



UNIVERSIDAD NACIONAL DE LA MATANZA

ESCUELA DE POSGRADO

MAESTRIA EN INFORMÁTICA

TESIS DE MAESTRÍA

**Título: "Desarrollo de un moderador automático
para salas de chat"**

Autor: Sergio Rafael Flores

Director: Dra. Silvana Vanesa Aciar

Codirector: Dra. Silvia Margarita Baldiris

Buenos Aires, 14 de noviembre de 2014

INDICE

	PAGINA
CAPITULO I – INTRODUCCION	5
CAPITULO II - Estado de la Cuestión	13
CAPITULO III - Planteamiento del Problema	45
CAPITULO IV – Solución	53
CAPITULO V – Validación	79
CAPITULO VI - Conclusiones y Futuros trabajos	101
7. CAPITULO VII – Bibliografía	107
8. ANEXOS	113

CAPITULO I: INTRODUCCION

Descripción del problema

El signo actual de nuestra sociedad es la irrupción de las Tecnologías de la Información y la Comunicación (TICs), que han alterado las formas habituales de relación social. Somos testigos del impacto político, económico y social que estas nuevas tecnologías están teniendo en nuestras sociedades a nivel mundial, dado su incorporación en los medios de comunicación en distintos ámbitos.

En Argentina, el ámbito educativo es uno de los espacios donde las nuevas tecnologías han incursionado favorecido por diferentes factores: por un lado, los estudiantes las han adoptado, y por otro, la presencia de políticas sociales que han favorecido el acceso de la población estudiantil, en especial en el ámbito de la educación media.

La irrupción de internet ha modificado las formas de comunicación y organización, ya que como sostiene Castells (1999, p.505), se ha convertido en un medio en el cual se basa la nueva sociedad en que vivimos, la "sociedad en red". En ese contexto, un desafío es cómo utilizar a internet en la educación a distancia o semi-presencial, y en especial, de que forma se puede lograr que su uso sea fructífero.

La posibilidad de aprovechar estos cambios depende de la penetración que internet tiene en los países. Argentina tenía para 2009 el 48,9% de la población conectada. En cambio Chile alcanzaba el 50,4% y Brasil con un 34,0 % de población conectada. Estos valores son bajos comparados con países desarrollados, como Noruega (90,9%) ó EU (74,1%) (Sunkel, Trujo y Espejo, 2014, p.45).

El aprovechamiento de internet en educación puede favorecer la comunicación entre docentes y estudiante y la alfabetización de estudiantes a distancia, en especial por ser un medio atractivo y conocido por los jóvenes.

Lo anterior supone la necesidad de desarrollar herramientas tecnológicas que puedan ser incorporadas a un ambiente de aprendizaje virtual que tenga en cuenta la necesidad de adaptación y personalización de los contenidos de manera que se ajusten a las preferencias y necesidades de todos los usuarios, incluidos profesores y alumnos (Rodríguez y otros, 2011).

Entre las formas más manejadas de comunicación dentro del uso de internet se encuentra el chat. En él los usuarios se envían mensajes y discuten sobre algún tema o tópico, o

simplemente conversan entre sí, utilizando el discurso escrito como representación del lenguaje con el cual se comunican.

En el chat la comunicación puede darse en un ambiente flexible, sin ningún tipo de insinuación previa de lo que se puede o se debería charlar, o en ambientes más acotados. En este último ambiente, los usuarios saben de antemano a que tema se deberían remitir o la hora a la cual deben enviar un mensaje, tal como sucede en plataformas de educación a distancia, o hasta en la comunicación interna de una empresa.

Cuando el ambiente donde se puede dar una comunicación es de tipo flexible, no resulta necesario aplicar un control sobre lo que los usuarios conversan, ya que justamente se está libre de cualquier tipo de limitaciones. En cambio si el ambiente es más más fijado, es de interés y necesario aplicar un cierto control a la conversación que se lleva a cabo.

Conforme una conversación se desarrolla en el tiempo es más fácil de desviarse del tema o tópico que se está discutiendo y emplear más tiempo en las discusiones sobre temas que no son tan relevantes para el objetivo de la discusión.

Es por ello que resulta útil aplicar a esta comunicación un observador al que se denominará “Moderador”, que tenga la capacidad de discernir si lo que conversan los usuarios es acorde para los propósitos de la herramienta utilizada, por ejemplo, un profesor puede actuar como Moderador en una sala de chat de alumnos en educación a distancia, analizando si lo que estos conversan es correcto para la materia, e interviniendo cuando le parezca necesario, para que los estudiantes no se desvíen de la temática planteada.

Es en este contexto donde surge la necesidad de investigar y desarrollar una herramienta moderadora de una conversación de una temática dada en un ambiente de aprendizaje virtual, que contextualice y canalice las interacciones de alumnos y docentes en un medio de interacción colaborativo de enseñanza-aprendizaje como lo es el chat.

Atendiendo a que en español no existe un software que modere en forma automática la conversación en ámbitos educativos, el **motivo de este trabajo** es proponer una solución a este problema mediante el desarrollo de un algoritmo que sirva de base para la creación de una herramienta moderadora de una conversación en una sala de chat, que “intercepte” la conversación establecida y detecte si está de acuerdo con la temática propuesta por el sitio, en caso contrario, el sistema computacional debería intervenir moderando la comunicación,

sugiriendo a los participantes que se trate la temática propuesta, de manera automática. Para ello, es preciso que el sistema (computadora) “comprenda” la conversación en español.

En un entorno de aprendizaje virtual, uno de los componentes más importantes es la comunicación e interacción en tiempo real mediante las herramientas que brinda el entorno virtual.

Según Adell (1998, p.13) la comunicación en tiempo real “implica que los participantes deben atender el canal de comunicación, coincidiendo en el tiempo”. En la actualidad se utilizan sistemas, cuyas diferencias están dadas en la codificación de la información: pueden utilizar el IRC (Internet Relay Chat) o chat. En este caso, la comunicación se establece sólo en modo texto.

Adell (1998) agrega que en el chat, los participantes pueden intercambiar fragmentos de texto en tiempo real: lo que escribe uno de ellos es enviado a la pantalla de la computadora de todos los demás. Los servidores de IRC están interconectados, formando redes, de tal manera que un usuario conectado a un servidor local puede “charlar” o “chatear” con otro usuario, conectado a su servidor local, si ambos están sintonizando el mismo canal.

El chat es una herramienta ampliamente utilizada por los alumnos y docentes para intercambiar conocimiento, experiencias y construir colaborativamente el aprendizaje.

El aprendizaje colaborativo (AC) hace referencia a la actividad de pequeños grupos desarrollada en el salón de clase o en un aula virtual, así lo expone (Johnson, 1999, p. 14).

Es preciso destacar que el AC es más que el simple trabajo en equipo por parte de los estudiantes. La idea que lo sustenta es sencilla: los alumnos forman "pequeños equipos" después de haber recibido las indicaciones del profesor. Dentro de cada equipo los estudiantes intercambian información y trabajan en una tarea hasta que todos sus miembros la han entendido y terminado, aprendiendo a través de la colaboración.

El IRC, es una clase de interacción síncrona. Según Blázquez (2001), es una interacción con coincidencia temporal, volviéndose la relación entre el hombre y la máquina más interactiva, así lo señala donde los estudiantes y/o profesores discuten o debaten sobre un tema particular.

Una dificultad que se presenta en muchas oportunidades, es que la conversación se desvía y los participantes de la misma comentan sobre otros temas que no coinciden al originalmente planteado y a resolver.

Esta dificultad conduce a que, en las sesiones de chat que se llevan actualmente se requiera de la intervención de un moderador humano, que generalmente suele ser el mismo docente o responsable del curso o tutores. Su función es controlar que las mismas se lleven a cabo dentro de la temática propuesta.

El uso generalizado del chat, como herramienta de comunicación en entornos virtuales de aprendizaje, requiere de moderadores humanos para cada sesión de chat establecida, a fin de evitar el desvío del tema tratado. Ello es algo casi imposible.

Por lo expuesto, se plantea la necesidad de contar con un moderador automático que intervenga en la sesión de chat entre participantes humanos, con el objeto de mantener los parámetros conversacionales en el entorno virtual educativo y sobre la temática planteada, a fin de alcanzar con éxito los propósitos de la comunicación en contextos educativos.

En el idioma español, actualmente no existe un moderador automático que intervenga en las sesiones de chat en entornos educativos. **Por ello, el trabajo de tesis se encuadra dentro del Área Inteligencia Artificial, en el campo referido al Procesamiento del Lenguaje Natural, de modo de permitir que el moderador automático pueda intervenir satisfactoriamente en la conversación de una temática dada en el chat de los participantes.**

Coincidiendo con lo que sostiene Hernández (2001), el chat ofrece un campo novedoso y fecundo para la investigación pedagógica, en tanto se indague en las posibilidades de su aprovechamiento óptimo, con el fin de mejorar la práctica didáctico-pedagógica. Además, investigaciones en esta área colaborarán en profundizar el conocimiento teórico y práctico sobre la naturaleza y el impacto de esta tecnología en la comunicación y la educación.

Importancia del problema

La importancia de la tesis radica en dos aspectos:

1. Está enmarcada en un proyecto internacional financiado por el Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED) denominado “Red Iberoamericana de Apoyo a los Procesos de Enseñanza Aprendizaje de competencias profesionales a través de entornos ubicuos y colaborativos” (Código 512RT0481), del cual participan diversas universidades de América y Europa: Universidad del Cauca, Colombia. Universidad Castilla-la Mancha, España. Universidad Tecnológica de Panamá. Universidad Nacional de Costa Rica.

Universidad Federal do Rio Grande do Sul, Brasil. Universidad Nacional de Colombia. Universidad Nacional de San Juan. Argentina. Universidad de Lleida, España. Pontificia Universidad Católica de Valparaíso, Chile. Universidad de Granada, España. Universidad de la Laguna, España.

También la tesis se vincula con el Proyecto de investigación “Desarrollo de Herramientas Tecnológicas de Soporte a la Educación Virtual”, del Instituto de Informática, Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan. Proyecto aprobado por el CICITCA en la convocatoria 2014.

2. Este trabajo analizará y desarrollará la implementación de un sistema de chat propio, con el propósito de poder incorporarle las capacidades de moderador. Esto se realiza para que esas capacidades puedan ser incorporadas a otros sistemas de chat a fin de poder controlar las conversaciones dadas en este contexto y determinar si son acordes o no a una temática pre-establecida y por la que se lleva a cabo la sesión de chat. El moderador puede intervenir automáticamente cuando detecte una desviación en la conversación de los participantes, que considere importante. La forma de esta intervención estará dada por el envío de un mensaje a todos los participantes, solicitando que no se desvíen del tema planteado. Todo esto de forma totalmente automática y valiéndose de la comprensión previa de los tópicos sobre los que se establece la temática para la conversación.

Objetivos

Objetivos Generales

1. Diseñar e implementar, en español, una arquitectura que posibilite moderar automáticamente la interacción en el chat, contextualizando el discurso escrito a una temática especificada.
2. Estimar el nivel de aceptación del sistema diseñado entre usuarios en el ámbito educativo de la Provincia de San Juan

Objetivos Específicos

- Analizar la problemática de la interacción y moderación de las conversaciones en un chat.
- Analizar el aporte del análisis del discurso en la construcción de la solución.

- Analizar en el Procesamiento del Lenguaje Natural las técnicas para la adquisición de información del texto en el discurso.
- Indagar en los procedimientos dentro de la minería de datos las técnicas para poder realizar text mining en el discurso.
- Proponer una solución en base al análisis realizado de la problemática y de las soluciones existentes en la contextualización del discurso escrito en un chat.
- Analizar el nivel de aceptación del sistema diseñado, dentro del nivel medio y universitario, a través de una encuesta aplicada a una muestra intencional de estudiantes y docentes.

Breve esbozo de la solución

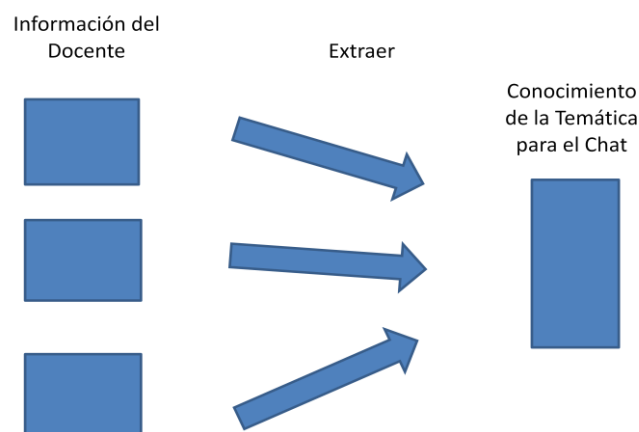
El estudio de la tecnología de Chat permite conocer el funcionamiento del mismo, en el cual se desarrolla una conversación, desde el punto de vista técnico se considera un discurso, es decir, el manifiesto de una persona en documento escrito en un entorno de chat, en el que se encuentra información.

La información contenida en el discurso debe ser analizada, para entender el discurso en lo que a estructura se refiere debe ser analizado, el proceso de análisis de las distintas partes del discurso se basa en lo que se conoce como minería de texto, que es una parte de una tecnología más general que es la minería de datos.

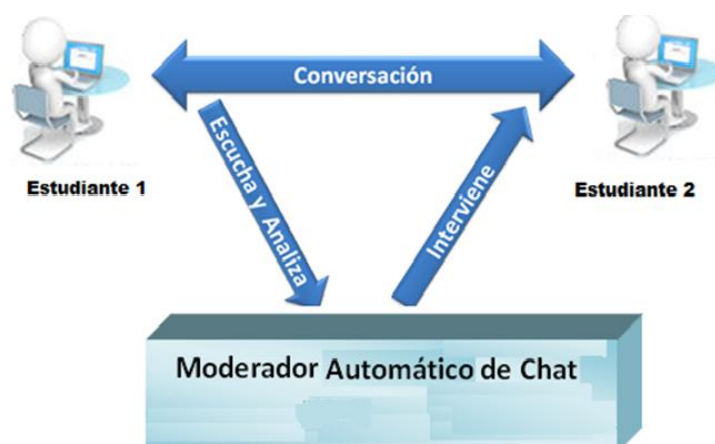
- Se analizó información desde el punto de vista de la teoría de la información para entender como se pone de manifiesto la cantidad de información en el discurso.
- Se estudió el análisis documental para ver la estructura del discurso y poder inferir que partes del mismo de deben estudiar y resguardar para generar nueva información.
- Se estudió el proceso de minería de datos y minería de texto para poder lograr nuevo conocimiento y poder actuar en consecuencia.
- Se estudio los lenguajes de programación para poder realizar las aplicaciones necesarias para modelar las distintas situaciones.

Gráficamente se pueden distinguir dos instancias:

1. Una previa al desarrollo del chat, donde el experto de la temática, en cada caso será el docente que desea realizar la sesión de chat moderado, que posee el conocimiento de la temática a discutir, debe proveer la información necesaria para el sistema.



2. Y la otra cuando se desarrolla la sesión de chat, donde los participantes conversan y el sistema vigila esas conversaciones y cuando detecta que hay desviaciones de la temática interviene con mensajes para que los participantes retomem el tema tratado.



CAPITULO II: Estado de la Cuestión

Atendiendo a que en la revisión bibliográfica no se encontraron desarrollos teóricos que trataran de modo puntual el problema del método que modere un discurso escrito en español (conversación), se debió recurrir a la tarea de ubicar la información en el discurso, el análisis del discurso, y el aporte desde la minería de datos como procedimiento de descubrimiento de nuevo conocimiento.

A continuación, se presentarán dichos aportes que contribuyeron a responder ciertos interrogantes que motivaron el estudio de lo que sucede en un chat dentro de una conversación: ¿Hay intercambio de información? ¿Hay flujo de información?, ¿Cuál es la cantidad de información que fluye? ¿En un texto donde está la información que hace referencia a los objetos de que se está hablando? ¿Cómo se puede obtener nuevos conocimiento a partir datos en un discurso? ¿Existen procesos para el análisis estructural de un discurso?

Nociones Básicas de Información

Según Cuevas (1981, p.15) la información, en el sentido ordinario de la palabra, es la transmisión a un ser consciente de una idea o una significación, por medio de un mensaje más o menos convencional, y por un soporte espacio-temporal: impresos, mensaje telefónico, etc.

La comprensión del mensaje es el objeto de la información. La comunicación del soporte, es decir el medio, es por donde viajará la información.

El mensaje conforma un discurso. Este término, discurso, proviene del latín *discursus*, que significa todo “escrito o tratado de no mucha extensión, en que se discurre sobre una materia para enseñar o persuadir” (RAE, 2014).

En un chat cada escrito que envían los participantes se genera un discurso, que tiene un contenido de información, y que a los demás participantes les provoca un estímulo.

Resulta razonable la relación entre información y conocimiento. Cuando se escucha un noticiero informativo es para aprender algo nuevo; si el periódico me enseña nada, puedo

decir que está desprovisto de interés. Admitimos, pues, el nexo entre *información* y *aportación de conocimiento*. Sin embargo, no se pueden identificar totalmente estos dos términos. Es evidente que la informática necesita conceptos objetivos y la noción de ganancia de conocimiento es fundamentalmente subjetiva.

Se puede precisar definición de información diciendo que es necesario hablar de *aportación eventual de conocimiento*. Pero esto no es todo. Todas las fuentes de información tienen una característica en común: se apoyan en el lenguaje hablado o escrito y siempre podemos escribir la información oral. Por tanto, damos el nombre de *información* sólo a aquellos datos que *se pueden escribir* (Cuevas, 1981).

Representación de la Información

La información en un discurso es una fórmula escrita susceptible de aportar un conocimiento, por lo tanto la información se puede representar mediante los lenguajes usuales, en particular el lenguaje español.

Así se puede representar información: numérica representada por los símbolos numéricos, alfabética ésta es representada por palabras formada por concatenaciones del alfabeto y que tienen un significado particular, para determinar el significado de las palabras, se toma como base el diccionario de la RAE, además hay otras representaciones como por ejemplo el grupo sanguíneo de las personas que con una codificación se puede establecer tipo y factor de la misma.

En este caso, las referencias simbólicas se eligieron por su valor mnemotécnico, al menos para los especialistas. Esta forma de representación se usa también para las informaciones compuestas.

La representación de la información mediante los lenguajes usuales es una forma de codificación por ejemplo: la ubicación de un domicilio permite designar una persona, una calle, una ciudad, un país.

Según Cuevas (1981) El enunciado de una palabra nos indica una imagen debido a que el aprendizaje de un lenguaje nos ha habituado a una asociación de este tipo.

Todo texto se puede considerar desde dos puntos de vista:

- Como una serie de caracteres que permiten reconocer este texto e identificarlo entre todos aquellos que se pueden escribir de la misma longitud y con el mismo alfabeto.
- Como un conjunto de frases que llevan un significado y que evocan imágenes más o menos precisas.

De estos dos aspectos del texto, tomamos el segundo para realizar el análisis. Así, en un diccionario se encuentran, clasificadas en orden alfabético las palabras de un lenguaje consideradas como referencias o cadenas de caracteres; a éstas van asociados textos formados por palabras de la misma lengua o de otro lenguaje que procesan las imágenes asociadas a las palabras de referencia. La palabra definida está tomada en su aspecto de referencia y las palabras de la definición con su valor semántico (significado), por ejemplo un diccionario.

Información y Conocimiento

Ante la cuestión capital de saber si la información es el conocimiento que se puede sacar eventualmente de una fórmula escrita o si es solamente esta fórmula considerada como una serie de caracteres, no hay otra actitud posible que una elección «a priori».

Cantidad de Información

Ciertas obras ofrecidas al gran público hablan de la información insistiendo no tanto en su naturaleza como en la medida de la cantidad de información contenida en un mensaje.

A menudo, se parte del hecho de que el texto utilizado para dar información es redundante y que se puede expresar casi siempre con nuevas palabras (un ejemplo típico son los telegramas).

Por consiguiente, el número de caracteres de un mensaje no es una medida de la cantidad de información que contiene.

Tampoco es posible intentar esta medida al nivel de las diferentes representaciones o codificaciones.

Lo que cuenta, en realidad, es el número de valores que puede tomar el fenómeno variable. *Se da más información diciendo la ciudad en que ha nacido un individuo que diciendo el país (hay muchas más ciudades que países en el mundo).*

Por otra parte, si bien está claro que no hay información a propósito de una cosa cierta (invariante), también es cierto que hay poca información a propósito de una cosa casi segura.

La cantidad de información está ligada, pues, no solamente al número de estados posibles, sino también a la posibilidad de que ocurra cada uno de ellos. C. Shannon ha propuesto una fórmula matemática que da la cantidad de información en función del número de estados y de sus probabilidades respectivas.

Desgraciadamente, en general, se ignora la probabilidad real de cada estado y, en muchos casos, es muy difícil diseñar la lista exacta de estados posibles.

Cuando se simplifica el problema, admitiendo que todos los estados son equi-probables, la cantidad de información no depende más que del número de estados posibles del fenómeno. Para acabar de *determinar esta medida, la unidad de información es, por definición, la que se obtiene designando uno de dos estados equi-probables.*

Medida de la cantidad de la información: Entropía

El primer ensayo para definir una medida de la información se remonta a 1927. En esta época, el americano R. Y. Hartley, en el Congreso Internacional de Telefonía y Telegrafía, demostró cómo era posible comparar las capacidades de los diferentes sistemas transmisores de información.

Por aspecto cuantitativo de una información se debe entender, en la teoría de la información, su estructura estadística. Es decir, admitiendo que la forma estándar en que se presenta una información es una sucesión de símbolos, lo único que interesa es la frecuencia con que ellos o sus combinaciones puedan aparecer.

En el caso de una lengua, se trata de determinar la frecuencia de las letras, de las palabras, etcétera. El contenido informacional de estas partes o de sus combinaciones, viene considerado en relación a la frecuencia con que aparecen en un texto. Como veremos, hay una estrecha relación entre el grado de novedad (adquisición de conocimiento) de la información y este aspecto estadístico.

El Tratamiento de la Información y sus Niveles

Se llama «tratamiento de conocimientos» a la actividad intelectual fundamental del hombre, mediante la cual confronta los conocimientos que posee, los reúne en teorías coherentes, los organiza en estructuras generales y los simplifica mediante mecanismos de deducción lógica.

El Lenguaje

El concepto de lenguaje tiene diferentes acepciones para cada uno de los especialistas (en lingüística, psicología, informática, etc.). En todo acto de comunicación se encuentran implicados un organismo emisor y uno o varios organismos receptores a los cuáles transmite algo el emisor. En este algo podemos diferenciar tres aspectos:

- 1) El soporte físico (materia).
- 2) La disposición formal o lógica, abstraída en lo posible del soporte físico, es decir, los signos o significantes.
- 3) Las ideas, objetos o acciones que pueden asociarse con la disposición formal, es decir el significado.

El lenguaje considerado como vehículo del pensamiento

El lenguaje es la expresión del pensamiento y de los sentimientos por medio de sonidos naturales o por símbolos gráficos. Es el principal medio utilizado para el intercambio de información.

Se nos presenta la duplicidad: significado y significante. Significante es la palabra definida como un conjunto de sonidos y símbolos gráficos que representa de un modo más o menos arbitrario algo inmaterial. Este algo inmaterial es el significado, que hace de la propia palabra la representación de un concepto.

Lingüística Matemática Determinista

Introducción al análisis estructural

Como su nombre indica, esta rama de la lingüística cuantitativa se propone estudiar las leyes generales que gobiernan los fenómenos de la lengua recurriendo a modelos matemáticos deterministas.

El instrumento matemático a que se recurre más a menudo en este planteamiento es la lógica matemática. A pesar de ser de reciente formación, los resultados alcanzados por la lingüística matemática determinística son más que satisfactorios. Basta pensar en la creación de los lenguajes informáticos con los que el hombre se comunica con las máquinas cibernéticas, especialmente con las computadoras, en la adquisición de nuevos elementos lógicos para el estudio de gramáticas formales, etcétera.

Contribuciones importantes en este campo de investigación fueron las de Chomsky, Turing, Barhillel, Kulagina, Marcus y otros.

Conviene precisar que todas estas investigaciones vienen estimuladas por el continuo progreso del área de la informática, que, como se sabe, caracterizan una nueva era en la civilización humana.

El análisis estructural y proyectividad tienen gran importancia en la construcción y traducción de lenguajes.

El *análisis estructural* refiere al orden en que se encuentran las palabras, en un discurso hablado o un texto escrito, es un orden lineal, esto es, un orden en que cada palabra tiene su predecesora (salvo la primera) y su sucesora (salvo la última).

Desde el punto de vista integral, estas relaciones (de vecindad lineal) dicen sin embargo bien poco. Un análisis más cuidadoso lleva a considerar otras relaciones más significativas desde el punto de vista arriba indicado.

Análisis Documental

El valor de la información y la incorporación de tecnologías para su procesamiento, hacen necesario ahondar en el análisis documental como proceso mediador, para ello es necesario profundizar en las dimensiones lingüística y psicológica del mencionado proceso, a fin de

explicar la dimensión lingüística del proceso y su relación con el análisis del discurso; tomando para ello el modelo propuesto por T. van Dijk., explicando los procesos mentales involucrados en la recepción y emisión de información. (Peña Vera y Pirella Morillo, 2007, p.58)

Para ello, Peña Vera y Pirella Morillo (2007, p.60) realizaron una investigación documental con un enfoque racional para integrar perspectivas teóricas de la Lingüística, Psicología, Psicolingüística y Documentación, e identificar conceptos vinculados con las labores de análisis documental. Los autores concluyen que algunos conceptos lingüísticos permiten el entendimiento y desmontaje de las estructuras gramaticales y, por ende, inciden en la aplicación del análisis documental. La metodología existente para analizar discursos constituye una valiosa herramienta teórico-metodológica para el análisis documental aplicado en las unidades de información; algunos conceptos psicológicos permiten entender los procesos mentales de los que reciben y emiten información, por consiguiente, su consideración es necesaria en los procesos analítico-sintéticos porque se recibe y crea nueva información.

El tema del análisis documental es complejo en esencia, por cuanto posee aristas de tipo lingüístico, psicológico-cognitivo, documental, social e informacional propiamente dicho. En él se encuentran involucrados los componentes de la tríada documento – sujeto – procesos, los cuales se afectan mutuamente y generan un entramado de relaciones en las que intervienen las características y particularidades de cada uno de ellos. De esta imbricación proviene, en parte, su complejidad ya que es un proceso plural que envuelve una serie de aspectos y elementos propios de otras disciplinas y ciencias; pero que irreductiblemente explica parte de las actividades y subprocesos activados cada vez que se ejercen los roles de analista y usuario de una información contenida en cualquier documento (Peña Vera y Pirella Morillo 2007, p.56)

A pesar de su complejidad innata, el valor del análisis documental se consolida cada vez con mayor fuerza debido a la avalancha informativa a la que se asiste en el presente tiempo. Esta no es comparable con los volúmenes documentales que se manejaban en décadas pasadas. Todo esto aunado al uso masivo de las tecnologías de información y comunicación (TIC) que han hecho posible el establecimiento de redes para el intercambio de información y, en general, han dado un vuelco a las posibilidades de procesamiento en las unidades y servicios de información; aliviando gran parte del peso que representa manejar grandes volúmenes de documentos (Peña Vera y Pirella Morillo 2007, p.57).

Sin embargo, los procesos medulares del tratamiento documental al que se someten las diversas fuentes y recursos de información siguen dependiendo de la capacidad y método aplicado por los analistas de información. Así lo corrobora Maniez (1993) (en Peña Vera y Pirella Morillo 2007, p.57) al señalar que detrás de cada discurso contenido en alguna fuente documental subyace una información solapada que amerita, para ser descubierta, de la capacidad intelectual y perspicacia de quien analiza la información. Por consiguiente, es necesario contar con metodologías que hagan posible conocer lo esencial del caudal de documentos que se generan en diversos formatos y en cada área de trabajo o del conocimiento; de manera que el hombre pueda aprehender lo que le haga falta en virtud de su formación, desempeño, inclinación natural, actividad de ocio a la que se dedique, entre otros. Todo ello con la finalidad de generar un impacto positivo en su aprendizaje y contribuir a su formación integral y, en consecuencia, a su adecuada inserción y desempeño social, académico, laboral.

Por otra parte, el tras tocamiento de los conceptos acabados y las formas rígidas, producto de la complejidad creciente, ha incidido también en los esquemas de procesamiento documental que de otrora se vienen aplicando. No solo es importante considerar los aspectos formales y los de contenido, para los cuales existe toda una plataforma de estándares de uso internacional, sino que además es necesario ahondar en las motivaciones e inclinaciones de los autores de las informaciones, sus formas de percibir la realidad, sus inclinaciones afectivas, ideológicas, cognitivas y con ellas se desprende información implícita que también afecta al contenido informativo de cualquier documento, y que puede resultar de interés en el momento de consultar alguna fuente documental. A esto se suma el surgimiento del paradigma de la complejidad que, según Wagensberg (1989) (en Peña Vera y Pirella Morillo 2007, p.57), propone entre otras cosas, que el hacer científico basado en la complejidad ha de estar en constante crecimiento, huir de la rotación trivial y, en cambio, manifestar un espiral en el que no existe punto de llegada ni de cierre, en el que nada queda acabado; lo cual constituye el estimulante principal para la generación de conocimiento. Esto hace necesario la ampliación del espectro teórico que se viene manejando, así como una mayor profundización en los aspectos subjetivos que ineludiblemente inciden en la producción de información y en el procesamiento de la misma. No es posible pues concebir el análisis documental como un proceso unidisciplinario. Ello limita su alcance y soslaya las variadas implicaciones que se derivan de su ejecución, sumergiendo su trascendencia en una simplista visión de procesos rutinarios que ignoran los múltiples factores intervinientes.

Peña Vera y Pirella Morillo (2007, p.58) profundizan en la interdisciplinariedad del análisis documental, y enfatizan en las dimensiones lingüística y psicológica de dicho proceso y, como objetivos específicos, en primer lugar, profundizar y puntualizar en sus aspectos característicos. Buscan explicar la dimensión lingüística y su relación con el análisis del discurso, tomando como base el modelo propuesto por T. van Dijk, y los procesos mentales involucrados en la recepción y emisión de información. Usan como vía para el alcance de estos propósitos la investigación documental de tipo racional, con la que integran perspectivas teóricas de la Lingüística, Psicología, Psicolingüística y Documentación; a la vez que identifican conceptos vinculados con las labores de análisis documental.

Definiciones de análisis documental: revisión de algunas posturas teóricas

La definición de análisis documental ha sido expuesta por diferentes investigadora y estudiosa del tema, exhibiendo distintos matices y aspectos medulares que han prevalecido a lo largo de algunas décadas. Este proceso se instaura a partir de la necesidad de facilitar el acceso de los individuos a las fuentes de información, las cuales han ido incrementando su volumen de producción, Vickery (1970, p.154) señaló que los métodos de recuperación, entre los que se cuenta el análisis documental, responden a tres necesidades informativas de los usuarios:

1. conocer lo que otros pares científicos han hecho o están realizando en un campo específico.
2. conocer segmentos específicos de información de algún documento en particular.
3. conocer la totalidad de información relevante que exista sobre un tema específico.

Con el transcurrir del tiempo, el proceso de análisis documental ha evolucionado partiendo del uso de esquemas de trabajos básicos y rudimentarios hasta el perfeccionamiento de técnicas y la incorporación de una visión más interdisciplinaria.

Mijáilov y Guiliarevskii (1974, p.18) señalaron que, luego de la recopilación de la información, era necesario iniciar un procesamiento analítico-sintético de los datos contenidos en un documento y el resultado de esta operación debía ser resumido brevemente o sintetizado. Pinto Molina (1992, p.89) añade que el análisis documental es «el complejo de operaciones que afectan al contenido y a la forma de los documentos originales, para transformarlos en otros documentos representativos de aquellos, que facilitan al usuario su

identificación precisa, su recuperación y su difusión». Por su parte, Solís Hernández (2003) lo define como la operación que consiste en seleccionar ideas informativamente relevantes de un documento, a fin de expresar su contenido sin ambigüedades para recuperar la información contenida en él. Obviamente que los propósitos del análisis documental trascienden la mera recuperación/difusión de la información. Ellos también se orientan a facilitar la cognición y/o aprendizaje del individuo para que este se encuentre en condiciones de resolver problemas y tomar decisiones en sus diversos ámbitos de acción.

Fox (2005, p.21) utiliza la expresión «tratamiento documental», dentro de la cual incluye al análisis que se aplica a los documentos cuando entran a formar parte de un fondo documental, y cuya finalidad es elaborar una nueva representación de su contenido para facilitar su consulta. La autora desglosa la definición de análisis documental partiendo de la denotación del término análisis, y concluye que es el examen, desglose, reconocimiento y estudio que se aplica a los documentos, considerando en ello sus elementos objetivos, tales como el tipo, autor, título, número de páginas, editorial, entre otros; y su significado general o la esencia del mensaje que contiene, el cual se constituye en elemento esencial para su recuperación intelectual. Perelló (1998) señala que el análisis documental comprende dos fases: una que consiste en la determinación del significado general del documento y la consiguiente transformación de la información contenida en él; y la otra que corresponde tanto a la descripción formal o exterior como a la elaboración de estrategias y métodos de búsqueda. El análisis documental se caracteriza por ser dinámico en el entendido que permite representar el contenido de un documento en una forma distinta a la original, generándose así un nuevo documento.

De las definiciones citadas es posible puntualizar que el análisis documental constituye un proceso ideado por el individuo como medio para organizar y representar el conocimiento registrado en los documentos, cuyo índice de producción excede sus posibilidades de lectura y captura. La acción de este proceso se centra en el análisis y síntesis de los datos plasmados en dichos soportes mediante la aplicación de lineamientos o normativas de tipo lingüístico; a través de las cuales se extrae el contenido sustantivo que puede corresponder a un término concreto o a conjuntos de ellos tomados aisladamente, o reunidos en construcciones discursivas. Por consiguiente, su finalidad es facilitar la aproximación cognitiva del sujeto al contenido de las fuentes de información.

En tal sentido, el análisis documental es un proceso en el que se encuentran involucradas las estructuras mentales de los individuos, tanto de los que analizan textos para otros, como de los usuarios finales de los textos analizados. T. van Dijk (1995) explica que durante este proceso entran en juego las estructuras discursivas y los significados por un lado, y las representaciones mentales como son el conocimiento general y los modelos específicos subjetivos, por el otro. Además de la fuerte y marcada influencia que ejerce el contexto y las convicciones ideológicas de quien elabora un discurso. Foucault (2003) añade que el análisis de las ideas, pensamientos y saberes que contienen los documentos ha de verse como el trabajo y la realización de una materialidad documental que presenta en cada sociedad formas específicas y organizadas; es decir, que está sujeta a transformaciones que se generan como resultado inevitable de la evolución de la historia del pensamiento y de los conocimientos, la cual, en sus diferentes épocas, da fe de la multiplicidad de rupturas y discontinuidades.

El estudio del análisis documental y de las implicaciones discursivas que subyacen en él, constituyen un asunto en el que confluyen lo psicológico, lo cognitivo, lo contextual, lo ideológico. De allí su complejidad, creándose así un entramado por naturaleza imbricado, que obliga a incursionar con una visión reticular del proceso. Esto lo corrobora T. van Dijk (1980) al señalar que cuando se analiza un discurso cualquiera, nos introducimos en una esfera interdisciplinaria, por cuanto le atañe a ciertas áreas de las Humanidades y de las Ciencias Sociales como la Psicología, la Sociología, la Comunicación de masas, el Derecho, la Teología, la Psiquiatría, la Antropología, la Filosofía, etc.

Todo ello en virtud del hecho de que el discurso determina y a la vez involucra funciones pragmáticas, cognoscitivas y sociales. También Foucault (2003) afirma que el análisis del discurso amerita el uso de un corpus de enunciados o de una colección de hechos que sirvan como muestra para derivar reglas de construcción de otros enunciados, que no deben pretender ser exhaustivas, pues ello entraría en contradicción con la naturaleza del pensamiento humano que es insospechada. La intención subyacente en T. van Dijk y, en parte, en Foucault se corresponde con el paradigma de la complejidad, según el cual el estudio de los fenómenos se hace desde una óptica plural, sobre todo desde aquellos fenómenos que atañen directamente a la acción del hombre. Al respecto Maass (2005) señala que con dicho paradigma los fenómenos se observan transdisciplinariamente, por cuanto surge una racionalidad en la que es necesario relacionar unos elementos con otros, y así tejer las tramas de relaciones que forman parte de la nueva dinámica, creatividad y reflexión que propone la complejidad. En este caso, se busca configurar un enfoque inter y transdisciplinar que permita

observar el análisis documental inspirado o formando parte de otros procesos más generales referidos a la esfera comunicativa del individuo.

Dimensión lingüística del análisis documental y sus puntos de encuentro con el análisis del discurso

Del hombre, como ser pensante y único posible autor de constructos racionales, es de quien provienen los diferentes contenidos discursivos, por lo cual es necesario considerar, a la par de las técnicas y/o procedimientos de análisis, los elementos de tipo lingüístico y psicológico que subyacen en cada discurso, bien sea de índole oral o escrita. Los discursos son entendidos, según Renkema (1999), como secuencias de oraciones o emisiones habladas conectadas entre sí, por medio de las cuales un emisor comunica un mensaje a un receptor.

Desde una perspectiva más antropocéntrica, es el espacio de exterioridad para el individuo, por cuanto le permite poner de manifiesto y fuera de sí elementos de su pensamiento, conocimiento, creatividad, en esencia, su cosmovisión.

Para este autor, el discurso actúa como símbolo, en el entendido de que se refiere a la realidad; es síntoma, porque su emisor trata de expresar algo a través de él, nunca se produce un discurso sin objetivo. Y es señal, porque los receptores deben hacer algo con él, es decir, el discurso siempre surte efectos.

El discurso es, pues, inherente a la existencia misma del hombre, puede decirse que constituye su instrumento de interacción por excelencia y, aunque su construcción atiende a normativas lingüísticas en primera instancia, también se manifiestan en su contenido importantes elementos psicológicos que expresan la interioridad del individuo, que se articulan mediante las primeras.

De manera que cuando el individuo crea documentos para perpetuar contenidos o información que son de su interés, o para representar y facilitar el posterior hallazgo de otros documentos, debe articular su lenguaje y la correspondiente plataforma normativa y metodológica que lo formaliza, es decir, el bagaje de herramientas lingüísticas mediante las cuales puede expresar adecuadamente su pensamiento. En este particular, toman parte diversos conceptos intervinientes señalados por T. van Dijk (1978 y 1980); Leahey y Harris (1998); Cantero y de Arriba (1997), tales como:

- La semántica, la cual se refiere al significado general y conceptual de las palabras, grupos de palabras y oraciones, y a la relación entre los signos y los objetos a los que hacen referencia.
- La sintaxis o estructura que supone las reglas para describir cómo pueden colocarse las palabras (sujetos, predicados, frases nominales y otros elementos) para formar frases aceptables en un idioma determinado.
- La estructura proposicional que corresponde a la ubicación y relación mutua entre las ideas o unidades semánticas que componen un discurso.
- La gramática explica sobre todo el sistema de normas que forma la base de la producción y la comprensión de los enunciados de una lengua determinada.
- La pragmática, estudia las relaciones entre texto y contexto.
- La oración que es la unidad de carácter ideal en función de la cual pueden explicarse las producciones lingüísticas reales.
- Las palabras son constructos mínimos cargados de significados que permiten a los hombres articular su lenguaje, formar la conciencia, concebir la realidad, transmitir conocimientos y regular la conducta.
- Los sustantivos son las palabras que permiten concebir la realidad en forma de unidades independientes y estáticas.
- Los adjetivos son palabras usadas para discriminar las cosas observadas, a través de ellos se señalan las características de estas.
- Los verbos son los elementos que ponen en movimiento y relacionan los significados estáticos de los sustantivos.
- Los adverbios permiten concebir las circunstancias en las que se desarrolla el proceso.

Existen otros factores influyentes tanto en el análisis documental como en el análisis del discurso, tales como la finalidad de uso de las formas lingüísticas o intención del autor, señaladas por Pinto Molina y Gálvez (1996) y Foucault (2003); el contexto situacional que refleja los vínculos socio/culturales entre emisor/ receptor; los intereses y conocimientos del analista, ya que es bien sabido que sus conocimientos previos influyen en la comprensión de

lo que analiza, a través de la vinculación que establece entre los nuevos y los existentes elementos informativos.

En líneas generales, puede apreciarse que en el interior de la esfera comunicativa del individuo subyacen, por una parte, los documentos como portadores de la huella de la actividad del hombre; por otra, los discursos como instrumentos de interacción humana por excelencia; y una serie de conceptos creados para explicar la dinámica que se establece en cada uno y entre ellos. A partir de aquí, se teje una relación estrecha entre el análisis documental, inserto en el análisis del discurso, y los conceptos lingüísticos mencionados; por cuanto el análisis documental requiere de elementos metodológicos que orienten la ejecución de los pasos o acciones que se han de seguir para desmontar el contenido de los documentos de la forma más idónea y efectiva posible, y lograr una representación abreviada y unívoca de los aspectos más relevantes de los mismos.

Y en cuanto al análisis del discurso, este dispone de lineamientos normativos y conceptos lingüísticos que sirven de apoyo y basamento teórico, es decir, este tipo de análisis posee toda una plataforma teórico-metodológica que le permite abordar distintos tipos de contenidos discursivos.

El análisis documental no adolece de métodos, estrategias o lineamientos propios para su ejecución pero, en virtud de la complejidad creciente y del consecuente agotamiento de los constructos acabados, es preciso cuestionar y replantear permanentemente lo que se hace, con el fin de mejorar. Una de las vías para lograr esto es mediante la integración y/o adaptación de métodos o conceptos manejados en ciencias o disciplinas afines. En líneas generales, el análisis del discurso constituye un marco metodológico para el análisis documental, en el entendido de que los documentos son discursos expresados a través de textos; por consiguiente, su estructura discursiva puede ser abordada a partir de los lineamientos y procesos aplicados en el análisis del discurso. No obstante, el elemento peculiar del análisis documental radica en su finalidad, que consiste en simplificar el contenido de los documentos y representarlos de una forma diferente a la original, tomando solo sus elementos esenciales o referenciales. Moreiro (1998) explica que la lingüística ha mantenido una relación inexcusable con la documentación, espectro científico en el que se enclava el análisis documental, fundamentada en el estudio del lenguaje, asumido por cada una con finalidades distintas, en el entendido de que para la primera ciencia es el objeto central de su estudio;

mientras que para la segunda constituye una herramienta fundamental para lograr su cometido social que es facilitar el acceso a las fuentes de información por vías descriptivas y analíticas.

Relación documento – texto – conocimiento

Existe también una relación importante en la tríada documento–texto–conocimiento; la esencia del primero radica en sus dos componentes fundamentales, que son la información y el soporte en el que esta se encuentra plasmada. Martínez Comeche (1996) lo define como la materialización de una información en un soporte, cuya finalidad es conservar, perpetuar y transmitir esa carga informativa, prolongando indefinidamente el acceso al mensaje que contiene, su misión social y cultural es trasladar en el tiempo y el espacio el conocimiento acumulado por el ser humano.

Se manifiesta como característica esencial la presencia de un contenido (información) y su respectivo registro en un soporte de cualquier índole, además de su intención o propósito, es decir, que surgen a consecuencia o por la necesidad de dejar constancia de alguna actividad institucional o personal.

En cuanto al texto, Pinto Molina y Gálvez (1996) señala que constituye una estructura discursiva que, como tal, posee encadenamiento y unidad, y está basada en la cohesión de sus elementos. Además, posee autonomía en su significación y en su comunicatividad, también carácter social por cuanto es el instrumento de la expresión y exteriorización de los hablantes. Moreira (2004: 28) lo define como la unidad comunicativo-documental, en cuanto secuencia completa y coherente de informaciones, su actividad comunicativa consiste en trasladar la información de manera ordenada. Y en cuanto al conocimiento, Pérez Montoro (2004) lo concibe como un estado mental que adquiere el individuo a partir de un proceso de asimilación o metabolización de información que, además, está relacionado con las acciones y con las decisiones de quien lo posee y, a la vez, es el factor crítico que permite la asimilación de nueva información y la creación de nuevo conocimiento por lo que es reestructurado continuamente; visto así, el conocimiento reside en la cabeza de los hombres pero también puede materializarse físicamente en documentos mediante los cuales se representa y difunde.

Se entiende, entonces, que el conocimiento es el resultado intelectual del procesamiento que el sujeto realiza sobre la información, basado en sus experiencias, conocimientos previos, datos e informaciones obtenidas sobre asuntos determinados; por consiguiente, su naturaleza

es más compleja pues se construye estableciendo ciertas relaciones y llevando a cabo diversas operaciones sobre un conjunto de informaciones o de otros conocimientos provenientes de diversos contextos y sin conexiones obvias, pero que al combinarlos proporcionan algún beneficio. Al parecer, existe una fuerte implicación entre estos conceptos porque el conocimiento se formaliza mediante el lenguaje escrito, el cual le permite adquirir concreción y saltar de una dimensión tácita a otra explícita; este hecho, a su vez, da lugar a la conformación de textos, constructos que hacen posible que todo individuo plasme y exteriorice el conocimiento propio o el de otros sujetos, haciendo de esta manera posible la comunicación y perpetuación de tales contenidos, y dando lugar así a la creación de documentos.

Con el propósito de consolidar la dimensión lingüística del análisis documental, se presenta la siguiente tabla que puntualiza los puntos de convergencia entre ambas esferas del conocimiento, es decir, la intersección existente entre los conceptos lingüísticos y el análisis documental.

Tabla N° 1: Vinculación de los conceptos lingüísticos con las fases del análisis documental

Conceptos de la Lingüística	Fase del análisis documental en la cual se aplican	Finalidad de uso
Lenguaje, lengua, habla, signo lingüístico, palabras, tipos de palabras.	Durante todo el proceso.	Comprender la estructura del sistema de comunicación humano
Semántica, sintaxis, estructura proposicional, pragmática, gramática, contexto	Interpretación del contenido discursivo. Generación de los productos resultantes del análisis documental (creación de documentos con valor agregado)	Identificar la forma en la que se encuentra estructurado el discurso y el significado de sus partes. Estructurar nuevos discursos que representen abreviadamente los textos analizados.
Oraciones, criterios de textualidad de las oraciones.	Interpretación del contenido discursivo. Generación de los productos resultantes del análisis documental.	Comprender el contenido discursivo y su relación con los factores que intervienen en su producción.

Fuente: Peña Vera y Pirella Morillo (2007, p.56)

Modelo propuesto por Teun van Dijk

T. van Dijk propone un modelo para el desmontaje de contenidos discursivos y su utilidad en el análisis documental. Dicho modelo de análisis de contenido basado en un enfoque deductivo del proceso en el que se señalan las llamadas macro-estructuras semánticas, que corresponden a la reconstrucción teórica de las nociones de tema o asunto del discurso. En el caso de la praxis documental, se refiere a la determinación de los descriptores o palabras claves a través de las cuales se puede representar el contenido de los documentos analizados.

Las macro-estructuras desempeñan un importante papel en la representación de los contenidos textuales en la memoria, según T. van Dijk (1978 y1980), su contenido es lo que con mayor seguridad recordarán los individuos de la puesta en contacto con alguna fuente de información, incidiendo por consiguiente en su recuperación, ya que será también la información más comúnmente usada en la comunicación e interacción posterior. Las macro-estructuras pueden derivarse de un párrafo, una página, un capítulo o del contenido total del documento. Es decir, se pueden establecer relaciones jerárquicas entre ellas.

El autor aclara que el uso de la noción de macro-estructura no es particular del análisis del discurso sino que, por el contrario, cada vez que se procesa información en cualquier ámbito se busca obtener un tema o idea general del texto, episodio o situación completa. De manera que las macro-estructuras son esenciales en cualquier modelo cognoscitivo en el que se produzcan o comprendan discursos, en el que se observen episodios, en el que se participe o se interpreten acciones, en el que se establezcan interacciones o se busquen soluciones a problemas, esto es, en todas las actividades del pensamiento humano.

De forma permanente, el individuo trata de reducir y organizar las grandes cantidades de información que percibe por sus sentidos, esto lo hace con el propósito de generar constructos globales coherentes bien sea de objetos, relaciones, hechos u otros, mediante los cuales pueda manejar la realidad con la cual interactúa.

T. van Dijk (1978 y1980) explica también la existencia de la microestructura que corresponde a la integración de las oraciones que conforman un texto, así como a su mutua conexión y coherencia sobre la cual se fundamenta el sentido global de los discursos. Este concepto atiende entonces a la secuencia proposicional que subyace, es decir, la relación entre las partes o segmentos discursivos.

Para determinar la macro-estructura, el autor plantea el uso de las macro-reglas que son el instrumento metodológico mediante el cual se puede reducir una secuencia de reglas estriba en el hecho de que la reducción de información semántica es necesaria para poder comprender, almacenar y reproducir discursos. En este sentido, las macro-reglas permiten reconstruir la capacidad lingüística del individuo, enlazando significados y convirtiéndolos en totalidades significativas más grandes; su característica principal es su naturaleza organizadora por cuanto le imprime unidad a las proposiciones entendidas como representación abstracta de la información.

T. van Dijk (1978 y 1980) (en Peña Vera y Pirella Morillo 2007, p.66) señala las siguientes macro-reglas:

1. **Omitir:** toda la información (o proposiciones) de poca importancia o no esencial puede ser suprimida.
2. **Seleccionar:** consiste en discriminar elementos del texto, lo cual conlleva también a omitir cierta información que está implicada, bien en los conocimientos generales del receptor, o bien en otros postulados existentes en el discurso.
3. **Generalizar:** esta regla también omite informaciones pero mediante la sustitución de una serie de conceptos o especificidades que se consideran irrelevantes, por el sobre-concepto compartido que define el conjunto, es decir, consiste en generar una proposición generalizante o abstracción que abarque otros conceptos.
4. **Construir o integrar:** consiste en generar una proposición o concepto más general que denote la totalidad del discurso, incluso utilizando términos no mencionados en el texto, pero que son razonablemente deducibles.

Se observa que, desde un punto de vista formal, las dos primeras son reglas de anulación, y las dos restantes son reglas de sustitución; sin embargo, en todas subyace una intención simplificadora derivada del hecho concreto de que el cerebro humano capta, procesa, memoriza con mayor facilidad constructos sintéticos, simplificados, concretos. Las macro-reglas propuestas por T. van Dijk pueden servir como herramientas metodológicas para abordar el proceso de análisis documental porque en ellas está presente la acción simplificadora necesaria para representar sintéticamente el contenido de los documentos analizados,

omitiendo y discriminando información poco relevante o implícita de algún modo en otras partes del texto; y también generando constructos que engloben lo esencial del discurso contenido en ellos y posean mejor calidad semántica.

Las macro-reglas incorporan la esencia del trabajo analítico, a la vez que exhiben una secuencialidad lógica en la ejecución de los procesos o etapas que deben cumplirse para producir idóneas representaciones conceptuales de los documentos. Pues en el primer acercamiento que el analista tiene con el documento debe deslastrar lo accesorio o redundante (información conocida o implícita), descubriendo a la vez lo esencial o medular del contenido; para luego avanzar hacia la configuración de conceptos que mejor expresen o dibujen en la mente de los usuarios los elementos nucleares de los documentos analizados.

La forma de aplicar estas macro-reglas varía en función de los intereses, conocimientos, deseos, normas y valores de cada receptor, es decir, de su estado cognoscitivo contextual, el cual determinará la forma de interpretar un texto.

Otro factor que incide en la aplicación de estas macro-reglas es el tipo de texto y su contexto. De estos depende la relevancia o no que pueda tomar algún aspecto de la información que se analiza. Se concluye pues que, a pesar de que las macro-reglas poseen un carácter general como principios para organizar y reducir globalmente la información, pueden aplicarse de distintas maneras.

Volviendo con las macro-estructuras, T. van Dijk (1978) señala que otro factor que influye en su determinación es la llamada superestructura, que corresponde al tipo o forma que posee el texto, es decir, el esquema al que este se adapta, el cual posee carácter convencional. Para identificar la superestructura, es preciso considerar las funciones que cumple el texto en el contexto comunicativo: describir, convencer, argumentar, dar a conocer opiniones, juicios de valor y otros. Las superestructuras más comunes, desde la óptica del autor son la narración, la argumentación y la demostración.

En líneas generales, las características que consideran para construir o identificar la macro-estructura y la super-estructura son: la presentación del texto (portada si la tuviere), la identificación del emisor del discurso, así como del lugar y fecha de emisión, inter-títulos o subtítulos (si los tuviere), indicadores referidos al tipo de texto y a la superestructura, textos acompañantes tales como anexos, introducción, prólogo, epílogo; la presencia de fórmulas para iniciar, desarrollar o concluir el discurso (como en el caso de la correspondencia

institucional), el tipo de letra, las demarcaciones de párrafos, la existencia de divisiones o partes del discurso, entre otros.

El modelo para el análisis de discursos propuesto por T. van Dijk permite un abordaje deductivo de los textos, considerando los aspectos resaltantes de la generalidad y posteriormente avanza hacia las especificidades del contenido, lo cual puede hacer y de hecho se realiza en parte cuando se ejecutan procesos de análisis documental. En estos, el analista captura los datos formales que identifican a los documentos y luego se aproxima a la esencialidad de su mensaje leyendo partes de su contenido, utilizando el auxilio terminológico de las listas de encabezamientos de materias o consultando descriptores establecidos en algún lenguaje documental creado para tal fin. Se puede decir que en el caso del análisis que se aplica a los documentos se avanza desde su exterior hacia su interior. Buscando siempre construir la representación más idónea, clara e intelectualmente accesible a los usuarios.

Peña Vera y Pirella Morillo (2007, p.80) consideran que la propuesta de T. van Dijk para realizar el análisis de discurso, funda teórica y metodológicamente la práctica del análisis documental. Desde las categorías de macro, micro y superestructura y desde las macro-reglas, se desarrolla el proceso analítico-sintético necesario para establecer los aspectos más relevantes de los documentos y construir representaciones abreviadas y esenciales para lograr una efectiva comunicación del conocimiento registrado.

“En el momento de recibir y emitir cualquier mensaje los individuos activan una serie de procesos de manera consciente e inconsciente que les permite captar, comprender, asimilar, memorizar, evocar recuerdos, expresar o construir nuevas informaciones. En todo ello cobra especial interés la carga subjetiva del emisor o receptor, según sea el caso, porque se produce un involucramiento de sus intereses y conocimientos previos, los cuales se erigen como trasfondo de los resultados obtenidos, tanto al decodificar como al elaborar nuevos contenidos”. (Peña Vera y Pirella Morillo 2007, p.80)

Minería de textos

La minería de texto es un proceso dentro de la minería de datos que permite extraer información a partir de colecciones de documentos de texto generalmente en lenguaje natural,

En un chat el mensaje enviado por cada participante está representado en un texto o discurso, es el que se considera como dato.

En general en la minería de textos lo primero a realizar es representar el texto en algún formato concreto que pueda ser adecuado para los algoritmos de aprendizaje (Ramírez C., Hernández Orallo J., & Ramírez J., 2004).

Esto se realiza en dos pasos:

1. El primero consiste en usar una representación más abstracta, siendo las más habituales en el recupero de la información las siguientes:
 - bolsas de palabras (*bag of words*, Saharni et al. 1996; Lagus et al. 1999): llamada también representación basada en vectores, ya que cada documento se representa como un vector, o basada en frecuencias (el número de veces que ha aparecido en el documento). Esta representación ignora el orden de aparición de las palabras en el texto y es una de las más empleadas en el área de la recuperación de la información.
 - frases (Frank et al. 1999): esta representación consiste simplemente en considerar el documento como un conjunto de frases sintácticas, tal y como se hace en el análisis del procesamiento de lenguaje natural. Esta representación permite mantener el contexto en el que ocurre una palabra, hecho que se pierde en la representación anterior.
 - *n-gramas* (Kargupta, 1997): permiten usar la información sobre la posición de la palabra en el texto, ya que éste se representa mediante secuencias de palabras de longitud máxima n , llamadas *n-gramas*. Permiten un mejor tratamiento de las frases negativas como " ... excepto ... " o " ... pero no ... " que de otra forma tomarían como relevantes las palabras que les siguen.
 - representación relacional (Cohen, 1995): la representación usando lógica de primer orden permite detectar patrones más complejos. Por ejemplo, cada palabra se puede representar mediante un átomo de la forma $w.(d,p)$, el cual es cierto cuando la palabra W ocurre en el documento d en la posición p .

- categorías de conceptos (Deerwester, 1990): también llamado Indexación Semántica Latente (*Laten Semanlic Tndexing*) ya que tiene como objetivo la reducción de la dimensión del vector de palabras inicial, reduciendo las palabras a su raíz morfológica, es decir, las palabras "informando", "información", "informado" e "informador" se representarían por su raíz "inform" y sólo esta palabra se usaría como componente del vector. Esta reducción tiene que ser cuidadosa, ya que otras palabras aparentemente con la misma raíz pueden no tener relación con el término. En el ejemplo anterior, "informal" e "informática" tienen poca relación.

Casi todas estas representaciones se enfrentan al problema del vocabulario (Furnas et al. 1987), es decir, tienen errores semánticos debido a la sinonimia (diferentes palabras con el mismo significado), la quasi-sinonimia (palabras relacionadas con la misma materia, como declaración y comunicado), la polisemia (palabras iguales con diferente significado), los lemas (palabras con el mismo radical, como descubrir y descubrimiento), etc. Aunque se han realizado algunos estudios comparando las distintas representaciones (como por ejemplo Scott & Matwin 1999), en general no se han encontrado diferencias sustanciales en cuanto a las prestaciones de los algoritmos usando una u otra, aunque en un problema específico sí que pueden aparecer diferencias.

2. El segundo paso consiste en reducir el conjunto de características original (reducción de la dimensionalidad en el área del reconocimiento de patrones), ya que el conjunto de características que resultan de las representaciones descritas puede ser de cientos de miles, algo inabordable para muchos de los algoritmos de aprendizaje inductivos. La primera aproximación consiste en eliminar palabras con poca semántica, como son los artículos, preposiciones y conjunciones. En (Moulinier, 1996) se describen dos maneras más elaboradas para reducir la dimensionalidad del vector basadas en el ámbito y en la naturaleza del problema. la reducción por ámbito tiene que ver con la universalidad del conjunto de características, mientras que la reducción por naturaleza describe cómo se seleccionan los atributos.

Se han empleado un gran número de técnicas del aprendizaje automático y estadísticas a la categorización de textos, incluyendo modelos de regresión multivariante (Yang & Chute 1994), clasificadores del vecino más próximo (Yang, 1994), modelos bayesianos (Joachims, 1996; Lewis & Ringuette, 1994), árboles de decisión (Lewis & Ringuette,

1994), redes neuronales (Schütze et al., 1995), aprendizaje de reglas simbólicas (Cohen 1995b) y máquinas de vectores soporte (Joachims, 1998).

La minería de datos y el proceso de descubrimiento de conocimiento en bases de datos

"Descubrimiento de conocimiento en bases de datos"

(Knowledge Discovery in Dataases, KDD).

Fayyad (1996) e define el KDD como "el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos". En esta definición se resumen cuáles deben ser las propiedades deseables del conocimiento extraído:

- **válido:** hace referencia a que los patrones deben seguir siendo precisos para datos nuevos (con un cierto grado de certidumbre), y no sólo para aquellos que han sido usados en su obtención.
- **novedoso:** que aporte algo desconocido tanto para el sistema y preferiblemente para el usuario.
- **potencialmente útil:** la información debe conducir a acciones que reporten algún tipo de beneficio para el usuario.
- **comprensible:** la extracción de patrones no comprensibles dificulta o imposibilita su interpretación, revisión, validación y uso en la toma de decisiones. De hecho, una información incomprensible no proporciona conocimiento (al menos desde el punto de vista de su utilidad).

Como se deduce de la anterior definición, el KDD es un proceso complejo que incluye no sólo la obtención de los modelos o patrones (el objetivo de la minería de datos), sino también la evaluación y posible interpretación de los mismos.

Así, los sistemas de KDD permiten la selección, limpieza, transformación y proyección de los datos; analizar los datos para extraer patrones y modelos adecuados; evaluar e interpretar los patrones para convertirlos en conocimiento; consolidar el conocimiento resolviendo posibles conflictos con conocimiento previamente extraído; y hacer el conocimiento disponible para su uso. Esta definición del proceso clarifica la relación

entre el KDD Y la minería de datos: el KDD es el proceso global de descubrir conocimiento útil desde las bases de datos mientras que la minería de datos se refiere a la aplicación de los métodos de aprendizaje y estadísticos para la obtención de patrones y modelos. Al ser la fase de generación de modelos, comúnmente se asimila KDD con minería de datos.

EL PROCESO DE EXTRACCIÓN DE CONOCIMIENTO

Como se ha mencionado la minería de datos no es más que un paso esencial de un proceso más amplio cuyo objetivo es el descubrimiento de conocimiento en bases de datos. Este proceso consta de una secuencia iterativa de etapas o fases.

Las fases del proceso de extracción de conocimiento

En la Figura N°1 se muestra que el KDD es un proceso iterativo e interactivo. Es iterativo ya que la salida de alguna de las fases puede hacer volver a pasos anteriores y porque a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad.

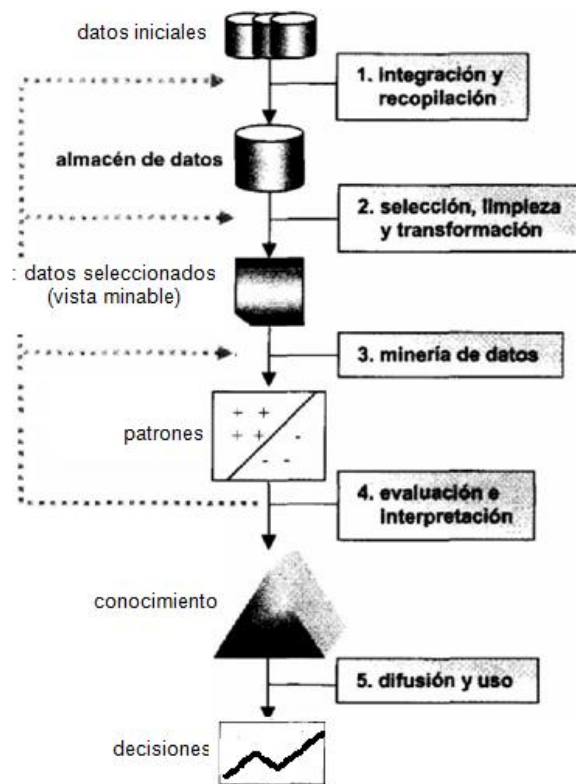
Figura N° 1: Proceso KDD



Fuente: "Introducción a la minería de datos" Orayo, J. (2007, p.13)

El proceso de KDD se organiza entorno a cinco fases como se ilustra en la Figura N° 2.

Figura N° 2: Proceso KDD

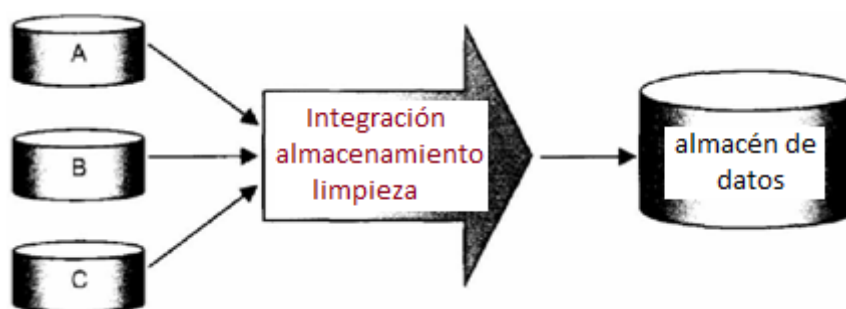


Fuente: "Introducción a la minería de datos" Orayo, J. (2007, p.20)

En la primera fase de integración y recopilación de datos se determinan las fuentes de información que pueden ser útiles y dónde conseguirlas. Para nuestro caso esta fase consiste en obtener y completar el diccionario de la RAE

En la segunda fase, se transforman todos los datos a un formato común, frecuentemente mediante un almacén de datos que consiga unificar de manera operativa toda la información recogida, detectando y resolviendo las inconsistencias. Estas situaciones se tratan en la fase de selección, limpieza y transformación, en la que se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos. Estas dos primeras fases se suelen englobar bajo el nombre de "preparación de datos".

Figura N° 3: Integración de datos



Fuente: "Introducción a la minería de datos" Orayo, J. (2007, p.21)

En la Figura N° 3 se muestra este proceso de integración de un almacén de datos para tres fuentes de datos originales (A, B, C).

La tercera fase, la fase de minería de datos, se decide cuál es la tarea a realizar (clasifica agrupar, etc.) y se elige el método que se va a utilizar. La calidad del conocimiento descubierto no sólo depende del algoritmo de minería utilizado, sino también de la calidad de los datos minados. Por ello, después de la recopilación, el siguiente paso en el proceso de KDD es seleccionar y preparar el subconjunto de datos que se va a minar, los cuales constituyen lo que se conoce como (*lista minable*). Este paso es necesario ya que algunos datos coleccionados en la etapa anterior son irrelevantes o innecesarios o duplicados para la tarea de minería que se desea realizar.

La fase de minería de datos es la más característica del KDD y, por esta razón, muchas veces se utiliza esta fase para nombrar todo el proceso. El objetivo de esta fase es producir nuevo conocimiento que pueda utilizar el usuario. Esto se realiza construyendo un modelo basado en los datos recopilados para este efecto. El modelo es una descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicciones. Para ello es necesario tomar una serie de decisiones antes de empezar el proceso:

- Determinar qué tipo de tarea de minería es el más apropiado.
- Elegir el tipo de modelo. Por ejemplo, para una tarea de clasificación podríamos usar un árbol de decisión, porque queremos obtener un modelo en forma de reglas.
- Elegir el algoritmo de minería que resuelva la tarea y obtenga el tipo de modelo que estamos buscando. Esta elección es pertinente porque existen muchos métodos para construir los modelos.

Tareas de la minería de datos

Dentro de la minería de datos hemos de distinguir tipos de tareas, cada una de las cuales puede considerarse como un tipo de problema a ser resuelto por un algoritmo de minería de datos. Esto significa que cada tarea tiene sus propios requisitos, y que el tipo de información obtenida con una tarea puede diferir mucho de la obtenida con otra.

Se puede hacer una clasificación de las distintas tareas pueden ser predictivas o descriptivas.

Entre las tareas predictivas se encuentran: la clasificación y la regresión, La clasificación es quizá la tarea más utilizada. Existen variantes de la tarea de la clasificación, como son el aprendizaje de "rankings", el aprendizaje de preferencias, el aprendizaje de estimadores de probabilidad, etc. La regresión es también una tarea predictiva.

Mientras que el agrupamiento (*clustering*), las reglas de asociación, las reglas de asociación secuenciales y las correlaciones son tareas descriptivas.

El agrupamiento (*clustering*) es la tarea descriptiva por excelencia y consiste en obtener grupos "naturales" a partir de los datos. Hablamos de grupos y no de clases, porque, a diferencia de la clasificación, en lugar de analizar datos etiquetados con una clase, los analiza para generar esta etiqueta. Los datos son agrupados basándose en el principio de maximizar la similitud entre los elementos de un grupo minimizando la similitud entre los distintos grupos. Es decir, se forman grupos tales que los objetos de un mismo grupo son muy similares entre sí y, al mismo tiempo, son muy diferentes a los objetos de otro grupo. Al agrupamiento también se le suele llamar segmentación, ya que parte o segmenta los datos en grupos que pueden ser o no disjuntos.

Técnicas de minería de datos

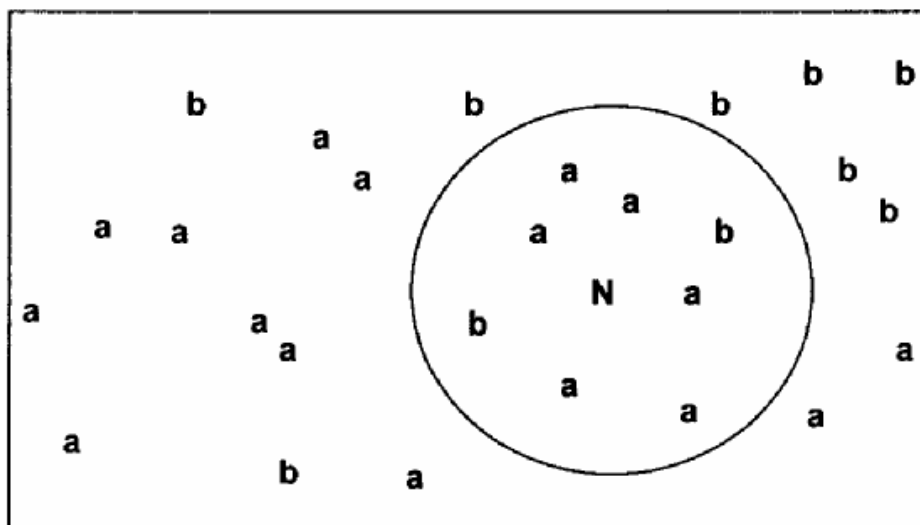
Dado que la minería de datos es un campo muy interdisciplinar, existen diferentes paradigmas detrás de las técnicas utilizadas para esta fase: técnicas de inferencia estadística, árboles de decisión, redes neuronales, inducción de reglas, aprendizaje basado en instancias, algoritmos genéticos, aprendizaje bayesiano, programación lógica inductiva y varios tipos de métodos basados en núcleos, entre otros. Cada uno de estos paradigmas incluye diferentes algoritmos y variaciones de los mismos, así como otro tipo de restricciones que hacen que la efectividad del algoritmo dependa del dominio de aplicación, no existiendo lo que podríamos llamar el método universal aplicable a todo tipo de aplicación.

En el aprendizaje basado en instancias o casos, las instancias se almacenan en memoria, de tal forma que cuando llega una nueva instancia cuyo valor es desconocido se intenta relacionar ésta con las instancias almacenadas (cuya clase o valor es conocida) buscando las que más se parecen, con el objetivo de usar los valores de estas instancias similares para estimar los valores a obtener de la nueva instancia en cuestión. Por lo tanto, más que intentar crear reglas, se trabaja directamente con los ejemplos.

Todo el trabajo en el aprendizaje basado en instancias se hace cuando llega una instancia a clasificar y no cuando se procesa el conjunto de entrenamiento. En este sentido se trata de un método retardado o perezoso, ya que retrasa el trabajo real tanto como sea posible, a diferencia de los otros métodos vistos hasta el momento que tienen un comportamiento anticipativo o voraz, produciendo generalizaciones en cuanto reciben los datos de entrenamiento.

En el aprendizaje basado en instancias, cada nueva instancia se compara con las existentes usando una métrica de distancia, y la instancia más próxima se usa para asignar su clase a la instancia nueva. La variante más sencilla de este método de clasificación es conocido como "el vecino más próximo" (r-zearest-neighbor). Otra variante, conocida como el método de los "k vecinos más próximos" (k-rzearest-neighbors), usa los k vecinos más próximos, en cuyo caso la clase mayoritaria de estos k vecinos se asigna a la nueva instancia. En la Figura N° 4 se observa que a la nueva instancia N se le ha de asignar la clase Q ya que, entre los vecinos más próximos (marcados con un círculo), hay más instancias de la clase a que de la b.

Figura N° 4: K-vecinos más próximos



Fuente: "Introducción a la minería de datos" Orayo, J. (2007, p.34)

El aprendizaje basado en instancias es muy útil para trabajar sobre tipos de datos no estándar, como los textos o multimedia. El único requerimiento para incluir un tipo de datos es la existencia de una métrica apropiada de distancia para formalizar el concepto de similitud. Calcular la distancia entre dos ejemplos o instancias es trivial cuando tienen un atributo numérico: basta con calcular la diferencia entre sus valores. Cuando hay varios atributos numéricos se puede usar la distancia euclídea, asumiendo que los atributos están normalizados y que son de igual importancia. Estas suposiciones no siempre se corresponden con la realidad y existen variantes que ponderan más los atributos importantes, así como otras métricas de distancia.

Los atributos nominales deben tratarse de forma especial definiendo "distancias" entre los diferentes valores. Normalmente, se asigna una distancia 0 si los valores son idénticos, y una distancia de 1 en cualquier otro caso. Así, la distancia entre *joven* y *joven* es 0, mientras que la distancia entre *joven* y *adulto* es de 1. En algunos casos es conveniente usar medidas más sofisticadas. Por ejemplo, podríamos usar una medida que estableciera que *joven* es más próximo a *adulto* que a *mayor*. Finalmente, algunos atributos son más importantes que otros, lo cual puede reflejarse en la métrica a través de pesos.

Construcción del modelo

Es en la construcción del modelo donde vemos mejor el carácter iterativo del proceso de KDD, ya que será necesario explorar modelos alternativos hasta encontrar aquel que resulte más útil para resolver nuestro problema. Así, una vez obtenido un modelo y a partir de los resultados obtenidos para el mismo, podríamos querer construir otro modelo usando la misma técnica pero otros parámetros o quizás usar otras técnicas o herramientas. En esta búsqueda del "buen modelo" puede que tengamos que retroceder hasta fases anteriores y hacer cambios en los datos que estamos usando o incluso modificar la definición del problema. Es más la elección de la tarea a realizar y del algoritmo a usar puede influir en la preparación de los datos (por ejemplo, un determinado algoritmo o técnica puede requerir que los datos se presenten en un formato determinado).

El proceso de construcción de modelos predictivos requiere tener bien definidas las etapas de entrenamiento y validación para asegurar que las predicciones serán robustas y precisas. La idea básica es estimar (o entrenar) el modelo con una porción de los datos (*training dataset*) y luego validarlo con el resto de los datos (*test dataset*), como pasamos a ver en la sección siguiente.

En la cuarta fase de evaluación e interpretación, se evalúan los patrones y se analizan por los expertos, y si es necesario: vuelve a las fases anteriores para una nueva iteración. Esto incluye resolver posibles conflictos con el conocimiento que se disponía anteriormente.

Medir la calidad de los patrones descubiertos por un algoritmo de minería de datos no es un problema trivial, ya que esta medida puede atañer a varios criterios, algunos de ellos bastante subjetivos. Idealmente, los patrones descubiertos deben tener tres cualidades: ser precisos, comprensibles (es decir, inteligibles) e interesantes (útiles y novedosos). Según las aplicaciones puede interesar mejorar algún criterio y sacrificar ligeramente otro.

Técnicas de evaluación

Para entrenar y probar un modelo se parten los datos en dos conjuntos: el conjunto de entrenamiento (*training set*) y el conjunto de prueba o de test (*test set*). Esta separación es necesaria para garantizar que la validación de la precisión del modelo es una medida

independiente. Algunos algoritmos de aprendizaje utilizan *internamente* un tercer conjunto que extraen del conjunto de aprendizaje, denominado conjunto de validación (*validation dataset*), para refinar el modelo o elegir entre posibles modelos antes de la salida final del algoritmo. No hemos de confundir esta pre-validación o evaluación interna con la verdadera evaluación y, por tanto, el conjunto de validación con el conjunto de test

Si no se usan conjuntos diferentes de entrenamiento y prueba, la precisión del modelo será sobreestimada, es decir, tendremos estimaciones muy optimistas.

En los modelos predictivos, el uso de esta separación entre entrenamiento y prueba es fácil de interpretar. La precisión es una buena estimación de cómo se comportará el modelo para datos futuros similares a los de test. Esta forma de proceder no garantiza que el modelo sea correcto, sino que simplemente indica que si usamos la misma técnica con una base de datos con datos similares a los de prueba, la precisión media será bastante parecida a la obtenida con éstos.

El método de evaluación más básico, la validación simple, reserva un porcentaje de la base de datos como conjunto de prueba, y no lo usa para construir el modelo. Este porcentaje suele variar entre el cinco por ciento y el 50 por ciento. La división de los datos en estos dos grupos debe ser aleatoria para que la estimación sea correcta.

El método que se usa normalmente es la validación cruzada con n pliegues.

Otra técnica para estimar el error de un modelo cuando se disponen de pocos datos es la conocida como *bootstrapping*.

Medidas de evaluación de modelos

Dependiendo de la tarea de minería de datos existen diferentes medidas de evaluación de los modelos. Por ejemplo, en el contexto de la clasificación, lo normal es evaluar la calidad de los patrones encontrados con respecto a su *precisión predictiva*, la cual se calcula como el número de instancias del conjunto de prueba clasificadas correctamente dividido por el número de instancias totales en el conjunto de prueba.

En el caso de que la tarea sea de reglas de asociación, se suele evaluar de forma separada cada una de las reglas con objeto de restringimos a aquellas que pueden aplicarse a un

mayor número de instancias y que tienen una precisión relativamente alta sobre estas instancias. Esto se hace en base a dos conceptos:

- *Cobertura* (también referida como soporte): número de instancias a las que la regla se aplica y predice correctamente.
- *Confianza*: proporción de instancias que la regla predice correctamente, es decir, la cobertura dividida por el número de instancias a las que se puede aplicar la regla.

Para la tarea de agrupamiento, las medidas de evaluación suelen depender del método utilizado, aunque suelen ser función de la cohesión de cada grupo y de la separación entre grupos. La cohesión y separación entre grupos se puede formalizar, por ejemplo, utilizando la *distancia* media al centro del grupo de los miembros de un grupo y la distancia media entre grupos, respectivamente. El concepto de distancia y de densidad son dos aspectos cruciales tanto en la construcción de modelos de agrupamiento como en su evaluación.

En la quinta fase, la fase de difusión se hace uso del nuevo conocimiento y se hace partícipe de él a todos los posibles usuarios. Para cada una de estas fases se emplean distintas técnicas de las diferentes disciplinas relacionadas.

Una vez construido y validado el modelo puede usarse principalmente con dos finalidades: para que un analista recomiende acciones basándose en el modelo y en sus resultados, o bien para aplicar el modelo a diferentes conjuntos de datos. También puede incorporarse a otras aplicaciones, es importante medir como el modelo evoluciona. Aun cuando el modelo funcione bien debemos continuamente comprobar las prestaciones del mismo. Esto se debe principalmente a que los patrones pueden cambiar. Por lo tanto, el modelo deberá ser monitorizado, lo que significa que de tiempo en tiempo el modelo tendrá que ser reevaluado, re-entrenado y posiblemente reconstruido completamente.

CAPITULO III - Planteamiento del Problema

Este capítulo presenta la vinculación entre los antecedentes teóricos sobre manejo de la información, análisis del discurso, minería de datos y proceso de obtención de nuevos conocimientos, con la solución planteada para resolver el algoritmo matemático que permita moderar las conversaciones en español en chat educativos.

Esquemáticamente, la vinculación planteada, remite a:

1. la información contenida en el mensaje:

- ✓ El método de T. van Dijk
- ✓ Los sustantivos como elementos que determinan los “que” de la cuestión en el mensaje

2. Cómo determinar el significado de las palabras

- ✓ La obtención de la base de datos de la RAE

3. El proceso de obtención de nuevo conocimiento KDD

- ✓ La aplicación de las fases del proceso

4. Cómo crear la estructura de referencia

- ✓ La información que suministra el Docente al momento de configurar la temática del chat
- ✓ La información adicional que debe proveer el docente para referenciar en el algoritmo.

5. Las condiciones de funcionamiento:

- ✓ El motivo de concentrarse en el algoritmo sin gastar energías en misceláneos como por ejemplo: errores de ortografía, sinonimia, emoticones, regionalismo, etc.

6. Cómo acceden los usuarios al sistema

- ✓ Sistema de login con previa registración

7. Cómo se desarrolla el chat

- ✓ Lo que sucede cuando un participante del chat envía un mensaje

Una vez ingresado un alumno a una sala de chat, puede comenzar a conversar con los demás participantes de la sala.

Al ingresar un mensaje a enviar, este será visto por todos los participantes de la sala, el sistema es el encargado de capturar el mensaje analizarlo y determinar el grado de similitud con la estructura de referencia para así de acuerdo con el punto de intervención determinado por el docente el sistema enviar un mensaje para que los participantes retome la temática planteada.

8. Las pruebas de validación del sistema

- ✓ Aplicación del método TAM: el Modelo fue propuesto por Fred Davis en 1985.

Detallaremos a continuación los aspectos del esquema anterior:

1. la información contenida en el mensaje:

- ✓ El método de T. van Dijk

Para analizar la información contenida en el mensaje de un participante de chat, se apela al método de T. van Dijk, que propone una secuencia de macro-reglas para el desmontaje del discurso.

El mensaje enviado por un participante del chat se considera como un pequeño discurso que se debe analizar para obtener la información sobre el asunto que se está queriendo referir.

De los estudios realizados por T. van Dijk (1978 y 1980); Leahey y Harris (1998); Cantero y de Arriba (1997) se infiere un modelo de análisis de contenido. Dicho modelo se basa en un enfoque deductivo del proceso, en el que se señalan macroestructuras semánticas que permiten la reconstrucción teórica de las nociones de tema o asunto del discurso propuesto por T. van Dijk. Este proceso reconocer las dos primeras macro reglas a saber:

a. Omitir: toda la información (o proposiciones) de poca importancia o no esencial puede ser suprimida: información numérica, conectores, etc.

b. Seleccionar: consiste en discriminar elementos del texto, de acuerdo a la descripción de los componentes que conforma el discurso, son los sustantivos los elementos que se consideran.

✓ Los sustantivos como elementos que determinan los “que” de la cuestión en el mensaje

En este punto se señala la necesidad de solo seleccionar los sustantivos, puesto que ellos son las palabras que permiten concebir la realidad en forma de unidades independientes y estáticas, es decir, son los que determinan los objetos de referencias que el discurso quiere transmitir, es decir, los sustantivos son los elementos que determinan los objetos a los que hace referencia el discurso, es por ello que se guardan para formar la base de conocimiento.

2. Cómo determinar el significado de las palabras

Para determinar el significado de las palabras, es necesario contar con una base de datos de las palabras bien formadas, con su significado y sus atributos, en el idioma español.

En el caso de esta tesis, esta base de datos es la versión N° 22 de la Real Academia Española (RAE), (2011).

3. El proceso de obtención de nuevo conocimiento KDD

En la primera fase de integración y recopilación de datos se determinan las fuentes de información que pueden ser útiles y dónde conseguirlas. Para nuestro caso, como ya dijimos, las palabras bien formadas se obtuvieron de la RAE.

En la segunda fase, se transforman todos los datos a un formato común. Si bien las palabras están disponibles en el sitio de la RAE, ellas pueden consultarse de manera aislada (palabra por palabra), pero no se puede obtener en forma directa todo el banco de datos. Para resolver este problema, se diseñó un algoritmo que permita bajar la base de datos mediante consultas sistemáticas de la totalidad de las palabras (aproximadamente 87,000), para tenerlas disponibles en una bases de datos local, y así contar con los datos preparados.

Un inconveniente que surgió fue el caso de las palabras en plural, pues el diccionario hace referencia a la palabra en singular dándole su significación. La solución fue realizar un proceso de singularización de las palabras que están en plural

Para la tercera fase, que es la fase de minería de datos, se decidió que la tarea de agrupar es la más indicada para realizar la minería de datos a realizar. Además se definió el método a utilizar: el método basado en casos.

Otra definición tomada fue con qué *técnica de minería de datos* se trabajaría: similitud basada en la distancia.

En el aprendizaje basado en instancias o casos, las instancias se almacenan en memoria, de tal forma que cuando llega una nueva instancia de valor desconocido, se la relaciona con las instancias almacenadas (cuya clase o valor es conocida) buscando las que más se parecen. Ello se hace con el objetivo de usar valores de estas instancias similares para estimar valores a obtener de la nueva instancia en cuestión. Por lo tanto, más que intentar crear reglas, se trabaja directamente con los ejemplos.

En el aprendizaje basado en instancias, cada nueva instancia se compara con las existentes usando una métrica de distancia, y la instancia más próxima se usa para asignar su clase a la instancia nueva.

La cuarta fase que es la de evaluación e interpretación se utilizaron para entrenar y probar el modelo dos conjuntos: el conjunto de entrenamiento (*training set*) y el conjunto de prueba o de test (*test set*).

La quinta fase se utilizó encuestas para medir el grado de aceptación del modelo.

4. Cómo crear la estructura de referencia

✓ La información que suministra el Docente al momento de configurar la temática del chat

Para formar la base de conocimiento es necesario que el experto, en este caso el docente que configura el chat para una sesión controlada, tiene que brindar al sistema una serie de datos como: título, palabras claves, resumen, objetivos y el artículos que los alumnos tiene que estudiar para discutir o intercambiar opinión en la sesión de chat.

Con esta información se aplica el método de T. van Dijk, para ello se aplica la primera macro regla que consiste en omitir información superflua, y se aplicó la segunda que es seleccionar la información que determine el objeto de referencia en el discurso.

✓ La información adicional que debe proveer el docente para referenciar en el algoritmo.

En el segundo paso de la configuración, el profesor se encontrará con ajustes para el “Moderador” automático, estas opciones de ajuste son en base a la información cargada del artículo en el paso anterior.

La configuración en este caso consiste en, primero, seleccionar las frases con las que el agente automático intervendrá en la conversación en caso de que esta se haya desviado de la temática planteada en el artículo. Además de las opciones de frases de intervención planteadas por el sistema, el profesor podrá agregar frases de intervención del sistema personalizadas que considere pertinentes para el control de la sesión de chat.

Luego, el sistema debe detectar cuál es el grado de tolerancia del profesor en el contexto que se analizará en el chat. Para esto, se seleccionan cinco grupos de palabras, ordenadas de acuerdo a su peso en la estructura de referencia. Las palabras del primer grupo serán las de mayor peso, las del segundo grupo de menor peso, y así hasta el último grupo que corresponde a palabras de peso mínimo. De acuerdo al criterio del profesor, el grupo de palabras elegido determinará de cierta forma la tolerancia que tendrá para la conversación, esta tolerancia será traducida en un porcentaje dependiendo de la selección del profesor, de modo que si selecciona el primer grupo (con las palabras de mayor peso) la tolerancia será la más estricta, casi sin dejar que los participantes se desvíen en ningún momento, con este criterio, si selecciona el siguiente grupo de palabras, o el siguiente, el porcentaje de tolerancia para intervenir irá bajando.

Una vez seleccionado este valor, se almacena, y luego se compara con el valor del grado de similitud que se irá obteniendo durante la conversación, si la similitud es menor al porcentaje de tolerancia se producirá la intervención del sistema o Moderador en la conversación con alguna de las frases elegidas en esta segunda etapa de configuración. La selección de la frase se realizará de forma aleatoria por cada intervención. Lo siguiente para el profesor es colocar las consignas de trabajo que serán visibles para los participantes durante la sesión del chat. Por último el docente debe ingresar la fecha, hora y duración de la sesión de chat.

5. Las condiciones de funcionamiento:

- ✓ El motivo de concentrarse en el algoritmo sin gastar energías en misceláneos como por ejemplo: errores de ortografía, sinonimia, emoticones, regionalismo, etc

En el presente trabajo tiene como objetivo principal crear las estructuras y algoritmos necesarios para construir un chat moderado a una temática específica, si bien es importante considerar la falta de ortografía en la escritura dentro de un chat la solución para este problema se puede resolver en futuras ampliaciones con una estructura de sinónimos, al igual que las contracciones y los emoticones.

6. Cómo acceden los usuarios al sistema

- ✓ Sistema de login con previa registración

Este trabajo está destinado a formar parte de una plataforma educativa, donde la validación del acceso está dada por la plataforma, mientras esto no ocurra, es necesario validar el ingreso al sistema, esto se logra con la registración de los usuarios.

7. Cómo se desarrolla el chat

- ✓ Lo que sucede cuando un participante del chat envía un mensaje

Una vez ingresado un alumno a una sala de chat, puede comenzar a conversar con los demás participantes de la sala. Cada participante queda identificado con un color diferente, que se establece de forma aleatoria al momento del ingreso del alumno a la sala de chat. Al enviar un mensaje, este será visto por todos los participantes de la sala.

El moderador captura todos los mensajes, los analiza y mediante un proceso decidirá la intervención.

1. Las pruebas de validación del sistema

- ✓ Aplicación del método TAM:

El modelo TAM tiene sus raíces en la Teoría de la Acción Razonada (TRA) de Ajzen y Fishbein (1975), donde las creencias están influidas por las actitudes, las cuales conducen a intenciones y resultan en algunos tipos de comportamiento.

La TRA es una teoría general que trata de explicar casi todo tipo de comportamiento humano, partiendo de la importancia de las creencias individuales, con el fin de predecir la conducta humana.

En el contexto de la aceptación tecnológica, esta teoría se utiliza para tratar de establecer los factores que condicionan al usuario hacia la innovación, la intención de comportamiento y la intensidad de uso del sistema (Fishbein y Ajzen, 1975; Ajzen y Fishbein, 1980).

El modelo TAM propone que el uso del Sistema es una respuesta influenciada por constantes estímulos externos de las capacidades y características del sistema actual.

CAPITULO IV - Solución

En este capítulo se presenta la Solución diseñada para resolver las desviaciones en las conversaciones que los participantes realizan en sesiones de chat moderado para contextos educativos.

El Problema se puede dividir en dos etapas:

1. La primera es cuando el docente configura la sesión de chat. En esa instancia, el docente proporciona de distintas formas la información al sistema, para que éste la procese y forme la base de conocimiento. Esta estructura de referencia está formada por todos los elementos del lenguaje que determinan de que se trata el tema a discutir en el chat.
2. La segunda etapa es cuando se desarrolla el chat, es decir, cada mensaje que se envía es capturado para su análisis, para luego poder determinar si se interviene enviando un mensaje a la sala.

Interesa mostrar cómo se resolvió el problema de moderar dichas conversaciones en español.

Para ello:

1. Se ha construido un sistema de chat propio a fin de tener acceso a todo el sistema, pues plataformas educativas tales como Moodle, Educativa, etc., poseen chat con acceso restringido a sus algoritmos.
2. Se obtuvo la base de datos completa de la RAE a fin de obtener el banco de palabras aceptadas en español, y así contar con la misma, sin necesidad de conexión on line de internet.
3. Se diseñó la estructura de referencia a fin de lograr que el docente que configura el chat, proporcione la información necesaria para la temática.
4. Se configuró la sesión de chat con el propósito de que los mensajes que envían los participantes del chat sean capturados y analizados, para que mediante el algoritmo, el sistema los compare con la estructura y decida automáticamente la intervención cuando dicho mensaje se desvía de la temática pautada por el docente.

A continuación se desarrolla en detalle los pasos enumerados arriba.

1. Se ha construido un sistema de chat propio

El sistema está desarrollado en lenguaje JAVA, haciendo uso de páginas JSP (Java Server Page), por tratarse de un sistema web.

Las páginas JSP, permiten contener código JAVA embebido en HTML, lo que da una gran potencial, por poder utilizar dicho lenguaje para crear sistemas de este tipo para la web. El motor de base de datos utilizada para el sistema es MySQL Server.

Además, se utilizaron otras herramientas para el desarrollo de la interfaz de la aplicación, como son hojas de estilo CSS y funciones JavaScript. Ellas se seleccionaron porque aportan efectos a la interfaz.

El servidor web utilizado para montar el sistema es Apache Tomcat.

El sistema, por ahora, está montado en un servidor local que atenderá de forma concurrente las peticiones de los clientes y los mensajes que se envíen desde todas las computadoras que se conecten al mismo. Esta situación variará cuando se integre el sistema a una plataforma educativa.

2. Se obtuvo la base de datos completa de la RAE

En el sitio web www.rae.es, de la Real Academia Española, es posible consultar las palabras de una por vez. Para obtener la base de datos de todas las palabras, se formuló e implementó un algoritmo a ese fin, Este algoritmo fue desarrollado en el lenguaje de programación Python.

2. Se obtuvo la base de datos completa de la RAE

Para obtener la base, se formuló e implementó un algoritmo a ese fin¹.

En el lenguaje de programación Python se realizó un algoritmo que permite realizar una consulta a la página de la Real Academia Española, a fin de obtener la definición de la palabra consultada. Luego, inserta la palabra con su definición en una base de datos local. Este proceso se realiza de forma sistemática para todas las palabras de la base de datos de la RAE, aproximadamente 85 mil palabras.

¹ Dicho algoritmo fue desarrollado por el alumno Gustavo Conturso, alumno vinculado con el proyecto de investigación marco de la tesis, bajo la Dirección del autor del presente trabajo de tesis.

A continuación se presenta el algoritmo que permite la obtención completa de la RAE.

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
#
# diccionario.py
#
# Copyright 2013 Gustavo E. Conturzo <gustavo@gmail.com>
#

import sqlite3
from funciones import *

def main():

# Abre el archivo que contiene el listado de palabras

    f1 = file("palabras.txt", "r")
    texto = f1.readlines()
    f1.close()

# Creamos o conectamos con la base de datos
    con = sqlite3.connect('./diccionario.db')
    con.text_factory = str
    cursor = con.cursor()

# Creamos la tabla
    try:
        cursor.execute('''CREATE TABLE PALABRAS
        (PALABRA TEXT PRIMARY KEY,
        DEFINICION TEXT)''')
    except:
        print "Tabla creada"

# Recorre el archivo limpiando texto e insertando en la BD
    for concepto in texto:
        concepto_limpio = limpiar_concepto(concepto) # limpia el
texto
        definido = concepto_limpio.split(".", 1) # divide
palabra y definición
        palabra = str(definido[0])
        definicion = str(definido.pop()).strip()
        if palabra != "": # acá va la inserción en la BD

            try:
                cursor.execute("INSERT INTO PALABRAS (PALABRA,
DEFINICION) VALUES (?, ?)", (palabra, definicion))
                con.commit()
            except:
                con.rollback()

        print "Palabra: " + palabra
        print "Definición: " + definicion + "\n"
```

```

con.close()

"""
cursor.execute("SELECT palabra, definicion from PALABRAS")
for i in cursor:
    print "Palabra " + i[0]
    print "Definicion: " + i[1] + "\n"
"""

if __name__ == '__main__':
    main()

#!/usr/bin/env python
# -*- coding: utf-8 -*-
#
# funciones.py
#
# Copyright 2014 Gustavo E. Conturzo <gustavo@gmail.com>
#

def limpiar_concepto(x):
    #x = x.replace("\n", " ")
    #x = x.replace(".\n", ".")
    #x = x.replace("\n", " ")
    #x = x.replace("\n", " ")
    #x = x.replace("_\n", " ")
    #x = x.replace("_", " ")
    #x =
x.replace("
", " ")
#x = x.replace(" ", " ")
#x = x.replace("'", " ")
x = x.replace(" ", " ")
x = x.replace(" ", " ")
x = x.replace(" ", " ")
x = x.replace(" ", " ")
x = x.replace(" ", " ")
x = x.replace(" ", " ")
x = x.replace(" ", " ")
x = x.replace(" ", " ")
x = x.replace(" ", " ")
x = x.replace("text", "")
x = x.replace("&acute;", "á")
x = x.replace("&oacute;", "ó")
#x = x.replace("\xc3\xa9", "é")
#x = x.replace("\xc3\xad", "í")
#x = x.replace("\xc3\xb3", "ó")
#x = x.replace("\xc3\xba", "ú")
#x = x.replace("\xc3\xb1", "ñ")
x = x.replace("&x00F1;", "ñ")
x = x.replace("\nReal Academia Española. Diccionario Usual.", "")
x = x.replace("Artículo enmendado.", "")
x = x.replace("Avance de la vigésima tercera edición", "")
x = x.replace("Real Academia Española © Todos los derechos
reservados", "\n")

```



```
x = x.replace("\Aviso", "")
x = x.replace("-----", "")
x = x.replace("No se ha podido procesar su información. Puede
intentarlo más tarde.\", "")
x = x.strip()
return x
```

3. Se diseñó la estructura de referencia

El docente configura la sesión de chat, como muestra la siguiente imagen.

Imagen N°1: Interfaz del primer paso en el proceso de configuración de una sala de chat

The image shows a web form titled "Configuración del artículo a tratar en la sesión de chat". The form is set against a light gray background and contains the following elements:

- Título del artículo:** A single-line text input field.
- Palabras Clave:** A vertical list of 10 numbered input boxes, each preceded by a number from 1 to 10.
- Abstract:** A large, empty text area with a small scroll indicator in the bottom right corner.
- Objetivos:** A large, empty text area with a small scroll indicator in the bottom right corner.
- Artículo a tratar:** A single-line text input field with a small scroll indicator in the bottom right corner.
- Guardar información:** A button located at the bottom right of the form.

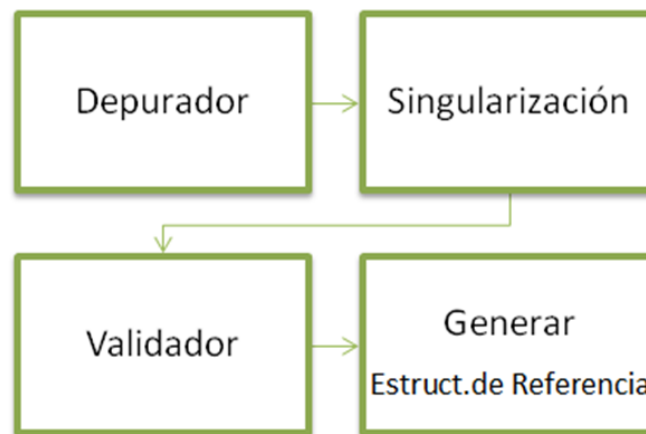
En la imagen anterior se muestra cómo se realiza el ingreso toda la información de la temática a tratar: el título de la temática, palabras claves (que son palabras muy vinculadas con la temática) un resumen de la temática, objetivos sobre el tema y el artículo o documento a tratar.

Con esta información, el sistema analiza palabra por palabra (a través de un método llamado depurador), luego se buscan las palabras en el diccionario.

Dado que en el diccionario se encuentra la significación de las palabras en singular, cuando las palabras están en plural, fue necesario detectarlas y aplicarles un proceso de singularización (que más adelante se verá en detalle), basado en las reglas de la formulación del plural. Con toda esa información se crea una base de conocimiento con la significación de la temática.

La construcción y configuración de la estructura de referencia sigue la siguiente secuencia:

Diagrama N°1. Procedimientos que forman parte del módulo “Analizador” del sistema



Fuente: Elaboración propia

Dicha secuencia consiste en:

- La primera función que se aplica, llamada “*Depurador*”, consiste en la toma de toda la información del artículo, para procesar todas las palabras, eliminando signos de puntuación, números y símbolos especiales (1° macro-regla de T. van Dijk). Luego, se convierten todas las palabras en minúscula para poder trabajar con este patrón durante todo el proceso y guardar de esta forma las palabras en la base de datos. Esta función obtiene como resultado un arreglo de palabras limpio.

b) La siguiente función, denominada “*Singularización*”, tiene como entrada el arreglo generado por la función anterior. Su objetivo es convertir al singular todas las palabras del arreglo o base de datos que no se encuentren de esta forma. La razón de esto es la necesidad de luego buscar las palabras en un diccionario, motivo por el cual se debe hacer este proceso de singularización.

El procedimiento realizado por este algoritmo hace uso de las reglas básicas del idioma español para las palabras en plural, convirtiendo al singular a las que corresponda y validando su existencia con el diccionario.

Las palabras que no se encuentren en el diccionario, son eliminadas del arreglo por tratarse de palabras mal escritas o incorrectas. Se obtiene como resultado entonces, el arreglo de palabras singularizado, y sin palabras incorrectas.

A continuación se detalla el algoritmo creado para resolver la singularización de palabras.

Algoritmo de singularización de palabras

Se detallará solamente este algoritmo con una breve descripción, se presentará un diagrama de flujo y el código correspondiente.

El algoritmo de singularización recibe como parámetro de entrada una palabra, y en caso de que la misma esté en plural, devuelve como salida su forma singular. Esto sucede analizando las terminaciones de las palabras y aplicando las reglas del lenguaje, posibilitando reconstruir el singular de cualquier plural del idioma Español.

En la lengua española la forma del número del plural presenta tres variantes:

- ✓ En -s (casas)
- ✓ En -es (profesores)
- ✓ En grado cero (atlas)

La variación del número es un fenómeno uniforme y fácil de describir o estudiar. Pero existen ciertos tipos de palabras que presentan complicaciones, ya que este proceso no es una tarea sencilla, como pueden ser los cambios ortográficos necesarios para singularizar adecuadamente algunas formas.

Estos cambios pueden ser:

- ✓ La eliminación del acento ortográfico (gérmenes -> germen),
- ✓ La adición de acento ortográfico (volcanes -> volcán)
- ✓ La sustitución de la letra C por Z (voces -> voz) (SÁNCHEZ, 1995).

El proceso de singularización comprueba la terminación de la palabra en plural. Para llevar adelante dicho proceso, se distinguen dos grandes grupos:

- las terminadas en –es

- las terminadas en –s

Estos dos grupos se dividen a su vez en subgrupos con condiciones de terminación más restrictivas.

Pueden existir terminaciones complejas (ejemplo: cipreses), y otras más sencillas (ejemplo: tés).

Dependiendo del tipo de terminación se efectúan las operaciones necesarias para singularizar la palabra. El conjunto de operaciones posibles es:

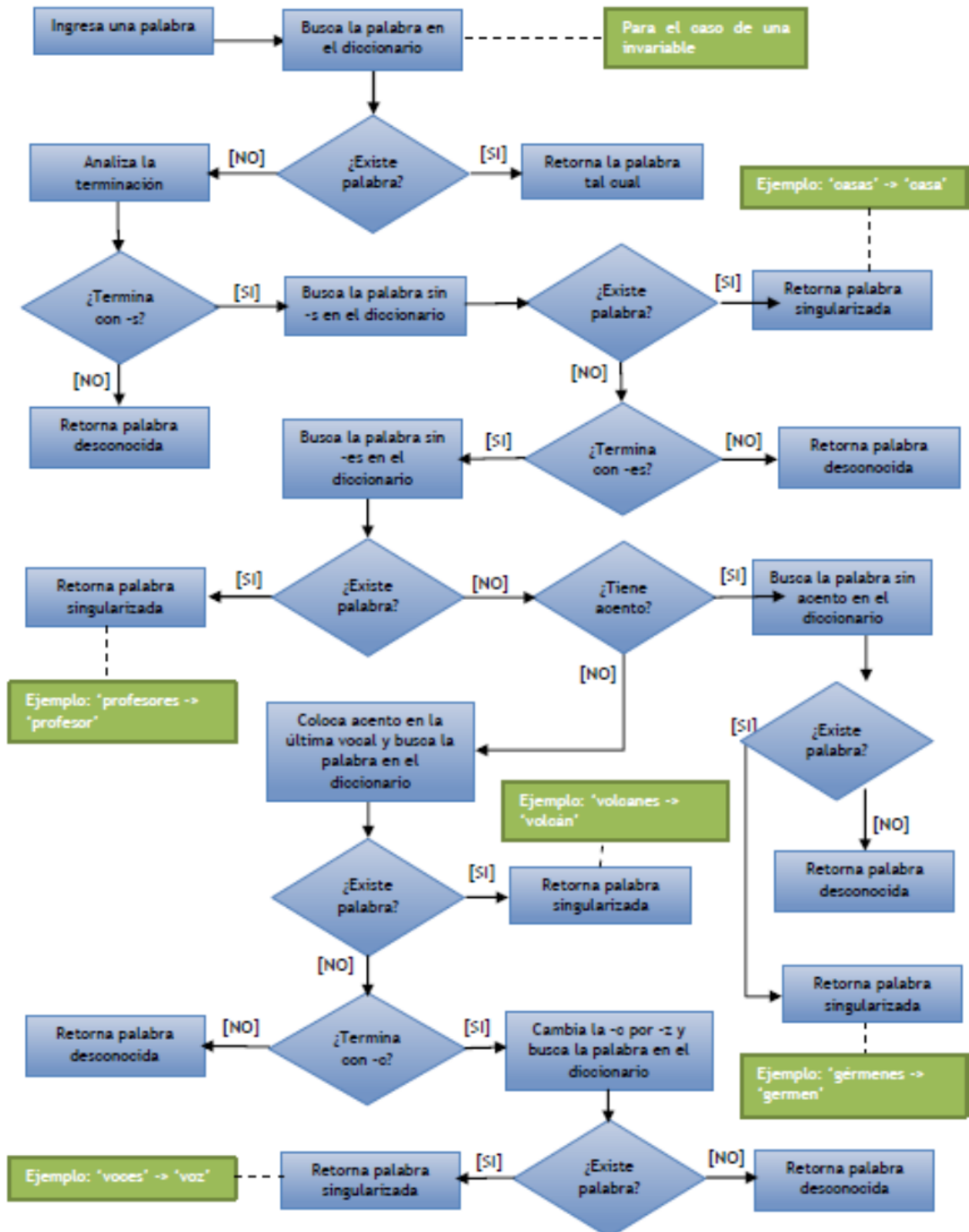
- quitar –s,
- quitar –es,
- quitar acento,
- poner acento en la última vocal,
- sustituir –c por –z.

En cada operación se realiza una búsqueda en el diccionario, el cual contiene todas aquellas palabras susceptibles de recibir morfemas (fragmento mínimo capaz de expresar su significado) de plural.

Quedan excluidas por tanto, las formas verbales (menos los participios), las preposiciones, las conjunciones, los adverbios, y cualquier otra palabra invariable que no admita plural. En este último grupo se encuentran los sustantivos invariables como: martes, oasis, crisis, tesis, y otros como tez, fénix, salud, etc. que no tienen forma plural.

A continuación, se presenta el **Diagrama de flujo** a fin de detallar el proceso de singularización arriba descrito.

Diagrama N°2. Diagrama de flujo del procedimiento de singularización



Fuente: Elaboración propia

A continuación se presenta el **Código en Java del proceso de singularización**:

```
package Conect;
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.PreparedStatement;
import java.sql.ResultSet;
import java.util.ArrayList;
import java.util.List;
import java.util.logging.Level;
import java.util.logging.Logger;
import java.sql.*;
/**
 *
 * @author Mathias Diaz
 */
public class Singularizacion
{
    static final String url = "jdbc:mysql://localhost:3306/chat";

    public static String Singulariza(String palabra)
    {
        String singular;
        String sinS;
        String sinE;
        String sinACNT;
        String conACNT;
        String cXz;
        char[] cadena = palabra.toCharArray();
        int ult = palabra.length();

        if(cadena[ult-1]=='s')
        {
            char[] cadena_sinS = new char[cadena.length-1];
            for(int i = 0; i<cadena.length-1; i++)
            {
```

```
        cadena_sinS[i]=cadena[i];
    }
    sinS = String.valueOf(cadena_sinS);
    //BUSCO LA PALABRA SIN S EN EL DICCIONARIO
    try
    {
        Class.forName("com.mysql.jdbc.Driver");
        Connection con = DriverManager.getConnection(url, "root", "");
        String consulta = "SELECT * FROM palabras where palabra='"+sinS+"'";
        PreparedStatement ps = con.prepareStatement(consulta);
        ResultSet rs = ps.executeQuery(consulta);
        rs.next();
        if(rs.getRow() != 0)//EXISTE LA PALABRA SIN S
        {
            singular = rs.getString("palabra");
            return singular;
        }
        else //NO EXISTE LA PALABRA SIN S
        {
            char[] cadena_sinE = new char[cadena_sinS.length-1];
            for(int i = 0; i<cadena_sinS.length-1; i++)
            {
                cadena_sinE[i]=cadena_sinS[i];
            }
            sinE = String.valueOf(cadena_sinE);
            //BUSCO LA PALABRA SIN S EN EL DICCIONARIO
            consulta = "SELECT * FROM palabras where palabra='"+sinE+"'";
            ps = con.prepareStatement(consulta);
            rs = ps.executeQuery(consulta);
            rs.next();
            if(rs.getRow() != 0)//EXISTE LA PALABRA SIN E
            {
                singular = rs.getString("palabra");
                return singular;
            }
            else//BUSCO ALGUNA VOCAL CON ACENTO EN LA PALABRA Y LA QUITO
```

```
{
char[] cadena_sinACNT = sinE.toCharArray();
for(int i = 0; i<cadena_sinACNT.length; i++)
{
switch(cadena_sinACNT[i])
{
case 'á': cadena_sinACNT[i]='a'; break;
case 'é': cadena_sinACNT[i]='e'; break;
case 'í': cadena_sinACNT[i]='i'; break;
case 'ó': cadena_sinACNT[i]='o'; break;
case 'ú': cadena_sinACNT[i]='u'; break;
}
}
sinACNT = String.valueOf(cadena_sinACNT);
//BUSCO LA PALABRA SIN ACENTO EN EL DICCIONARIO
consulta = "SELECT * FROM palabras where palabra='"+sinACNT+"'";
ps = con.prepareStatement(consulta);
rs = ps.executeQuery(consulta);
rs.next();
if(rs.getRow() != 0)//EXISTE LA PALABRA SIN ACENTO
{
singular = rs.getString("palabra");
return singular;
}
else//COLOCO EL ACENTO EN LA ULTIMA VOCAL Y BUSCO ESTA PALABRA
{
char[] cadena_conACNT = sinE.toCharArray();
ult = sinE.length();
int b = 0;
for(int i = ult-1; i>=0; i--)
{
switch(cadena_conACNT[i])
{
case 'a': cadena_conACNT[i]='á'; b = 1; break;
case 'e': cadena_conACNT[i]='é'; b = 1; break;
case 'i': cadena_conACNT[i]='í'; b = 1; break;
```



```
        case 'o': cadena_conACNT[i]='ó'; b = 1; break;
        case 'u': cadena_conACNT[i]='ú'; b = 1; break;
    }
    if(b==1){break;}
}
conACNT = String.valueOf(cadena_conACNT);
//BUSCO LA PALABRA CON EL ACENTO EN EL DICCIONARIO
consulta = "SELECT * FROM palabras where palabra='"+conACNT+"'";
ps = con.prepareStatement(consulta);
rs = ps.executeQuery(consulta);
rs.next();
if(rs.getRow() != 0)//EXISTE LA PALABRA CON EL ACENTO
{
    singular = rs.getString("palabra");
    return singular;
}
else//SUSTITUIR C POR Z
{
    char[] cadena_cXz = sinE.toCharArray();
    ult = sinE.length();
    if(cadena_cXz[ult-1]=='c')
    {
        cadena_cXz[ult-1]='z';
    }
    cXz = String.valueOf(cadena_cXz);
    //BUSCO LA PALABRA CON C X Z EN EL DICCIONARIO
    consulta = "SELECT * FROM palabras where palabra='"+cXz+"'";
    ps = con.prepareStatement(consulta);
    rs = ps.executeQuery(consulta);
    rs.next();
    if(rs.getRow() != 0)//EXISTE LA PALABRA CON LA C POR LA Z
    {
        singular = rs.getString("palabra");
        return singular;
    }
    else// NO QUEDAN MAS OPCIONES SE RETORNA LA PALABRA TAL CUAL ES
```

```
        {  
            return palabra;  
        }  
    }  
}  
}  
}  
}  
}  
catch (Exception ex) {}  
}  
return palabra;  
}  
}
```

En resumen: el proceso de singularización descripto permite luego validar el tipo gramatical de una palabra. Para ello hay que realizar una consulta al diccionario, y este no es más que un listado de todas las palabras en singular.

c) El proceso sigue ahora con la función denominada “**Validador**”, que tiene como entrada el arreglo resultante de la “Singularización”.

Esta función realiza la validación gramatical de las palabras del arreglo, descartando conectores, artículos, adjetivos, verbos, preposiciones, adverbios y aposiciones.

Esto lo hace recorriendo el arreglo y buscando la definición de cada palabra en el diccionario.

El diccionario que se utiliza para hacer las consultas, es el perteneciente a la Real Academia Española (RAE).

Se obtiene como resultado un arreglo que representará la estructura de conocimiento del artículo analizado (2º macro-regla de T. van Dijk).

d) Por último, se ejecuta la función “**GeneraOntología**”.

Esta función obtiene como entrada el arreglo de palabras final, resultante del paso anterior, y realiza dos tareas:

1. La primera: guardar, sin repetir, cada palabra del arreglo en el campo “palabra” de la tabla de la base de datos “Arreglo Ontologico”, haciendo persistente estos datos, para su posterior uso.
2. La segunda: coloca un peso a cada palabra, en el campo denominado con el mismo nombre de la tabla mencionada. Este peso determina la importancia de la palabra en la información cargada.

El cálculo de este valor, depende del lugar de origen de la palabra, para ello se toma el siguiente criterio de pesos:

- Si la palabra tiene origen en palabras claves: peso = 0,6
- Si la palabra tiene origen en título: peso = 0,4
- Si la palabra tiene origen en abstract: peso = 0,3
- Si la palabra tiene origen en objetivos: peso = 0,3
- Si la palabra tiene origen en documento: peso = 0,05

Por cada repetición de la palabra en se aumenta el peso en 0,05, tomando como máximo valor 1.

El resultado obtenido varía entre 0 (poco pertinente) a 1 (muy pertinente).

Se obtiene entonces como resultado de la función la estructura de referencia generada para la temática y almacenada en la base de datos.

En la siguiente imagen se muestra la interfaz de carga de datos adicionales por parte del docente para lograr una personalización del sistema en la configuración de la sesión.

Imagen 2: Interfaz del segundo paso en el proceso de configuración de una sala de chat

Configuración del artículo a tratar en la sesión de chat

Seleccione las frases de intervención:

Puede agregar otras frases si desea:

Según lo escrito por sus alumnos:
¿Cuándo intervendría usted, si considero que se están desviando de tema?
Teniendo en cuenta las siguientes palabras:

Consignas de trabajo:

Fecha de comienzo:

Hora de comienzo y duración:

Alumnos por favor no se desvíen del tema del artículo
 Alumnos creo que la conversación no es adecuada a la temática de la clase
 La conversación no tiene nada que ver con el artículo. Por favor vuelvan al tema
 La discusión está fuera de contexto, hacer uso del chat solo para tratar acerca del artículo

tecnología, formación, educación
 conectividad, pedagogía, obstáculo
 proyecto, capacitación, formulación
 dicho, debe, taller
 haber, hallazgo, ruido

Comienzo:
 00 : 00
 Duración:
 00 : 00

En la imagen anterior se muestra cómo se establece esta interfaz:

- Las frases que utilizará el sistema para intervenir cuando la temática de la conversación en el chat se desvíe de la estructura de conocimiento, pudiendo personalizar con más frases de preferencia del docente.
- El punto de intervención de acuerdo al criterio del Docente: el sistema debe detectar cuál es el grado de tolerancia del profesor en el contexto que se analizará en el Chat, éste punto también hace referencia al punto de corte o punto de intervención, que necesita el sistema para poder intervenir.

Para esto, se seleccionan 5 grupos de 3 de palabras, ordenadas de acuerdo a su peso en la estructura de referencia.

Las palabras del primer grupo serán las de mayor peso, las del segundo grupo de menor peso, y así hasta el último grupo que corresponde a palabras de menor peso. De acuerdo al criterio del profesor, el grupo de palabras elegido determinará de cierta forma la tolerancia que tendrá para la conversación, esta tolerancia será traducida en un porcentaje dependiendo de la selección del profesor, de modo que si selecciona el primer grupo (con las palabras de mayor peso) la tolerancia será la

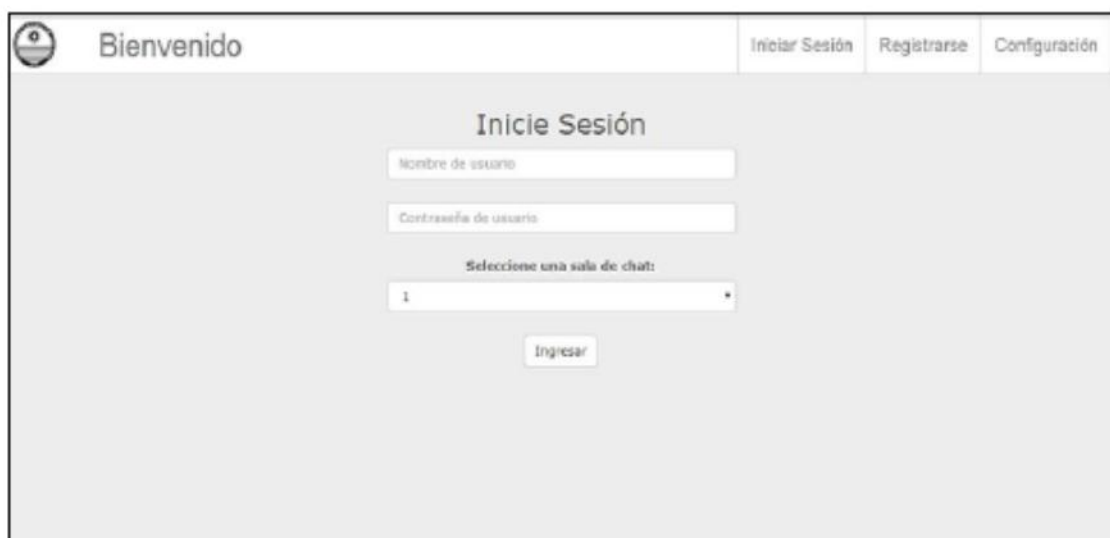
más estricta, casi sin dejar que los participantes se desvíen en ningún momento, con este criterio, si selecciona el siguiente grupo de palabras, el porcentaje de tolerancia para intervenir irá disminuyendo.

- Las consignas de trabajo, para que los alumnos en la sesión de chat las tengan presente, estas consignas podrán ser visibles para los participantes durante la sesión del Chat.
- Fecha, hora y duración de la sesión de chat, esto se debe a que el chat es un medio de comunicación sincrónico, y como tal se debe sincronizar el acceso de los alumnos al chat.

4. Se configuró la sesión de chat

Para ingresar al sistema, los participantes iniciar la sesión de chat.

Imagen 3: Interfaz de login de alumnos



La primera vez, previo al inicio de la sesión, los participantes deben registrarse tal como se muestra en la Imagen N° 3.

Cada vez que un participante de la sesión de chat, envía un mensaje el cual será guardado en la base de datos en la tabla "MensajesPúblicos".

Cabe aclarar que el sistema solo reconoce palabras bien formadas, es decir que abreviaturas, errores de ortografía y regionalismos serán consideradas como palabras desconocidas. Además al iniciar el chat se establecen cuatro puntos temporales basados en la configuración, que impactan sobre el “Moderador”:

1. En un primer momento, cuando el chat comienza a la hora establecida el sistema envía un mensaje dando la bienvenida a todos los participantes y aclarando que tienen tres minutos para poder presentarse, a partir de ahí comienzan a correr esos minutos en donde el Moderador no realizará ningún análisis sobre lo escrito por los participantes, ya que se corresponde a la presentación normal antes de iniciar una conversación.
2. Seguido a esto, una vez finalizado estos tres minutos, el sistema envía un segundo mensaje avisando que ya pueden comenzar a trabajar sobre la temática establecida en el artículo, aclarando que se remitan a las consignas de trabajo presentes en la sala en caso de tener alguna duda. A partir de aquí, el Moderador queda habilitado para poder analizar todo lo que se escriba durante la conversación.
3. Cuando restan tres minutos para la finalización de la sesión de chat, el sistema avisan este evento, mediante un mensaje que se envía a todos los estudiantes. Se deja de moderar para la despedida.
4. Cuando el chat llega al punto de fin establecido, el sistema envía un último mensaje y bloque la sesión para la sala.

El funcionamiento del moderador para detectar las desviaciones considera el llenado de una ventana temporal de palabras.

Esta ventana temporal es un tipo de dato abstracto cola (primero en llegar, primero en salir), almacenada en la base de datos en una tabla con el nombre de “Ventana”.

Esta ventana se diseño para tener presente siempre los mensajes actuales y que a medida que van ingresando a la ventana mensajes nuevos, estos desplazan a los más antiguos.

Su función es almacenar, temporalmente, las palabras que se vayan ingresando en los mensajes del chat, para que el moderador haga uso de esta estructura luego para el cálculo del grado de similitud.

Cada vez que un participante envía un mensaje el Moderador toma el mensaje y lo analiza con las siguientes dos funciones:

- “ControlaFrecuencias”: tiene como objetivo rellenar la “Ventana Temporal” de palabras.

Para ello, cada vez que un alumno escribe y envía un mensaje en la sala de chat, esta función lo toma como entrada, y procesa cada palabra del mismo, aplicando dos de las funciones que se describieron para el Analizador.

Primero toma las palabras del mensaje y depura los signos de puntuación, números y símbolos especiales, y luego singulariza estas palabras, según corresponda, para luego finalmente agregarlas a la cola, que se mencionaron anteriormente.

Esta función obtiene como resultado de su ejecución elementos (palabras) ingresando a la cola temporalmente, mientras se escriban mensajes en el chat.

- La siguiente función es “VerificaConversación”, que se ejecuta cada tres minutos, teniendo en cuenta la configuración temporal que realizó el profesor para la sala chat.

La función realiza el cálculo del grado de similitud.

El cálculo del grado de similitud que realiza el Moderador fue inspirado por la función matemática del coseno de un ángulo, y el artículo “La similitud léxico-semántica en artículos de investigación científica en español: Una aproximación desde el Análisis Semántico Latente” de René Venegas (VENEGAS, 2006 pag.11).

Venegas (2006) analiza y aplica la similitud entre vectores (palabras), ésta es calculada usando medidas de coseno, cuyos valores van de 1 para vectores con la misma dirección (esto significa que lo medido es igual) a 0 para aquellos vectores ortogonales (perpendiculares en el espacio multivectorial, es decir, que lo medido es completamente distinto)

El cálculo de la similitud se realiza del siguiente modo:

- primero controla la condición de que la Ventana Temporal esté llena.
- Si esta condición es verdadera, toma las palabras que estén en ese momento en la Ventana Temporal y guarda en un arreglo las mismas.
- Luego para cada palabra del arreglo verifica si se encuentra en el Arreglo Ontológico, si es así, cuenta la frecuencia de esta palabra en la ventana,

obtiene el peso de la palabra en dicho arreglo y realiza el cálculo del grado de similitud.

- Repite este proceso para cada palabra (sin repetir las anteriormente analizadas) y finalmente obtiene el promedio de estos cálculos, que será el valor final del grado de similitud, a continuación se darán detalles de este proceso.

Primero se calcula la distancia: la función matemática que se usa para este cálculo es la siguiente:

$$distancia = \frac{frec * peso}{\sqrt{frec^2 - peso^2}}$$

Al resultado de la distancia de la fórmula se lo divide por el peso de la palabra para poder limitar el valor que pueda tomar, y a ese valor se lo divide por la cantidad de elementos (N) de la ventana.

La función matemática que se utiliza para calcular el valor de similitud es la siguiente:

$$similitud = \frac{\frac{distancia}{peso}}{N}$$

Al valor de similitud se lo guarda en un sumador, y el proceso es igual para cada palabra.

$$sumador += similitud.$$

Pero si una palabra se repite en la ventana, y ya fue procesada anteriormente no se vuelve a procesar, solo se aumenta un contador de “aciertos” de palabras de la estructura de conocimiento, presentes en dicha ventana.

Cuando se ha procesado toda la ventana, se obtiene el promedio de estos valores guardados en el sumador, dividido en la cantidad de palabras que poseen el valor de similitud, y ese valor será multiplicado por la cantidad de elementos (contador) de la ventana que correspondían a cada una de estas palabras, es decir la cantidad de aciertos de palabras del arreglo que hay en la ventana.

$$\text{Grado de similitud final} = \frac{\text{sumador}}{\text{cant}} * \text{aciertos}$$

Finalmente, se obtiene como resultado un valor de grado de similitud para la conversación en ese momento. Cabe aclarar que este valor va a variar permanentemente conforme se desarrolle el chat entre los alumnos, pudiendo tomar diferentes valores, cercanos a cero si la conversación se desvía de la temática establecida en la estructura de conocimiento, o cercanos a uno si es más pertinente para la misma.

En base al resultado obtenido para el grado de similitud, el Moderador decide si la conversación se aleja del valor de tolerancia que se guardó cuando el profesor configuró el chat y de esa forma intervenir en la conversación, solicitando a los alumnos que retomen la temática planteada para esa sala de chat, utilizando de forma aleatoria alguna de las frases de intervención seleccionadas también por el profesor en la etapa de configuración.

El contenido de la siguiente imagen fue capturado en una de las pruebas realizadas al sistema, en el Colegio Superior N° 1 de Rawson.

Imagen N° 4: Interfaz de usuario alumno en una sala de chat del sistema.



A continuación, la siguiente Imagen N°5 presenta una Visualización de las consignas de trabajo para la sala en la misma escuela arriba mencionada. En la prueba realizada que muestra la captura, se trabajó con un artículo llamado “Tensiones y obstáculos de la integración de las tics a la educación”.

Imagen N° 5: Interfaz de usuario alumno en una sala de chat (con las consignas)



En la imagen siguiente, se observa que el sistema Moderador Automático de Chat, provee una interfaz para que el profesor si quisiera, pueda observar en tiempo real la conversación y ver como fluctúan los valores de la “Ventana Temporal” como así también del grado de similitud, permitiéndole verificar también que el Moderador se ejecute conforme a los ajustes realizados en la configuración.

Imagen N°6: Interfaz de usuario profesor (con los valores de similitud)



A MODO DE CIERRE:

En este capítulo se ha presentado la **Solución diseñada** para resolver las desviaciones en las conversaciones que los participantes realizan en sesiones de chat moderado para contextos educativos.

El punto central de la solución diseñada consiste en la conjunción de los aportes de la inteligencia artificial, la minería de datos y la aplicación de aspectos de la lingüística para construir un algoritmo que calcule el grado final de similitud, aspecto vital para resolver la intervención del sistema de modelación automática en el chat.

El “grado final de similitud” es un valor numérico que determina cuanto se desvía la conversación con respecto a la estructura de referencia, (que es la que contiene el conocimiento de la temática tratada), su cálculo se pudo realizar por medio de la combinación de métodos de distintas ramas de conocimiento, tales como la inteligencia artificial, la lingüística y la minería de datos.

Con respecto a la inteligencia artificial, su aporte consiste en que el sistema puede aprender a partir de un conjunto de datos que el docente guarda en el momento de la configuración del chat, y construir una estructura de conocimiento. Luego, a medida que transcurre la sesión, el sistema toma la información de los mensajes, la procesa y calcula el grado final de similitud con respecto a la base de conocimientos. Esto configura un conocimiento adquirido desde los mensajes que envían los estudiantes, para luego cotejarlo con el punto de intervención configurado por el docente.

Con respecto a la minería de datos, ella aporta la definición del proceso KDD. Dicho proceso proporciona las fases para obtener nuevo conocimiento a partir de los datos. En el KDD, se decidió en la fase de minería de datos:

- por un aprendizaje basado en instancias
- por la técnica de similitud basada en distancia
- por el cálculo de la métrica de distancia fue por medio del coseno.

Con respecto a la lingüística la aplicación de las dos primeras macro-reglas de T.van Dijk para el desmontaje del discurso, dan un método sistemático para el análisis documental para la comprensión del discurso.

CAPITULO V: Validación

Como ya se planteó en los capítulos anteriores., el trabajo central de esta tesis permitió el desarrollo de un software.

Dado que la tesis se vincula con una investigación más amplia que le da un contexto al desarrollo logrado, se consideró que una vez que el sistema de software se encontrara funcionando, era necesario comprobar o validar que lo hace ajustado a lo programado. Es decir, si cumple con las expectativas tanto de los diseñadores y como de los usuarios finales. Sobre todo, que el software, cuando se implemente, satisfaga las expectativas del usuario. En este caso, los docentes.

Para poder lograr esta comprobación, se uso un modelo para entender por qué los usuarios aceptan o usan las tecnologías: el Modelo de Aceptación Tecnológica propuesto por Davis.

Instrumento de validación

Para el caso particular de esta tesis, el algoritmo desarrollado sirvió para la construcción de un software que modera las conversaciones en salas de chat, en el marco de un proyecto más amplio. El desarrollo del software que se llevó a cabo, se validó a través de una Encuesta aplicada a alumnos y docentes, a fin de conocer si el sistema creado cumple con las expectativas de diseño e implementación para lo cual fue programado.

Existe una gran preocupación acerca de asegurar la adopción exitosa de tecnologías dentro de las organizaciones. Un modelo clave para lograr este objetivo ha sido el TAM (Modelo de Aceptación Tecnológica) propuesto por Davis, que es el más probado y validado de todos los existentes (Chuttur, 2009).

En la mayoría de las situaciones donde se desea conocer y validar el uso de una determinada herramienta tecnológica, el TAM ha confirmado su solidez y capacidad para predecir la adopción tecnológica entre los usuarios de una organización. El modelo de aceptación tecnológica, es el más aprobado por los estudiosos de las Tecnologías de la Información y comunicación, porque ha sido efectivo cuando se ha probado para predecir el uso de una tecnología en particular.

Este modelo se utiliza para predecir el uso de una tecnología, dos características principales:

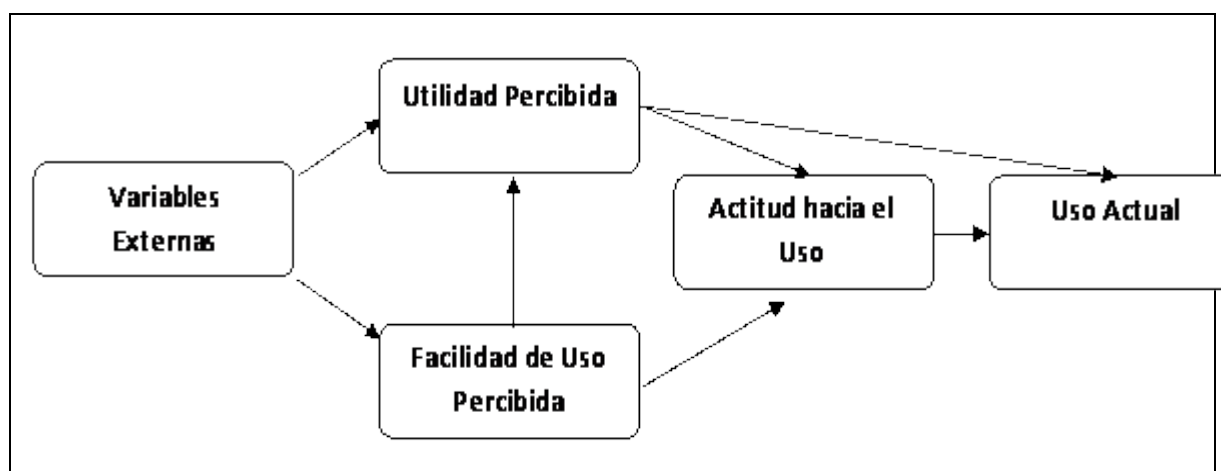
- Utilidad percibida (Perceived Usefulness-PU).
- Facilidad de uso percibida (Perceived Ease of Use-PEOU).

La utilidad percibida (PU) se refiere al grado en que una persona cree que usando un sistema en particular mejorará su desempeño en el trabajo.

La facilidad de uso percibida (PEOU) señala hasta qué grado una persona cree que usando un sistema en particular realizará menos esfuerzo para desempeñar sus tareas.

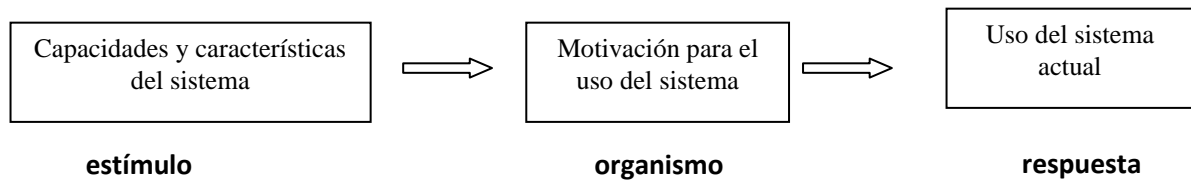
Según Davis, el propósito del TAM es explicar las causas de aceptación de las tecnologías por los usuarios (Chuttur, 2009). Ese modelo propone que las percepciones de un individuo en cuanto a la utilidad y la facilidad de uso percibidas de un sistema de información sean concluyentes para determinar su intención de usar un sistema. El modelo se muestra en la siguiente Figura.

Figura 1: Modelo de Aceptación Tecnológica propuesto por Davis(1989).



Fuente: Elaboración propia

El modelo TAM, tiene sus raíces en la Teoría de la Acción Razonada (TRA) de Ajzen y Fishbein (Chuttur, 2009). Para este modelo, las creencias están influidas por las actitudes, las cuales conducen a intenciones, resultando en algunos tipos de comportamiento. La TRA es una teoría general que trata de explicar casi todo tipo de comportamiento humano, partiendo de la importancia de las creencias individuales, con el fin de predecir la conducta humana. En el contexto de la aceptación tecnológica, esta teoría se utiliza para tratar de establecer los factores que condicionan al usuario hacia la innovación, la intención de comportamiento y la intensidad de uso del sistema (Chuttur, 2009). El modelo TAM propone que el uso del Sistema es una respuesta influenciada por constantes estímulos externos de las capacidades y características del sistema actual, esto se ilustra en la siguiente Figura.

Figura 2: Uso de un sistema según el Modelo de Aceptación Tecnológica.

Fuente: Elaboración propia

Se propone el Modelo TAM como instrumento de validación para el trabajo con el sistema el Moderador Automático de Chat (MAC) para evaluar el grado de aceptación que tienen los docentes y los estudiantes de educación superior al usar el sistema MAC, sistema propuesto como una tecnología a implementarse en el ámbito educativo en escenarios virtuales.

Los modelos del instrumento según TAM, para docente y estudiantes se presentan en la sección Anexos.

Para predecir el uso del sistema MAC, se utilizó al Modelo TAM como guía para la construcción del instrumento de validación. El mismo se ha subdividido en las siguientes características o dimensiones, para el perfil Docente. La siguiente Tabla N° 1 presenta dichas dimensiones y los enunciados del instrumento que corresponden a cada característica, para el perfil docente.

Tabla N° 1: Dimensiones o características y enunciados del Instrumento para docentes del MAC

Perfil: Docente		
Dimensión: Utilidad Percibida	Dimensión: Facilidad de Uso Percibida	Dimensión: Intención de Uso
1. Al usar el Chat con Moderador Automático, le permitió monitorear el debate de los alumnos más eficientemente en comparación como lo hacía utilizando otra herramienta.	3. Aprender a utilizar la herramienta le fue fácil.	5. Utilizaría esta herramienta cada vez que tenga que dictar un curso en una plataforma virtual.
2. Considera esta herramienta importante para favorecer el proceso-aprendizaje en sus alumnos.	4. Encuentra fácil el uso del Chat con Moderador Automático.	6. Utilizaría esta herramienta cada vez que tenga que gestionar el dictado de curso en una plataforma.
14. Considera que usar el Chat con Moderador Automático mejora el estudio de su asignatura.	8. El usar el Chat en el aula virtual no le representa ningún estrés.	7. Incluiría esta herramienta cada vez que tenga que gestionar el dictado de curso en una plataforma.
15. Encuentra muy útil el Chat con Moderador Automático.	9. Encuentra fácil que el haga lo que Usted quiere que haga.	20. Intenta usar el Chat para planear interacciones entre sus alumnos.
16. Usar el Chat con Moderador Automático, le facilita la tarea de monitorear las interacciones de los alumnos.	10. Es fácil para Usted, recordar cómo se realizan las tareas usando el Chat con Moderador Automático.	
17. Usar el Chat con Moderador Automático le permite aclarar dudas, enriquecer y fortalecer los conocimientos en sus alumnos.	11. Es fácil y entendible percibir las interacciones de los alumnos usando el Chat con Moderador Automático.	
18. Usar el Chat en el aula virtual mejora la calidad del trabajo que realiza en su asignatura.	12. Es fácil obtener la información que Usted quiera del Chat.	
19. El usar el Chat le permite desarrollar habilidades de Comunicación y/o manejo de conflictos.	13. Considera el Chat con Moderador Automático una herramienta fácil de usar en la interacción de los alumnos.	

Fuente: Elaboración propia.

Con el mismo Modelo TAM y de acuerdo al instrumento diseñado para predecir el uso del sistema MAC, se subdividió las siguientes características o dimensiones, para el perfil Alumno. La Tabla N° 2 presenta dichas dimensiones y los enunciados del instrumento que corresponden a cada característica, para el perfil Alumno.

Tabla N° 2: Dimensiones o características y enunciados del Instrumento para alumnos del MAC.**Perfil: Alumno**

Dimensión: Utilidad Percibida	Dimensión: Facilidad de Uso Percibida	Dimensión: Intención de Uso
1. Al usar el Chat con Moderador Automático, le permitió interactuar con sus compañeros más eficientemente en comparación como lo hacía utilizando otra herramienta.	3. Aprender a utilizar la herramienta le fue fácil.	5. Utilizaría esta herramienta cada vez que tenga que recibir un curso.
2. Considera esta herramienta importante para favorecer el proceso-aprendizaje.	4. Encuentra fácil el uso del Chat con Moderador Automático.	16. Intenta usar el Chat para planear interacciones entre sus compañeros en una materia.
11. Considera que usar el Chat con Moderador Automático mejora el estudio de una asignatura.	6. El usar el Chat no le representa ningún estrés.	
12. Encuentra muy útil el Chat con Moderador Automático.	7. Encuentra fácil que el Chat haga lo que Usted quiere que haga.	
13. Usar el Chat con Moderador Automático le permite aclarar dudas, enriquecer y fortalecer sus conocimientos de la materia.	8. Es fácil para Usted, recordar cómo se realizan las tareas usando el Chat con Moderador Automático.	
14. Usar el Chat mejora la calidad del trabajo que realiza en la asignatura.	9. Es fácil obtener la información que Usted quiera del Chat.	
15. El usar el Chat le permite desarrollar habilidades de Comunicación.	10. Considera el Chat con Moderador Automático una herramienta fácil de usar en la interacción con sus compañeros sobre una asignatura.	

Fuente: Elaboración propia.

Pruebas experimentales

Las pruebas experimentales usando el sistema MAC fueron dos.

1. En el ámbito de la UNSJ
2. En el ámbito de un colegio no universitario: el Colegio Superior N° 1 de Rawson

Ambos sitios fueron seleccionados intencionalmente por que los miembros del equipo de investigación forman parte de sus cuerpos docentes, razón por la cual estaba asegurada la accesibilidad a los docentes y alumnos de la prueba.

1. En el ámbito de la UNSJ

La primera prueba corresponde a **Docentes**, quienes fueron intencionalmente seleccionados dentro del Departamento de Informática e Instituto de Informática de la Facultad de Ciencias Exactas, Físicas y Naturales de la UNSJ. A los docentes seleccionados se les aplicó una Encuesta (Instrumento para Docentes) luego de haber usado el software.

La prueba se realizó en el mes de abril de 2014, y se registró en video. La cantidad total de participantes fue diez (10).

Los docentes fueron invitados y accedieron cordialmente a realizar la prueba.

La prueba se llevó a cabo de mañana y se destinó de una hora reloj (de 9 a 10 hs) para que su desarrollo.

A los docentes seleccionados se les asignó *roles*: de docentes y de alumnos.

Tres docentes asumieron un rol docente y el resto el rol de estudiantes de las Carreras Licenciatura en Sistemas de Información y Licenciatura en Ciencias de la Computación.

A los participantes con el rol de docente, se les entregó con anterioridad el documento con el cual debatirían los docentes con rol de estudiantes en la prueba. El artículo tratado fue: "*Comunicación y pedagogía. Tecnologías de la Información y la comunicación. Estrategias didácticas. Internet en educación del autor Jordi Adell*".

La prueba se desarrolló con normalidad y se destaca la muy buena predisposición para utilizar el Sistema y luego para completar el Instrumento de Validación.

Se utilizaron equipos portátiles personales de cada participante, para realizar la prueba. Se dispuso de un Punto de Acceso configurado de modo tal de proveer la conexión inalámbrica con el equipo servidor del sistema MAC, hacia los clientes. El sistema MAC, se podía ejecutar como una aplicación web y desde cualquier navegador: Internet Explorer, Mozilla Firefox y Chrome.

La muestra de docentes que desempeñaron el **rol de docentes** en la primera experiencia se caracterizó por:

- el 100% son mujeres
- el 100% tiene título universitario técnico
- el 50% tiene título docente
- el 100% trabaja con plataformas virtuales
- el 100% posee equipo informático
- el 100% tiene conexión a internet domiciliaria

Algunos rasgos de los docentes seleccionados - que desempeñaron el rol docente- son que registran:

- un nivel de experiencia medio en el uso de internet y TIC: el 50% registra un nivel intermedio, y el otro 50% un nivel de experto.
- un nivel medio de experiencia en el uso de chat: el 50% si usa chat
- un nivel medio de experiencia en el uso de chat como herramienta de apoyo en el proceso educativo: el 50% si usa chat para su actividad didáctica

La muestra de docentes que desempeñaron el *rol de alumnos* en la primera experiencia se caracterizó por:

- el 100% son mujeres
- el 100% tiene título universitario técnico
- el 70% trabaja con plataformas virtuales
- el 80% tiene conexión a internet domiciliaria

Los rasgos de los docentes seleccionados - que desempeñaron el rol de alumno- son que registran:

- un nivel de experiencia medio en el uso de internet y TIC: el 50% registra un nivel de iniciación, y el otro 50% un nivel de intermedio.
- un nivel alto de experiencia en el uso de chat: el 100% si usa chat
- un nivel bajo de experiencia en el uso de chat como herramienta de apoyo en el proceso educativo: sólo el 20% si usa chat para su actividad didáctica

Los docentes participantes realizaron sugerencias una vez finalizada la prueba. Ellas son:

- los docentes con rol docente:

- el sistema debería manejar familia de palabras (sinónimos y palabras relacionadas),
- se tendría que agregar la publicación de las consignas en el sistema.

- los docentes con rol de alumnos:

- el sistema podría publicar las consignas de trabajo,
- se podría mejorar la interfaz de usuario y permitir la presentación de los estudiantes sin que intervenga el sistema MAC,
- se podría mejorar el tiempo de respuesta: deja muchas pausas y lo escrito no se ve de modo inmediato en la ventana de texto,
- resaltar los llamados de atención del profesor
- introducir distinto color en las intervenciones de los participantes para diferenciarlos.

2. En el ámbito de un Instituto de enseñanza de Nivel Superior No Universitaria

La segunda prueba experimental se realizó en el Colegio Superior N° 1 de Rawson de San Juan.

Se realizó en el mes de junio de 2014, y se registró en video.

La cantidad total de participantes fue de treinta: diez fueron docentes y veinte fueron estudiantes de Segundo Año del Profesorado en Tecnologías. Todos ellos fueron invitados y donde accedieron muy cordialmente.

La prueba se llevó a cabo de tarde y se destinó de una hora reloj (de 18 a 19 hs), para su desarrollo.

Se les entregó, con anterioridad, a los docentes un documento con el cual debatirían los estudiantes en la prueba. El artículo tratado fue: *"Documento extraído de Formación Docente y TIC: Logros, tensiones y desafíos de 17 experiencias en América Latina. Magaly Robalino Campos. UNESCO Chile"*.

La prueba se desarrolló con normalidad y, al igual que la anterior prueba, se destaca la muy buena predisposición para utilizar el Sistema y luego para completar el Instrumento de Validación. Se utilizaron equipos portátiles del Programa Nacional Conectar Igualdad pertenecientes a la Biblioteca del Instituto y que fueron prestadas para realizar la prueba.

Se dispuso de un Punto de Acceso configurado de modo tal de proveer la conexión inalámbrica con el equipo servidor del sistema MAC, hacia los clientes. El sistema MAC, se podía ejecutar como una aplicación web y usando cualquier navegador: Internet Explorer, Mozilla Firefox y Chrome. Es preciso destacar, que para cuando se llevó a cabo esta segunda prueba piloto, las sugerencias que fueron enumeradas, tanto de docentes como de los alumnos en la primera prueba, fueron implementadas en su mayoría. Estas fueron mejoras al Sistema MAC. En la Tabla N° 3 se detalla lo que se mejoró para la segunda prueba experimental.

Tabla N°3: Implementaciones realizadas en base a las sugerencias

Cambios introducidos con base en las sugerencias	Implementado Si/No
El sistema maneja familia de palabras (sinónimos y palabras relacionadas).	No
Publicación de las consignas en el sistema.	Si
Mejor interfaz de usuario.	Si
Permite la presentación de los estudiantes sin que intervenga el sistema MAC.	Si
Mejor el tiempo de respuesta para no dejar tantas pausas.	Si
Resalta los llamados de atención del profesor se.	No
Distinto color de las intervenciones de los participantes para diferenciarlos.	Si

Fuente: Elaboración propia.

La **muestra de docentes** (diez) en la segunda experiencia se caracterizó por:

- el 100% son mujeres
- el 100% tiene título universitario
- el 80% tiene título docente
- el 50% trabaja con plataformas virtuales
- el 100% posee equipo informático
- el 100% tiene conexión a internet domiciliaria
- un nivel de experiencia medio en el uso de internet y TIC: el 40% registra un nivel de iniciación, y el otro 50% un nivel intermedio.
- un nivel bajo de experiencia en el uso de chat: el 20% usa chat y el 80% no.
- un nivel bajo de experiencia en el uso de chat como herramienta de apoyo en el proceso educativo: sólo el 10% si usa chat para su actividad didáctica

La muestra **de alumnos** (20 alumnos) en la segunda experiencia se caracterizó por:

- el 60% son mujeres
- el 100% no tiene título universitario
- el 80% trabaja con plataformas virtuales
- el 60% tiene conexión a internet domiciliaria

Los rasgos de los alumnos es que registran:

- un nivel de experiencia bajo en el uso de internet y TIC: el 80% registra un nivel de iniciación, y el otro 20% un nivel de intermedio.
- un nivel alto de experiencia en el uso de chat: el 80% si usa chat
- nulo nivel de experiencia en el uso de chat como herramienta de apoyo en el proceso educativo: ninguno usa chat para su actividad didáctica

Las sugerencias que se detectaron es la segunda prueba son:

- los docentes proponen:
 - que el sistema pueda controlar que alumno no ha participado e intervenga,
 - que el sistema pueda evaluar la participación de los alumnos.
- los alumnos proponen:
 - que se mejore más la interfaz de usuario,
 - que se permita subir fotos y utilizar emoticones,
 - mejorar el tiempo de envío de mensajes.

RESULTADOS DE LAS PRUEBAS

1. PARA DOCENTES

A continuación se presenta el análisis estadístico de los resultados de una Encuesta aplicada a 14 docentes participantes en ambas pruebas experimentales llevadas a cabo. La encuesta buscó captar la opinión sobre el sistema MAC (Ver modelo del instrumento en Anexos).

Los enunciados de la encuesta se agruparon en 3 bloques o dimensiones: **Utilidad Percibida**, **Facilidad de Uso Percibida** e **Intención de Uso**.

Para llevar a cabo el procesamiento estadístico se utilizó el software SPSS 11.5.

Primeramente se midió la fiabilidad del instrumento diseñado, para ello, se utilizó como medida de fiabilidad el coeficiente “alfa de Cronbach” (Cronbach, 1951), que es un índice de consistencia interna que toma valores entre 0 y 1 y que sirve para comprobar si el instrumento que se está evaluando recopila información defectuosa, que llevaría a conclusiones equivocadas; o si se trata de un instrumento fiable que hace mediciones estables y consistentes. Se considera una fiabilidad respetable a partir de 0,70.

En función de los datos obtenidos, el coeficiente es: $\text{Alpha} = 0,9466$.

A partir de este valor se puede afirmar que el cuestionario es fiable, es decir, mide lo que se pretende medir.

Seguidamente, se realizó un análisis estadístico para cada dimensión o característica del instrumento. La metodología aplicada fue:

- a. Cada uno de las dimensiones fueron resumidas en una variable resumen, que es la mediana de los resultados que integran cada bloque.
- b. Para cada dimensión se calculó el alfa de Cronbach.
- c. Como existen diversos factores que afectan a la fiabilidad, se debe analizar si los ítems o enunciados que la integran están bien ubicados o no y además, si es necesario que queden. Esto se puede lograr utilizando correlaciones parciales, en este contexto se las denomina correlaciones artículo-total. Se considera que una correlación artículo-total $\geq 0,5$ es suficiente.

A continuación se mostrará el análisis realizado para cada una de las dimensiones que fueron tenidas en cuenta en el instrumento y respondiendo al modelo TAM seleccionado y aplicado.

Dimensión 1: Utilidad percibida.

Este bloque abarca los enunciados 1, 2, 14, 15, 16, 17, 18 y 19 de la Encuesta, los cuales se presentan a continuación:

1. Al usar el Chat con Moderador Automático, le permitió monitorear el debate de los alumnos más eficientemente en comparación como lo hacía utilizando otra herramienta.
2. Considera esta herramienta importante para favorecer el proceso-aprendizaje en sus alumnos.
14. Considera que usar el Chat con Moderador Automático mejora el estudio de su asignatura.
15. Encuentra muy útil el Chat con Moderador Automático.
16. Usar el Chat con Moderador Automático, le facilita la tarea de monitorear las interacciones de los alumnos.
17. Usar el Chat con Moderador Automático le permite aclarar dudas, enriquecer y fortalecer los conocimientos en sus alumnos.

18. Usar el Chat en el aula virtual mejora la calidad del trabajo que realiza en su asignatura.

19. El usar el Chat le permite desarrollar habilidades de Comunicación y/o manejo de conflictos.

De los ítems anteriores, se resumieron en una nueva variable denominada *Utilidad percibida*, la cual fue encontrada como la mediana entre los enunciados que integran el bloque.

Para esta dimensión se obtuvo Alpha = 0,8771.

A continuación se analizaron las correlaciones artículo-total, que informan el grado en que cada enunciado o artículo está unido a la puntuación general de la dimensión con la que deberá estar altamente conectado.

Si todos los artículos o enunciados están diseñados para medir la misma dimensión, deben estar correlacionados positivamente entre sí y se espera ver altas correlaciones artículo-total (valores cercanos a 1). Se considera que una correlación artículo-total $\geq 0,5$ es suficiente.

En la Tabla N° 4, se presentan los cálculos de la matriz de correlaciones teniendo en cuenta la dimensión: Utilidad Percibida.

Tabla N° 4: Matriz de correlaciones para la dimensión Utilidad Percibida

	1	2	14	15	16	17	18	19	UTILIDAD PERCIBIDA
1	1,000	0,224	0,528	0,301	0,611	0,444	0,382	0,818	0,640
2	0,224	1,000	0,754	0,291	0,465	0,767	0,828	0,322	0,808
14	0,528	0,754	1,000	0,159	0,528	0,689	0,940	0,498	0,820
15	0,301	0,291	0,159	1,000	0,139	0,170	0,159	0,434	0,521
16	0,611	0,465	0,528	0,139	1,000	0,702	0,382	0,320	0,686
17	0,444	0,767	0,689	0,170	0,702	1,000	0,678	0,225	0,913
18	0,382	0,828	0,940	0,159	0,382	0,678	1,000	0,489	0,808
19	0,818	0,322	0,498	0,434	0,320	0,225	0,489	1,000	0,571
UTILIDAD PERCIBIDA	0,640	0,808	0,820	0,521	0,686	0,913	0,808	0,571	1,000

Fuente: Elaboración propia con base a la Encuesta aplicada a docentes. San Juan. 2014.

Se presenta a continuación la tabla de frecuencias del variable resumen para el bloqueo dimensión: Utilidad percibida.

Tabla N 5: Distribución de frecuencias de la variable Utilidad Percibida

	Frecuencia	Porcentaje
3,50	2	14,3
4,00	7	50,0
4,50	1	7,1
5,00	4	28,6
Total	14	100,0

Fuente: Elaboración propia con base a la Encuesta aplicada a docentes. San Juan. 2014.

Se observa que el 85,7% (suma de las 3 últimas filas) de los encuestados docentes cree que el Sistema de Moderador Automático de Chat (MAC) mejorará su desempeño y su intervención pedagógica en ambientes virtuales de enseñanza-aprendizaje y logrará que sus estudiantes no se desvíen del tema y aprovechen tiempo de debate utilizando la herramienta Chat (lo que corresponde a la puntuación 4 y 5) con la utilidad percibida. También se puede destacar que hay 0 % para los valores restantes.

Dimensión 2: Facilidad de uso percibida

Esta dimensión abarca los enunciados 3, 4, 8, 9, 10, 11, 12 y 13 de la encuesta, los cuales se presentan a continuación:

3. Aprender a utilizar la herramienta le fue fácil.
4. Encuentra fácil el uso del Chat con Moderador Automático.
8. El usar el Chat en el aula virtual no le representa ningún estrés.
9. Encuentra fácil que el haga lo que Usted quiere que haga.
10. Es fácil para Usted, recordar cómo se realizan las tareas usando el Chat con Moderador Automático.
11. Es fácil y entendible percibir las interacciones de los alumnos usando el Chat con Moderador Automático.
12. Es fácil obtener la información que Usted quiera del Chat.

13. Considera el Chat con Moderador Automático una herramienta fácil de usar en la interacción de los alumnos.

Estos ítems se resumieron en una nueva variable denominada **Facilidad de uso percibida**, la cual fue encontrada como la mediana entre los enunciados que integran el bloque.

Para esta dimensión se obtuvo $\text{Alpha} = 0,8960$. Se muestran a continuación la matriz de correlaciones y la tabla de frecuencias de la variable resumen para este bloque (o dimensión).

En la siguiente Tabla se presentan los cálculos de la matriz de correlaciones teniendo en cuenta la dimensión: Facilidad de uso percibida.

Tabla N° 6: Matriz de correlaciones teniendo en cuenta la dimensión: Facilidad de uso percibida

	3	4	8	9	10	11	12	13	FACILIDAD DE USO PERCIBIDA
3	1,000	0,616	0,464	0,347	0,747	0,602	0,534	0,585	0,723
4	0,616	1,000	0,608	0,475	0,747	0,236	0,680	0,585	0,825
8	0,464	0,608	1,000	0,743	0,391	0,591	0,732	0,616	0,851
9	0,347	0,475	0,743	1,000	0,347	0,524	0,601	0,625	0,653
10	0,747	0,747	0,391	0,347	1,000	0,155	0,496	0,277	0,657
11	0,602	0,236	0,591	0,524	0,155	1,000	0,400	0,671	0,530
12	0,534	0,680	0,732	0,601	0,496	0,400	1,000	0,626	0,798
13	0,585	0,585	0,616	0,625	0,277	0,671	0,626	1,000	0,686
FACILIDAD DE USO PERCIBIDA	0,723	0,825	0,851	0,653	0,657	0,530	0,798	0,686	1,000

Fuente: Elaboración propia con base a la Encuesta aplicada a docentes. San Juan. 2014.

Se presenta a continuación la tabla de frecuencias de la variable resumen para el bloque dimensión: Facilidad de uso percibida.

Tabla N° 7: Distribución de frecuencias de la variable Facilidad de uso percibida

	Frecuencia	Porcentaje
3,50	2	14,3
4,00	5	35,7
4,50	1	7,1
5,00	6	42,9
Total	14	100,0

Fuente: Elaboración propia con base a la Encuesta aplicada a docentes. San Juan. 2014.

Se observa que el 85,7% (suma de las 3 últimas filas) de los encuestados docentes, opinan que usando el sistema de Moderador Automático de Chat, realizará menos esfuerzo para desempeñar sus tareas. Es decir que podrá proponer con mayor facilidad situaciones pedagógicas de debate entre sus estudiantes sin que éstos se desvíen del tema que los convoca utilizando el sistema MAC (lo que corresponde a la puntuación 4 y 5) con la facilidad de uso percibida.

Dimensión 3: Intención de uso

Este bloque abarca los enunciados 5, 6, 7 y 20 de la encuesta, los cuales se presentan a continuación:

5. Utilizaría esta herramienta cada vez que tenga que dictar un curso en una plataforma virtual.
6. Utilizaría esta herramienta cada vez que tenga que gestionar el dictado de curso en una plataforma.
7. Incluiría esta herramienta cada vez que tenga que gestionar el dictado de curso en una plataforma.
20. Intenta usar el Chat para planear interacciones entre sus alumnos.

Los ítems anteriores, se resumieron en una nueva variable denominada **Intención de uso**, la cual fue encontrada como la mediana entre los enunciados que integran el bloque. Para esta dimensión se obtuvo $\text{Alpha} = 0,9453$

Se presentan a continuación la matriz de correlaciones en la siguiente Tabla, que corresponde a la dimensión Intención de uso.

Tabla N° 8: Matriz de correlaciones teniendo en cuenta la dimensión: Intención de Uso.

	5	6	7	20	INTENCIÓN DE USO
5	1,000	0,155	0,496	0,277	0,672
6	0,155	1,000	0,400	0,671	0,550
7	0,496	0,400	1,000	0,626	0,684
20	0,277	0,671	0,626	1,000	0,566
INTENCIÓN DE USO	0,672	0,550	0,684	0,566	1,000

Fuente: Elaboración propia con base a la Encuesta aplicada a docentes. San Juan. 2014.

Seguidamente, se presenta la tabla de frecuencias de la variable resumen para este bloque (o dimensión) denominada Intención de Uso.

Tabla N° 9: Distribución de frecuencias de la variable Intención de Uso.

	Frecuencia	Porcentaje
3,00	3	21,4
3,50	1	7,1
4,00	4	28,6
4,50	1	7,1
5,00	5	35,7
Total	14	100,0

Fuente: Elaboración propia con base a la Encuesta aplicada a docentes. San Juan. 2014.

Se observa que el 71,4% (suma de las 3 últimas filas) de los encuestados docentes poseen una actitud positiva de uso con respecto al sistema de Moderador Automático de Chat en ambientes didáctico-pedagógicos virtuales (lo que corresponde a la puntuación 4 y 5) con la intención de uso. Es preciso aclarar que la dimensión de Intención de Uso, es consecuencia de las características o dimensiones de Utilidad Percibida y Facilidad de uso Percibida, las cuales brindan valores en los cuales se puede concluir que existiría una Intención de Uso del sistema mayor al 70 % de los docentes encuestados.

2. PARA ALUMNOS

A los estudiantes de enseñanza de Nivel Superior Universitaria y Superior No Universitaria que utilizan Plataformas Virtuales Educativas y sus herramientas para su formación se les aplicó una Encuesta (Ver modelo del instrumento en Anexos). La cantidad de alumnos encuestados fueron treinta.

De idéntico modo que la Encuesta aplicada a docentes, los enunciados de la encuesta se agruparon en 3 bloques o dimensiones: **Utilidad Percibida, Facilidad de Uso Percibida e Intención De Uso**. Respecto a la fiabilidad del instrumento, dado los datos obtenidos, el coeficiente Alpha es 0,9203. Dicho valor permite afirmar que el cuestionario es fiable, es decir, mide lo que se pretende medir y para lo cual fue diseñado.

Para el procesamiento estadístico se utilizó el software SPSS 11.5.

Al igual que en el instrumento de los docentes, se analizaron las correlaciones artículo-total en el instrumento de los estudiantes, que informan el grado en que cada enunciado o artículo está unido a la puntuación general de la dimensión con la que deberá estar altamente conectado. Si todos los artículos o enunciados están diseñados para medir la misma dimensión, deben estar correlacionados positivamente entre sí y se espera ver altas correlaciones artículo-total (valores cercanos a 1). Se considera que una correlación artículo-total $\geq 0,5$ es suficiente.

Dimensión 1: Utilidad percibida

Este bloque abarca los enunciados 1, 2, 11, 12, 13, 14 y 15 de la encuesta, los cuales se presentan a continuación:

1. Al usar el Chat con Moderador Automático, le permitió interactuar con sus compañeros más eficientemente en comparación como lo hacía utilizando otra herramienta.
2. Considera esta herramienta importante para favorecer el proceso-aprendizaje.
11. Considera que usar el Chat con Moderador Automático mejora el estudio de una asignatura.
12. Encuentra muy útil el Chat con Moderador Automático.
13. Usar el Chat con Moderador Automático le permite aclarar dudas, enriquecer y fortalecer sus conocimientos de la materia.
14. Usar el Chat mejora la calidad del trabajo que realiza en la asignatura.
15. El usar el Chat le permite desarrollar habilidades de Comunicación.

Los ítems antes presentados se resumieron en una nueva variable denominada **Utilidad percibida**, la cual fue encontrada como la mediana entre los enunciados que integran el bloque. Para esta dimensión se obtuvo $\text{Alpha} = 0,8415$

En la Tabla N° 10, se ilustra los cálculos de la matriz de correlaciones teniendo en cuenta la dimensión: Utilidad Percibida.

Tabla N° 10: Matriz de correlaciones teniendo en cuenta la dimensión: Utilidad Percibida.

	1	2	11	12	13	14	15	UTILIDAD PERCIBIDA
1	1,000	0,163	0,569	0,297	0,484	0,244	0,400	0,504
2	0,163	1,000	0,511	0,527	0,282	0,418	0,210	0,578
11	0,569	0,511	1,000	0,694	0,642	0,536	0,295	0,671
12	0,297	0,527	0,694	1,000	0,456	0,603	0,334	0,784
13	0,484	0,282	0,642	0,456	1,000	0,413	0,569	0,620
14	0,244	0,418	0,536	0,603	0,413	1,000	0,559	0,738
15	0,400	0,210	0,295	0,334	0,569	0,559	1,000	0,611
UTILIDAD PERCIBIDA	0,504	0,578	0,671	0,784	0,620	0,738	0,611	1,000

Fuente: Elaboración propia con base a la Encuesta aplicada a estudiantes. San Juan. 2014.

Se presenta a continuación la tabla de frecuencia de la variable resumen: Utilidad percibida.

Tabla N° 11: Distribución de frecuencias de la variable Utilidad percibida.

	Frecuencia	Porcentaje
3,00	1	3,3
4,00	19	63,3
5,00	10	33,3
Total	30	100,0

Fuente: Elaboración propia con base a la Encuesta aplicada a estudiantes. San Juan. 2014.

Se observa que el 96,6% (suma de las 2 últimas filas) de los encuestados alumnos opinan que el Sistema de Moderador Automático de Chat (MAC) mejorará su desempeño en su formación en ambientes virtuales de enseñanza-aprendizaje (lo que corresponde a la puntuación 4 y 5) con la utilidad percibida.

Dimensión 2: Facilidad de uso percibida

Este bloque o dimensión abarca los enunciados 3, 4, 6, 7, 8, 9 y 10 de la encuesta, los cuales se presentan a continuación:

3. Aprender a utilizar la herramienta le fue fácil.
4. Encuentra fácil el uso del Chat con Moderador Automático.
6. El usar el Chat no le representa ningún estrés.
7. Encuentra fácil que el Chat haga lo que Usted quiere que haga.
8. Es fácil para Usted, recordar cómo se realizan las tareas usando el Chat con Moderador Automático.
9. Es fácil obtener la información que Usted quiera del Chat.
10. Considera el Chat con Moderador Automático una herramienta fácil de usar en la interacción con sus compañeros sobre una asignatura.

Estos ítems se resumieron en una nueva variable denominada **Facilidad de uso percibida**, la cual fue encontrada como la mediana entre los enunciados que integran el bloque. Para esta dimensión se obtuvo $\text{Alpha} = 0,8582$. En la Tabla N° 12 se presentan los cálculos de la matriz de correlaciones teniendo en cuenta la dimensión: Facilidad de uso percibida.

Tabla N° 12: Matriz de correlaciones teniendo en cuenta la dimensión: Facilidad de uso percibida.

	3	4	6	7	8	9	10	FACILIDAD DE USO PERCIBIDA
3	1,000	0,474	0,075	0,303	0,328	0,162	0,073	0,558
4	0,474	1,000	0,457	0,364	0,493	0,475	0,432	0,681
6	0,075	0,457	1,000	0,796	0,575	0,556	0,757	0,709
7	0,303	0,364	0,796	1,000	0,574	0,609	0,692	0,728
8	0,328	0,493	0,575	0,574	1,000	0,671	0,476	0,642
9	0,162	0,475	0,556	0,609	0,671	1,000	0,639	0,735
10	0,073	0,432	0,757	0,692	0,476	0,639	1,000	0,739
FACILIDAD DE USO PERCIBIDA	0,558	0,681	0,709	0,728	0,642	0,735	0,739	1,000

Fuente: Elaboración propia con base a la Encuesta aplicada a estudiantes. San Juan. 2014.

Se presenta a continuación la tabla de frecuencias de la variable resumen para la dimensión: Facilidad de uso percibida.

Tabla N° 13: Distribución de frecuencias de la variable Facilidad de uso percibida

	Frecuencia	Porcentaje
2,00	1	3,3
3,00	3	10,0
4,00	14	46,7
5,00	12	40,0
Total	30	100,0

Fuente: Elaboración propia con base a la Encuesta aplicada a estudiantes. San Juan. 2014

Se logra observar que el 86,7% (suma de las 2 últimas filas) de los encuestados alumnos opinan que usando el sistema de Moderador Automático de Chat, realizarán menos esfuerzo para desempeñar sus tareas de formación (lo que corresponde a la puntuación 4 y 5) con la facilidad de uso percibida.

Dimensión 3: Intención de uso

Este bloque abarca los enunciados 5 y 16 de la encuesta, los cuales se presentan a continuación:

5. Utilizaría esta herramienta cada vez que tenga que recibir un curso.
16. Intenta usar el Chat para planear interacciones entre sus compañeros en una materia.

Estos ítems se resumieron en una nueva variable denominada **Intención de uso**, cual fue encontrada como la mediana entre los enunciados que integran el bloque. Para esta dimensión se obtuvo $\text{Alpha} = 0,7220$. En la Tabla N° 14 se ilustra los cálculos de la matriz de correlaciones teniendo en cuenta la dimensión: Intención de uso.

Tabla N° 14: Matriz de correlaciones teniendo en cuenta la dimensión: Intención de uso.

	5	16	INTENCIÓN DE USO
5	1,000	0,567	0,875
16	0,567	1,000	0,895
INTENCIÓN DE USO	0,875	0,895	1,000

Fuente: Elaboración propia con base a la Encuesta aplicada a estudiantes. San Juan. 2014.

Se presenta en la Tabla N° 15, la tabla de frecuencias de la variable resumen para la dimensión: Intención de uso.

Tabla N° 15: Distribución de frecuencias de la Intención de uso.

	Frecuencia	Porcentaje
2,00	1	3,3
3,00	6	20,0
3,50	4	13,3
4,00	9	30,0
4,50	4	13,3
5,00	6	20,0
Total	30	100,0

Fuente: Elaboración propia con base a la Encuesta aplicada a estudiantes. San Juan. 2014.

Se observa que el 63,3% (suma de las 3 últimas filas) de los encuestados alumnos opinan poseen una actitud positiva de uso del sistema de Moderador Automático de Chat en ambientes didáctico-pedagógicos virtuales (lo que corresponde a la puntuación 4 y 5) con la utilidad percibida. Recordar que la dimensión de Intención de Uso, es consecuencia de las características o dimensiones de Utilidad Percibida y Facilidad de uso Percibida, las cuales brindan valores de los cuales se puede concluir que existiría una Intención de Uso del sistema MAC mayor al 70 % de los alumnos encuestados.

En Síntesis:

Tanto en docentes como en alumnos, las variables medidas indican una **alta aceptación del sistema** por parte de los usuarios. Si bien, hay algunas variantes cuando se discrimina la opinión de docentes y alumnos:

- para el caso de los docentes:

- ✓ Es alta la utilidad percibida por parte de los docentes, lo mismo sucede con la facilidad de uso percibida.
- ✓ También es alta la intención de uso, es decir los docentes poseen una actitud positiva de uso con respecto al sistema de Moderador Automático de Chat en ambientes didáctico-pedagógicos virtuales.

- para el caso de los alumnos:

- ✓ Visualizan con mayor frecuencia que los docentes la utilidad percibida.
- ✓ La facilidad de uso percibida por parte de los alumnos es alta.
- ✓ Si bien la intención de uso es superior al 50%, es menor a la detectada entre los docentes.

CAPITULO VI: Conclusiones y Futuros trabajos

Partiendo de la realidad que no existe un moderador automático para salas de chat en ambientes educativos en español, la presente Tesis se planteo como un Objetivo: “Diseñar e implementar, en español, una arquitectura que posibilite moderar automáticamente la interacción en el chat, contextualizando el discurso escrito a una temática especificada.”

Respecto a dicho objetivo, los capítulos precedentes dan cuenta del proceso que se siguió para el diseño e implementación del Sistema Automático de Chat (SAC), que permitió contextualizar discursos en español para ambientes escolares. El sistema creado permite capturar las conversaciones de los participantes, las analiza y las compara con una base de conocimiento cargada previamente por el docente. Mide el grado de desviación de las conversaciones respecto a la base de conocimiento y según el criterio de intervención (definido y cargado por el docente), el SAC interviene de manera automática con mensajes previamente definidos y cargados por el docente.

En el proceso de construcción del SAC surgió la necesidad de vincular el área de la lingüística al área de la informática, a fin de comprender la problemática del lenguaje en general, y del discurso en particular debido a su ambigüedad. Esto significó un esfuerzo adicional, al tener que incursionar en un área desconocida, lo cual requirió lecturas de ese ámbito de conocimiento. Ellas llevaron a autores destacados, tales como T Van Dick, de quien se recuperan sus macro-reglas para el desmontaje del discurso.

Reflexionando sobre los aspectos favorables del diseño y construcción de los algoritmos que permitieron moderar el chat e implementar un sistema Moderador Automático de Chat como tecnología novedosa para la educación, se puede destacar que el SAC:

- Facilita la intervención didáctico-pedagógica del docente en ambientes virtuales, debido a que se "lo libera" del trabajo repetitivo de escuchar e intervenir si los estudiantes se desvían del tema a debatir.
- Libera al docente de tareas repetitivas, permite que el mismo pueda disponer de tiempo para replantear prácticas más creativas e intervenciones didácticas más motivadoras.

- Permite atender grandes cantidades de alumnos de modo personalizado.
- Flexibiliza los horarios de intervención de los alumnos, ayudándolos a ajustar su ritmo de estudio.
- Permite monitorearse la participación del alumno de manera más personal.
- Incorpora los avances del área de inteligencia artificial y los del procesamiento del lenguaje natural a mejorar el proceso de enseñanza-aprendizaje en ambientes virtuales.

Otro aspecto relevante es que el proceso de enseñanza a distancia resulta favorable cuando no es posible contar con la presencia física del docente, en este sentido, la herramienta que se aporta fortalece la educación a distancia. En este sentido el desarrollo de SAC, favorece la inclusión social.

También, pensando desde la perspectiva del destinatario de esta herramienta (los alumnos), la propuesta de uso de chat en el proceso de enseñanza y aprendizaje, resulta beneficioso en tanto el chat es una forma de comunicación muy utilizada por los alumnos fuera del ámbito educativo. Ello favorece la adopción de la propuesta del uso de este medio como práctica de enseñanza.

Cabe destacar que al inicio del trabajo de tesis, uno de los tantos interrogantes era cómo tener en cuenta la forma de escribir de los alumnos, es decir, de qué modo arman los discursos que luego el profesor debe evaluar. Al respecto, es común que los alumnos usen ciertas contracciones, por ejemplo la q por la palabra que, lo cual llevó a decidir que en este primer prototipo se considerarían las palabras bien formadas según la RAE.

Esta situación planteada, originó la necesidad de contar con condiciones de funcionamiento, para asegurar un uso eficiente del sistema.

Entonces, para brindar la solución a la problemática de monitorear las conversaciones de los participantes alumnos en la sala de Chat, se tuvo en cuenta las siguientes condiciones de funcionamiento, es decir, el alcance del actual sistema:

- ✓ La carga del artículo a tratar en la conversación, será por copy-paste.

- ✓ Se espera que los participantes, escriban “palabras bien formadas”, es decir, sin errores de ortografía ni errores semánticos.
- ✓ El diálogo en el Chat debe ser sin “contracciones”, lo que implica que las palabras deberán ser escritas en su totalidad, por ejemplo: “q” debe ser escrita “que”.
- ✓ La sesión de Chat no soporta “emoticones”, es decir, símbolos que expresan emociones. Por ejemplo 😊 “significa que estoy contento ó estoy bien”.
- ✓ El sistema analiza a los sinónimos como palabras distintas.
- ✓ El “regionalismo” o "variantes del idioma" o "localismo", no se tendrá en cuenta por ahora. solo aquellas incluidas en la RAE (Real Academia Española).
- ✓ El lenguaje a considerar será solo el Español.
- ✓ Solo se analizan y se procesan los sustantivos. Este tipo de palabras, de acuerdo a un algoritmo de procesamiento, representan desde un 30% a un 40 % de la cantidad total de palabras de un texto.

Estas condiciones de funcionamiento que restringen el lenguaje usual de los alumnos, deben ser replanteadas a futuro, como ajustes a realizar en el SAC.

El otro Objetivo de la tesis fue: “Estimar el nivel de aceptación del sistema diseñado entre usuarios en el ámbito educativo de la Provincia de San Juan”.

Para alcanzarlo, se apeló a una metodología ampliamente aceptada para validar nuevas tecnologías: el método TAM. Al respecto, se definieron dos ámbitos escolares donde aplicar este método, y por medio de encuestas a docentes y alumnos se logró evaluar el nivel de aceptación del SAC.

Los resultados de las encuestas, tanto en docentes como en alumnos, indican una **alta aceptación del sistema** por parte de los usuarios. Si bien, hay algunas variantes cuando se discrimina la opinión de docentes y alumnos:

- para el caso de los docentes:

- ✓ Es alta la utilidad percibida por parte de los docentes, lo mismo sucede con la facilidad de uso percibida.

✓ También es alta la intención de uso, es decir los docentes poseen una actitud positiva de uso con respecto al sistema de Moderador Automático de Chat en ambientes didáctico-pedagógicos virtuales.

- para el caso de los alumnos:

- ✓ Visualizan con mayor frecuencia que los docentes la utilidad percibida.
- ✓ La facilidad de uso percibida por parte de los alumnos es alta.
- ✓ Si bien la intención de uso es superior al 50%, es menor a la detectada entre los docentes.

Posibles futuros trabajos

El desarrollo alcanzado permite pensar en algunos cambios al sistema para mejorarlo. Entre ellos podemos mencionar:

- Incorporar los algoritmos a una plataforma educativa: si bien la construcción ha sido con un chat propio, lo ideal una vez monitoreado su normal funcionamiento, es incorporarlo a alguna plataforma educativa como por ejemplo moodle o educativa.
- Adicionar el uso de sinónimos: actualmente cuando varias palabras distintas que tienen el mismo significado ingresan al sistema, son tomadas como palabras distintas, a pesar de que hacen referencia a la misma cosa. Una manera de mejorar la eficiencia del sistema es incorporar el manejo de sinónimos, esto es, que si dos más palabras tienen el mismo significado, considerarlas como la misma palabra.
- Reconocimiento de usuarios con un bajo nivel de aporte a la conversación: cuando en la sesión de chat un alumno no presenta un nivel aceptable de participación, el sistema debería contar con alarmas que lo detecte, a fin de enviar mensajes motivándolo a mejorar la cantidad de sus intervenciones. Por ejemplo, haciéndole preguntas puntuales de la temática.
- El sistema podría incorporar un modo automático de corrección de errores de ortografía.
- Incorporar el reconocimiento de abreviaturas: diseñar un algoritmo que reconozca las palabras abreviadas de mayor uso entre los usuarios alumnos, para ser incorporadas como sinónimos de las palabras bien formadas.

- Reconocer regionalismos: diseñar un algoritmo que permita reconocer la ubicación geográfica para recopilar las palabras propias de una región.
- Gestión de emoticones: incorporar al sistema el manejo de emoticones muy utilizados en los chat, que establecen el estado de ánimo de los alumnos. Podría funcionar estos emoticones como motivación en el proceso de enseñanza-aprendizaje.
- Utilizar el sistema como medio de evaluación: dado que se tiene una estructura de conocimiento incorporado por el docente por un lado, y a los discursos de los alumnos por otro lado, el sistema podría cotejarlos y dar una puntuación que funcione como evaluación cuantitativa, en tanto puede medir el grado de conocimiento del alumno de los temas tratados en el documento aportado por el docente. Para alcanzar esto, se debería aplicar los algoritmos del SAC a cada alumno.
- Incorporar las etapas del proceso de T. van Dijk: el sistema creado sólo utiliza las dos primeras macro-reglas (omitir-seleccionar), dejando afuera las macro reglas de generalizar e integrar. Ello permitiría superar que sólo sean evaluados los sustantivos, para considerar a todas las palabras del discurso que construyen los alumnos en las sesiones de chat. Ello requiere la generación de algoritmos que permitan analizar todo tipo de palabras (artículos, verbos, adjetivos) y de esa manera poder lograr un método más eficiente para evaluar las intervenciones.

Para concluir, cabe recordar que la tesis presentada forma parte de un proyecto de investigación más amplio, con lo cual la posibilidad de profundizar los aspectos seleccionados es viable en tanto podrán ser objetivos de futuras investigaciones del equipo del cual es parte el autor de esta tesis.

7. Bibliografía

Adell, J. (1998). Redes y Educación. Universitat Jaume I. Barcelona, España. Disponible en:
<http://edutec.rediris.es/documentos/1998/redes.html>

Cantero, J. y de Arriba, J. (1997) Psicolingüística del discurso. Barcelona: Octaedro.

Casali, A. (2010). Sistema inteligente para la recomendación de objetos de aprendizaje.
LACLO.

Castells, M. (1999) Economía, Sociedad y Cultura. Siglo XXI Editores. Madrid

Cohen. (1995). Text Categorization and Relational Learning. International Conference of
Machine Learning. Chomsky, N. (2005). Problemas actuales en teoría lingüística.
Temas teóricos de gramática generativa. (Octava ed.). Argentina: Siglo XXI.

Cronbach, L. (1951). Coefficient alpha and the internal structure of test. Psychometrika.

Cuevas, G. (1975). Teoría de la información, codificación y lenguajes. Madrid.

Chuttur, M. (2009). Overview of the Technology Acceptance Model: Origins, Developments
and Future Directions. Indiana, USA: Indiana University. Obtenido de
<http://sprouts.aisnet.org/9-37>

Dijk, T.V. (1978) La ciencia del texto: un enfoque interdisciplinario. Barcelona: Paidós.

----- (1980). Estructuras y funciones del discurso. México D.F.: Siglo Veintiuno.

----- (1995). Texto y contexto: semántica y pragmática del discurso. 5a. ed. Madrid:
Cátedra.

----- (2000). El discurso como estructura y proceso. (Elena Marengo, Trad.) Barcelona:
Gedisa.

Deerwester. (1990). Indexing by latent semantic analysis. Journal of the American Society for
Information Science.

- Deitel, H. M., & Deitel, Paul J. (2004). Como Programar en Java. Pearson Educación.
- Diaz Ogás, M. (2014) “Diseño e implementación de un Moderador Automático para una Sala de Chat” Tesis de Licenciatura en Ciencias de la Computación. FCEFYN. UNSJ.
- Fayyad. (1996). From Data Mining to Knowledge Discovery: An Overview. Advances in Knowledge Discovery and Data Mining. AAAI/MIT .
- Frank. (1999). Domain-Specific Keyphrase Extraction. 16th International Joint Conference on Artificial Intelligence.
- Hernández, N. (2001). El Chat como herramienta de comunicación en la educación a distancia: usos y potencialidades para fomentar el aprendizaje cooperativo. Docencia Universitaria, Volumen II. Disponible en: http://saber.ucv.ve/ojs/index.php/rev_docu/article/viewFile/4475/4286
- Fayyad. (1996). From Data Mining to Knowledge Discovery: An Overview. Advances in Knowledge Discovery and Data Mining. AAAI/MIT
- Foucault, M. (2003) La arqueología del saber. 23a. ed. México: Siglo Veintiuno.
- Fox, V. (2005) Análisis documental de contenido: principios y prácticas. Buenos Aires. Alfagrama.
- Furnas (1987). The vocabulary problem in human system communication. Communications of the ACM.
- Hernández, A. y Velilla (2001) “Dimension reduction in nonparametric discriminant analysis” Working Paper 01-39, Statics and Econometric Series 25, Universidad Carlos III de Madrid.
- Joachims (1998) “Text Categorization with support vector machines: Learning with many relevant features”, in Proceedings of 10^o European Conference on Machine Learning, pp. 137-142.

- Johnson, D., Johnson, R.T., & Holubec, E. J.(1999). El aprendizaje cooperativo en el aula. Barcelona: Paidós. España.
- Kabalen, D. y Sánchez, M. (1997) La lectura analítico-crítica: un enfoque cognoscitivo aplicado al análisis de la información. 3a. ed. México D.F.: Trillas.
- Leahey, T. y Harris, R. (1998) Aprendizaje y cognición. 4a. ed. Madrid: Prentice Hall.
- Lewis, & Ringuette. (1994). A comparison of two learning algorithms for text categorization. 3th Symposium on Document Analysis and Information Retrieval.
- Maass, M. (2005) Laboratorio de investigación y desarrollo en comunicación compleja: una propuesta para pensar la complejidad. En Andamios: revista de investigación social. Vol. 1, no. 2, 79-96.
- Martínez Comeche, J. (1996). El mensaje documentario y el documento. En: López Yepes, José, ed. Manual de información y documentación. Madrid: Pirámide. p. 48-62.
- Maniez, J. (1993) Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales. Madrid: Fundación
- Mijáilov, Alexandre I. y Rudzhero S. Guiliarevskii. (1974) Curso introductorio de informática/documentación. Caracas: Fundación Instituto Venezolano de Productividad.
- Moreiro, J. (1998) Introducción al estudio de la información y la documentación. Medellín: Universidad de Antioquia.
- Moulinier. (1996). A Framework for Comparing Text Categorization Approaches. International Conference on Maching Learning.
- Pérez Montoro, M. (2004) Identificación y representación del conocimiento organizacional: la propuesta epistemológica clásica. Disponible en:
<http://www.uoc.edu/in3/dt/20390/index.html>

Pérez, T. Vera, B. & Sánchez, S. (2012). Experiencias de educación mediada por tecnología en la FaCiMed. Estrategias de consolidación de procesos de alfabetización digital. Disponible en: <http://seadiuncoma.files.wordpress.com/2012/06/010-experiencias-de-educacion3b3n-mediada-por-tecnolog3ada-en-la-facimed.pdf>

Perelló, J. (1998) Sistemas de indización aplicados en bibliotecas: clasificaciones, tesauros y encabezamientos de materias. En: Magán Wals, José Antonio, ed. Tratado básico de biblioteconomía. 3a. ed. Madrid. Complutense. p. 200-203.

Peña Vera, T. y Pirela Morillo, J. (2007) "La complejidad del análisis documental" Revista Información, cultura y sociedad, versión On-line ISSN 1851-1740. Disponible en: http://www.scielo.org.ar/scielo.php?pid=s1851-17402007000100004&script=sci_arttext

Pinto Molina, M. y Gálvez, C. (1996) Análisis documental de contenido: procesamiento de la información. Madrid: Síntesis. (Bibliotecología y Documentación).

Pinto Molina, M. (1992) El resumen documental: principios y métodos. Madrid: Pirámide.

RAE (2014) Diccionario Español de la Real Academia Española. Avance de la 23ª edición Disponible en: <http://lema.rae.es/drae/srv/search?id=f3dTLsHfMDXX2T7UDfT7&s=6am>

Ramírez C., Hernández Orallo J., & Ramírez J. (2004). Introducción a la Minería de Datos. (Primera ed.). Pearson.

Renkema, J. (1999). Introducción a los estudios sobre el discurso. Barcelona. Gedisa.

Rodríguez, A. y otros (2011).. Inclusive scenarios to evaluate an open and standards-based framework that supports accessibility and personalisation at higher education. Volume Part IV (UAHCI'11), Constantine Stephanidis (Ed.), Vol. Part IV. Springer-Verlag, Berlin, Heidelberg, 612-621. Disponible en: http://link.springer.com/chapter/10.1007%2F978-3-642-21657-2_66#page-1

Schütze et al. (1995), "A comparison of classifiers and document representations for the routing problem", in Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval., pp. 229-237

Scott, & Matwin. (1999). Feature engineering for text classification. International Conference on Machine Learning.

Solís Hernández, I. (2003) El análisis documental como eslabón para la recuperación de información y los servicios. Disponible en:

<http://www.monografias.com/trabajos14/analisisdocum/analisisdocum.shtml>

Sunkel G., Trucco, D. y Espejo, A. (2014) "La integración de las tecnologías digitales en las escuelas de América Latina y el Caribe. Una mirada multidimensional" CEPAL, Santiago de Chile.

Vickery, B. (1970) Techniques of information retrieval. Londres. Butterworths.

Wagensberg (1989) (en Peña Vera y Pirella Morillo 2007, p.57

Yang, & Chute. (1994). An example-based mapping method for text categorization and retrieval. ACM Transactions on Information Systems. .

Yang. (1994). Expert Network: Effective and efficient learning from human decisions in text categorization and retrieval. 17th ACM SIGIR Conference.

ANEXOS

Anexo 1: Instrumento de evaluación para Docentes**Cuestionario para Docentes****Género**

Femenino Masculino

Rango de edad

Menos de 20 20-30 31-40 41-50 Más de 50

¿En qué nivel trabaja como docente?

- Educación Primaria
 Educación Secundaria
 Educación Superior No Universitaria
 Educación Universitaria

Otro:

Antigüedad en la Docencia:

0 a 5 años 5 a 10 años 10 a 15 años Más de 15 años

Experiencia Profesional y en Docencia:

Usted es profesional en el área de: _____

Ha sido docente en cursos presenciales en los últimos 12 meses? Si ___ No ___

¿Posee alguna titulación en docencia? Si ___ No ___

En caso afirmativo, cuéntenos cual: _____

Trabaja con plataformas virtuales educativas:

si no

Si la respuesta anterior fué afirmativa, que plataforma virtual educativa utiliza:

Moodle E-ducativa Otra:

¿En qué asignatura/s trabaja y en que modalidad? Por favor especifique:

Asignatura	Modalidad			Institución
	Presencial	Virtual	Semi-Presencial	

¿Durante cuánto tiempo ha sido docente?

Modalidad	1-2 Años	2-5 Años	5-10 Años	>10 Años
Presencial				
Virtual				
Semi-Presencial				

¿Cuántos cursos (sistema de educación formal o no formal) diferentes ha desarrollado :

Modalidad	1-5 Cursos	5-10 Cursos	10-20 Cursos	>20 Cursos
Presencial				
Virtual				
Semi-Presencial				

¿Coloca sus materiales de enseñanza en internet? Si___ No___

En caso afirmativo, usa alguna de las siguientes plataformas:

Moodle__ dotLRN__ WebCT/Blackboard__ Atutor__
 Educativa____Otras_____

¿Cuáles son las principales ventajas o inconvenientes que usted encuentra en el uso de plataformas de educación virtuales como las antes mencionadas?

Ventajas	Inconvenientes

¿Qué tipo de equipo posee? Por favor, indique:

- Netbook del Programa Conectar Igualdad.
- Tengo otra computadora PC (o de escritorio).
- La netbook de Conectar y otra computadora.
- No tengo computadora.

Otro:

¿Cómo se conecta a internet?

- Tengo internet en casa.
 - Voy a un cyber.
 - Me conecto en el lugar de trabajo.
- Otro:

Uso de Internet y TIC en general. Por favor indique:

¿Qué nivel de experiencia tiene en el uso de computadoras?

Iniciación__ Intermedio__ Experto__

¿Qué nivel de experiencia tiene en el uso de internet?

Iniciación__ Intermedio__ Experto__

¿Cuánto tiempo lleva usando internet?

Menos de 1 año__ 1-2 Años__ Más de 3 años__

¿Cuántas horas a la semana navega en internet?

No todas las semanas__ Menos de 5 horas/sem __ 6-15 horas/sem __ Mas de 15 horas/sem __

¿Usa el chat?

Si__ No__

En caso afirmativo, cuánto tiempo lo usa:

No todas las semanas__ Menos de 5 horas/sem __ 6-15 horas/sem __ Mas de 15 horas/sem __

¿Usa el chat como herramienta de apoyo al proceso educativo?

Si__ No__

En caso afirmativo, cuánto tiempo lo usa:

No todas las semanas__ Menos de 5 horas/sem __ 6-15 horas/sem __ Mas de 15 horas/sem __

¿Qué herramientas de Internet usa de forma regular?, marque solo tendencia de uso principal?

Herramienta	Para aprender	Uso personal	Para enseñar	Otros usos
Correo Electrónico				
Foros				
Mensajería Instantánea - Chat (msn, skype)				
Wikis				
Blogs				
Redes Sociales				
Redes Profesionales				
Otros				

¿Qué otras herramientas usa?

Herramienta	Para aprender	Uso personal	Para enseñar	Otros usos
Procesador de textos				
Hoja de calculo				
Editores de Video/fotos				
Otros				

Contestar los siguientes enunciados en base a la herramienta de Moderador Automático de Chat:

Pregunta	Muy de acuerdo	De acuerdo	Indeciso	Poco de acuerdo	Desa cuerdo
1. Al usar el Moderador Automático de Chat, le permitió monitorear el debate de los alumnos más eficientemente en comparación como lo hacía utilizando otra herramienta.					
2. Considera esta herramienta importante para favorecer el proceso-aprendizaje en sus alumnos.					
3. Aprender a utilizar la herramienta le fue fácil.					
4. Encuentra fácil el uso del Moderador Automático de Chat.					
5. Utilizaría esta herramienta cada vez que tenga que ofrecer un curso en una plataforma virtual.					
6. Utilizaría esta herramienta cada vez que tenga que gestionar el dictado de curso en una plataforma.					
7. Incluiría esta herramienta cada vez que tenga que gestionar el dictado de curso en una plataforma					
8. El usar el chat no le representa ningún estrés.					
9. Encuentra fácil que el haga lo que Usted quiere que haga.					
10. Es fácil para Usted, recordar cómo se realizan las tareas usando el Moderador Automático de Chat.					
11. Es fácil y entendible percibir las interacciones de los alumnos usando el Moderador Automático de Chat.					
12. Es fácil obtener la información que Usted quiera del chat.					
13. Considera el Moderador Automático de Chat una herramienta fácil de usar en la interacción de los alumnos.					
14. Considera que usar el Moderador Automático de Chat mejora el estudio de su asignatura.					

15. Encuentra muy útil el Moderador Automático de Chat.					
16. Usar el Moderador Automático de Chat, le facilita la tarea de monitorear las interacciones de los alumnos.					
17. Usar el Moderador Automático de Chat le permite aclarar dudas, enriquecer y fortalecer los conocimientos en sus alumnos.					
el chat en el aula virtual mejora la calidad del trabajo que realiza en su asignatura.					
19. El usar el chat le permite desarrollar habilidades de Comunicación y/o manejo de conflictos.					
20. Intenta usar el chat para planear interacciones entre sus alumnos.					

Si Usted tiene alguna sugerencia sobre el Sistema Moderador Automático de Chat la puede expresar a continuación:

Muchas gracias por su colaboración!

Anexo 2: Instrumento de evaluación para Estudiantes:

Cuestionario para Alumnos

Género

Femenino Masculino

Rango de edad

Menos de 20 20-30 31-40 41-50 Más de 50

¿En qué nivel estudia como alumno?

Educación Superior No Universitaria

Educación Universitaria

Otro:

Estudia con plataformas virtuales educativas:

si no

Si la respuesta anterior fue afirmativa, que plataforma virtual educativa utiliza:

moodle

E-ducativa

Otra:

¿En qué asignatura/s estudia y en que modalidad? Por favor especifique:

Asignatura	Modalidad			Institución
	Presencial	Virtual	Semi-Presencial	

Si he recibido cursos ¿Cuántos cursos (sistema de educación formal o noformal) diferentes ha recibido:

Modalidad	1-5 Cursos	5-10 Cursos	10-20 Cursos	>20 Cursos
Presencial				
Virtual				
Semi-Presencial				

¿Cuáles son las principales ventajas o inconvenientes que encuentra en el uso de plataformas de educación virtuales?

Ventajas	Inconvenientes

¿Qué tipo de equipo posee? Por favor, indique:

- Tengo la netbook del Programa Conectar Igualdad.
- Tengo otra computadora PC (o de escritorio).
- Tengo la netbook de Conectar y también otra computadora.
- No tengo computadora.

Otro:

¿Cómo se conecta a internet?

- Tengo internet en casa.
- Voy a un cyber.
- Me conecto en el lugar de estudio o trabajo.

Otro:

Uso de Internet y TIC en general. Por favor indique:

¿Qué nivel de experiencia tiene en el uso de computadoras?

Iniciación__ Intermedio__ Experto__

¿Qué nivel de experiencia tiene en el uso de internet?

Iniciación__ Intermedio__ Experto__

¿Cuánto tiempo lleva usando internet?

Menos de 1 año__ 1-2 Años__ Más de 3 años__

¿Cuántas horas a la semana navega en internet?

No todas las semanas__ Menos de 5 horas/sem __ 6-15 horas/sem __ Mas de 15 horas/sem __

¿Usa el chat?

Si__ No__ En caso afirmativo, cuánto tiempo lo usa:

No todas las semanas__ Menos de 5 horas/sem __ 6-15 horas/sem __ Mas de 15 horas/sem __

¿Usa el chat como herramienta de apoyo al proceso educativo?

Si__ No__

En caso afirmativo, cuánto tiempo lo usa:

No todas las semanas__ Menos de 5 horas/sem __ 6-15 horas/sem __ Mas de 15 horas/sem __

¿Qué herramientas de Internet usa de forma regular?, marque solo tendencia de uso principal

Herramienta	Para aprender	Uso personal	Otros usos
Correo Electrónico			
Foros			
Mensajería Instantánea - Chat (msn, skype)			
Wikis			
Blogs			
Redes Sociales			
Redes Profesionales			
Otras			

¿Qué otras herramientas usa?

Herramienta	Para aprender	Uso personal	Otros usos
Procesador de textos			
Hoja de calculo			
Editores de video/fotos			
Otras			

Contestar los siguientes enunciados en base a la herramienta de Moderador Automático de Chat usado:

Pregunta	Muy de acuerdo	De acuerdo	Indeciso	Poco de acuerdo	Desacuerdo
1. Al usar el Moderador Automático de Chat, le permitió interactuar con sus compañeros más eficientemente en comparación como lo hacía utilizando otra herramienta.					
2. Considera esta herramienta importante para favorecer el proceso-aprendizaje.					
3. Aprender a utilizar la herramienta le fue fácil.					
4. Encuentra fácil el uso del Moderador Automático de Chat.					
5. Utilizaría esta herramienta cada vez que tenga que recibir un curso.					
6. El usar el chat no le representa ningún estrés.					
7. Encuentra fácil que el chat haga lo que Usted quiere que haga.					
8. Es fácil para Usted, recordar cómo se realizan las tareas usando el Moderador Automático de Chat.					
9. Es fácil obtener la información que Usted quiera del chat.					
10. Considera el Moderador Automático de Chat una herramienta fácil de usar en la interacción con sus compañeros sobre una asignatura.					

11. Considera que usar el Moderador Automático de Chat mejora el estudio de una asignatura.					
12. Encuentra muy útil el Moderador Automático de Chat.					
13. Usar el Moderador Automático de Chat le permite aclarar dudas, enriquecer y fortalecer sus conocimientos de la materia.					
14. Usar el chat mejora la calidad del trabajo que realiza en la asignatura.					
15. El usar el chat le permite desarrollar habilidades de Comunicación.					
16. Intenta usar el chat para planear interacciones entre sus compañeros en una materia.					

Si Usted tiene alguna sugerencia sobre el Sistema de Moderador Automático de Chat la puede expresar a continuación:

Muchas gracias por su colaboración!