulo 1. Fundamentación teórica

1.5.2.2 PHP

PHP (acrónimo de "PHP: Hypertext Preprocessor") es un lenguaje interpretado de alto nivel embebido en páginas HTML y ejecutado en el servidor (Sæther y otros, 2001). Este lenguaje permite procesar la información de formularios, generar páginas con contenidos dinámicos, o mandar y recibir cookies. Según (Palomo y Montero, 2007) este presenta características que lo han hecho alcanzar una gran popularidad en la generación dinámica de páginas web, ellas son:

- Es un lenguaje libre el cual puede descargarse desde su sitio web oficial: <http://www.php.net>.
- Está disponible para muchos sistemas (GNU/Linux, Windows, UNIX²⁴, etc.).
- Tiene una extensa documentación oficial en varios idiomas (disponibles libremente en <http://www.php.net>).
- Tiene una amplia gama de extensiones: para conectar bases de datos, para manejo de sockets, para generar documentos PDF, para generar dinámicamente páginas en Flash, etc.
- Al ejecutarse en el servidor, todo tipo de máquinas con todo tipo de sistemas operativos pueden usar las aplicaciones PHP.
- En caso de fallo del cliente por cualquier motivo se puede seguir utilizando el sistema desde cualquier otro que tenga un navegador web con conexión al servidor.

1.5.3 Lenguaje de modelado

El lenguaje de modelado define (González Cornejo, 2008), es la notación (principalmente gráfica) que usan los métodos para expresar un diseño. El proceso indica los pasos que se deben seguir para llegar a un diseño. En la investigación se hacía necesario el empleo de un lenguaje que fuese gráfico, a fin de especificar y documentar la herramienta en desarrollo, de un modo estándar incluyendo aspectos conceptuales como los procesos del modelo de dominio y las funciones de la herramienta. Fue UML el lenguaje unificado que cumplía con estos requisitos.

²⁴ Unix, (registrado oficialmente como UNIX®) es un sistema operativo portable, multitarea y multiusuario; desarrollado, en principio, en 1969, por un grupo de empleados de los laboratorios Bell de AT&T,

Capítulo 1. Fundamentación teórica

1.5.3.1 UML

Según (González Cornejo, 2008) el Lenguaje Unificado de Modelado (por sus siglas en inglés UML: Unified Modeling Language) emplea métodos de análisis y diseño orientados a objetos que aparecen a fines de los 80's y principios de los 90s. UML incrementa la capacidad de lo que se puede hacer con otros métodos de este enfoque.

Este lenguaje de modelado unido a una gestión de calidad, evita malos entendidos y entrega ciertas precauciones en la evolución y mantención de programas. Principalmente en lo que se refiere a los requerimientos asociados al levantamiento y diseño funcional de un sistema. Permite mostrarles a los clientes de que trata el dilema que abordan, quienes no piensan que el cambio que están solicitando es pequeño, cuando detrás de la petición existe una enorme cantidad de tareas relacionadas al requerimiento.

El lenguaje está dotado de múltiples herramientas para realizar las especificaciones del modelo de dominio presentado, pero solo trabajara de manera simplificada con las siguientes:

- Modelamiento de Clases.
- Casos de Uso.
- Diagrama de Interacción.

1.5.4 Herramienta CASE

Son diversas aplicaciones informáticas destinadas a aumentar la productividad en el desarrollo de software reduciendo el coste de las mismas en términos de tiempo y de dinero. Estas herramientas pueden ser de ayuda en todos los aspectos del ciclo de vida de desarrollo del software en tareas como el diseño de proyectos, cálculo de costes, implementación de parte del código automáticamente con el diseño dado, Compilación automática, documentación o detección de errores entre otras. (ITESCAM, 2014)

1.5.4.1 Visual Paradigm

Visual Paradigm para UML es una herramienta para desarrollo de aplicaciones utilizando modelado UML. Permite a Ingenieros de Software, Analistas de Sistemas y Arquitectos de sistemas la confiabilidad y

Capítulo 1. Fundamentación teórica

estabilidad en el desarrollo orientado a objetos que necesitan en la construcción de sistemas a gran escala (Targetware, 2007).

Propicia un conjunto de ayudas para el desarrollo de programas informáticos, desde la planificación, pasando por el análisis y el diseño, hasta la generación del código fuente de los programas y la documentación.

Entre sus características se encuentran:

- Disponibilidad en múltiples plataformas (Windows, Linux).
- Diseño centrado en casos de uso y enfocado al negocio que generan un software de mayor calidad.
- Uso de un lenguaje estándar común a todo el equipo de desarrollo que facilita la comunicación.
- Capacidades de ingeniería directa e inversa.
- Modelo y código que permanece sincronizado en todo el ciclo de desarrollo.
- Disponibilidad de múltiples versiones, con diferentes especificaciones.
- Licencia: gratuita y comercial.
- Soporta aplicaciones Web.
- Varios idiomas.
- Generación de código para Java y exportación como HTML.
- Fácil de instalar y actualizar.
- Compatibilidad entre ediciones.
- Soporte de UML versión 2.1.
- Diagramas de Procesos de Negocio - Proceso, Decisión, Actor de negocio, Documento.

Permite aumentar la calidad del desarrollo de la Herramienta, a través de la mejora de la productividad en el desarrollo y mantenimiento del software. Aumenta el conocimiento informático de una empresa ayudando así a la búsqueda de soluciones para los requisitos. También permite la reutilización del software, portabilidad y estandarización de la documentación, además del uso de las distintas metodologías propias de la Ingeniería de Software (Pressman, 2002; pág. 559-573).

Capítulo 1. Fundamentación teórica

1.5.5 Sistema gestor de base de datos

Según (Franco, 2009) un sistema gestor de base de datos se define como el conjunto de programas que administran y gestionan la información contenida en una base de datos. Ayuda a realizar las siguientes acciones:

1. Definición de los datos
2. Mantenimiento de la integridad de los datos dentro de la base de datos
3. Control de la seguridad y privacidad de los datos
4. Manipulación de los datos

Son varios los Sistemas gestores de base de datos existentes, entre ellos los más usados en la Universidad son PostgreSQL y MySQL. Este último es el empleado por la herramienta informática, debido a que los datos se obtienen de una base de datos del Sistema LimeSurvey que emplea este gestor, de esta manera se mantiene la integridad y compatibilidad de los mismos.

1.5.5.1 MySQL

MySQL es un Sistema de Administración de Bases de Datos Relacional (en inglés RDBMS). Se trata de un programa capaz de almacenar una enorme cantidad de datos de gran variedad y de distribuirlos para cubrir las necesidades de cualquier tipo de organización. Incluye todos los elementos necesarios para instalar el programa, preparar diferentes niveles de acceso de usuario, administrar el sistema y proteger y hacer volcados de datos. Utiliza el lenguaje de consulta estructurado (SQL), permite crear bases de datos, así como agregar, manipular y recuperar datos en función de criterios específicos.

MySQL dispone de muchas de las funciones que exigen los desarrolladores profesionales, como funciones SSL²⁵ e integración con la mayor parte de los entornos de programación. Así mismo, se desarrolla y actualiza de forma mucho más rápida que otros gestores.

²⁵ Secure Sockets Layer (SSL; en español «capa de conexión segura») y su sucesor Transport Layer Security (TLS; en español «seguridad de la capa de transporte») son protocolos criptográficos que proporcionan comunicaciones seguras por una red, comúnmente Internet.

Capítulo 1. Fundamentación teórica

Características más importantes de este gestor de base de datos (Oracle, 2011):

- Interioridades y portabilidad
 - Escrito en C y en C++
 - Probado con un amplio rango de compiladores diferentes
 - Funciona en diferentes plataformas.
 - Usa GNU Automake, Autoconf, y Libtool para portabilidad.
 - APIs disponibles para C, C++, Eiffel, Java, Perl, PHP, Python, Ruby, y Tcl.
 - Uso completo de multi-threaded mediante threads del kernel. Pueden usarse fácilmente multiple CPUs²⁶ si están disponibles.
 - Proporciona sistemas de almacenamientos transaccionales y no transaccionales.
 - Usa tablas en disco B-tree (MyISAM²⁷) muy rápidas con compresión de índice.
 - Relativamente sencillo de añadir otro sistema de almacenamiento. Esto es útil si desea añadir una interfaz SQL para una base de datos propia.
 - Un sistema de reserva de memoria muy rápido basado en threads.
 - Joins muy rápidos usando un multi-join de un paso optimizado.
 - Tablas hash en memoria, que son usadas como tablas temporales.

²⁶ La Unidad Central de Procesamiento (del inglés: Central Processing Unit, CPU) o procesador, es el componente principal del ordenador y otros dispositivos programables, que interpreta las instrucciones contenidas en los programas y procesa los datos.

²⁷ MyISAM es el mecanismo de almacenamiento de datos usado por defecto por el sistema administrador de bases de datos relacionales MySQL.

Capítulo 1. Fundamentación teórica

- Las funciones SQL están implementadas usando una librería altamente optimizada y deben ser tan rápidas como sea posible. Normalmente no hay reserva de memoria tras toda la inicialización para consultas.
- El código MySQL se prueba con Purify (un detector de memoria perdida comercial) así como con Valgrind, una herramienta GPL (<http://developer.kde.org/~sewardj/>).
- El servidor está disponible como un programa separado para usar en un entorno de red cliente/servidor. También está disponible como biblioteca y puede ser incrustado (linkado) en aplicaciones autónomas. Dichas aplicaciones pueden usarse por sí mismas o en entornos donde no hay red disponible.

Por las razones planteadas anteriormente fue seleccionado este gestor de base de datos para el almacenamiento de la información obtenida por la herramienta tras el análisis de los datos.

1.6 Conclusiones del capítulo

El análisis de los antecedentes realizado aporta una visión más concreta acerca del tema de la investigación, así como de las técnicas y pruebas que se utilizan para el proceso de caracterización a estudiantes y grupos de la antigua Facultad 7 en la Universidad. El análisis de los sistemas homólogos estudiados permite concluir que no pueden ser usados como solución pues no dan respuesta a las necesidades de la investigación, pero aportan experiencia y funcionalidad a la herramienta. El estudio de las herramientas, técnicas, metodologías y tecnologías empleadas en el presente trabajo permitió adquirir y ampliar los conocimientos y habilidades necesarias para desarrollar la herramienta.

Capítulo 2. Características de la Herramienta

Capítulo 2. Características de la Herramienta

2.1 Introducción

En este capítulo se realiza una descripción de los procesos del negocio a través de la definición del modelo de dominio en el análisis de datos en el proceso de caracterización de los estudiantes y grupos universitarios de la Facultad 2. Consecutivamente se hace una definición de los conceptos asociados al modelo de dominio y son definidos los requisitos funcionales y no funcionales que rigen el desarrollo de la herramienta.

2.2 Propuesta de solución

En la actualidad, el análisis de datos en el proceso de caracterización de los estudiantes y grupos universitarios de la Facultad 2 a partir de la aplicación de la técnica psicológica de los 10 Deseos, no es realizado de manera ágil y eficiente, se presentan los siguientes problemas:

- El análisis de los datos obtenido se debe hacer de manera manual con cada uno de los estudiantes de cada grupo de la Facultad 2.
- No todos los profesores tienen el nivel psicopedagógico para realizar el mismo con el rigor que lo requiere, pues en su mayoría son muy jóvenes y carecen de experiencia.
- Los datos obtenidos en la encuesta de los 10 Deseos están en formato sólido, papel, el cual tiene una elevada probabilidad de deterioro debido a las condiciones de almacenamiento, como la humedad.
- Estos documentos con los datos obtenidos para el análisis también tienen una elevada posibilidad de ser perdidos durante la manipulación de los mismos por parte de los profesores guías que aplican la encuesta, la receptan y la almacenan.

Partiendo de lo evidenciado anteriormente se propone, una Herramienta informática aplicada al análisis de datos en el proceso de caracterización de la Facultad 2 usando la minería de texto, como solución a la problemática expresada. Para este análisis se tomará como base para el desarrollo de la herramienta informática, las técnicas clásicas de la minería de texto, para extraer la información de los datos obtenidos de la encuesta de los Diez Deseos.

Capítulo 2. Características de la Herramienta

2.2.1 Técnica seleccionada para la exploración de los principales motivos del estudiante

2.2.1.1 Técnica de los Diez Deseos

Es con esta técnica psicológica que van a trabajar los estudiantes ya que permite examinar cuales son los motivos que rigen la personalidad de los mismos, lo cual posibilita a los profesores guías, tutores y especialistas de los años valorar cuales son los principales motivos que guían a los estudiantes en la Universidad. Ver Tabla 2.

Instrucciones para su aplicación: se solicita que escriba sus 10 principales deseos o aspiraciones que quiere lograr y sean significativos. Debe hacerlo por orden de importancia para él.

Orientaciones para su interpretación: debe organizar los datos en la siguiente tabla para facilitar el análisis y las conclusiones de la información obtenida.

Tabla 2. Formulario para la Técnica de los 10 Deseos

Contenido de los motivos	Polaridad afectiva			Inmediato	Mediato
	+	-	+-		
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

A partir de los datos organizados en la tabla se realizará el análisis y valoración de la información. En las conclusiones de esta técnica se debe tener en cuenta:

- Contenido del primer deseo (que debe ser el más importante). Su relación con los demás. La reiteración o no de este contenido en otros deseos
- Los contenidos referidos. Cuáles se repiten para valorar los principales motivos teniendo en cuenta el orden en que se expresan (jerarquía); la relación entre ellos; la riqueza o amplitud con la que se expresan estos contenidos. La presencia de algún contenido interesante, poco común, extraño (en función de la etapa del desarrollo del sujeto). Los últimos deseos pueden ser interesantes.
- Polaridad afectiva. Analizar el predominio de la polaridad afectiva en los deseos expresados. La polaridad positiva revela la tendencia hacia la búsqueda de lo agradable para el sujeto, cierto

Capítulo 2. Características de la Herramienta

optimismo; la polaridad negativa expresa la tendencia a la evitación, temor hacia el contenido expresado; la polaridad ambivalente refleja la presencia de conflicto con respecto al contenido expresado.

- Inmediatez o mediatez: se refiere a la dimensión temporal de los contenidos expresados donde el predominio de una u otra revela la proyección o no hacia el futuro. Puede haber equilibrio
- Se debe considerar también la observación del sujeto a cierta distancia durante la realización de la técnica, ya que la persona debe sentir privacidad para realizarla. Apremiar si muestra cooperación o no, si le resulta fácil, si se muestra tenso o tranquilo, seguro o inseguro.

En sí, constituye una técnica valiosa para la exploración de la esfera motivacional afectiva de la personalidad que valora la jerarquía motivacional en un momento determinado. (Collazo Delgado, y otros, 1992)

2.3 Descripción del análisis de los datos

La herramienta comienza el análisis obteniendo los datos que son respuestas de las encuestas realizadas a los estudiantes de un determinado grupo de la Facultad 2 de la UCI, una vez obtenidas pasan a ser el conjunto de documentos que van a ser preprocesados, ver Figura 4. En esta etapa es empleado el algoritmo de etiquetado TreeTagger para el procesamiento del lenguaje natural contenido en cada uno de los documentos, el cual obtiene la colección de documentos procesados, esta colección contiene los elementos categorizados, han sido identificadas y etiquetadas las palabras claves para la realización de las operaciones básicas de minería de texto, están representados en un árbol jerárquico que soportar mejor estas operaciones básicas.



Figura 4. Preprocesamiento de los documentos colectados y representación jerárquica. (Elaboración propia)

Capítulo 2. Características de la Herramienta

Luego de esto, las palabras claves etiquetadas son comparadas con la base de conocimientos (ontología), ver Figura 5, a través de rutinas de análisis para identificar la pertenencia de los conceptos principales a cada una de las categorías presentes en la ontología.

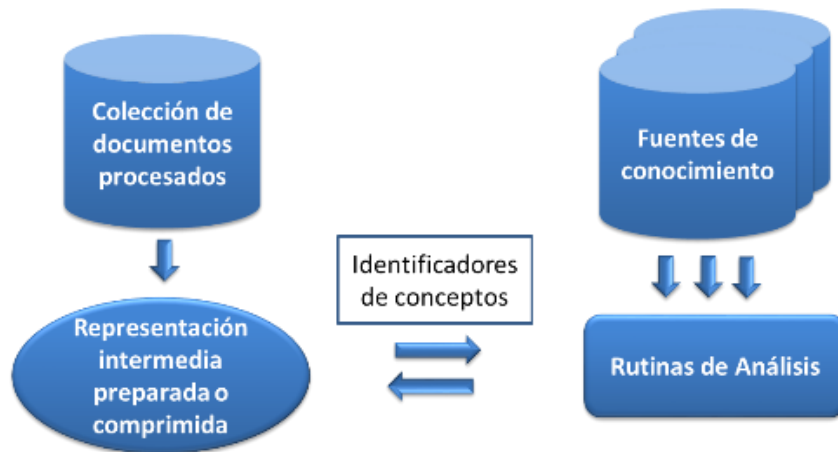


Figura 5. Identificación de conceptos con la base de conocimientos. (Elaboración propia)

Una vez que se obtienen los conceptos principales orientados al dominio de la caracterización psicosocial, se obtiene una base de conocimientos previos con aquellos con las categorías que prevalecen dentro de la ontología, se identifican los patrones y se analizan las tendencias a través de algoritmos de descubrimiento de la minería de texto, ver Figura 6. Para ello por cada documento analizado de cada estudiante se tiene un vector de frecuencias relativas, en el cual se representa el grado de pertenencia de cada concepto a cada una de las categorías presentes en la base de conocimientos.



Capítulo 2. Características de la Herramienta

Figura 6. Etapa de obtención del histograma de frecuencia relativa y de la base de conocimientos previos.
(Elaboración propia)

Esto se realiza contando las veces que cada concepto se asocia en cada categoría y se divide entre la cantidad de palabras claves identificadas luego del preprocesamiento de los datos, quedando como lo muestra la Tabla 3 y 4.

Tabla 3. Vector de pertenencias de los conceptos a cada categoría de un estudiante

Nombre	Adquisición BM	Reconocimiento	Estudio	Profesión	Familia	Salud	Amigos	Pareja	Sociales	Recreación
Pedro	1	3	0	1	2	3	0	1	0	1

Tabla 4. Vector de frecuencias relativas de un estudiante

Nombre	Adquisición BM	Reconocimiento	Estudio	Profesión	Familia	Salud	Amigos	Pareja	Sociales	Recreación
Pedro	0,083333	0,25	0	0,083333	0,166667	0,25	0	0,083333	0	0,083333

Luego de obtener el vector de frecuencias relativas de cada estudiante por cada categoría de la ontología, son tomados en consideración aquellas categorías que contienen algún valor y son dispuestas de manera jerárquica de modo que la categoría que mayor valor tenga sea la que más caracterice al estudiante, analizando los rasgos de manera individual. Al concluir el análisis de las respuestas de cada uno de los estudiantes de un grupo, y obtenido este vector general de frecuencias relativas que contenga a todos los estudiantes del grupo se puede hacer una clasificación global analizando los datos por cada columna, es decir, por cada categoría. Como se muestra en la Tabla 5, en la cual se obtiene las frecuencias relativas de cada categoría, lo que permite caracterizar al grupo de estudiantes.

Tabla 5. Matriz de frecuencias relativas de un grupo de estudiantes

Nombre	Adquisición BM	Reconocimiento	Estudio	Profesión	Familia	Salud	Amigos	Pareja	Sociales	Recreación
Pedro	0,083333	0,25	0	0,083333	0,166667	0,25	0	0,083333	0	0,083333
Fefa	0,333333	0	0,25	0,25	0,083333	0	0,083333	0,083333	0	0,083333
Osvaldo	0	0	0	0	0	0	0	0	0	0
Ana	0,166667	0,083333	0,166667	0,25	0,166667	0	0,083333	0	0	0,25

Capítulo 2. Características de la Herramienta

Nena	0	0	0	0	0	0	0	0	0	0
Mario	0,083333	0,083333	0,333333	0,25	0,166667	0	0	0,083333	0,083333	0,166667
...
Yanio	0,083333	0,083333	0	0,166667	0,416667	0,083333	0	0,083333	0	0
Total	2,5	1,416667	2,666667	3,75	3,75	1,583333	1,083333	1,25	0,75	2,583333
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
Totales	0,1	0,056667	0,106667	0,15	0,15	0,063333	0,043333	0,05	0,03	0,103333

De esta manera la herramienta informática emite, en un documento de análisis psicosocial, la información requerida para ser consumida por el usuario en el proceso de caracterización psicosocial. Además el usuario en la siguiente etapa, ver Figura 8, a través de la interfaz gráfica en donde tendrá acceso a la herramienta y sus funcionalidades, podrá realizar el refinamiento de la misma a través del método k-means para eliminar aquellos datos redundantes y agrupar los conceptos que se ubiquen en más de una categoría, y graficar la información obtenida empleando gráficos de estrella o radar para una mejor comprensión de las diferencias existentes entre cada estudiante o grupo caracterizado, como se muestra en el gráfico de radar de la figura:

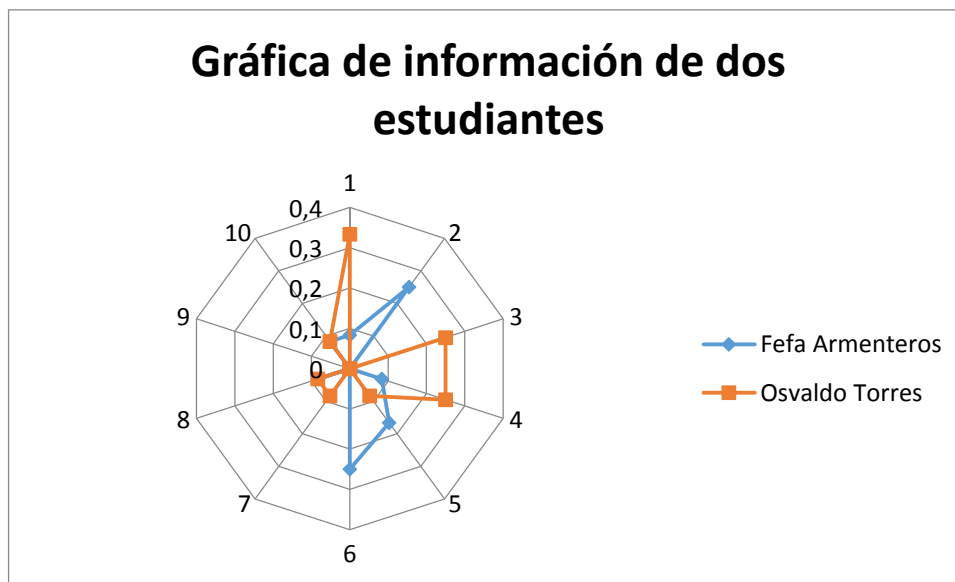


Figura 7. Gráfica de información obtenida de dos estudiantes (Elaboración propia)

Capítulo 2. Características de la Herramienta

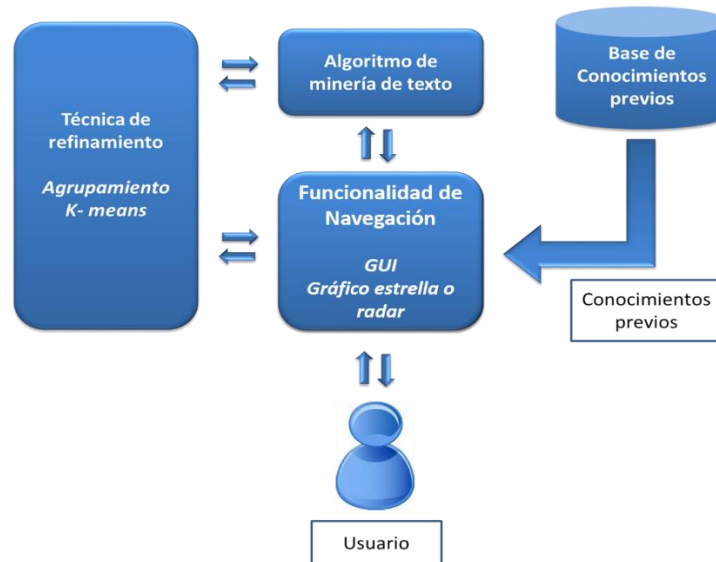


Figura 8. Presentación y refinamiento de la información obtenida. (Elaboración propia)

Este análisis que hace la herramienta accederá a la identificación y extracción de los conceptos principales de los datos contenidos en las encuestas de los 10 Deseos aplicadas a los estudiantes de la Facultad 2, permitirá agruparlos y categorizarlos por cada uno de las diez categorías presentes en la base de conocimientos que consiste en una ontología con términos de psicología. En sí, hará el trabajo que hacen los especialistas, tutores y profesores guías (trabajadores de la herramienta) de manera automática lo cual hará más ágil el proceso de análisis de los datos.

2.4 Ontología

La base de conocimiento que contiene las diez categorías mencionadas anteriormente, en las que se van a clasificar cada uno de los conceptos extraídos luego del análisis, fue elaborada a través de una ontología en el lenguaje de programación XML, lo cual permitió establecer una jerarquía entre los conceptos pertenecientes a cada categoría. Donde existe un conjunto de conceptos relacionados con las mismas, y por cada concepto puede existir un conjunto de otros conceptos que se asocian al término que este representa.

Capítulo 2. Características de la Herramienta

Luego de realizar la comparación de los conceptos identificados después del análisis con cada categoría que se encuentra en la ontología, puede corresponderle más de un concepto a cada una de ellas y un concepto puede estar asociado a más de una categoría, como se muestra en la figura:

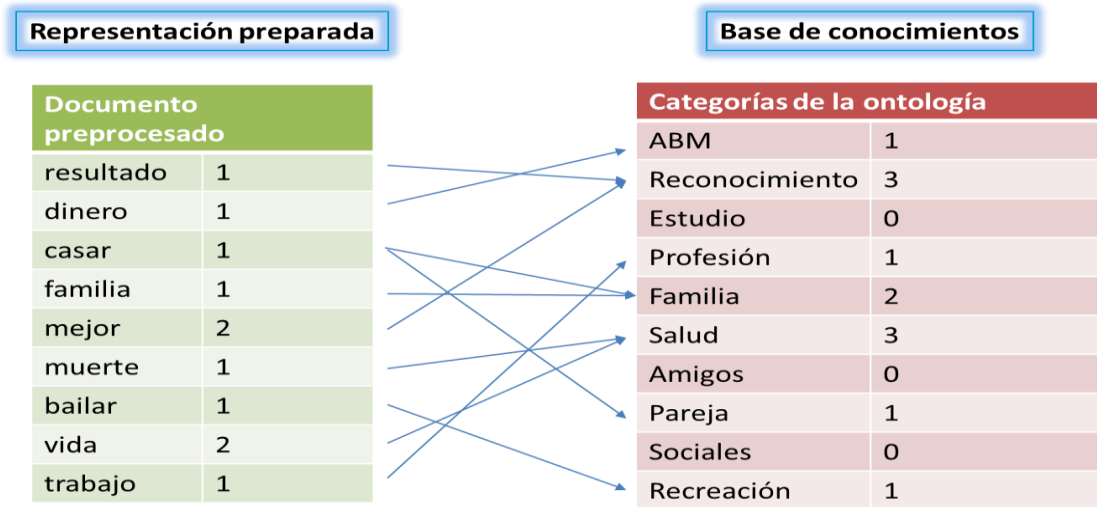


Figura 9. Identificación de conceptos por cada categoría. (Elaboración propia)

2.5 Modelo de Dominio

Debido a que el entorno organizacional en el cual está inmersa la descripción anterior no se puede clasificar como un negocio pues no cumple con las características determinantes del mismo planteadas por (Ciudad Ricardo, y otros, 2007) las cuales son:

- Presencia de flujos de información interconectados
- Posibilidad de determinar subsistemas
- No solapamiento de responsabilidades
- Fácil determinación de responsabilidades: mando, ejecución, recolección, etc.
- Existencia de reglas de funcionamiento

Es realizado un modelo de dominio que tiene por finalidad estipular una especificación del dominio del problema y los requerimientos desde el punto de vista de la clasificación por objetos y desde la perspectiva de entender los términos empleados en el dominio. Para descomponer el dominio del problema hay que

Capítulo 2. Características de la Herramienta

identificar los conceptos, los atributos y las asociaciones del dominio más importantes. (Larman, 2004, pág. 85-103)

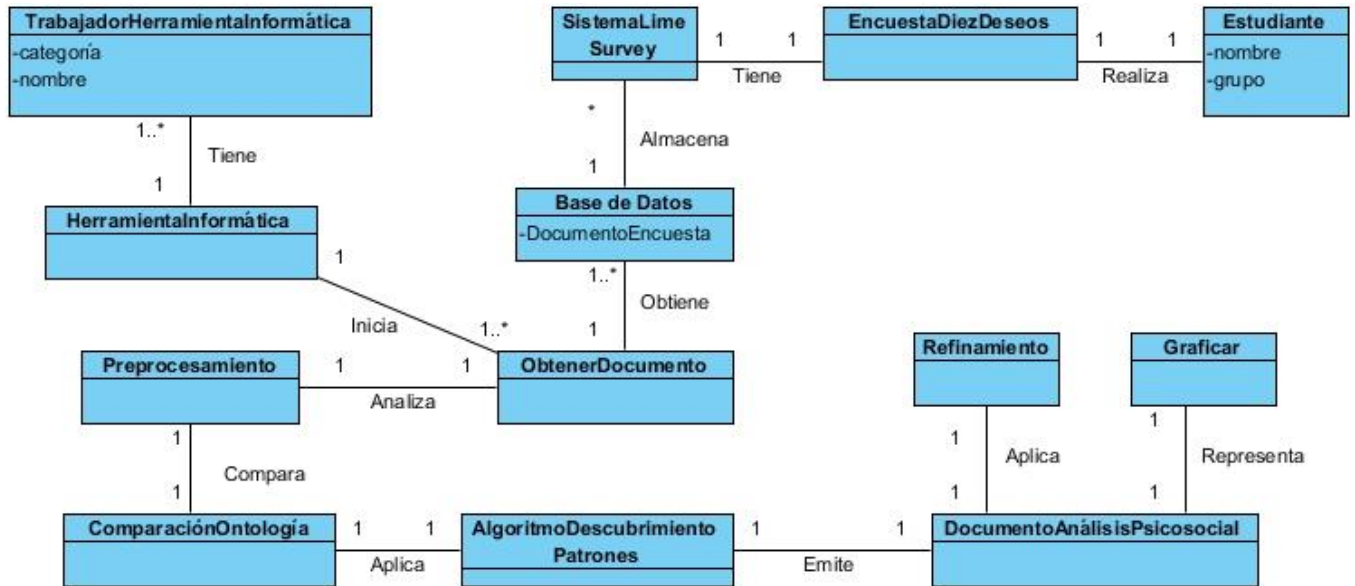


Figura 10. Modelo del Dominio. (Elaboración propia)

2.5.1 Definición y descripción de los conceptos del Modelo del Dominio:

- **Estudiante:** es el estudiante, que en conjunto con otros, componen los grupos de la Facultad 2 de la UCI a los cuales se les aplicará la encuesta de los 10 Deseos.
- **Sistema LimeSurvey:** sistema de encuesta en línea donde será publicada la encuesta de los 10 Deseos para su realización.
- **Encuesta de los 10 Deseos:** técnica psicológica con la que trabajaran los estudiantes para la exploración de la esfera motivacional.
- **Documento Encuesta:** contiene las respuestas de la encuesta de los 10 Deseos realizada por el estudiante.
- **Base de Datos:** contiene almacenados los distintos documentos con los datos de las respuestas de los estudiantes encuestados en el Sistema LimeSurvey para el análisis de los mismos, está alojada en el SGBD MySQL.

Capítulo 2. Características de la Herramienta

- **Herramienta Informática:** desarrollada en la investigación para que con el uso de la minería de texto realice el análisis de los datos obtenidos de las encuestas.
- **Trabajador de la Herramienta Informática:** es el usuario (profesor, tutor o especialista) que hará uso de la herramienta para analizar los datos en el proceso de caracterización del estudiante y del grupo de estudiantes universitarios.
- **Análisis de datos:** proceso por el cual los datos obtenidos pasaran por cada una de las fases de la minería de texto para la obtención de la información requerida, ellas son:
 - **Obtener documento:** permite obtener el documento por la herramienta informática para su análisis.
 - **Preprocesamiento:** con el que se realiza las rutinas, procesos y métodos requeridos para preparar los datos para las operaciones básicas de descubrimiento de conocimiento.
 - **Comparación con Ontología:** compara los conceptos extraídos luego del preprocesamiento con los términos asociados a cada categoría para clasificar cada concepto identificado por cada una de ellas y de esta forma poder realizar una caracterización.
 - **Algoritmos de descubrimiento de patrones:** son los que se emplean para el descubrimiento de la información en los datos textuales obtenidos.
- **Documento análisis psicosocial:** es el documento emitido por la herramienta como resultado del proceso de análisis y contiene la información sobre las principales categorías a las que más se inclina el estudiante o el grupo de estudiantes universitarios lo cual permite que permiten su caracterización.
- **Refinamiento:** es la que filtra información redundante y agrupa cerradamente los datos relacionados, en la cual se hará uso del k-means.
- **Graficar:** permite graficar la información obtenida a través de gráficos de estrella.

2.6 Especificación de los requerimientos

La especificación de los requerimientos consiste en una de las tareas más importantes y difíciles dentro de la ingeniería del software. Establece el mecanismo que ayuda a comprender y analizar que necesita el cliente y de esta manera proporcionarle una solución que esté acorde con sus necesidades. Esta es la base del desarrollo de la herramienta pues determina las funciones que realizará la misma, el modo en que serán realizadas y con qué finalidad.

Capítulo 2. Características de la Herramienta

2.6.1 Obtención de los requerimientos

Para la identificación y obtención de los requisitos existen varias técnicas que se pueden emplear según (Nicolás, 2003), ellas son:

- Entrevistas
- Sesiones de tormentas de ideas (Brainstorming)
- Revisión de la documentación técnica
- Cuestionarios
- Observación
- Prototipos
- Casos de uso

Para obtener los requisitos de la herramienta propuesta se utilizaron las siguientes técnicas de las mencionadas anteriormente:

- Entrevistas: fueron realizadas a los especialistas en el área de la psicopedagogía, a los que se le realizaron diversas preguntas para lograr un mejor entendimiento sobre la problemática a resolver y partiendo de ellas identificar los requisitos funcionales.
- Sesiones de tormentas de ideas (Brainstorming): se realizaron encuentros con profesores especializados del CICE para propiciar que se generaran ideas respecto al funcionamiento de la herramienta.
- Revisión de la documentación técnica: fueron revisadas y estudiada documentaciones técnicas de otros sistemas que pudieran aportar ideas para el funcionamiento de la herramienta.

2.6.2 Requisitos funcionales

Los requisitos funcionales son las condiciones que un software debe cumplir. La herramienta desarrollada cuenta con los siguientes:

RF.01: Autenticar usuario.

El trabajador de la herramienta informática se identifica en la misma, a través de un formulario de acceso que se le muestra para poder acceder a sus funcionalidades.

Capítulo 2. Características de la Herramienta

RF.02: Obtener documentos con las respuestas de las encuestas de la Base de Datos del Lime Survey.

Funcionalidad que permite obtener los datos del nombre, grupo y las respuestas de los estudiantes de la encuesta de los Diez Deseos, realizada en el Sistema Lime Survey y que están almacenados en la base de datos del propio sistema, los cuales serán la entrada para el análisis por parte de la herramienta informática.

RF.03: Analizar los datos contenidos en los documentos adquiridos.

Funcionalidad que permite realizar el proceso de análisis a los datos de cada documento obtenido, para realizar la extracción de información a partir de estos datos.

RF.04: Comparar los conceptos claves identificados con las categorías presentes en la ontología.

Funcionalidad que permite clasificar cada uno de los conceptos claves identificados en el análisis por cada categoría a la cual es asociado como término.

RF.05: Almacenar los vectores de frecuencia relativa en la base de datos de la herramienta.

Funcionalidad que permite guardar en la base de datos de la herramienta, cada uno de los vectores de frecuencia relativa que se obtienen luego del proceso de análisis anterior, los cuales funcionan como el documento de análisis psicosocial de cada estudiante.

RF.06: Refinar la información obtenida mediante el método de agrupamiento.

Funcionalidad que le permite al trabajador de la herramienta, optimizar la información obtenida, realizando agrupaciones a través del método de agrupamiento k means.

RF.07: Graficar la información obtenida.

Funcionalidad que permite representar la información obtenida a través de gráficos de estrella o de radar.

RF.08: Emitir el documento de análisis psicosocial.

Documento que contiene el vector de frecuencia relativa asociado a cada estudiante, el cual representa las categorías en la que cada se manifiesta con mayor incidencia cada estudiante.

Capítulo 2. Características de la Herramienta

2.6.3 Requisitos no funcionales

Los requisitos no funcionales son las propiedades o cualidades que debe presentar el producto. La herramienta informática posee los siguientes:

- Usabilidad:

RNF.01: La herramienta debe ser fácil de usar por parte de los usuarios.

Midiéndose este requisito funcional a través de tres factores principales (Markessinis, 2010), ellos son: la eficacia (el usuario logra lo que quiere), la eficiencia (lo logra rápidamente y sin obstáculos) y la satisfacción (después de completar la tarea).

RNF.02: Empleo de menús para hacer más fácil el acceso a las distintas funcionalidades de la herramienta.

RNF.03: Las terminologías usadas en la herramienta deben adaptarse al lenguaje y términos que emplean los usuarios en la rama donde se desempeñan para una mayor comprensión de la misma.

- Rendimiento:

RNF.04: Se deben respetar las buenas prácticas de la programación para incrementar el rendimiento en las operaciones costosas en tiempo y recursos.

- Soporte:

RNF.05: Se debe dar acceso a la documentación necesaria para la comprensión del sistema y acelerar el proceso de mantenimiento.

- Portabilidad:

RNF.06: La herramienta es multiplataforma, pues se puede ejecutar tanto en la plataforma de Windows como en Linux, está basado en Perl que es un lenguaje interpretado.

- Seguridad:

RNF.07: El acceso a la herramienta informática solo podrá ser a través de la autenticación de los usuarios pertenecientes a grupos de trabajadores definidos previamente por el administrador de la misma en la base de datos de la herramienta.

- Políticos-culturales:

Capítulo 2. Características de la Herramienta

RNF.08: La herramienta no debe contener ni presentar informaciones ni materiales que atenten contra la integridad de la ideología e idiosincrasia nacional.

RNF.09: Deben ser respetadas la ética y la moral al tratar con los datos de los estudiantes almacenados en la base de datos.

- Legales:

RNF.10: La utilización de la herramienta será puramente con fines educativos del sistema de enseñanza-aprendizaje, dejando excluida completamente su utilización para alguna actividad con fines lucrativos.

- Confiabilidad:

RNF.11: El acceso a los datos obtenidos de los estudiantes encuestados que son almacenados en la base de datos solo podrá ser por el personal designado y previamente autenticado en la herramienta.

- Interfaz:

RNF.12: La herramienta debe presentar una interfaz amigable que permita el rápido aprendizaje de los usuarios para trabajar con la misma.

RNF.13: La herramienta debe ser lo más sencilla posible de manera que la interacción del usuario con la misma pueda realizarse con conocimientos básicos de informática.

- Ayuda y documentación en línea:

RNF.14: La documentación por parte del grupo de desarrollo debe mantenerse actualizada.

- Hardware:

RNF.15: El PC Servidor para montar la herramienta informática, debe tener al menos 1GB de memoria RAM, Disco Duro de 40GB, Microprocesador Dual-Core a 1.86GHz, una tarjeta de red para la conexión LAN, una tarjeta de video de 128MB (puede ser On-Board).

- Software:

RNF.16: Plataforma Windows con sistema operativo Windows Vista, 7 u 8, o plataforma Linux en las distribuciones de Ubuntu 12.04, Debian Wheezy, Nova 2.0, como mínimo.

RNF.17: Sistema Gestor de Base de Datos MySQL v.4 o superior.

Capítulo 2. Características de la Herramienta

RNF.18: Módulos Perl::File::Type, Lingua::TreeTagger para Perl.

RNF.19: Servidor de aplicaciones Web Apache.

2.6.4 Definición de los casos de uso

2.6.4.1 Definición de los actores

Tabla 6. Definición de los actores de la herramienta

Actor	Justificación
Trabajador de la herramienta	Es el profesor guía, tutor o especialista del año encargado de trabajar con la herramienta para realizar el análisis de los datos en el proceso de caracterización.

2.6.4.2 Diagrama de casos de uso de la herramienta

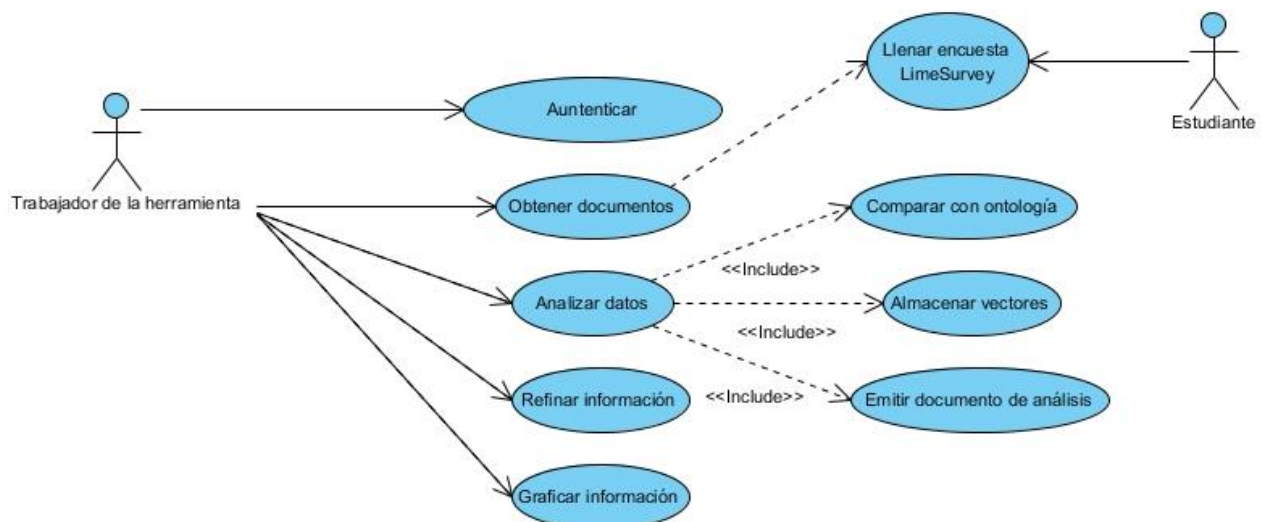


Figura 11. Diagrama de casos de uso de la herramienta. (Elaboración propia)

2.6.4.3 Descripción de los casos de uso

Descripción de los casos de uso de la herramienta, definiendo los eventos que efectúa el actor para trabajar con las funcionalidades de la herramienta. A continuación se muestra la descripción para el caso de uso **Auntenticar**, el resto de las descripciones para cada uno de los casos de usos pueden ser vistos en el [Anexo 1](#).

Capítulo 2. Características de la Herramienta

Tabla 7. Descripción del caso de uso Autenticar

Caso de uso		
CU-1	Autenticar	
Objetivo	Permitir la entrada del usuario Trabajador de la herramienta informática a la misma	
Actores	Trabajador de la herramienta	
Resumen	El caso de uso inicia cuando el Trabajador de la herramienta informática desea entrar al sistema con las credenciales que les fueron asignadas, el sistema brinda la posibilidad de introducir los parámetros, el Trabajador de la herramienta introduce los datos, el sistema le permite la entrada, el caso de uso termina.	
Complejidad	Baja	
Prioridad	Alta	
Precondiciones	El Trabajador de la herramienta debe tener una credencial perteneciente al sistema	
Postcondiciones	El Trabajador de la herramienta entra al sistema satisfactoriamente	
Requisitos Funcional	RF.01	
Flujo de eventos		
Flujo básico Autenticar		
	Actor	Sistema
1.	El caso de uso inicia cuando el Trabajador de la herramienta informática desea entrar al sistema.	Brinda la posibilidad de introducir los parámetros elementales: <ul style="list-style-type: none"> • Usuario • Contraseña
2.	El Trabajador de la herramienta informática introduce los parámetros obligatorios: <ul style="list-style-type: none"> • Usuario • Contraseña 	Si se introduce el usuario o contraseña incorrecta ver Alternativa 1: "Usuario, clave o aplicación incorrecta" Si se introduce la aplicación incorrecta ver Alternativa 2: "Campos vacíos"
3.		El caso de uso termina.
Flujos alternos		
Nº 1 Usuario o clave incorrecta		
	Actor	Sistema
1.	El Trabajador de la herramienta introduce valores incorrectos.	El sistema muestra un mensaje: "Usuario o clave incorrectos."
Prototipo no Funcional		
Nº 2 Campos vacíos		
	Actor	Sistema
1.	El Trabajador de la herramienta deja campos obligatorios vacíos.	El sistema señala los campos vacíos con un * y muestra un mensaje: "Existen campos obligatorios vacíos"

Capítulo 2. Características de la Herramienta

Prototipo no Funcional		
Prototipo no Funcional del Flujo básico Autenticar		
Relaciones	CU incluidos	
	CU extendidos	
Requisitos no funcionales		Software, Interfaz, Seguridad.

2.7 Conclusiones del capítulo

Se logró obtener una visión clara de las características del tema en cuestión, permitiendo la realización de un modelo de dominio, que representa las características del entorno donde se desenvuelve la problemática para una mejor comprensión de la misma, lo cual hizo más fácil el proceso de obtención de los requisitos funcionales y no funcionales.

Capítulo 3. Análisis y diseño del sistema

Capítulo 3. Análisis y diseño del sistema

3.1 Introducción

En este capítulo se realiza la descripción de la arquitectura seleccionada para desarrollar la herramienta y su fundamentación, son definidos los patrones de diseño a emplear y es descrita la forma en que se relacionan las clases de la herramienta informática para realizar la extracción de información a partir de la obtención de los datos que se encuentran en el Sistema Lime Survey.

3.2 Arquitectura de la Herramienta

Según (Hofmeister, y otros, 2000) la arquitectura del software sirve como un plan de diseño que se usa para negociar los requerimientos del sistema y como una forma de estructurar las discusiones con los clientes, desarrolladores y gestores. Sugieren que es una herramienta esencial para la gestión de la complejidad. Esta oculta detalles y permite a los diseñadores centrarse en las abstracciones clave del sistema. A su vez, (Bosch, 2000) manifiesta que la arquitectura del sistema afecta al rendimiento, solidez, grado de distribución y mantenibilidad de un sistema. El estilo y el diseño que se elige para una aplicación pueden depender de los requerimientos no funcionales del sistema: rendimiento, protección, seguridad, disponibilidad y mantenibilidad.

Entre las diversas arquitecturas que existen se encuentran las arquitecturas cliente-servidor. En esta arquitectura una aplicación se modela como un conjunto de servicios proporcionados por los servidores y un conjunto de clientes que usan estos servicios (Orfali, y otros, 1992). Uno de los tipos de esta arquitectura es el Cliente-Servidor de dos capas (es la más sencilla), en la que una aplicación se organiza como un servidor o múltiples servidores idénticos y un conjunto de clientes.

Pero para el desarrollo de la Herramienta se necesita de tres capas: lógica-presentación, procesamiento de la aplicación y gestión de los datos y resulta engorroso asociarlas en solo dos computadoras, el cliente y el servidor pues puede ocasionar problemas con la escalabilidad y rendimiento si se elige el cliente ligero, o problemas con la gestión del sistema si se usa el modelo de cliente rico, que son los dos modelos en que se clasifica esta capa. Para evitar estos problemas se emplea como solución usar la arquitectura Cliente-Servidor de tres capas, en esta, las tres capas mencionadas anteriormente son procesos lógicamente separados que se ejecutan sobre procesadores diferentes. (Sommerville, 2005)

Capítulo 3. Análisis y diseño del sistema

Para el análisis de los datos por parte de la Herramienta con minería de texto se hace necesario contar con una arquitectura funcional que diseñe la estructura en que se realizaran cada uno de los procesos que incurren en el mismo. Existen varios tipos de arquitecturas funcionales en la minería de texto:

- Simple modelo de entrada-salida para minería de texto.
- Ciclo iterativo de entrada-salida del usuario.
- Arquitectura funcional para minería de texto de alto nivel.
- Sistema de Arquitectura para sistema de minería de texto genérico.
- Sistema de arquitectura para un sistema de minería de texto avanzado y orientado a dominio.
- Sistema de arquitectura para un sistema de minería de texto avanzado con una base de conocimientos previos.

3.2.1 Cliente-servidor en tres capas

El uso de esta arquitectura permite optimizar las transferencias de información entre el servidor web y el servidor de la base de datos, reducir el tráfico de la red, el procesamiento de aplicación al ser la parte más volátil del sistema puede ser fácilmente actualizada debido a que está localizada centralmente, el procesamiento puede ser distribuido entre la lógica de la aplicación y servidor de gestión de datos, en cuyo caso las respuestas a las peticiones del usuario son más rápidas. Esta arquitectura será la base para implementar el Sistema de arquitectura para un sistema avanzado de minería de texto con base de conocimientos previos, ver Figura 12.

Esta arquitectura consta de tres niveles:

- **Capa de Presentación (o interfaz del usuario):** esta capa es la interfaz con el usuario y consiste de hardware como un PC o una estación de trabajo y un navegador de red. En ella se mostrara la funcionalidad de navegación para el usuario (Trabajador de la herramienta).
- **Capa Lógica de Funcionalidad / Negocio:** esta capa proporciona la funcionalidad al extremo de los usuarios y contiene la lógica del negocio (Herramienta). Es donde se harán las tareas de preprocesamiento y las diferentes operaciones de la minería de texto para el análisis de los datos. Proporciona el puente entre la primera y la tercera capa.

Capítulo 3. Análisis y diseño del sistema

- **Capa de Datos:** esta capa incluye la Base de Datos del LimeSurvey, cuyos datos serán los de entrada para el análisis de la Herramienta, la cual guardará los vectores de frecuencias relativas obtenidos en una Base de Datos en esta capa.



Figura 12. Arquitectura a emplear para el desarrollo de la herramienta. (Elaboración propia)

3.3 Patrón arquitectónico

Definen (Buschmann et al., 1996) a patrón como una regla que consta de tres partes, la cual expresa una relación entre un contexto, un problema y una solución. En líneas generales, un patrón sigue el siguiente esquema:

- Contexto. Es una situación de diseño en la que aparece un problema de diseño
- Problema. Es un conjunto de fuerzas que aparecen repetidamente en el contexto
- Solución. Es una configuración que equilibra estas fuerzas. Ésta abarca:
 - Estructura con componentes y relaciones
 - Comportamiento a tiempo de ejecución: aspectos dinámicos de la solución, como la colaboración entre componentes, la comunicación entre ellos, etc.

Partiendo de la definición anterior, propone los patrones arquitectónicos como descripción de un problema particular y recurrente de diseño, que aparece en contextos de diseño específico, y presenta un esquema genérico demostrado con éxito para su solución. El esquema de solución se especifica mediante la descripción de los componentes que la constituyen, sus responsabilidades y desarrollos, así como también la forma como estos colaboran entre sí.

Capítulo 3. Análisis y diseño del sistema

Algunos tipos de patrones arquitectónicos:

- Layers.
- Pipes and filters.
- Blackboard.
- Broker.
- Model-View-Controller.

Este último por sus características funcionales y ventajas, mostradas a continuación, es el empleado para el desarrollo de la herramienta.

3.3.1 Modelo Vista Controlador

El patrón Modelo Vista Controlador (MVC) es una propuesta de diseño de software utilizada para implementar sistemas donde se requiere el uso de interfaces de usuario. Surge de la necesidad de crear software más robusto, donde se potencie la facilidad de mantenimiento, se reutilice el código y exista la separación de conceptos (Álvarez, 2014).

Modelo

Es la capa donde se trabaja con los datos, contiene los mecanismos para acceder a la información y también para actualizar su estado. Para ello se emplea el gestor de base de datos MySQL, donde se encontrarán todas las funciones que accederán a las tablas y harán los correspondientes selects, updates, inserts, entre otras para realizar las consultas.

Vista

La vista contiene el código de la herramienta que va a producir la visualización de la interfaz de usuario. En ella tenemos los códigos HTML y PHP que nos permite mostrar la salida. En la vista generalmente trabajamos con los datos, sin embargo, no se realiza un acceso directo a éstos. La vista requerirá los datos al modelo y ella generará la salida, tal como nuestra herramienta requiere.

Controlador

Capítulo 3. Análisis y diseño del sistema

Contiene el código necesario para responder a las acciones que se solicitan en la herramienta, como analizar los datos, comparar el resultado con la base de conocimientos, mostrar la información obtenida, entre otros. Esta capa sirve de enlace entre la vista y el modelo, respondiendo a los mecanismos que puedan requerirse para implementar las necesidades de la herramienta. Su responsabilidad no es manipular directamente datos, ni mostrar ningún tipo de salida, sino servir de enlace entre las otras dos capas para implementar las diversas necesidades del desarrollo.



Figura 13. Patrón arquitectónico Modelo Vista Controlador. (Reenskaug, 1979)

3.4 Patrón de diseño

Los Patrones de Diseño son la base para la búsqueda de soluciones a problemas comunes en el desarrollo de software y otros ámbitos referentes al diseño de interacción o interfaces (Larman, 1999), introduciéndose el trabajo con patrones de diseño para asignar responsabilidades.

3.4.1 Patrones Generales de Software para Asignar Responsabilidades (GRASP)

Los patrones GRASP (por sus siglas en inglés del General Responsibility Assignment Software Patterns) son usados en el diseño de la herramienta para asignar las responsabilidades de las diferentes clases presentes en la misma y mantener el bajo acoplamiento y la alta cohesión entre ellas. Estos son:

- Experto

Capítulo 3. Análisis y diseño del sistema

Para asignar la responsabilidad de realizar la labor de conectar con la base de datos y obtener los mismos, de preprocesar los datos, de comparar con la ontología, a las diferentes clases de la herramienta informática que contienen los datos involucrados con cada una de estas labores, es decir los atributos.

- Controlador

Para asignar la responsabilidad de controlar el flujo de eventos de la herramienta, como los de extraer las palabras clave del texto, etiquetarlos, compararlos con la ontología, a clases específicas. Lo cual facilita la centralización de las actividades.

- Bajo Acoplamiento

Para mantener la menor dependencia entre las clases de la herramienta informática para impedir que cambios en las clases relacionadas ocasionen cambios en la clase loca y sean más difíciles de entender.

- Alta Cohesión

Para que cada una de las clases presentes en la herramienta informática realice la labor única que se le ha sido asignada, así se evita que las clases con responsabilidades estrechamente relacionadas no realicen un trabajo enorme.

3.5 Extracción de los datos del Sistema Lime Survey

Para realizar la extracción de los datos que se encuentran en la base de datos del Sistema Lime Survey, es realizada una conexión desde la herramienta informática hasta esta base de datos, mediante la clase controladora CC_contraladora que le asigna la responsabilidad a la clase C_conexión, la cual contiene como atributos los parámetros de identificación y configuración para poder hacer la conexión, como lo son, la dirección del servidor donde se encuentra alojada la base de datos, el puerto de red a través del cual se realizará la conexión, el nombre de usuario y la contraseña para identificarse y tener acceso a la base de datos.

Luego de realizada la conexión se hacen consultas a la base de datos, para obtener los datos correspondientes a los campos que se encuentran en la tabla contenedora de las respuestas de los estudiantes encuestados, los cuales identifican al mismo como son el id (identificador dentro de la tabla donde se encuentra, el nombre y el grupo), y los que contienen los diez deseos del estudiante. Estos datos obtenidos mediante las consultas son insertados en la base de datos propia de la herramienta informática,

Capítulo 3. Análisis y diseño del sistema

los cuales son los datos de entrada para el análisis que esta realiza en la extracción de información, que posteriormente se efectúa como se explicó en el Epígrafe 2.3 del Capítulo 2.

La clase controladora CC_controladora es la encargada de asignar las siguientes responsabilidades: de realizar la extracción de los Tokens a la clase C_extraerTokens, del etiquetado de los Tokens obtenidos mediante sus raíces morfológicas a la clase C_etiquetarTokens, del conteo de las veces que se repiten a la clase C_enumerarTokens, de la comparación con la ontología para la clasificación en cada una de las diez categorías a la clase C_compararOntología, del cálculo del vector de pertenencias de los Tokens a cada categoría de un estudiante y de la determinación del vector de frecuencias relativas de cada estudiante y grupo universitario a la clase C_calcularVectores, la cual es encargada finalmente de generar el Documento de análisis psicosocial que contiene esta información extraída y es almacenado en la CE_Vector de la base de datos de la herramienta informática.

3.6 Conclusiones del capítulo

Se obtuvo una arquitectura que permitió realizar un plan de diseño que estructuró de forma funcional como se efectuaría la minería de texto sobre la base de una arquitectura cliente servidor en tres capas, se logró determinar la forma en que serían asignadas las diferentes responsabilidades de cada clase en el proceso de desarrollo a través de los patrones de diseño seleccionados y se describió cómo funcionan entre sí, las clases encargadas de obtener los datos de la base de datos del Sistema Lime Survey y realizar el análisis de los mismos para la extracción de la información requerida en el proceso de caracterización psicosocial.

Capítulo 4. Implementación y validación

Capítulo 4. Implementación y validación

4.1 Introducción

En el presente capítulo se define el modelo de datos definido y se describen sus tablas, así como el diagrama de despliegue, son determinadas las estrategias de codificación y se establece la forma en que será realizada la validación de la herramienta y que técnicas se emplearan para realizarla.

4.2 Modelo de datos

El modelo de datos, permite , haciendo uso de un diagrama entidad-relación (DER) definir los objetos de datos primarios que va a procesar la herramienta informática, la composición de cada objeto de datos y los atributos que describe el objeto, donde residen actualmente los objetos y la relación entre los objetos y los procesos que los transforman (Pressman, 2002). En la siguiente figura se muestra el modelo de datos de la herramienta informática:

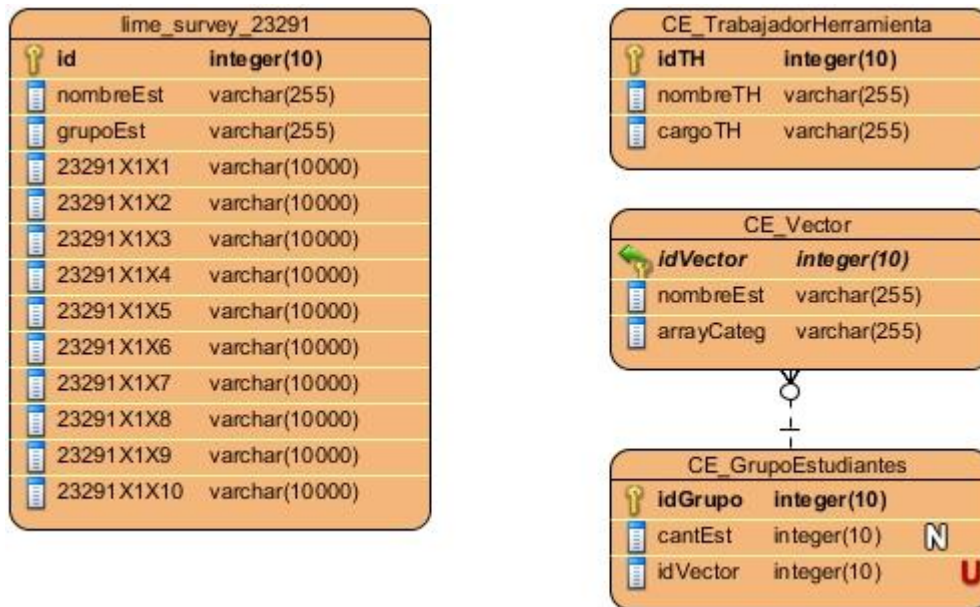


Figura 14. Diagrama del Modelo de Datos (Elaboración propia)

Capítulo 4. Implementación y validación

Descripción de las tablas

Tabla 8. Descripción de la tabla lime_survey_23291 del Sistema Lime Survey

Nombre: lime_survey_23291		
Descripción: tabla autogenerada por el Sistema Lime Survey que contiene la respuesta de las encuestas de los estudiantes y de donde se obtendrán las mismas para el análisis por parte de la herramienta.		
Atributo	Tipo	Descripción
Id	Integer	Identificador único para cada estudiante con sus respuestas.
nombreEst	Varchar	Nombre del estudiante encuestado.
grupoEst	Varchar	Grupo al que pertenece el estudiante encuestado.
23291X1X1	Varchar	Campo que contiene la respuesta asociada a la pregunta 1.
23291X1X2	Varchar	Campo que contiene la respuesta asociada a la pregunta 2.
23291X1X3	Varchar	Campo que contiene la respuesta asociada a la pregunta 3.
23291X1X4	Varchar	Campo que contiene la respuesta asociada a la pregunta 4.
23291X1X5	Varchar	Campo que contiene la respuesta asociada a la pregunta 5.
23291X1X6	Varchar	Campo que contiene la respuesta asociada a la pregunta 6.
23291X1X7	Varchar	Campo que contiene la respuesta asociada a la pregunta 7.
23291X1X8	Varchar	Campo que contiene la respuesta asociada a la pregunta 8.
23291X1X9	Varchar	Campo que contiene la respuesta asociada a la pregunta 9.
23291X1X10	Varchar	Campo que contiene la respuesta asociada a la pregunta 10.

Tabla 9. Descripción de la tabla CE_TrabajadorHerramienta

Nombre: CE_TrabajadorHerramienta		
Descripción: tabla que contiene a los trabajadores profesores guías, tutores y especialistas) que trabajan con la herramienta para la realización del análisis de los datos.		
Atributo	Tipo	Descripción
idTH	Integer	Campo con un valor entero único que identifica a cada trabajador de la herramienta.
nombreTH	Varchar	Nombre del trabajador de la herramienta.
cargoTH	Varchar	Cargo docente que ocupa dentro de la facultad el trabajador de la herramienta.

Capítulo 4. Implementación y validación

Tabla 10. Descripción de la tabla CE_Vector

Nombre: CE_Vector		
Descripción: tabla que contiene los vectores generados por la herramienta, los cuales contienen la información del resultado del análisis de las encuestas.		
Atributo	Tipo	Descripción
idVector	Integer	Campo con un valor entero único que identifica a cada vector.
nombreEst	Varchar	Nombre del estudiante al que pertenece el vector.
arrayCateg	Varchar	Arreglo de Categorías que contienen las categorías que identifican los motivos del estudiante.

Tabla 11. Descripción de la tabla CE_GrupoEstudiantes

Nombre: CE_GrupoEstudiantes		
Descripción: tabla que contiene los vectores asociados a todos los estudiantes de un mismo grupo.		
Atributo	Tipo	Descripción
idGrupo	Integer	Número que identifica al grupo, el mismo que tiene dentro de la facultad.
cantEst	Integer	Cantidad de estudiantes que contiene el grupo.
idVector	Integer	Identificador que relaciona al vector de la tabla CE_Vector.

4.3 Diagrama de despliegue

El diagrama de despliegue permite mostrar cómo se configuran las instancias de los componentes y los procesos para la ejecución run-time (en tiempo de ejecución) en las instancias de los nodos de proceso de la herramienta informática (Larman, 2003). Como se muestra en la siguiente figura:

Capítulo 4. Implementación y validación

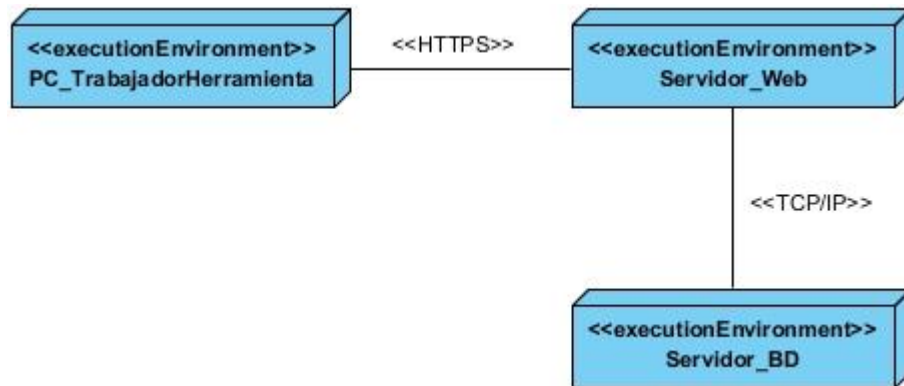


Figura 15. Diagrama de Despliegue

Tabla 12. Descripción del Diagrama de despliegue

Nodos y enlaces de comunicación	Descripción
PC_TrabajadorHerramienta	Ordenador desde donde el trabajador de la herramienta accede a la aplicación. Sistema operativo GNU/Linux o Windows, navegador Web Mozilla Firefox 20 o superior.
Servidor_Web	Servidor Web Apache 2.0
Servidor_BD	Servidor donde se encuentra ubicada la base de datos, debe estar montado el SGBD MySQL5.0 en un sistema operativo GNU/Linux o Windows.
HTTPS	Protocolo empleado para establecer la conexión entre las PC cliente y el servidor web.
TCP/IP	Protocolo empleado para realizar la conexión entre el servidor web y el servidor donde se encuentra ubicada la base de datos.

4.4 Estrategias de codificación

4.4.1 Estándar de codificación

Es de gran importancia para la calidad de la herramienta y para que obtenga un buen rendimiento emplear técnicas de codificación concretas y realizar buenas prácticas de programación con vistas a generar un código de alta calidad.

Capítulo 4. Implementación y validación

4.4.1.1 Denotación, llaves de apertura, cierre y tamaño de líneas

Se debe usar la indentación sin tabulaciones, con un equivalente a 4 espacios, para mantener integridad en las revisiones. El uso de las llaves “{}” será en una nueva línea. La longitud de las líneas de código es aproximadamente de 75-80 caracteres, para mantener la legibilidad del código.

```
while (<FILE>
{
    $_ = lc;
    s/--/ -- /g;
    s/[.,:;?"'""""'!()]//g;
    @words = split(/\s+/);
    foreach $word (@words)
    {
        print OUT "$word\n";
    }
}
```

Figura 16. Identación y llaves.

4.4.1.2 Conversión de nomenclatura

Las variables se rigen por la nomenclatura camelCase. Siempre comienzan con minúscula y en caso de nombres compuestos la primera letra de cada palabra comienza con mayúscula.

```
@ficherosEtiqu = glob("*.txt");
print "@ficherosEtiqu\n";

foreach $fichero (@ficherosEtiqu){
```

Figura 17. Nomenclatura de variables.

Las funciones se rigen por la nomenclatura camelCase. Siempre comienzan con minúscula y en caso de nombres compuestos la primera letra de cada palabra comienza con mayúscula. Los parámetros son separados por espacio luego de la coma que los separa.

Capítulo 4. Implementación y validación

```
sub preprocesarDatos
{
    foreach $fichero (@ficheros)
    {
        #Fichero que contiene el documento encuesta

        open (FILE, "/opt/lampp/htdocs/herramienta/estudiantes/$fichero") or die("Fichero no encontrado");
    }
}
```

Figura 18. Nomenclatura de funciones.

Los ficheros siempre se escriben en minúscula y en caso de nombres compuestos se usa el carácter subrayado "_".

- Vistas: intuitivo y relacionado con el formulario y/o vista que representa.
- Modelos: con el mismo nombre de la clase que representa que contiene en el nombre el prefijo CE_.
- Controladoras: con el mismo nombre de la clase que representa.

4.4.1.3 Estructuras de control

Se incluye un espacio entre las estructuras de control (if, for, foreach, while, switch) y los paréntesis. Se recomienda utilizar siempre llaves de apertura y cierre, incluso en situaciones en las que técnicamente son opcionales. Esto aumenta la legibilidad y disminuye la probabilidad de errores lógicos.

```
#Para mostrar solamente el lemma
foreach my $token ( @{ $tagged_text->sequence() } )
{
    if ( defined $token->lemma() )
    {
        print OUT $token->lemma(), "\n";
    }
}
```

Figura 19. Nomenclatura de estructuras de control.

4.5 Validación de requisitos

Se determina si los requisitos cumplidos durante el desarrollo son consistentes con las necesidades del problema. Son examinados para asegurar que han sido establecidos sin ambigüedad, sin inconsistencias,

Capítulo 4. Implementación y validación

sin omisiones, que los errores detectados hayan sido corregidos y que el resultado del trabajo se ajusta a los estándares establecidos.

Las técnicas empleadas para la validación de los requisitos fueron las siguientes:

- Revisiones de requerimientos: para la lectura, revisión y corrección de la definición de requisitos, de este modo se obtienen los posibles errores que pueden existir en la especificación de los requisitos de la herramienta o validar la correcta interpretación de la información.
- Construcción de prototipos: consiste en construir una maqueta de la futura herramienta a partir de los requisitos recogidos en la especificación, permitan al trabajador tener una idea de la estructura de la interfaz del sistema; así este corrige errores o añade aspectos para su completitud.
- Criterio de expertos (variante Delphi): para la evaluación del grado de similitud de las respuestas emitidas por la herramienta con las respuestas que daría el análisis de los datos de un experto en el entorno donde se desarrolla el problema.

4.5.1 Criterio de expertos

Para una correcta aplicación de la técnica de criterio de expertos se debe garantizar la selección de estos especialistas. Para ello método Delphi de esta técnica consta de dos etapas, la selección de los indicadores y la medición. En la primera etapa son seleccionados los indicadores que podrán ser evaluados por los expertos referentes al análisis de los datos en la caracterización psicosocial, para ello se les realiza una encuesta, la cual se podrá ver en más detalle en el [Anexo 2](#), como ejemplo para la elaboración de la misma se toma la encuesta confeccionada por (Verdecia, 2011). En la etapa de medición se confecciona la lista que a criterio del investigador cumplen con los criterios de expertos y que han trabajado durante un prolongado tiempo sobre asuntos afines al problema.

Medición

Para la selección de los expertos se escogieron 5 especialistas pertenecientes al departamento CICE de la Facultad 2, a los que se les envió el cuestionario. Luego se confeccionó una lista con 3 de ellos, dos con grado de Doctor en Ciencias Pedagógicas y un Máster en Ciencias Pedagógicas, que a criterio del investigador cumplían los requisitos de expertos y que además llevaban un largo tiempo trabajando sobre actividades relacionadas con la caracterización psicosocial de estudiantes y grupos universitarios en la UCI.

Capítulo 4. Implementación y validación

Fueron tomados los siguientes aspectos en consideración: título universitario, grado científico, categoría académica, año de experiencia en la docencia, años de experiencia en la rama de la psicología, el nivel de dominio sobre el tema de las encuestas y las fuentes de argumentación.

1. Determinación del coeficiente de conocimiento (Kc): se les planteo a los especialistas que evaluaran siguiendo una escala del 1 al 10 su nivel de conocimiento sobre la caracterización psicosocial de estudiantes y grupos universitarios. De acuerdo a los valores seleccionados por los expertos se calcula Kc dividiendo el valor marcado entre 10. A continuación un resumen de los valores obtenidos de Kc, registrados en la Tabla 25 del [Anexo 3](#):

Tabla 13. Resumen de ubicación de los expertos según Kc

Coeficiente de conocimiento	1.00	0.90
Cantidad de expertos	2	1

2. Coeficiente de argumentación (Ka): fue solicitado a cada experto autoevaluarse a través del cuestionario sobre ciertas fuentes de argumentación, los resultados son recogidos en la siguiente Tabla 15, registrándose en la Tabla 26 del [Anexo 3](#).

Tabla 14. Ubicación de los expertos en las fuentes de argumentación

No.	Fuentes de argumentación	Alto	Medio	Bajo
1	Análisis teóricos realizados	0.30	0.20	0.10
2	Experiencia obtenida	0.50	0.40	0.30
3	Conocimiento de trabajos de autores nacionales	0.05	0.04	0.03
4	Su intuición	0.05	0.04	0.03

El resumen de la ubicación de los expertos según Ka:

Tabla 15. Resumen de la ubicación de los expertos según Ka

Coeficiente de argumentación	1	0.9	0.8	0.7	0.6	0.5
Cantidad de expertos	0	1	1	1	0	0

Posteriormente se calcularon los coeficientes de competencias (K) según la fórmula $K = \frac{Kc+Ka}{2}$

Capítulo 4. Implementación y validación

Intervalos para definir el nivel de competencia de un experto:

Tabla 16. Intervalos para medir la competencia de un experto

Nivel de competencia		
Alto	Medio	Bajo
$0.8 < K < 1.0$	$0.5 < K < 0.8$	$K < 0.5$

La tabla con K_c , K_a , K , y el correspondiente nivel de competencia para los expertos se muestra en la Tabla 27 que se encuentra en el [Anexo 3](#).

Valoración de forma general de los expertos y selección final

Distribución de los expertos según su nivel de competencia:

Tabla 17. Distribución según la competencia de los expertos

Nivel de competencia	Cantidad	%
Alta	3	100
Media	0	0
Baja	0	0
Total	3	100

Como se puede observar, los tres especialistas clasifican en el nivel alto de competencia por lo cual se puede garantizar la confiabilidad de las respuestas que emitan sus análisis de los datos.

4.5.2 Proceso de validación

Luego de haber aplicado el método anterior y garantizado la selección de los tres expertos, dos Doctores en Ciencias Pedagógicas y un Máster en la misma disciplina, les son entregados los datos obtenidos como resultado de las respuestas emitidas por el Sistema Lime Survey de las encuestas de los Diez Deseos de cada estudiante, luego de que realicen el análisis, les son entregadas las respuestas que emitió la herramienta tras analizar los mismos datos obtenidos del Sistema Lime Survey, y estos comprueban el grado de similitud entre las respuestas. De esta forma se comprueba que tan bueno es el análisis que realiza la herramienta. El proceso se muestra en el siguiente ejemplo:

Capítulo 4. Implementación y validación

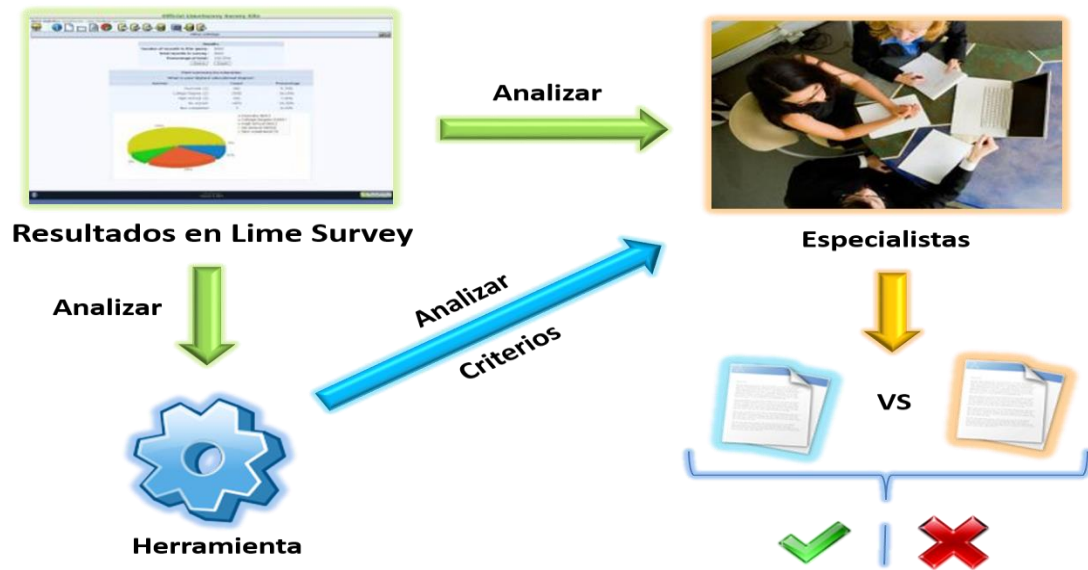


Figura 20. Validación del análisis de la herramienta.

4.6 Conclusiones del capítulo

Fue obtenida la estructura que tendrán los datos almacenados en la base de datos a partir del modelo de datos confeccionado así como la forma en que será desplegada la herramienta para su uso. La implementación del código se hizo de forma fácil y legible empleando las buenas prácticas de la programación que propiciaron las normas de las nomenclaturas establecidas a partir de las estrategias de codificación seleccionadas. Se obtuvieron los expertos que cumplían los requisitos necesarios para poder realizar una correcta y confiable validación de la herramienta.

Beneficios

La información que se obtiene luego de realizado el análisis por la herramienta informática desarrollada, permitirá a los profesores guías, tutores y especialistas de los años obtener conocimiento acerca de cuáles son los motivos que guían a los estudiantes en la Universidad. Por ejemplo, sobre cuál es el motivo que guía al estudiante a terminar la carrera universitaria, si es por exigencia de sus padres o si es por su propia voluntad de superación personal como futuro profesional, o si el estudiante se mantiene en la Universidad porque aquí se encuentra en mejores condiciones que en su casa, por problemas familiares que existen, o detectar el potencial de un estudiante sobre determinadas materias que cursa y de esta forma podría ser incorporado a la práctica profesional en las áreas donde mejor se desenvuelva, mejorando de esta forma su productividad. De esta forma el profesor podrá saber sobre que rasgos trabajar para mejorar la vinculación del estudiante a la institución y a sus actividades.

Aplicar la encuesta de los Diez Deseos a los estudiantes a inicios del curso escolar y culminando este, le permitirá a los profesores guías con la obtención de las informaciones emitidas por la herramienta desarrollada luego del análisis de los datos, conocer cómo ha variado la esfera motivacional del estudiante, y el profesor podrá evaluar el trabajo que ha realizado para incidir positivamente en los motivos del estudiante.

Conclusiones generales

- El análisis realizado sobre los distintos sistemas homólogos tanto a nivel internacional como nacional, permitió adquirir los conocimientos necesarios para una mejor comprensión del objeto de estudio y sentar las bases de la investigación.
- El estudio del proceso de caracterización psicosocial que se realiza en la Facultad 2, permitió establecer las características de la herramienta propuesta y las nociones para realizar el análisis y diseño del desarrollo de la misma.
- La elaboración de la base de conocimientos basándose en una ontología, permitió agrupar los términos con los cuales se realiza la comparación de los Tokens obtenidos, en diez categorías a las cuales son asociados los conceptos identificados luego de esta comparación.
- El uso de los patrones de diseño y de los estándares de codificación empleados proporcionaron los resultados deseados, logrando maximizar la organización y uniformidad durante la implementación de la herramienta.
- La selección de los especialistas para realizar la validación del prototipo funcional realizado garantizó que se puedan obtener criterios confiables acerca de la exactitud del análisis de los datos que realiza la misma.

Recomendaciones

Extender este análisis de datos realizado por la herramienta informática desarrollada, a los procesos donde se realice la evaluación de profesores de la Universidad donde se busca obtener un perfil psicológico de los mismos, como las encuestas que se le hacen a los profesores recién graduados para su inserción en determinadas especialidades como la Seguridad Informática y en el área de los proyectos relacionados con el desarrollo de software para el Ministerio del Interior y las FAR.

Continuar con la automatización de otras técnicas psicológicas (completamiento de frases, composición, inventario de problemas, adivina quién es, test socio métrico, expectativas y propósitos) que son empleadas en el proceso de caracterización en la Facultad 2 para lograr la realización de un sistema donde estén todas coleccionadas.

Referencias bibliográficas

Alarcón, Rodrigo. 2009. *Extracción automática de contextos definitorios en copus especializados.* Barcelona : Universidad Pompeu Fabra, 2009. Tesis de Doctorado.

Álvarez, Miguel Ángel. 2014. Desarrollo Web. *Que es MVC.* [En línea] 2014. [Citado el: 10 de abril de 2014.] <http://www.desarrolloweb.com/articulos/que-es-mvc.html>.

Armstrong. 2014. ArmSTRONG-Sistema de Evaluación Integral. *ArmSTRONG.* [En línea] Armstrong Intellectual Capital Solutions, S.A. de C.V., 2014. [Citado el: 20 de marzo de 2014.] <http://www.armstrong.com.mx/Producto/armstrong.php>.

Baron, Robert A. y Byrne, Donn. 1998. *Psicología Social.* 1998.

Basso Mesa, Anileidy y Gómez Almaguer, Dayaima. 2009. *Sistema automatizado para el proceso de caracterización de los estudiantes.* Universidad de las Ciencias Informáticas. La Habana : s.n., 2009. Trabajo de diplomado.

Beck, Kent. 1999. *Extreme Programming Explained: Embrace Change.* 1a Edición. s.l. : Addison-Wesley Pub Co., 1999.

Bellman, Richard. 1978. *Introduction to artificial intelligence. Can computers think?* San Francisco : s.n., 1978.

Bermúdez Morris, Raquel y Pérez Martín, Lorenzo M. 2002. *La facilitación de la dinámica grupal en educación.* 2002.

Bosch, J. 2000. *Design & Use of Software Architectures.* s.l. : Addison-Wesley, 2000.

Bosch, M. 2012. *Las ontologías del Web semántico: su uso como espacios para la pluralidad y la diversidad.* 2012.

Buschmann, F., y otros. 1996. *Pattern - Oriented Software Architecture. A System of Patterns.* [Libro] Inglaterra : John Wiley & Sons, 1996.

—. 1996. *Pattern-Oriented Software Architecture. A System of Patterns.* Inglaterra : John Wiley & Sons, 1996.

Cabrera Ardanás, Maudel, y otros. 2013. *Utilización de la aplicación AsDi en el programa de caracterización psicológica: aportes, limitaciones y retos.* La Habana : FORDES, 2013.

Castellanos Simons, Doris, y otros. 2001. *Hacia una concepción del aprendizaje desarrollador.* 2001.

Referencias bibliográficas

- Charniak, E. A. y McDemott, D. V. 1985.** *Información bibliográfica.* s.l. : Addison-Wesley Pub. Co., 1985.
- Charniak, Eugene y McDermott, Drew V. 1985.** *Introduction to Artificial Intelligence.* s.l. : Addison-Wesley Pub. Company, 1985.
- Ciudad Ricardo, Febe Ángel y Herrera Martínez, Yosnel. 2007.** *DoMet COMO PROPUESTA PARA LA MODELACIÓN DE ENTORNOS ORGANIZACIONALES COMPLEJOS Y DIFUSOS.* La Habana : UCI, 2007.
- Collazo Delgado, Basilia y Puentes Albá, María. 1992.** *La orientación en la Actividad Pedagógica.* s.l. : Pueblo y Educación, 1992. pág. 137.
- CPAN. 2013.** CPAN. *Comprehensive Perl Archive Network.* [En línea] Perl.org, 2013. [Citado el: 14 de mayo de 2014.] <http://search.cpan.org/>.
- Cueli, J., y otros. 1990.** *Teorías de la Personalidad.* s.l. : Trillas, 1990.
- Davyt Dávila, Nicolás. 2003.** *Ingeniería de Requerimientos: Una guía para extraer, analizar, especificar y validar los requerimientos de un proyecto.* s.l. : Universidad ORT Uruguay, 2003.
- de la Puente Viedma, Carlos. 1995.** *SPSS/ PC+ Una guía para la investigación.* Salamanca, Madrid : Editorial Complutense, 1995. pág. 490.
- Feldman, Ronen y Sanger, James. 2007.** *The Text Mining Handbook.* USA : Cambridge University Press, 2007.
- Foo, Y. y Ding, S. 2002.** *Ontology. Research and Development. Part-1 A Review of Ontology Generation.* s.l. : Journal of Informatic Science., 2002.
- Franco, A. C. 2009.** *Sistemas de Gestión de Bases de Datos.* España : s.n., 2009.
- González Cornejo, José Enrique. 2008.** *¿Qué es UML? El Lenguaje de Modelado Unificado. DocIRS.* [En línea] 2008. [Citado el: 24 de marzo de 2014.]
- González Soca, Ana María y Reynoso Cápiro, Carmen. 2002.** *Nociones de Sociología Psicología y Pedagogía.* La Habana : Pueblo y Educación, 2002.
- Haugeland, J. 1985.** *Artificial Intelligence: The Very Idea.* Massachussets Institute of Technology Cambridge : MIT Press, 1985.
- Hofmeister, C., Nord, R. y Soni, D. 2000.** *Applied Software Architecture.* s.l. : Addison-Wesley, 2000.
- Hurtado, Jimena, Bonilla, Elssy y Jaramillo, Christian. 2009.** *La Investigación - Aproximaciones a la Construcción del Conocimiento Científico.* s.l. : Alfaomega, 2009.

Referencias bibliográficas

- Introducción a Perl (Parte 1)*. **Montez Díaz, María José**. 2013. 3, 2013, Hackers & Developers, págs. 45-50.
- ITESCAM**. 2014. ¿Qué es una herramienta CASE? *ITESCAM*. [En línea] 2014. [Citado el: 4 de abril de 2014.] <http://www.itescam.edu.mx/principal/sylabus/fpdb/recursos/r88017.PDF>.
- Kurzweil, Ray**. 1990. *The Age of Intelligent Machines*. 1990.
- La lematización en español: una aplicación para la recuperación de información*. **Gómez Díaz, Raquel**. 2005. s.l. : Ediciones Trea, 2005.
- Larman, Craig**. 1999. *UML y Patrones. Introducción al análisis y diseño orientado a objetos*. Primera. s.l. : Prentice Hall, Hispanoamerica S.A., 1999. Vols. I, Capítulo 18.
- . 2003. *UML y patrones. Una introducción al análisis y diseño orientado a objetos y al proceso unificado*. Segunda. s.l. : Prentice Hall, 2003.
- Las ontologías en la representación y organización de la información*. **González Pérez, Y**. 2006. 4, s.l. : Acimed, 2006, Vol. 14.
- LimeSurvey**. 2014. LimeSurvey. [En línea] 2014. [Citado el: 20 de marzo de 2014.] <http://www.limesurvey.org/es/>.
- Lizama, Uriel**. 2008. Perl en Español. [En línea] 2008. [Citado el: 16 de mayo de 2014.] <http://perlenespanol.com/articulos/archivo/000134.html>.
- Markessinis, Andreas**. 2010. Como medir la usabilidad de una Web. *AndreasMarkessinis*. [En línea] 2010. [Citado el: 14 de abril de 2014.] <http://andreasmarkessinis.com/blog/como-medir-la-usabilidad-de-una-web/>.
- Martínez Rivera, Anigleidis y Estrada Rodríguez, Maisel Luis**. 2012. *Solución informática para la gestión del proceso de caracterización estudiantil y generación de reportes del Sistema de Caracterización Integral*. La Habana : s.n., 2012. Trabajo de diploma.
- Ministerio de Educación Superior**. 2012. *Objetivos de Trabajo para el año 2013 y hasta el 2016*. La Habana : MES, 2012.
- Nasukawa, Kuniya y Masayuki, Oishi**. 2001. *Inaccessibility of the domain-initial nucleus in high-pitch agreement*. 2001.
- Nilsson, Nils J**. 1998. *Artificial Intelligence: A New Synthesis*. s.l. : Morgan Kaufmann, 1998.
- Oracle Corp**. 2011. Las principales características de MySQL. MySQL 5.0 Reference Manual. *MySQL*. [En línea] 2011. [Citado el: 16 de marzo de 2014.] <http://dev.mysql.com/doc/refman/5.0/es/features.html>.

Referencias bibliográficas

- Orfali, R. y Harkey, D. 1992.** *Client/Server Programming with OS/2*. New York : Van Nostrand Reinhold, 1992.
- Otero Vidal, Mari Carmen. 2007.** *Capítulo 4. Diagramas de Interacción. UML*. 2007.
- Palomo Duarte, Manuel y Montero Pérez, Ildefonso. 2007.** *Programación en PHP a través de ejemplos*. s.l. : Autoedición, 2007.
- Paquete tecnológico para la evaluación psicológica en salud ocupacional. Grupo de autores. 2012.* 2012, Revista Cubana de Salud y Trabajo, págs. 62-74.
- Perspectivas para la integración de la minería de textos y la gestión del conocimiento. Bordón, L. y D'Avanzo, E. 2004.* 85, 2004, The IPTS Report.
- PHP. 2014.** *PHP*. s.l. : The PHP Group, 2014.
- Poole, David, Mackworth, Alan y Goebel, Randy. 1998.** *Computational Intelligence: A Logical Approach*. 1998.
- Pressman, Roger S. 2002.** *Ingeniería de Software, un enfoque práctico*. Quinta. s.l. : McGraw-Hill Companies, 2002. págs. 559-573.
- Reyes Cueto, Dreidys R. y Ferrera Ches, Dany. 2012.** *Registro de información para la caracterización de los estudiantes de la Universidad de las Ciencias Informáticas*. Universidad de las Ciencias Informáticas. La Habana : s.n., 2012. Trabajo de diploma.
- Rich, Elaine y Knight, Kevin. 1991.** *Artificial intelligence*. s.l. : McGraw-Hill, 1991.
- . **1994.** *Inteligencia Artificial*. s.l. : McGraw-Hill, 1994.
- Robles Martínez, Gregorio y Ferrer Zarzuela, Jorge. 2002.** *Programación eXtrema y Software Libre*. Madrid : s.n., 2002.
- Rouse, Margaret. 2012.** *Principios de la analítica de datos: una guía esencial*. 2012.
- Russel, Stuart Jonathan y Norvig, Peter. 2010.** *Artificial Intelligence: A Modern Approach*. s.l. : Prentice Hall, 2010.
- Sæther Bakken, Stig, y otros. 2001.** *Manual de PHP*. s.l. : Grupo de documentación de PHP, 2001.
- Sierra, Gerardo, y otros. 2008.** *Definitional verbal patterns for semantic relation extraction. Terminology*. 2008.
- Snow, Rion, y otros. 2007.** *Learning to merge word senses. In Proceeding of EMNLP*. 2007.

Referencias bibliográficas

- Sosa, Eduardo. 1997.** Procesamiento del lenguaje natural: revisión del estado actual, bases teóricas y aplicaciones. *El profesional de la información*. [En línea] 1997. [Citado el: 10 de abril de 2014.] http://www.elprofesionalde lainformacion.com/contenidos/1997/enero/procesamiento_del_lenguaje_natural_revisin_del_estado_actual_bases_tericas_y_aplicaciones_parte_i.html.
- Targetware Informática S.A.C. 2007.** Targetware Informática S.A.C. [En línea] Software.com.ar, 2007. [Citado el: 21 de febrero de 2014.] <http://www.software.com.ar/visual-paradigm-para-uml.html>.
- Torruella, J. y Llisterri, J. 1999.** *Diseño de corpus textuales y orales. Filología e informática. Nuevas tecnologías en los estudios filológicos*. Filología Española, Universidad Autónoma de Barcelona. Barcelona : Milenio, 1999. págs. 45-77, Seminario de Filología e Informática.
- Valdivia Mena, Leticia y Torres Almarales, Yoandri. 2010.** *Herramienta de apoyo al proceso de caracterización psicológica para la entrada a proyectos informáticos*. Universidad de las Ciencias Informáticas. La Habana : s.n., 2010. Trabajo de diploma.
- Verdecia Martínez, Edistio Yoel. 2011.** *Metodología para la certificación formativa de roles desde la práctica profesional*. Universidad de las Ciencias Informáticas. La Habana : s.n., 2011.
- Winston, P. H. 1992.** *Artificial Intelligence*. s.l. : Addison-Wesley, 1992.
- Xanthos, Aris. 2012.** CPAN. [En línea] perl.org, 2012. [Citado el: 15 de mayo de 2014.] <http://search.cpan.org/dist/Lingua-TreeTagger/lib/Lingua/TreeTagger.pm>.

Anexo 1

Tabla 18. Descripción del Caso de Uso Obtener documentos

Caso de uso		
CU-2	Obtener documentos	
Objetivo	Obtener los documentos que contienen las respuestas de las encuestas almacenadas en el Sistema Lime Survey.	
Actores	Trabajador de la herramienta	
Resumen	El caso de uso inicia cuando el actor accede a la opción Obtener documento, la herramienta informática lista los documentos que se encuentran en la base de datos del Sistema Lime Survey y le permite al actor almacenar en la base de datos de la herramienta informática estos documentos. La herramienta obtiene y almacena los documentos de las respuestas de las encuestas y el caso de uso termina.	
Complejidad	Media	
Prioridad	Alta	
Precondiciones	Deben estar creadas en la base de datos de la herramienta informática las tablas de las entidades en las que deben ser almacenados los documentos obtenidos.	
Postcondiciones	Los documentos de las encuestas son almacenados satisfactoriamente.	
Requisito Funcional	RF.02	
Flujo de eventos		
Flujo básico Obtener documentos		
	Actor	Sistema
1.	El caso de uso inicia cuando el Trabajador de la herramienta selecciona la opción Obtener documento.	
2.	El Trabajador de la herramienta selecciona los documentos de las encuestas que desea almacenar.	Almacena los documentos seleccionados en la base de datos propia. Cancelar operación. Ver Alternativa 1: "Cancelar operación." Si el usuario no selecciona ningún documento. Ver Alternativa 2: " El usuario no seleccionó documentos"
3.		El caso de uso termina.
Flujos alternos		
Nº 1 Cancelar operación		
	Actor	Sistema
1.	Presiona el botón Cancelar.	La herramienta informática vuelve al estado anterior.
Prototipo no Funcional		
Nº 2 El usuario no seleccionó documentos		

Actor		Sistema
1.	El usuario presiona el botón Aceptar	La herramienta muestra el mensaje: "Seleccione los documentos a almacenar"
Prototipo no Funcional		
Prototipo no Funcional del Flujo básico Obtener documentos		
Relaciones	CU incluidos	
	CU extendidos	
Requisitos no funcionales		Software, Interfaz, Confiabilidad, Usabilidad.

Tabla 19. Descripción del caso de uso Analizar datos

Caso de uso		
CU-3	Analizar datos	
Objetivo	Analizar los datos de los documentos obtenidos para la extracción de la información.	
Actores	Trabajador de la herramienta	
Resumen	El caso de uso inicia cuando el actor presiona el vínculo Analizar datos, el caso de uso termina.	
Complejidad	Media	
Prioridad	Alta	
Precondiciones	Deben existir documentos de encuestas en la herramienta informática.	
Postcondiciones	Se crea el Documento de análisis psicosocial determinado.	
Requisito Funcional	RF.03	
Flujo de eventos		
Flujo básico Analizar datos		
Actor	Sistema	
1.	El caso de uso inicia cuando el Trabajador de la herramienta accede a la opción Analizar datos.	La herramienta informática muestra los documentos de encuestas que serán analizados.
2.	El trabajador de la herramienta presiona el botón Analizar.	La herramienta informática emite el Documento de análisis psicosocial. Brinda la posibilidad de guardar o abrir el Documento análisis psicosocial emitido. Cancelar operación. Ver Alternativa 1: "Cancelar operación." Si al buscar documentos para realizar el análisis la herramienta informática no encuentra ver Alternativa 2: "No existen documentos para analizar"

3.		El caso de uso termina.
Flujos alternos		
Nº 1 Cancelar operación		
	Actor	Sistema
1.	Presiona el botón Cancelar.	La herramienta informática vuelve al estado anterior.
Prototipo no Funcional		
Nº 2 No existen documentos para analizar		
	Actor	Sistema
1.		La herramienta no encuentra documentos de encuestas en su base de datos y muestra un mensaje "No existen documentos para analizar"
Prototipo no Funcional		
Prototipo no Funcional del Flujo básico Analizar datos		
Relaciones	CU incluidos	1. Comparar con Ontología. 2. Almacenar vectores. 3. Emitir documento de análisis.
	CU extendidos	
Requisitos no funcionales		Software, Interfaz, Rendimiento.

Tabla 20. Descripción del caso de uso Comparar con ontología

Caso de uso	
CU-4	Comparar con ontología
Objetivo	Comparar los términos obtenidos del análisis con la ontología.
Actores	-
Resumen	El caso de uso inicia cuando la herramienta concluye la etapa del preprocesamiento de los datos y obtiene los conceptos que serán comparados con la ontología, para ubicar los mismos en cada categoría de evaluación psicosocial, se calcula el vector de pertenencias de los conceptos a cada categoría de un estudiante y luego se determina el vector de frecuencias relativas de cada estudiante, el caso de uso termina.
Complejidad	Alta
Prioridad	Alta
Precondiciones	Debe haber concluido el preprocesamiento de los datos.
Postcondiciones	Se obtiene el vector de frecuencias relativas de cada estudiante y de cada grupo de estudiantes.

Requisito Funcional	RF.04	
Flujo de eventos		
Flujo básico Comparar con ontología		
	Actor	Sistema
1.		La herramienta informática compara los conceptos preprocesados con los términos presentes en cada una de las categorías presentes en la ontología.
2.		Calcula los vectores de pertenencias de cada concepto a cada una de las categorías de la ontología por cada estudiante.
3.		Calcula los vectores de frecuencias relativas de cada estudiante y de cada grupo de estudiantes.
4.		El caso de uso termina.
Flujos alternos		
Nº 1		
	Actor	Sistema
Prototipo no Funcional		
Nº 2		
	Actor	Sistema
Prototipo no Funcional		
Prototipo no Funcional del Flujo básico Comparar con ontología		
Relaciones	CU incluidos	
	CU extendidos	
Requisitos no funcionales	Software, Interfaz, Rendimiento, Usabilidad.	

Tabla 21. Descripción del caso de uso Almacenar vectores

Caso de uso	
CU-5	Almacenar vectores
Objetivo	Almacenar los vectores de frecuencias relativas por cada estudiante y grupo obtenidos en la Base de Datos de la Herramienta.

Actores	-	
Resumen	El caso de uso inicia cuando la herramienta obtiene los vectores de frecuencias relativas de cada estudiante y grupo de estudiante para almacenarlos en la base de datos de la herramienta, el caso de uso termina.	
Complejidad	Alta	
Prioridad	Alta	
Precondiciones	Se deben tener los vectores de frecuencia relativas de cada estudiante y grupo de estudiantes.	
Postcondiciones	Se encuentran almacenados los vectores de frecuencias relativas de cada estudiante y grupo de estudiantes en la base de datos de la herramienta informática.	
Requisito Funcional	RF.05	
Flujo de eventos		
Flujo básico Almacenar vectores		
	Actor	Sistema
1.		La herramienta informática realiza la conexión a la base de datos de la misma para acceder a sus tablas.
2.		Realiza consultas de inserción para insertar los vectores de frecuencias relativas de cada estudiante y grupo de estudiantes en la tabla correspondiente de la base de datos.
3.		El caso de uso termina.
Flujos alternos		
Nº 1		
	Actor	Sistema
Prototipo no Funcional		
Nº 2		
	Actor	Sistema
Prototipo no Funcional		
Prototipo no Funcional del Flujo básico Almacenar vectores		
Relaciones	CU incluidos	
	CU extendidos	
Requisitos no funcionales	Software, Interfaz, Rendimiento, Usabilidad.	

Tabla 22. Descripción del caso de uso Emitir documento de análisis

Caso de uso		
CU-6	Emitir Documento de análisis	
Objetivo	Generar el documento de análisis psicosocial que contiene la información obtenida luego del análisis realizado.	
Actores	-	
Resumen	El caso de uso inicia cuando la herramienta informática obtiene todos los vectores de frecuencias relativas de cada documento analizado en su base de datos, para generar un documento de análisis psicosocial con la información que contiene cada uno de estos vectores de frecuencias relativas por cada estudiante y grupo de estudiantes, es caso de uso termina.	
Complejidad	Media	
Prioridad	Alta	
Precondiciones	Se deben tener los vectores de frecuencia relativas de cada estudiante y grupo de estudiantes almacenados en la base de datos de la herramienta informática.	
Postcondiciones	Se obtiene el Documento de análisis psicosocial por cada estudiante y grupo de estudiantes analizados.	
Requisito Funcional	RF.06	
Flujo de eventos		
Flujo básico Emitir documento de análisis		
	Actor	Sistema
1.		La herramienta informática genera el Documento de análisis psicosocial a partir de la información contenida en la tabla Vector de su base de datos.
2.		Muestra al Trabajador de la herramienta el documento generado con la información obtenida del análisis a través de la interfaz gráfica de usuario.
3.		El caso de uso termina.
Flujos alternos		
Nº 1		
	Actor	Sistema
Prototipo no Funcional		
Nº 2		
	Actor	Sistema

Prototipo no Funcional	
Prototipo no Funcional del Flujo básico Emitir documento de análisis	
Relaciones	CU incluidos
	CU extendidos
Requisitos no funcionales	Software, Interfaz, Rendimiento, Usabilidad.

Tabla 23. Descripción del caso de uso Refinar información

Caso de uso		
CU-7	Refinar información	
Objetivo	Refinar la información obtenida para eliminar la información redundante.	
Actores	Trabajador de la herramienta informática	
Resumen	El caso de uso inicia cuando el Trabajador de la herramienta informática selecciona la opción de Refinar información, para realizar un agrupamiento de la información obtenida empleando el método K means para optimizar la misma eliminando la información redundante, el caso de uso termina.	
Complejidad	Media	
Prioridad	Alta	
Precondiciones	Debe existir un Documento de análisis psicosocial al cual se le realizará el refinamiento.	
Postcondiciones	Se obtiene la información optimizada por grupos de interés comunes.	
Requisito Funcional	RF.07	
Flujo de eventos		
Flujo básico Refinar información		
	Actor	Sistema
1.	El caso de uso inicia cuando el Trabajador de la herramienta informática selecciona la opción Refinar información.	La herramienta informática muestra los Documentos de análisis psicosocial que contienen la información a refinar.
2.	Selecciona los Documentos de análisis psicosocial que serán refinados y presiona el botón Refinar.	Realiza el refinamiento de la información que contienen los Documentos de análisis psicosocial seleccionados. Muestra el resultado del refinamiento a través de la interfaz gráfica de usuario. Cancelar operación. Ver Alternativa 1: "Cancelar operación."

		Si el usuario no selecciona ningún Documento de análisis psicosocial. Ver Alternativa 2: “El usuario no seleccionó Documentos”
3.		El caso de uso termina.
Flujos alternos		
Nº 1 Cancelar operación		
	Actor	Sistema
1.	Presiona el botón Cancelar.	La herramienta informática vuelve al estado anterior.
Prototipo no Funcional		
Nº 2 El usuario no seleccionó Documentos		
	Actor	Sistema
1	El usuario presiona el botón Refinar.	La herramienta informática muestra el mensaje: “Seleccione el Documento de análisis psicosocial a refinar”.
Prototipo no Funcional		
Prototipo no Funcional del Flujo básico Refinar información		
Relaciones	CU incluidos	
	CU extendidos	
Requisitos no funcionales		Software, Interfaz, Rendimiento, Usabilidad.

Tabla 24. Descripción del caso de uso Graficar información

Caso de uso	
CU-8	Graficar información.
Objetivo	Mostrar la información obtenida luego del análisis de los datos en forma de gráfico.
Actores	Trabajador de la herramienta informática
Resumen	El caso de uso inicia cuando el actor presiona el vínculo Graficar información, la herramienta informática grafica la información y le muestra la gráfica, el caso de uso termina.
Complejidad	Media
Prioridad	Alta
Precondiciones	Debe existir el Documento de análisis psicosocial, puede ser luego de refinada su información.
Postcondiciones	Se obtiene el gráfico de la información del Documento de análisis psicosocial.

Requisito Funcional	RF.08	
Flujo de eventos		
Flujo básico Graficar información		
	Actor	Sistema
1.	El caso de uso inicia cuando el Trabajador de la herramienta informática selecciona la opción Graficar información.	La herramienta informática muestra los Documentos de análisis psicosocial que contienen la información a graficar.
2.	Selecciona los Documentos de análisis psicosocial que serán graficados y presiona el botón Graficar.	Realiza el gráfico de la información que contienen los Documentos de análisis psicosocial seleccionados. Muestra el gráfico a través de la interfaz gráfica de usuario. Cancelar operación. Ver Alternativa 1: "Cancelar operación." Si el usuario no selecciona ningún Documento de análisis psicosocial. Ver Alternativa 2: "El usuario no seleccionó Documentos"
3.		El caso de uso termina.
Flujos alternos		
Nº 1 Cancelar operación		
	Actor	Sistema
1.	Presiona el botón Cancelar.	La herramienta informática vuelve al estado anterior.
Prototipo no Funcional		
Nº 2 El usuario no seleccionó Documentos		
	Actor	Sistema
1	El usuario presiona el botón Graficar.	La herramienta informática muestra el mensaje: "Seleccione el Documento de análisis psicosocial a graficar".
Prototipo no Funcional		
Prototipo no Funcional del Flujo básico Graficar información		
Relaciones	CU incluidos	
	CU extendidos	
Requisitos no funcionales	Software, Interfaz, Rendimiento, Usabilidad.	

Anexo 2

Cuestionario a expertos

Estimado experto(a), la presente encuesta forma parte de una investigación pedagógica que está dirigida al perfeccionamiento del proceso docente educativo. Por cuanto estamos convencidos de que sus valoraciones acerca de los asuntos que sometemos a su consideración nos servirán de considerable ayuda, le solicitamos la más responsable atención a esta consulta.

1-Datos generales del encuestado:

Institución y departamento donde labora: _____

Título universitario: _____

Categoría científica: _____ Categoría docente: _____

Años de experiencia en la docencia: _____

Años de experiencia en los estudios de la rama de la psicología: _____

Según resultados obtenidos en diferentes estudios realizados con anterioridad, hemos extraído los siguientes indicadores que someteremos a su consideración en esta consulta para juzgar su efectividad en la medición de las variables de la presente investigación.

Es por ello que el objetivo de la presente encuesta consiste en que usted evalúe cada uno de los indicadores que se le presentarán. Para expresar su evaluación, por favor, luego de analizar cuidadosamente el material que se adjunta, evalúe a cada uno de los indicadores que se le presentan en la tabla de la subsiguiente sección II, marcando con una cruz en la casilla correspondiente y teniendo en cuenta para ello el siguiente código de categorías de clasificación: 5: manifestación **MUY ADECUADO**; 4: manifestación **BASTANTE ADECUADO**; 3: manifestación **ADECUADO**; 2: manifestación **POCO DECUADO**; 1: manifestación **INADECUADO**.

2-Listado de indicadores a valorar:

No.	Indicador	5	4	3	2	1
-----	-----------	---	---	---	---	---

1	Cómo evalúa el concepto de caracterización psicosocial.					
2	Cómo evalúa el empleo de la técnica de los Diez Deseos para realizar esta caracterización.					
3	Cómo evalúa el concepto de los motivos del individuo como objeto conductor de su actividad.					
4	Cómo evalúa el empleo del concepto de los motivos y las necesidades como los elementos de base de la motivación humana, los constituyentes básicos sobre los cuales se conforma todo el proceso de desarrollo del proceso motivacional.					

3-Si desea exponer cualquier otra opinión, por favor, expréselo en el espacio disponible a continuación.

4-Como parte del método de procesamiento de los datos obtenidos por medio de la presente encuesta, necesitamos caracterizar estadísticamente la competencia del conjunto de expertos del cual usted forma parte, por lo que finalmente le rogamos nos ayude respondiendo lo más fielmente posible al siguiente TEST DE AUTOVALORACIÓN DEL CONSULTADO:

a) Evalúe su nivel de dominio acerca de la esfera sobre la cual se le consultó marcando con una cruz sobre la siguiente escala (1: dominio mínimo; 10: dominio máximo)

1	2	3	4	5	6	7	8	9	10	

b) Evalúe la influencia de las siguientes fuentes de argumentación en los criterios valorativos aportados por usted.

Fuentes de argumentación	Grado de influencia de las fuentes de argumentación		
	Alto	Medio	Bajo
Análisis teóricos realizados			
Experiencia obtenida			
Conocimiento de trabajos de autores nacionales			
Su intuición			

Es todo, muchas gracias por su colaboración.

Código _____

Anexo 3

Tabla 25. Cálculo del coeficiente de conocimiento para los expertos

Experto	Escala										Kc
	1	2	3	4	5	6	7	8	9	10	
1										x	1
2									x		0,90
3										x	1

Tabla 26. Matriz de coeficientes de argumentación por expertos

Experto	Escala				Ka
	1	2	3	4	
1	0,20	0,40	0,04	0,04	0,68
2	0,20	0,50	0,04	0,05	0,79
3	0,30	0,50	0,05	0,05	0,90

Tabla 27. Niveles de competencia de los expertos

Experto	Kc	Ka	K	Nivel de competencia
1	1	0.70	0,85	Alto
2	0.90	0.80	0.85	Alto
3	1	0.90	0.95	Alto