

Aplicación Web para el análisis de Distribución de Frecuencias empleando Netnografía

Norma Leticia Hernández Chaparro^{1,1}, Luis Ángel Reyes Hernández¹

¹ Instituto Tecnológico de Orizaba, Av. Oriente 9 No. 852
Col. Emiliano Zapata, 94320 Orizaba, Veracruz
{norma.hdez16, l.a.reyes.h}@gmail.com

(*Paper received on June 30, 2013, accepted on August 15, 2013*)

Abstract. El presente artículo propone una aplicación *Web* dirigido al sector educativo tomando como caso de estudio la materia de Probabilidad y Estadística, puntualmente el tema de Distribución de Frecuencias impartida en todas las carreras en el Instituto Tecnológico de Orizaba donde existe un déficit de atención por parte de los alumnos ya que hace uso de datos ficticios y en consecuencia un impacto en los índices de aprobación. El objetivo es proporcionar una herramienta de ayuda para los maestros del instituto, utilizando Netnografía para establecer qué datos extraer de una Red Social como Facebook o Twitter y trabajar con datos de una fuente real. La aplicación *Web* propuesta es desarrollada mediante el lenguaje de programación JSP que se encarga de hacer la interfaz entre el usuario y la Red Social a elegir, extraer un conjunto de parámetros específicos y realizar el análisis de Distribución de Frecuencias.

Keywords: *Aplicación Web, Distribución de Frecuencias, Netnografía, JSP.*

1 Introducción

El gran impacto que tienen las aplicaciones *Web* usadas en muchos servicios hace que cada día desarrolladores de este tipo de aplicaciones provean mejoras e implementen interfaces cada vez más interesantes y creativas a la vista del usuario. Las aplicaciones *Web* permiten el envío de información importante, en conjunto con los colores, imágenes, videos y demás, genera interés en los usuarios para continuar con la navegación dentro de estas aplicaciones, factor muy importante puesto que sin ella el usuario queda fuera o perdido en la aplicación sin darse cuenta. Dentro del programa de la materia de Probabilidad y Estadística impartida en el Instituto Tecnológico de Orizaba, existe un tema en especial que es la Distribución de Frecuencias, en la cual se analiza una serie de datos que posteriormente de forma ordenada y agrupada permite hacer predicciones a futuro. Estos datos normalmente son ficticios no provienen de una fuente real y provoca la falta de interés en el alumno ya que no observa el alcance que el análisis de Distribución de Frecuencias pueda tener, además del desconocimiento de las áreas de aplicación. La Netnografía es el estudio de las redes sociales, proporciona información de estos grupos basada en ciertos criterios, la importancia de la

Netnografía ha incrementado conforme el paso del tiempo y da a conocer ventajas de llevar acabo alguna investigación basada en esta rama de la ciencia de la Etnografía.

2 Trabajos Relacionados

La Etnografía más que una metodología, es una forma de vida [1], en un principio utilizada para analizar grupos de aborígenes. La función de la Etnografía es el análisis del comportamiento así como prácticas de un grupo en específico de personas, sugiere la participación con este grupo para generar un contraste de lo que las personas hacen y dicen. Internet es uno de los recursos más poderosos disponible para el uso de la Etnografía, un estudio profundo de los múltiples usos de Internet, encuestas online, el manejo de sitios *Web*, videoconferencias, acceso a revistas en línea y materiales relacionados. Indiscutiblemente hoy en día las herramientas basadas en *Web* son de gran utilidad e indispensables para los etnógrafos, aunque realmente el objeto de estudio son los humanos. Aplicar la interpretación cultural o la perspectiva interna, genera una sensibilidad en cuanto a las necesidades, temores y aspiraciones de la gente.

La investigación Netnográfica descrita en [2] basada en los métodos clásicos de la investigación etnográfica, denota la sociabilidad online por la comunicación y la cultura presente en comunidades o redes sociales, la Netnografía es una disciplina nueva y en proceso de construcción creada para entender lo que millones de personas expresan a lo largo del día en las diversas redes sociales existentes, obtener el carácter normal de las ciberculturas es la parte modular de la Netnografía que es la referencia a la cultura dentro de las redes sociales o mejor dicho el ciberespacio, analizada por muchos netnógrafos con el fin de comprender también el ciberespacio, cómo se dió el origen de las comunidades nativas online y el fenómeno que Manuel Castells ha denominado la autocomunicación de las masas que consiste en que este tipo de comunicación no es unidireccional si no que permite a una persona qué quiere saber y qué información desea obtener.

El reconocido experto Robert V. Kozinets en [3] emplea un término basado en la Etnografía conocido como Netnografía que consiste en seis pasos: la planificación de la investigación, introducción, recopilación de los datos, interpretación, garantizar estándares éticos y representación de la investigación, además se enfoca en la recolección de información como el comportamiento de los individuos en las redes sociales exponiendo conocimientos, sentimientos y el por qué de la existencia de esos grupos que se forman por tener identidad compartida o ciertos intereses en común, estos datos son importantes y de alto impacto para el marketing y otras áreas, aunado a la investigación cualitativa la cual busca obtener datos no cuantitativos con el fin de explorar las relaciones sociales y describir la realidad que los integrantes de dichas redes sociales viven. Se ha designado a Internet como objeto de estudio que la Netnografía utiliza para investigar cómo se comporta un “consumidor” de servicios en una comunidad virtual mediante filtros inteligentes logrando la esencia de la opinión sincera y espontánea que ellos tienen sobre cualquier tema, el análisis a obtener depende de la finalidad que se persigue o para qué realizar la investigación, es de

suma importancia para empresas preocupadas por conocer que piensan los usuarios de los productos o servicios que ofrecen, esto permite a las empresas medir, juzgar, valorar las opiniones y realizar mejoras.

El proceso de la Netnografía se manifiesta en dos fases especificadas en [4], la primera (proceso automático) las computadoras conectadas en red indagan en foros, chats, listas de distribución, grupos de noticias y todas las vivencias relacionadas con el tema de análisis de consumo, posteriormente se convierten en información filtrada de acuerdo a interés y credibilidad. Programas inteligentes analizan la información y seleccionan la de mayor utilidad para el estudio. En la segunda fase (proceso manual) un grupo de especialistas clasifican la información en comentarios favorables, neutros o no favorables acerca del objeto de análisis. Gracias a lo anterior los expertos emiten conclusiones, recomendaciones y el informe final a la empresa que solicitó la investigación. La ventaja que tiene la Netnografía sobre la Etnografía o cualquier otra técnica es que los errores se eliminan en la encuesta directa para presentar resultados únicos de la opinión del usuario de los servicios o productos que consumió.

La cadena de hoteles Sharm El Sheikh en Egipto solicitó una investigación para medir y analizar los comentarios de los usuarios que se hospedaron en alguno de los hoteles de esta cadena, en [5] se publicaron los resultados obtenidos y arrojaron ocho dimensiones: confort, educación, hedonista, la novedad, el reconocimiento, la relacional, la seguridad y el sentido de belleza. Es importante mencionar el interés producido por realizar una investigación de los antecedentes y de las consecuencias poco investigadas para la construcción de la experiencia del cliente que tiene una gran lealtad con la marca. El turismo juega un papel importante en el tema de la Netnografía, siendo ejemplo pionero de la economía y la experiencia que muestran los consumidores.

De acuerdo a cualquier ciencia aplicada como la informática, la ingeniería, la física y demás, el objetivo principal en [6] es la explicación de teorías de probabilidades, procesos estadísticos y procesos estocásticos donde su comportamiento es no determinista. Temas como frecuencias relativas, probabilidad, análisis combinatorio, probabilidad condicional, independencia, variables aleatorias, función de distribución, la importancia de las variables aleatorias discretas, estimación y pruebas son explicadas ampliamente por el autor. Realiza una aplicación a un caso de estudio tomando en cuenta similares propuestas detalladas.

Actualmente las redes sociales representan un cúmulo de información conformado por opiniones, sentimientos y otros datos que los usuarios inscritos en estas redes proporcionan; en [7] describe algunas de las redes sociales más utilizadas por las personas en todo el mundo como Facebook, Twitter, Linked In y algunas otras más. El cómo combinar datos de estas redes, el uso de las técnicas de análisis y visualización, ayudará a que la búsqueda de información sea la idónea y correcta. La minería de datos está presente por las técnicas empleadas para las áreas de la *Web social*, una de las técnicas más avanzadas es TF-IDF (Term Frequency-Inverse Document Frequency, Término de Frecuencia – Documento de Frecuencia Inversa) basada en la observación de la frecuencia de aparición de los términos. La opción del

uso de tecnologías *Web* como HTML5 y JavaScript toolkits es propuesta para crear visualizaciones interactivas mediante estas tecnologías. Gracias al uso de XFN (XHTML Friends Network, eXtensible HyperText Markup Language Redes de Amigos) mediante los enlaces de una página *Web* se representa las relaciones entre personas, logra explorar las conexiones sociales existentes. La finalidad entonces es descubrir y obtener el máximo provecho de las redes sociales que son una inmensa fuente de información, donde la minería de datos juega un papel importante para lograr ese objetivo.

El gran número de sistemas de autenticación de usuarios y de acceso a los recursos y datos personales se ha convertido en un gran problema mencionado en [8], sobre todo en los servicios en internet relacionados con las redes sociales. Actualmente existen diversos estándares y protocolos para satisfacer necesidades de autenticación y autorización. Debido a que un usuario hace uso de diversas aplicaciones como la consulta de su correo electrónico, redes sociales, entre otras, la mayoría requiere que el usuario se autentique, es decir, se identifique ante la aplicación que normalmente lo hace mediante un usuario y una contraseña. Estas aplicaciones cuentan con su propio sistema de autenticación y gestión de datos personales lo que ocasiona al usuario tener que recordar diversos identificadores y contraseñas para cada aplicación. El uso de los protocolos y tecnologías permite precisamente solventar esas situaciones a las que son orillados los usuarios. Es importante identificar la diferencia entre autenticar, que es la certificación de un usuario en un sistema y autorizar que significa otorgar acceso a un usuario o sistema sobre la información o recursos de algún usuario en concreto. *OAuth 2.0* (Open Authorization, Autorización Abierta) es un protocolo que define como un tercero (consumidor) accederá a los recursos propios de otro usuario, su propósito es que un usuario proporcione acceso a un tercero sobre los recursos que tenga en un servidor sin que el consumidor conozca al usuario ni su contraseña.

En la Tabla 1 se muestra una comparativa de las aportaciones de cada trabajo, técnicas o algoritmos empleados, aplicación *Web*, aplicación estadística y dominio. De los trabajos mencionados la mayoría son empleados para la enseñanza, emplean técnicas o algoritmos y algunos realizan la aplicación de estadísticas pero ninguno hace uso de una aplicación *Web*.

Tabla 1. Comparativa de trabajos relacionados.

Autor	Técnicas o Algoritmos	Aplicación <i>Web</i>	Aplicación Estadística	Dominio
David M. Fetterman, Leonard Bickman.	No aplica	No aplica	Aplica	Enseñanza
Miguel del Fresno.	Aplica	No aplica	No aplica	Enseñanza
Gary M. Bowler Jr.	Aplica	No aplica	No aplica	Enseñanza
O. W. Turpo Gebera.	Aplica	No aplica	No aplica	Marketing
Ismail A. R., Prof. T.C.	Aplica	No	Aplica	Marketing

Melewar, Prof. Arch W.		aplica		
Russell, Matthew A.	Aplica	No aplica	No aplica	Enseñanza
Norma L. Hdez. Ch.	Aplica	Aplica	Aplica	Enseñanza

3 Desarrollo

La Netnografía proporciona un método de dos fases [4], la primera consta de un listado de redes sociales a emplear, recolección de datos y su duración definido en días, semanas y meses. La segunda fase corresponde a la lectura y clasificación de todos los datos obtenidos de acuerdo a la lista de criterios establecidos y por último una evaluación global para la interpretación de las estadísticas obtenidas y redacción de conclusiones que serán entregadas junto con el informe final. El siguiente paso realizado es el desarrollo del estudio netnográfico que permite identificar el objeto de estudio: las redes sociales Facebook y Twitter con sus respectivos criterios de búsqueda mostrados en la Tabla 2, la elección de estos criterios es debido a que se requiere información que pueda contabilizarse, para realizar las operaciones propias de la Distribución de Frecuencias mencionada en [6]. La frecuencia relativa (*f.r.*) es la representación de cada intervalo expresada en porcentaje. La frecuencia acumulada (*f.a.*) indica cómo se van acumulando los datos conforme pasa de un intervalo al siguiente. La frecuencia relativa acumulada (*f.r.a.*) representa el porcentaje acumulado hasta un determinado intervalo.

Tabla 2. Objeto de estudio con sus respectivos criterios de búsqueda.

Twitter		Facebook	
Usuario	Persona concreta	Usuario	Persona concreta
Seguidor	Usuario que sigue a una usuario concreto	Lista de amigos	Listado de los contactos o amistades pertenecientes a un Usuario
Siguiendo	Persona a la que el usuario concreto sigue	Intereses	Temas importantes para el usuario
Tweet	Mensaje formado por mínimo 1 y máximo 140 caracteres publicado por un usuario	Likes	Calificación positiva de usuario hacia un comentario, una fotografía, un tema de interés, una página que proporcione un servicio o de algún otro usuario
Hashtag	Palabra denotada como muy importante precedida del símbolo #	Estado	Conjunto de caracteres que permiten expresar un sentimiento o pensamiento en un instante dado.
		Fotos	Imagen compartida por un usuario

El análisis realizado de la Distribución de Frecuencias permite mostrar los datos recolectados y ordenados en una tabla, la frecuencia absoluta (*f.i.*) es el número de veces que aparece cada dato, la suma de las frecuencias absolutas se representa con *N*,

siendo igual al número total que existe de datos. En la Tabla 3, se muestra un ejemplo de datos analizados mediante Distribución de Frecuencias, tomando como dato extraído el número de “likes” publicados por un usuario de Facebook durante una semana, se observan los días de la semana que representan las siete clases que agrupan y ordenan la información.

Tabla 3. Análisis de Distribución de Frecuencias de una semana para los “likes” de un usuario.

Semana 1	Clase	f.i.	f.a.	f.r. %	f.r.a. %
Lunes	1	70	70	35	35
Martes	2	40	110	20	55
Miércoles	3	30	140	15	70
Jueves	4	10	150	5	75
Viernes	5	5	155	2.5	77.5
Sábado	6	40	195	20	97.5
Domingo	7	5	200	2.5	100
$\Sigma=200 \text{ likes}$					

Una vez identificado el objeto de estudio, se obtuvo la definición de parámetros que son los componentes de cada clase existente de las redes sociales. La definición del universo se determinó para cada una de las redes sociales, en común mantienen la clave del usuario registrado en una cuenta, pero específicamente en el caso de Facebook se determinó también como universo las amistades que tiene un usuario, sus intereses, los “Likes” que provea y el número de estados publicados, para el caso de Twitter, el universo se conforma por el número de seguidores, el número de personas que sigan a este usuario, el número de los “Tweets” enviados así como la cantidad de los “Hashtag” publicados. El desarrollo de la aplicación *Web* está basado en el lenguaje de programación *JSP (JavaServer Pages)*, fue necesario crear una aplicación independiente de la aplicación *Web* en Facebook para solicitar los permisos necesarios al usuario y extraer su información, hay datos que no requieren un permiso en especial que sea registrado en la aplicación de Facebook como el nombre del usuario, la foto de su perfil, su dirección, entre otros, pero para obtener el correo electrónico del usuario, es necesario indicar en la aplicación de Facebook que requiere solicitar permiso especial con el consentimiento del usuario para obtener esa información. La aplicación de Facebook entabla conexión con el protocolo de autorización *OAuth* [8] estándar abierto para incluir seguridad en la autenticación de aplicaciones *Web* y de escritorio, es el puerto de autenticación de credenciales de Facebook y Twitter por mencionar algunas redes sociales, permite a una aplicación cliente actuar en nombre de un usuario concreto para obtener recursos de ese usuario, estos recursos se encuentran en un determinado servidor de recursos. Una vez autenticado el usuario, el protocolo *OAuth* obtiene los permisos otorgados por el usuario y la aplicación *Web* puede acceder a los recursos del usuario para ser extraídos y utilizados para el análisis de *Distribución de Frecuencias*. Cabe mencionar que este protocolo empleado en Facebook, evita el mal uso de una cuenta de usuario. Los elementos básicos de la arquitectura de *OAuth* son los siguientes:

Aplicación cliente: Aplicación que solicita y utiliza la información del usuario para un fin concreto.

Servidor de autorización: Servidor que recoge la información de la aserción, de la aplicación cliente y del recurso al cual se desea acceder y devuelve un código de acceso a los recursos para ser utilizado en el servidor de recursos.

Servidor de recursos: Servidor que permite acceder a los recursos con un código de acceso determinado. En la Figura 1, se muestra el comportamiento del protocolo OAuth en [8].

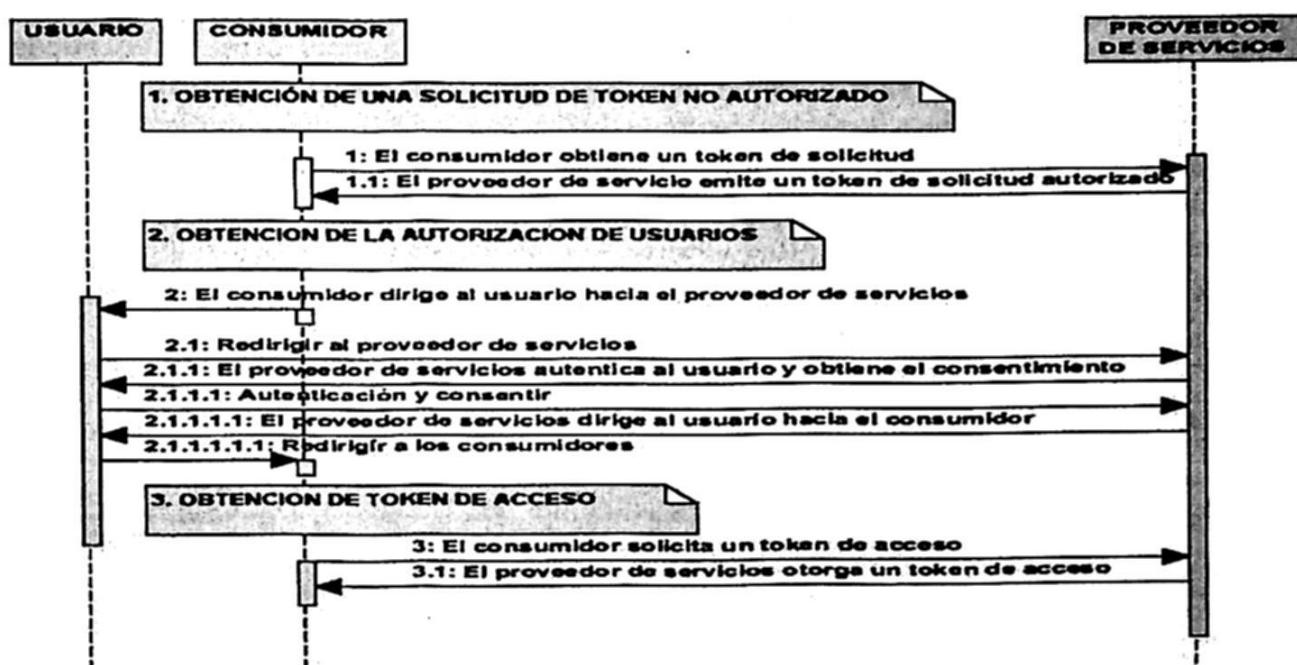


Fig. 1. Secuencia de autorización *OAuth*.

4 Trabajo a Futuro

Realizar la codificación en *JSP* para efectuar la conexión y solicitud de permisos para extraer información de un usuario registrado en Twitter. Aplicar la Distribución de Frecuencias en los datos obtenidos de Twitter y mostrar los resultados correspondientes del análisis.

5 Conclusiones

El uso de las aplicaciones *Web* ha incrementado y con ello la necesidad de contar cada día con mejores aplicaciones para solventar diversos problemas que surgen actualmente como el caso especial observado en el Instituto Tecnológico de Orizaba en cuanto a la materia que se imparte de Probabilidad y Estadística y puntualmente en el tema de Distribución de Frecuencias, debido a que no existe interés por parte de los alumnos al realizar los análisis correspondientes y emplear datos ficticios que no

proveen ningún foco de atención hacia los alumnos, se presenta como solución a este problema utilizar la Netnografía para obtener datos provenientes de una fuente real, como son las redes sociales donde la mayoría o casi todos los alumnos están familiarizados con ellas puesto que son parte de su vida cotidiana, la aplicación *Web* desarrollada permite la interacción con redes sociales, la extracción de información y el desarrollo completo del análisis de Distribución de Frecuencias, de esta manera los alumnos de esta institución podrán valorar la importancia del uso de la Distribución de Frecuencias, ampliar su panorama de las áreas donde pueden utilizar estos conocimientos y la disminución del nivel reprobatorio de los alumnos se observará en un porcentaje significativo y se prevé que con el uso de esta aplicación, los mismos alumnos generen ideas propias para enriquecer la herramienta e inclusive puedan crear aplicaciones con este enfoque para otras áreas empleando la Distribución de Frecuencias. Así mismo esta herramienta busca ser una ayuda completa para los maestros ya que especifican que el promedio de tiempo empleado para enseñar a sus alumnos todo el análisis de Distribución de Frecuencias es alrededor de 3 a 4 semanas de esta forma la aplicación disminuiría ese tiempo de enseñanza a 1 semana aproximadamente.

Agradecimientos

Agradecemos a CONACyT (Consejo Nacional de Ciencia y Tecnología). El gran apoyo del M.S.C. Luis Ángel Reyes Hernández, al Dr. Giner Alor Hernández y a la M.C.E. Beatriz Alejandra Olivares Zepahua por el aporte en el proyecto.

Referencias

1. Fetterman D.M., and Bickman L., "Ethnography: Step-by-Step (Applied Social Research Methods)", Nashville: Debra J. Rog, Westat. 2010
2. del Fresno M., "Netnografía Investigación, análisis e intervención social online", Editorial UOC, Barcelona, 03/2011.
3. Bowler G.M., "Netnography: A Method Specifically Designed to Study Cultures and Communities Online", The Qualitative Report 15, no. 5, 2010, 1270-1275.
4. Turpo Gebera W.O., "La Netnografía: un método de investigación en Internet", edited by la Ciencia y la Cultura (OEI) Organización de Estados Iberoamericanos para la Educación, Revista Iberoamericana de Educación, no. 47/2, 2008.
5. Ahmed Rageh I., Prof. Melewar T.C., Prof. Woodside A., "A Netnography study to uncover the underlying dimensions of customer experience with resort brands". Brunel Business School, Brunel University, UK. 2010
6. Lefebvre M., "Applied Probability and Statistics", 2006, 358 p.
7. Russell M.A., "Mining the Social Web: Analyzing Data from FaceBook, Twitter, LinkedIn, and Other Social Media Sites", O'REILLY, United States of America, 2011.
8. Sánchez J., "Sistemas de Autenticación y Autorización en Internet", UdL, Máster en Interacción Persona-Ordenador, Universidad de Lleida, 2010-2011.