

---

## PROYECTO NO. 3

---

201904490 – Diego Aldair Sajche Avila

### Resumen

Para este proyecto la empresa Tecnologías Chapinas S.A fue contratada por la red social ChapinChat para desarrollar una herramienta capaz de analizar el contenido de dicha red social, así como establecer perfiles de sus usuarios para poder conocerlos mejor y tengan una experiencia más personalizada.

La empresa Tecnologías Chapinas S.A creara un sistema que consiste en construir en base a la información de los mensajes recibidos, analizar el contenido del mensaje y establecer por cada usuario el porcentaje de contenido que esta asociada a cada perfil que haya sido configurado en el sistema. Los perfiles serán conjuntos de palabras clave, que permitirán calcular el porcentaje de palabras en relación al total de palabras de dicho perfil, para una mayor exactitud también se definirá un listado de palabras que no será considerado como palabras clave a analizar.

### Palabras clave

Tecnologías Chapinas, ChapinChat, perfiles, usuarios, porcentajes.

### Abstract

*For this project the company Tecnologías Chapinas S.A. was contracted by the social network ChapinChat to develop a tool capable of analyzing the content of this social network, as well as to establish profiles of its users in order to know them better and have a more personalized experience.*

*The company Tecnologías Chapinas S.A. will create a system that consists of building based on the information of the messages received, analyze the content of the message and establish for each user the percentage of content that is associated with each profile that has been configured in the system. The profiles will be sets of keywords, which will allow to calculate the percentage of words in relation to the total number of words in that profile, for greater accuracy will also define a list of words that will not be considered as keywords to be analyzed.*

### Keywords

*Tecnologías Chapinas, ChapinChat, profiles, users, percentages..*

## Introducción

El siguiente proyecto busca crear un servicio web consumido por la red social ChapinChat con la ayuda de Tecnologías Chapinas S.A, el consistirá en construir un sistema de información que reciba mensajes, analice contenido y establezca por cada usuario el porcentaje de contenido que está asociado a cada perfil que haya sido configurado en el sistema.

La empresa Tecnologías Chapinas S.A define un perfil como un conjunto de palabras que pueden presentar interés del usuario que son de utilidad para ChapinChat, los perfiles son dinámicos y pueden cambiar con el tiempo.

Finalmente, Tecnologías Chapinas S.A. creo un algoritmo capaz de establecer el peso de un perfil en un usuario de la red social, este algoritmo se basa en los porcentajes de probabilidad de los perfiles en los mensajes que el usuario genere en la red social.

## Desarrollo del tema

Para desarrollar esta herramienta, podemos mencionar los objetivos principales del proyecto, los cuales serían que sea capaz de analizar el contenido de dicha red social y establezca perfiles de los usuarios para así poder brindar una mejor experiencia a los usuarios.

La empresa ChapinChat generara los menajes basándose en la siguiente estructura:

*Lugar y Fecha: Guatemala, dd/mm/yyyy hh24:mi*

*Usuario: nombre\_usuario\_sin\_espacios*

*Red social: ChapinChat*

*Mensaje escrito por el usuario en la red socia*

La estrategia consiste en construir el sistema para recibir los mensajes, analizar el contenido y establecer el porcentaje de contenido asociado a cada perfil configurado en el sistema. Como se menciono anteriormente, existirá una lista de palabras clave las cuales servirán para comparar con los perfiles y así determinar el porcentaje de palabras iguales, también habrá una lista de palabras descartadas, las cuales servirán para excluir del mensaje y así no tendrá validez al analizar el contenido.

Para validar la cantidad de palabras dentro del mensaje que se encuentran asociadas a los perfiles se realiza de la siguiente manera: se divide la cantidad de palabras iguales con cada perfil y ese valor se divide entre las palabras contabilizadas en el mensaje (este valor es sin incluir las palabras excluidas) y ese resultado se multiplica por el cien por ciento, esto para obtener el porcentaje relacionado a cada perfil.

Es importante mencionar que las palabras mayúsculas y minúsculas no son relevantes al momento de determinar si pertenecen a una palabra clave, cuando una palabra clave esta formada por varias palabras, suma esa cantidad de palabras para obtener el porcentaje; y en el caso de que las palabras estén formadas por varias palabras no importa si están separadas por tabulaciones, espacios o cambios de línea.

Podemos mencionar unas consideraciones a tomar en cuenta al momento de recibir los mensajes:

- Lugar y Fecha: El nombre de la ciudad siempre se encuentra al inicio, se le agrega una coma y seguidamente se coloca la fecha y la hora como se mencionó anteriormente.
- Usuario: El nombre de usuario puede contener cualquier carácter excepto espacios, signos de tabulación o cambios de línea. También se pueden incluir los correos electrónicos.

- Red Social: En este campo siempre será el nombre de la red, el cual es ChapinChat
- Mensaje: El mensaje es un texto libre lo cual significa que puede contener cualquier contenido, básicamente es un tema de cualquier interés.

Finalmente se ha creado un algoritmo para establecer el peso de cada perfil del usuario, se basa en los porcentajes de probabilidad que el usuario ha generado en la red social, de tal manera que, se promedien por cada perfil los porcentajes de todos los mensajes, y este promedio corresponderá a 1 peso de cada perfil.

Este es un ejemplo de como serian las tablas generadas del calculo de pesos para cada usuario:

Tabla I: Mensajes enviados por usuario y porcentajes de probabilidad de los perfiles

Usuario: map001		
Mensaje	% perfil 1	% perfil 2
1/4/2023 09:31	10.71	25
1/4/2023 12:07	5.65	40.33
1/4/2023 15:21	0	21.5

Fuente: IPC2\_Proyecto3.pdf

Para calcular los pesos de cada perfil se realiza de la siguiente manera: se suman todos los porcentajes de cada perfil y se dividen entre la cantidad de porcentajes de cada perfil. Prácticamente es un promedio el que se utiliza para determinar el peso de cada perfil.

## Conclusiones

1. El sistema serio capaz de analizar el contenido de dicho sistema a través de los mensajes de la red social y así poder determinar los porcentajes de palabras iguales a las palabras claves.

2. Se logro una mayor exactitud de en los porcentajes debido al listado de palabras descartadas, las cuales

no fueron consideras para analizar dichos mensajes de la red social.

3. Se realizo el sistema con la ayuda de archivos xml, los cuales sirvieron de entrada para determinar los perfiles a utilizar, y también se realizaron archivos de salida; los cuales exportaron la información ya procesada en la red social con los porcentajes, los usuarios verificados y las palabras que se analizaron.

4. Dicho proyecto también se realizo en el lenguaje de Python, donde se utilizó la arquitectura HTTP/Flask, para los cuales en el Frontend se realizó con el framework Django y para el Backend se utilizó el framework Flask; siguiendo el patrón MVT (Modelo-Vista-Template).

## Referencias bibliográficas

Corbucci, S. (2011). *Django. En 100 Cult Films* (pp. 65–66). British Film Institute.

Tutorial Point. (2020). *Html tutorial* (Tutorial Point, Ed.). Independently Published.

Welcome to flask — flask documentation (2.3.X). (s/f). Palletsprojects.com. Recuperado el 3 de mayo de 2023, de <https://flask.palletsprojects.com/en/2.3.x/>

Facultad de Ingenieria. (2023). *Proyecto No. 3. Ciudad de Guatemala: Universidad San Carlos de Guatemala. Recuperado el 03 de Mayo de 2023*