



## 4. Métodos

En esta sección se describen tanto las unidades de análisis como los diferentes modelos cuanti- cualitativos que serán utilizados para la evaluación de las nuevas ofertas académicas de la Universidad Santo Tomás, en relación con las diferentes necesidades a las que se busca dar respuesta. Así, en este capítulo se han estructurado cuatro secciones, la primera presenta la selección de las unidades de análisis o unidades de muestreo que serán utilizadas en el modelo. En la segunda se realiza la edificación del modelo de evaluación de la pertinencia de las propuestas académicas. En la tercera y última secciones, se presenta el procesamiento de los productos de CTel e I+D+I a partir de algoritmos de procesamiento de lenguaje natural, computación y ciencia de redes.

### 4.0.1 Unidades de análisis

Para la definición de las unidades de muestreo o análisis es necesario considerar la estructura organizacional que tiene la Universidad Santo Tomás. Es decir, como el Movtea intenta promover insumos para la gestión estratégica de los aspectos académicos, es preciso conservar la estructura jerárquica definida por la universidad. Así, los resultados obtenidos por el Movtea podrán ser escalados a los diferentes niveles de la estructura organizacional. En conformidad con lo anterior, los niveles de las unidades de análisis utilizados por el Movtea (desde el nivel inferior al superior) son:

4. Programas
5. Facultades

## 6. Divisiones

Otro factor que justifica el orden de las unidades de análisis se debe a la diferenciación de las denominaciones de las propuestas académicas de la Universidad Santo Tomás, con programas académicos internacionales. Por ejemplo, a nivel nacional existen programas de mercadeo, pero a nivel internacional se encuentran en algunas escuelas de administración de negocios como línea de especialización o investigación en *marketing*.

### 4.0.2 Análisis de clúster

El análisis de clúster o *clustering* es un proceso de minería de datos cuyo objetivo principal es encontrar grupos (o clústeres) con elementos internos con porcentajes importantes de similitud (Distefano y Mindrila, 2013). Por lo tanto, sus algoritmos tienen como objetivo maximizar la similitud dentro de cada clúster y minimizar la similitud entre estos.

Para esta investigación, los clústeres serán utilizados para evaluar el nivel de productividad científica y académica, así como su posición en los modelos de clasificación mencionados en la sección anterior. Dentro de los principales resultados que se espera obtener con esta técnica, se destacan dos aspectos importantes: el primero está referido a la conformación de grupos de universidades con el mayor volumen de productividad científica y académica por área del conocimiento. Incluso, se puede establecer análisis intra e intergrupales, con el objeto de describir las brechas que genera la heterogeneidad en el nivel de productividad científica de las instituciones. En segundo lugar, una vez creados estos clústeres, se realizará un estudio detallado de la productividad académica de dichas universidades a través de bases de datos textuales obtenidas de bases de datos indexadas y de un proceso de *Web Scraping*.

### 4.0.3 Procesamiento de la productividad

Una vez obtenida la clasificación de las instituciones en cada clúster, el siguiente paso es analizar la productividad de las universidades conforme a su posicionamiento e impacto en el área del conocimiento.

La información sobre la productividad se recolectará de los docentes e investigadores con mayor impacto y visibilidad en esos programas, facultades y divisiones que tienen grandes porcentajes de similitud con los objetivos, misión y visión de cada una de las propuestas académicas de la Universidad Santo Tomás. Para la selección de los docentes e investigadores más significativos se tendrán en cuenta los siguientes criterios:

1. Índice H (Huelva, 2020) superior a 30.
2. Índice de Scopus (Scopus, 2020a) superior a 10.
3. Más de 1000 citaciones en los últimos tres años en Google Académico (Scholar, 2020).

#### 4.0.4 Proceso de *Web Scrapping*

El proceso de *Web Scrapping* o *websrapping* (Marti, 2020) es el tercer procedimiento de la sistematización metodológica de esta investigación, y consiste en extraer información de diferentes sitios web a través de *script* o algún *software* de manera automática.

Una de las razones más importantes por las cuales fue seleccionada esta herramienta para esta investigación es que facilita la recolección de información de los principales sitios web (o repositorios web) donde se registra la productividad científica y académica de los docentes e investigadores de la Universidad Santo Tomás, como también de diferentes instituciones educativas locales, regionales, nacionales e internacionales (cuyos programas tengan similitudes significativas con las propuestas). Esto permite construir una base de datos textual robusta, con información importante para estudiar a profundidad la configuración de los patrones de las tendencias de formación, CTel e I+D+I desde los productos académicos y científicos divulgados.

#### 4.0.5 Procesamiento de Lenguaje Natural

Es una herramienta computacional que permite realizar análisis textual, considerando aspectos como la morfología del lenguaje, patrones

en los textos de ciertos autores, entre otros. Para esta investigación, se utilizará la detección de patrones en los textos que presentan los resultados de las diferentes investigaciones realizadas por los autores más importantes (de acuerdo con los criterios presentados en el epígrafe de procesamiento de la productividad) de los programas, escuelas, departamentos, institutos y facultades. En el Movtea se diseñarán algoritmos con las siguientes funciones:

1. Distribución de palabras clave de artículos y obras científicas de los investigadores.
2. Asociaciones de palabras centrales de los productos científicos a grupos de investigación e investigadores.
3. Superposición de palabras y conceptos claves de investigaciones por territorios, grupos, centros e institutos de investigación.
4. Construcción de redes de conocimiento generadas por las palabras y los conceptos claves de los productos de investigación.
5. Todos los análisis anteriores se aplican a algunas de las otras fuentes descritas en este documento, por ejemplo, planes y/o apuestas estratégicas de los sectores productivos.

#### **4.0.6 Índice de similitud y distancias**

Esta es una medida de similitud que compara los miembros de dos conjuntos para mirar cuáles elementos son compartidos y cuáles son distintos. Tiene un rango de valores entre cero y uno, por lo cual se puede reportar como un porcentaje, además es muy fácil de interpretar. Es una medida muy utilizada cuando se pretende comparar datos textuales.

#### **4.0.7 Análisis de redes**

Primero es necesario definir el concepto de *Grafo* como una estructura matemática que permite estudiar las comparaciones dos a dos de

objetos y entidades definidas previamente (Luke, 2015). Formalmente tenemos que:

- **Grafo** es una pareja de conjuntos  $G = (V, E)$  donde  $V$  es el conjunto de vértices y  $E$  el conjunto de aristas.  $E$  se compone de pareja de elementos de  $V$  (parejas no ordenadas) (Luke, 2015).
- **Dígrafo** se define como todas las parejas de conjuntos ordenados  $D = (V, A)$ . Con  $V$  el conjunto de vértices y  $A$  el conjunto de conexiones.  $A$  está conformado por las parejas de elementos de  $V$  (parejas ordenadas) (Luke, 2015).

El origen de la teoría de grafos se genera con el problema de los puentes de Königsberg, en 1730. En 1840, A. Möbius proporciona una idea de lo que es un grafo completo y un grafo biparticionado. El concepto de *árbol* fue definido por Gustav Kirchhoff, en 1845, como un grafo conectado sin ciclos. Este concepto fue aplicado en las teorías asociadas a las corrientes y campos eléctricos (Luke, 2015).

En 1852, Thomas Guthrie elaboró el famoso problema de 4 colores que sería resuelto cerca de 100 años después y, posteriormente, en 1856, Thomas P. Kirkman y William R. Hamilton estudiaron los ciclos en los grafos y definieron el concepto de *grafo hamiltoniano*, para estudiar el número de sitios que era posible visitar exactamente una vez en un viaje. Con la resolución del problema de 4 colores, por parte de Kenneth Appel y Wolfgang Haken, se dio inicio a lo que se conoció como la Teoría de Grafos.

En 1941, Ramsey trabajó en diferentes coloraciones y con esto pudo identificar una nueva rama de la teoría de grafos denominada teoría extremal de grafos. Posteriormente, en 1969, el problema de 4 colores fue resuelto computacionalmente por Heinrich. El estudio de la conectividad asintótica de grafos permitió la creación de la teoría grafos aleatorios. Es importante mencionar que las historias de la teoría de grafos y la topología están muy relacionadas.



### III RESULTADOS OBTENIDOS



