# Informe tarea 1 sistemas inteligentes

Profesor: Alejandro Figueroa  
Profesor Ayudante: Jean Contreras Leyton  
Alumno: Andrés Valenzuela González

# **Introducción**

Para comprender el siguiente informe, es necesario primero conocer las comunidades de Pregunta-Respuesta;

* Las comunidades Pregunta-Respuesta (*cQA*) son sitios web donde sus miembros pueden realizar preguntas con el propósito de que sean respondidas por el resto de la comunidad, así como también poder responder a preguntas hechas por otros miembros. Un ejemplo de estas comunidades es *Yahoo! Answers*.

Dentro de estas *cQA* podemos identificar los siguientes elementos:

1. Preguntas: Usualmente presentan un título y un cuerpo, estas pueden pertenecer a cualquier categoría disponible en la comunidad.
2. Respuestas: Respuestas de usuarios a las preguntas realizadas dentro de la comunidad. Usualmente siguen ciertas normas al momento de responder que dependen de cada comunidad.
3. Categorías: Las preguntas realizadas caen en categorías existentes en la comunidad, lo que permite una óptima búsqueda de preguntas dentro de ella.
4. Calificaciones: Dentro de la comunidad es posible evaluar las respuestas de una pregunta, lo cual permite “*rankear”* las respuestas útiles de las respuestas erróneas.

Gracias a estas comunidades, tanto un usuario externo como interno puede encontrar respuestas existentes a preguntas ya hechas por otros usuarios, lo cual atrae a personas provocando que la comunidad crezca.

# **Descripción del Problema**

Para la Tarea 1 el alumnado debe etiquetar manualmente un conjunto de perfiles de la *cQA* *Yahoo! Answers*. Para esto, se le brindo acceso al **grandioso** servidor de la *Universidad Nacional Andrés Bello,* con el cual el alumno deberá identificar y seleccionar desde una lista de etiquetas *(USA only – Non-USA – World – Undetermined)* la localidad a la que el usuario describe ser dentro de su perfil.

Luego de la realización del etiquetado de 2000 perfiles, el alumno deberá obtener la probabilidad de cada etiqueta, la entropía, la información mutua y responder a las preguntas *¿Qué observó?*, *¿Qué dificultades encontró?, ¿Qué observa de las preguntas que fueron en la clase “undetermined”?, ¿Qué características tienen los usuarios relacionados con cada ubicación geográfica (las más prominentes)? ¿Cuál es la distribución en su conjunto de datos? ¿Cuál es la probabilidad de cada uno de los valores de variables?*, entre otras.

# Figura 1

Figura . Ecuaciones de Entropía e Información Mutua

# Siendo:

# , Entropía de una variable.

# , Información Mutua entre dos variables.

# , Entropía condicional y porcentaje de una tabla con dos variables.

# *Recordar que para la realización de este informe el conjunto de datos consistía de 2000 perfiles.*

# **Distribución del Conjunto de Datos**

|  |  |
| --- | --- |
| Etiqueta | Total |
| *USA only* | **827** |
| *Non-USA* | **545** |
| *World* | **78** |
| *Undetermined* | **550** |

Entropia del dataset: 1.732657