

Para la definición del proyecto **Webscraping Institucional** es necesario la definición del tipo de entrada de datos; el algoritmo de minado de datos y el resultado.

Se presenta la siguiente definición de datos de entrada; resumiendo este como cualquier dato recibido de la aplicación en el cual se le quiere realizar la búsqueda de información. Dicha entrada se dividirá en la aplicación de esta manera:

Ubicación

- Referente a todas las ubicaciones geográficas, municipio, estado, zona, etc dentro del Estado de Hidalgo.

Temática

- Referente a una palabra clave que defina una temática, área de estudio o eje dentro del Estado de Hidalgo.

Actor/Institución

- Referente a los actores como personal, instituciones, organismos dentro del Estado de Hidalgo.

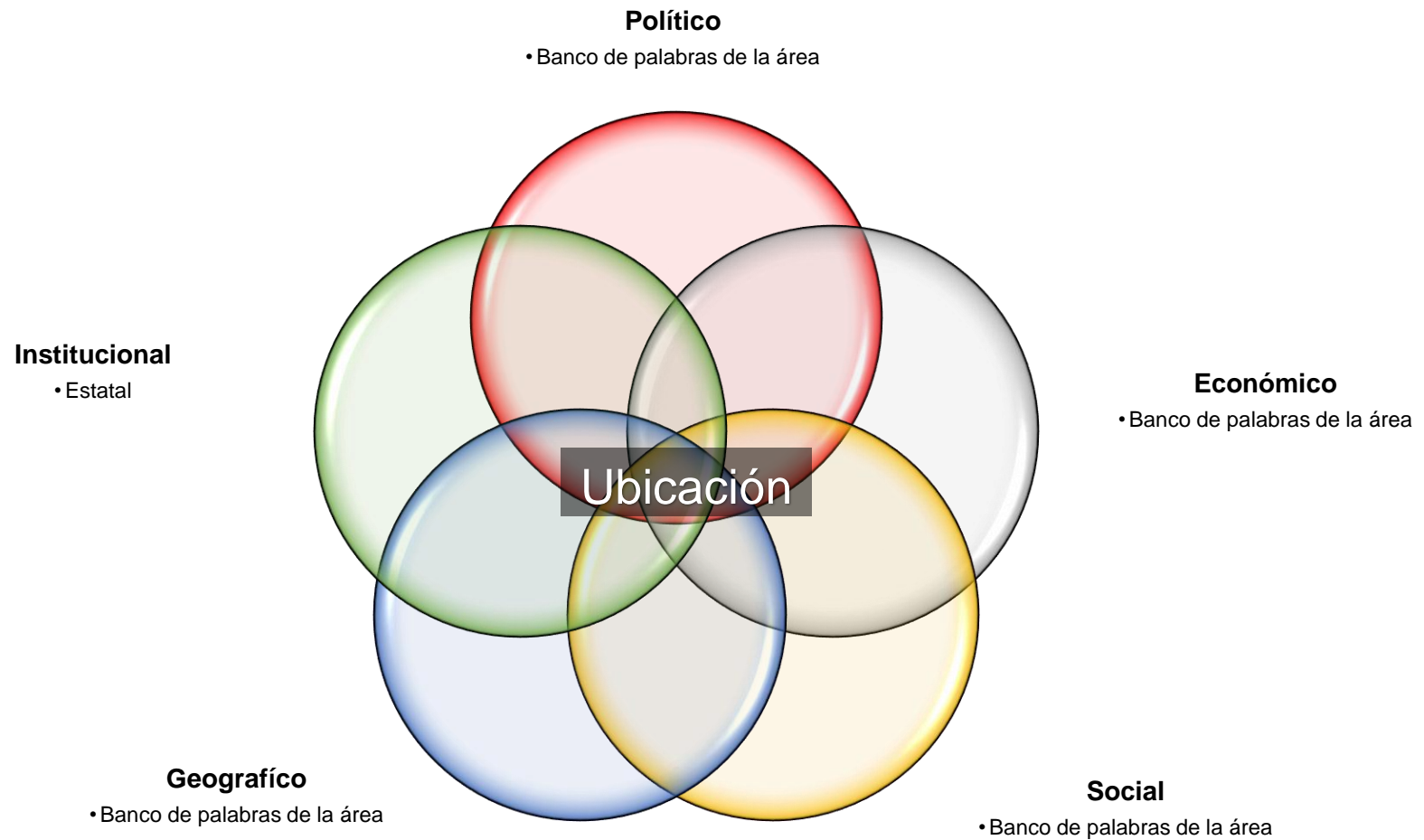
Dicha categorización permitirá definir en el área de definición del algoritmo las variables para el análisis y presentación de la información.

Definición de algoritmo:

Dicho algoritmo para la clasificación de información obtenida se basará en la implementación de una minería de datos basada en la teoría de Veen para su definición.

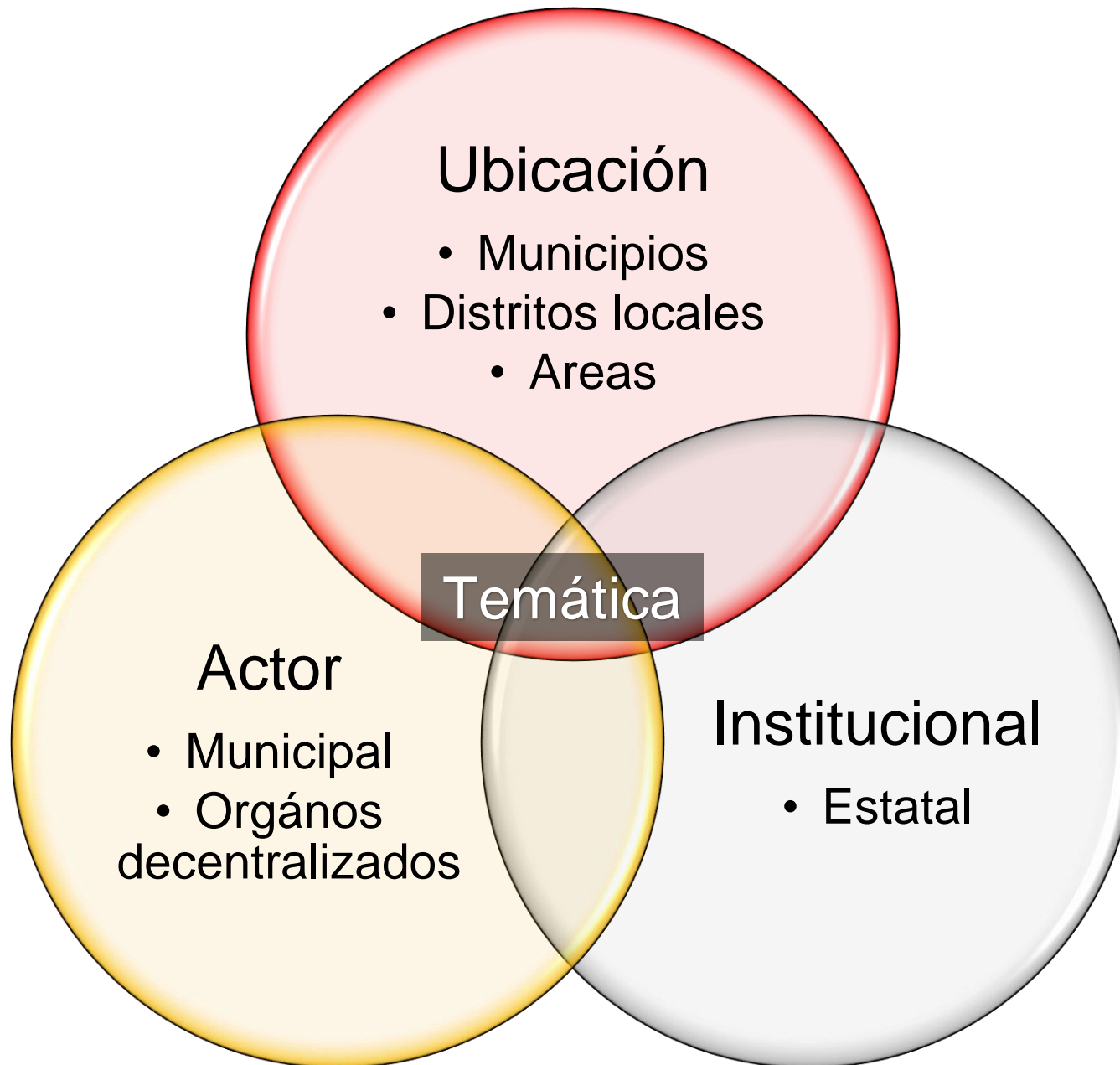
Algoritmo de Ubicación

En el proceso de análisis de información bajo una entrada de ubicación se clasificará los valores de criterio de la siguiente manera:



Algoritmo de Temática

En el proceso de análisis de información bajo una entrada de ubicación se clasificará los valores de criterio de la siguiente manera:



Algoritmo de Actor

En el proceso de análisis de información bajo una entrada de ubicación se clasificará los valores de criterio de la siguiente manera



