

NUDIST VERSUS AQUAD: DOS POSIBILIDADES INFORMÁTICAS PARA EL ANÁLISIS DE DATOS CUALITATIVOS

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

Desde que en 1966 apareciera el primer programa para el análisis de textos, el denominado *General Inquirer* (Stone et al., 1966), hasta la fecha muchos han sido los avances habidos en este campo, sobre todo en los últimos años de la década de los 80 y estos primeros 90.

Tesch (1988) presentaba ya un primer análisis comparativo de los programas de análisis de datos cualitativos disponibles en el mercado en ese momento, haciendo referencia a programas como THE ETHNOGRAPH, QUALPRO, QUALOG, TAP... Esta misma autora, en 1990 nos presenta un excelente trabajo en el que sintetiza éstos y otros programas, recogiendo sus principales funciones (Tesch, 1990). En esta segunda revisión puede observarse una rápida evolución en cuanto a la cantidad de programas, así como en la prestaciones que ofrecen y su disponibilidad. Tan sólo tres años después de la publicación del texto de Tesch, algunos programas han sufrido una gran transformación; éste es el caso de programas como NUDIST¹ (*Non-numerical Unstructured Data Indexing, Search and Theorizing* —Indización, búsqueda y elaboración de teorías de datos no numéricos y no estructurados) o AQUAD (*Analysis of Qualitative Data*). El primero se presenta en 1992 en su versión para ordenadores personales, en versiones para PCs y Macintoshes, y tiene ya anunciada su versión 3.0 para finales de este mismo año. El segundo, referido al caso concreto de nuestro país, en su versión 3.0 en castellano presenta mejoras sustanciales con respecto a las anteriores.

A lo largo de este trabajo vamos a centrar nuestra atención en estos dos programas (NUDIST y AQUAD) por ser considerados actualmente los dos programas que nos sirven como medio para la construcción y comprobación de teoría (Huber, 1990; Tesch, 1990; Richards, 1992).

Aunque AQUAD surge con posterioridad a NUDIST, en el caso de nuestro país es el programa que más se ha venido utilizando y ello por varios motivos. En primer lugar, por la difusión que se ha hecho del mismo: su creador —Günter Huber— ha impartido cursos en varias universidades españolas, exponiendo los fundamentos del programa e iniciando al manejo del mismo; ha participado en reuniones científicas celebradas en nuestro país (Huber, 1988) o ha publicado en revistas nacionales artículos sobre el tema (Huber y Marcelo, 1990). En segundo lugar por ser un programa que exige pocos requerimientos de hardware y está en castellano.

¹ Para una visión más amplia sobre NUDIST puede consultarse la comunicación de A. Corrales et al. (1993) «NUDIST»: *Una herramienta informática para el análisis de datos cualitativos* presentada en este Seminario.

NUDIST, por contra, hasta 1992 no ha dispuesto de una versión para ordenadores personales, lo cual ha limitado su expansión. Asimismo, en esta nueva versión los requerimientos de hardware son bastante elevados, representando ello un obstáculo para su utilización.

Independientemente de estas primeras diferencias, lo que realmente nos interesa en estos momentos son las posibilidades y limitaciones de cada uno de los programas de cara al análisis de datos cualitativos, y en ello centraremos nuestra atención, aunque con las limitaciones propias que la longitud fijada en este trabajo nos impone, para lo cual presentaremos un breve análisis comparativo de ambos programas.

1. ¿CÓMO TRABAJAN ESTOS PROGRAMAS?

En una primera aproximación podríamos decir que son bastante similares, por cuanto que a través de los mismos podemos llevar a cabo los principales procesos que todo análisis de datos cualitativos exige (García et al., 1993). NUDIST y AQUAD facilitan la tarea de análisis de datos cualitativos permitiendo codificar unidades de significado diferenciables en el texto, efectuar recuentos de tales unidades, examinar el contenido de las categorías resultantes o comprobar hipótesis. No obstante, desde el comienzo del trabajo con uno u otro programa existen ya diferencias apreciables.

a) Preparación del texto para el análisis

La base del análisis la constituye el texto con el que se trabaja, que para ambos programas debe presentarse en ficheros con formato ASCII, aunque con exigencias diferentes según veremos a continuación.

En el caso de AQUAD, es preciso numerar las líneas de texto, operación que el programa realiza automáticamente. Las líneas del texto han de estar formateadas a una longitud máxima de 48 caracteres (pulsaciones) por línea, de tal forma que quede en la pantalla espacio suficiente para agregar códigos a los márgenes, o bien sin limitación cuando se opta por introducir la localización de los códigos directamente en un fichero de códigos. El texto puede tener una longitud de 9999 líneas, alcanzando el límite de 64 KBytes de Turbo Prolog para strings (cadenas de signos). No obstante, pueden ser analizados múltiples textos a la vez, por lo que bastaría con segmentar en varios ficheros aquellos textos que rebasen el tamaño fijado.

NUDIST, por su parte, no tiene casi limitación alguna en la longitud de la unidad de texto, que puede llegar a ser de 32000 caracteres por párrafo. Si se opta por tomar como unidad de texto la línea habría que formatear el texto a 69 caracteres por línea; pero también puede optarse por tomar como unidad de texto la frase, el párrafo, la intervención (en un diálogo) o cualquier otra forma que se desee. A diferencia de AQUAD, NUDIST no permite realizar la codificación de forma directa sobre el propio texto en pantalla.

Otra diferencia por lo que se refiere al tipo de texto que permite analizar, es lo que NUDIST denomina ficheros «off-line», es decir, fichero donde se hace referencia a archivos, fotos, cintas o cualquier otro objeto que no se encuentra físicamente en el fichero del ordenador.

b) Codificación del texto

La codificación del texto implica asignar a cada unidad de significado diferenciable en el texto, un código que haga referencia al contenido de dicha unidad, y he aquí donde surge la primera gran diferencia entre ambos programas.

En NUDIST este proceso es denominado «sistema de indización», por cuanto que lo que se crean

son índices donde se ordenan jerárquicamente todas las categorías, y no se pone restricción alguna en cuanto al número de categorías del sistema o al número de categorías de una unidad de significado, pudiendo comenzar por unas pocas e ir construyendo el sistema poco a poco ya que pueden modificarse en cualquier momento.

Cuando codificamos un texto mediante el programa AQUAD, utilizamos códigos y metacódigos. Los códigos son asignados a determinados fragmentos de texto añadiendo junto a la línea donde comienza el fragmento codificado una etiqueta de identificación seguida del número de línea donde finaliza el fragmento. Por ejemplo, si quisiéramos codificar un fragmento de texto comprendido entre las líneas 15 y 19, en el que se habla acerca de la evaluación, utilizando para ello el código EVA, escribiríamos al final de la línea 15 la expresión VA-19. Un mismo fragmento puede ser codificado con varios códigos, puede estar incluido o superpuesto parcialmente con un fragmento correspondiente a otro código. Los fragmentos correspondientes a una serie de códigos vinculados por algún criterio, pueden ser codificados automáticamente por un metacódigo.

Tanto NUDIST como AQUAD permiten efectuar un recuento de las apariciones de códigos en los textos analizados, haciendo posible el análisis de los datos mediante procedimientos tales como el análisis secuencial, los modelos lineales logarítmicos, el análisis factorial de correspondencias y el escalamiento no métrico, el análisis de correlación canónica, los modelos logit, etc. (Ver Ponencia sobre «Análisis de datos cualitativos en la investigación sobre la diferenciación educativa»).

c) Examinar el contenido de las categorías y los sistemas de codificación

En NUDIST las recuperaciones de texto son ilimitadas. No hay restricciones en el número de veces que un pasaje de texto puede ser indizado, y virtualmente ninguna restricción en las preguntas que puede hacerse sobre la indización de líneas de un texto, o la indización en archivos o la información sobre el propio archivo. Se pueden utilizar operadores booleanos en cualquier combinación de unos con otros, y combinados con operadores no booleanos que establecen relaciones de los pasajes en el texto y estructuración de indización.

Además, NUDIST permite examinar los propios documentos y el sistema de indización de forma integrada, permitiendo de esta forma al usuario el control para reorganizar y reformar todos los aspectos del sistema. En NUDIST, el sistema de indización puede ser ojeado, explorado, reorganizado y cambiado, creciendo a medida que nuevos conceptos emergen de los datos.

El principal objetivo de AQUAD no es la organización de sistemas de categorías jerarquizados, por lo que su flexibilidad para renombrar códigos es mucho menor. En cambio, sí contribuye eficazmente a la revisión del contenido de las categorías, permitiendo la recuperación de todos los fragmentos etiquetados con determinados códigos.

d) La comprobación de hipótesis

Considerando la comprobación de hipótesis con las limitaciones que han de entenderse desde el análisis de datos cualitativos por ordenador —búsqueda de evidencia que apoye una afirmación—, tanto AQUAD² como NUDIST permiten esta comprobación, aunque con diferencias.

AQUAD ofrece un repertorio de hipótesis sobre vínculos existentes entre categorías, basadas todas ellas en relaciones de secuenciación, inclusión, superposición o coaparición dentro de una distancia

² En el caso concreto del programa AQUAD puede consultarse la comunicación de J. Gil et al. (1993): *Formulación y comprobación de hipótesis cualitativas con ayuda del programa AQUAD*, presentada en este Seminario.

CUADRO 1
 CARACTERÍSTICAS DE LOS PROGRAMAS NUDIST Y AQUAD

	NUDIST 2.3	AQUAD 3.0
Lenguaje de programación	• LISP	• Turbo-Prolog
Versiones	• VAX / PC-386 /Mac	• PC
Hardware necesario	• PC-386DX o SX • 4 Mb (RAM); 2 Mb (HD)	• PC XT • 640 K (RAM)
Entrada	• Fichero ASCII	• Fichero ASCII
Límites de los ficheros	• Ilimitado	• 64 K por fichero
Comandos	• Todos los comandos están disponibles en pantalla • Ficheros de comandos	• Comandos disponibles en pantalla
Principales procedimientos	• Indizar, codificar, recuperar, buscar, contar, numerar líneas, comprobación de hipótesis.	• Numerar líneas, codificar, recuperar, contar, comprobar hipótesis
Códigos	• Sistema de indización en árbol • Nodos ilimitados	• Sistema de códigos y metacódigos • Posibilidad de codificación múltiple
Procedimiento de codificación	• Directamente • Por búsqueda de texto • A través de comandos	• Directamente • A través de comandos
Modificación de la codificación	• Borrar, modificar, cambiar, sustituir, añadir...	• Añadir metacódigos
Salida de datos	• Pantalla • Fichero ASCII	• Pantalla • Impresora • Fichero ASCII
Ventajas fundamentales	• Flexibilidad • Inicio de codificación automática	• Formulación de hipótesis propias
Inconvenientes más graves	• Manual y programa en Inglés • Requerimiento de Hardware	• Excluye comprobación de relaciones no basadas en coaparición de códigos
Manual	• Guía para comenzar • Manual del usuario • Manual de Referencia	• Manual
Precio	±24.000 Ptas	±20.000 Ptas.

máxima de los fragmentos correspondientes a dos o más códigos. Permite además al analista formular sus propias hipótesis en lenguaje turbo-prolog, para buscar evidencia de la aparición de diferentes constelaciones de códigos que se repiten a lo largo del texto.

En NUDIST la comprobación de hipótesis puede llevarse a cabo de forma fácil a través de los ficheros de comandos, permitiendo contrastar la proximidad de determinadas categorías o estableciendo ciertas relaciones causales o de otro tipo.

e) La rapidez en el trabajo

A fin de evitar un trabajo monótono en el que sería preciso repetir muchas operaciones, NUDIST prevee la ejecución de tales tareas a partir de la realización de ficheros de comandos, a través de los cuales se puede ejecutar acciones como introducir los documentos que se van a analizar, ejecutar procesos de búsqueda guardando los resultados en un nodo, crear nodos y darles títulos, crear árboles enteros, indizar un documento, ejecutar operaciones de creación de nodos más complejas que las realizadas de forma interactiva. En definitiva, a través de este sistema se pueden realizar las operaciones más fundamentales e importantes, sin necesidad de tener que estar interactuando continuamente con el ordenador.

En el caso de AQUAD, es necesario ir seleccionando las operaciones que deseamos realizar a partir de un menú visualizado en la pantalla.

2. A MODO DE CONCLUSIÓN

En forma de síntesis, en el cuadro 1 ofrecemos una información básica sobre las principales diferencias existentes entre los dos programas analizados, y aun cuando consideramos que es el propio usuario quien debe tomar las decisiones en función de las características y objetivos de su investigación, podemos decir que NUDIST es un programa más flexible que AQUAD, que nos permite trabajar con mayor cantidad y diversidad de formatos de texto y de forma más rápida. Pero también adolece de unas limitaciones como son el estar íntegramente en inglés y con unas presentaciones en pantalla que no aprovecha el entorno gráfico de Windows, limitación esta última que parece que se subsanará en la nueva versión 3.0 de la que se dispondrá a finales de 1993, lo cual le hará situarse entre los programas que trabajan como una aplicación bajo Windows. Y en cuanto al requerimiento de hardware, el que precisa ya se viene convirtiendo poco a poco en estándar.

Como hemos podido comprobar a lo largo del presente trabajo, las operaciones que se suelen realizar en los procesos de análisis de datos cualitativos quedan suficientemente facilitadas por cualquier de los dos programas informáticos analizados. No obstante, si el investigador está interesado en llevar a cabo una asignación de códigos jerarquizados, con lo que ello conlleva de supuestos implícitos, no cabe duda que su elección ha de dirigirse hacia el programa NUDIST.

REFERENCIAS BIBLIOGRÁFICAS

- CORRALES, A. et al. (1993): «*NUDIST*»: *Una herramienta informática para el análisis de datos cualitativos*. Comunicación presentada al VI Seminario de Métodos de Investigación, El Escorial.
- 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.
- HUBER, G. (1988): *Análisis de datos cualitativos: la aportación del ordenador*. En C. MARCELO (Coord.): *Avances en el Estudio del Pensamiento de los Profesores*. Sevilla: Servicio de Publicaciones de la Universidad de Sevilla, (pp. 77-85).

- HUBER, G. y MARCELO, C. (1990): Algo más que recuperar palabras y contar frecuencias: la ayuda del ordenador en el análisis de datos cualitativos. *Enseñanza*, (8), 69-84.
- RICHARDS, L. y RICHARDS, T. J. (1991): The Transformation of Qualitative Method: Computational Paradigms and Research Processes. En N. Fielding y R. L. (Eds.): *Using Computers in Qualitative Research*. Berkeley: Sage, (pp. 38-53).
- RICHARDS, T. et al. (1992). *NUDIST 2.3. User Manual*. Eltham: Replee P/L.
- RICHARDS, T. et al. (1992): *NUDIST 2.3. Reference Manual*. Eltham: Replee P/L.
- STONE, P. J., DUNPHY, D. C., SMITH, M. S. y ORGILVIE, D. M. (1966): *The GENERAL INQUIRER: A computer approach to content analysis*. Cambridge, MC: The M.I.T. Press.
- 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. Fielding y R. L. (Eds.): *Using Computers in Qualitative Research*. Berkeley: Sage, (pp. 16-37).