Ordnung im Informations-Chaos

Hyper-G als sinnvolle Ergänzung zum World Wide Web

Auch die Humboldt-Universität ist vom World Wide Web - Fieber erfaßt worden: Der Aufbau und die Nutzung von WWW-Diensten haben explosionsartig an den verschiedenen Einrichtungen zugenommen. Von der TU Graz kommt jetzt ein interessanter Diskussionsbeitrag mit der Frage: Ist denn WWW das Informationssystem auch in der Zukunft? Oder wird das WWW auf Grund seiner Design-Schwächen einmal nicht mehr beherrschbar sein? Ist Hyper-G als sinnvoller Ersatz oder als Ergänzung zum World Wide Web nutzbar? Herr Frank Kappe von der TU Graz gibt mit diesem Artikel einen ersten Einblick in die Funktionalität und Kompatibilität zu WWW von Hyper-G.

Frank Kappe
Tu Graz

Hyper-G ist ein verteiltes Hypermedia-System, das speziell für die Wartung größerer Informationsbestände Vorteile gegenüber World Wide Web (WWW) bietet. Ein Vergleich der charakteristischen Strukturen von Hyper-G und WWW zeigt, daß es sinnvoll ist, in vielen Fällen einen Hyper-G Server zu betreiben, auf dem mit WWW-Clients zugegriffen werden kann.

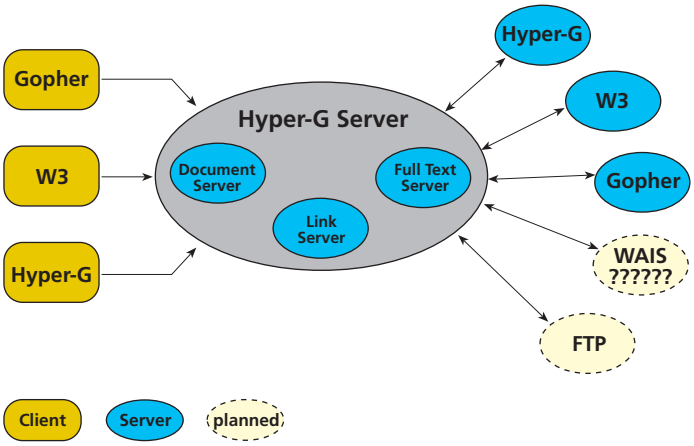
World Wide Web

Das World Wide Web – auch als WWW, W3 oder „Web“ bekannt – wurde 1989 am CERN als Informationssystem für Teilchenphysiker entworfen. WWW ist ein verteiltes Hypermedia-System: Durch Aktivieren von klickbaren Bereichen in Dokumenten („Hyperlinks“) kann zu Dokumenten verschiedenen Typs (Text, Bild, Ton, Video, etc.) verzweigt werden, wobei der Ort der physikalischen Speicherung des Dokuments belanglos ist.

Inzwischen hat sich WWW - hauptsächlich durch einfach zu bedienende Clients wie Mosaic und Netscape - tatsächlich weltweit verbreitet. Leider hat sich der geographische Schwerpunkt der WWW-Entwicklung inzwischen zu amerikanischen Firmen verlagert. Auch der inhaltliche Trend ändert sich: WWW wird inzwischen immer weniger als Informationssystem gesehen, sondern zunehmend als Werbe- und Präsentationssystem, wo die Optik entscheidend ist. Die Zukunft liegt im Transaktionssystem, z.B. als „virtuelles Einkaufszentrum“, in dem nicht nur im elektronischen Versandkatalog geblättert, sondern auch gekauft werden kann; übrigens ebenfalls eine aus Europa importierte Idee (Bildschirmtext, Minitel).

WWW = Informationssystem?

Tatsächlich besitzt das Design von WWW einige Schwächen, die es für Informationssysteme mit mehr als ein paar Dokumenten wenig geeignet erscheinen lassen: Es gibt keine Strukturierungsmöglichkeiten für Information außer den Hyperlinks, was zu Orientierungslosigkeit der Benutzer führt, dem sog. „lost in hyperspace“ Syndrom. WWW selbst besitzt keine Suchmöglichkeit, sondern verläßt sich auf externe Suchmaschinen (z.B. WAIS), wodurch - wenn es der Betreiber überhaupt vorgesehen hat - bei jedem Server die Suche etwas anders funktioniert und eine server-übergreifende Suche unmöglich ist. Dieser Mangel an Suchfunktionen macht ein zielgerichtetes Auffinden von Informationen kaum möglich.



Auf Hyper-G Server kann auch mit WWW- und Gopher-Clients zugegriffen werden. Der Server agiert als Proxy/Cache und greift seinerseits auf andere Hyper-G-, WWW- oder Gopher-Server zu.

Links

Die Links in WWW sind unidirektional, d.h. es gibt keine Möglichkeit festzustellen, welche Links auf ein bestimmtes Dokument zeigen, da die Linkinformation in den Dokumenten gespeichert ist. Das vereinfacht zwar die Programmierung des Servers, hat aber zwei große Nachteile: Erstens ist es damit unmöglich, dem Benutzer dynamisch generierte „Link-Karten“ zu zeigen, die seine augenblickliche Position im Netz der Hyperlinks zeigen. Dies würde das „lost in hyperspace“ Problem verringern. Zweitens kann bei der Modifikation oder beim Löschen eines Dokuments nicht festgestellt werden, welche ande-

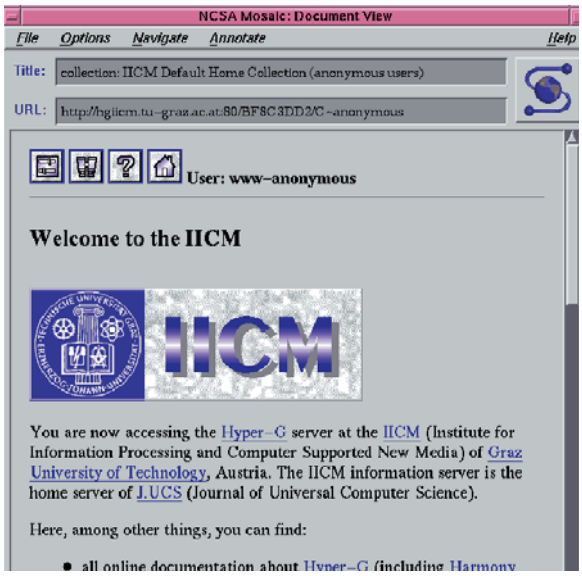
ren Dokumente darauf Bezug nehmen und daher gegebenenfalls modifiziert werden müssen. Dadurch entstehen mit der Zeit in einem großen Informationsangebot immer mehr ungültige Querverweise.

URLs

Das Problem wird noch dadurch verstärkt, daß WWW derzeit zur Bezeichnung von Dokumenten anstelle logischer Namen „Uniform Resource Locators“ (URLs) verwendet, die im Prinzip die physikalische Position des Dokuments angeben. Das ist etwa so, als würde ein Buch nicht mit Autor und Titel, sondern mit seiner Position in einem bestimmten Regal einer bestimmten Bibliothek bezeichnet werden. Der Effekt ist, daß die physikalische Position eines Dokumentes nachträglich kaum mehr geändert werden kann, ohne die Konsistenz zu gefährden.

Server

Normale WWW-Server bieten kaum Komfort zur Verwaltung von größeren Datenbeständen durch mehrere Informationsanbieter (z.B. Schreibrechte), und die normalen WWW-Clients sind nur zum Lesen geeignet. Daher ist es die Regel, daß bei größeren Institutionen (z.B. Universitäten) gleich mehrere Server betrieben werden. Ted Nelson, einer der Hypertext-Pioniere, hat diesen Effekt als „Balkanisierung“ bezeichnet: Es entsteht eine Unzahl von „WWW-Fürstentümern“ – natürlich mit verschiedener Struktur und Benutzerschnittstelle. Die Navigation wird weiter erschwert.



Zugriff auf Hyper-G mit XMosaic
Die Icons links oben im Textbereich erlauben verschiedene Einstellungen (Sprache, Benutzeridentifikation, mehr Info über Documents), Suche, Hilfe, und Navigation zum Haupteinstiegspunkt des Servers.

Die angesprochenen Schwächen sind bei kleinen, von einer einzelnen Person gewarteten WWW-Servern noch zu verkraften, erschweren aber den Betrieb eines größeren Informationssystems erheblich. Wie kann man nun trotzdem ein großes Informationssystem für WWW aufbauen, ohne auf Benutzerkomfort und Datenkonsistenz verzichten zu müssen?

Hyper-G

Eine interessante Möglichkeit besteht in der Verwendung eines Hyper-G Servers für die Datenhaltung. Von außen, aus der Sicht des Endbenutzers, sieht dieser aus wie ein normaler WWW-Server. Man kann mit WWW-Clients wie Mosaic darauf zugreifen, ebenso mit Gopher-Clients und mit „echten“ Hyper-G Clients. Der (die) Informationsanbieter pflegen die Informationsstruktur mit Hyper-G Clients, z.B. Harmony. Wo liegen nun die Unterschiede zum „normalen“ WWW Server?

Collections

Hyper-G verwendet zur zusätzlichen Strukturierung von Information sogenannte Collections, die Dokumente oder andere Collections enthalten, ähnlich wie ein Directory. Die entstehende Collection-Hierarchie dient gleichzeitig drei Zwecken:

- Sie vereinfacht die Navigation.

Der Benutzer kann nun wahlweise hierarchisch, wie in Gopher, oder über Hyperlinks navigieren. Oft erweist sich die hierarchische Organisation der Daten als logischer und leichter durchschaubar. Zudem bieten die Hyper-G Clients „Location Feedback“: Die Position des gerade aktuellen Dokuments innerhalb der Collection-Hierarchie wird angezeigt, unabhängig davon, wie man zu diesem Dokument gelangte, z.B. auch nach einer Suche oder dem Verfolgen eines Links. Dieses Feature hat sich als besonders wertvoll zur Wahrung der Orientierung im System herausgestellt.

- Dokumente in einem Hyper-G Server sind automatisch auch suchbar.

Die Suche kann auf vom Benutzer bestimmte Teilmengen der Collection-Hierarchie eingeschränkt werden. Die Collections können dabei durchaus über mehrere Server verteilt sein.

- Zugriffsrechte können bestimmten Collections zugeordnet werden.

Dadurch kann ein Server problemlos von mehreren Informationsanbietern und Benutzergruppen verwendet werden. (keine Balkanisierung!). Benutzer können persönliche Home-Collections anlegen, in denen sie ihre eigene Sicht der Informationsstruktur aufbauen können.

Harmony - der UNIX Hyper-G Client. Links oben der „Collection Browser“ der immer die aktuelle Position in der Collection-Hierarchie zeigt, links unten die „LocalMap“, die die Linkstruktur rund um das aktuelle Dokument anzeigt (besonders interesant, wenn das aktuelle Dokument modifiziert oder gelöscht werden soll).

Eine spezielle Art von Collections - sogenannte „Cluster“ - binden mehrere Dokumente zu einem logischen Dokument zusammen. Damit können Multimedia-Dokumente erzeugt werden, z.B. ein Ton, der zusammen mit einem Bild abgespielt werden soll, sowie mehrsprachige Dokumente, indem z.B. ein deutscher und ein englischer Text in einen Cluster gruppiert werden. Der Client wählt dann je nach Benutzerpräferenz einen Text aus.

Objekt-Datenbank

Der Hyper-G Server verwendet eine Datenbank zur Verwaltung seiner Objekte (Dokumente, Links, Collections, Benutzer, usw.). Modifikationen der Informationsstruktur sind nur über diese Datenbank möglich, um die Konsistenz zu gewährleisten. Alle Dokumente im Hyper-G Server sind dadurch automatisch suchbar, und zwar einerseits nach Attributen wie Titel (mehrsprachig), Autor, Erzeugungs-, Modifikations- oder Ablaufdatum und beliebigen Schlüsselworten. So kann z.B. nach Dokumenten gesucht werden, die nach einem bestimmten Termin neu eingespielt wurden. Bei Textdokumenten kann nach Inhalt gesucht werden.

Zu den Attributen gehören auch Zugriffsrechte. Hyper-G unterstützt ein hierarchisches System von Benutzergruppen und Zugriffsrechten.

Links

Auch die Links werden nicht in den Dokumenten, sondern in der Datenbank gespeichert. Das hat folgende Vorteile:

- Hyper-G ist ein echtes Hypermediasystem.

D.h. Links können zwischen Multimedia-Dokumenten gesetzt werden – nicht nur vom Text ausgehend wie in WWW. So kann z.B. in einem MPEG-Video eine klickbare Fläche spezifiziert werden, die zu einer bestimmten Stelle in einem Tondokument springt. Das wäre sonst, ohne Änderung des MPEG-Standards, nicht möglich. Besonders für den Dokumenttyp PostScript ist dies wesentlich, denn es ermöglicht das elektronische Publizieren von qualitativ hochwertigem Material als Hypertext. Die elektronische Zeitschrift J.UCS verwendet dieses Verfahren.

- Links sind bidirektional.

Das System weiß, welche Links auf ein bestimmtes Dokument zeigen. Diese Information wird dazu verwendet, dem Benutzer eine Übersicht über die Links rund um das aktuelle Dokument zu geben.

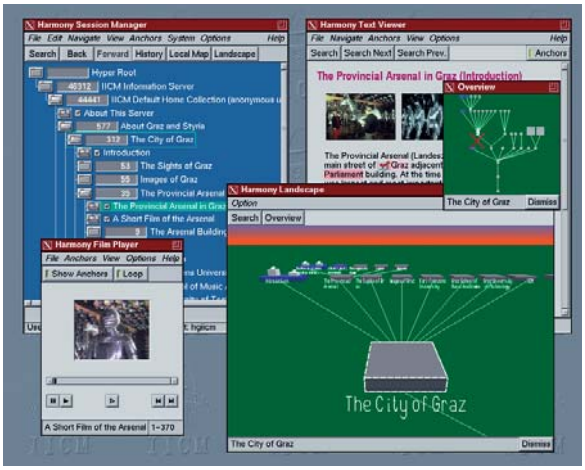
- Die Konsistenz der Linkstruktur wird automatisch gewährleistet.

Beim Löschen eines Dokuments werden auch alle Verweise darauf entfernt.

Hyper-G Clients

Informationen werden - sofern das nicht von Scripts erledigt wird - über Hyper-G Clients interaktiv eingespielt bzw. gewartet. Der UNIX/X11-Client heißt Harmony, der PC-Client heißt Amadeus. Es gibt auch einen Client für VT100-artige Terminals. Ein Power-Mac-Client ist in Vorbereitung.

Harmony ist der komfortabelste Client für Hyper-G. Er unterstützt interaktives Editieren von Links, Dokumenten und der Collection-Hierarchie. Eigene „Viewer“ für SGML-Text, Rasterbilder, Digital Video und Audio, 3D-Szenen und PostScript erlauben nicht nur die Aktivierung von Links in diesen Dokumenttypen, sondern auch die Plazierung von Quell- und Zielbereichen der Links. Für andere



Die dreidimensionale Informationslandschaft verschafft Überblick. Links unten der Harmony Film Player, der auch Hyperlinks im Film unterstützt.

Dokumenttypen können externe Programme gestartet werden. Diese können das Dokument nur anzeigen oder abspielen.

Während aber die Informationsanbieter Harmony oder Amadeus verwenden werden, können die Endbenutzer durchaus Mosaic oder Netscape nutzen, sofern sie Informationen lediglich konsumieren und nicht editieren wollen.

Zukünftige Entwicklungen

Obwohl Hyper-G heute schon ein sehr mächtiges System darstellt, wollen wir in Zukunft in folgende Richtungen intensiv weiterforschen:

- D-Visualisierungstechniken für komplexe Informationsstrukturen

In der 3D-Informationslandschaft kann viel mehr Information angezeigt werden als im 2D-Collection Browser, ohne überladen zu wirken. Zum Beispiel ist es leichter möglich, unter hunderten Dokumenten neu eingespielte zu finden, wenn z.B. die Farbe das Alter des von Dokuments angibt.

- Skalierbare Algorithmen

Skalierbare Algorithmen zur effizienten Verteilung von Update-Informationen zwischen Hyper-G Servern, um die Konsistenz auch über Server-Grenzen hinweg automatisch gewährleisten zu können.

Informationen über Hyper-G

Hyper-G Clients sind für jedermann gratis verfügbar, der Server nur für „Educational Institutions“. Von kommerziellen Anbietern wird erwartet, sich an den Wartungskosten zu beteiligen. In Deutschland spiegeln die folgenden FTP-Server die Hyper-G Software und Dokumentation:
ftp://ftp.ask.uni-karlsruhe.de/pub/infosystems/Hyper-G
ftp://elib.zib-berlin.de/pub/Hyper-G

Im WWW sind Informationen unter http://info.iicm.tu-graz.ac.at/Chyperg. (das ist natürlich ein Hyper-Server).•

TU Graz
Institut for Information Processing and
Computer Supported New Media (IICM)
Frank Kappe
Schießstadtgasse 4a
A-8010 Graz/Austria

Tel.: 0043/316 - 832 551 - 22
Fax:0043/316 - 824 394

E-Mail: fkappe@iicm.tu-graz.ac.at

Der Nachdruck erfolgt mit freundlicher Genehmigung des DFN e. V. aus den DFN-Mitteilungen 3/95.

Benutzungsordnung für das Zusammenwirken der Anwender der DFN-Kommunikationsdienste

Das Datennetz der Humboldt-Universität ermöglicht Wissenschaftlern und Studenten die freizügige Nutzung der im Deutschen Wissenschaftsnetz WiN und weltweit im Internet angebotenen Kommunikationsdienste. Als Betreiber des WiN hat der DFN-Verein eine “Benutzungsordnung für das Zusammenwirken der Anwender der DFN-Kommunikationsdienste” erlassen, die nachfolgend, entsprechend Absatz 4 der Ordnung, den Benutzern der Humboldt-Universität zur Kenntnis gegeben wird. Sie können diesen Text auch auf dem WWW-Server des DFN-Vereins unter http://www.dfn.de/dfn/nutzord-inhalt.html nachlesen.

- beschlossen vom Vorstand am 16.05.1994 -

Ziel der Benutzerordnung ist es, die Zusammenarbeit der Anwender untereinander zu regeln. Um dieses Ziel zu erreichen, werden im folgenden eine Reihe von unterstützenden organisatorischen Maßnahmen durch die nutzenden Einrichtungen gefordert und Verhaltensregeln für einen sinnvollen Umgang mit den Netzressourcen und zur Vermeidung mißbräuchlicher Nutzung aufgestellt.

Die Benutzerordnung richtet sich in erster Linie an Personen, die für die Bereitstellung von Kommunikationsdiensten in den Mitgliedseinrichtungen des DFN-Vereins verantwortlich sind. Es wird erwartet, daß jede Einrichtung ihre Endnutzer von dieser Benutzerordnung in Kenntnis setzt. Darüber hinaus wird empfohlen, für die lokal angebotenen Kommunikationsdienste eine eigene Benutzerordnung zu erstellen, die mit den in diesem Dokument aufgestellten Richtlinien in Einklang steht oder auf sie verweist.

Das Einhalten dieser Ordnung liegt im gemeinsamen Interesse aller Beteiligten, da die Verschwendung von Netzressourcen oder deren Mißbrauch zu einer Erhöhung der Nutzungsentgelte und zu Unregelmäßigkeiten bei der Nutzung der Dienste führen könnte.