



CPU-e, Revista de Investigación
Educativa

E-ISSN: 1870-5308

cpu@uv.mx

Instituto de Investigaciones en Educación
México

Hernández Maldonado, María Luisa; Domínguez Palmeros, Herminia; Jarvio Fernández,
Antonia Olivia

Análisis estadístico de datos textuales aplicado al uso de redes sociales

CPU-e, Revista de Investigación Educativa, núm. 21, julio-diciembre, 2015, pp. 1-27

Instituto de Investigaciones en Educación
Veracruz, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=283140301002>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto



Revista de Investigación Educativa 21

julio-diciembre, 2015 | ISSN 1870-5308 | Xalapa, Veracruz

© Todos los Derechos Reservados

Instituto de Investigaciones en Educación | Universidad Veracruzana

Análisis estadístico de datos textuales aplicado al uso de redes sociales

Dra. María Luisa Hernández Maldonado

Docente de tiempo completo

Facultad de Estadística e Informática, Universidad Veracruzana, México

lhernandez01@uv.mx

Mtra. Herminia Domínguez Palmeros

Docente

Instituto Tecnológico de la Cuenca del Papaloapan, México

hdomp@hotmail.com

Dra. Antonia Olivia Jarvio Fernández

Coordinadora de Desarrollo de Colecciones

Dirección General de Bibliotecas, Universidad Veracruzana, México

ojarvio@uv.mx

Las preguntas abiertas son frecuentes en un cuestionario. Su utilización es necesaria particularmente cuando se busca conocer las motivaciones y la opinión de los entrevistados sobre algún asunto. Tradicionalmente el tratamiento de este tipo de respuestas consiste en cerrar la variable, complicándose el proceso a medida que el tamaño de la muestra aumenta. En estos casos, se utiliza un análisis estadístico de datos textuales que con ayuda del análisis de correspondencias, permite descubrir tendencias, desviaciones y asociaciones entre individuo y palabras. Para mostrar la metodología se presenta un estudio sobre el uso de redes sociales, tomando en cuenta una pregunta abierta y las características individuales de los entrevistados. El análisis muestra que, eventualmente, algunas personas forman parte de una red social por conocer gente y por comunicarse, por diversión o trabajo, por facilidad, hacer tareas o comunicarse con personas lejanas y maestros. A otros no les gustan pero reconocen la necesidad de utilizarlas.

Palabras clave: Análisis Estadístico de Datos Textuales, Análisis de Correspondencias, Redes sociales.

Recibido: 27 de marzo de 2014 | **Aceptado:** 23 de septiembre de 2014

Open questions are frequent in a questionnaire. Its use is necessary when it is intended to know motivations and interviewees' opinion about some matter. Traditionally, the treatment to this type of responses consists of closing the variable, the process gets complicated as soon as the sample increases. In these cases, it is used a statistical analysis of textual data that, with help of the correspondence analysis, allows to discover tendencies, deviations and associations between individual and words. To show the methodology, it is presented a study of social networks use, considering an open question and individual characteristics of the interviewees. The analysis shows that, eventually some people are part of social networks for meeting people and for communicating, for fun or work, for ease, for doing homework or to communicate with distant people and teachers. Others do not like them but they recognize the need of using them.

Keywords: Statistical Analysis of Textual Data, Correspondence Analysis, Social networks.

Análisis estadístico de datos textuales aplicado al uso de redes sociales

Introducción

A partir de la década de los noventa en que surge el teléfono móvil digital se ha tenido un constante desarrollo tecnológico que ha permitido la instalación de aplicaciones que incrementan la posibilidad de comunicación, como la mensajería instantánea, la videollamada y el uso de redes sociales que surgieron una década después. Estas formas de comunicación mediante un dispositivo móvil, se hicieron posibles con el lanzamiento del *smartphone* (teléfono inteligente).

ComScore, líder en medición del mundo digital, presentó en 2010 un estudio acerca del estado de internet en Latinoamérica. Menciona que en 2009 la red social favorita en América Latina fue Facebook, con un promedio de navegación de tres horas al mes. En particular menciona que en México más del 80% de los usuarios de redes sociales la utilizan (Sutter, 2010).

En la Universidad Veracruzana (UV) antes del presente trabajo no se encontró ningún estudio sobre el uso de redes sociales que permita ver el crecimiento del uso de este medio de comunicación en la comunidad universitaria, aunque se conoce el

interés de algunos investigadores sobre el tema. En un estudio más reciente, realizado al 19% de la población estudiantil de la Universidad Autónoma de Nayarit, Ahuacatlán, se menciona que 98.5% usa una red social y de ellos, 89.5% la red que más usan es Facebook. Sobre las horas diarias de conexión, 43.2% de estudiantes permanece conectado más de una hora y a veces hasta perder la noción del tiempo (Varela, Montaña & Ramírez, 2013, p. 2).

Como parte del estudio sobre la lectura digital en el ámbito de la UV, realizado por la Dra A. Olivia Jarvio Fernández en 2010, se aplicó una encuesta cuyo instrumento contenía una serie de preguntas relacionadas con el uso de redes sociales. Este trabajo tiene como objetivo conocer las opiniones de la comunidad universitaria respecto al uso de redes sociales. La pregunta de investigación que se plantea es: ¿qué razones exponen los miembros de la comunidad universitaria para usar las redes sociales, tomando en cuenta sus características individuales?

Cuando el interés del estudio radica en conocer las motivaciones de los encuestados, se recurre al uso de preguntas abiertas porque ofrece a cada uno de los entrevistados la oportunidad de expresar libremente su propia motivación (Álvarez, 2003, pp. 46-47). La información que se obtiene al realizar preguntas abiertas es diferente a la recogida a través de preguntas cerradas, por lo que ambos tipos de preguntas se consideran complementarias, mas nunca semejantes. Además, tómese en cuenta que “de cada pregunta nacen una o varias variables estadísticas” (Abascal & Grande, 2005, p. 51) por lo que el estudio se ve enriquecido al utilizar ambos tipos de preguntas.

Una de las alternativas en el tratamiento estadístico de este tipo de estudio forma parte de la Minería de Datos Textuales (*Text mining*), cuya finalidad radica en descubrir conocimientos que no existían de manera implícita en un conjunto de textos, sino que se generan al relacionar el contenido de algunos de ellos (Hearst, 1999, p. 3). Los autores que han realizado las mayores contribuciones de la estadística aplicada en la minería de datos textuales llaman a esta alternativa de diferentes formas: Análisis Estadístico de Datos Textuales (AEDT) (Bécue, Lebart & Rajadell, 1992), Lexicometría, Estadística textual (Lebart & Salem, 1994), Análisis estadístico de textos (Lebart, Salem & Bécue, 2000). En este trabajo se le referirá como AEDT.

El AEDT surge “de la relación que se ha dado entre el estudio cuantitativo de los textos literarios y la corriente de la estadística moderna llamada análisis de datos” (Bécue et al., 1992, p. 2); parte del número de ocurrencias de las palabras contenidas en el conjunto de textos provenientes de libros, artículos periodísticos o a partir de las respuestas a preguntas abiertas y su relación con características propias de los encuestados, entre otros. La cadena de tratamiento estadístico habitualmente sigue cuatro etapas: problema, datos, tratamiento e interpretación; este esquema marca las

actividades que un estadístico debe seguir aunque, en la práctica las situaciones que se presentan se pueden enriquecer con un Análisis multivariado (Lebart et al., 2000 pp. 30-31, 73).

Materiales y método

Fuente de datos

A través del correo electrónico se invitó a participar en la encuesta sobre Prácticas Lectoras en los Nuevos Soportes Digitales en la UV a personas registradas en la base de datos del módulo de préstamo de libros del sistema Unicornio, utilizado en las bibliotecas de la UV, “dado que a las bibliotecas de la UV acude el 94% de la comunidad universitaria” (Jarvio, 2011, p. 100). El tamaño de la muestra fue de 641 individuos; 86% corresponde a usuarios que solicitaron préstamo de libros y textos académicos sobre la disciplina en que se desenvuelven; el 14% restante se identifica como lector, pues se caracteriza por solicitar obras literarias además de las académicas requeridas para el desempeño de su disciplina.

Se realizó un muestreo post-estratificado por región, observándose una distribución aproximada a la poblacional, “con sólo pequeñas diferencias en las proporciones, sobre todo en las regiones que tienen el menor tamaño” (Jarvio, 2011, p. 102). Posteriormente se aplicó un filtro sobre la pregunta ¿formas parte de alguna red social?, seleccionando sólo los casos que contestaron de manera afirmativa, quedando el corpus (conjunto de respuestas libres) formado por respuestas de 343 miembros de la comunidad universitaria.

Variables de estudio

El cuestionario se encuentra constituido por 26 preguntas cerradas, siete de las cuales, dependiendo de la respuesta proporcionada por el encuestado, lo hacía pasar a una u otra pregunta abierta. Se generaron para este estudio nueve variables; seis variables categóricas: región, edad, ocupación, área académica, sexo y pertenencia a una red social; dos de tipo textual: horas de conexión y redes sociales a las que pertenece, tratadas de forma tradicional, (transformadas en categóricas) y la variable textual de estudio tratada a través de un AEDT, obtenida a través de las respuestas a la pregunta abierta: ¿Por qué te gusta formar parte de una red social?

Análisis de variables socio-demográficas

Se realizó previo al análisis textual un análisis descriptivo y exploratorio de las variables categóricas que consistió en la construcción de tablas de frecuencias absolutas y gráficas de barras que incluyen frecuencias relativas porcentuales para las variables de tipo categórico.

Se trataron tradicionalmente dos de las preguntas abiertas debido a lo concreto de las respuestas, acorde al siguiente procedimiento: se depuraron las respuestas leyendo e interpretando lo que querían expresar los encuestados, se definieron los criterios de interpretación para clasificar las respuestas, por último se codificó cada una de ellas en cuatro categorías. A partir de este tratamiento se obtuvieron dos nuevas variables categóricas que forman parte del análisis mencionado, permitiendo de esta manera contextualizar la variable textual objeto de estudio.

Selección del vocabulario

El AEDT parte de la construcción de una base de datos que contiene un corpus depurado de acuerdo con lo propuesto por Lebart et al. (2000, pp. 43-51). Posteriormente se procede a la selección del vocabulario con el que se va a trabajar: se aplica un filtro dependiendo del volumen de palabras distintas para eliminar las de más baja frecuencia, se identifican y eliminan palabras herramientas (artículos, preposiciones, etc.), y si es necesario, con la finalidad de evitar ambigüedades se crean multipalabras, por ejemplo, “No_me_gusta”. Se obtiene el vocabulario a estudiar de acuerdo con el recuento de las formas gráficas (comúnmente palabras) y al procedimiento mencionado; además, se seleccionan segmentos de dos o más palabras que tienen una frecuencia importante, de acuerdo con el criterio del investigador (Lebart et al., 2000, pp. 73-94).

Palabras-segmentos característicos y respuestas modales

La unidad léxica (palabra o segmento) característica, es aquella que se encuentra sobre o infrautilizada en el texto. En términos más explícitos, es aquella unidad léxica que presenta una frecuencia significativamente mayor o menor en una parte del texto, comparada con la frecuencia global de la misma unidad léxica presente en el corpus. Para realizar la selección de las palabras sobre o infrautilizadas, se utiliza el modelo hipergeométrico (Figura 1), como herramienta únicamente descriptiva, ya que en este proceso no es posible asumir los supuestos del modelo (Lebart et al., 2000, pp. 167-180).

Se fija un umbral, que puede ser cualquier valor habitual utilizado en los test estadísticos; se eligió el valor 0.1 (por tratarse de un corpus pequeño), motivo por el cual se seleccionaron las palabras y segmentos sobreutilizados cuyo valor test resultó igual o superior a 1.6 y los infrautilizados con valores test iguales o inferiores a -1.6.

$$p(k, k_i, k_j, n) = \frac{\binom{k_i}{n} \binom{k - k_i}{k_j - n}}{\binom{k}{k_j}}$$

	k_{ij}		k_i
	k_j		$k_{..}$

Donde n = subfrecuencia de la palabra i en la parte j del corpus;

k_i = frecuencia de la palabra i en la totalidad del corpus;

k_j = tamaño (número de ocurrencias) de la parte j del corpus;

$k_{..}$ = tamaño (número de ocurrencias del corpus) (o simplemente k)

“Se denota $\text{Prob}(k, k_i, k_j, n)$ la probabilidad así calculada de obtener exactamente n objetos marcados al efectuar la extracción sin reposición de una muestra de tamaño k_j , a partir de una población de efectivo total k , sabiendo que dicha población comporta k_i objetos marcados”.

Figura 1. Modelo hipergeométrico aplicado al análisis de texto

Fuente: Lebart et al., 2000, p. 169.

También se seleccionaron respuestas modales que son respuestas genuinas que por su alta representatividad caracterizan a los individuos. Es decir, se caracteriza a los individuos de acuerdo con las respuestas expresadas en cada una de las categorías que compone la variable elegida, pudiendo elegir para caracterizar a los individuos cualquiera de las variables categóricas. Se busca contextualizar a los individuos de acuerdo con sus respuestas a la pregunta abierta; de esta manera, esta técnica ofrece una solución al aspecto fragmentario y desarticulado de los estudios que se limitan al análisis de palabras aisladas.

La selección de las respuestas modales se basa en dos criterios: el primero corresponde a los valores test más altos de las palabras sobreutilizadas contenidas en las respuestas, el cual favorece respuestas cortas y el segundo criterio, se basa en las dis-

tancias χ^2 y se enfoca a la selección de respuestas largas. Para realizar una revisión minuciosa del tema, se recomienda consultar Lebart et al. (2000).

Análisis de correspondencias (AC)

De acuerdo con Lebart et al. (2000),

el análisis de correspondencias es una técnica de descripción de tablas de contingencia (o tablas cruzadas) y de ciertas tablas binarias (llamadas tablas de 'presencia-ausencia'). Dicha descripción se hace esencialmente bajo la forma de una representación gráfica de las asociaciones entre filas y entre columnas. (p. 75)

Cada dimensión de la tabla cruzada hace posible el cálculo de distancias entre los elementos de la otra dimensión; así, el conjunto de columnas permite definir distancias entre filas y el conjunto de filas define distancias entre columnas. De esta manera, se obtienen tablas de distancias que sirven de base para la representación geométrica y que describen las semejanzas entre las filas y columnas de la tabla cruzada que se analiza.

El AC se aplicó a una tabla léxica agregada (Tabla 1), la cual es una tabla de contingencia, cuyas filas corresponden a las palabras o segmentos repetidos que forman parte del vocabulario seleccionado; las columnas de la tabla se conforman por cada una de las categorías de una de las variables, sobre la cual se tiene el interés de contextualizar las respuestas libres (en este caso, la variable región). La tabla léxica agregada, contiene las frecuencias absolutas con que cada palabra (o segmento repetido) del vocabulario seleccionado, fue pronunciada por el grupo de personas que pertenecen a cada una de las categorías de la variable de interés que complementa el estudio (Lebart et al., 2000, p. 142).

Tabla 1. Tabla léxica a la cual se aplicó un análisis de correspondencia

	Xalapa	Veracruz	Poza Rica- Tuxpan	Córdoba- Orizaba	Coatzacoalcos- Minatitlán
ACTIVIDAD	3	1	0	0	0
ADEMAS	1	5	1	0	0
AMIGO	49	55	14	11	7
AMISTAD	13	3	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮
TENER CONTACTO	4	2	0	0	1
TENGO AMIGO	1	1	1	0	0

Fuente: elaboración propia.

Resultados

Los resultados se presentan de acuerdo con la metodología expuesta. Inicialmente se muestran los resultados del análisis de las variables socio-demográficas; posteriormente, los resultados de la aplicación del AEDT, que incluyen las palabras y segmentos característicos y respuestas modales, y por último, el estudio se complementa con un AC.

Variables Socio-demográficas

La Figura 2 muestra que el mayor número de personas (146 participantes) que contestaron el cuestionario pertenece a la región Xalapa, en contraste con Coatzacoalcos-Minatitlán que tuvo la menor participación, con sólo 20 personas encuestadas. En 2010 estas regiones correspondieron a las de mayor y menor matrícula respectivamente (Jarvio, 2011, p. 103).

Región	Frecuencias absolutas
Xalapa	146
Veracruz	102
Poza Rica-Tuxpan	25
Córdoba-Orizaba	30
Coatzacoalcos-Minatitlán	20
Totales	323

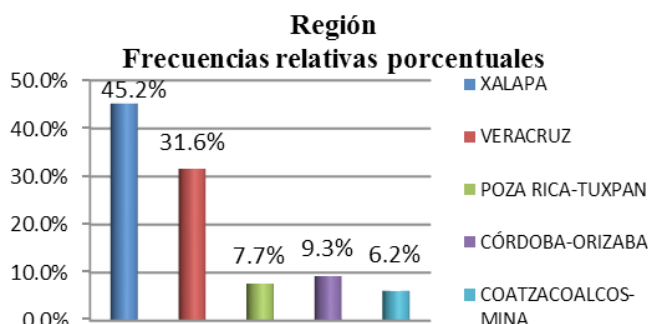


Figura 2. Respuestas obtenidas por región

Fuente: Elaboración propia.

Variable sexo. De los 323 encuestados que respondieron a la pregunta de estudio, 178 son de sexo femenino (55%) y 145 del sexo masculino (45%).

Variable edad. El mayor número de participantes en la encuesta pertenece al grupo entre 17 y 24 años de edad, del cual se obtuvieron 257 respuestas (79.6%); en la categoría entre 25-34 años respondieron 45 encuestados (13.9%). De personas mayores de 45 años se obtuvieron 21 respuestas (6.5%).

Variable ocupación. Se tienen 10 categorías: funcionario de entidad académica, funcionario de administración universitaria, investigador, académico de tiempo completo, académico de asignatura, técnico académico, estudiante de licenciatura, estudiante de nuevo ingreso, bibliotecario y otra ocupación. El mayor porcentaje de participación se encontró en los estudiantes de licenciatura (83.9%). El siguiente porcentaje más alto corresponde a la categoría estudiantes de nuevo ingreso, con 4.6%; en la categoría bibliotecario contestó un 3.4%; las modalidades restantes sumaron 8.1%.

Variable área académica. La mayor participación se dio en el área Económico-Administrativa (32.5%), seguida de cerca por Ciencias de la Salud (27.9%), decreciendo el porcentaje hasta el 2.2%, en la categoría Artes (Figura 3).

Área académica	Frecuencias absolutas
Artes	7
Biológico agropecuaria	21
Ciencias de la salud	90
Económico administrativa	105
Humanidades	47
Técnica	53
Total	323

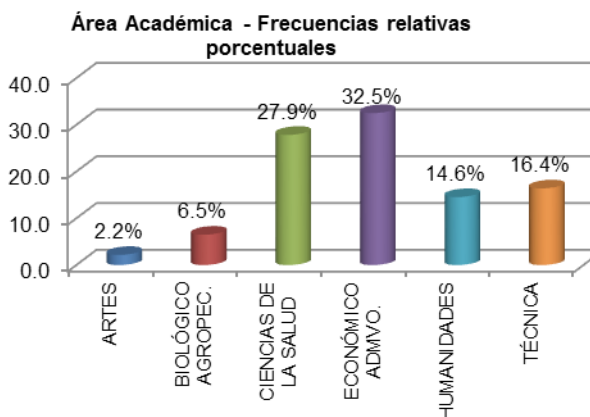


Figura 3. Participación por área académica

Fuente: Elaboración propia.

Variables respectivas al uso de redes sociales

Variable tiempo de membresía. Los datos reflejan que el porcentaje mayor de encuestados (34%) ha formado parte de una red social en el periodo de uno a menos de dos años. El menor de los porcentajes (15%) ha sido miembro de una red social de dos años a menos de tres años (Figura 4).

Tiempo de membresía	Frecuencias absolutas
<1	91
>=1 Y <2	104
>=2 Y <3	47
>=3	67
Total	309

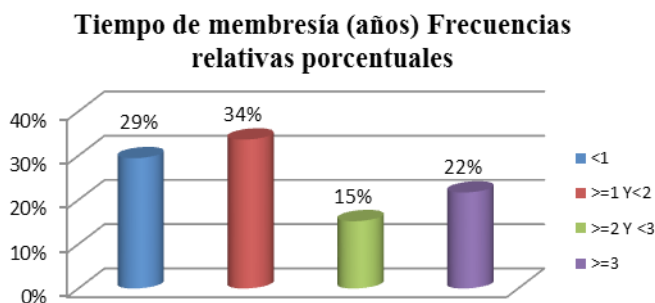


Figura 4. Participación en la encuesta de acuerdo con el tiempo de membresía

Fuente: Elaboración propia.

Variable horas diarias de conexión a la red social. Veintiocho por ciento de los encuestados se conecta diariamente de una a menos de dos horas; 26% manifiesta conectarse más de tres horas, incluyendo algunos casos extremos de conexión de 24 horas (Figura 5).

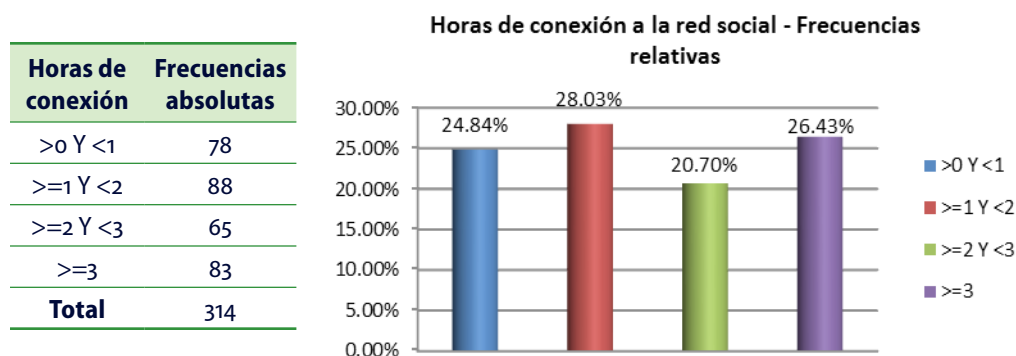


Figura 5. Participación en la encuesta de acuerdo con el tiempo de conexión

Fuente: Elaboración propia.

Resultados del Análisis Textual

El corpus que se creó a partir de las respuestas a la pregunta ¿Por qué te gusta formar parte de una red social?, se conforma por un total de 3,930 palabras contenidas en 343 respuestas; 723 son palabras distintas que corresponden 18.4% con respecto al total.

Selección del vocabulario

Se seleccionó el vocabulario con el que se realizó el AEDT de acuerdo con los siguientes criterios:

1. Se realizaron 45 agrupaciones de palabras provenientes de la misma raíz en una misma forma gráfica (por ejemplo, la palabra *familiares*, mencionada 25 veces, se agrupó con la palabra *familia*, mencionada 14 veces; en este caso, la frecuencia de la palabra *familiares* cambió a 39).

2. Las palabras que por sí solas no tienen un significado claro y que podían generar ambigüedad, se identificaron y se les dio algún tratamiento específico; por ejemplo, la palabra *no*, que fue utilizada en diferentes contextos, cuando se utilizó ligada al *me gusta* se trató como una multipalabra: *no_me_gusta*.
3. Se eliminaron 47 palabras herramienta que suman un total de 1603 frecuencias (por ejemplo: la palabra *de* se utilizó 225 veces). Comúnmente las palabras herramienta tienen altas frecuencias, sin embargo, aportan poca información importante.
4. Se aplicó un filtro para eliminar del vocabulario inicial las formas gráficas con frecuencia < 4. Un fragmento del vocabulario final se muestra en la Tabla 2.

El vocabulario seleccionado es de 1,629 palabras, 122 de ellas diferentes. El estudio se realizó con 41.45% del vocabulario inicial. Posteriormente, se identificaron los segmentos repetidos con una longitud mínima de dos palabras y una frecuencia mínima de tres; el número de segmentos diferentes con esas características fue de 37 y se muestran en la Tabla 3.

Tabla 2. Vocabulario seleccionado

Palabra/segmento	Longitud	Frecuencia
AMIGOS	6	136
PORQUE	6	129
CONTACTO	8	90
PERSONAS	8	50
NO	2	47
FAMILIARES	10	39
COMUNICACIÓN	12	34
⋮	⋮	⋮

Fuente: Elaboración propia.

Tabla 3. Segmentos repetidos

No.	Frecuencias	Palabras	Caracteres	Segmento
1	3	2	13	AMIGO LEJANOS
2	3	2	14	COMPARTIR COSA
3	7	2	13	CONOCER GENTE
4	5	2	15	CONOCER PERSONA
5	3	2	13	CONOCES GENTE
⋮	⋮	⋮	⋮	⋮
37	3	2	11	TENGO AMIGO

Fuente: Elaboración propia.

Palabras, segmentos característicos y respuestas modales

Se obtuvieron palabras y segmentos característicos para cada una de las categorías de las variables sexo, área académica y región. Se eligió el nivel de significancia $\alpha = 0.1$, por lo que se seleccionan las palabras con valor test mayor o igual a 1.6, para palabras sobreutilizadas (palabras de frecuencia significativamente mayor) y menor o igual a -1.6 para palabras infrautilizadas (las de frecuencia significativamente menor). Se eligieron las primeras respuestas modales de cada una de las categorías correspondientes a las variables cerradas antes mencionadas, que contextualizan a la pregunta de estudio.

Variable sexo. Los varones se conectan a la red social porque les parece interesante, se comunican con compañeros, comparten cosas y hablan del tiempo. En las respuestas modales, utilizando el primer criterio de clasificación, se interpreta que los hombres utilizan las redes sociales como un medio de distracción o sólo para pasar el tiempo. De acuerdo con el segundo criterio, se refuerzan las respuestas anteriores expresando que les gusta socializar, hacer nuevas amistades y compartir vivencias. Se puede apreciar en la Tabla 4 que la palabra tiempo es utilizada en dos contextos diferentes; en el primer criterio, está asociada a la “pérdida de tiempo” y en el segundo criterio además de estar asociada a la “pérdida de tiempo” también se relaciona con establecer contacto con personas que hace tiempo no ven.

Las mujeres expresan que el uso de las redes sociales obedece principalmente al gusto por estar en comunicación con su familia; bajo el primer criterio resalta este hecho como una necesidad; en el segundo, además de estar en contacto con familiares que pueden estar lejos físicamente, también se habla de volver a encontrarse con amigos (Tabla 4).

Tabla 4. Respuestas modales correspondientes a la variable sexo

Criterio de Clasificación	SEXO: MASCULINO
Primer: V-test	
1.624	Pérdida de tiempo
1.369	Sólo para promover mis aficiones
1.208	Sólo por diversión
1.105	Un poco de ocio no le hace mal a nadie y conoces personas
Segundo: Chi²	
0.936	Sólo por socializar, contactar y hacer nuevas amistades, perder el <i>tiempo</i> en alguna otra actividad
0.937	Porque veo a personas que tiene <i>tiempo</i> que no veo y desde mucho <i>tiempo</i> conozco, a más personas y chicas guapas
0.942	Porque <i>comparto</i> vivencias e información interesante así como poder socializar con más personas
0.957	Para <i>compartir</i> algunas ideas y conocer a personas
	SEXO: FEMENINO
Primer: V-test	
1.259	Para estar en contacto con mi familia
1.220	Conectarme con mis amigos y familiares
1.215	Por necesidad de comunicación
0.981	Se está más en contacto con la gente sin necesidad de un celular
Segundo: Chi²	
0.856	Porque es un medio para estar en contacto con mis amigos y familiares
0.856	Es una fuente de comunicación para mí y mis amigos además de que puedo estar en contacto con amigos y familiares que están lejos
0.863	Puedo estar en contacto con amigos y familia, volver a encontrarme con amigos
0.869	Para estar en contacto con familia y amigos

Fuente: Elaboración propia.

Variable área académica. En este apartado se presentan como ejemplo dos de las seis categorías que contiene la variable área académica: Biológico-agropecuaria y Técnica, por ser en las que se identifican de manera más clara los contextos.

La Tabla 5 presenta a las personas del área **Biológico-agropecuaria** como aquellas que usan las redes sociales por el gusto de conocer y comunicarse con otras personas, además como medio de compartir fotos. En las respuestas modales (Tabla 6) en ambos criterios, expresan los encuestados que usan las redes sociales como un pasatiempo y porque a través de ellas conocen personas, se comunican con compañeros, amigos y también comparten fotos.

Entre las personas del área **Técnica** se encuentran los que comentaron que no les gusta formar parte de una red social y se observa que tienden a evitar la palabra amigo; sin embargo, expresan usar la red social porque la necesitan para estar al día (Tabla 5).

Tabla 5. Palabras y segmentos característicos de las áreas académicas Biológico-agropecuaria y Técnica de la UV

Palabras o segmentos característicos	Frecuencia interna	Frecuencia global	Valor - Test	Probabilidad
Área académica: Biológico-agropecuaria				
Sobreutilizadas				
CONOCES	3	11	2.004	0.023
COMUNICO	2	5	1.881	0.030
FOTOS	3	13	1.801	0.036
HACE	2	6	1.714	0.043
COMPARTO	2	6	1.714	0.043
Infrautilizadas	No se identificaron palabras infrautilizadas			
Área académica: Técnica				
Sobreutilizadas				
TODO	6	11	2.658	0.004
ESTÁN	7	18	2.106	0.018
NO_ME_GUSTA	3	5	1.838	0.033
DÍA	3	5	1.838	0.033
OTRO	4	9	1.707	0.044
Infrautilizadas				
AMIGO	15	136	-1.652	0.049

Fuente: Elaboración propia.

De acuerdo con las respuestas modales proporcionadas por los encuestados que pertenecen al área **Técnica**, en el primer criterio (Tabla 6), ellos mencionan que no les gusta pertenecer a las redes sociales pero les brinda información y les permite relacionarse socialmente. Estas razones se complementan en el segundo criterio de clasificación, señalando los participantes que no les gustan las redes sociales debido al *spam*, pero les mantiene informados y comunicados con personas lejanas o incluso de la misma ciudad, además de que encuentran amigos allí con los que hacía tiempo no se habían comunicado.

Tabla 6. Respuestas modales en las categorías Biológico-Agropecuaria y Técnica de la variable área académica

Criterio de Clasificación	Área académica: Biológico Agropecuaria
Primer: V-test	
0.965	Un poco de ocio no le hace mal a nadie y conoces personas
0.941	Me comunico con compañeros
0.899	Porque por ese medio me comunico con mis amigos comparto fotos , etc.
0.879	Porque comparto comentarios con compañeros de la facultad y fotos sobre nosotros en prácticas de campo y fiestas
0.705	Para encontrar amigos que no veía hace tiempo
Segundo: Chi²	
0.793	Porque conoces a muchas personas y te enteras de muchas cosas por medio de esas redes además conoces a personas que te pueden ayudar
0.800	Porque por ese medio me comunico con mis amigos, comparto fotos , etc.
0.822	Para comunicarme con mis alumnos y mis hijos que son jóvenes y así trabajo con ellos y no estoy descontextualizada, veo sus fotos , lo que piensan, los cambios en la escritura y más cosas que me ayudan para formarlos y apoyarlos, pues ahora en las redes sociales todo escriben y hasta mal y de más
0.879	Porque comparto comentarios con compañeros de la facultad y fotos sobre nosotros en prácticas de campo y fiestas
0.881	Un poco de ocio no le hace mal a nadie y conoces personas

Criterio de Clasificación	Área académica: Técnica
Primer: V-test	
1.427	No_me_gusta solo lo hice por curiosidad
1.215	Es una fuente de información y de relación social
1.144	Por los juegos
1.101	Es sólo por ocio
1.080	Sólo por diversión
Segundo: Chi²	
0.747	Realmente no_me_gusta del todo por el spam que provoca, pero de cierta manera es más cómodo para hablar de detalles que se pasaron por alto cuando viste a x persona o incluso para mantenerte comunicado y al día con las personas que no están cerca
0.763	Porque mis amigos también están en la red social y porque es entretenido
0.804	Por el contacto con otras personas sobre todo las de otros estados del país u otros países y además lamentablemente se volvió esencial para la comunicación con gente de nuestra misma ciudad
0.862	Porque es un chat pero mejorado
0.864	Por contacto con personas de otros países y organización social

Fuente: Elaboración propia.

Variable Región. Las personas de la región Xalapa utilizan las redes sociales con fines de socialización, conocen gente en algunos casos de otra ciudad y fomentan nuevas amistades; además, tienden a evitar la palabra *amigo*. Aun cuando la palabra sobreutilizada *amistad* y la palabra *amigo* (infrautilizada) tienen la misma raíz son expresadas en contextos culturales diferentes: la intención de la primera se enfoca hacia la obtención de nuevas amistades o simplemente conocidos y la segunda se refiere a amigos con los que se cuenta de hace tiempo. Considerando el primer criterio de clasificación, se obtienen respuestas modales que expresan que el motivo de los encuestados para utilizar las redes sociales es socializar, convivir con amistades y también por cuestiones escolares. Las respuestas modales obtenidas mediante el segundo criterio señalan como motivo la comunicación con amistades y familiares, así como conocer nuevas personas (Tabla 7).

Tabla 7. Respuestas modales de las regiones Xalapa y Veracruz

Criterio de Clasificación	REGIÓN: XALAPA
Primer: V-test	
1.795	Para socializar
1.287	Por convivir con mis amistades y compartir un gusto así como por esparcimiento
1.251	Por motivos escolares amistades y trabajo
0.882	Sólo por socializar , contactar y hacer nuevas amistades, perder el tiempo en alguna otra actividad
0.876	Estar en comunicación y compartir archivos
Segundo: Chi²	
0.851	Pues porque es una forma de comunicación con mis amigas y puedo conocer a nueva gente
0.880	Porque de esa manera tengo más contacto con familiares y amigos que son de otras ciudades
0.881	Por el contacto con otras personas sobre todo las de otros estados del país u otros países y además lamentablemente se volvió esencial para la comunicación con gente de nuestra misma ciudad
0.890	Porque conozco gente y me agrada estar en contacto con mis amistades
Criterio de Clasificación	REGIÓN: VERACRUZ
Primer: V-test	
2.124	Contacto con amigos
1.416	Por estar en contacto con mis amigos
1.416	Para estar en contacto con amigos
1.317	Porque me mantengo en contacto con mis amigos
1.317	Encuentro a mis amigos y me mantengo en contacto con ellos
Segundo: Chi²	
0.811	Contacto con amigos
0.811	Porque estoy en contacto con mis amigos, familia y puedo conocer nuevos amigos, es una forma de comunicación
0.816	Puedo estar en contacto con amigos y familia, volver a encontrarme con amigos
0.822	Para estar en contacto con familia y amigos
0.825	Para estar en contacto con amigos

Fuente: Elaboración propia.

En la región **Poza Rica-Tuxpan** se detecta que el gusto por formar parte de una red social radica en que les permite conocer nuevas personas. Las respuestas modales basadas en el criterio del valor test permiten observar que les agrada pertenecer a redes sociales por los *amigos*. De acuerdo con el segundo criterio, las redes sociales se consideran una herramienta de comunicación económica que facilita estar en contacto con amigos y familiares lejanos, además de atender cuestiones académicas (Tabla 8).

A las personas de la región **Veracruz** les gusta formar parte de una red social porque conocen personas e intercambian información con sus contactos. Tomando en cuenta las respuestas modales en ambos criterios se observa que les gusta pertenecer a las redes sociales por estar en contacto con los amigos, aunque en el segundo criterio también destacan respuestas que incluyen a la familia como razón para conectarse (Tabla 7).

En la región **Córdoba-Orizaba** no es posible identificar de manera clara un contexto ya que aparecen como palabras sobreutilizadas: *distancia*, *no me gusta*, *día*, *alguna* y *poco*; se deduce que, aun cuando no les guste formar parte de una red social, la usan para mantenerse al día. La conjetura realizada para la región **Córdoba-Orizaba** se reafirma y complementa al revisar las respuestas modales, las cuales en el primer criterio mencionan como razón del gusto por las redes que les permiten comunicarse con personas distantes, aunque aclaran que no les gustan las redes sociales. Las respuestas modales basadas en el segundo criterio también expresan que en realidad no les gusta usar las redes sociales pero los mantienen comunicados e informados (Tabla 8).

Tabla 8. Respuestas modales de Poza Rica-Tuxpan y Córdoba-Orizaba

Criterio de Clasificación	REGIÓN: POZA RICA-TUXPAN
Primer: V-test	
1.062	Para buscar amigos
1.062	Por mis amigos
0.636	Porque conoces amigos y estás conectado con tus amigos lejanos
0.561	Facebook te permite contactarte con gente que tiene algún tiempo que no ves o sabes de ella
0.531	Para conocer amigos

Segundo: Chi ²	
0.789	Porque en ocasiones es útil para comunicarme económicamente con familiares que se encuentran en el extranjero, además porque me permite conocer a nuevas personas que podrían convertirse en mis amigos
0.848	Porque te mantienes en contacto con amigos que se encuentran lejos y conoces nuevas personas
0.859	Porque me permite mantenerme comunicado, es cierto, sirve para eventos sociales pero a través de ellas enviamos invitaciones para cursos o eventos relacionados a la escuela, facilita la comunicación y disminuye los gastos por contactar gente
0.867	Me permite estar en contacto con otras personas e intercambiar conocimientos y opiniones en temas en común
0.891	Facebook te permite contactarte con gente que tiene algún tiempo que no ves o sabes de ella
Criterio de Clasificación	REGIÓN: CÓRDOBA-ORIZABA
Primer: V-test	
0.789	Comunicación con algunas personas que es poco probable que vea en el lugar donde viven
0.546	No_me_gusta solo por estar a la moda
0.546	Porque interactúo con amigos míos a pesar de la distancia
0.519	Por enterarme de trabajos escolares
0.457	Solo lo hice para conocer un poco más sobre esta red social aunque en realidad no lo utilizo mucho
Segundo: Chi ²	
0.860	Realmente no_me_gusta del todo por el spam que provoca, pero de cierta manera es más cómodo para hablar de detalles que se pasaron por alto cuando viste a x persona o incluso para mantenerte comunicado y al día con las personas que no están cerca
0.871	Comunicación con algunas personas que es poco probable que vea en el lugar donde viven
0.893	Sólo lo hice para conocer un poco más sobre esta red social aunque en realidad no la utilizo mucho
0.907	Compartir fotos con familia y amigos a distancia
0.922	Por la publicidad que encuentras algunas cosas son de interés y es necesario mantenerse actualizado

Fuente: Elaboración propia.

Para la región **Coatzacoalcos-Minatitlán** se puede decir que utilizan las redes sociales porque su uso es fácil y se comunican con personas lejanas; llama la atención que también las utilizan para atender tareas que es necesario realizar. A través del primer criterio de clasificación se obtienen respuestas modales que implican, como motivo del gusto por las redes sociales, la amistad y el atender asuntos académicos. En el segundo criterio mencionan el distraerse conociendo gente y lugares, tener contacto con antiguos compañeros de escuela e incluyen también asuntos académicos (Tabla 9).

Tabla 9. Respuestas modales de la región Coatzacoalcos-Minatitlán

Criterio de Clasificación	REGIÓN: COATZACOALCOS-MINATITLÁN
Primer: V-test	
0.696	Para tener más amistad
0.672	Porque se puede tener acceso a antiguos compañeros de escuela con los que se había dejado de tener contacto
0.581	Porque es fácil de enviar y recibir tareas y comunicarse con los maestros
0.435	Porque interactúo con personas que están lejos de mí y encuentro diversos puntos de opinión
Segundo: Chi²	
0.865	Porque conoce gente de diferentes lugares, se conocen otros lugares sin necesidad de ir hasta ellos y es una forma de distracción
0.868	Porque se puede tener acceso a antiguos compañeros de escuela con los que se había dejado de tener contacto
0.875	Porque es fácil enviar y recibir tareas y comunicarse con los maestros
0.890	Porque así puedo comunicarme con mis amigos no solamente por los eventos a los que podamos ir como amigos, también por la tarea y dudas que tenemos sobre ella

Fuente: Elaboración propia.

Análisis de Correspondencias

De manera complementaria a los análisis para la variable “región”, se aplicó un AC a la tabla léxica agregada conformada por cada una de las palabras del vocabulario como filas y las categorías de la variable región como columnas; la intersección entre filas

y columnas contiene las frecuencias absolutas con que cada palabra es mencionada por los individuos de cada una de las cinco regiones, La Tabla 10 muestra los valores propios, el porcentaje de inercia explicado por cada uno de los factores y el porcentaje de inercia acumulado. El primer plano factorial se compone de los dos primeros ejes factoriales que explican el 55.68 % de la variabilidad.

Tabla 10. Valores propios del análisis de correspondencias

Eje factorial	Valores propios	Porcentaje	Porcentaje acumulado
1	0.0918	29.22	29.22
2	0.0831	26.46	55.68
3	0.0773	24.60	80.28
4	0.0619	19.72	100.00

Fuente: Elaboración propia.

La Tabla 11 presenta la suma de las contribuciones (CTR) a la formación de los primeros dos ejes factoriales (primer plano factorial), de cada una de las regiones y la calidad de representación, medida a través del coseno cuadrado (Cos^2). Se observa que la región con mayor contribución es Coatzacoalcos-Minatitlán (58.36) y la que tiene mayor calidad de representación (Cos^2) es Veracruz (0.70). Esta información constituye una ayuda en la interpretación de la representación gráfica del AC.

Tabla 11. Ayudas a la interpretación en el primer plano factorial

Región	CTR	Cos^2
Xalapa	33.06	0.65
Veracruz	42.00	0.70
Poza Rica-Tuxpan	20.33	0.28
Córdoba-Orizaba	46.25	0.55
Coatzacoalcos-Minatitlán	58.36	0.65

Fuente: Elaboración propia.

La Figura 6 muestra la representación de las palabras y regiones en el 1er. plano factorial. El Factor 1 opone a los individuos que usan las redes sociales por diversión y por intercambiar información con amigos, con aquéllos que las usan por compartir, socializar y conseguir nuevas amistades, pero también por cuestiones de trabajo. Los primeros se encuentran asociados a las regiones Veracruz y Poza Rica, los segundos se asocian a la región Xalapa.

En el Factor 2 se observa oposición entre las regiones Coatzacoalcos-Minatitlán y Córdoba-Orizaba. Las personas de Coatzacoalcos-Minatitlán utilizan las redes sociales con intención de conocer gente y distraerse, pero también con finalidades académicas: por las tareas y para estar en contacto con los maestros; a diferencia de la región Córdoba-Orizaba que aun cuando usan las redes sociales también por entretenimiento, no les agradan mucho, mas por cuestiones de trabajo y por comunicarse con personas distantes se ven en necesidad de utilizarlas.

Se muestran en la Figura 6 las palabras con mayor Cos^2 ; más claramente, las que tienen la mejor calidad de representación en el primer plano factorial. Así mismo, aparecen palabras que se seleccionaron por su mayor contribución a la formación de los ejes factoriales. Para llevar a cabo la selección de palabras más contributivas se calculó el promedio de contribución a la formación de cada factor, eligiendo las que estaban arriba del promedio (CTR mayor a 0.82). Además, estas palabras tienen la característica de estar bien representadas, con un Cos^2 superior a 0.45.

Subrayadas las palabras con contribución y Cos^2 más altos. Las demás palabras sólo tienen una calidad de representación alta. Enmarcados los segmentos de dos palabras, que fueron considerados como variables suplementarias.

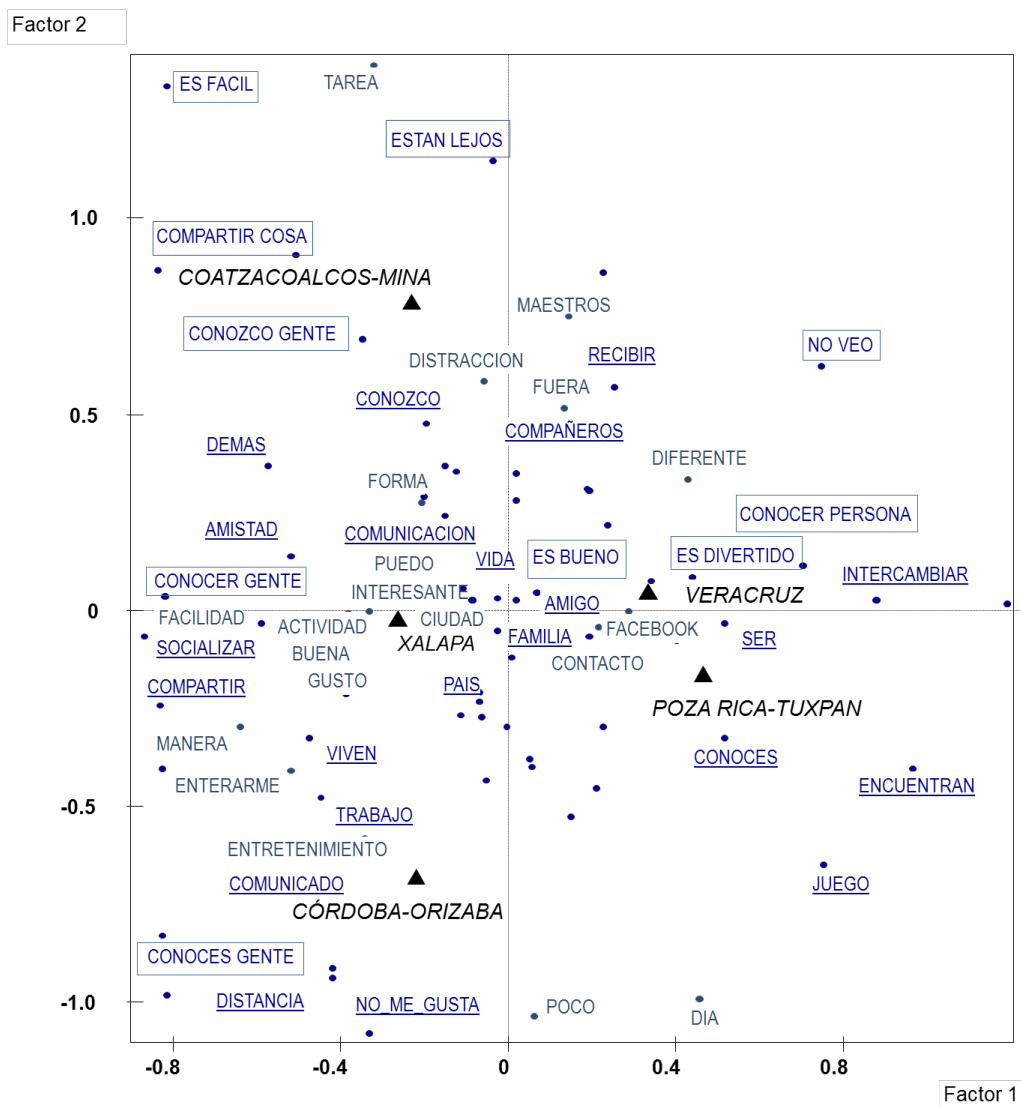


Figura 6. Representación de las palabras y regiones en el 1er. plano factorial

Fuente: Elaboración propia.

Conclusiones

A través de la selección de las respuestas modales se encontraron resultados interesantes que permiten diferenciar los gustos entre hombres y mujeres sobre el uso de las redes sociales. Es notorio que a las mujeres les gusta pertenecer a una red social para comunicarse con su familia, mientras que los varones usan este medio para comunicarse con compañeros y porque les parece interesante su uso y aprovechan o pasan el tiempo entretenidamente.

De manera similar, cuando se aplica este análisis de acuerdo con el área académica a la que pertenecen, se observa que el gusto por usar las redes sociales no depende de la carrera que estudien o el área a la que pertenecen; es decir, el uso de las redes sociales está dirigido a una necesidad de estar informados y comunicarse con amistades y familiares. Particularmente, se destaca que las personas del área Biológico-agropecuaria manifiestan el interés por compartir fotos con sus amigos; los del área Técnica, aun cuando les es necesario, no les gusta pertenecer a una red social.

El AC aplicado a la tabla léxica agregada permitió observar que los individuos que se agrupan en las regiones Xalapa y Veracruz comparten el gusto de formar parte de una red social porque conocen gente y están comunicados; complementariamente, se deja ver que detrás de este mismo interés, los motivos que los impulsan son diferentes: en Veracruz el interés principal es la diversión y en Xalapa además de socializar les motiva el trabajo. También se observa que se encuentran en oposición los intereses de los individuos de Coatzacoalcos-Minatitlán y Córdoba-Orizaba; los primeros hacen uso de las redes sociales por las facilidades que les brindan de comunicación entre compañeros y maestros, los segundos manifiestan usar las redes sociales por necesidad y no por gusto. En general el estudio permite, a través de la opinión fiel de los individuos, tener una clara información sobre el uso y gusto que tienen por las redes sociales.

En términos generales, el AEDT ha hecho una aportación importante en el campo de la investigación en el uso de redes sociales de la UV. Las herramientas estadísticas empleadas permitieron contextualizar las opiniones que eventualmente manifestó la comunidad universitaria sobre su gusto en el uso de las redes sociales, de acuerdo con el perfil de los individuos definido por las variables categóricas.

Lista de referencias

- Abascal, E., & Grande, I. (2005). *Análisis de encuestas*. España: ESIC.
- Álvarez, R. (2003). Las preguntas de respuesta abierta y cerrada en los cuestionarios. *Análisis estadístico de la información. Metodología de Encuestas*, 5(1), 45-54. Recuperado de <http://casus.usal.es/pkp/index.php/MdE/article/view/932/873>
- Bécue, M., Lebart, L., & Rajadell, N. (1992). El análisis estadístico de datos textuales. La lectura según los escolares de enseñanza primaria. *Anuario de Psicología*, 55, 7-22. Recuperado de <http://www.raco.cat/index.php/AnuarioPsicologia/article/view/61168>
- Hearst, M. A. (1999). Untangling Text Data Mining. En University of Maryland (Ed.), *Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics*. California: Editor. Recuperado de <http://people.ischool.berkeley.edu/~hearst/papers/acl99/acl99-tdm.html>
- Jarvio, A. O. (2011). *La lectura digital en el ámbito de la Universidad Veracruzana* (Tesis doctoral inédita). Universidad de Salamanca, España.
- Lebart, L., & Salem, A. (1994). *Statistique Textuelle*. Paris: Dunod.
- Lebart, L., Salem, A., & Bécue M. (2000). *Análisis Estadístico de Textos*. España: Milenio.
- Sutter, J. (2010). *Comscore presenta el "Estado de Internet en Latinoamérica"*. Recuperado de <http://www.iabcolombia.com/noticias/comscore-presenta-el-estado-de-internet-en-latinoamerica/>
- Varela, V. M., Montaña, O. G., & Ramírez, J. A. (2013). Influencia de las redes sociales en estudiantes de la Unidad Académica de Contaduría y Administración extensión Sur. En Universidad Autónoma de Sinaloa (Ed.), *XVI Congreso Internacional sobre Innovaciones en Docencia e Investigación en Ciencias Económico Administrativas*. México: Editor. Recuperado de <http://fca.uach.mx/apcam/2014/04/04/Ponencia%2082%20UAN-Ahuacatlan.pdf>

Glosario de términos

Distancia Chi ²	$d^2(i, i') = \sum_{j=1}^p \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{.i}} - \frac{f_{i'j}}{f_{.i'}} \right)^2$ <p>Distancia entre dos puntos fila o dos puntos columna. Es una distancia parecida a la euclídea (suma de los cuadrados de las diferencias entre los componentes de los perfiles), excepto una ponderación ($1/f_{.j}$).</p>
Forma gráfica o palabra	Secuencia de letras delimitada a la izquierda y a la derecha por un blanco o signo de puntuación.
Inercia	Equivalente a la noción estadística de varianza explicada. Da idea de la dispersión en el espacio de una nube de puntos. El porcentaje de inercia se obtiene dividiendo el valor propio del factor entre la suma de los valores propios de todos los factores. Sumando los porcentajes de los factores se obtiene el porcentaje de inercia acumulado.
Lema	Entrada del diccionario a la cual corresponde cada forma gráfica (raíz de la palabra).
Muestreo post-estratificado	Estratificación o clasificación de las unidades después de haber llevado a cabo el muestreo.
Pregunta abierta	Pregunta en la que el encuestado tiene la oportunidad de expresar libremente su opinión.
Pregunta cerrada	Pregunta cuyas posibles respuestas se presentan en el cuestionario para que una de ellas sea elegida por el encuestado.
Respuesta modal	Respuestas auténticas seleccionadas en razón a su carácter representativo para una categoría dada de individuos. Contienen un gran número de palabras características de la parte del corpus a la cual pertenecen.
Tabla de contingencia	Tabla que almacena las frecuencias absolutas de ocurrencia de dos o más variables cualitativas en un conjunto de elementos. La primera variable se representa mediante las filas de la tabla y la segunda variable por medio de las columnas de la misma.
Unidad léxica	Forma gráfica, lema o unidad compleja (compuesta de dos o más unidades).
Variable categórica o nominal	Una variable es categórica o nominal cuando los valores que toma son los nombres de categorías. Por ejemplo, el estado civil, con sus categorías: casado, soltero, viudo, divorciado, unión libre.
Variable textual	Variable originada a partir de las respuestas a una pregunta abierta.