# Informe tarea 1 sistemas inteligentes

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

# **Introducción**

El propósito de este informe es superar la confrontación ante el etiquetado de un conjunto de 2000 datos provenientes de la comunidad Pregunta-Respuesta (cQA) *Yahoo!* y los problemas que conlleva esta tarea. Además, se revisaran conceptos aprendidos anteriormente como el de *Probabilidad* y de la materia dictada en el ramo de *Sistemas Inteligentes.* Sumado a esto, se describirá la experiencia de etiquetado y se responderán las preguntas “*¿Qué dificultades encontró?, ¿Qué observa de las preguntas que fueron en la clase “undetermined”?.”* entre otras.

# **Descripción del Problema**

Durante el trabajo del etiquetado, se vislumbraron diversos problemas, los cuales correspondieron a más de una consulta por parte del alumnado a su profesor ayudante, pero los más destacables fueron:

* Mala ortografía de algunos perfiles
* Pésima redacción de numerables perfiles
* Imposibilidad de poseer conocimiento absoluto de todos los países del mundo
* Imposibilidad de poseer conocimiento absoluto de todas las regiones de los países antes contemplados
* Barrera de idiomas ante perfiles provenientes de occidente (Coreano, Japonés, Chino, Árabe) y algunos otros de habla portuguesa
* Varios perfiles cumplían rol de descripción de personaje, pero sin ubicación del sujeto
* Ciertos perfiles se dedicaban a describir a una empresa
* Otros perfiles solo contenían pasajes de la biblia
* Gran cantidad de perfiles de la clase *USA only* hacían uso de abreviaciones para el nombre del estado en donde viven/vivieron (CL, California – cali, California – TX, Texas – NC, North Carolina – etc)

Además, las preguntas que fueron de la clase *“undetermined”* poseen

Idea: aleatoriedad, mala redacción, son empresas, son pasajes de la biblia, no especifican de donde vienen, son personas mayores de 30-40 años, son empresas, son personas con poco o nulo interés en los demas