

Cómo  
funciona  
la  
Web

# Capítulo 1

## La Web como espacio de información universal

Claudio Gutiérrez

Todo estaría en sus ciegos volúmenes. Todo: la historia minuciosa del porvenir, *Los egipcios* de Esquilo, el número preciso de veces que las aguas del Ganges han reflejado el vuelo de un halcón, el secreto y verdadero nombre de Roma, la enciclopedia que hubiera edificado Novalis, mis sueños y entre-sueños en el alba del catorce de agosto de 1934, la demostración del teorema de Pierre Fermat, los no escritos capítulos de *Edwin Drood*, esos mismos capítulos traducidos al idioma que hablaron los garamantas, las paradojas de Berkeley acerca del tiempo y que no publicó, los libros de hierro de Urizen, las prematuras epifanías de Stephen Dedalus que antes de un ciclo de mil años nada querrían decir, el evangelio gnóstico de Basílides, el cantar que cantaron las sirenas, el catálogo fiel de la Biblioteca, la demostración de la falacia de ese catálogo. Todo, ...

J. L. Borges, *La Biblioteca Total.*

El sueño de la biblioteca infinita se ha hecho realidad: la Web hoy contiene lo que soñó Borges y bastante más. De hecho, se estima que la pieza promedio de información en la Web hoy día nunca será vista más que por su productor y sus amigos cercanos, y uno no puede ver más que un porcentaje minimal de lo que está publicado.

¿Cómo se logró esta fantástica biblioteca infinita? En este breve capítulo revisaremos los fundamentos conceptuales y técnicos que están en la base de la Web, y discutiremos sus alcances y limitaciones.

Es común que los términos *Web*, *Red* e *Internet* se usen intercambiablemente. Desde el punto de vista técnico son objetos completamente diferentes. Internet hace referencia a la red física que conecta diferentes computadores y lugares. Sus preocupaciones son protocolos de transmisión de datos (TCP IP), manejo de nombres de dominio, etc. y que lo tratamos en detalle en el capítulo 3. La Web hace referencia a la arquitectura lógica de la información que ha sido posible construir sobre esa red física. Confundirlos es como confundir el cerebro (una red neuronal) con el conocimiento que posee una persona. Todos tenemos casi el mismo material cerebral, pero los conocimientos y la información que cada uno posee difieren vastamente.

## De fuentes aisladas a redes de información

La evolución del procesamiento de información ha ido desde unidades aisladas hasta una interconexión mundial hoy día a través de la Web.

Probablemente la mejor metáfora sea de nuevo la de una biblioteca. Allí hay información restringida al lugar físico donde funciona. Por un momento olvidemos los catálogos globales (¡productos de la Web también!), y pense-

mos cómo hace 50 años alguien buscaba información. Debía recorrer biblioteca por biblioteca, y correlacionar o comparar la información a mano. Por ejemplo, determinar los títulos de libros que estudian la vida de Andrés Bello. No podía “navegar” a través de la imagen virtual de todos los libros de todas las bibliotecas del mundo juntas. Sin embargo, la Web hizo posible esa realidad.

El desarrollo de la tecnología computacional ha sido clave en este proceso. Los computadores en sus inicios eran gigantescos armatostes que ocupaban pisos enteros de edificios, “centros” de procesamiento de información. La gente, técnicos, usuarios, etc. giraba en torno a ellos. La conexión entre dos de estos gigantescos aparatos era escasa o nula. Con el advenimiento de los computadores personales, llegó también la idea de que cada usuario poseedor de un PC pudiera “conectarse” con otros cercanos. Nacieron las redes locales. De esta idea hay un paso a pensar una red más grande, y finalmente una red “global”. Y con esto, aparece el problema de cómo coordinar, integrar la información que está en cada uno de los nodos (computadores) de esta gigantesca red.

A comienzos de los noventa, Tim Berners-Lee [1] tuvo una idea genial: diseñar este sistema global de información de tal forma que cada usuario en un nodo pudiera navegar por el resto de forma totalmente automática, es decir, sin tener idea de cómo funciona el sistema del otro, qué sistema operativo tiene, qué lenguajes de programación usa, qué aplicaciones corre. Su experiencia en el CERN (ver figura 1.1) fue la gatilladora de esta simple idea, que es el origen de la Web. En palabras de Berners-Lee: “El concepto de la Web integró muchos sistemas de información diferentes, por medio de la formación de un espacio imaginario abstracto en el cual las diferencias entre ellos no existían. La Web tenía que incluir toda la información de cualquier tipo en cualquier sistema.”

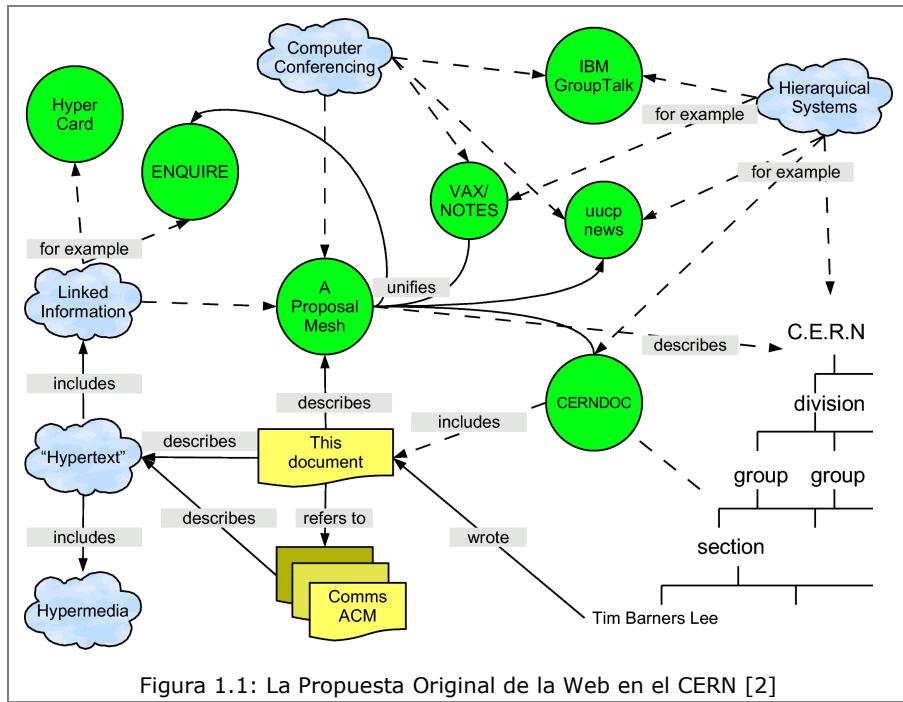


Figura 1.1: La Propuesta Original de la Web en el CERN [2]

Es así como la Web es hoy un gran espacio de información universal, una vitrina de acceso a casi -excluimos la de organizaciones como el Pentágono, etc.- toda la información existente en el mundo hoy en día.

## Las bases lógicas de la Web

Desde el punto de vista técnico, los tres pilares básicos sobre los que se sustenta la arquitectura lógica de la Web son:

1. *Identificadores únicos (URI)*: en un mundo ideal, la suposición básica necesaria para poder referirse (referenciar) y hablar de (describir) todos los objetos, es que éstos tengan su nombre propio, que en términos técnicos se llama *identificador*. En la Web estos nombres propios se llaman *Identificadores Universales de Recursos* (URI por sus siglas inglesas).

Una versión más elemental de URI es la URL (*Localizador universal de recursos*), que corresponde a una *dirección* en la Web. La dirección es una de las formas de identificar un objeto, pero es bueno señalar que la noción de identificador es más amplia que la de dirección, por ejemplo para recursos móviles que no tienen dirección fija.

2. *Lenguaje universal para describir HTML*: Otra suposición básica para la comunicación universal es un lenguaje único, entendible por todos. Tim Berners-Lee diseñó el lenguaje HTML (siglas del inglés *Hyper Text Markup Language*, que a sus características de simplicidad de uso, suma una característica clave: el ser un lenguaje de *hipertexto*, es decir, que tiene un forma de anclar o redirigir al lector desde un punto cualquiera del texto a otro lugar. Estos son los famosos *links* o enlaces en la Web.

3. *Protocolo de transmisión de datos HTTP*: Desde un punto de vista más técnico, uno necesita un protocolo que permita *enviar y traer* información en HTML desde un lugar (sitio) a otro en esta gigantesca red que es la Web.

El protocolo HTTP (sigla del inglés Hyper Text Transfer Protocol) tiene varias características distintivas que lo han hecho muy perdurable. HTTP es un protocolo de transmisión entre clientes y servidores. El cliente, que puede ser un browser, un agente, o cual-

quier herramienta. El servidor es el que almacena o crea recursos como archivos HTML, imágenes, etc. Entre ellos puede haber varios intermediarios, como proxies, gateways y túneles. A través de instrucciones simples, pero poderosas, el cliente indica al servidor qué acciones realizar para recibir o entregar datos. Ver más detalles en capítulo 3.

## La filosofía de la Web como espacio de información: la W3C

La Web fue creada con una cierta filosofía, una posición de principios frente a los desarrollos que se venían dando en materia de publicaciones, de desarrollo de software, de derechos de autor y de difusión. Esta filosofía puede resumirse en tres principios básicos: *todos pueden publicar, todos pueden leer, nadie debe restringir*.

¿Cómo lograr esto técnicamente? En esta dirección, se creó el Consorcio de la Web (W3C), una organización internacional que se propuso como sus dos objetivos primordiales el impulsar la *interoperabilidad* y *evolutividad* de la recientemente creada red universal de información. Para esto se comenzaron a generar estándares y protocolos. ¿Qué significan estos dos requerimientos en más detalle? En un famoso artículo, *Explorando la Universalidad* [3], Tim Berners-Lee desglosaba sus aspectos básicos:

- *Independencia de Dispositivo*. La misma información debe ser accesible desde diversos dispositivos. Esto significa, por ejemplo, que la visualización debe tener estándares que permitan acceder a la información desde casi cualquier formato de pantalla y audio. Una de

las bases para implementar esta *desiderata* es la separación de contenido y forma en la información.

- *Independencia de Software.* Hay muchos y diversos programas de software que se usan. Ninguno debe ser crítico para el funcionamiento de la Web. El desarrollo descentralizado del software ha sido clave para su crecimiento. Además, tema no menor, este postulado previene que la Web misma caiga bajo el control de una comunidad dada o algún gobierno usando el control del software.
- *Internacionalización.* Desde sus inicios, la Web no ha estado cargada a ningún país. Con la introducción de UNICODE, la última barrera que cargaba su desarrollo hacia los lenguajes occidentales ha sido barrida. (La diferencia clave entre el viejo HTML y el nuevo estándar XHTML, aparte de mejoras técnicas relacionadas con XML, es que XHTML está basado en UNICODE.)
- *Multimedia.* Los formatos disponibles para publicar deben estar abiertos a todas las facetas de la creatividad humana capaces de representar. En este sentido, soportar multimedia no representa sólo un par de avances tecnológicos, sino una filosofía de desarrollo de la Web.
- *Accesibilidad.* La gente difiere en múltiples cosas, en particular, en sus capacidades. La universalidad de la Web debe permitir que ella sea usada por la gente independientemente de sus discapacidades. De nuevo aquí la separación de contenido y forma de la información es un pilar básico.
- *Ritmo y razón.* Como dice TBL, la información varía desde un poema hasta una tabla en una base de datos. El balance entre procesamiento automático y humano debe estar presente. Por un lado, por

las cantidades y tipo de información actualmente disponible es impensable que ésta sea procesada sólo por seres humanos: se necesitan agentes automáticos. Por otra parte, es absurdo pensar que en algún momento los humanos serán prescindibles en el desarrollo y enriquecimiento de la Web. Hay que buscar los justos términos para cada aplicación.

- *Calidad.* Las nociones de calidad son subjetivas e históricas. Por ello es impensable que algún día *toda* la información vaya a ser de calidad. Aquí hay otro compromiso, y es que la tecnología de la Web debe permitirnos navegar y vivir entre información con diferentes niveles de calidad.
- *Independencia de escala.* La armonía a gran escala supone armonía en sus componentes. La Web debe soportar grandes y pequeños grupos. Debe permitir que la privacidad de la información de individuos y grupos pueda ser negociada por ellos mismos, y permitir que cada grupo se sienta seguro en el control de su espacio. Hay que lograr un balance entre un gigante monolítico y una diversidad que pueda llevar al aislamiento completo de cada uno.

## La Web Semántica

Uno de los problemas más importantes que aparece con la Web es el de determinar qué “significa” cada dato que está en la Web. Es prácticamente imposible para un usuario chileno entender una página en chino o tailandés. Y viceversa. El problema es aún más dramático: es muy difícil para un humano encontrar la información que necesita. Los buscadores funcionan de manera puramente “sintáctica”, es decir, no “entienden” las palabras. ¿Qué hacer?

Tradicionalmente eso era resuelto por *catalogadores*, personas especializadas que agregaban *metadatos* (etiquetas que explicitan información) a los libros: qué tema trata, dónde está ubicado, cuál es el autor, etc. Estos metadatos están accesibles en un catálogo en las bibliotecas. En la Web, como ya veíamos, no tenemos catálogo, ni menos catalogadores. Con los volúmenes de información que cada día crecen, es imposible que humanos se preocupen de clasificar la información. Además, porque el modelo de la Web es distribuido, quienes publican tienen diversas visiones sobre cómo clasificar sus objetos.

Para los profesionales de la información, el principal desafío hoy es cómo manejar esta extraordinaria cantidad de datos que crece día a día. Estamos comenzando a ver los problemas: los motores de búsqueda a menudo no contestan lo que buscamos; hay dificultades para filtrar la información; la heterogeneidad de los datos y los contenidos; desde el punto de vista de quien publica, se ha convertido en un problema hacer visible la visible, tanto en formato como en contenido. Han habido avances en los niveles estructurales y sintácticos con el estándar XML y sus tecnologías aledañas. Desafortunadamente, al nivel del significado (semántica) aún estamos muy por debajo de las necesidades. Estamos lejos de responder preguntas como “todos los museos que exhiban trabajos de Guayasamín” o “¿Cuál es la biblioteca que tiene la mejor colección de los escritos de Gandhi?” o “¿Cuál es la compañía que ofrece el mejor mapa de Isla de Pascua desde el punto de vista precio/resolución?” Un motor de búsqueda estándar (como Google, Yahoo!, etc.) no puede responder tales consultas. Pero tampoco ningún agente las podría responder hoy en día. Sin embargo, la información está allí: hay que relacionarla y agregarla. La limitación obedece a la falta de habilidad de las máquinas para entender el significado y las relaciones entre las partes de información que recolectan. Hoy en día los humanos agregamos el contexto, interpretamos y damos sentido a la información que existe

en la Web. En otra dirección, otro ejemplo de estas limitaciones es la dificultad para diseñar e implementar una tarea tan natural como organizar todos los recursos educacionales de un país, de tal forma que resulte sencillo para cada estudiante y profesor el publicar y obtener la información que requieran. Se necesitan vocabularios comunes, descripción precisa de los datos expuestos, publicación distribuida, búsquedas automatizadas. En una frase: debido a las enormes dimensiones, la Web se ha convertido en una torre de Babel no sólo al nivel del lenguaje natural, sino esencialmente al nivel del significado, contradiciendo las ideas por las cuales fue creada. ¿La solución? Pavimentar el camino para la construcción de agentes de software que puedan procesar información de la Web por nosotros. La noción de *Web Semántica* [4] es transformar la Web actual de tal forma que la información y los servicios sean entendibles y usables tanto por computadores como por humanos. La Web Semántica creará el ambiente necesario donde los agentes de software puedan rápidamente realizar tareas sofisticadas y ayudar a los humanos a encontrar, entender, integrar, y usar la información en la Web.

## ***Metadatos y RDF***

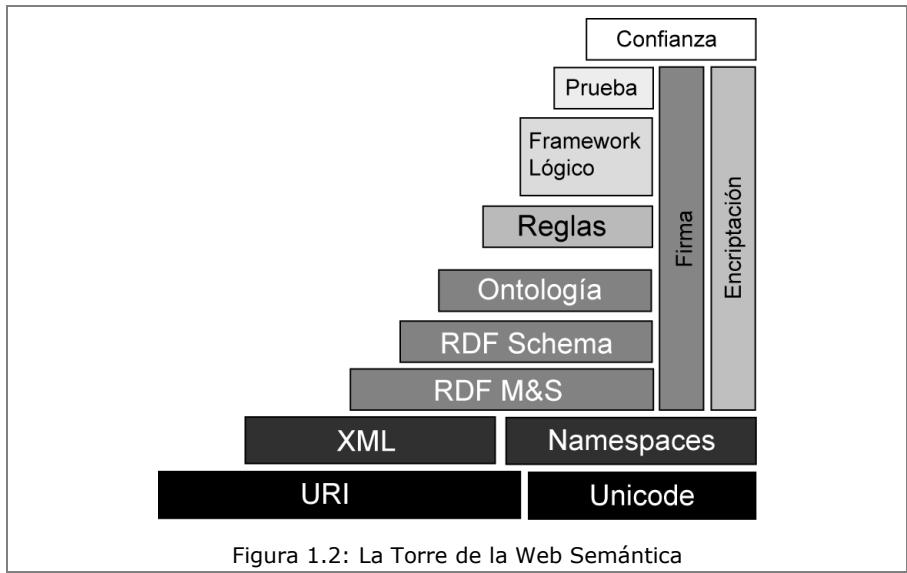
La característica distintiva de la Web Semántica será un *lenguaje estándar de metadatos y ontologías*, que permitirán que agentes de software encuentren el significado de la información en páginas Web, siguiendo enlaces a las definiciones de términos claves y reglas para razonar acerca de ellas lógicamente. Los *metadatos* son datos descriptivos acerca de un objeto o recurso, sea éste físico o electrónico. Las *ontologías* son especificaciones formales de vocabulario y conceptos compartidos para un dominio.

Aunque el concepto de metadatos es relativamente nuevo, los conceptos subyacentes han estado rondando desde que se organizaron grandes colecciones de información. En áreas tales como catalogación en bibliotecas y

museos han sido usados por décadas, por ejemplo, el DCC (Dewey Decimal Classification), OCLC (On Line Computer Library Center), Dublin Core. Una manera útil de pensar acerca de los metadatos es “la suma total de lo que uno puede decir acerca de cualquier objeto de información a cualquier nivel de agregación”. Hay muchos tipos de metadatos, y los usos más comunes se refieren a documentación de *copyrights* y accesos legales, versionamiento, ubicación de información, indización, descripción de condiciones físicas de recursos, documentación de software, autenticación, etc.

En la Web, los metadatos también han jugado un rol importante en áreas como catálogos de propósito general (Dublin Core, Open Directory Project, Wikipedia), sindicación y *rating* (Rich Site Summary RSS, Platform for Internet Content PICS), colecciones personales (música, fotos), privacidad, etc. Y los más populares hoy son simplemente *tags*, es decir, etiquetas; un lenguaje que no tiene verbos ni adjetivos. Simplemente nombres. Todos estos metadatos son sectoriales y usan una diversidad de modelos y lenguajes.

Por el contrario, se necesita un lenguaje de significados (de metadatos) universal. Este es RDF [5] (del inglés *Resource Description Framework*), que es un lenguaje diseñado para soportar la Web Semántica, de la misma manera que HTML es el lenguaje que ayudó a iniciar la Web. El modelo de RDF es simple: el universo a modelar (la Web) es un conjunto de *recursos* (esencialmente todo puede tener una URL); el lenguaje para describirlo es un conjunto de *propiedades* (técnicamente predicados binarios); las descripciones son *oraciones* similares en estructura al modelo sujeto-predicado-objeto, donde el predicado y el objeto son recursos o cadenas de caracteres. Así, por ejemplo, uno puede afirmar “El creador de <http://www.picarte.cl> es Claudio Gutiérrez”. El vocabulario de las *propiedades* para este lenguaje pue-



de ser definido siguiendo las líneas dadas en los esquemas RDF (*RDF Schema*), y básicamente son codificaciones de ontologías a diferentes niveles.

## El Futuro de la Web

No es fácil predecir los desarrollos futuros de la Web. El proyecto inicial de Tim Berners-Lee incluía el desarrollo de capas sucesivas para permitir el intercambio global de información y conocimiento. Luego de la estructura básica que conocemos, vendrá una capa de *semántica*, de metadatos. Esta capa permitiría procesar la información semi-automáticamente, es decir, permitiría a agentes de software procesar la información en paralelo a los humanos. (Nótese que la Web actual está hecha casi en su totalidad para que seres humanos la naveguen.)

La Web por supuesto ha evolucionado en miles de direcciones, muchas no previstas, como redes sociales, blogs, etc. Muchos han llamado al conjunto de estos desarrollos “novedosos” no previstos *Web 2.0*. En los capítulos siguientes trataremos varias de estas facetas.

El futuro está abierto. Hoy en día no es posible predecir los usos futuros de la Web, y aquí ya entramos al campo de la ciencia ficción.

### ***Para saber más***

- ◆ Tim Berners-Lee, *Tejiendo la Red*, Siglo Veintuno Eds., España, 2000.
- ◆ Tim Berners-Lee, Ora Lassila *La Web Semántica*, Scientific American, 2002.
- ◆ La World Wide Web Consortium (W3C) ha dispuesto una breve guía introductoria, en español, sobre la web semántica:  
<http://www.w3c.es/Divulgacion/Guiasbreves/WebSemantica>

### ***Referencias***

1. "CERN: Where the web was born." Page at the CERN.  
<http://public.web.cern.ch/public/en/About/Web-en.html>
2. Tim Berners-Lee. Information Management: A Proposal (1989).  
<http://info.cern.ch/Proposal.html> -  
<http://www.w3.org/History/1989/proposal.html>
3. Tim Berners-Lee. The World Wide Web - Past Present and Future: Exploring Universality. <http://www.w3.org/2002/04/Japan/Lecture.html>
4. W3C Semantic Web Activity: <http://www.w3.org/2001/sw/>
5. Resource Description Framework (RDF) / W3C Semantic Web Activity:  
<http://www.w3.org/RDF/>

# Capítulo 2

# Anatomía de la Web

Ricardo Baeza Yates

## Introducción

¿Qué estructura tiene la telaraña mundial de computadores o World Wide Web? (la Web de ahora en adelante, aunque no me queda claro si es femenino o masculino). Nadie sabe. Crece más rápido que la capacidad de ella misma para detectar sus cambios. Sus conexiones son dinámicas y muchas de ellas quedan obsoletas sin ser nunca actualizadas. El contenido de la Web es hoy de miles de terabytes (un terabyte o Tb es un billón de megabytes) de texto, imágenes, audio y video. Para aprovechar esta gran base de datos no estructurada es importante poder buscar información en ella, adaptándose al crecimiento continuo de la Web.

Al igual que Internet, la red de computadores que interconecta el globo, que ya sobrepasó los 430 millones de computadores conectados en más de 220 países durante 2006, los servidores Web también crecen en forma exponencial desde 1993 (un servidor Web es el software que administra un sitio Web). Lamentablemente nadie sabe su número exacto, pues no es posible a partir de un nombre de dominio saber si es o no un servidor Web (la mayoría comienza con `www`, pero muchos lugares no siguen esta convención). Además un mismo computador puede manejar distintos servidores y también existen servidores virtuales (un mismo servidor Web puede manejar

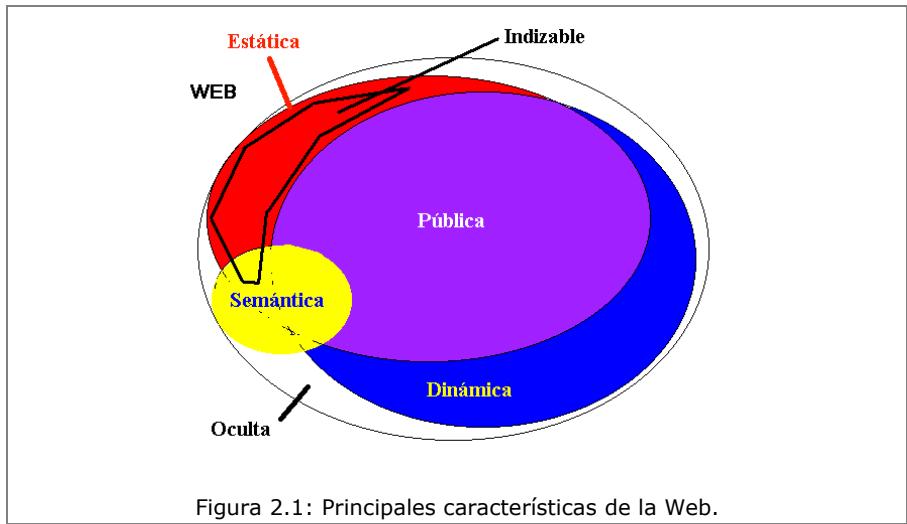


Figura 2.1: Principales características de la Web.

lógicamente otros servidores). En el año 2000, el número de servidores sobrepasó los 10 millones y en mayo de 2007 ya llegaban a los 120 millones.

## Conceptos Básicos

La Web es compleja: hay páginas estáticas y dinámicas, públicas y privadas, con o sin metadatos, que representan la semántica de la Web, tal como se muestra en la Figura 2.1.

Las páginas *estáticas* son aquellas que existen todo el tiempo en un archivo en algún servidor Web. Las páginas *dinámicas* son aquellas que se crean cuando una persona interactúa con un servidor Web, por ejemplo la respuesta a una consulta en un buscador o el resultado de llenar un formulario en un sitio de comercio electrónico. Actualmente, la mayor parte de la Web es dinámica, y como en algunos sitios se puede generar un número no

acotado de páginas dinámicas (por ejemplo, un calendario), la Web que podemos crear es infinita.

Las páginas *públicas* son las que todas las personas pueden ver y las *privadas* son las que están protegidas por una clave o se encuentran dentro de una Intranet. Como cada persona tiene acceso a distintas páginas privadas, la Web pública depende del observador. En particular cada buscador refleja una Web pública distinta. Algunos sitios tienen información semántica que ayuda a los buscadores y se estima que un 5% de ellos tiene información fidedigna. Sin embargo, más son los sitios que tienen información falsa, lo que se llama *spam de Web*.

## **Minería Web**

Para caracterizar la Web debemos realizar un proceso de minería de datos de la Web, lo que en inglés se llama *Web mining*. Una metáfora sería excavar la Web y es posible hacerlo en distintas partes de ella: en su contenido, en su estructura y en su uso. El contenido y la estructura se recolectan con un software que recorre las páginas de la Web y siguen sus enlaces, un programa que en inglés se llama *crawler*. El uso se obtiene de la información que dejan las personas al usar un sitio Web, que se almacena en una bitácora. A continuación detallamos brevemente cada uno de estos casos.

### **Excavando el Contenido**

Lo más simple es recuperar información a través de buscadores como Google o Yahoo!. Pero es posible también usar análisis de lenguaje natural para entender parcialmente la semántica del texto, extraer otros objetos como imágenes o audio, aprovechar las marcas de HTML para transformar el contenido o extraer datos específicos, o mejorar los resultados de los bus-

cadores agrupando páginas similares. Uno de los problemas principales es cómo encontrar las páginas que poseen el contenido que necesitamos, pues sólo encontrar todas las páginas que son indexables ya es difícil (ver figura 2.1).

### ***Desenredando la Estructura***

La estructura de la Web es compleja y evoluciona en el tiempo. Hay desde sectores altamente conectados hasta islas que sólo conocen algunos buscadores. La estructura puede ser usada por los buscadores para jerarquizar los resultados (en base a las páginas más referenciadas usando heurísticas como Pagerank) o para encontrar grupos de páginas que se apuntan entre sí y representan comunidades de personas con intereses similares. El problema principal en este caso es entender el proceso de evolución y su relación con las personas que participan en él.

### ***Analizando el Uso***

Analizar las bitácoras de acceso (*logs*) a un sitio Web es lo más interesante desde el punto de vista comercial. Por ejemplo, una página que nunca es visitada tal vez no tiene razón de ser, o si páginas muy visitadas no están en los primeros niveles, esto sugiere mejorar la organización y navegación del sitio. Por lo tanto, es importante detectar patrones de acceso y sus tendencias. Esta detección puede ser genérica o para un usuario específico (lo que permite personalizar sitios en forma dinámica) y los resultados pueden ser usados para recomendar servicios o productos. El problema principal en este caso es poder diferenciar a los usuarios y cuándo se conectan o desconectan (determinar sesiones).

## *El Principio del Mínimo Esfuerzo*

George Kipling Zipf era un lingüista de Harvard y publicó en 1949 su libro sobre el principio del mínimo esfuerzo un año antes de su deceso, a la prematura edad de 40 años. El descubrimiento inicial de Zipf en 1932 fue que si uno contaba el número de veces que se usaba cada palabra en distintos textos en inglés, y las ordenaba de más frecuente a menos frecuente, se cumplía que la frecuencia de la palabra  $i$ -ésima, multiplicada por  $i$ , era igual a una constante  $C$ , y la constante  $C$  dependía del texto escogido. Actualmente es necesario elevar  $i$  a un exponente  $t$  mayor que 1 y cercano a 2 para muchos textos existentes, en particular en la Web. Graficando esta curva usando una escala logarítmica en ambos ejes, ella se convierte en una recta con pendiente negativa  $t$  [1].

Zipf explica estos resultados empíricos como una condición humana, donde siempre es más fácil escribir una palabra conocida que una menos conocida. Fenómenos similares aparecen en otros ámbitos como el número de citas bibliográficas a un artículo dado o las poblaciones de las ciudades. Diversos autores, entre ellos Mandelbrot y Miller, argumentaron más tarde que en realidad la ley de Zipf representa la consecuencia de las leyes de las probabilidades en procesos asociados a codificación de información donde hay mucho de azar. Sin querer tomar partido en esta disputa científica, cierta o no cierta, la ley de Zipf aparece frecuentemente en la práctica y refleja bien la actitud natural de minimizar el esfuerzo, exceptuando los casos extremos, que serían en el ejemplo inicial, usar muy pocas palabras o usar muchas. Tal vez esta ley sólo explica la diversidad humana, la que se inclina más por la pereza que por la erudición. De hecho, que  $t$  sea ahora alrededor de 1.8 para textos en inglés, indica un mayor sesgo en esa diversidad, y una degradación en el tiempo de la riqueza del vocabulario que usamos al escribir.

## *La Web como un Proceso Humano*

La Web es el producto del trabajo colaborativo de millones de personas. Si hay algún fenómeno donde el principio del mínimo esfuerzo aparecería si existiera, es la Web. Aparte de la distribución de palabras en la Web, las siguientes medidas siguen una curva de Zipf:

- Tamaños de las páginas o de otros tipos de archivos (imágenes, audio, etc.). En este caso la ley no se ajusta bien al comienzo, porque hacer páginas con muy poco texto produce el pudor de la vergüenza que contrarresta al mínimo esfuerzo.
- Número de enlaces que salen de una página. En este caso la curva no se ajusta muy bien en los extremos, porque hacer una página con muy pocos enlaces cae en el caso del punto anterior y, por otra parte, hay páginas con muchos enlaces producidas en forma automática.
- Número de enlaces que llegan a una página. La mayoría de las páginas tienen sólo un enlace a ellas y hay pocas páginas con muchos enlaces.
- Fecha de actualización de las páginas, existen más páginas nuevas o modificadas que viejas.
- Número de componentes conexos de distinto tamaño. Es decir, grupos de páginas en las que se puede navegar de cualquier página a otra página. Esto representa en cierta medida el número de páginas de un sitio Web: muchos sitios tienen pocas páginas, pocos sitios muchas páginas.
- Uso de las palabras en las consultas a un buscador (confirmado experimentalmente en TodoCL.cl). El resultado es que la mayoría de las preguntas son muy simples.

Lo anterior se propaga a otras medidas, como tráfico en la red, uso de proxies, etc. ¿Es todo esto una casualidad producto del azar o un fenómeno del comportamiento humano?. Sin duda la respuesta es que esta ley es resultado del proceso humano de creación de la Web.

## Caracterizando la Web

### *Estructura y Visibilidad*

¿Cuántas referencias tiene una página HTML? (HTML es un acrónimo para Hyper Text Markup Language; el lenguaje usado para estructurar páginas Web). Más del 75% de las páginas tiene al menos una referencia, y en promedio cada una tiene entre 5 y 15 referencias. La mayoría de estas referencias son a páginas en el mismo servidor. De hecho, la conectividad entre sitios distintos no es muy buena. En particular, la mayoría de las páginas no son referenciadas por nadie y las que sí son referenciadas, lo son por páginas en el mismo servidor.

Considerando sólo referencias externas (entre sitios distintos), más del 80% de las páginas tienen menos de 10 referencias a ella. Otros sitios son muy populares, teniendo decenas de miles de referencias a ellos. Si contamos sitios que referencian a sitios, aparecen ODP ([www.dmoz.org](http://www.dmoz.org)), el directorio abierto, y el directorio de Yahoo! en los dos primeros lugares. Estos sitios son los que conectan la Web. Por otro lado, hay algunos sitios que no son referenciados por nadie (están porque fueron incluidos mediante el envío directo de una dirección Web a Yahoo! u otros buscadores, pero que realmente son islas dentro de la Web). En este mismo sentido, las páginas personales también se pueden considerar como entes aislados en muchos casos. Asimismo, la mayoría de los sitios (80%) no tiene ninguna referencia

hacia páginas en otros servidores. Esto significa que una minoría de los servidores mantiene toda la carga navegacional de la red. Estadísticas recientes indican que el 1% de los servidores contienen aproximadamente el 50% del volumen de datos de la Web, que se estimaba mayor a 20,000 millones de páginas durante 2006.

### *Tamaños y características*

¿Cómo es una página Web promedio? Una página de HTML promedio tiene alrededor de 5 a 7 kilobytes (alrededor de mil palabras). Si agregamos audio o video, este promedio aumenta. De hecho, la distribución de tamaños sigue una distribución de Zipf. En otras palabras, aunque la mayoría de los archivos son pequeños, existe un número no despreciable de archivos grandes; y hasta 50 kilobytes predomina el volumen de las imágenes. Desde allí hasta 300 kilobytes son importantes los archivos de audio. Más allá de este límite, llegando a varias decenas de megabytes, tenemos archivos de video. Los formatos más populares (en base a la extensión del nombre de archivo) son HTML, GIF, TXT, PDF, PS y JPG, entre otros.

¿Cómo es una página HTML? Alrededor de la mitad de ellas no tiene ninguna imagen. Un 30% no tiene más de dos imágenes y su tamaño promedio es de 14Kb. Por otra parte hay un porcentaje no despreciable (mayor al 10%) de páginas con más de 10 imágenes. La razón es que son imágenes tipográficas, como por ejemplo puntos rojos, líneas de separación de color, etc. La mayoría de las páginas usan HTML simple. Sólo un porcentaje pequeño sigue todas las normas y otro porcentaje mayor (alrededor del 10%) es sólo texto. Finalmente, la calidad del texto deja mucho que desear, pues hay errores de tipeo, errores que viene de la conversión de imágenes de documentos a texto, etc. Más aún, la información contenida puede estar obsoleta,

puede ser falsa o engañosa. Hay que tener esto en mente cuando usamos una página Web como fuente de información o la referenciamos.

## ***Los Sitios Impenetrables***

Estos sitios son aquellos que contienen una o más páginas donde un buscador no puede extraer los enlaces a las páginas internas porque no usan HTML sino un diseño gráfico basado en un programa. Es decir, la estética es prioritaria pero por ignorancia mata su contenido. Según el último estudio de la Web Chilena [4], estos son el 21% de los sitios, es decir más de 25 mil sitios. Esto incluye sitios que usan Flash en su portada, otros que son o hacen una llamada a un programa y unos pocos que usan mapas de imágenes anticuados. Muchos de estos sitios tienen una portada impenetrable de más de 100Kbs de código, sin contar imágenes, así que además son poco visibles, pues en un módem normal tardarían al menos 30 segundos en cargarse.

Uno puede perdonar que la mayoría de las empresas chilenas no sepan que Flash o Javascript mal usado convierte sus sitios en bóvedas de seguridad. Sin embargo, hay casos en que esto es imperdonable:

- La empresas de tecnologías de la información no pueden apelar a la excusa de ser ignorantes.
- Los sitios de gobierno deben ser los más públicos, visibles<sup>1</sup> y fáciles de encontrar de la Web.
- Las empresas donde la información es uno de sus valores fundamentales. ¡Y sin mencionar las empresas de este tipo que no tienen sitio Web!

---

<sup>1</sup> Por ejemplo, cuando su portada hace difícil la navegación.

## ***La Web como un Grafo***

Imaginemos que por cada persona que conocemos existe una conexión directa entre ella y sus amigos. Por ejemplo, un número telefónico. Si hacemos esto para todas las personas del mundo, tenemos un *grafo* (como los de la Figura 2.2) muy grande. En ese grafo podemos ahora medir “distancias” entre dos personas usando el número mínimo de llamadas telefónicas que necesita una persona para contactar con otra. Por ejemplo, si la persona que quiero contactar está en China es posible que si yo conozco una persona que conoce a una persona en China, el número de llamadas sea pequeño (en el mejor caso, sólo tres llamadas). La distancia máxima entre dos personas se llama el *diámetro del grafo*, usando una analogía geométrica. A mediados de los sesenta, Milgram realizó un famoso experimento utilizando paquetes de correo y estimó que el diámetro dentro de Estados Unidos era 6.

Para que un grafo tenga un diámetro pequeño debe tener muchas conexiones. Si todas las conexiones existen, el diámetro es 1. Por otra parte, un grafo aleatorio tiene un diámetro mucho mayor. Un modelo de grafo que representa bien este fenómeno es aquel en el que cada persona está conectada con todas las personas cercanas (geográficamente) y sólo con algunas lejanas de manera aleatoria y con una distribución de probabilidad uniforme. Este modelo se llama *small-world* o mundo pequeño, valga la redundancia, y también representa bien la red neuronal de un gusano y la red eléctrica del oeste de Estados Unidos, entre otros casos [2].

A finales de los 90, Albert, Jeong y Barabási midieron la distancia (número mínimo de enlaces para llegar de una página a otras) entre 330 mil páginas de la Web [5]. Con esto aproxiaron el diámetro con una función logarítmica en el número de páginas. Al extrapolar esta función, considerando que el número de páginas Web es de más de mil millones de páginas, obtu-

vieron que el diámetro de la Web es aproximadamente 19. Es decir, con 19 clicks del ratón llegamos a cualquier página Web del planeta. Ellos y otros autores sugieren que un buscador podría aprovechar esto para encontrar rápidamente la página deseada. Sin embargo, esto significa saber qué enlace seguir, un problema que no es trivial.

Aunque el modelo de mundo pequeño podría ser válido en la Web, este modelo no explica cómo una persona que sólo tiene conocimiento local puede saber a quién contactar para encontrar a otra persona. Recientemente, Kleinberg [6] ha modificado el modelo original, de tal modo que las conexiones lejanas no siguen una distribución uniforme, sino que una que es inversamente proporcional al cuadrado de la distancia. Esta distribución es óptima en el sentido que minimiza el número promedio de llamadas que haría una persona para contactar a otra, y explica lo que ocurre en la práctica.

La Web es más que un simple conjunto de documentos en distintos servidores, ya que existen relaciones de información entre los documentos mediante los enlaces que establecen entre ellos. Esto presenta muchas ventajas, tanto para los usuarios, a la hora de buscar información, como para los programas que recorren la Web a la hora de buscar contenido para recolectar (probablemente para un motor de búsqueda). Debido a esto se plantea la Web como un modelo de grafo dirigido, en el que cada página es un nodo y cada arco representa un enlace entre dos páginas.

En general las páginas enlazan a páginas similares, de modo que es posible reconocer páginas mejores que las demás, es decir, páginas que reciben un número mayor de referencias que lo normal. En base a esto la Web tiene una estructura que se puede clasificar como *red libre de escala*. Dichas redes, al contrario de las redes aleatorias, se caracterizan por una distribución dispareja de enlaces y porque dicha distribución sigue una ley de Zipf. Los nodos altamente enlazados actúan como centros que conectan muchos de los

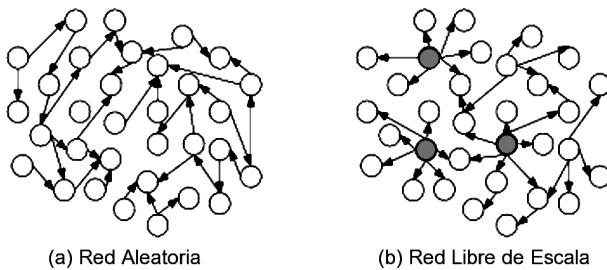


Figura 2.2: Ejemplos ilustrativos de una red aleatoria y una red libre de escala.  
Cada grafo tiene 32 nodos y 32 enlaces.

otros nodos a la red, como se ilustra en la Figura 2.2. Esto quiere decir que la distribución de los enlaces es muy sesgada: unas pocas páginas reciben muchos enlaces mientras que la mayoría recibe muy pocos o incluso ninguno.

## Conejividad

Para conocer qué páginas Web apuntan a una página dada es necesario recorrer toda la Web, algo que los grandes buscadores hacen periódicamente. El primer estudio de la estructura del grafo de la Web fue realizado a partir de dos recorridos de Altavista en Mayo y Octubre de 1999, cada uno de más de 200 millones de páginas (entre un 20% y un 25% de la Web en esa época) y 1.500 millones de enlaces. Sólo almacenar y procesar el grafo equivalente es todo un desafío.

Los resultados de este estudio mostraron que la fracción de páginas de la Web que son apuntadas por  $i$  páginas es proporcional a  $1/i^{2.1}$ , mientras que la fracción de páginas que tienen  $i$  enlaces es proporcional a  $1/i^{2.7}$ . Esto significa que el número de páginas muy apuntadas (populares) y el número

de páginas con muchos enlaces es muy pequeño. Estos valores son casi los mismos para los dos recorridos, pese a que entre ellos pasaron 6 meses.

## ***Estructura***

Para analizar la estructura de la Web se buscan las partes del grafo que están conectadas entre sí. El estudio ya mencionado, y el único realizado a nivel global, muestra que el núcleo o centro de la Web lo constituyan más de 56 millones de páginas, existiendo un camino para ir de cualquier página a otra, con un largo máximo (diámetro) de al menos 28. En otras palabras, el camino más corto entre dos páginas en el peor caso implicaba visitar 28 de ellas. Esto contrasta con el modelo del mundo pequeño mencionado al comienzo que predecía un diámetro máximo de 20 páginas para toda la Web. En la práctica se encontraron caminos hasta de largo 900, lo que indica que el diámetro de la Web es mucho mayor. De todos modos, este número no es tan grande considerando que son cientos de millones de páginas.

La Figura 2.3 muestra la estructura de la Web de acuerdo al estudio mencionado. A la izquierda había 43 millones de páginas desde las cuales se podía llegar al centro, pero no viceversa. Del mismo modo, a la derecha había otras 43 millones que podían ser accedidas desde el centro, pero que no enlazaban páginas del núcleo. Alrededor de estos dos grupos hay tentáculos que contienen 44 millones de páginas y que son caminos sin salida, con la excepción de algunos tubos, que conectan el grupo de la izquierda con el de la derecha. Finalmente, tenemos 17 millones de páginas que están agrupadas en islas que no están conectadas al centro de la Web. Muchos se preguntarán cómo Altavista conocía estas islas si no están conectadas al resto de la Web y no pueden ser recorridas siguiendo enlaces. Es muy simple: estos son sitios Web que fueron directamente enviados al buscador y por lo tanto están en su índice aunque el resto del mundo no las conozca.

Los autores del estudio no hacen ninguna interpretación sobre esta estructura. En las investigaciones que hemos realizado en Chile, que muestran una estructura similar, el grupo de la izquierda son páginas más nuevas que aún no son demasiado conocidas y que si tienen éxito pasarán al centro de la Web, donde están las páginas consolidadas. En cambio, en el grupo de la derecha son páginas antiguas que no enlazan al centro de la Web porque en su época esas páginas no existían, pero sí fueron enlazadas por nuevas páginas. También incluyen muchos sitios Web que no tienen enlaces externos pero que se han preocupado de tener un enlace desde un buen sitio, por ejemplo vía enlaces publicitarios.

En Chile hemos encontrado que la proporción de sitios que son islas es muy alta, mucho mayor que en el estudio original, gracias a que conocemos todos los dominios .cl.

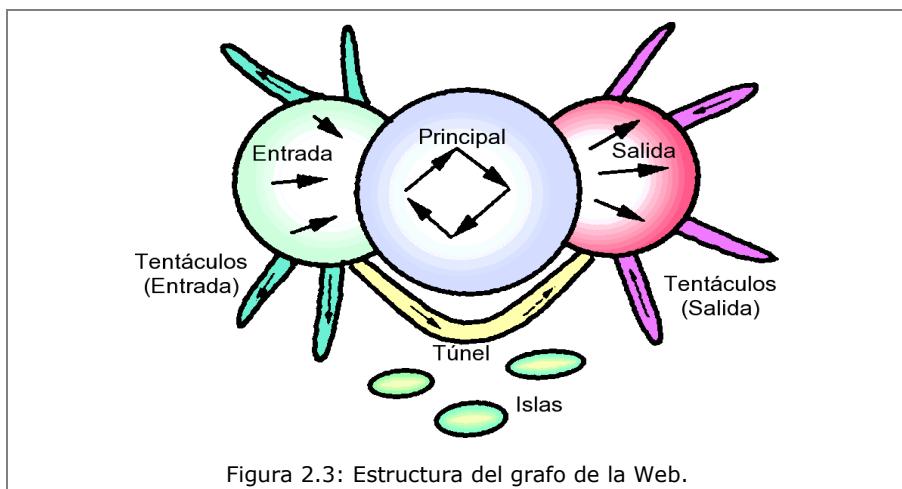


Figura 2.3: Estructura del grafo de la Web.

## *Dinámica de la Web*

Más de la mitad de la Web ha nacido o ha sido actualizada en los últimos seis meses. Parte de ese crecimiento, alrededor de 20%, es replicándose a través de sitios espejos o mirrors u otros tipos de copias (en algunos casos plagio). Al mismo tiempo gran parte de la Web muere. Se estima que el tiempo promedio de vida de una página es alrededor de tres meses. Otra parte de la Web muta, ya sea a través de cambios de nombres de dominio, sitios, directorios o archivos. Es como un organismo caótico, como una colonia de bacterias que está sobrealimentada en algunas partes y en otras agoniza.

La dinámica violenta de la Web y su volatilidad tiene consecuencias importantes. Por ejemplo, sitios Web nuevos serán difíciles de encontrar sin campañas de publicidad, correo electrónico o a través de la comunicación verbal entre personas. Lo mismo para los buscadores. Además, los sitios nuevos tendrán menos sitios que los referencien, con los que son menos importantes para buscadores como Google o Yahoo! que usan los enlaces a un sitio para evaluar su importancia.

Un sitio nuevo generalmente comienza en ISLAS o IN. Luego, si es conocido, pasa al centro de la Web o MAIN. Si luego decide no apuntar a un sitio importante o no es actualizado pasa a la derecha u OUT, o peor aún, se convierte nuevamente en isla. Los componentes más estables en Chile están en MAIN y OUT que tienen el 35% de todos los sitios. En la figura 2.4, mostramos la dinámica de la estructura de la Web Chilena. Los tonos claros indican la procedencia de los sitios antiguos, mientras que los oscuros representan los sitios nuevos. El tamaño del tarro de basura indica la cantidad de sitios que desaparecen.

# Capítulo 3

## Internet

José Miguel Piquer

### El desarrollo de Internet<sup>2</sup>

En las décadas de 1970 y 1980 los computadores se desarrollaban rápidamente mientras iba siendo claro que existía la necesidad de interconectarlos en redes mundiales, básicamente para poder enviar *mail* desde una parte del mundo a cualquier otra; necesidad básica de la comunidad científica que hasta ese momento sólo disponía de un lento y poco confiable sistema de cartas internacionales para intercambiar ideas y trabajos escritos.

Sin embargo, estas redes se desarrollaban en torno a un tipo determinado de computador: existían la redes de computadores IBM (BITNET), Digital (DECNET), Unix (UUCP), etc. En Chile nos conectamos a la red BITNET y a la red UUCP en 1986. Ambas conexiones llegaban a la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, pero BITNET llegaba al

---

2 En el lenguaje coloquial, muchas veces el lego usa intercambiablemente las nociones de “Internet” y “Web”. Desde un punto de vista técnico es necesario diferenciarlas. Una analogía puede ayudar a aclarar la intuición de esta diferencia: el sistema de transporte de pasajeros terrestre está basado en una red de carreteras. Pero el transporte de pasajeros y la red de carreteras son dos cosas completamente diferentes, con problemas diferentes. Lo mismo ocurre para la Web respecto de Internet.

## Capítulo 3 Internet

Centro de Computación (en el segundo piso en Blanco Encalada 2120) y UUCP al Departamento de Ciencias de la Computación (en el primer piso de la misma dirección). Estas redes eran incompatibles entre sí, y no teníamos forma de enviar mails desde la una hacia la otra, por lo que tuvimos por un tiempo un sistema de interconexión que consistía de una persona con un disquette que subía y bajaba las escaleras con el mail de una red hacia la otra.

La necesidad clara de construir un sistema interconectado mundial entre todas estas redes fue uno de los motores fundamentales de Internet. El mismo nombre lo indica: el objetivo era construir una inter-red; una red de redes. Internet conquistó el mundo a través de dos tecnologías clave: el protocolo Internet (IP), que permitía conectar a Internet a cualquier tecnología de red existente; y al sistema de nombres de dominio que permitió tener direcciones de correo electrónico únicas e independientes de la tecnología usada. En 1986, en la Universidad de Chile teníamos varias direcciones de mail, las que ocupaban la casi totalidad de la superficie de nuestras tarjetas de visita. Si el nombre de usuario era jperez, en la tarjeta figuraba la siguiente lista:

```
UUCP: ...!seismo!uchdcc!jperez
BITNET: jperez@uchcecm.BITNET
DECNET: uchvax.DECNET::jperez
X.400: S=jperez; P=uchdcc; A=d400; C=c1;
```

Al comenzar a usar nombres de dominio, la dirección de correo se volvió única (`jperez@dcc.uchile.cl`) y se ha mantenido así por 20 años, a pesar de que la tecnología física de interconexión ha cambiado múltiples veces. Para lograr esto, la Universidad de Chile tuvo que inscribirse como la organización a cargo de administrar el dominio `.CL`, ya que fue la primera en requerir un nombre de este tipo en Chile.

Hoy resulta difícil imaginar la informalidad de esos años, pero todo esto ocurría sin apoyo oficial de ningún tipo, y era simplemente el esfuerzo de un grupo de investigadores motivados tanto en Chile como en el extranjero para que Internet funcionara y se desarrollara.

Durante muchos años el dominio .CL creció muy lentamente (ver figura 3.1-b). Al cabo de 10 años, comenzaron a aparecer las inscripciones masivas de nombres y hubo que crear una organización formal que administrara los nombres (NIC Chile), un sistema de cobros por dominio y un sistema de administración de los conflictos que surgen en torno a un nombre. NIC Chile continúa operando el dominio .cl bajo el alero de la Universidad de Chile hasta el día de hoy.

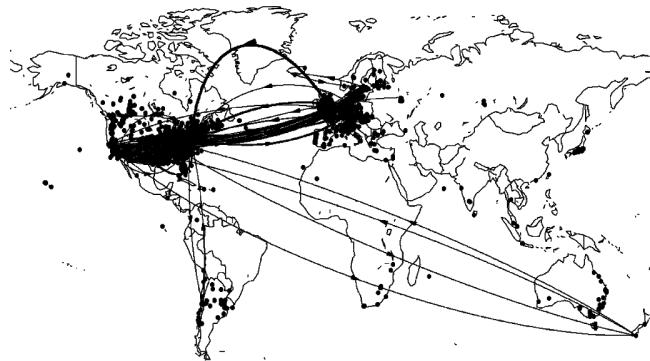
En el mundo, los nombres de dominio han sido uno de los principales puntos de conflicto entre el sector privado, el público y la comunidad internacional. Aunque se ha ido avanzando mucho y se han creado organizaciones con bastante apoyo para administrarlos a nivel mundial, aun persisten muchas discusiones en torno a la operación del sistema, su relación con las marcas y la propiedad intelectual y el rol de los gobiernos en los dominios de país.

## Arquitectura

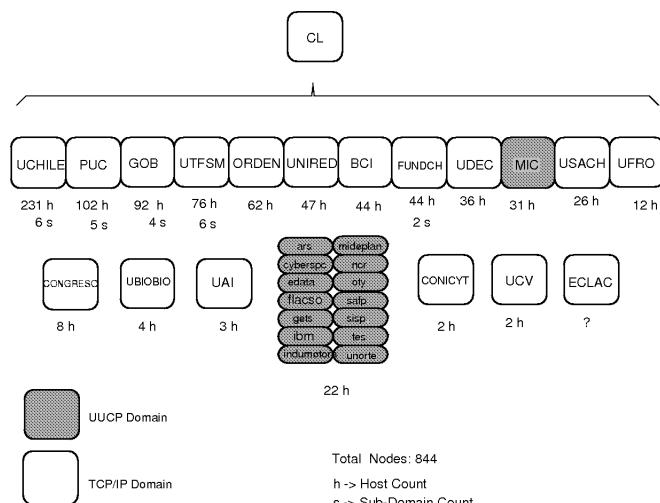
Para que la Web funcione, se requiere de una Internet que provea básicamente la funcionalidad que permita que cualquier computador conectado a Internet pueda conectarse a un servidor identificado por la URL utilizada.

Parte de esa funcionalidad la provee el ISP (Internet Service Provider) y otra parte la provee mi computador y otra el servidor web de destino.

## Capítulo 3 Internet



(a)



(b)

Figura 3.1: (a) Flujo de las News en 1993, (b) nombres inscritos bajo el dominio .cl en 1993. Tomados de [1].

La arquitectura Internet divide esta funcionalidad en cuatro servicios:

1. Traducción de nombre de dominio a dirección IP (DNS)

Este es el servicio inicial que se invoca para traducir un nombre de dominio (como `www.ciw.cl`) a una dirección IP (como `146.83.4.11`), que es básicamente un número único que se requiere para poder llegar al computador destino. Este servicio es crucial para el funcionamiento eficiente de la Web, puesto que todo nombre debe ser traducido antes de poder conectarnos al servidor. La operación requiere de varios servidores de nombres (DNS) que responden por cada dominio, proveiendo redundancia y rapidez en las respuestas.

Este servicio es provisto en parte por el ISP, quien debe proveernos de un servidor de nombres inicial a quien enviarle nuestras consultas, y en parte por servidores por cada dominio. En el ejemplo, hay un grupo de servidores para `.cl` y otro para `ciw.cl`, los que responden con la dirección IP de `www.ciw.cl`.

2. Conexión y Transporte (socket)

Una vez obtenida la dirección IP del servidor establecemos una conexión con él, que permite enviar y recibir datos en forma confiable. Esto se hace a través de un `socket` que es la parte más compleja del sistema porque implementa un protocolo de corrección de errores que permite transmitir información sobre una Internet que pierde datos, que los desordena y a veces incluso los duplica.

La inteligencia del `socket` radica sólo en los extremos de la conexión: el navegador y el servidor. El resto de la red no interviene en este servicio, y eso es fundamental para mantener a Internet como

un servicio barato y eficiente, dado que la complejidad principal la ejecutan los computadores en los extremos. Esto contrasta con la red telefónica que es todo lo contrario: los teléfonos son tontos y toda la inteligencia y complejidad radica en la red misma, lo que la hace mucho más cara.

Este servicio no es provisto por el ISP.

### 3. Ruteo de paquetes IP

El servicio básico que me debe proveer un ISP es el ruteo de los datos que fluyen entre el navegador y el servidor, los que van en paquetes separados los unos de los otros y que deben pasar a través de varias redes potencialmente en países y continentes diferentes.

Este es el servicio fundamental que me provee el ISP.

### 4. Protocolo HTTP

Este es el diálogo que se establece entre el navegador (Internet Explorer, Mozilla Firefox, Opera, etc.) y el servidor web en el otro extremo una vez que están conectados. El protocolo permite intercambiar contenidos de todo tipo, como texto, páginas web, imágenes, audio, video, etc. Toda la web está basada en HTTP.

El protocolo original fue desplegado en Internet en 1991 y rápidamente le cambió la cara a Internet; pasó de terminales de texto a navegadores muy parecidos a los actuales.

En resumen, el navegador envía una URL al servidor, quien le responde con el contenido almacenado para esa URL de manera que el navegador lo interprete y decida qué hacer con éste. El diálogo HTTP termina al terminar esa transferencia.

## El gobierno de Internet

En inglés se habla de “Internet Governance”, que más que un gobierno es una forma de control y supervisión del sistema que nos dé garantías de que esto funcione en forma estable para todos.

En un inicio, cuando Chile se conectó a Internet en 1992, un par de personas controlaban los servicios y asignaban recursos casi sin formalidad alguna. Solicitamos<sup>3</sup> a Jon Postel, quien manejaba los nombres de dominio, que nos asignara la administración de .cl ya que estaba vacante. Nos dio la respuesta positiva rápidamente.

Esto ha cambiado mucho y hoy es muy complejo el tema de la administración y asignación de responsabilidades en Internet. En esto participa la comunidad Internet completa, los gobiernos y los organismos internacionales como las Naciones Unidas. Al ser de alcance global, Internet no debe ser controlada por ningún país en particular, pero la comunidad le teme mucho a una administración burocrática tipo Naciones Unidas.

Por ahora, el organismo que intenta administrar esta discusión y los recursos de Internet es ICANN, que es una fundación sin fines de lucro con residencia en California, Estados Unidos. Su autoridad es bastante cuestionada, pero todos respetan sus procedimientos para garantizar la estabilidad operacional de Internet. A modo de ejemplo, .cl es uno de los pocos dominios de país que tiene un acuerdo marco firmado con ICANN especificando las responsabilidades de cada parte.

Existe una gran batalla de poder en torno a Internet en la actualidad [2]. Algunos opinan que los países deben tomar control sobre sus recursos al ser un servicio básico, los organismos internacionales consideran que deben

---

<sup>3</sup> Jorge Olivos, Patricio Poblete y yo.

## Capítulo 3 Internet

existir leyes globales para regirla y los usuarios sólo queremos que siga funcionando. Afortunadamente, a estas alturas no es fácil tomar acciones locales para ninguno de los actores y se requiere un cierto consenso para llevar a cabo cualquier cambio, lo que da algunas garantías de que el sistema siga operando en forma estable por muchos años más.

### ***Para saber más***

- ◆ Para saber más sobre el gobierno de Internet, visite el sitio de ICANN:  
<http://www.icann.org>
- ◆ NIC Chile (<http://www.nic.cl>) se encarga de administrar los nombres de dominio en Chile.

### ***Referencias***

1. Ricardo Baeza-Yates, José M. Piquer, Patricio V. Poblete. "The Chilean Internet Connection or I Never Promised You a Rose Garden." INET '93.  
<http://www.nic.cl/inet93/paper.html>
2. .CL. Wikipedia the Free Encyclopedia: <http://en.wikipedia.org/wiki/.cl>
3. Internet Governance. Wikipedia the Free Encyclopedia:  
[http://en.wikipedia.org/wiki/Internet\\_governance](http://en.wikipedia.org/wiki/Internet_governance)