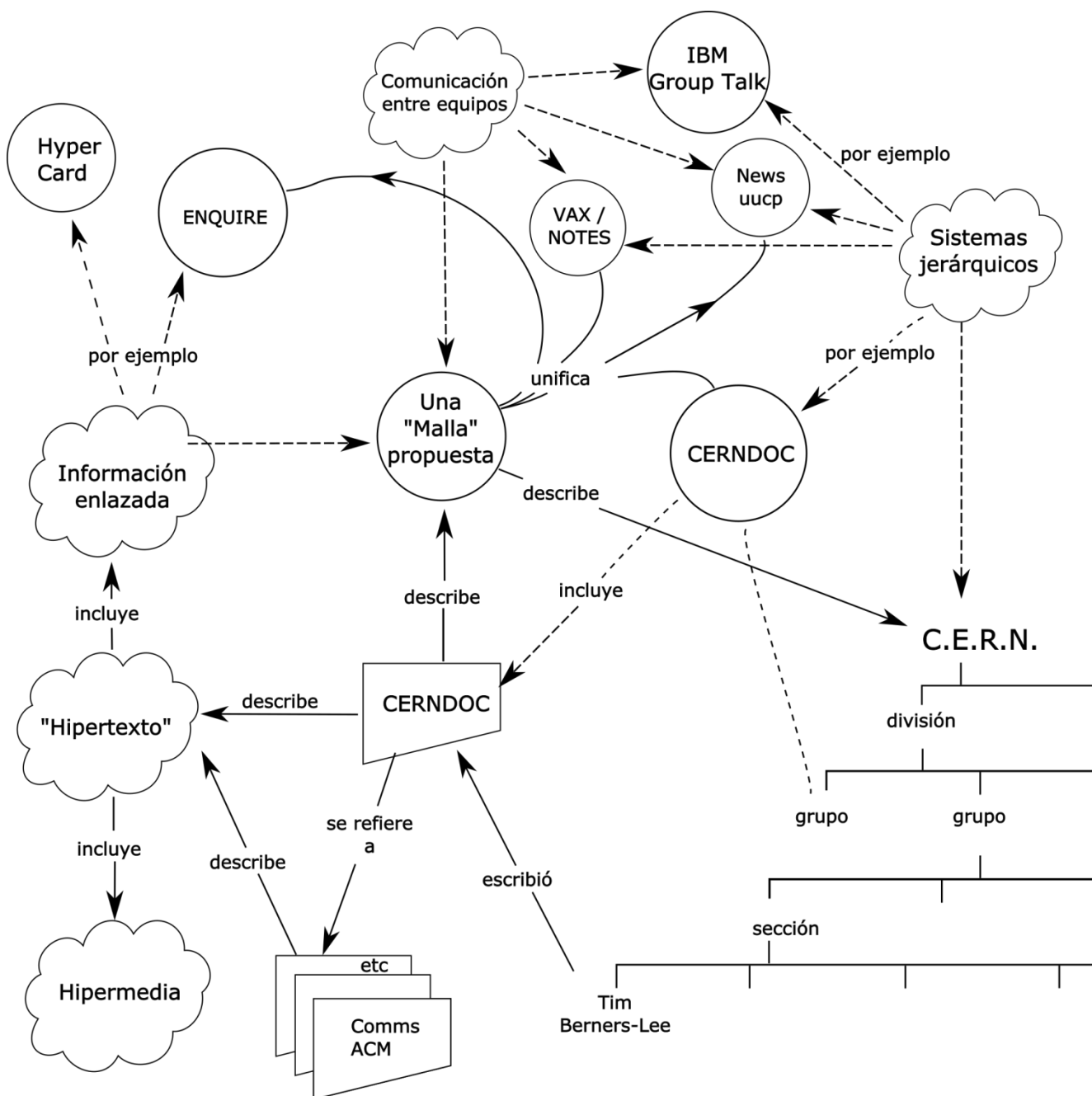


Gestión de la información: una propuesta

Tim Berners-Lee, CERN
Marzo 1989, Mayo 1990

Esta propuesta trata sobre la gestión de información general sobre aceleradores y experimentos en el CERN. Se discute el problema de la pérdida de información en sistemas complejos en desarrollo, y se propone una solución basada en un sistema de hipertexto distribuido.



Visión general

Muchas de las discusiones sobre el futuro en el CERN y en la época del LHC acababan con misma la pregunta: “Sí, pero ¿cómo mantendremos el seguimiento de un proyecto tan grande?”. Esta *Propuesta* ofrece una respuesta a dichas preguntas. En primer lugar, se discute el problema del acceso a la información en el CERN. Más adelante, se introduce la idea de los sistemas de información enlazada, y se comparan con otras formas, menos flexibles, de encontrar información.

A continuación, se resume mi corta experiencia con sistemas de texto no lineales, que se conocen como *hipertexto*; se describe lo que necesita el CERN de un sistema de este tipo, y lo que la industria puede proporcionar al respecto. Finalmente, sugiere los pasos que deberíamos dar para involucrarnos en este momento con el hipertexto, de tal forma que individual y colectivamente podamos entender lo que estamos creando.

Pérdida de información en el CERN

El CERN es una organización maravillosa. En ella están implicadas miles de personas, muchas de ellas muy creativas, trabajando todas juntas con objetivos comunes. Aunque no están organizados nominalmente en una estructura jerárquica, eso no impide que la gente se comunique y comparta información, equipos y software entre los diferentes grupos.

La estructura de trabajo actual de la organización es una *web* multiconectada, cuyas interconexiones evolucionan con el tiempo. En este entorno, una persona que acabe de llegar, o cualquiera que empiece una nueva tarea, recibe normalmente algunos consejos que le serán de utilidad para comunicarse con otras personas. La información sobre estas ventajas y cómo encontrarlas, circula por los pasillos, y los detalles sobre lo que se necesita hacer se difunde de forma parecida. Teniendo esto en cuenta, el resultado es notablemente satisfactorio, a pesar de algunos malentendidos ocasionales y esfuerzos duplicados.

Sin embargo, el gran movimiento de personas que se produce en el CERN representa un problema.

El tiempo medio de estancia en el CERN es de dos años. El resultado es que la información se pierde de forma continua. La entrada de nuevas personas implica una considerable parte de su tiempo, así como el de otras personas, antes de que sepan qué es lo que se cuece dentro de la Organización. Los detalles técnicos de proyectos previos a veces se pierden para siempre, o se recuperan sólo tras una investigación detectivesca en caso de emergencia. A menudo, si la información no se ha grabado, no se puede encontrar de ninguna forma.

Si un experimento del CERN fuera un desarrollo estático, de una sola persona, toda la información generada se podría escribir en un gran libro. Sin embargo, el CERN cambia continuamente, a medida que se generan nuevas ideas, según tenemos acceso a la tecnología, y hacemos frente a los problemas que van surgiendo. Cuando se necesita hacer un cambio, normalmente sólo afecta a una pequeña parte de la organización; por ejemplo, cuando se modifica una parte del experimento o del detector. En este punto, se tiene que imaginar qué otras partes o personas serán afectadas por dicho cambio. Mantener un registro actualizado es algo impracticable, y la estructura del libro debería revisarse constantemente.

El tipo de información que estamos discutiendo responde, entre otras, a preguntas como las siguientes:

- ¿Dónde se usa este módulo?
- ¿Quién escribió este código? ¿Dónde se usa?
- ¿Qué documentos hay disponibles sobre este concepto?
- ¿Qué laboratorios están implicados en el proyecto?

- ¿Qué sistemas dependen de este dispositivo?
- ¿Qué documentos hacen referencia a éste?

Los problemas que plantea la pérdida de información son particularmente agudos en el CERN, pero en este caso (como en otros) el CERN es un modelo en miniatura del resto del mundo en el transcurso de unos pocos años. El CERN se enfrenta ahora a algunos problemas a los que pronto se tendrá que enfrentar el mundo. En 10 años, puede que haya muchas soluciones comerciales a los problemas de los que estamos hablando, mientras nosotros necesitamos algo *ahora* que nos permita continuar¹.

¹ Lo mismo es aplicable, por ejemplo, a las pasarelas de correo electrónico, a la preparación de documentos, y a los sistemas heterogéneos de programación distribuida.

Sistemas de información enlazada

En un sistema de información para manipular este tipo de información, sería deseable que posibilitara la creación de un fondo de información a desarrollar, que pudiera crecer y evolucionar al mismo tiempo que la organización y los proyectos en que está involucrada. Para que esto sea posible

el método de almacenamiento no debe imponer sus propias restricciones a la información.

Esta es la razón por la que una “telaraña”² de notas con enlaces (como referencias) entre ellas, es muchísimo más útil que un sistema jerárquico fijo. Cuando se describe un sistema complejo, mucha gente recurre a diagramas con círculos y flechas, porque nos dejan libertad para describir las relaciones que hay entre las cosas, de una forma que las tablas, por ejemplo, no pueden proporcionar. El sistema que necesitamos es como un diagrama de círculos y flechas, en el que éstos pueden representar cualquier cosa.

A los círculos, podemos llamarlos *nodos*, y a las flechas, *enlaces*. Supongamos que cada nodo es como una pequeña nota, un artículo resumen, o un comentario. No importa si se trata de texto, gráficos, o ambos. Idealmente, representa o describe una persona u objeto concreto. Ejemplos de nodos pueden ser:

- Personas
- Módulos de software
- Grupos de personas
- Proyectos
- Conceptos
- Documentos
- Tipos de hardware
- Objetos de hardware específicos

Las flechas que enlazan el círculo A con el círculo B pueden significar, por ejemplo, que A...

- depende de B
- es parte de B
- hizo B
- hace referencia a B
- usa B

² Berners-Lee utiliza la expresión web (‘telaraña’), una afortunada expresión que fue acogida por la comunidad desde el principio, siendo hoy en día de uso común. La www fue llamada, en consecuencia, ‘telaraña mundial’. [N. del T.]

- es un ejemplo de B

Estos círculos y flechas, nodos y enlaces³, tienen significados diferentes según el diagrama en los distintos tipos de diagramas convencionales:

Diagrama	Los nodos son	Las flechas significan
Árbol genealógico	Personas	"Es pariente de"
Diagrama de flujo	Módulos de software	"Pasa datos a"
Dependencia	Módulo	"Depende de"
Diagrama PERT	Tareas	"Debe hacerse antes"
Organigrama	Personas	"Encarga un informe a"

El sistema debe permitir introducir todo tipo de información. Otra persona debe poder encontrar la información, y no necesariamente tiene por qué saber qué es lo que está buscando.

En la práctica, resulta útil para el sistema tener en cuenta los tipos de enlaces entre los elementos (por ejemplo: dependencias), y el tipo de nodos (personas, cosas, documentos, ...) sin imponer limitación alguna.

El problema de los árboles

Muchos sistemas se organizan jerárquicamente. El sistema de documentación CERNDOC es un ejemplo, igual que lo es el sistema de ficheros Unix, o el sistema VMS/HELP. Un árbol tiene la ventaja práctica de que cada nodo tiene un nombre único. Sin embargo, esto no permite al sistema modelar un mundo real. Por ejemplo, en un sistema jerárquico como VMS/HELP, a menudo tenemos una hoja como

```
HELP COMPILER SOURCE_FORMAT PRAGMAS DEFAULTS
```

sólo para acabar encontrando una referencia a otra hoja: "Por favor, véase

```
HELP COMPILER COMMAND OPTIONS DEFAULTS PRAGMAS"
```

siendo necesario abandonar el sistema, para volver a entrar de nuevo. Lo que necesitábamos, en realidad, era un enlace de un nodo a otro, porque en este caso la información no estaba organizada de forma natural en un árbol.

³ Los sistemas de información enlazada poseen entidades y relaciones. Hay, sin embargo, muchas diferencias entre un sistema de este tipo y un sistema de bases de datos "Entidad-Relación". Por un lado, la información almacenada en un sistema basado en enlaces está orientada a lectores humanos; por el otro, los nodos no tienen tipos [de datos] estrictos, que definen exactamente las relaciones que debe haber entre ellos. Los nodos de tipos similares no tienen por qué almacenarse en el mismo lugar.

Otro ejemplo de sistema estructurado en forma de árbol es el sistema uucp News (prueba a teclear 'rn' bajo Unix). Es un sistema jerárquico de discusiones (“newsgroups”), en el que cada grupo contiene artículos escritos por muchas personas diferentes. Representa un método muy útil para reunir a los expertos, pero sufre de la falta de flexibilidad de un árbol. Típicamente, una discusión bajo un *newsgroup* desarrollará un tema, de modo que en un momento determinado deberíamos estar en un lugar diferente del árbol. (Ver la *Figura 1*).

```

From mcvax!uunet!pyrdc!pynrj!rutgers!bellcore!geppetto!duncan Thu Mar...
Article 93 of alt.hypertext:
Path: cernvax!mcvax!uunet!pyrdc!pynrj!rutgers!bellcore!geppetto!duncan
>From: duncan@geppetto.ctt.bellcore.com (Scott Duncan)
Newsgroups: alt.hypertext
Subject: Re: Threat to free information networks
Message-ID: <14646@bellcore.bellcore.com>
Date: 10 Mar 89 21:00:44 GMT
References: <1784.2416BB47@isishq.FIDONET.ORG> <3437@uhccux.uhcc...
Sender: news@bellcore.bellcore.com
Reply-To: duncan@ctt.bellcore.com (Scott Duncan)
Organization: Computer Technology Transfer, Bellcore
Lines: 18

Doug Thompson has written what I felt was a thoughtful article on
censorship -- my acceptance or rejection of its points is not
particularly germane to this posting, however.

In reply Greg Lee has somewhat tersely objected.

My question (and reason for this posting) is to ask where we might
logically take this subject for more discussion. Somehow alt.hypertext
does not seem to be the proper place.

Would people feel it appropriate to move to alt.individualism or even
one of the soc groups. I am not so much concerned with the specific
issue of censorship of rec.humor.funny, but the views presented in
Greg's article.

Speaking only for myself, of course, I am...
Scott P. Duncan (duncan@ctt.bellcore.com OR ...!bellcore!ctt!duncan)
(Bellcore, 444 Hoes Lane RRC 1H-210, Piscataway, NJ...)
(201-699-3910 (w) 201-463-3683 (h))

```

Figura 1. Un artículo dentro de la estructura UUCP News.

El campo Subject (Asunto) permite a las notas del mismo tema enlazarse todas dentro de un “newsgroup”. El nombre del grupo (alt.hypertext) es un nombre jerárquico. Esta nota concreta expresa un problema con la estructura de árbol estricta de este esquema: la discusión está relacionada con varias áreas. Obsérvese que los campos “References” (Referencias), “From” (Emisor) y “Subject” (Asunto) pueden ser usados para generar enlaces.

El problema de las palabras clave

Las palabras clave son un método muy utilizado para acceder a datos cuyas coordenadas exactas no se conocen. El problema de las palabras clave, sin embargo, es que dos personas nunca elegirán las mismas. Son realmente útiles, pero sólo si la persona conoce bien la aplicación que está usando. Los sistemas prácticos de palabras clave (como, por ejemplo, VAX/NOTES) necesitan palabras clave para poder grabar registros, lo que significa un paso adelante en la dirección adecuada.

Un sistema de enlaces avanza hacia el siguiente nivel lógico. Las palabras clave pueden ser nodos que representan un concepto. Un nodo palabra clave, entonces, no es diferente de cualquier otro nodo. Se pueden enlazar documentos, etc., a las palabras clave, de tal forma que se encontrarán al localizar cualquier nodo con el que estén relacionadas. De esta forma, los documentos sobre temas similares estarán enlazados, indirectamente, a través de sus conceptos clave.

La búsqueda de una palabra clave se convierte, así, en una tarea que se inicia en una cantidad pequeña de nodos con nombre, y que continúa encontrando los nodos cercanos.

Estas son las razones que me llevaron a construir un pequeño sistema de información enlazada, sin darme cuenta de que el término para esta idea ya había sido acuñado: "hipertexto".

Una solución: Hipertexto

Experiencia personal con el hipertexto

En 1980, escribí un programa para mantener un registro del software con el que estaba relacionado en el sistema de control del PS⁴. Se llamaba *Enquire*, y permitía almacenar fragmentos de información, y de alguna forma, enlazarlos entre sí. Para encontrar la información, se avanzaba a través de enlaces de una a otra hoja, de un modo similar al que usaba el antiguo juego de ordenador “adventure”⁵. Lo usé para mi propio registro de personas y módulos de software. Era similar a la aplicación *Hypercard*⁶, que Apple desarrolló más recientemente para el Macintosh. Una diferencia era que *Enquire*, a pesar de no disponer de gráficos, funcionaba en un sistema multiusuario, lo que permitía a mucha gente acceder a los mismos datos⁷.

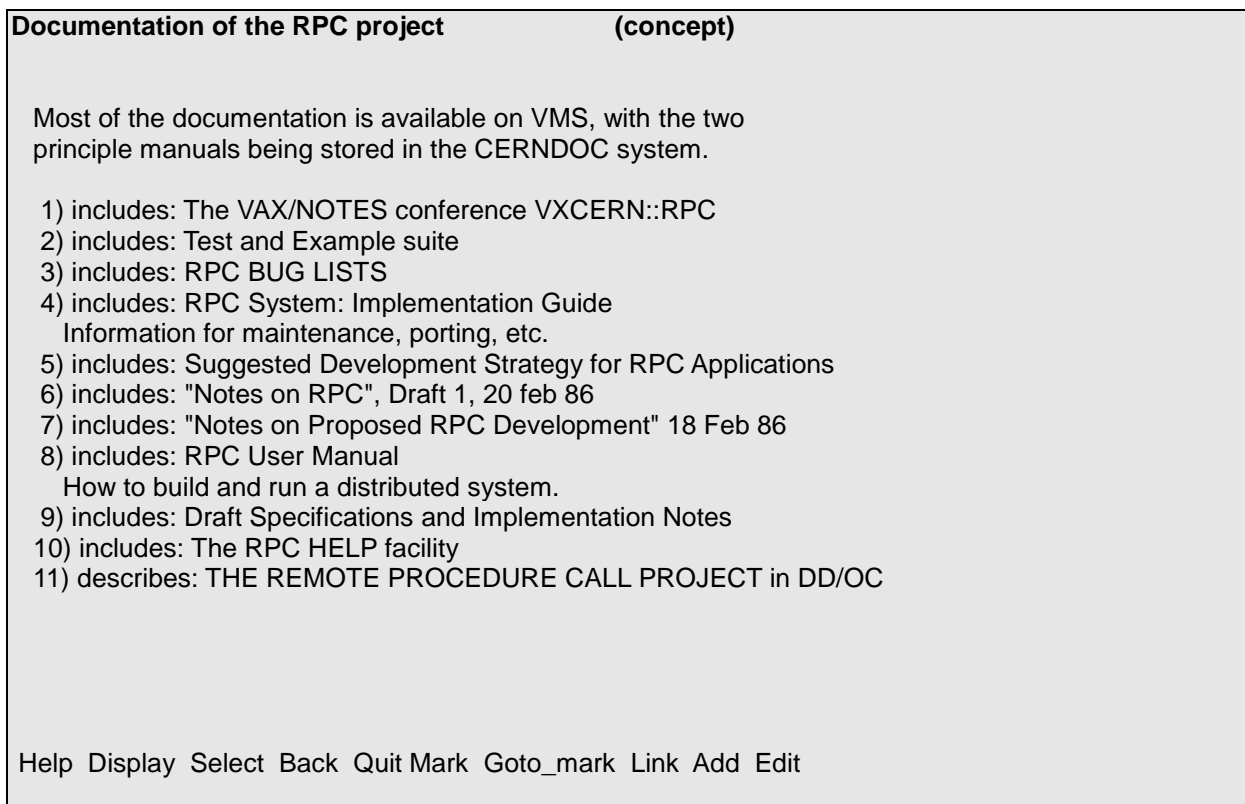


Figura 2. Una pantalla en el sistema *Enquire*.

⁴ Protón Sincrotón del CERN. Se trata de un acelerador de partículas, construido en 1959. [N. del T.]

⁵ *Adventure*, el primer juego de ordenador basado en texto, un tipo de juego que en España fueron conocidos como “aventuras conversacionales”. [N. del T.]

⁶ HyperCard era una aplicación informática, pionera en lo que se refiere a sistemas hipermedia. Lanzada en 1987 por Apple Computer, acabó retirándose en 2004 de forma definitiva. [N. del T.]

⁷ Aunque en el texto no se dice de forma expresa, hay que entender que el acceso a los datos se hacía de forma *simultánea*. [N. del T.]

Este ejemplo es básicamente una lista, de modo que la lista de enlaces es más importante que el texto del nodo mismo. Obsérvese que cada enlace tiene un tipo (“includes”, por ejemplo) y puede también tener un comentario asociado. (La línea inferior es una barra de menús.)

Poco después de mi vuelta al CERN⁸, en la división DD⁹, encontré que el entorno era similar al del PS, así que eché de menos a *Enquire*. Entonces escribí una versión para el VMS, y la usé para mantener un registro de los proyectos, personas, grupos, experimentos, módulos de software y dispositivos de hardware con los que había trabajado. Personalmente, la encontré muy útil. No había hecho un gran esfuerzo para dejarla disponible para su uso público, pero observé que unas pocas personas la habían utilizado exitosamente para investigar los proyectos, y para buscar todo tipo de cosas por decisión propia.

Puntos calientes (Hot spots)

Entretanto, se han desarrollado varios programas que exploran estas ideas, tanto desde un aspecto comercial como desde el académico. La mayoría de ellos, usan “puntos calientes” (hot spots) en los documentos, tales como iconos, o frases realzadas, a modo de áreas sensibles. Si pulsamos un punto caliente con el ratón, aparecerá la información relevante, o se expandirá en pantalla el texto que la incluye. Imaginemos, pues, las referencias en este documento, todas asociadas con la dirección de red a la que hacen referencia, de modo que mientras lees este documento podrías saltar entre ellas pulsando con el ratón.

“Hipertexto” es un término acuñado en la década de 1950 por Ted Nelson [...], que se ha convertido en alguien popular por estos sistemas. Sin embargo, el hipertexto agrupa dos ideas diferentes. Una idea (la relevante para nuestro propósito) es el concepto:

“Hipertexto”: Información, legible por las personas, enlazada de una forma que no tiene restricciones.

La otra idea, que es independiente, y una cuestión de tiempo y tecnología, la constituyen los documentos multimedia, que incluyen gráficos, datos de voz y video. Este concepto no será tratado aquí, aunque usaré la palabra “Hipermedia” cuando quiera hacer referencia a objetos no textuales.

Es difícil valorar el efecto de un gran sistema hipermedia en una organización, a menudo porque estos sistemas nunca se han utilizado seriamente a gran escala. Por esta razón, necesitamos gran cantidad de información existente para que sea accesible al adoptar un nuevo sistema de gestión de información.

⁸ Berners-Lee dejó el CERN en 1980, para trabajar en la empresa Image Computer Systems Ltd., volviendo de nuevo en 1984. [N. del T.]

⁹ DD es el acrónimo de la Data Handling Division (División de Gestión de Datos). [N. del T.]

Requerimientos del CERN

Para adoptar un sistema en el entorno del CERN, tenemos una serie de requerimientos prácticos claros.

Acceso remoto entre redes

El CERN es distribuido, de modo que el acceso desde máquinas remotas es esencial.

Heterogeneidad

Se necesita acceder a los mismos datos desde diferentes sistemas (VM/CMS, Macintosh, VAX/VMS, Unix).

Descentralización

Los sistemas de información empiezan siendo pequeños, pero tienden a crecer. Además, al principio se encuentran aislados, y más adelante se fusionan con otros. Un sistema nuevo debe permitir unirse a los ya existentes, sin requerir ningún control central ni coordinación.

Acceso a los datos existentes

Si proporcionamos acceso a bases de datos existentes como si estuvieran en formato hipertexto, el sistema despegará más rápido. Esta cuestión se discute más adelante.

Enlaces privados

Debe ser posible añadir nuestros propios enlaces a y desde la información pública. Además, debería poderse anotar los enlaces, así como los nodos, de forma personal.

De todo...

El almacenamiento de texto en formato ASCII, y su visualización en pantallas de 24x80 caracteres, a corto plazo es suficiente, y esencial. Añadir gráficos será una tarea extra, que por el momento tendrá mucha menor presencia.

Análisis de datos

Una intrigante posibilidad, dada una gran base de datos hipertextual con enlaces, es que permite algo de análisis automático. Es posible buscar, por ejemplo, anomalías como software no documentado, o departamentos en los que no haya personas. Se puede generar listas de personas o dispositivos para otros propósitos, como listas de correo a las que se podría enviar información cuando se produzca algún cambio.

Es también factible ver la topología de una organización o un proyecto, de modo que se puedan tomar decisiones sobre cómo debería gestionarse, y cuál debería ser su evolución. Resulta particularmente útil en los casos en que la base de datos se hace muy grande, pues algunos grupos de proyectos llegan a estar tan imbricados entre sí, que

es fácil perder la perspectiva, como cuando los árboles impiden ver el bosque.

En un sitio tan complejo como el CERN, no siempre resulta obvio cómo dividir a las personas en grupos. Imaginemos que construimos un gran modelo tridimensional, con las personas representadas por pequeñas esferas, y cadenas entre las personas que tienen algo en común en el trabajo.

Ahora, imaginemos que cogemos la estructura y la sacudimos, hasta que nos hacemos una idea del enredo: quizá verías cómo algunos grupos se forman en algunos puntos, y en otros lugares, áreas débiles de comunicación que sólo agrupan a unas pocas personas. Quizá un sistema de información enlazada nos permitirá ver la estructura real de la organización en la que trabajamos.

Enlaces vivos

Los datos a los que se refiere un enlace (o un *punto caliente*) pueden ser estáticos, o temporales. En muchos casos, en el CERN la información sobre el estado de los sistemas está cambiando continuamente. El hipertexto permite a los documentos enlazar a datos “vivos”, de tal forma que cada vez que se sigue un enlace, se recupera la información actual. Si se sacrifica la portabilidad, es posible hacer que un enlace apunte a una aplicación específica, de modo que un programa de diagnóstico, por ejemplo, podría enlazarse directamente a la guía de mantenimiento.

Lo no necesario

Las discusiones sobre el hipertexto, en ocasiones han abordado el problema del copyright y de la seguridad de los datos. En el CERN, estos puntos son de una importancia secundaria, pues aquí el intercambio de información lo consideramos aún más importante que los secretos. Posiblemente se puedan construir sistemas de autorización y de cuentas¹⁰ para el hipertexto con un diseño muy sofisticado, pero no se proponen en este documento.

En los casos en que la referencia debe hacerse a datos que están efectivamente protegidos, será suficiente con los sistemas de protección de ficheros ya existentes.

¹⁰ Se refiere a las cuentas de usuario en los sistemas informáticos, donde cada persona accede a sus aplicaciones y documentos personales. [N. del T.]

Aplicaciones específicas

Los siguientes son tres ejemplos de lugares específicos en los que el sistema propuesto podría ser útil de forma inmediata. Aunque hay muchos otros.

Desarrollo de la documentación del proyecto

El proyecto *Llamada a Procedimiento Remoto* tiene una descripción de su esqueleto que usa *Enquire*. Aunque limitado, es muy útil para grabar quién hizo qué, dónde está, qué documentos hay, etc. Además, se puede mantener un seguimiento de los usuarios, y añadir algunos datos extra de información disponibles, que no se sabe dónde guardar. Serían muy útiles los enlaces cruzados a otros proyectos, o a bases de datos que contengan información sobre personas y documentos.

Recuperación de documentos

El sistema CERNDOC proporciona los mecanismos para almacenar e imprimir documentos. Un sistema basado en enlaces nos permitiría navegar por los conceptos, documentos, sistemas y autores, permitiendo además referencias entre documentos que se vayan a almacenar. (Una vez que se encuentra un documento, se podría usar el sistema ya existente para mostrarlo en pantalla o imprimirlo).

El "Inventario de habilidades personales"

Las habilidades personales y la experiencia son sólo una parte en la que se necesita la flexibilidad que proporciona el hipertexto. Las personas pueden ser vinculadas a los proyectos en los que han trabajado, y enlazarse a máquinas concretas, lenguajes de programación, etc.

Estado actual de Hipermedia

En la actualidad, se produce una actividad creciente en la investigación sobre hipermedia en las universidades y en los laboratorios de investigación de las empresas, y como resultado de ello, han aparecido algunos sistemas comerciales. Se han celebrado dos congresos, *Hypertext '87* e *Hypertext '88*, y en Washington DC, el National Institute of Standards and Technology (NST, Instituto Nacional de Estándares y Tecnología) mantuvo un taller sobre estandarización en el hipertexto, una investigación de lo que ocurrirá en 1990.

La edición especial sobre el hipertexto de *Communications of the ACM*¹¹, contiene muchas referencias a documentos sobre este tema. Hay una bibliografía sobre hipertexto en [NIST90], y un grupo de noticias (newsgroup) dedicado a este tema, denominado `alt.hypertext`. En cualquier caso, no ofrezco aquí la lista.

Técnicas de exploración

Gran parte de la investigación académica se centra en el aspecto de la interfaz humana de la navegación en un espacio de información complejo. Los problemas que se han apuntado son los que tienen que ver con la facilidad en la navegación¹², y con evitar la sensación de estar “perdidos en el hiperespacio”¹³. Mientras los resultados de la investigación son interesantes, muchos usuarios del CERN acceden al sistema usando terminales primitivas, de modo que los estilos de ventanas avanzadas no son muy importantes para nosotros en este momento.

¿Interconexión o publicación?

Muchos sistemas actuales usan una sola base de datos, a la que acceden muchos usuarios usando un sistema de ficheros distribuido. Hay pocos productos que tengan en cuenta la idea del “docuverso”, de Ted Nelson¹⁴, que permite literalmente enlaces entre nodos de distintas bases de datos. Para poder hacerlo, debería llevarse a cabo algún tipo de estandarización. Sin embargo, en el taller de la estandarización, el

¹¹ *Communications of the ACM* (CACM) es la revista mensual de la Association for Computing Machinery (ACM), que se publica desde 1957, y continúa en activo. [N. del T.]

¹² Aunque Berners-Lee utiliza en todo el texto la expresión *browse* ('ojetear', 'explorar'), en este punto es donde usa por primera vez, literalmente, el término *navigation* ('navegación'). En otras palabras: es a él a quien se debe el uso de este término, y no a Netscape, ni a Microsoft, ni a ninguna otra empresa. [N. del T.]

¹³ Personalmente, que se trata de un juego de palabras en el que Berners-Lee hace referencia a la serie de televisión *Lost in space* (1965-1968), que en España se llamaba *Perdidos en el espacio*. En 1998 se produjo una adaptación cinematográfica, con el mismo título. Y por si no hubiera sido ya suficiente, este mismo año 2018 ha visto estrenar una nueva serie, titulada de igual forma. [N. del T.]

¹⁴ El filósofo y sociólogo Ted Nelson fundó el proyecto *Xanadu* en 1960, cuyo núcleo era "docuverse", un documento global y único, y que, gracias a un gran número de ordenadores conectados entre sí, pretendía cubrir todo lo escrito en el planeta. La información se relacionaría entre sí, en forma de hipertexto. [N. del T.]

énfasis se puso en cuanto a homogeneizar el formato de los medios de intercambio, y no en la red misma. La razón es el fuerte impulso dado a la publicación de información hipermedia, por ejemplo con el disco óptico. Parece haber un consenso general sobre el modelo abstracto de datos que se debería usar en un sistema hipertextual.

Desgraciadamente, muchos sistemas se han montado con muy poca o ninguna visión de su portabilidad. Otros, aunque se han hecho públicos, están basados en software propietario¹⁵ que no se ha diseñado para usar fuera de su entorno. Sin embargo, hay varios proyectos interesantes, y continuamente están apareciendo otros. Compound Document Architecture, de Digital (CDA), por ejemplo, es un modelo de datos que puede hacerse extensible al modelo hipermedia, y hay rumores de que Digital podría estar interesada en ello.

Incentivos y CALS

El Departamento de Defensa de Estados Unidos ha promovido la investigación hipermedia, en efecto, especificando la documentación hipermedia para su consulta futura. Esto significa que todos los manuales de equipos de defensa deben proporcionarse en formato hipermedia. El acrónimo CALS significa Computer-aided Acquisition and Logistic Support¹⁶. Hay también mucha ayuda de la industria editorial, así como de los bibliotecarios, cuyo trabajo es precisamente organizar la información.

¿Qué aspecto tendrá el sistema?

Veamos qué componentes debe tener un sistema de hipertexto en el CERN.

La única forma en la que se puede incorporar una flexibilidad suficiente, es separar el software de almacenamiento, de la información del software que la muestra, y diseñando un buen interfaz entre ambos. Puesto que se requiere acceso de red, es natural que se permita que este interfaz limpio coincida con la división física entre el usuario y el ordenador remoto en el que se encuentra la base de datos¹⁷.

Esta división es también importante en tanto que permite la heterogeneidad requerida en el CERN (y sería una bendición para el mundo en general).

¹⁵ En oposición al software libre. [N. del T.]

¹⁶ Soporte y Adquisición Logística asistidos por ordenador. [N. del T.]

¹⁷ Una división cliente/servidor a este nivel facilita el acceso múltiple, en el que un solo proceso servidor puede dar servicio a muchos clientes, evitando los problemas del acceso simultáneo de multitud de usuarios a una base de datos.

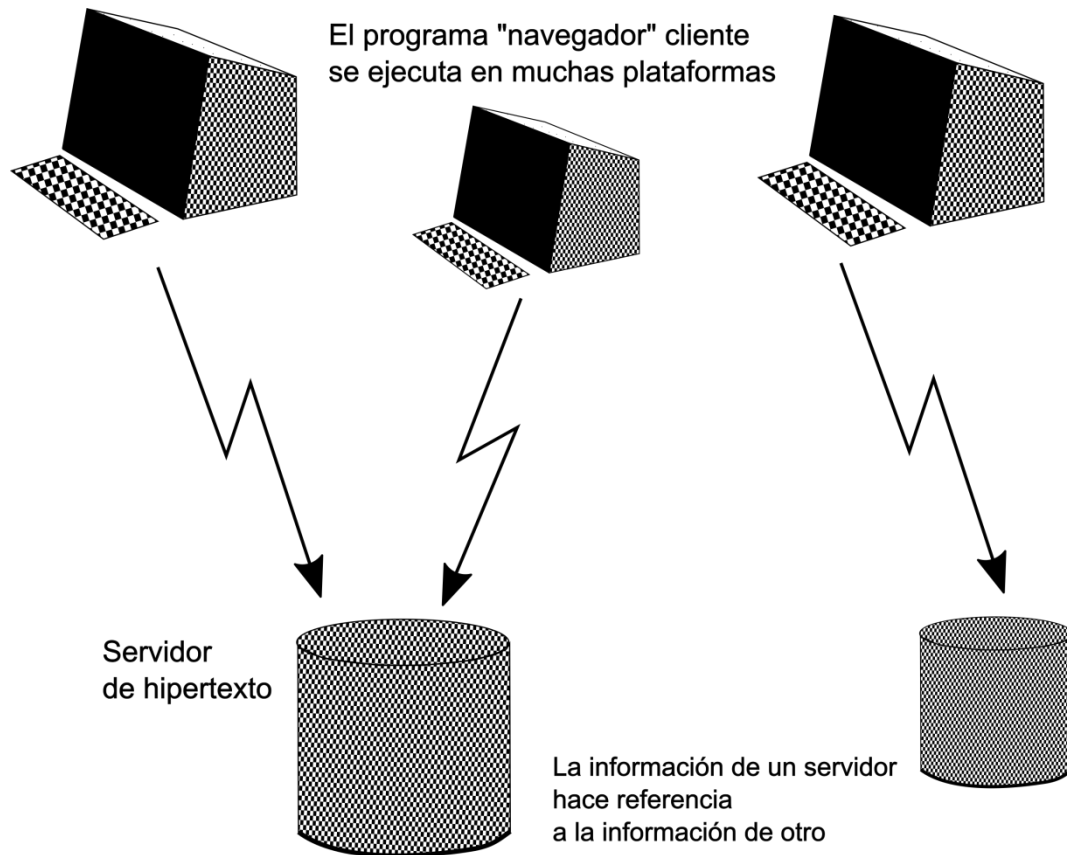


Figura 2. Un modelo cliente/servidor para un sistema de hipertexto distribuido.

Por ello, **una fase importante del diseño del sistema es la definición de su interfaz**. Después de todo, el desarrollo de los distintos formularios y del servidor de datos puede efectuarse en paralelo. El sistema estará bien construido si se pueden mapear muchas fuentes de información diferentes, tanto pasadas, como presentes y futuras, y si se pueden escribir, a lo largo de los años, diferentes interfaces humanas a programas informáticos para obtener ventaja de la nueva tecnología y los estándares.

Acceso a datos existentes

El sistema debe conseguir una utilidad manifiesta desde el principio. Los sistemas de hipertexto existentes han tenido que justificarse a sí mismos basándose sólo en datos nuevos. Si, en cambio, ya había una base de datos del personal, por ejemplo, a la que pudieran enlazarse nuevos datos, el valor de cada componente de los datos sería muy superior.

Lo que se necesita es un programa que haga de pasarela, para mapear una estructura ya existente en el modelo de hipertexto, y que permita un acceso limitado (quizá de sólo lectura). Sería un *servidor de hipertexto*¹⁸, escrito para ofrecer la información existente, de forma que satisfaga un interfaz estándar. La verdad es que no me imagino el servidor generando una base de datos hipertextual a partir de una ya existente; más bien se generaría una vista hipertextual a partir de una base de datos activa.

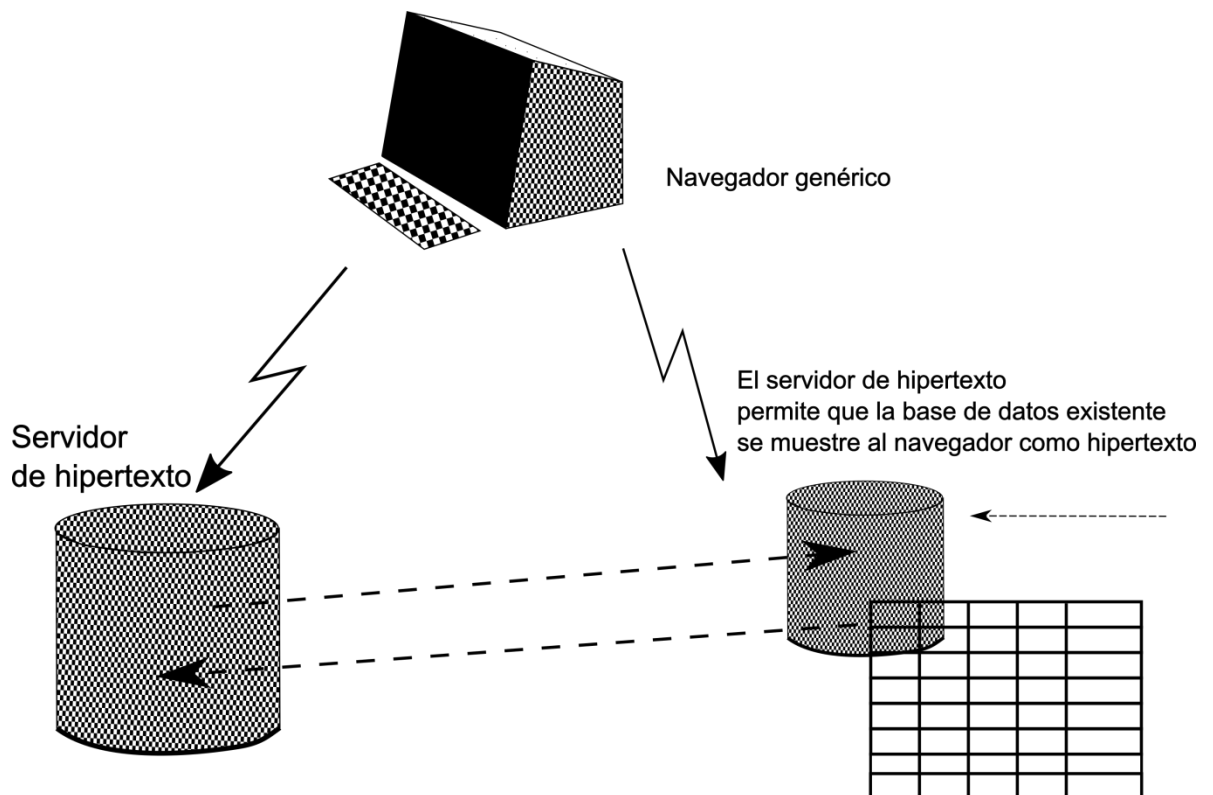


Figura 3. Una pasarela de hipertexto permite que los datos existentes se vean en forma de hipertexto con un navegador hipertextual.

¹⁸ Subrayado del traductor.

Algunos ejemplos de sistemas que se podrían conectar de esta forma son:

- Noticias uucp** Se trata de un sistema de conferencias (charlas) electrónicas, basado en Unix. Un servidor uucp de news podría generar enlaces entre las notas con el mismo asunto, y también mostrar la estructura de las conferencias.
- VAX/Notes** Sistema de conferencias de Digital. Se ha usado ampliamente en FermiLab, pero mucho menos en el CERN. La topología de una conferencia es bastante restringida.
- CERNDOC** Este es un sistema de registro y distribución de documentos, que se ejecuta en una máquina VM del CERN. Al igual que los documentos, categorías y proyectos, las palabras clave y los autores se prestan a ser representados como nodos de hipertexto.
- Sistemas de ficheros** Permitiría que cualquier fichero fuera enlazado desde otros documentos del sistema hipertextual.
- La Agenda telefónica** Incluso la agenda podría ser vista como hipertexto, con enlaces entre las personas y las secciones, las secciones y los grupos, las personas y las plantas de los edificios, etc.
- El Manual unix** Este es un gran cuerpo de texto legible por ordenador, que en la actualidad está organizado de forma plana¹⁹, pero que también contiene información enlazable, aunque en un formato estándar (“Véase también...”).
- Bases de datos** Se podría, quizá, crear una herramienta genérica que permitiera que cualquier base de datos que use un gestor de bases de datos (DBMS) pueda mostrar una vista hipertextual.

En algunos casos, escribir estos servidores significaría tener que desentrañar y obtener los detalles precisos de protocolos y formatos de ficheros existentes. Puede que no sea práctico proporcionar una funcionalidad completa del sistema original mediante hipertexto. En realidad, lo importante será permitir la lectura al público en general: quizá sea un número limitado de personas las que proporcionen la información, y que están satisfechas de usar las facilidades existentes.

A veces, es posible mejorar el sistema de almacenamiento existente codificando la información en forma de hipertexto, si se sabe que un servidor generará una representación hipertextual. En los artículos 'nuevos', por ejemplo, se podría usar (en el texto) un formato estándar para una referencia a otro artículo. La pasarela del sistema de hipertexto lo podría tomar para generar un enlace a dicha nota. Este tipo de mejora permitirá una mayor integración entre el sistema antiguo y el nuevo.

¹⁹ Se refiere a formato de texto plano, por oposición a los ficheros con un formato determinado. [N. del T.]

Siempre habrá un gran número de sistemas de gestión de la información –obtendríamos un montón de utilidad añadida si fuera posible hacer referencias cruzadas (mediante enlaces). Sin embargo, lo perderemos si intentamos restringirlos excluyendo sistemas y obstaculizando la evolución del hipertexto.

Conclusión

Deberíamos trabajar en pos de un sistema de información enlazada universal, en el que la generalidad y la portabilidad sean más importantes que las técnicas de gráficos impactantes y las complejas funcionalidades extra.

El propósito sería permitir un lugar en el que se pudiera encontrar cualquier información o referencia que alguien crea que es importante, y una forma de encontrarla de forma inmediata. El resultado debería ser suficientemente atractivo como para que se quisiera usar; la información crecería pasado un umbral crítico, de tal forma que la propia utilidad del sistema animaría a incrementar su utilización.

El traspaso de este umbral se aceleraría al permitir que las grandes bases de datos ya existentes se enlazaran conjuntamente con las nuevas.

Un proyecto práctico

A continuación indico los pasos prácticos a dar en el CERN para encontrar una solución real. Tras una discusión preliminar sobre los requerimientos ya citados, se necesita una investigación de lo que hay disponible en el mercado. En este punto, buscaremos sistemas que estén diseñados teniendo en cuenta sus futuros usos:

- Portable, o soportado por muchas plataformas,
- Extensible a nuevos formatos de datos

Podemos encontrar que una pequeña adaptación, parte del sistema que necesitamos, se puede combinar a partir de varias fuentes: por ejemplo, un navegador con un origen, y una base de datos con otro.

Pienso que dos personas, durante un período de 6 a 12 meses, serían suficientes para esta fase del proyecto.

Una segunda fase tendría que ver, casi con seguridad, con algo de programación, para poder implantar un sistema real en muchas máquinas del CERN. Una parte importante, discutida más abajo, es la integración de un sistema hipertexto con datos existentes, para poder ofrecer un sistema universal, y lograr una utilidad crítica ya en los primeros estadios del proceso.

(... y ¡sí, esto proporcionaría un excelente proyecto con el que probar nuestras nuevas técnicas de programación orientada a objetos!).

Tim Berners-Lee, Marzo 1989—Mayo 1990

Referencias

- [NEL67] Nelson, T.H. "Getting it out of our system" in *Information Retrieval: A Critical Review*, G. Schechter, ed. Thomson Books, Washington D.C., 1967, 191-210.
- [SMISH88] Smish, J.B and Weiss, S.F,"An Overview of Hypertext",in *Communications of the ACM*, July 1988 Vol 31, No. 7, y otros artículos en el mismo número especial dedicado al "Hypertexto".
- [CAMP88] Campbell, B and Goodman, J,"HAM: a general purpose Hypertext Abstract Machine", en *Communications of the ACM* July 1988 Vol 31, No. 7.
- [ASKCYN88] Akscyn, R.M, McCracken, D and Yoder E.A,"KMS: A distributed hypermedia system for managing knowledge in orignations", en *Communications of the ACM* , July 1988 Vol 31, No. 7.
- [HYP88] *Hypertext on Hypertext*, una versión hipertextual de la edición especial de Comms of the ACM, está disponible para Macintosh y PC.
- [RN] Bajo el sistema operativo Unix, teclea `man rn` para obtener información sobre el comando `rn`, que se usa para leer newsgroup (grupos de noticias).
- [NOTES] Bajo el sistema operativo VMS, teclea `HELP NOTES` para ver información sobre el sistema VAX/NOTES.
- [CERNDOC] Bajo CERNVM, escribe `FIND DOCFIND` para obtener información de cómo acceder a los programas del CERNDOC.
- [NIST90] J. Moline et. al. (ed.) *Proceedings of the Hypertext Standardisation Workshop January 16-18, 1990*, National Institute of Standards and Technology, pub. U.S. Dept. of Commerce.

Traducción, adaptación y notas: Carlos Menéndez García
DATOS A TUTIPLÉN · <https://datosatutiplen.wordpress.com/>
URL del documento:
Primera versión de la traducción: 17/07/2007
Revisión: 07/05/2018
Título original: *Information Management: A Proposal*