

Caracteres permitidos

Es posible visualizar de color rojo los caracteres discrepantes con los nombres "oficiales".

El tamaño de las letras no tiene relevancia acá.

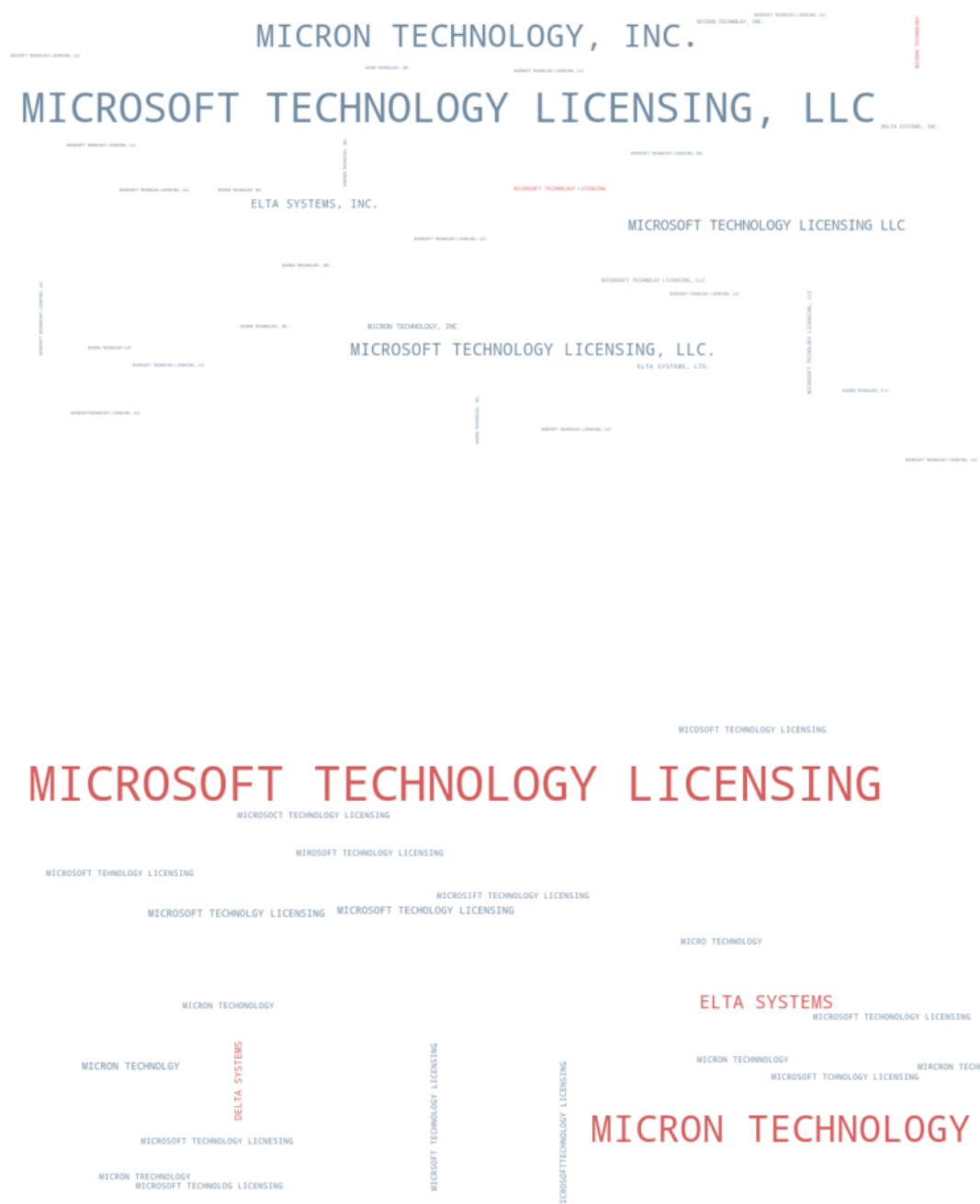
SEENMHR
HGAFCY

DSIT / RFG
LMTCYN
LWK R OYI H.

Primera limpieza

En rojo se puede ver la coincidencia con los nombres "oficiales", antes y después.

El tamaño de los nombres sí tiene relevancia acá.



Datos adicionales

Se muestra la lista de abreviaturas de empresa encontradas en el dataset.

[' INC', ' LLC', ' LTD', ' LLP', ' PC', ' KKC']

Según enunciado, se asumen estas como las posibles palabras especiales para remover.

Luego de la limpieza básica, el porcentaje de datos coincidentes con un nombre original es de 99.51%.

La cantidad de etiquetas existentes en el dataset es de 4/4.