

ANÁLISIS DOCUMENTAL. INDIZACIÓN. RESÚMENES. INDIZACIÓN AUTOMATIZADA.

I. INTRODUCCIÓN

El interés por el análisis de la información, más allá de sus rasgos puramente formales, comenzó a vislumbrarse a principios del siglo XX, con la publicación del *Traité de la documentation* de Paul Otlet.

El **final de la II Guerra Mundial** trajo consigo el inicio de la llamada “**explosión documental**” y el incremento de publicaciones fue tal que los sistemas tradicionales de control bibliográfico se vieron desbordados. Fue entonces cuando el **ANÁLISIS DOCUMENTAL**, entendido como proceso encaminado a extraer la información contenida en los documentos, hizo su aparición definitiva.

II. EL ANÁLISIS DOCUMENTAL

Es uno de los eslabones clave de la **CADENA DOCUMENTAL**, integrada por procesos de entrada, tratamiento, difusión y mantenimiento de la información.

María Rosa Garrido, autora de una de las mejores definiciones del concepto de análisis documental, lo concibe como el «conjunto de operaciones que conducen a representar el CONTENIDO y FORMA del documento primario de modo distinto al original, elaborando el llamado documento secundario, dotado de todos los elementos informativos necesarios, que permiten tanto su identificación como su recuperación».

FASES

Atendiendo a la **CONCEPCIÓN INTEGRAL** del análisis documental, defendida por autores como **María Pinto Molina** o **López Yepes**, éste consta de **dos fases**:

- **ANÁLISIS FORMAL O EXTERNO**

Estudia el **CONTINENTE o soporte** del documento y se corresponde, fundamentalmente, con la parte de **descripción bibliográfica** del proceso de catalogación. En ella, se recopilan los elementos informativos que permiten identificar inequívocamente un documento (título, autor, edición, etc.) y se redactan conforme a unas normas internacionalmente reconocidas: las **ISBD** y, en menor medida, las AACR. En España, contamos con las **Reglas de Catalogación** del Ministerio de Cultura, que siguen el modelo ISBD, pero con peculiaridades propias.

Esta descripción bibliográfica se completa posteriormente con **puntos de recuperación o acceso** (encabezamiento y asientos secundarios) y con **datos de localización física** del documento (signatura topográfica).

- **ANÁLISIS DE CONTENIDO O INTERNO**

Persigue la **REPRESENTACIÓN CONCEPTUAL** de la información y, en última instancia, la **recuperación colectiva de documentos que aborden la misma materia**. Las **tres operaciones** que posibilitan este análisis son:

- **Clasificación:** Consiste en la agrupación de los materiales según su temática para facilitar su emplazamiento y posterior búsqueda. Puede establecerse mediante sistemas de clasificación **alfabéticos**, como las listas de encabezamientos de materia, o **sistemáticos**, como la CDU.
- **Indización:** Es la técnica utilizada para extraer las ideas más representativas de un documento y traducirlas a unos términos adecuados, procedentes del **lenguaje natural** o de un **lenguaje documental** previamente elaborado.
- **Resumen:** Es la representación del contenido intelectual de un documento de forma más precisa, explícita y **extensa**. Aunque no requiere de ningún lenguaje documental, sí existe una cierta normalización en cuanto a su redacción.

Así, según este enfoque, el análisis documental implica la **identificación del documento original**, tanto en sus aspectos formales como de contenido.

Otros autores, como **Chaumier** y **Gardin**, defienden una **CONCEPCIÓN MÁS RESTRICTIVA**, según la cual el análisis documental comprendería sólo la fase de análisis del contenido y, básicamente, los procesos de **indización** y **resumen**, que paso a detallar con más detenimiento a continuación:

III. INDIZACIÓN

Según la norma **ISO 5127**, la indización es la representación del contenido de un documento mediante la **selección de palabras significativas** extraídas de su contexto o mediante la **atribución de términos** pertenecientes a un lenguaje controlado. Tal labor ayuda a almacenar y localizar la información presente en el fondo bibliográfico.

El grado de profundidad en la indización dependerá de la política de cada centro, pero, en todos ellos, se buscará el equilibrio entre exhaustividad y concreción, para evitar **dos fallos graves** en el momento de la recuperación:

- **Ruido:** Se produce cuando los resultados obtenidos tras la búsqueda no se corresponden con la demanda de información planteada. La aparición de documentos no pertinentes se debe a una indización con más términos de los debidos.
- **Silencio:** Ocurre cuando el sistema no recupera ningún documento asociado a la búsqueda, a pesar de formar parte del fondo documental. No los selecciona por haber sido indizados de manera demasiado específica.

TIPOS DE INDIZACIÓN

Como se deduce de la definición anterior, se diferencia entre:

- **INDIZACIÓN EN LENGUAJE LIBRE O NATURAL**

Es aquella que se realiza mediante **palabras clave** o mediante **descriptores libres**, es decir, sin utilizar un vocabulario rígido preestablecido. En el primer caso, el indizador **extrae** directamente del documento las palabras (no vacías) más representativas. En el segundo, es el propio indizador el que **asigna** términos del lenguaje natural, que no vienen explícitamente en el texto, aunque sí implícitamente, puesto que reflejan su contenido.

Las LISTAS DE PALABRAS CLAVE O DE DESCRIPTORES LIBRES se elaboran **sobre la base de indizaciones previas en lenguaje libre**, como herramientas de referencia para conseguir una cierta uniformidad descriptiva.

- **INDIZACIÓN EN LENGUAJE CONTROLADO**

Es aquella que se realiza mediante **términos recopilados de antemano en un lenguaje documental** elaborado con tal fin. En los centros de documentación, la herramienta de indización por excelencia es el TESAURO y, en este caso, las unidades léxicas asignadas para representar el contenido conceptual de los documentos reciben el nombre de **descriptores** (así llamados por **Mooers** en los años 50). Con este tipo de indización se busca huir de la riqueza del lenguaje natural y, sobre todo, de los accidentes lingüísticos inherentes a éste: sinonimia, polisemia, ambigüedad, etc.

MÉTODOS DE INDIZACIÓN

Desde la implantación de las TIC, es necesario distinguir entre:

- **INDIZACIÓN HUMANA**

Basada exclusivamente en el **trabajo intelectual del indizador**, con independencia de que éste seleccione los términos del lenguaje libre o de uno controlado, según el caso.

- **INDIZACIÓN AUTOMÁTICA O AUTOMATIZADA**

Apoiada en **aplicaciones informáticas**, para tratar de mecanizar una actividad esencialmente intelectual. Existen diferentes niveles de automatización:

1. **Indización asistida por ordenador**, en la que la informática ayuda al documentalista a almacenar términos previamente indizados por él. Gracias a su base de datos, el ordenador permite la asignación de tales términos sin necesidad de teclearlos.
2. **Indización semiautomática**, en la que los programas informáticos realizan sus propuestas de indización, tras haber analizado los documentos de modo automático, pero es el documentalista el que las elige y valida.
3. **Indización completamente automática**, en la que los términos propuestos por el ordenador se almacenan directamente como descriptores, sin validación alguna por parte de un profesional.

En principio, una indización perfecta no puede ser completamente automatizada, ya que los sistemas informáticos aún presentan importantes **deficiencias**, como la incapacidad de reconocer diferentes significados en una misma palabra o de distinguir conceptos en vez de cadenas de palabras no vacías. No obstante, sí representan un importante **ahorro de tiempo** y ofrecen **alternativas atractivas, sin carga subjetiva** alguna.

En España, la mayoría de bibliotecas y centros de documentación siguen llevando a cabo la indización intelectual o humana. En cambio, en países como Francia o Estados Unidos, existe una gran tradición investigadora en este campo. Algunos programas de indización automática destacables son, por ejemplo, **SPIRIT** (de la sociedad SIEMENS), basado en fundamentos lingüísticos y estadísticos (frecuencia de aparición de los términos más representativos) o **INDEXICON** (de la corporación norteamericana ICONOVEX), que se encarga de leer documentos para localizar los términos más importantes y generar un listado.

IV. RESUMEN

ISO 214:1976

Según la norma **UNE 50-103:1990**, el resumen constituye la **representación abreviada, precisa y objetiva** del contenido de un documento original. Se considera la operación más complicada del análisis de contenido, por el gran esfuerzo intelectual que comporta, pero, a la vez, la más valorada por el usuario, dado su enorme interés informativo.

El resumen, como producto documental, es un **documento secundario en sentido amplio**, que condensa información completa y representativa del documento primario, para:

- Servir de **anticipo**, o incluso de **sustituto, del documento original**, ya que su rápida lectura permite al usuario decidir sobre la conveniencia o no de leer el texto completo.
- Facilitar la **comprensión de los documentos**, especialmente de aquellos escritos en idiomas desconocidos para el usuario.
- Dar a conocer las **novedades** en el campo científico.
- Ayudar en las **labores de búsqueda** retrospectiva y **recuperación** de la información.
- Facilitar las **tareas de indización**, puesto que el resumen recoge las ideas y términos más importantes del documento primario.

TIPOS DE RESUMEN

Existen diversos criterios para establecer las distintas categorías de resumen, si bien los dos más frecuentes son la **autoría** y la **densidad informativa**:

Según su origen o autoría, pueden ser:

- **De autor original**. Su valor reside en que éste conoce a la perfección el contenido conceptual, pero, como contrapartida, suele desconocer la técnica y caer a menudo en la subjetividad.
- **De expertos**. Una vez más, el redactor es especialista en la materia, pero suele desconocer la metodología. A diferencia del anterior, presenta mayor objetividad.
- **De documentalistas**. Son los resúmenes óptimos, sobre todo si los profesionales encargados de elaborarlos están familiarizados con la materia, pues conocen las técnicas de resumir y cuentan con la debida preparación (capacidad analítica y de síntesis, amplia cultura, etc.).

Según su **densidad informativa o nivel de análisis**, los tres más representativos son:

- **Indicativo (descriptivo)**. No suele superar las **50 palabras** y sirve para alertar al usuario de la existencia del documento y de los temas generales que aborda. Se utiliza en centros con pocos recursos o con excesiva carga de trabajo.
- **Informativo**. Su extensión oscila entre las **100 y 300 palabras** y recoge todos los aspectos relevantes del documento primario (objetivos, métodos y resultados o conclusiones). Es el **resumen documental por excelencia**.
- **Analítico**. Representa el nivel más alto de descripción posible y puede alcanzar o exceder las **500 palabras**, en función del texto original y del personal disponible. Es poco habitual por el alto consumo de tiempo y recursos que supone.

Existen, no obstante, otros, como el **informativo-indicativo**, que combina los dos primeros, o el **crítico**, que juzga de alguna manera el contenido, método o resultados del original, etc.

METODOLOGÍA

El proceso de elaboración de un resumen presenta **dos etapas básicas**: análisis y síntesis:

1. El **ANÁLISIS** comienza con la **lectura del documento**, que permite identificar su contenido cognitivo y comprender el mensaje. A partir de ahí, se **selecciona lo sustancial y se elimina lo accidental**, dando lugar a un proceso intelectual en el que se desmenuzan las ideas fundamentales que irán incluidas en el resumen. Sobre esta base, se cimentará la síntesis.

2. La **SÍNTESIS** conlleva gran dificultad, pues supone elaborar información nueva a partir de los datos extraídos del documento fuente. El fin es obtener un **documento secundario autónomo y fiel al original**, que no altere el mensaje, ni lo interprete o complete con elementos semánticos ajenos. El resumidor deberá, por tanto, desplegar todas sus artes para recomponer el texto desmenuzado en la fase de análisis de forma más concisa e intentando respetar la estructura y principios básicos expuestos a continuación:

ESTRUCTURA

En la medida de lo posible, el resumen deberá dar respuesta a los siguientes puntos y en el siguiente orden:

- **Objetivos y alcance**, es decir, el propósito del autor al escribir el documento.
- **Metodología** utilizada por el autor: técnicas, procesos, etc.
- **Resultados** obtenidos y/o **conclusiones** extraídas, que constituyen la esencia de la investigación.

Alternativamente, podrá citarse **información complementaria**, como la existencia de gráficos explicativos, si se considera relevante para el estudio.

PRINCIPIOS

En la redacción de todo resumen han de tenerse en cuenta, además, una serie de **PRINCIPIOS BÁSICOS**, tales como la **entropía** o economía de signos, **pertinencia** o fidelidad a la fuente, **concisión**, **coherencia** (estructura organizativa lógica y ordenada), **claridad**, uso preferible de la **tercera persona o pasiva refleja** y del **tiempo presente** para lograr mayor objetividad, **corrección gramatical**, redacción en **un solo párrafo**, etc.

Todas estas pautas relativas a la presentación y al estilo de los resúmenes responden a un deseo de **NORMALIZACIÓN**, plasmado en normas internacionales, como la **ISO 214:1976**, y nacionales, como la **UNE 50-103:1990**.

Por último, en cuanto a los **intentos de automatización en el campo de los resúmenes**, cabe señalar que, a pesar de los esfuerzos realizados, los resultados obtenidos no han sido demasiado satisfactorios. Los sistemas ideados hasta el momento logran extraer frases significativas del documento, pero se muestran incapaces de confeccionar resúmenes estructurados y lógicos.

V. CONCLUSIÓN

El **análisis documental** actúa de intermediario entre el productor y el usuario en el proceso de circulación de la información, únicamente posible gracias a la labor del documentalista. En este contexto, operaciones tan importantes como la indización o el resumen se convierten en piezas clave para la difusión del conocimiento, más aún ante el gran número de publicaciones existentes. Por eso, ambas tareas han de ser objeto de constante evaluación e innovación para evolucionar hacia mejor cada día.

ANÁLISIS DOCUMENTAL. INDIZACIÓN. RESÚMENES. INDIZACIÓN AUTOMATIZADA.

I. INTRODUCCIÓN

II. EL ANÁLISIS DOCUMENTAL

FASES

- ANÁLISIS FORMAL O EXTERNO
- ANÁLISIS DE CONTENIDO O INTERNO

} **CONCEPCIÓN INTEGRAL** del análisis documental
(Pinto Molina, López Yepes) ≠ Concepción más restrictiva (Chaumier, Gardin)

III. INDIZACIÓN

TIPOS DE INDIZACIÓN

- INDIZACIÓN EN LENGUAJE LIBRE O NATURAL
- INDIZACIÓN EN LENGUAJE CONTROLADO

MÉTODOS DE INDIZACIÓN

- INDIZACIÓN HUMANA
- INDIZACIÓN AUTOMÁTICA O AUTOMATIZADA

IV. RESUMEN

TIPOS DE RESUMEN

- Según su origen o autoría
- Según su densidad informativa o nivel de análisis

METODOLOGÍA

1. Análisis
2. Síntesis

ESTRUCTURA

1. Objetivos y alcance
2. Metodología
3. Resultados y/o conclusiones

PRINCIPIOS

V. CONCLUSIÓN

ANÁLISIS DOCUMENTAL. INDIZACIÓN. RESÚMENES. INDIZACIÓN AUTOMÁTICA.

I. INTRODUCCIÓN

- Interés análisis de la info. ppios s. XX → *Traité de la documentation*, Paul Otlet.
- Fin II GM “**explosión documental**” → análisis documental hizo su aparición definitiva.

II. EL ANÁLISIS DOCUMENTAL

- Eslabón clave de la CADENA DOCUMENTAL.
- Def. **María Rosa Garrido**.

FASES

- **CONCEPCIÓN INTEGRAL** (María Pinto Molina, López Yepes) → 2 fases:

A) ANÁLISIS FORMAL O EXTERNO

- Estudia CONTINENTE, se corresponde fundamentalmente con la parte de DESCRIPCIÓN FÍSICA en catalogación.

B) ANÁLISIS DE CONTENIDO O INTERNO

- Persigue REPRESENTACIÓN CONCEPTUAL info. y **recuperación colectiva** docs que comparten materia / 3 operaciones:
 1. **Clasificación** = agrupación de materiales según su temática (puede ser **alfabética** o **sistemática**).
 2. **Indización** = técnica para extraer ideas ppales y traducirlas a unos términos adecuados (lenguaje **libre** o **documental**).
 3. **Resumen** = representación contenido intelectual docs de forma más precisa, explícita y extensa (≠ lenguaje documental).
- Este enfoque → **identificación doc. original**, tanto en sus aspectos formales como de contenido.
- Otros (Chaumier, Gardin) → **CONCEPCIÓN MÁS RESTRICTIVA**: sólo fase análisis de contenido (indización y resumen).

III. INDIZACIÓN

- Def. **norma ISO 5127** (selección vs. atribución).
- Equilibrio entre exhaustividad y concreción para no dar lugar a RUIDO o SILENCIO en la recuperación.

TIPOS

A) INDIZACIÓN EN LENGUAJE LIBRE

- Realizada mediante **palabras clave** o mediante **descriptores libres** (≠ lenguaje rígido y preestablecido). LISTAS.

B) INDIZACIÓN EN LENGUAJE CONTROLADO

- Realizada mediante **términos recopilados de antemano en un lenguaje documental** elaborado con tal fin. TESAUROS.

A) INDIZACIÓN HUMANA

- Resultado exclusivo del **trabajo intelectual del indizador**.

B) INDIZACIÓN AUTOMATIZADA O AUTOMÁTICA

- Apoyada en **aplicaciones informáticas**. Distintos niveles:
 1. **Indización asistida por ordenador**.
 2. **Indización semiautomática**.
 3. **Indización completamente automática**.
- Una indización perfecta no puede ser completamente automatizada (deficiencias vs. ventajas). España vs. FR, EEUU.

IV. RESUMEN

- Def. **norma UNE 50-103:1990** (representación abreviada, concisa y objetiva del contenido doc.). Labor más complicada.
- Producto documental → **documento secundario en sentido amplio**, FUNCIONES.

TIPOS

Según su origen o autoría:

- a) **De autor** (☺ contenido ☹ técnica y subjetividad).
- b) **De expertos** (☺ redactor especialista en la materia ☹ desconocimiento metodología / mayor objetividad).
- c) **De documentalistas** (☺ resúmenes óptimos: conocimiento técnica de resumir y preparación adecuada).

Según su densidad informativa o nivel de análisis:

- a) **Indicativo (descriptivo)**. ~ 50 palabras (existencia doc. y temas grales).
- b) **Informativo**. ~ 100 y 300 palabras (todos los aspectos relevantes doc., **resumen documental por excelencia**).
- c) **Analítico**. + 500 palabras (nivel más alto de descripción posible, en función texto original y personal, poco frec.).
- d) **Otros** (informativo-indicativo, crítico...).

MÉTODOLOGÍA

1. **ANÁLISIS**. Lectura doc. / **Seleccionar lo sustancial**, eliminar lo accidental → proceso intelectual: desmenuzar ideas.
2. **SÍNTESIS**. Elaborar info. nueva / **Doc. secundario autónomo y fiel** al original / Recomposición ideas desmenuzadas.

ESTRUCTURA

1. **Objetivos y alcance** = propósito autor.
2. **Metodología** = técnicas, procesos... empleados por el autor.
3. **Resultados y/o conclusiones** = esencia de la investigación.

PRINCIPIOS

- Ppios básicos: entropía, pertinencia, claridad, concisión, coherencia, 3ª persona, pasiva refleja, presente, un solo párrafo...
- Estas pautas presentación y estilo responden a deseo de NORMALIZACIÓN (**ISO 214**, **UNE 50-103:1990**).
- Estado de la automatización de resúmenes.

V. CONCLUSIÓN

- Análisis documental = **intermediario entre productor y usuario** / Indización y resumen claves en la difusión info. / Evolución.