

EL ANALISIS DE DOCUMENTOS

El análisis de documentos es el método de recoger información para la investigación histórica, y también es útil para saber sobre la situación del momento. Es el primer paso a dar después de haberse planteado un problema.

La forma de trabajar esta técnica tiene diferentes puntos de vista:

- Métodos clásicos: sacan información mientras lo leen, surgió de la crítica literaria e histórica. Primero se hace un análisis interno que muestra el contenido, el método externo nos muestran el contexto y nos muestran el grado de veracidad y su resonancia. El método histórico de entre los métodos clásicos es el más común, pero en él es fundamental la autenticación y la veracidad del documento.
- El método cuantitativo. Para Duverger es complementario al método clásico. Busca eliminar el subjetivismo de la persona. Y hoy día tiene la ventaja del uso de la informática.

De entre sus métodos estará :

- o La semántica cuantitativa: busca analizar el vocabulario , el etilo y las formas de expresión. Se cuentan las palabras se clasifican según su función sintáctica y por cálculos matemáticos miden la asociación de palabras para analizar el estilo del autor.
- o El análisis de contenido Es objetivo y lo veremos desarrollado en el punto numero 2
- o Documentación bibliográfica nos muestra que el autor puede sacar información de sus propios estudios o de otros autores, que debe de citar, por medio de las notas y la bibliografía .

2- ANALISIS DE CONTENIDO

Para Berelson es una técnica de investigación que describe objetiva , sistemática y cuantitativa del contenido . Y más adelante se introdujo el contenido latente , el no evidenciado. Busca convertir en números los datos. Y se basa en la objetividad , cualquiera puede repetir el estudio. Es sistematizado la hora de recoger información, se deben tratar los datos de manera cuantificable. Debe ser manifiesto , o sea evidente en si mismo.

Krippendorff considera a esta como una técnica destinada a formular, a partir de ciertos datos inferencias reproducibles y válidas que pueden aplicarse en el contexto. O sea que a partir de cierta información que se puede extraer de un documento intentar extrapolarlo a otras situaciones similares.

Los pasos de este método son:

- Decidir se analiza todos los documentos o una parte de ellos y si solo se escoge algunos tienes que escoger los más significativos.
- Definir la unidad de análisis , como vamos a evaluarlo si por palabras , frases , oraciones , párrafos, temas...
- Categorización , identificar seleccionar y estructurar las categorías del análisis, dentro de lo que se va a analizar palabras frases , este paso busca mostrar todas las opciones posibles.
- Fiabilidad, correlación entre categorizaciones de varios codificadores
- Análisis estadístico, frecuencias, tablas de contingencias, medidas de asociación, análisis discriminante, análisis de cluster, análisis de correspondencias.

Este método puede buscar describir tendencias, analizar técnicas persuasivas, analizar el estilo, relacionar características , relacionar atributos.

Las ventajas de este método será que no es muy intrusiva y es muy barata pero como inconveniente, no coinciden muchas veces el propósito de inicio con el final.

Los documentos serán limitados en algunos temas concretos

TIPOS DE DOCUMENTOS

Los documentos son muy variados pueden ser :

- actas circulares
- Informes, programas
- Cartas, diarios , discursos
- Revistas periódicas

PROCEDIMIENTO A SEGUIR EN EL ANÁLISIS DE CONTENIDO

Los pasos que se realizan será :

1. Objetivos y contextos:
2. Unidades de análisis: se debe explicar que hay que contar. Así trabajaremos en unidades de muestreo cosas a observar, unidades de contexto da significado a unidades de registro, unidades de registro, categoría del texto que pueden ser :
 - Léxica por términos o por palabras
 - Propositionales: de oraciones o proposiciones
 - La narrativa : acciones de personajes en un tiempo y en un espacio
 - Temática: conceptos referencias y valoraciones
 - Argumentativas fruto de enunciaciones
 - Discursivas fruto de discursos de sujetos.

REGLAS DE NUMERACIÓN O RECUENTO

Hay muchos tipos de forma de formas de numerar, que serían:

- Presencia o ausencia de un código
- Frecuencia de aparición de un código
- Frecuencia ponderada, según la importancia del código,
- Orden de aparición de códigos
- Densidad de textos. Es la frecuencia absoluta del código dividido entre el total de las palabras. Que se puede dar en proporción o en tanto por ciento.
- Nivel de concentración número de códigos distintos divididos por el sumatorio de las frecuencias de los códigos.

CATEGORIZACIÓN

1. Es cuando clasifico los elementos del texto según unos criterios establecidos antes. Se suele llamar cuello de botella, hoy se hace por procedimientos automáticos informáticos, aunque finalmente el analista debe marcar la categorización definitiva según los intereses de investigación. Condiciones que debe cumplir:
 - a. Exclusión mutua, no puede haber dos categorías conexas.
 - b. Homogeneidad, para el contenido que refleja la categorización
 - c. Productividad, pensado en proporcionar índices para inferencias
 - d. Significativas, relevantes
 - e. Replicables, cualquier investigador puede realizar la misma categorización.
2. **Codificación:** asignación de códigos a cada categoría para facilitar el tratamiento de datos, esta codificación debe permitir que las categorías que representan cumplan las condiciones de exclusión y consistencia.
3. **Fiabilidad:** se puede estimar calculando el porcentaje de veces que varios codificadores independientes coinciden cuando codifican el mismo material. Para realizar el estudio se acostumbra a que sean como poco dos codificadores los que realicen el proceso.
4. **Validez:** reflejada por la exhaustividad cuando las categorías agotan todo el contenido del documento y por la productividad o capacidad de que el conjunto de categorías proporcionen resultados ricos para plantear inferencias e hipótesis.
5. **Análisis de los datos:** para ello es necesario:
 - a. **Fase Descriptiva**, que analizara la frecuencia de aparición de las categorías y sus relaciones internas o externas con otras variables.
 - b. **Fase Inferencial**, se intentará sacar conclusiones, ver el poder de generalización.
 - c. **Fase Multivariante**, analizará las estructuras complejas de las categorías y las relaciones de los bloques de contenidos con estas.

5. TIPOLOGÍAS DE ANÁLISIS DE CONTENIDO.

5.1 Análisis de Contenido Temático: se centra en la presencia de términos o conceptos con independencia de las relaciones entre los mismos. Se utiliza como técnica de recuento, la identificación y clasificación temática y la búsqueda de palabras en contexto. Al buscar una temática en un contexto se ha de definir antes de comenzar el análisis:

El **Recuento de Palabras**, es muy simple y fácil de realizar con procedimientos informáticos, aunque será necesario tener en cuenta cuestiones morfológicas: singular/plural, preposiciones, conjunciones, artículos... para evitar duplicidad de significado y otras por no aportar información relevante.

La **Identificación y Clasificación Temática**, se puede ejecutar después de conocer la frecuencia de aparición de palabras, llevará a fijar términos relevantes para el objetivo de investigación. Otra forma es trabajar con una **"lista abierta"** de temas, para agregar descriptores a medida que aparecen nuevos temas en un corpus (todo el conjunto de texto del discurso) bajo estudio. Se puede hacer de forma manual o bajo aplicaciones destinadas al análisis temático automático (**Pasw Text**), que se apoyan en un diccionario donde registran todos los términos a tener en cuenta así como los sinónimos relacionados, para así poder "extraer" los temas tratados y su frecuencia. Para el éxito de este proceso es crucial la calidad del diccionario. Existe dificultad de interpretación automática de palabras homógrafas para lo que estos programas han desarrollado procedimientos que eligen un significado en función del contexto.

Las palabras en contexto tratan de seleccionar palabras no solo por su frecuencia sino en función del contexto. Programas como **Anatex** permite la creación de palabras registradas junto con su contexto de forma al aparecer la palabra se seleccionara el contexto más apropiado y se facilita la creación de núcleos de contenido. La técnica de **KWIC** permite la búsqueda de cadena de caracteres de diferente amplitud, desde raíces de palabras hasta palabras compuestas o frases.

5.2 Análisis de contenido Semántico: pretende estudiar las relaciones entre los temas tratados en un texto, en consecuencia se debe definir los patrones de relaciones que se tomarán en cuenta: relaciones gramaticales o relaciones semánticas. Existen **varias técnicas:** *análisis de matriz gramatical, análisis de matriz semántica, análisis de componentes, análisis de valorativo, análisis de núcleos y análisis actancial.*

El **Análisis de Matriz Semántica** es el más utilizado, exige como paso previo que se definan los modelos de relaciones que pueden ser codificados, elaborando la llamada **"Gramática Semántica"** o conjunto de modelos con la forma semántica de las relaciones semánticas que interesan. La matriz definida depende directamente de la estructura del fenómeno estudiado y debe permitir recoger diversas formas de expresión del mismo, para lo que es importante comprender la intención del hablante y tener en cuenta el contexto. La unidad de análisis es la clausula (fragmento de texto) que encaja en la matriz definida. Un software capaz de reconocer algunas estructuras es KEDS.

Ejemplo: "expresión de problema – solución requerida" para estudiar las demandas de los ciudadanos ante un gobierno o programa electoral que implica una estructura (matriz) binaria del tipo VERBO (acción a realizar = solución) Vs OBJETO (problema al cual corresponde)

5.3 Análisis de contenido de Redes: se centra en la ubicación relativa de ciertos componentes. Se pueden realizar análisis más complejos partiendo de un análisis de matriz semántica, teniendo en cuenta el tipo de relaciones entre los términos, conceptos o temas.

Por ejemplo, existen trabajos que han construido secuencias de relaciones de causa a efecto (es posible calcular el peso propio de cada causa en relación a otras), también se puede construir redes a partir de estructuras gramaticales, como por ejemplo con **SemNet**, software de propósito general en materia de representación de redes semánticas pues sirve para el análisis de contenido como para organizar ideas, representar estructuras de conocimiento u orientar el estudio.