

NUDIST: UNA HERRAMIENTA INFORMÁTICA PARA EL ANÁLISIS DE DATOS CUALITATIVOS

Ana Corrales, Gregorio Rodríguez, Javier Gil y Eduardo García
Universidad de Sevilla

Desde la perspectiva de la metodología cualitativa la introducción del ordenador en el proceso de investigación ha supuesto un cambio radical y una nueva forma de pensar en el método cualitativo (Richards y Richards, 1991; García et al., 1993). Efectivamente, el uso del ordenador en el análisis de datos cualitativos de acuerdo con unas instrucciones (programas) que respondan a las necesidades de operaciones conceptuales que en cada caso se requieren, ofrece avances que van desde una mayor velocidad y capacidad de memoria, hasta las habilidades para localizar frecuencias de cadenas de caracteres en los textos pudiendo recuperarlas con rapidez mediante el uso de códigos, o llegar a la construcción de teorías, pasando por la comprobación de hipótesis cualitativas (Gil et al., 1993).

Actualmente existen numerosos programas informáticos que permiten realizar las operaciones requeridas en los procesos de análisis de datos cualitativos y que hasta la aparición de los mismos debían ejecutarse manualmente (Tesch, 1988, 1989, 1990, 1991; García et al., 1993). En el presente trabajo vamos a centrar nuestra atención en la descripción de las principales características y aportaciones que uno de ellos, concretamente el programa NUDIST (en su versión 2.3 para ordenadores compatibles), brinda al análisis de datos cualitativos.

EL SISTEMA NUDIST

Creado en la Universidad de La Trobe (Melbourne, Australia) NUDIST es un programa informático que nos permite manejar, organizar y realizar procesos de investigación en los que se manejan datos cualitativos de carácter textual, producto de entrevistas, observaciones, documentos históricos o literarios, notas de campo, noticias de periódicos, etc.

Siempre que se aborda el análisis de datos cualitativos es preciso acometer los siguientes procesos (Richards y Richards, 1992):

- Agrupar todos los documentos que se desean analizar en el proyecto;
- Indizar¹ segmentos del texto en varias categorías de indización;
- Buscar palabras y frases en los documentos;

¹ Los autores del programa insisten en la diferencia fundamental entre «indizar» y «categorizar», dando al primer término el sentido de una jerarquía de categorías, como podremos ver más adelante.

— Emplear la indización y búsqueda de texto como pilares básicos para encontrar párrafos en los textos y las ideas que en ellos se expresan;

— Elaborar notas y memos sobre las ideas principales y teorías a medida que se desarrolla el proyecto; y,

— Reorganizar y establecer la indización a medida que nuestros conocimientos y teoría aumentan en el proceso de investigación.

El programa NUDIST realiza estos procesos de forma eficaz y contiene, además, muchas otras características que no se encuentran en otros programas de análisis de datos cualitativos. De hecho, debido a su gran capacidad para indizar, buscar y teorizar —partes fundamentales del análisis cualitativo— este programa se ha llamado «NUDIST» que son la iniciales de «Non-Numerical Unstructured Data Indexing, Search and Theorizing» («Indización, búsqueda y elaboración de teorías de datos no numéricos y no estructurados»).

Para que el programa NUDIST trabaje eficazmente se crean un sistema de documentación y un sistema de indización, que son la base y el soporte de los distintos procesos que se ejecutan posteriormente.

En el **sistema de documentación** se puede almacenar, en primer lugar, cualquier cantidad de información de documentos textuales «on-line» (esto es, archivos en el propio disco del ordenador), con datos sobre los mismos que se deseen analizar a lo largo de la investigación. En segundo lugar se pueden almacenar documentos «off-line» (es decir, documentos que no están en el disco sino fuera del ordenador, como libros, notas, fotografías, etc.).

Ambos documentos (on-line y off-line) están divididos en unidades de texto como líneas, frases, diálogos, o cualquier otra opción se desee, siempre que se especifique claramente al programa NUDIST qué consideramos como unidad de texto. Asimismo, estos documentos y los datos acerca de los mismos, se pueden estudiar y visualizar de muy diversas formas. Podemos optar por su visualización a través de la pantalla del ordenador o salvarlos en un archivo para imprimirlo posteriormente, siempre con la cantidad de texto deseada.

Además, los documentos se pueden indizar bajo las categorías de indización que se deseen, pudiendo colocar encabezamientos (tan largos como se estime) para describir el contenido del documento, su origen ...etc., y subencabezamientos en el caso de los documentos on-line, también con cualquier longitud. Tanto el encabezamiento como el subencabezamiento aparecen con cualquier recuperación del documento del que procede.

El **sistema de indización**, por su parte, permite almacenar referencias a ideas, conceptos y hechos que surgen a lo largo de la investigación, y referencias a unidades de texto de los documentos on-line y off-line que muestren esas ideas, conceptos o hechos. El sistema de indización puede ser tan sencillo o complejo como el investigador lo desee o necesite, pudiendo contener cualquier cantidad de categorías de indización y, en cada categoría, las referencias de indización a las unidades de texto de los documentos de datos on-line y off-line que se deseen. Estas categorías, que pueden, además, tener subcategorías que a su vez pueden tener otras y así sucesivamente, se pueden crear durante los procesos de indización de los documentos, o bien mientras se avanza en la investigación, en un proceso continuo de mejora y adaptación del sistema de indización a las necesidades de la propia investigación.

Comprender lo que es un sistema de indización es fundamental para poder trabajar con NUDIST, aunque el programa lo ofrece como una posibilidad pudiendo realizarse los procesos de análisis tomando los códigos que se hayan elegido sin establecer ninguna relación entre los mismos, de ahí que profundicemos en algunos conceptos que son necesarios tener en cuenta a la hora de proceder al análisis de datos con este programa.

NUDIST permite organizar las categorías de indización en «árboles» de categorías, sub-categorías,

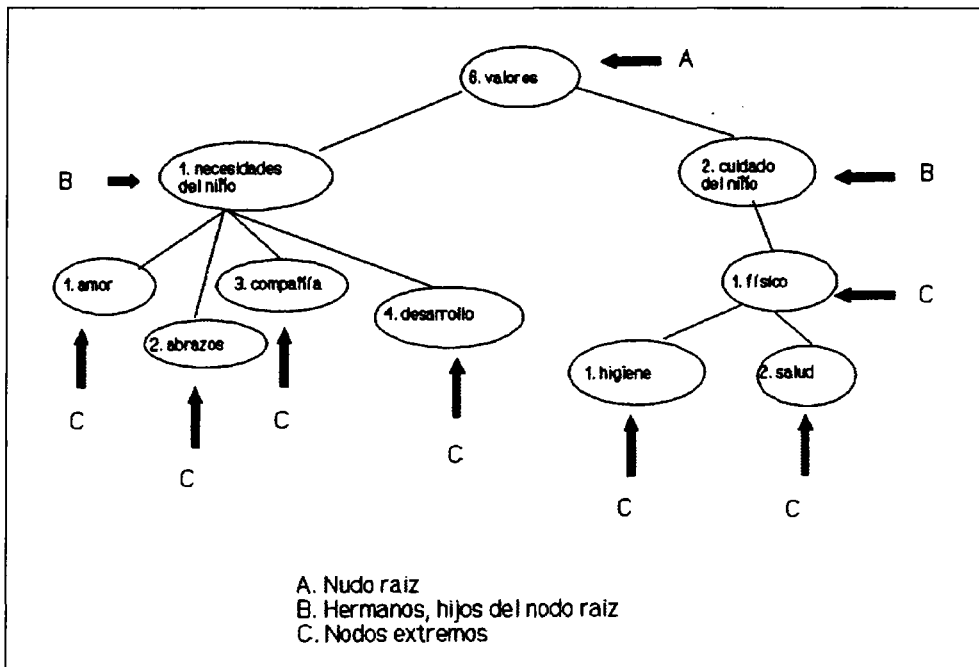


Figura 1. Ejemplo de árbol de indización.

sub-sub-categorías, etc. Los puntos que se encuentran al final y en la unión de las ramas son denominados **nodos**, no existiendo límite alguno en cuanto al número.

En la Figura 1 podemos contemplar una parte de un sistema de indización, en el que podemos observar cómo el nodo que se encuentra en el nivel superior recibe el nombre de **raíz**, no tiene a nadie por encima, pero sí puede tener «hijos» por debajo. Estos últimos que parten de los nodos raíces son conocidos como **hermanos**. Por último, los nodos que carecen de «hijos» son denominados **terminales o extremos**.

Cada nodo tiene un número, conocido como **número del nodo**, que van a permitir construir la **dirección del nodo**, a partir de la estructura jerárquica. Así, el nodo 1 2 1, se corresponde con el título higiene. En lugar de números, también cada nodo puede identificarse por el **título**, de esta forma al nodo 1 2 1 le corresponde el título *valores/cuidado del niño/higiene*.

La creación de las categorías de indización se puede llevar a cabo de tres formas diferentes:

- Tecleando las referencias de indización de las unidades de texto de un documento.
- Mediante la realización de una búsqueda automática de texto, localizando palabras, patrones de palabras o expresiones en los documentos on-line, empleando NUDIST para añadir las referencias resultantes a la base de datos de indización.
- Mediante el uso de una extensa gama de órdenes que permite crear nuevas categorías de indización a partir de las ya existentes.

La base de datos de indización puede ser de cualquier complejidad, desde una simple lista de códigos hasta estructuras en árbol de categorías y subcategorías de indización de una enorme complejidad. Esta estructuración permite organizar y manejar los conceptos indizados como sistemas teóricos, no simplemente como etiquetas.

Mediante el sistema de indización se puede:

- a. Asignar categorías de indización referidas a documentos completos, pudiendo almacenar información ilimitada sobre el documento.
- b. Almacenar comentarios en cualquier categoría de indización, grabar ideas sobre esa categoría o, por ejemplo, relaciones de esa categoría con otras.
- c. Recorrer el sistema de indización a través de la pantalla y guardar en un archivo cualquier operación realizada para imprimirlo posteriormente.
- d. Cambiar la indización de cualquier documento o cambiar libremente las categorías de indización cuando se desee.
- e. Recuperar cualquier parte de los documentos empleando órdenes que permiten cambiar las categorías de indización para expresar preguntas e hipótesis.
- f. Los resultados de la búsqueda de texto se pueden salvar en una categoría de indización. Esto significa que los resultados de búsqueda de texto, la indización manual y la construcción de nodos están completamente integrados en la base de datos de indización, y que pueden ser utilizados como base para estudios posteriores.

LOS PROCESOS QUE NUDIST NOS FACILITA

1. Navegar a través del sistema de indización

El sistema de indización es la vida y el alma de un proyecto de investigación que se sirva de NUDIST, de tal forma que el programa nos permite utilizar distintas herramientas para poder analizarlo, estudiarlo y, en su caso, modificarlo. En cualquier etapa del proceso de investigación en el que nos encontremos se puede hojear cualquier categoría o grupos de categorías en el sistema de indización, lo que se indiza y qué otras categorías se refieran a éstas.

2. Integrar la búsqueda de texto y la indización

La mayoría de los programas especializados en el análisis de datos cualitativos, o bien permiten la codificación y recuperación, o bien realizan búsqueda de texto. Sin embargo, el programa NUDIST combina la exploración de información de indización con búsquedas de expresiones que ocurren en el texto, así se puede buscar cualquier secuencia de caracteres en los documentos y recuperarla con la cantidad de texto deseada (la sección completa o incluso el documento entero en donde aparece).

Además, NUDIST permite guardar el resultado de una búsqueda de texto en un nodo del sistema de indización como indización de referencias de lo encontrado. Como cualquier otro nodo, se le puede asignar un nombre, incluir comentarios en él y contenido con información de indización sobre las unidades de texto de los documentos. Además, se puede hojear, alterar mediante la supresión o inclusión de algunas unidades de texto, mover o combinar en cualquier operación para construir nuevas categorías a partir de las ya existentes.

3. Comentarios en el sistema de indización

Lo importante al apoyar el desarrollo de teorías sobre los datos es la posibilidad de realizar comentarios sobre las categorías de indización, y éstos pueden proceder de dos fuentes principales: el investigador y el propio programa. Añadiendo, o borrando, en cualquier categoría de indización comentarios de textos el investigador dispone de una forma adecuada de almacenar notas de trabajo, memos, opiniones y pensamientos acerca de las categorías de indización. A su vez, NUDIST añade

automáticamente su propio comentario a las categorías de indización para grabar el momento de su creación o cualquier otro evento importante en su historia, proporcionando de esta forma una revisión continua del desarrollo de las ideas en el proceso de investigación.

4. Modificar el sistema de indización

Mientras la mayoría de los sistemas manuales y los programas de codificación-recuperación permiten, en cierta medida, recodificar los documentos, NUDIST está diseñado para controlar con una gran flexibilidad el sistema de indización y todas las indizaciones realizadas en el mismo. Efectivamente, el sistema de indización en NUDIST se puede hojear, explorar, reorganizar, cambiar y crecer a medida que nuevos conceptos e ideas sobre los datos emergen.

El objetivo es, pues, la máxima flexibilidad y control sobre la información. El investigador puede modificar el sistema de indización sin que tenga que recurrir de nuevo a los documentos previos; toda la información del sistema de indización puede recuperarse junta para ser recolocada en nuevos patrones de indización. Es más, determinados conceptos específicos se pueden agrupar automáticamente para poder explorar conceptos generales, incluso cuando estos últimos no han sido codificados explícitamente. Si la localización de un nodo (categoría de indización) en el sistema global no está en el sitio adecuado o deseado, NUDIST permite recolocarlo, con la información de indización, en el lugar donde proceda, y todos los nodos y subnodos siguientes pueden ser combinados, borrados o recolocados.

Así pues, esta flexibilidad permite que NUDIST se muestre como una útil herramienta de apoyo a las investigaciones que se inscriben en línea con la *teoría fundamentada* (Glaser y Strauss, 1967; Straus, 1987; Strauss y Corbin, 1990), método que se centra en la captura e interrogación de los significados emergentes de los datos.

5. Buscar y recuperar

Mediante la utilización combinada de una amplia variedad de operadores booleanos (Y, O y NO, etc.) y, a su vez, combinados con una serie de operadores no booleanos, se pueden establecer relaciones entre las unidades del texto y las estructuras de indización. NUDIST examina cuidadosamente los documentos recuperados para ayudar a manejar eficazmente la gran cantidad de información que se haya guardado en los documentos. Con cada recuperación de texto aparece el encabezamiento (en el que se ha grabado información sobre el documento), subencabezamientos referidos a cada pasaje recuperado (en los que se han podido anotar cuestiones como quién está hablando, las preguntas realizadas, etc., por ejemplo, en el caso de que se tratase de una entrevista), estadísticos numéricos sobre la recuperación y documentación que NUDIST ofrece sobre la historia de cada recuperación realizada.

6. Crear nuevas categorías de indización a partir las anteriores

NUDIST permite búsquedas indefinidas de ideas que surgen con posterioridad porque el resultado final de cualquier recuperación es un nuevo nodo en el sistema de indización, que puede ser explorado cuando se desee. Cada nuevo nodo que se construye está dotado de toda la información que acompaña a cualquier nodo: título, comentarios, fecha, facilidades de hojear, estadísticas, ...etc. De esta forma, al igual que en el modelo de la teoría fundamentada, el análisis que nos permite NUDIST se convierte en un proceso de ir explicitando las ideas emergentes, permitiendo tanto la construcción de teoría, como la comprobación de la misma.

7. Ficheros de comando para grandes cantidades de procesos

En lugar de realizar las operaciones a través de los menús que NUDIST facilita, éstas se pueden controlar y ejecutar a través de ficheros de comandos que el investigador construya. Por ejemplo, si se han de introducir numerosos archivos, o realizar búsquedas de texto complejas en varios archivos a la vez, o asignar códigos a determinadas unidades de texto, NUDIST permite hacerlo siguiendo las instrucciones escritas en un archivo de comandos en vez de tener que realizarlo a través de las selecciones del menú.

Esencialmente los archivos de comando pueden utilizarse para introducir documentos, crear nuevos árboles o sub-árboles de indización, buscar en los archivos on-line textos o patrones de texto y clasificar automáticamente lo hallado o ejecutar comandos para construir nodos y así crear nuevas categorías de indización a partir de la anteriores.

8. Integrar análisis de datos textuales y numéricos

Además de la lógica integración que podemos establecer entre, por ejemplo, los datos de identificación como sexo, edad, etc., que se suelen recoger en numerosas investigaciones, un aspecto importante de NUDIST es la posibilidad que nos ofrece de una serie de datos «estadísticos», como producto de las búsquedas realizadas, que posteriormente pueden ser utilizados para su análisis a través de procedimientos estadísticos.

9. Comprobación de hipótesis

Una de la posibilidades que nos ofrecen los ficheros de comandos es la comprobación de hipótesis de causalidad o de relaciones a lo largo de los textos objeto de investigación, entendiendo por hipótesis «las relaciones entre tópicos identificados en un texto a partir de las cuales es posible reconstruir un todo estructurado que nos aproxima a una comprensión global de los datos» (Gil et al. 1993: 1).

TRABAJANDO EN EQUIPO CON NUDIST

Aunque NUDIST 2.3 se sirve de menús y diálogos para que el investigador pueda controlar sus acciones, empleando sofisticadas formas para hojear, todas sus presentaciones se ofrecen por pantalla, simulando a un terminal de salida simple. Incluso empleando un Macintosh o Window 3.0 en un ordenador compatible, la ventana simple en la que NUDIST 2.3 funciona mantiene la misma estructura, de tal forma que el trabajo se asemeje con independencia del tipo de ordenador que se esté utilizando.

El programa NUDIST puede ejecutarse bien de forma individual o bien en equipo de trabajo. Tanto en versión para PC/Windows o Macintosh, se puede utilizar de forma individual, de tal forma que un único investigador puede controlar todo el proceso de investigación.

Pero si se utiliza una red local, o bien la versión para Unix o VAX/VMS, entonces existe la posibilidad de que varios investigadores compartan el mismo proyecto de investigación, aún cuando cada uno puede asumir distintas responsabilidades dentro del mismo.

BIBLIOGRAFÍA

GARCÍA, E. et al. (1993): *Análisis de datos cualitativos en la investigación sobre la diferenciación educativa*. Ponencia presentada al VI Seminario de Métodos de Investigación, El Escorial.

- GIL, J. et al. (1993): *Formulación y comprobación de hipótesis cualitativas con ayuda del programa AQUAD*. Comunicación presentada al VI Seminario de Métodos de Investigación, El Escorial.
- GLASER, B. G. y Strauss, A. L. (1967): *The Discovery of Grounded Theory*. Nueva York: Aldine Publishing Company.
- RICHARDS, L. y RICHARDS, T. J. (1991): The Transformation of Qualitative Method: Computational Paradigms and Research Processes. En N. Fielding y R. Lee (Eds.): *Using Computers in Qualitative Research*. Berkeley: Sage, (pp. 38-53).
- RICHARDS, T. et al. (1992a): *NUDIST 2.3. User Manual*. Eltham: Replee P/L.
- RICHARDS, T. et al. (1992b): *NUDIST 2.3. Reference Manual*. Eltham: Replee P/L.
- STRAUSS, A. L. (1987): *Qualitative Analysis for Social Scientist*. MA: Cambridge University Press.
- STRAUSS, A. L. y Corbin, J. (1990): *Basics of Qualitative Research. Grounded Theory Procedures and Techniques*. Newbury Park: Sage.
- TESCH, R. (1988): *The Impact of the Computer on Qualitative Data Analysis*. Comunicación presentada en la convención anual de la AERA, Nueva Orleans.
- TESCH, R. (1989): Basic Qualitative Analysis with QUALPRO. *International Journal of Qualitative Studies in Education*, 2 (4), 367-371.
- TESCH, R. (1990): *Qualitative Research: Analysis Types and Software Tools*. Hampshire: Falmer Press.
- TESCH, R. (1991): Software for Qualitative Researchers: Analysis Needs and Program Capabilities. En N. G. Fielding y R. M. Lee (Eds.): *Using Computers in Qualitative Research*. Londres: Sage Publications.