

Terapia de exposición con realidad virtual evaluada mediante minería de textos

Exposure therapy with virtual reality evaluated by text mining

Rosa Flores Morales¹

RESUMEN

El artículo presenta un estudio acerca de la aplicación de Minería de Textos en la evaluación de la psicoterapia. Nuestros voluntarios fueron sujetos diagnosticados con fobia social y tratados con terapia de exposición en Realidad Virtual. El cálculo de la polaridad emocional, desde el enfoque del análisis sintáctico, fue aplicado al material narrativo producido por los propios voluntarios. Para su valoración cuantitativa se utilizó el lexicon de sentimientos LIWC (*Linguistic Inquiry and Word Count*). Los resultados obtenidos muestran una relación proporcional entre el nivel emocional global de los textos y el avance de la práctica psicológica, lo que sugiere su consideración como evidencia para medir la efectividad de la terapia.

PALABRAS CLAVE

Evaluación de psicoterapia, Fobia social, Minería de textos, Polaridad sentimental, Realidad virtual.

ABSTRACT

This article presents a study about the application of text mining in psychotherapeutic evaluations. Our volunteers are subjects diagnosed with social phobia and treated with virtual reality exposure therapy. Evaluación de psicoterapia Sentiment analysis, through syntactic analysis, was applied over narrative material the volunteers produced themselves; for the quantitative assessment, we used the sentiment lexicon LIWC (*Linguistics Inquiry and Word Count*). The results show a proportional relation between the emotional level of the narrative material and the advancement of therapy, which suggests the consideration of text mining as an effectiveness indicator for psychotherapies.

1 Rosa Flores Morales es licenciada en Informática por la Universidad Mayor de San Andrés (La Paz-Bolivia), con maestría en Ciencias de la Computación, Diplomada en Proyectos de Investigación (I+D+i). Docente de Postgrado y docente emérito de pregrado, investigadora titular del Instituto de Investigaciones en Informática. UMSA. Líneas de trabajo: Simulación de Sistemas, Inteligencia Artificial, Innovación y tecnología. *Correo electrónico.: rflores@umsa.bo y rfloresmorales@gmail.com*

KEYWORDS

Psychotherapy evaluation, Sentiment analysis, Social phobia, Text mining, Virtual reality therapy

RESUMO

O artigo apresenta um estudo sobre a aplicação do mineração de texto na avaliação da psicoterapia. Nossos voluntários foram pacientes com diagnóstico de fobia social e tratada com terapia de exposição em Realidade Virtual. O cálculo da polaridade emocional, a partir da abordagem de análise, foi aplicada no material narrativa produzida pelos próprios voluntários. Foi utilizada para a avaliação quantitativa o léxico de sentimentos LIWC (Linguistic Inquiry e Word Count). Os resultados mostram uma relação proporcional entre o nível emocional general de textos e o avanço da prática psicológica, sugerindo-se sua consideração como prova para medir a eficácia da terapia.

PALAVRAS-CHAVE

Avaliação da psicoterapia, Fobia social, Mineração de texto, Polaridade sentimental, Realidade virtual

1. Introducción

Una fobia está presente cuando una persona siente un miedo exagerado irracional ante una situación o estímulo. La terapia de exposición en vivo frente al objeto temido, permite reducir de forma efectiva este temor (Labrador & Ballesteros, 2011), sin embargo, en muchos casos, el objeto fóbico impide su exposición ya sea por problemas de acceso o de costo. Esta dificultad en casos de fobia social (FS) es más compleja: exponer al paciente en un escenario que genera ansiedad implica la presencia de personas, como organizar una entrevista de trabajo, disponer de una sala de reuniones con un número de personas dispuestas a intercambiar opiniones, etc. Ante estas dificultades surge como alternativa la Realidad Virtual (RV), tecnología que “provee entornos sintéticos, tridimensionales, multisensoriales, interactivos e inmersivos” (Stuart, 2001) (citado en (Luque, 2009), p. 3), que permiten ver, oír y sentir como se lo haría en un ambiente real.

El uso de la RV tuvo sus orígenes en los años 90, cuando algunas compañías aéreas y programas aeroespaciales utilizaron esta tecnología en el entrenamiento de los astronautas. En la actualidad, su incorporación en la terapia psicológica ha tenido mucho éxito, tanto en los procesos experimentales, como en los clínicos (Bornas, Rodrigo, Barceló, & Toledo, 2002). El año 2004 se reporta un caso de estudio en el tratamiento de la acrofobia con RV, donde los investigadores apoyan la eficacia del uso de esta tecnología en tratamientos por exposición a personas que sufren este trastorno (Choi, Jang, Ku, Shin, & Kim, 2004). Vincelli et al, el año 2003 publican otro artículo de tratamiento con RV en casos de pánico con resultados favorables (Vincelli, Anolli, Bouchard, Wiederhold, Zurloni, & Riva, 2004). Los trabajos de Botella et al (Botella, García-Palacios, Baños, & Quero, 2016) muestran las experiencias de la RV y la Realidad Aumentada en el tratamiento psicológico.

El tratamiento psicoterapéutico en los casos de FS, hoy en día, ha incorpora-

do la RV con resultados satisfactorios. Al respecto, existen muchos reportes de investigación que evidencian la efectividad de esta tecnología en la terapia cognitiva conductual. El grupo del Dr. Slater (Slater, Pertaud y Steed, 1999; Pertaud, Slater y Baker, 2002) ha trabajado en el diseño y validación de software de RV para el tratamiento del miedo a hablar en público. Sus trabajos demostraron que los escenarios virtuales, en los que el sujeto se enfrentaba a la situación de hablar ante audiencias virtuales, producían una respuesta de ansiedad similar a la que produciría una audiencia real. Citado por García et al (García, Rosa-Alcazar, & Olivares, 2011).

La incorporación de la RV en la psicología clínica no implica la propuesta de una nueva terapia, sino de añadir tecnología en la terapia cognitiva tradicional (Luque, 2009). Behar (Behar, 1993) afirma al respecto que, la aplicación de las nuevas tecnologías se apoyan en el recurso de la terapia cognitiva tradicional y que no se trata de una nueva psicoterapia, sino de modificaciones en el dispositivo terapéutico tradicional y en su implementación. Sin embargo, desde la perspectiva de la Práctica Psicológica Basada en la Evidencia (PPBE) esta innovación tecnológica requiere una evaluación sobre su efectividad.

2. Psicología basada en evidencias

La PPBE postula la necesidad de contar con elementos de medición de tal manera que la eficacia se remita a la evaluación de la evidencia que pretende establecer relaciones causales entre tratamientos y trastornos (Rivera & Bernal, 2008). En otras palabras, surge la pregunta de si un determinado tratamiento cuenta con apoyo empírico, si hay evidencia de que dicha intervención funciona.

La necesidad de la práctica clínica basada en la evidencia se remonta a los años 30 (Rivera & Bernal, 2008), cuando Otto Fenichel publica los resultados de sus estudios

realizados en el Instituto Psicoanalítico de Berlín, evaluando los expedientes y ubicándolos bajo las categorías de: “curado, bastante mejorado, mejorado, no curado”. Esta clasificación nos habla de la efectividad en la práctica clínica y de algún proceso de medición empleado. Según Rivera y Bernal (Rivera & Bernal, 2008), mediante la medición se puede obtener información concreta de cómo llegó la persona al inicio del proceso terapéutico y cómo se encuentra al finalizarlo. La medición permite advertir ciertas características de la persona, síntomas apremiantes y patrones de conducta o de pensamientos que pueden afectar el proceso terapéutico o que deben ser atendidos para asegurar la efectividad del mismo. A su vez, la medición durante el proceso terapéutico permite obtener información valiosa para identificar si la persona está respondiendo a la terapia de acuerdo a lo esperado o si se debe realizar algunas modificaciones para observar un mejor efecto.

Los autores Rivera y Bernal (Rivera & Bernal, 2008) señalan que un asunto importante dentro de la medición del cambio terapéutico es la utilización de los instrumentos apropiados que permitan obtener información más específica y cuantificable a fin de medir los avances y retrocesos de las personas durante la terapia.

En la práctica psicológica, según Behar (Behar, 1993) las entrevistas clínicas, de selección y la misma psicoterapia se fundamentan en el intercambio verbal entre el terapeuta y el paciente, “donde la palabra es soberana”. Este material verbal resultante, muy pocas veces es objeto de análisis sistemático y riguroso. Lo que sucede en general, es que el psicólogo deduce conclusiones, interpretando desde su subjetividad el sentido del material verbal recogido. La autora no resta importancia a esta línea de trabajo pero enfatiza la necesidad de utilizar todo tipo de material verbal para extraer información objetiva me-

dian­te el análisis del texto, como el análisis de contenido o la semántica analítica. Considerando esta necesidad de medir el progreso de los pacientes en la terapia, con el objeto de evaluar lo que sucede durante el proceso apoyado con escenarios de RV, algunas investigaciones estudiaron las reacciones psicofisiológicas, en un intento por lograr medidas cuantitativas y de mayor objetividad. Estas medidas consideraron los signos vitales como el ritmo cardíaco, la presión arterial, la frecuencia fundamental de la voz y otros. En un caso concreto se estudió la acrofobia, tratada con terapia de exposición con escenarios de RV (Plaza, 2008), este trabajo reporta que “un 47% de las personas que habían manifestado tener miedo a las alturas informaron sentirse un poco afectadas por el ambiente virtual, mientras que un 5% de estos individuos se sintieron muy afectados”. Al considerar el ritmo cardíaco como un indicador, el autor encuentra que, “para individuos con miedo a las alturas, la frecuencia cardíaca presenta un valor similar en reposo y en el ambiente virtual (reposo: 83.3 bpm o pulsaciones por minuto; ambiente virtual: 84.4 bpm, sig: 0.754)”. Al analizar la temperatura corporal en sujetos que padecen de miedo a las alturas, encuentra que la diferencia no es significativa al exponerlos en ambientes reales y virtuales. El trabajo llega a la conclusión final que algunos parámetros, como la temperatura, la variabilidad cardíaca y la frecuencia fundamental de la voz, permitieron distinguir el estado de reposo y ambiente virtual. Solamente el parámetro de temperatura y frecuencia fundamental de la voz son útiles para evaluar el estado de los sujetos en el ambiente virtual de alturas.

3. Minería de textos

La Minería Textual (MT) es una disciplina relativamente joven, en la que se integran técnicas desarrolladas desde la lingüística computacional, la lexicometría, la minería

de datos, el procesamiento del lenguaje natural y la psicometría. Su objetivo es el de conseguir información a partir de patrones no triviales y tendencias que pueden obtenerse al analizar textos no estructurados de gran tamaño (Eíto & Senso, 2004). Tiene su origen en la aplicación de técnicas estadísticas del análisis de textos de discurso, propuesta por el profesor Benzécri, fundador de la escuela francesa de análisis estadístico de datos (1973).

La lingüística computacional es una disciplina nueva, cuyo objetivo consiste en reconocer, comprender el lenguaje natural humano mediante algoritmos computacionales (Castillo, Cárdenas, Curti, & Casco, 2015).

La minería de textos se entenderá como el proceso que permite la organización, clasificación, extracción de información y evaluación con la mínima intervención humana en la detección automática de temas predominantes en un conjunto de documentos y búsqueda de textos relevantes mediante consultas de grandes prestaciones (Césari, 2007).

4. Análisis de sentimientos

El análisis de sentimientos, también conocido como minería de opinión (MO), es una técnica de la MT que permite identificar y extraer información subjetiva del texto escrito. En términos generales, el análisis de sentimientos intenta determinar la actitud de un interlocutor o un escritor con respecto a algún tema o la polaridad contextual general de un documento. La actitud puede ser su juicio o evaluación, estado afectivo (o sea, el estado emocional del autor al momento de escribir), o la intención comunicativa emocional (o sea, el efecto emocional que el autor intenta causar en el lector) (Perez & Araujo).

El análisis de sentimientos surgió con la proliferación de críticas, calificaciones, re-

comendaciones y otras formas de expresión manifestadas en las redes sociales. La opinión en línea se ha convertido en objeto de estudio para evaluar nichos de mercado, indagar sobre la valoración de clientes respecto a un servicio o producto. Esta tarea implica un volumen de datos que sería imposible su tratamiento de forma manual.

Un método para determinar el sentimiento en un texto es el uso de un sistema de escalado donde las palabras son asociadas con un sentimiento: negativo, neutral o positivo; cuando una pieza de texto no estructurado es analizada usando procesamiento de lenguaje natural, se identifica el concepto principal, los subsecuentes conceptos son analizados para un entendimiento de estas palabras y de cómo se relacionan con el concepto substancial. A cada concepto se le otorga una puntuación basada en la forma en que las palabras se asocian.

El proceso de establecer la polaridad sentimental de lo escrito puede realizarse, en general, de dos formas. La primera utiliza las técnicas del aprendizaje automático de la Inteligencia Artificial, en las cuales, a partir de un procesamiento de entrenamiento se logra obtener un clasificador automático. La segunda consiste en analizar la estructura sintáctica de una oración mediante grafos y el uso de diccionarios (lexicón sentimental), donde cada palabra tiene asociado un valor numérico de polaridad sentimental.

El enfoque basado en el aprendizaje automatizado consiste en entrenar un clasificador usando un algoritmo de aprendizaje supervisado a partir de una colección de textos anotados. Cada texto habitualmente se representa con un vector de palabras, n-gramas, en combinación con otro tipo de características semánticas que intentan modelar la estructura sintáctica de las frases. Los sistemas utilizan diversas

técnicas, aunque las más populares son los clasificadores basados en SVM (*Support Vector Machines*), *NaiveBayes* y KNN (*K-Nearest Neighbor*). En las investigaciones más recientes se han empezado a utilizar otras técnicas más avanzadas, como LSA (*Latent Semantic Analysis*) e incluso *Deep Learning* (Perez & Araujo).

Los enfoques semánticos se caracterizan por el uso de diccionarios de términos (lexicón) elaborados con este propósito. Los sistemas automáticos basados en computadoras, preprocesan el texto y lo dividen en palabras, con la apropiada eliminación de las palabras de parada y una normalización lingüística por tematización, y luego comprueban la aparición de los términos del lexicón para asignar el valor de polaridad del texto mediante la suma de valores. Generalmente los sistemas incluyen además, un tratamiento más o menos avanzado de términos modificadores (como muy, poco, demasiado), que aumentan o reducen la polaridad del o los términos a los que acompañan, y términos inversores o negadores (como no, tampoco), que invierten la polaridad de los términos a los que afectan (Vilares, Alonso, & Gómez-Rodríguez, 2013).

Este preprocesamiento, aclara Abregú (Abregú, 2010) es ineludible “porque no se puede realizar un tratamiento lexicométrico sobre formas brutas. Hay que desambigüizar, es decir, eliminar los homógrafos y si es posible, sustituirlos por un sinónimo heterógrafo. Otro tratamiento necesario es la *lematización*, que consiste en producir las formas standard, normalizar las palabras, desnudándolas de sus especificidades de flexión verbal, declinación, desinencias, género, número, etc. En español, para aislar la raíz o lexema de un verbo, se lo lleva al infinitivo, y se extrae la desinencia verbal: *ar*, *er* o *ir*. Los sustantivos y adjetivos deben ser puestos en singular y sin la terminación de género. Además, es conveniente separar otros sufijos y prefijos

jos. Los lemas no siempre coinciden con los lexemas o raíces. No se trata de hacer una descomposición etimológica sino de encontrar, bajo un lema, todas las palabras que lo contienen” (p.370).

LIWC es un programa (software) con un enfoque semántico, desarrollado para analizar las implicaciones psicológicas del uso de las palabras, fue creado por J. Pennebaker, R. Booth y M. Francis. Ha sido utilizado con diferentes propósitos pero ha arrojado buenos resultados en el área de las ciencias sociales y en psicología.

El programa, que viene acompañado con un diccionario de palabras que utiliza en su funcionamiento, ofrece una herramienta eficaz para estudiar los componentes emocionales, cognitivos y estructurales contenidos en un texto. En su construcción participaron psicólogos, sociólogos y lingüistas. Está compuesto por un diccionario en español de 7.515 palabras. Cada palabra puede ser clasificada en 80 categorías; las categorías se clasifican en cuatro dimensiones: procesos lingüísticos estándar, procesos psicológicos, relatividad, y asuntos personales (tabla1).

Tabla 1.
Ejemplo de categorías de LIWC

Dimensión	Abreviatura	Ejemplos
II. PROCESOS PSICOLÓGICOS		
Procesos afectivos o emocionales	Afectivo	feliz, feo, amargo
Emociones Positivas	Emopos	feliz, bonito, bueno
Sentimientos Positivos	Sentpos	feliz, felicidad, amor
Optimismo y Energía	Optime	certeza, orgullo, ganar
Emociones Negativas	Emoneg	odio, enemigo, sin sentido
Ansiedad o miedo	Ansiedad	nervioso, miedo, tenso
Enojo	Enojo	odio, matar, enojo
Tristeza o depresión	Tristeza	llorar, tristeza
Procesos cognitivos	MecCog	causa, saber, debería
Causa	Causa	porque, efecto, por lo tanto
Insight	Insight	Pensar, saber, considerer
Discrepancia	Discrep	debería, podría
Inhibición	Inhib	bloquear, prohibir
Tentativo	Tentat	Talvez, a lo mejor, adivina
Certeza	Certeza	siempre, nunca
Procesos Sensoriales y Perceptuales	Sentidos	ver, tocar, escuchar
Ver	Ver	vista, ver, observer
Escuchar	Oír	escuchó, escucha, sonido
Sentir	Sentir	tocar, sostener, sentir

Procesos Sociales	Social	hablar, nosotros, amigos
Comunicación	Comu	hablar, compartir, conversar
Referencia a otras personas	Refotro	primera persona plural, segunda y tercera
Amigos	Amigos	colega, amigo
Familia	Familia	mamá, hermano, primo
Humanos	Humanos	niño, mujer, grupo
III. RELATIVIDAD		
Tiempo	Tiempo	hora, día, noche
Verbo de tiempo pasado	Pasado	caminé, hice, tuve
Verbo de tiempo presente	Presente	camino, hago, tengo
Verbo de tiempo futuro	Futuro	caminaré, haré, voy a
Espacio	Espacio	alrededor, encima, arriba
Arriba	Arriba	arriba, encima, sobre
Abajo	Abajo	abajo, debajo
Inclusivo	Incl	con, y, incluyendo
Exclusivo	Excl	pero, excepto, sin
Movimiento	Moción	mover, caminar, ir
IV. ASUNTOS PERSONALES		
Ocupación	Ocupa	trabajar, clase, jefe
Escuela	Escuela	clase, estudiante, colegio
Trabajo	Trabajo	empleo, jefe, carrera
Logro	Logro	tratar, objetivo, ganar

El diccionario incluye la raíz de las palabras para facilitar el proceso de lematización, por ejemplo: la raíz “hamb*” permite que cualquier palabra indicativa que coincida con las primeras 5 letras sea contada como una palabra de comer (esto incluiría hambre, hambriento, hambrienta), el asterisco denota la aceptación de todas las letras siguientes. Las palabras pueden pertenecer a una o varias categorías, por

ejemplo, la palabra ‘lloré’ es parte de cuatro categorías de palabras: tristeza, emoción negativa, afecto total y un verbo en tiempo pasado (LIWC, 2011). Todas las palabras que expresan la ira, por definición, serán clasificadas como emoción negativa y, en general, como palabras emocionales. Las emociones asociadas a cada palabra se asignan con valores cuantitativos (tabla2).

Tabla 2.
Sentimientos y valores
cuantitativos asociados

Valor numérico	Sentimiento asociado
-2	Muy negativo
-1	Negativo
0	Neutro
1	Positivo
2	Muy positivo

Existen muchos otros diccionarios de sentimientos, algunos construidos para una temática específica, como por ejemplo, el cine, hotelería, autos, etc. El objetivo de estos diccionarios es facilitar el análisis de opinión en las redes sociales o páginas web donde los internautas se expresan mediante comentarios. Otros diccionarios con clasificaciones complejas como: débil, fuerte, activa, pasiva, placer, dolor, virtud, vicio, etc. están disponibles sólo en inglés. Es el caso de *"The General Inquirer"*, *"MPQA Subjectivity Cues Lexicon"*, que es otro lexicón que, además de clasificar las palabras en positivas o negativas, especifica la intensidad de cada una de ellas (fuerte, débil). *"Senti Word Net"*, *"Bing Liu Opinion Lexicon"* son compilaciones de palabras con carga de valor sentimental igual que las anteriores.

La precisión de un sistema de análisis de sentimientos es corroborada por el juicio humano. Sin embargo, de acuerdo a las investigaciones, "los evaluadores humanos normalmente están de acuerdo en un 79% de las veces con la asignación automática. La evaluación de los programas computacionales llegan a un nivel de aciertos de un 70%" (Perez & Araujo). "Si un programa estuviera "bien" el 100% de las veces, los humanos aún estarían en desacuerdo con él en alrededor de un 20% de las veces, puesto que, de todos modos, están en desacuerdo con cualquier respuesta" (Wikipedia, 2015).

5. Diarios personales

En relación con la escritura impregnada en los documentos personales como cartas, diarios, mensajes, entre otros, la valoración epistémica del lenguaje y la perspectiva de Freud, permiten acceder al inconsciente y convertir en consciente lo inconsciente, por lo que según Santander (Santander, 2011), se justifica considerar el lenguaje como material de análisis. Este autor advierte que en el "análisis del discurso, el lenguaje no es transparente, los signos no son inocentes, que la connotación va con la denotación; que el lenguaje muestra; pero también distorsiona y oculta, que a veces lo expresado refleja directamente lo pensado y a veces sólo es un indicio ligero, sutil, cínico"(p. 208). Sin embargo, los diarios personales se consideran como objetos de estudio desde el principio de la espontaneidad, la asociación libre de ideas y la escritura para uno mismo, características que elevan el valor de significancia de su contenido.

El psicólogo estadounidense James Pennebaker, profesor de la Universidad de Texas, lleva más de 30 años investigando los efectos terapéuticos de la escritura. Este profesor considera que los textos escritos de una persona pueden ayudar a descubrir emociones y pautas de pensamiento más inconscientes.

La forma en que las personas hablamos y escribimos nos permite conocer nuestros mundos emocionales y cognitivos. Las investigaciones a este respecto sugieren que la salud física y mental de las personas puede ser predicha por las palabras que usan. El análisis de textos basados en estos estudios, indican que aquellos individuos que más se benefician al escribir, tienden a usar relativamente altos niveles de palabras emocionales positivas, un número moderado de palabras emocionales negativas y un número creciente de palabras cognitivas (e. g., Pennebaker & Fran-

cis, 1996; Pennebaker, Mayne, & Francis, 1997) (citado por (Meaning Cloud, 2015).

II. Objetivo

Behar(Behar, 1993)considera que la lexicometría cronológica que analiza los datos de un mismo emisor recogidos a través del tiempo se viene utilizando en el análisis de textos políticos, podría ser empleada en el estudio de procesos psicológicos, siempre y cuando se dispusiera de datos textuales de los sujetos implicados, recogidos a lo largo del tiempo de manera sistemática. Para este fin, podrían servir diarios personales, entrevistas grabadas, etc.

Siguiendo esta línea de trabajo, en la investigación se propone evaluar con técnicas de Minería de Textos la eficacia de la terapia de exposición con Realidad Virtual en casos de FS.

- Establecer los términos más frecuentes en los textos contenidos en los diarios digitales de pacientes tratados con la terapia de exposición en ambientes de RV.
- Evaluar con la técnica de polaridad sentimental los diarios digitales de los pacientes.
- Analizar la polaridad sentimental global de los diarios digitales desde el inicio, durante y al final de la terapia.

III. Metodología

La MT ha desarrollado varias técnicas de análisis de datos. En la presente investigación se utilizaron técnicas lexicométricas y el enfoque de la MO.

La MO fue procesada siguiendo las etapas de la figura 1, desde el análisis de la estructura sintáctica de las oraciones y el uso de diccionarios del LIWC.

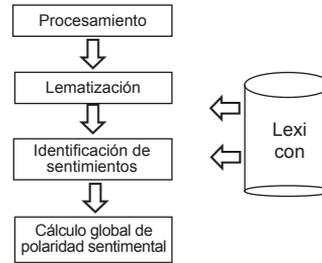


Figura 1: Etapas del análisis de Minería de Opinión

La primera fase consiste en determinar las palabras más frecuentes a fin de depurar el texto eliminando palabras y signos sin sentido. La eliminación no sólo se basó en la frecuencia y la longitud, sino en la importancia de la palabra evaluada en el texto, por ejemplo las palabras ‘yo’, ‘no’ a pesar de su frecuencia no se eliminaron por ser consideradas muy relevantes.

La segunda fase destinada a la lematización del texto, consiste en sustituir una palabra por su lema, a fin de concentrar la información semántica del texto, al mismo tiempo transformar las formas verbales al infinitivo, los sustantivos al singular y los adjetivos al masculino singular.

La tercera fase, consiste en analizar la polaridad de la oración utilizando LIWC (versión en español). En esta etapa se realiza un análisis sintáctico de las oraciones para establecer una relación entre sus componentes, formando un árbol de vínculos. Cada texto es segmentado en oraciones y éstas, a su vez, en palabras (*tokens*) para luego etiquetarlas según la morfosintaxis de cada una de ellas. Ejemplo: sea la oración: “ella quería que le contaras la verdad” el análisis de la oración muestra el árbol de la figura 2.

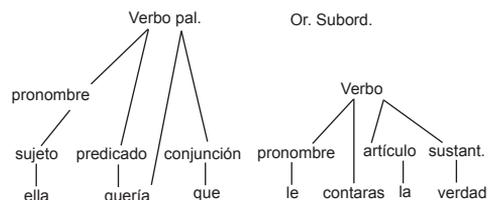


Figura 2. Árbol morfosintáctico de oración El proceso final, destinado al análisis semántico, se apoya en el lexicon a fin de considerar la polaridad de cada palabra y calcular el promedio global, dando como resultado un valor de polaridad: muy positivo, positivo, negativo, muy negativo o neutro; éste último en el caso en que la oración tenga igual número de palabras positivas y negativas. Existe un proceso particular al tratar palabras de negación a fin de interpretar su alcance y la implicación en la oración resultante. De igual forma, existen procesos para tratar las palabras intensificadoras como 'muy', 'poco', 'demasiado', etc.

Como herramientas de software se utilizaron SPSS Analitycs, LIWC2001 y librerías de Phytion.

IV. Muestra

La investigación se centró en el proyecto "Medidas Heurísticas para el Tratamiento de Fobias en Ambientes de Realidad Virtual" de la Universidad Mayor de San Andrés, donde la población voluntaria estuvo constituida por 179 estudiantes universitarios mayores de 16 años diagnosticados con distintos grados de FS. La selección de la población voluntaria se trabajó en tres etapas. La primera orientada a la preselección a partir del uso de cuestionarios e inventarios; se presentaron dos inventarios y un cuestionario validados internacionalmente: escala de Ansiedad Social de Liebowitz (*Liebowitz Social Anxiety Scale LSAS*), Inventario de Fobia Social (*Social Phobia Inventory SPIN*), Cuestionario de confianza para hablar en público (*Personal Report of Confidence as Speaker PRCS*). La segunda etapa orientada a la confirmación de voluntarios a partir del uso de entrevistas y una prueba complementaria, utilizó la historia clínica y MINI (Entrevista

Neuropsiquiátrica Internacional). La tercera etapa, se orientó al establecimiento de una línea base del comportamiento fóbico, fue completada sólo por aquellos sujetos seleccionados en las etapas anteriores. Los instrumentos utilizados con este propósito fueron los siguientes: Cuestionario de Evaluación Multimodal (Lazarus), Formulario de autorregistro y Jerarquía de ansiedades. De este grupo se eligieron de manera intencionada 48 sujetos para el presente estudio.

Todos los pacientes seleccionados fueron tratados con terapia de desensibilización y exposición en ambientes de RV. Durante el proceso de la terapia, éste se complementó con tareas fuera de las sesiones. Estas tareas tuvieron el objetivo de practicar lo aprendido mediante exposiciones graduales en diferentes situaciones reales. Las experiencias fueron narradas por los propios pacientes en diarios digitales, instalándose en sus celulares una aplicación (App) que facilitó el registro y envío automático a la Base de Datos.

Corpus

El corpus está compuesto por dos versiones. La primera (corpus I) se construyó a partir de 570 narrativas correspondientes a todos los pacientes seleccionados para el estudio. Este corpus en total presenta 114.560 palabras. La segunda versión (corpus II) es una colección de los escritos de cada paciente, de tal manera que se tiene un corpus cronológico por cada uno de ellos. Este corpus individualizado tuvo en promedio 4500 palabras con un 35 % de palabras distintas. En el proceso de lematización se descartaron algunas palabras innecesarias llegando a un total aproximado de 4000 palabras para el análisis. Las estadísticas del corpus general se observan en las tablas 3 y 4

Tabla 3.
Estadística general del corpus

Palabras	Corpus
Tokens	114560
Tipos	147
long promedio de palabras	5
Oraciones	1289
long oraciones	10
Paráfrasis	14500
Long paráfrasis	6
palabras de long1	2410
palabras de long2	56023
palabras de long3	56892
palabras de long4	45286
palabras de long5	36895
palabras de long6	5897
palabras de long7	3586
palabras de long8	2897
palabras de long9	1589
palabras de long10	2589
palabras de long11	562
palabras de long12	456
palabras de long13	789
palabras de long >	452

Tabla 4.
Estadística del corpus individual

Criterio	Número
Número total de palabras promedio por paciente	4500
Número de palabras distintas	1105
Número de palabras retenidas	895
Número total de textos	480

V. Resultados

En la tabla 5 se observan el número de sesiones de evaluación y el número de sesiones de tratamiento brindado a la población voluntaria.

Del total de pacientes tratados, en un periodo de dos años, el 58% fue dado de alta, el 26% abandonó la terapia y un 16% no finalizó con éxito el tratamiento (figura 3).



Figura 3. Resultados de la terapia

Tabla 5.
Media, mediana, moda y rango de sesiones de evaluación y tratamiento

	Sesiones de evaluación	Sesiones de tratamiento
Media	3,5	14
Rango	2 a 5	1 a 21
Mediana	4	10
Moda	3	8

Resultados Corpus I

Analizar la frecuencia de las palabras permite elegir las formas gráficas a eliminar por considerarse innecesarias, además, muestra la tendencia en el corpus. El análisis del corpus general presenta las palabras más frecuentes en la tabla 6.

Tabla 6.
Palabras más frecuentes en el Corpus I

Palabra	Frecuencia	Palabra	Frecuencia	Palabra	Frecuencia
Aburrir	79	entender	83	nervioso	259
Actuar	101	error	108	no	205
agachar	51	estrés	251	pensamiento	136
alrededor	24	excusa	189	peor	286
amigo	68	exposición	89	pertenecer	138
ansiedad	195	fallar	147	poder	75
asistir	96	familia	73	producir	159
atención	114	ganar	46	respirar	194
atrever	95	ganas	52	sanción	279
caras	56	hablar	98	sentir	107
cerca	41	incómodo	251	sonrojar	306
correcto	256	inferior	428	temblor	453
debate	26	integrar	108	temor	689
decir	179	mal	205	tenso	659
detrás	16	micrófono	28	tonto	155
disminuido	276	miedo	163	tranquilidad	205
dolor	98	nadie	82	ver	305
enfrentar	103	negativo	322	vergüenza	504

El análisis de las palabras que componen el Corpus I, según las categorías del léxico, se presentan en la figura 4, donde se puede observar las categorías más utiliza-

das al inicio de la terapia y su diferencia con las categorías empleadas al finalizar el proceso.

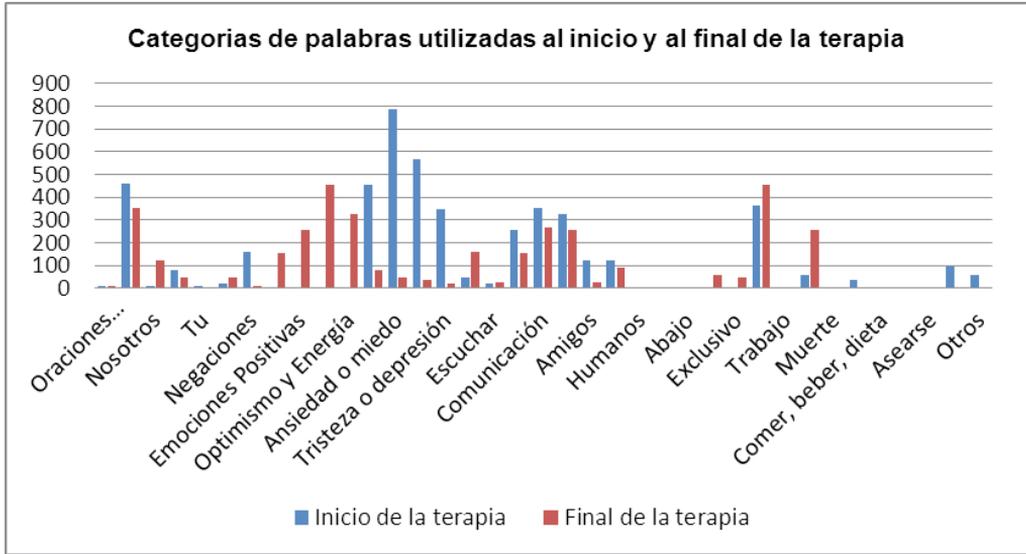


Figura 4. Categorías de palabras utilizadas en el Corpus I

Resultados Corpus II

De forma individual, para cada paciente, se identificaron las categorías utilizadas siguiendo la cronología de las narraciones

(figura 5); es decir, se analizaron 48 diarios personales según la categorización de las palabras utilizadas.

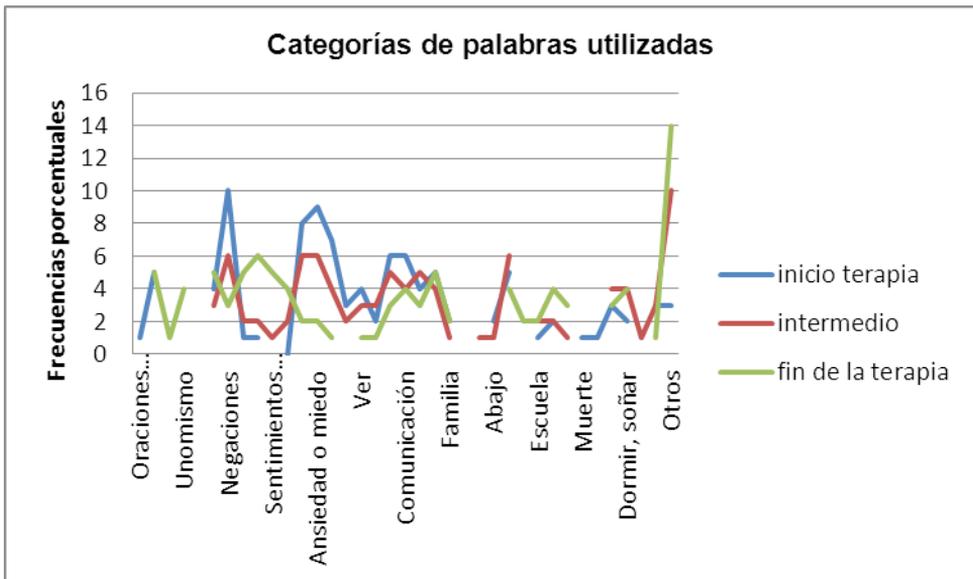


Figura 5. Categorías utilizadas por un sujeto en 3 sesiones

Cada una de las narraciones fue analizada siguiendo la metodología descrita. Por razones de confidencialidad, no se exponen los textos originales, sin embargo, para fines de comprensión, se resume uno de ellos a manera de ejemplo. El texto corresponde al inicio del tratamiento y se trata del temor que se experimentó en las aulas universitarias al momento de exponer un trabajo frente al docente y los compañeros, tarea que no se cumplió por el problema de ansiedad y el dolor de cabeza manifestado. El texto está compuesto por 623 palabras. El análisis estadístico identificó las palabras más frecuentes y su relevancia en el contexto (tabla 7).

Tabla 7.
Frecuencia de palabras en el Corpus II

Frecuencia y Relevancia de palabras		
Palabra	Frecuencia	Relevancia
Docente	3	100
Cabeza	2	66
Crítico	2	66
Frente	2	60
Comparación	3	67
Clase	2	55
Compañero	1	45

En el mismo texto se identificaron los temas tratados y su relevancia porcentual (tabla 8).

Tabla 8.
Temas identificados

Identificación de temas tratados	
Tema	Relevancia
educación-profesor	100%
salud-cefalea	33%
Sociedad	33%

Cada oración fue analizada según su estructura sintáctica. A su vez, para cada una de las palabras que componen la oración se identificó su categoría y se asoció el sentimiento encontrado en el lexicon de LIWC (tabla 9).

Tabla 9.
Lista de palabras, categorías y sentimientos

Polaridad emocional asociada a las palabras		
Concepto	Categoría	Sentimiento
Exposición	evento	negativo
Escenario	lugar	positivo
Comparación		positivo
Docente	entidad	negativo
Cuenta	producto	neutral
Dolor	sentimiento	negativo
Clase	entidad	
Estado	lugar	
Cálculo		neutral
Cabeza	cuerpo	negativo
Trabajo	entidad	muynegativo
Resultado		positivo
Cara	cuerpo	negativo
Critic		negativo
Salida	lugar	positivo
Sorteo		positivo
Frente		negativo

El cálculo de polaridad emocional basado en las palabras que conforman el texto y considerando su estructura, establece que el sentimiento global es negativo en un 86%.

Este mismo sujeto produjo 18 narraciones en total. El análisis cronológico de los textos, según la valoración global, muestra la reducción de la polaridad negativa a medida del avance de la terapia (figura 6).

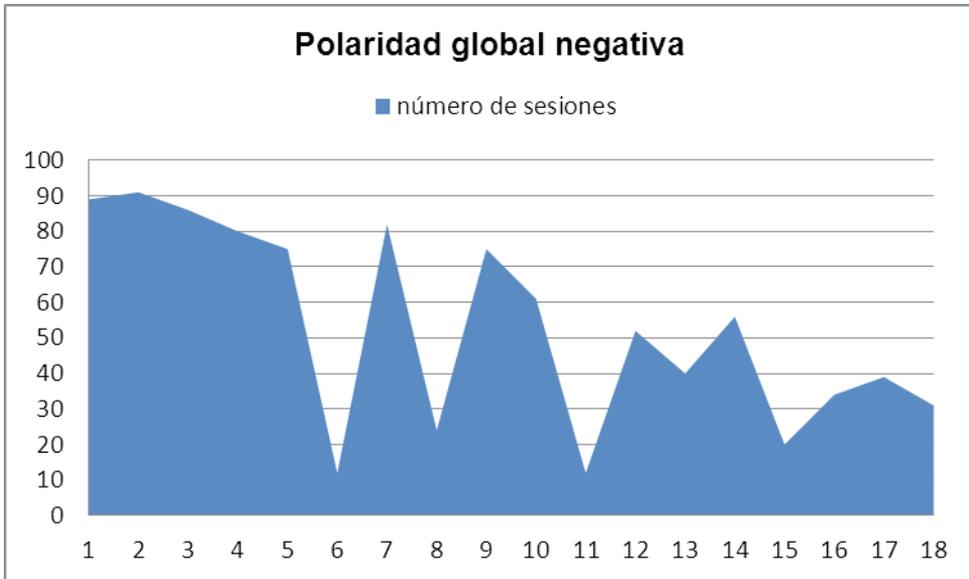


Figura 6. Polaridad global negativa de un texto

Siguiendo con el análisis del mismo sujeto, en la sesión 8, el texto comenta una actividad de tipo taller, donde se expuso un material para la reflexión, aunque sintió que “las piernas le temblaban” salió adelante y concluyó con el trabajo. En este texto se observan los conceptos asociados y el valor sentimental de cada palabra (tabla 10). La evaluación global del mismo se asocia con una negatividad de un 28%, por lo que se observa la presencia de un mayor número de palabras positivas.

Tabla 10.
Lista de palabras, categorías y sentimientos

Polaridad emocional asociada a las palabras		
Concepto	Categoría	Sentimiento
interactivo		positivo
estreslaboral		negativo
disponibilidad		positivo

Grupo	organización	positivo
asistencia		neutral
Espacio	localidad	neutral
Colegio	organización	
Merienda	product	positivo
Pierna	cuerpo	negativo
Taller	localidad	negativo
Actividad	proceso	positivo
Trabajo	otra entidad	
Persona	persona	negativo

Para comparar la polaridad negativa de tres sujetos, se tomaron los valores promedio del área de negatividad a fin de evitar los valores máximos o mínimos. Al superponer las líneas de estas curvas, claramente se puede observar que la polaridad negativa tiende a disminuir en relación al avance terapéutico (figura 7).

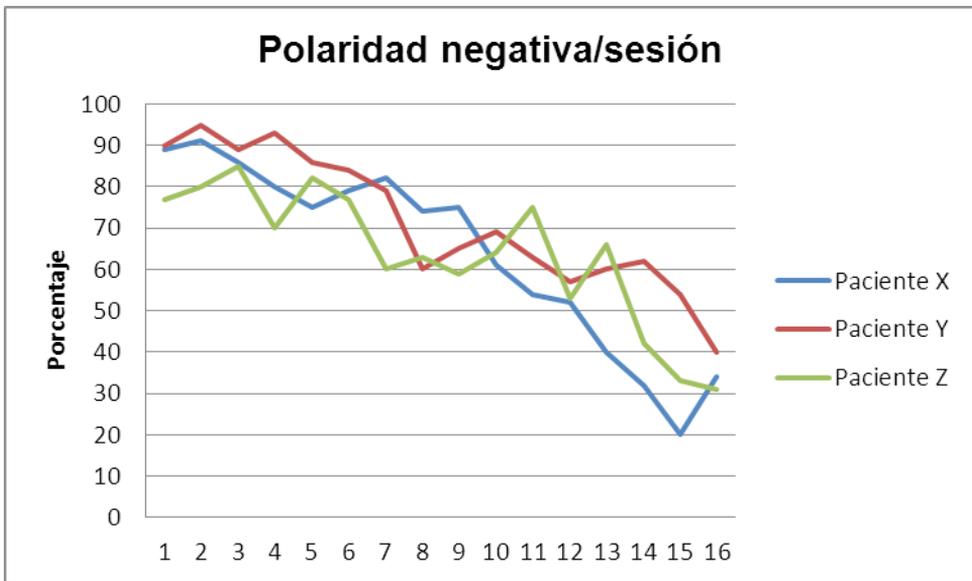


Figura 7. Polaridad emocional global de tres sujetos

VI. Discusión

El proceso de clasificación de las palabras en las diferentes categorías del lexicon, presenta algunos errores de interpretación. Por ejemplo, en la oración “la clase estaba totalmente llena” se interpretó la palabra “estado” como un lugar, asociada al concepto de nación, país, lugar geográfico. Estas palabras ambiguas han dado lugar a 27% de interpretaciones erróneas. La variabilidad de la polaridad emocional durante el proceso terapéutico muestra un espectro muy dinámico, sin embargo, se observa una tendencia clara respecto a la diferencia entre las evaluaciones de inicio y final del proceso coincidiendo los resultados con el post test y el criterio de alta por parte del terapeuta.

La escritura mediante un *smatphone* implica cierta incomodidad, brevedad, añadidura de nuevas expresiones como abreviaturas (qps), repetición de caracteres (noooo) y se evitan reglas de acentuación y de puntuación, lo que deriva en un análisis sintáctico más complejo, razón por la

que no se automatizó en su totalidad este paso. Otra limitación es el hecho de no considerarse el estilo de la escritura de mensajes en medios digitales como el uso de mayúsculas, que implica elevar el tono de voz. Otro tanto sucedió con el proceso de desambiguación del significado de las palabras.

Existen muchas expresiones que actúan como una sola unidad de significado y no deberían separarse. Entre éstas se tienen: ‘estrés laboral’, ‘de manera que’, ‘en absoluto’, ‘por lo tanto’, etc. El tratamiento automatizado de estos casos es complejo por lo que, una vez más, se necesitó la intervención humana para generar otro diccionario conveniente de palabras compuestas considerando el contexto del estudio.

VII. Conclusiones

La muestra del estudio es referencial, no existen datos estadísticos ni estudios de casos de fobia social en Bolivia para su comparación y generalización.

Los datos demuestran una relación significativa entre el número de sesiones y el decremento de la polaridad negativa, estableciéndose una relación cuantitativa entre el progreso de la terapia y la reducción de los síntomas de ansiedad, síntomas interpretados desde el uso del vocabulario en la producción escrita. Asimismo, el cambio de categorías en el espectro de clasificación de las palabras puede ser interpretado como el cambio emocional en la escritura.

El cálculo de la polaridad emocional y en general la Minería de Textos podría presentar ciertas ventajas con relación a otras herramientas que tienen el propósito de medir algún aspecto del campo psicoterapéutico. Estas técnicas pueden ser utilizadas en poblaciones diferentes sin necesidad de validar su aplicación. Los diccionarios con el valor sentimental de cada palabra están disponibles en diferentes idiomas por lo que no requieren la adaptación cultural para establecer la validez de la medida.

Referencias Bibliográficas

- Abregú, F. (2010). *Elementos para un análisis del discurso*. Entrevista a Evo Morales La Paz: Instituto de Estudios Bolivianos.
- Behar, J. (1993). *Anuario de Psicología, 1993, No 59, 61-78, Universidad de Barcelona*. Retrieved Julio 11, 2014, from Aproximación al análisis textual informatizado: <https://dialnet.unirioja.es/servlet/articulo?codigo=2946600>
- Bornas, X., Rodrigo, T., Barceló, F., & Toledo, M. (2002, Febrero). *International Journal of Clinical and Health Psychology, vol. 2, núm. 3, febrero-marzo, 2002, pp. 533-541*. Retrieved Abril 23, 2014, from Las nuevas tecnologías en la terapia cognitivo-conductual: una revisión: <http://www.redalyc.org/articulo.oa?id=33720311>
- Botella, C., García-Palacios, A., Baños, R., & Quero, S. (2016, Septiembre). *ResearchGate*. Retrieved Septiembre 17, 2016, from Cuadernos de Medicina Psicosomática y Psiquiatría de Enlace: [file:///C:/Users/rv6/Downloads/Cuad%20N%C2%BA%2082%20trabajo%202%20\(1\).pdf](file:///C:/Users/rv6/Downloads/Cuad%20N%C2%BA%2082%20trabajo%202%20(1).pdf)
- Castillo, J., Cárdenas, M., Curti, A., & Casco, O. (2015). *Repositorio Institucional de la Universidad Nacional de La Plat*. Retrieved Julio 20, 2016, from Software para asistir en la creación de corpus para sistemas de análisis de texto no estructurado: <http://sedici.unlp.edu.ar/handle/10915/45466>
- Césari, M. (2007). *Cartografiado de textos: protocolo de exploración y visualización de datos textuales aplicados a la minería de textos*. Mendoza: Tesis de Maestría no publicada.
- Choi, Y., Jang, D., Ku, J., Shin, M., & Kim, S. (2004, Julio). *CyberPsychology & Behavior. July 2004, 4(3): 349-354*. Retrieved Mayo 10, 2014, from Short-Term Treatment of Acrophobia with Virtual Reality Therapy (VRT): A Case Report: <http://online.liebertpub.com/doi/abs/10.1089/109493101300210240>
- Eíto, R., & Senso, J. (2004). *El profesional de la información, ISSN 1386-6710, Vol. 13, Nº 1, 2004, págs. 11-27*. Retrieved Octubre 10, 2014, from Minería Textual: <https://dialnet.unirioja.es/servlet/articulo?codigo=801453>
- García, E., Rosa-Alcazar, A., & Olivares, P. (2011, Diciembre). *Terapia Psicológica 2011, vol. 29, No 8, 233-243*. Retrieved Enero 14, 2015, from Terapia de Exposición Mediante Realidad Virtual e Internet en el Trastorno de Ansiedad/Fobia Social: Una Revisión Cualitativa: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-48082011000200010

- Labrador, F., & Ballesteros, F. (2011). *Psicothema*. Retrieved Febrero 10, 2014, from *Psicothema* 2011, vol.23, nro. 4, pp. 560-565: <http://www.psicothema.com/psicothema.asp?id=3922>
- LIWC. (2011). *Linguistic Inquiry and Word Count*. Retrieved Junio 25, 2015, from La aplicaci{on de LIWC: <http://www.liwc.net/liwcespanol/liwcdescription.php>
- Luque, L. (2009, Mayo 12). *Ciberterapia y realidad virtual en la terapia psicológica*. Retrieved Junio 25, 2014, from *Revista Psicología científica*: <http://www.psicologiacientifica.com/ciberterapia-realidad-virtual-psicologia/>
- Meaning Cloud. (2015, Marzo). *Introducción al análisis de sentimientos (minería de opiniones)*. Retrieved Enero 12, 2016, from *Introducción al análisis de sentimientos (minería de opiniones)*: <https://www.meaningcloud.com/es/blog/introduccion-al-analisis-de-sentimientos-mineria-de-opinion>
- Perez, S., & Araujo, M. (n.d.). *Universidad Central e Venezuela*. Retrieved Noviembre 15, 2014, from *Análisis de sentimientos*: <https://www.overleaf.com/articles/analisis-de-sentimientos/knhvsjtbnyv/viewer.pdf>
- Plaza, M. (2008, Junio). *Uso de ambientes virtuales y selección de parámetros de medidas en la aplicación para el tratamiento de fobias*. Retrieved abril 26, 2009, from *Scielo*: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0122-34612008000100003&Ing=es&nrm=iso
- Rivera, C., & Bernal, G. (2008, Enero). *Revista Puertorriqueña de Psicología v.19 San Juan 2008*. Retrieved Agosto 20, 2016, from *La medición en el cambio psicoterapéutico: hacia una práctica psicológica basada en la evidencia y la medición*: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1946-20262008000100010
- Santander, P. (2011, Septiembre). *Cinta de Moebio, núm. 41, septiembre, 2011, pp. 207-224*. Retrieved Enero 15, 2015, from *Por qué y cómo hacer Análisis de Discurso*: <http://www.redalyc.org/articulo.oa?id=10119954006>
- Vilares, D., Alonso, M. A., & Gómez-Rodríguez, C. (2013). *Sociedad española para el procesamiento del lenguaje natural. vol50. 2013*. Retrieved Marzo 22, 2014, from *Clasificación de polaridad en textos con opiniones en español mediante análisis sintáctico de dependencias*: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4701>
- Vincelli, F., Anolli, L., Bouchard, S., Wiederhold, B., Zurloni, V., & Riva, G. (2004, Julio). *CyberPsychology & Behavior. July 2004, 6(3): 321-328*. Retrieved Junio 15, 2014, from *Experiential Cognitive Therapy in the Treatment of Panic Disorders with Agoraphobia: A Controlled Study*: <http://www.ncbi.nlm.nih.gov/pubmed/12855090>
- Wikipedia. (2015). *Wikipedia*. Retrieved mayo 12, 2015, from *Análisis de sentimiento*: https://es.wikipedia.org/wiki/An%C3%A1lisis_de_sentimiento

Fecha de entrega: 22 de Septiembre de 2016

Manejado por: Prof. Luis Hugo Jiménez Carrasco

Fecha de dictamen: 4 de Octubre de 2016