

Informática aplicada a las ciencias sociales

JORDI COLOBRANS

INTRODUCCIÓN

A principios de junio de 1998 me llegó un email: “Hola Jordi. Encontrarás la continuación una consulta que según creo puedes resolver. Saludos. François”.

Me comprometí a intentarlo. Se trataba de construir una base de datos para introducir topónimos y analizar cartas portulanas medievales (mapas manuscritos del Mediterráneo) para la Unidad Docente de Paleografía y Diplomática de la Facultad de Geografía e Historia de la Universidad de Valencia. El problema era el siguiente: En la British Library de Londres y otras bibliotecas del mundo había varias cartas portulanas medievales, unas pocas eran de autor y fecha conocido, el resto estaban por determinar. ¿Era posible averiguar la autoría o la escuela de cartógrafos desde la que se habían trazado las cartas y ubicar cronológicamente los documentos?

François tenía razón. La informática podía ayudar a resolver aquel problema. A mediados de junio se había aprobado el presupuesto y empecé a colaborar con Ramón J. Puchades Bataller, el investigador al frente del proyecto.

A la aplicación se le dio el nombre de ‘Portolanes’. La primera tarea fue distinguir los tipos de prestaciones que tenía que ofrecer esta base de datos. Básicamente tres: introducción, comparación y presentación de datos. Para la introducción y dado el volumen de información, un centenar de cartas y entre unos 1.500 y 1.700 topónimos por carta (± 150.000 registros), se tuvieron que resolver varias cuestiones relativas a las preferencias para la automatización y la autointroducción de datos. La tarea más importante de esta fase fue la de conceptualizar la aplicación y determinar las categorías analíticas. De esta fase de categorización depende, en gran medida, el posterior desarrollo de la aplicación.

Una vez introducidos los datos, éstos debían poder dar los perfiles de cada una de las cartas portulanas. Se seleccionaba el código de la carta, se pulsaba un botón y al cabo de unos instantes aparecía definido el perfil de la carta seleccionada que quedaba almacenado en formato numérico y por porcentajes. Con los datos introducidos y los perfiles ya podía hacerse la pregunta para la que había estado construida la base de datos: Comparar una carta con otra y determinar las probabilidades de que dos o más cartas pertenecieran a un mismo autor-escuela/ fecha.

Toda la aplicación se planteó de manera gráfica y visual, pensando en que el usuario no tenía que tener grandes conocimientos de informática, es decir, que la informatización del trabajo no debía suponerle un problema, sino resolver los que ya tenía. Se asignó una pantalla a cada tarea y botones de navegación y comandos en todos ellos, salvo los específicos para cada caso. Como usuario debía poder ser autosuficiente, no depender del desarrollador. La aplicación debía prever los posibles errores que podían producirse y corregirlos antes de que el usuario tropezara con ellos.

Ramón Puchades participó activamente durante todo el proceso de desarrollo. Primero definiendo y perfilando los objetivos de la investigación y estableciendo las categorías analíticas. Posteriormente, cuando empezó a trabajar sobre el prototipo y a partir de su experiencia con datos reales, planteando modificaciones que agilizaran sus tareas de introducción de datos. Finalmente, y pensando en los anexos de su tesis, definiendo el tipo de listados con su contenido, tipo de ordenación y formato para argumentar las conclusiones.

Desde que se inició el desarrollo de esta aplicación hasta su plena funcionalidad pasaron dos meses. El primer mes se dedicó al planteamiento, desarrollo, diseño y el segundo a optimizar, ajustar el interface y modificar la presentación de datos. La aplicación se desarrolló en FileMaker Pro 4.0, compatible con Mac y PC. El canal de comunicación más utilizado fue el correo electrónico y el teléfono.

La experiencia con 'Portolanes' demostraba que la relación entre informática y ciencias sociales es, en muchos casos, no un problema técnico sino una cuestión de diferencias entre lenguajes especializados. Ramón Puchades sabía lo que quería y cómo lo quería, pero lo expresaba en su propio lenguaje, el de un historiador. Mi tarea consistía en trasladar y adaptar su manera de pensar a la forma y manera de procesar la información característica de las bases de datos. En realidad yo hacía de traductor, conjugando mi formación en Ciencias Sociales con la formación informática que adquirí posteriormente

INFORMÁTICA Y CIENCIAS SOCIALES

Los cambios producidos por el impacto de las nuevas tecnologías plantean nuevos desafíos, nuevas posibilidades. El trabajo que se realizó con 'Portolanes' difícilmente hubiera podido llevarse a cabo sin el apoyo de las herramientas informáticas. Sólo con las funciones de autointroducción de datos se ahorró la friolera de más de 2.000 horas de trabajo. Para obtener el retrato robot de una carta basta con hacer un clic en un botón y apenas un minuto de espera. Cruzar los datos de dos cartas es otro clic en un botón y un par de

minutos. Manualmente estas operaciones hubieran supuesto horas, muchas. La informática ofrece nuevas condiciones de trabajo tanto en la empresa, la oficina, como en casa o la universidad.

Investigar supone negociar con diversas fuentes de información. En la actualidad y en muchos proyectos de investigación –por su alcance y complejidad, por trabajar en coordinación con otros grupos internacionales, p.e.– estas fuentes, por su volumen, desbordan la capacidad humana de gestión de la información. Sólo con las tareas de búsqueda y selección de la información relevante, se emplea una parte preciosa del tiempo disponible. Este tiempo es un tiempo ‘escasamente intelectual’, si se nos permite introducir este término, es decir, es un tiempo de espera y toma de decisiones preliminares relativamente poco creativo, destinado a diferenciar los materiales sobre los que se podría iniciar un análisis. Piénsese en la actualidad, por ejemplo, en una búsqueda en Internet. Si dispongo de tres horas, ¿cómo puedo consultar 500 páginas web, cada una de las cuales quizá encierre algún link relevante que desconozco y que me oriente hacia el objeto/ sujeto de la investigación? O en un ejemplo más característico de los antropólogos: A la vuelta de un trabajo de campo etnográfico se han acumulado 1.500 hojas de diario de campo y 50 entrevistas en profundidad. La condición es la siguiente: 4 meses para vaciar los materiales y presentar su informe y, a la vez, por si fuera poco, 10 horas semanales de carga académica docente.

Evidentemente, en este último caso, el problema de fondo va mucho más allá que el de la interdisciplinariedad y el de establecer puentes y enlaces entre distintas especialidades académicas. Se solapa con la dinámica propia de los departamentos universitarios, el tipo de relaciones establecidas, su cultura organizativa, las dificultades presupuestarias y demás factores contextuales. Pero estas condiciones de trabajo son habituales en la cultura académica española, especialmente en los departamentos de Ciencias Sociales y Humanidades

Por otro lado, el propio proceso de investigación conlleva un sinfín de tareas que pueden rutinizarse y que absorben un porcentaje importante del tiempo del investigador (desde el diseño y presentación del proyecto, hasta las correcciones tipográficas y el formato final del artículo / libro, pasando especialmente por la fase de recogida de datos, ordenación, catalogación y clasificación de documentos y materiales, y su análisis). En todas estas fases, los investigadores acostumbran a racionalizar el proceso con anterioridad, –planificación a medio y largo plazo–, aunque otros prefieren depositar sus expectativas en la intuición y solucionar los problemas en la medida que vayan surgiendo –planificación a corto plazo–. Son dos maneras de trabajar. Hay más. Estas intuiciones unas veces funcionan de manera inexplicable, pero otras generan una serie de conflictos que se van arrastrando y van creciendo a lo largo de la investigación –p.e. no definir claramente el tema, el problema, la pregunta, o el propósito de la investigación; o no profundizar en las categorías analíticas y descubrir que después de 350 horas de trabajo éstas no resultan operativas–.

La informática puede resolver algunos de estos problemas, no todos, tanto en lo referido al diseño y planteamiento de la investigación, la recopilación y análisis de los datos como en la redacción y el control de los procesos de investigación. Sin embargo he aquí el handicap: Los científicos sociales no

siempre poseen conocimientos informáticos, no siempre pueden acceder a los servicios informáticos, y, cuando acceden, no siempre consiguen que los informáticos les entiendan o que los informáticos construyan lo que les piden. Otras veces sí y el intercambio tiene un final feliz.

BASES DE DATOS PARA INVESTIGADORES

En la actualidad, la variada oferta de hardware y software puede hacer que la compra de un equipo informático o la elección de un software sea un verdadero problema. La incertidumbre del 'qué equipo necesito', si con 4 ó 6 gigas de ROM, 64 ó 128 RAM, o con un microprocesador de 330 ó 450 Mhz, y la variedad y volatibilidad del desembolso a realizar, generan un tipo de angustia, en muchos casos, gratuita.

El equipo informático se debe elegir una vez que se sabe para qué se va a utilizar. Con el software ocurre algo parecido. Los científicos sociales generalmente utilizan procesadores de texto y, algunas veces, especialmente las nuevas generaciones, programas de gráficos y dibujo, y bases de datos. Otros por su especificidad incorporan programas de manipulación de vídeo o sonidos y retoque de imágenes. Unos pocos, aprenden a utilizar aplicaciones del tipo Nud*ist, Childes, Etnograph, Folio, Intext, Aquad, Ceta, Diction, Ask-Sam, Martin, WinMax, AnthroPack, -todos en inglés- para el análisis fundamentalmente cualitativo de documentos.

De todos estos, los procesadores de texto fueron los primeros en incorporarse al equipo de herramientas de los investigadores sociales. Más tarde, las bases de datos y los programas de gráficos y dibujo, junto con los de estadística.

Mi experiencia me está demostrando que las bases de datos son una herramienta de trabajo muy cómoda y con muchas posibilidades. En la actualidad, las bases de datos relacionales tipo Oracle, FileMaker, Acces o dBase, ofrecen una serie de prestaciones muy útiles a los investigadores y grupos de investigación. La madurez a la que están llegando estas bases de datos-entornos de programación ofrece desde las típicas funciones de clasificación, ordenación, búsqueda, importación/ exportación de datos, impresión, hasta la automatización de tareas, creación de gráficos, cálculos sofisticados, interfaz visual, funciones típicas de los procesadores de texto, publicación en web, etc. Todo un complejo de recursos de los cuales las Ciencias Sociales y las Humanidades pueden aprovechar una parte para sus tareas habituales.

Sin embargo, antes de pedir algo a una base de datos, repasemos cómo funcionan. A grandes rasgos todas las bases de datos relacionales comparten una serie de características, por ejemplo, siempre se distingue entre un modo de 'trabajo' desde el que se construye la base de datos y un modo de 'visualización', desde el que la estructura de la base de datos no puede ser modificada y desde el que los usuarios introducen o consultan datos. Para esta descripción voy a tomar como terminología de referencia la del FileMaker, la que utilizo habitualmente y que, de momento, está resolviendo los problemas que se me plantean.

Una base de datos relacional puede tener varios 'archivos' conectados entre sí, cada uno de los cuales pueden contener datos a distintos niveles jerárquicos (una tesis posee varias partes, cada parte, a su vez, varios capítulos, y

cada capítulo distintos apartados). Dentro de cada archivo se crean ‘registros’ –uno para cada nueva entrada, p.e. distintas referencias bibliográficas–. Desde el modo ‘trabajo’, en cada registro pueden introducirse ‘campos’. Cada campo tiene una etiqueta y un espacio para la información (El campo ‘autor’ poseerá información acerca de los autores que se introduzcan en él). Los campos pueden ser de varios tipos: de texto, numéricos, contenedores –cuando almacenan un sonido, una imagen o una película–, de cálculo o de sumario –cuando realizan operaciones–, y globales, si el contenido del campo debe ser el mismo para todos los registros. Los registros creados pueden verse, o no, en distintas ‘presentaciones’. Cada presentación puede tener su propio diseño. Los elementos gráficos pueden tener asignadas determinadas tareas automatizadas que se activan al pulsar sobre ellos. Estas tareas o funciones se configuran aparte. Es una de las tareas que requiere más experiencia y habilidad y el lugar en el que se establecen las principales diferencias entre desarrolladores.

Por otra parte, las bases de datos no sólo son una herramienta para la manipulación de datos, también implican una manera de pensar los datos y relacionarse con ellos. Planificar una investigación pensando en qué se va a pedir a los datos, cómo se van a manejar, cómo se van a presentar, es decir, yendo de la estructura del resultado al inicio del proceso de investigación, facilita su planificación y diseño. Preparar un archivo para la introducción de datos, supone hipotetizar con categorías analíticas. En este sentido, la flexibilidad para modificar la estructura del análisis o ampliar las categorías permite ensayar, o testar, distintas hipótesis de trabajo y simular modelos y categorías de análisis posibles. Se trata de un proceso interactivo en el que se van introduciendo cambios hasta llegar a un punto de ‘satisfacción suficiente’. De esta manera se fusionan las funciones de usuario y desarrollador dando forma a un nuevo perfil, el del ‘usuario espabilado’. En estas condiciones, las bases de datos muestran una dimensión interactiva que hace de ellas una herramienta de trabajo altamente satisfactoria.

EL PROYECTO ODYSSEY

Empecé a construir bases de datos para llevar a cabo tareas propias de las ciencias sociales y las humanidades en 1994, dos años después de haber iniciado mi proyecto de investigación doctoral. Los principios no fueron fáciles. ¿Qué hace un antropólogo metido en eso de la informática?, me preguntaban los antropólogos y los sociólogos con los que me relacionaba por un lado y los informáticos por el otro. ‘Hago interdisciplinariedad’, sigue siendo mi respuesta. Después de la defensa de mi tesis, a mediados de 1996, y a raíz de mi experiencia con las bases de datos me concentré en la creación de herramientas informáticas para investigadores y grupos de investigación.

Di a este proyecto el nombre de ‘proyecto ODYSSEY’. Como tal, se propone informatizar las distintas fases del itinerario de los intelectuales, es decir, informatizar la gestión del trabajo intelectual a distintos niveles. En la actualidad la actividad se concentra en cuatro grandes áreas de desarrollo: a) Diseño y planificación de la investigación; b) Recogida de datos; c) Análisis de datos; y d) Divulgación de resultados y docencia. Cada una de estas áreas contiene varios apartados. Por ejemplo, el apartado de recogida de datos incluye un programa

de gestión bibliográfica –BQAS–, varios módulos de gestión de archivos y documentos, y un pack para la recogida de historias de vida –Memorias–. El apartado de análisis de datos incluye proyectos del tipo Ikra –para el análisis cualitativo de textos–. El de divulgación de resultados proyectos como ‘Papers’ –para la redacción de artículos–, ‘Currículum’ –para la gestión del currículum–, o ‘TecnoTeacher’ para la gestión docente –planificación y diseño de cursos, contenidos, calendario de sesiones, control de calificaciones, etc.–. En estos proyectos acostumbran a colaborar investigadores de distintos departamentos de distintas universidades, grupos y/o investigadores individuales. En octubre de 1998 sumaban un total de 12 subproyectos en curso. Pueden consultarse estos y otros afines en <http://www.dsigsoftware.com>

Como ejemplo de proyecto en curso tomaré el de *DiarioDeCampo* (octubre-diciembre 1998), un software para el vaciado y análisis de entrevistas, diarios de campo y documentos varios que se está desarrollando en colaboración con la Dra. Cristina Larrea, del Departamento de Antropología de la Universidad Central de Barcelona, a raíz de su trabajo de campo en Brasil. Cristina Larrea trabajó como antropóloga dentro del macroproyecto interdisciplinar: ‘Projecto da avaluaéao epidemiológica du programa Bahia Azul’, en concreto, en el subproyecto ‘Acompanhamento das mudanéas das percepções do população resultante das açoes du programa Bahia Azul’. La información recopilada a analizar incluye cuatro diarios de campo elaborados por cuatro investigadoras (‘pesquisadoras’) y un total de 120 entrevistas en profundidad.

Técnicamente, *DiarioDeCampo* resuelve varias necesidades: introducir y almacenar textos completos (importándolos cuando convenga), fragmentar textos para su análisis, disponer de espacios y categorías para el análisis, utilizar herramientas para establecer y consultar campos semánticos, llevar a cabo búsquedas complejas e indirectas (hallar elementos no presentes en un texto pero asociados con otros), cruzar datos e imprimir distintos tipos de listados (componentes de un campo semántico, textos fragmentados analizados, resultados del cruce de datos, nombre de los personajes, escenarios y situaciones que aparecen en los documentos, listados de términos y en qué parte del documento aparecen), varias comodidades de uso (preferencias de autointroducción de datos, pantallas de ayuda, cuadros de diálogo, preferencias de búsqueda, de ordenación y de impresión) y automatización de tareas rutinarias o especializadas.

Las categorías analíticas se han estructurado en función de los objetivos predeterminados por la investigación. Aquí, dado que las bases de datos se han creado con posterioridad a la recopilación de datos, ha exigido a los investigadores un tiempo adicional para reorganizar y homologar la información. En relación a la cuestión del diseño de la base de datos y a su amigabilidad se han seguido los mismos criterios que en *Portolanes* y en el resto de proyectos desarrollados hasta el momento. Quizá alguna versión futura de *DiarioDeCampo* haga las cosas sola, pero de momento las personas siguen siendo las que dirigen la investigación.

CONCLUSIÓN

El proyecto ODYSSEY es un proyecto abierto a los retos, los desafíos y las colaboraciones. Los proyectos realizados tipo BQASv2.5 –Bibliographical

Qualitative Analysis System- testimonian su alcance y envergadura, pero aún quedan muchas islas por visitar y muchas aventuras que vivir en el transcurso de la travesía.

El objetivo de la relación entre informática y ciencias sociales debería ser el de establecer un puente, un lenguaje común entre distintas concepciones del mundo y maneras de interpretar la realidad propias de tradiciones académicas distintas.

Las bases de datos no son sólo una manera de almacenar y gestionar información, son también una manera de pensar e interpretar la realidad. Son, a la vez, una técnica y un estilo cognitivo, una herramienta y una metáfora que permite pensar la realidad del trabajo del intelectual y del investigador de manera distinta. Una herramienta que instaure nuevas condiciones de trabajo. No es lo mismo diseñar y planificar una investigación pensando en dónde, quién y cómo se van a recopilar los datos que, además de lo anterior, pensando en cómo van a organizarse los datos para su rápida disponibilidad, localización, listado, consulta y/o impresión.

Introducir las bases de datos como elemento esencial de la caja de herramientas de intelectuales e investigadores, presupone contemplar una nueva variable en la programación de las actividades. De la misma manera que se elaboran encuestas o entrevistas, deberían construirse bases de datos para trabajar específicamente en cada investigación. Hay veces que, resolviendo la estructura de una aplicación de este tipo, se resuelve buena parte del problema de diseño de la investigación y de los procesos de recopilación, análisis y presentación de datos.

Los bases de datos, interpretadas desde la perspectiva del 'usuario espabilado', es decir, como una herramienta para el diseño y la organización de la investigación, como una herramienta de reflexión interactiva sobre los procesos de investigación y análisis, y, en tercer lugar, como una fuente de consulta, no constituyen la panacea universal, pero sí que pueden simplificar y agilizar buena parte de la gestión del trabajo intelectual. Como consecuencia, el investigador puede ejercer como tal: concentrándose en las tareas creativas e interpretativas, y no empleando su tiempo en tareas rutinarias, manuales o administrativas.