

## Sistemas Inteligentes

### Trabajo 1 Etiquetado – Características Demográficas y Sicográficas en Usuarios de Plataformas cQA

**Profesor:** Alejandro Figueroa

[alejandro.figueroa@unab.cl](mailto:alejandro.figueroa@unab.cl)

**Ayudante:** Jean Contreras

[j.contrerasleyton@uandresbello.edu](mailto:j.contrerasleyton@uandresbello.edu)

**Fecha de Publicación:** lunes 6 de marzo de 2017

**Fecha de Entrega:** viernes 31 de marzo de 2017

**Lugar:** Horario de clases/ayudantía/con Haydeé Vidal

**Horario:** miércoles 14:10-15:40  
Jueves 12:10-13:40

#### Aspectos Generales

- El trabajo es individual.
- La entrega del informe impreso debe ser realizada de manera presencial, de acuerdo a lo señalado en el cuadro superior.
- Lea atentamente las indicaciones esbozadas en el syllabus del curso.

#### Comunidades de Pregunta-Respuesta

Las comunidades de Pregunta-Respuesta (cQA) son sitios en donde, a grandes rasgos, sus miembros responden a preguntas realizadas por otros usuarios. Algunos ejemplos de estas pueden ser StackExchange, Quora o Yahoo! Answers. Los elementos centrales de una cQA, podrían ser identificados como; preguntas, respuestas, categorías, calificaciones hechas a la calidad de las respuestas, etc. Una de las razones que hace importante el uso de estas comunidades es que los usuarios pueden obtener respuestas específicas a sus preguntas, adecuadas a su necesidad puntual. Por ende, este tipo de comunidades son atractivas para una gama amplia de personas. En estas comunidades existen distintas formas de visualizar, agrupar y encontrar cosas en común entre sus miembros. Dos de ellas están siendo muy populares hoy en día en el análisis de datos en diversas aplicaciones industriales: mediante sus condiciones demográficas y sicográficas.

La [demografía](https://es.wikipedia.org/wiki/Demograf%C3%ADa)<sup>1</sup> (del griego δῆμος dēmos 'pueblo' y γραφία grafía 'trazo, descripción' -estudio de la población-) es una ciencia que estudia las poblaciones humanas, su dimensión, estructura, evolución y características generales. Dimisiones típicas que se estudian son: Edad, sexo biológico, orientación sexual, tamaño de la familia, ciclo de vida familiar, ingresos

<sup>1</sup> <https://es.wikipedia.org/wiki/Demograf%C3%ADa>

## 1er Semestre 2017 - Sistemas Inteligentes

familiares, profesión, nivel educativo, estatus socio-económico, religión, nacionalidad, culturas, raza, generación. En cambio la [sicografía](#)<sup>2</sup> se focaliza en estilos de vida, personalidades y aspectos culturales como intereses, actitudes y opiniones.

Para el ejercicio de esta tarea, el estudiante deberá etiquetar manualmente un conjunto de miembros cQA. Para esto, se le presentará la actividad pública del usuario en la plataforma: sus preguntas, respuestas, su foto de perfil, su nickname y su auto-descripción si está disponible. Las variables demográficas y sicográficas a identificar están dadas en la siguiente tabla:

Variable	Valores posibles
<i>Edad</i>	Under 12 years old 12-17 years old 18-24 years old 25-34 years old 35-44 years old 45-54 years old 55-64 years old 65-74 years old 75 years or older no identificable (caso que no se pueda inferir del perfil del usuario)
<i>Etnia</i>	White/Blanco Hispanic or Latino Black or African American Native American or American Indian/Aborigen americano Asian / Pacific Islander/Asiático/Aborigen de Oceanía Other (cualquier otro caso no mencionado) no identificable (caso que no se pueda inferir del perfil del usuario)
<i>Estado Civil</i>	Single, never married/Soltero que nunca se ha casado Married or domestic partnership/casado o vive con alguien Widowed/viudo Divorced/divorcedo Separated/separado (pero aún casado) no identificable (caso que no se pueda inferir del perfil del usuario)
<i>Religión</i>	Protestant/Other Christian/Protestante/Evangélico Catholic Mormon Jewish/Judio Muslim/Musulmán Other non-Christian (Buddhist, Hindu, Native American, Sikh, Wiccan, Pagan, Spiritualist, Sikh, Unitarian/Universalist) No religion identity (Atheist, Agnostic, Humanist, No religion) no identificable (caso que no se pueda inferir del perfil del usuario)
<i>Clase Social</i>	TOP-UPPERS representa la elite "social". Su fortuna se pasa a través de las generaciones, familiarmente, y son conocidos. En general, compran joyas, casas, vacaciones, símbolos de estatus, educación "superior" para

<sup>2</sup><https://en.wikipedia.org/wiki/Psychographic>

## 1er Semestre 2017 - Sistemas Inteligentes

	<p>sus hijos.</p> <p>BOTTOM-UPPERS: comienzan su vida en la clase media, y ganan su fortuna en su profesión o negocio. Los productos que normalmente compran son casas caras, automóviles, educación para sus hijos, joyas, piscinas y otras cosas que dan estatus.</p> <p>TOP-MIDDLES: no son extremadamente ricos, tampoco tienen estatus familiar. Viven extremadamente cómodos, disfrutando de casas finas, ropa fina, muebles finos, vinos, etc. Aquí están los que enfocan sus carreras como gerentes de empresas, dueño de empresas exitosas, y otros puestos profesionales.</p> <p>BOTTLE-MIDDLES: Usualmente son "white collar" (trabajadores de oficina, dueños de Pymes) "gray collar" (repartidores de correos), y algunos "blue collar" (eléctricos, gasfiter, supervisores). Se conforman a los estándares y normas de la sociedad, y rara vez se desvían de lo que se espera. Sus casas convencionales son de gran importancia para ellos y tienen como objetivo mantenerlas limpias y ordenadas. Compran bienes estándar, muebles y ropas tradicionales y ordenadas.</p> <p>TOP-LOWERS: representa trabajadores "blue collar" incluyendo educados y semi-educados trabajadores de fábrica. Generalmente buscan respeto, protegen sus posesiones y buscan seguridad. El hombre suele beber cerveza, le gusta mirar deportes y disfruta de salir a pasear (outdoors). La mujer tiene puestos part y full time para proveer ingresos extras para la familia.</p> <p>BOTTLE-LOWERS: Consiste en los educados pobremente, trabajadores sin educación. Están usualmente sin trabajo y requieren de ayuda pública. Sus compras son generalmente espontaneas y a crédito. Su casa es en su mayor parte bajo el estándar.</p> <p>no identificable (caso que no se pueda inferir del perfil del usuario)</p>
Estilo de vida	<p><u>Activism</u><sup>3</sup> desarrollan una acción intencional para producir cambios políticos o sociales. Normalmente, esta acción se hace en favor de o en contra de un argumento polémico.</p> <p><u>Asceticism</u><sup>4</sup> es una vida caracterizada por la abstinencia de placeres mundanos (austeridad), normalmente para alcanzar una espiritualidad mayor. Muchos creen que el acto de purificar el cuerpo ayuda a purificar el alma, ergo una mejor conexión con lo divino o una mayor paz interior. Eso puede ser mediante rituales, mortificación propia, renuncia del placer. Mantienen que estas restricciones les provoca una libertad mayor en sus vidas.</p> <p><u>Modern Primitivism</u><sup>5</sup> es un movimiento subcultural en los países</p>

<sup>3</sup><http://psychology.wikia.com/wiki/Activism>

<sup>4</sup><http://psychology.wikia.com/wiki/Asceticism>

## 1er Semestre 2017 - Sistemas Inteligentes

	<p>desarrollados y se relaciona estrictamente con aquellas personas que deciden realizar modificaciones corporales como parte de su vida, también porque participan en los ritos de las culturas tribales tradicionales, pero adaptados con modernos conocimientos y procedimientos</p> <p><u>Back to the land</u><sup>6</sup> cubre variados movimientos agrarios a través de los tiempos. El común denominador es un llamado a la gente a tener granjas pequeñas, a sembrar tu propia comida en escala pequeña (para uno mismo o para otros).</p> <p><u>Bohemianism</u><sup>7</sup> es la práctica de un estilo de vida no convencional, normalmente en compañía de gente que tiene un parecer similar en términos de música, arte, literatura. Tienen pocas relaciones permanentes. Pueden ser vagabundos, aventureros, etc.</p> <p><u>Clothes</u><sup>8</sup> frees un movimiento cultural y político que promueve y defiende el nudismo personal, familiar, y social.</p> <p><u>Communal</u><sup>9</sup> es la vida intencional en comunidad donde la mayoría de los recursos son compartidos. Casi no hay propiedad privada/personal. Va más allá de solo compartir hogar.</p> <p><u>Hippies</u><sup>10</sup> buscan liberarse de las restricciones de la sociedad, buscar su propio camino y un Nuevo significado a la vida. Una forma de hacerlo es mediante un estándar diferente de vestimenta que les permite reconocerse entre ellos mismos.</p> <p><u>Nomadism</u><sup>11</sup> se mueven de un lugar a otro, no tienen un lugar fijo de vida.</p> <p>A <u>quirkyalone</u><sup>12</sup> es una persona que prefiere estar solo que resuelto. Escucha su voz interior para mantenerse veraz consigo mismo.</p> <p><u>Simple living</u><sup>13</sup> incluye práctica que buscan simplificar la vida. Buscan eliminar posesiones o aumentar la suficiencia propia. Están satisfechos</p>
--	---

5 [https://es.wikipedia.org/wiki/Primitivo\\_moderno](https://es.wikipedia.org/wiki/Primitivo_moderno)

6 [https://en.wikipedia.org/wiki/Back-to-the-land\\_movement](https://en.wikipedia.org/wiki/Back-to-the-land_movement)

7 <http://psychology.wikia.com/wiki/Bohemianism>

8 <https://en.wikipedia.org/wiki/Naturism>

9 [http://psychology.wikia.com/wiki/Commune\\_%28intentional\\_community%29](http://psychology.wikia.com/wiki/Commune_%28intentional_community%29)

10 [https://en.wikipedia.org/wiki/Hippie#Ethos\\_and\\_characteristics](https://en.wikipedia.org/wiki/Hippie#Ethos_and_characteristics)

11 <http://psychology.wikia.com/wiki/Nomad>

12 <http://quirkyalone.net/>

## 1er Semestre 2017 - Sistemas Inteligentes

	<p>con lo que tienen, no con lo que quieren. Es parecido al ascetecismo en que promueven en vivir simple y restringirse los lujos, no todos los simple living son practican el ascetecismo. Es diferente a los que viven en la pobreza forzadamente, ya que es un estilo de vida voluntario.</p> <p>A <a href="#">groupie</a><sup>14</sup> es una persona cuya devoción a otra persona (usualmente a una celebridad, en especial a una estrella de rock) se aproxima al nivel de fanatismo.</p> <p>Rural lifestyle (vida de campo/rural) Traditional lifestyle (vida típica occidental, clase media, working family) Other no identificable (caso que no se pueda inferir del perfil del usuario)</p>
Partido Político	<p>Republicano Democrata Libertarian/Libertario Independiente Green/Verde Other no identificable (caso que no se pueda inferir del perfil del usuario)</p>
Auto descripción	<p>Affectionate, passionate, loving, romantic Amicable, amiable, affable, benevolent Awkward, absent-minded, forgetful, careless Brave, courageous, daring, adventuresome Broad-minded, open-minded, liberal, tolerant Creative, inventive, imaginative, artistic Dominating, authoritarian, demanding, aggressive Efficient, organized, diligent, thorough Egocentric, vain, self-centered, narcissistic Frank, straightforward, outspoken, candid Funny, humorous, amusing, witty Intelligent, smart, bright, well-informed Kind, good-hearted, warm-hearted, sincere Refined, gracious, sophisticated, dignified Reserved, conservative, quiet, conventional Self-assured, confident, self-sufficient, secure Sociable, friendly, cheerful, likeable Stubborn, hard-headed, headstrong, obstinate Tense, nervous, high-strung, excitable Trustworthy, competent, reliable, responsible</p>

Nótese que el valor de la variable debe estar explícitamente o implícitamente en el texto. Nótese que en el caso de la auto-descripción, ud. debe considerar lo que el miembro dice de si mismo, no la impresión que ud. se genere de su actividad. Bien puede ser con las palabras indicadas o de otra forma.

<sup>13</sup>[https://en.wikipedia.org/wiki/Simple\\_living](https://en.wikipedia.org/wiki/Simple_living)

<sup>14</sup><http://www.urbandictionary.com/define.php?term=groupie>

## Objetivos

La tarea apunta a confrontar al alumno con la etiquetación de un conjunto de datos, y los problemas que conlleva esta tarea. Además, a revisar conceptos aprendidos en asignaturas anteriores como el de probabilidad. Sumado a esto, se pondrán en práctica conceptos que serán discutidos en clases como el de entropía.

## Desarrollo

A cada estudiante se le asignará un número de 100 usuarios donde deberá identificar el valor de las variables demográficas y sicográficas listadas anteriormente. En caso de no poder relacionarlo con ninguna de las categorías, el alumno puede asignarlo a una clase "OTROS", "NO IDENTITY" o bien "NO IDENTIFICABLE" para el caso que no pueda obtenerse del perfil. El alumno deberá entregar un informe donde:

- Describe extensamente su experiencia de etiquetado: ¿Qué observó? ¿Qué dificultades encontró? ¿Qué observa de las preguntas que fueron en la clase "OTROS"/"NO IDENTITY"? Discuta esto en extenso y ejemplificando sus observaciones.
- ¿Qué características tienen los usuarios relacionados con cada valor de variable? ¿Cuál es la distribución en su conjunto de datos? ¿Cuál es la probabilidad de cada uno de los valores de variables?
- ¿Hay perfiles comunes? ¿Logra identificar algún tipo de usuario dentro de la comunidad Yahoo! Answers? Elabore sobre este último punto.
- ¿Hay relación entre ocurrencias de valores para las primeras cuatro variables (demográficas) y el segundo grupo (sicográfico)?
- Calcule la entropía<sup>15</sup> de sus conjuntos de datos para cada variable. ¿Cuál es la que tiene mayor certidumbre en la comunidad?
- Calcule la Información mutua para los valores de la variable edad y estilos de vida<sup>16</sup>.

Nótese que debe entregar un informe conforme al esquema descrito en el Syllabus del curso, es decir con introducción, descripción del problema, conclusiones, etc.

## Entregable

El conjunto de datos finalizado el proceso de etiquetación. Este debe estar en el servidor.

<https://ba21e628.ngrok.io/SI/index.php>

<sup>15</sup>[http://gerardnico.com/wiki/data\\_mining/entropy](http://gerardnico.com/wiki/data_mining/entropy)

<sup>16</sup><http://nlp.stanford.edu/IR-book/html/htmledition/mutual-information-1.html>

## ***1er Semestre 2017 - Sistemas Inteligentes***

El no término del etiquetado en los plazos estipulados, implica la nota 1.0. Un informe que desarrolle los puntos anteriormente nombrados. El alumno necesitará este conjunto de datos para tareas posteriores.

### **Bonus**

El alumno podrá adquirir un punto de bonus para la tareas, mediante el etiquetado de una porción adicional de 100 miembros. Su análisis deberá abarcar los 200 miembros. El caso de sobrepasar el 7.0, el resto podrá ser cobrado en alguna tarea futura.