

Das Hyper-G Projekt

Thomas Dietinger

InterNet das Netz der Netze

Das Internet fand seinen Ursprung vor ungefähr 20 Jahren in einem Forschungsprojekt, namens 'ARPAnet', des amerikanischen Verteidigungsministeriums. Hauptzweck dieses Konzeptes war die Entwicklung eines möglichst ausfallsicheren Kommunikationssystems. Man ging davon aus, daß Teile des Netzwerkes zu jedem Zeitpunkt aussetzen können und trotzdem die Verbindung zwischen den einzelnen Rechnern aufrecht erhalten werden muß. Die Daten wurden dafür vor der Übertragung in Pakete zerstückelt, mit einer Empfängeradresse, der eindeutigen IP-Nummer, versehen und danach über unter Umständen verschiedene Teile des Netzes verschickt. Diesen Mechanismus bezeichnete man später als das Internet-Protokoll.

Um auch das richtige Zusammenfügen der einzelnen Komponenten zu gewährleisten und zu garantieren, daß eventuell verlorengegangene Datenpakete nochmals gesendet werden, kam ein zusätzliches Datensicherungs- und Übertragungssteuerungsprotokoll oder 'Transmission Control Protocol', kurz TCP hinzu. Vor allem unter UNIX wurde TCP/IP später zu einem der populärsten Netzwerkprotokollen.

Besonders Universitäten zeigten sehr bald reges Interesse an dem nationalen Netz, um ihre Daten und damals noch sehr teure Computerressourcen miteinander auszutauschen. Als Anfang der achtziger Jahre lokale Netze weltweit große Verbreitung fanden, bestand rasch der Wunsch auch diese miteinander, über das nunmehrige internationale Netz, zu verbinden.

InterNet - das Netz der Netze - war entstanden !

InterNet - Ein Chaos an Daten

Man schätzt, daß es mittlerweile mehr als 5 Millionen Rechner und mehr als 30 Millionen Benutzer im Internet gibt. Diese können auf Ummengen von Terabyte an Informationen zugreifen. Universitäten stellen wissenschaftliche Dokumente zur Verfügung, vermehrt kommerzielle Organisationen bieten Kundenunterstützung über das Netzwerk an und nicht zuletzt immer mehr Privatpersonen, man denke nur an Public Domain und Shareware Software, beteiligen sich aktiv daran.

Die Art und Weise des Zuganges zu diesen Daten und das Auffinden der gewünschten Informationen ist meist sehr schwierig und unübersichtlich. Es gibt zahlreiche Dienste, wie FTP, Telnet, News, Email, etc. deren Bedienung sich stark unterscheiden. Um diese jedoch auch nutzen zu können muß man zusätzlich die Internet-Adresse des Dienstleisters, in Form einer 32 Bit Nummer oder des eindeutigen Internet-Names, kennen. Um dem Benutzer trotzdem noch einen einigermaßen übersichtlichen Zugang zu gewähren ist eine Art Telefonbuch, nach Themengebieten geordnet, notwendig. Zusätzlich wäre es wünschenswert, wenn die Bedienung nicht nur einfach, sondern auch einheitlich, und im Besten Fall von einem einzigen Programm zu bewältigen wäre.

Diese Ansprüche versucht man mit Informationssysteme, zumindest ansatzweise, zu befriedigen.

Informationssysteme

Informationssysteme haben primär die Aufgabe einen strukturierten Zugang zu den Daten (aller Art, Stichwort 'Multimedia'), über verschiedene Systeme und Plattformen hinweg, anzubieten.

Zusätzlich sollen sie auch die Möglichkeit bieten, Informationen ansprechend zu präsentieren (Werbung).

Eine wesentliche Voraussetzung dafür ist das sogenannte *Client-Server Konzept*.

Hierbei stellt der *Server* die Daten, mithilfe eines möglichst leistungsfähigen Protokolls über das Netzwerk dem *Client*, zur Abfrage, bereit. Die Benutzerschnittstelle wird ausschließlich auf Client-Seite implementiert. Das garantiert einerseits geringe Netzbelastung (das Netzwerk wird nur für den reinen Datenverkehr verwendet), andererseits bietet es die Möglichkeit die Bedienung an die gewohnte Oberfläche und Eigenschaften des Betriebssystems des Abfrage-PCs anzupassen. Damit ist ein

schnelles Einarbeiten und eine einfache Benutzung der neuen Software möglich.

Der Server wiederum benötigt (theoretisch) überhaupt keine Benutzerschnittstelle und kann auf die Umgebung desselben optimiert sein.

Es haben sich in den vergangenen 5 Jahren im Wesentlichen 3 Informationssysteme etabliert, die diese Bezeichnung auch wirklich verdienen:

- Gopher
- WorldWideWeb (WWW, oder W3)
- Hyper-G

Daneben gibt es noch eine Datenbank, deren Zugang nur über eine Suchabfrage möglich ist, namens

- WAIS

Da sie allerdings keine Strukturierung aufweist und keine Verweise auf andersartige Protokolle möglich sind, habe ich sie in der obigen Aufzählung nicht angeführt.

Gopher

Einige Abteilungen der University of Minnesota stellten Informationsdienste für ihre Studenten mittels menügesteuerten Abfrageterminals zur Verfügung. Als sich schließlich immer mehr Institute dieser Idee anschlossen, ging man dazu über, diese in einem universitätsweiten System zusammenzufassen. Um die Verwaltung dennoch dezentral zu organisieren, wurden lediglich Verweise vom Gesamtsystem zu den einzelnen Abteilungssystemen hergestellt.

Daraus entstand ein hierarchisches System, dessen Konzept auch sehr bald im Internet Anwendung fand - 'Gopher' war geboren !

Ein wesentlicher Aspekt von Gopher ist, daß dieses System auch Zugang zu anderen traditionellen Datensystemen anbietet. Zur Navigation durch das Internet-Chaos und als Einstiegspunkt zu anderen Diensten verwendet man allerdings das einheitliche, menügesteuerte Gopher-System.

Zusätzlich zur hierarchischen Navigation bietet Gopher die Möglichkeit nach den Titeln von allen Dokumenten im System zu suchen. Diese Funktionalität wurde später durch sogenannte Suchmaschinen, wie zum Beispiel 'Veronica', auf mehrere Server ausgedehnt. Erreicht wurde dies dadurch, daß diese Suchroboter ständig andere Server nach Informationen 'abgrasten' und diese in einem eigenen Index speicherten. Veronica wurde dabei direkt in den Gopher-Server integriert, der Benutzer mußte deshalb trotz des erweiterten Funktionsangebot nichts Neues hinzulernen und die Clients mußten dafür nicht speziell angepaßt werden.

Weiters ist wichtig zu erwähnen, daß die Gopher -'Datenbank' prinzipiell mehrere Arten von Dokumenttypen (Text, Bild, Ton) direkt unterstützt. Trotzdem fehlte bis vor kurzem ein wesentliches Element im Gopher-Konzept: 'Links' wurden erst im Konzept des nächsten Informationssystems (WWW) vorgesehen und begründeten damit das Sinken der Popularität von Gopher.

Allerdings ist Gopher kein totes System, sondern befindet sich in stetiger Weiterentwicklung (Gopher+, Virtual-Reality in Gopher, ...).

Das World-Wide-Web

Das World-Wide-Web wurde am Kernforschungszentrum 'CERN' entwickelt und verwendete als erster Internet-Dienst das Konzept von 'Hypertext'.

Hypertext besteht aus normalen Text, erweitert um Querverweise, die man zur Navigation durch die Datenwelt nutzen kann ('Linkverfolgung'). Später wurde Hypertext um eingebettete Grafiken ('Inline Images') erweitert und begründete, durch diese zusätzlichen Präsentationsmöglichkeiten, die Popularität von WWW.

Ein bis jetzt ungelöstes Problem ist das "Lost-in-HyperSpace Syndrom": Es passiert sehr leicht, daß man sich in der WWW Welt verirrt, da dem

System keine (hierarchische) Struktur zugrunde liegt. Außerdem bietet das System von sich aus keine Suchmöglichkeiten an. Segen und zugleich auch Fluch von WWW ist das besonders einfache Server-Konzept. Ein reiner WWW-Server verbraucht zwar sehr wenig Systemressourcen und kann deshalb schon auf einem kleinen PC betrieben werden, bietet jedoch auch nicht sehr viel Funktionalität. Dem wird meistens durch Erweiterung um einen nachgeschalteten WAIS-Server, um wenigstens Suchfunktionalität zu erhalten, oder durch direkte Manipulation am Server abgeholfen. Der Nachteil dabei ist allerdings der größere Installationsaufwand, die schwierigere Serververwaltung und außerdem geht dadurch der Vorteil der standardisierten Benutzung (jeder Provider kocht ein anderes Süppchen) für den 'Infonauten' verloren. Zusätzlich dazu ist das WWW-Protokoll nicht sehr effizient ('verbindungsloses' Protokoll: vor dem Holen eines jeden einzelnen Datenpaketes wird die Verbindung zum Server neu aufgebaut) und erhöht dadurch die Netzauslastung.

Weitere Konzeptschwächen sind das nicht zu Ende gedachte Linkkonzept (Links sind nur in Texten möglich) und das Fehlen einer Benutzerverwaltung um das Einbringen von Daten in das System einer größeren Anzahl von Usern zu erlauben. (Nur dem Systemverwalter ist es normalerweise möglich, den Datenbestand einfach zu erweitern)

Trotzdem hat WWW Gopher an Popularität weit überholt und ist zur Zeit DAS führende Informationssystem. Wesentlich dazu beigetragen haben die hervorragenden Clients, allen voran Mosaic und NetScape. Diese gestatten nicht nur eine sehr einfache Bedienung von WWW, sondern unterstützen auch Gopher und andere einfachere, aber sehr gebräuchliche, Internet-Dienste wie News, Email und FTP.

Hyper-G

Hyper-G wurde in Zusammenarbeit mit dem Joanneum Research Institut IHM (Institut für HyperMedia) und der technischen Universität (IICM - Institut für Informationsverarbeitung und computergestützte neue Medien) als 'second generation information system' entwickelt und soll zwar alle Vorteile der 'first generation' Systeme (Gopher, WWW) beinhalten, nicht aber deren Nachteile.

Wesentliche Aspekte sind:

- echtes verteiltes Konzept (z.B.: Suche über mehrere Systeme, ...)
- standardisierte Suche nach allen Titeln, Attributen und Textinhalten
- Zugriffsmechanismen zur verteilten Bearbeitung der Daten
- Mehrsprachigkeit
- alle Dokumente können nun Links enthalten (nicht nur Text)
- echte Datenbank, Objekte hierarchisch strukturiert
- Bidirektionales Gateway für Gopher und WWW (zusätzliche Kompatibilität mit Gopher und WWW Clients !)

Das Hyper-G Konzept

Die Client-Server Architektur

DER Hyper-G Server besteht in Wirklichkeit aus drei kooperierende Prozessen:

- Link Server: speichert Hyperlinks und Collections
- Volltext Server: verwaltet den Volltextindex und die Volltextsuche
- Dokument Server: speichert und cached die eigentlichen Dokumente

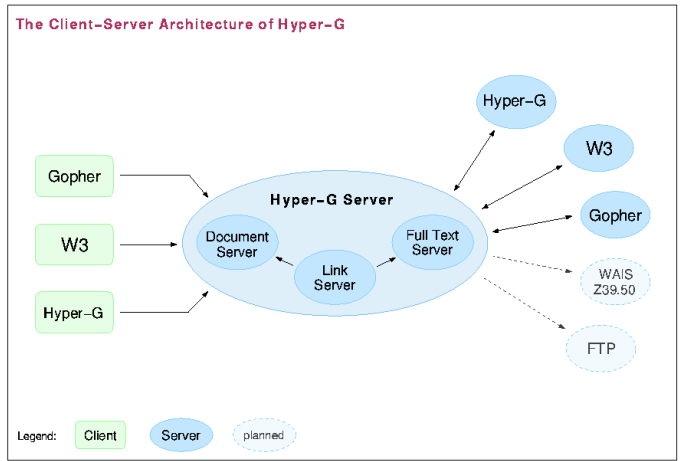


Bild 1: Die Client-Server Architektur von Hyper-G

Hyper-G Clients sprechen immer nur mit einem Server, der auch den Zugriff auf andere (Hyper-G und derzeit WWW und Gopher) Server realisiert (Server/Server-Kommunikation im Gegensatz zur Client/Server-Kommunikation).

Dadurch ist es Hyper-G Clients möglich auf WWW und Gopher (später auch FTP und WAIS) Diensten zuzugreifen, ohne, daß sie deren Protokolle 'sprechen' müssen ('schlanke' aber leistungsfähige Clients).

Der Zugriff von Nicht-Hyper-G Clients (z.B. Mosaic) geschieht über Hyper-G Gateways, die beim Hyper-G Server liegen.

Das Datenmodell

Hyper-G beinhaltet eine objektorientierte Datenbank. Jedes Objekt hat bestimmte Fähigkeiten und Eigenschaften ('Attribute'), die an davon abgeleitete Objekten 'weitervererbt' werden können. Hier eine kurze Übersicht der wichtigsten Objekttypen:

- Dokument: ist die kleinste Informationseinheit, spezielle davon abgeleitete Typen sind:
 - text (zur Zeit nur HTF, das Hyper-G interne Textformat, später auch HTML 3.0)
 - image (GIF, JPEG, TIFF, ...)
 - movie (MPEG)
 - sound (au, WAV)
 - scene (Wavefront), kann zur Zeit nur auf bestimmten Plattformen (z.B.: Windows NT, Silicon-Graphics) abgerufen werden.
 - Postscript (formatiertes Dokument). Dieser Dokumenttyp wird zum Beispiel zum Herausgeben einer elektronischen wissenschaftlichen Zeitschrift (J.UCS) verwendet.
 - remote (Verzweigung zu anderen Internet-Diensten wie WWW, Gopher, FTP, Telnet, ...)
 - generic (*Benutzerdefiniertes* Objekt wie z.B.: Excel Tabelle, PowerPoint Dokument, Midi-Datei, ...)
- Cluster: ist eine logische Gruppe von zusammengehörigen Dokumenten, welche eine Einheit repräsentieren. Ein typisches Beispiel hierfür ist ein Cluster bestehen aus einem Text in englisch und deutsch (mit gleichem Inhalt), einem Tondokument und einem Bild.
- Collection: enthält Dokumente, Clusters oder Sub-collections vergleichbar mit den Verzeichnissen in einem normalen Dateisystem. Allerdings handelt es sich hier um einen 'azyklischen Graphen'. Dies bedeutet, daß eine Sub-Collection in mehreren Collections (z.B. aus Gründen der Übersichtlichkeit) vorkommen kann, ohne daß sie physikalisch mehrfach vorhanden wäre. Technisch gesprochen wird dies dadurch erreicht, daß Beziehungen zwischen Collections durch Verlinkung (ähnlich wie im HyperText) erreicht wird. Diese Art der Hierarchie hilft bei Bewältigung des 'Lost-in-Hyperspace'-Syndroms.

- HyperLinks: sind auch eigene Datenbankobjekte und werden nicht innerhalb eines Dokumentes gespeichert (wie im WWW HyperText), sondern stehen lediglich in Relation zu diesen. Dadurch ergeben sich wesentliche Vorteile:
 - innerhalb JEDEN Dokumentes ist eine Verlinkung möglich (z.B.: Film, PostScript-Dokument, sogar Ton). Da man das zu verlinkende Objekt nicht zu modifizieren braucht, ist es auch möglich in Dokumenten, die sich auf einem schreibgeschützten Teil der Datenbank (z.B.: CD-ROM) befinden, Links einzufügen.
 - Links können *bidirektional* sein (und sind es auch): das bedeutet, daß man zu einem Dokument auch abfragen kann, welche Links zu diesem Dokument zeigen. Dies kann zur Implementierung einer bedeutenden Zusatzhilfe zur Navigation (sogenannte 'local map') im Client verwendet werden. In Harmony, dem UNIX Hyper-G Client, ist dies bereits verfügbar, Amadeus wird dieses Feature bis Herbst beinhalten. Diese 'local map' zeigt die *Umgebung* eines Dokumentes)
 - Ein Dokument auf dem Links zeigen, kann gelöscht werden und durch ein neues mit gleichem Titel ersetzt werden und die Links werden dabei konsistent gehalten ! (Das bedeutet, daß die Links nach dem Dokumentaustausch wieder auf das neue Dokument zeigen.)

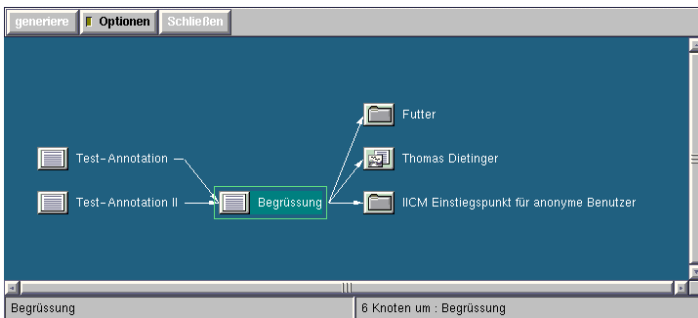


Bild 2: 'Local Map' in Harmony zeigt die Umgebung eines Dokumentes

- Links, die auf bereits gelöschte Dokumente zeigen werden automatisch deaktiviert, dadurch werden 'dangling links' (Links die nirgendwohin zeigen) vermieden.

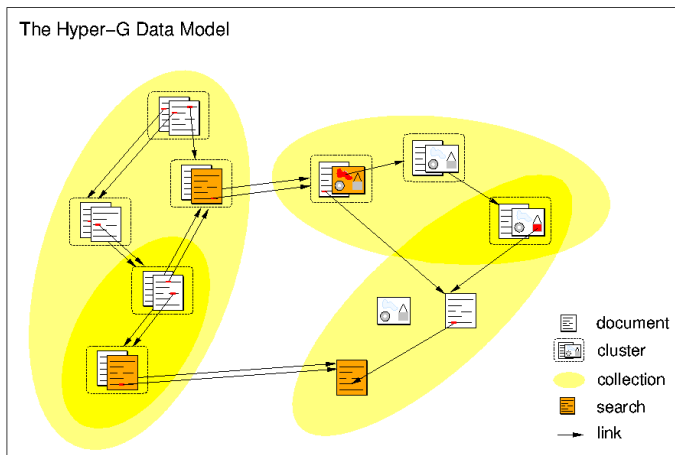


Bild 3: Das Hyper-G Datenmodell

Die Suchfunktionen

Hyper-G stellt den Clients eine Fülle von Suchfunktionen und Selektierungsmechanismen zur Verfügung.

Gesucht werden kann generell nach:

Objekt-Titeln. Erlaubt sind hier die üblichen regulären Ausdrücke.

speziellen Objekt-Eigenschaften wie Autor, Erstellungsdatum, Schlüsselwörter etc.

- nach **jedem** Teil eines Textdokumentes (Volltext-Suche). Dabei werden verschiedene Textpartien unterschiedlich bewertet. So bekommen zum Beispiel Überschriften eine höhere Wertigkeit, als normaler Text. Dies wird in der Sortierreihenfolge des Suchergebnisses zum Ausdruck gebracht (wenn nach den sogenannten 'Score' (Wertigkeit) sortiert wird). Die Suchbegriffsyntax ist von der Funktionalität an der von WAIS angelehnt. Die Dokumente werden dabei nicht zur Laufzeit durchsucht, sondern beim Einfügen der Texte in Hyper-G (also nur einmal) 'indiziert', dies garantiert ein effizientes und schnelles Finden der gesuchten Dokumente.

Zusätzlich zur Art des Suchens, kann auch angegeben werden, wo gesucht werden soll. Dabei unterscheidet man grundsätzlich:

- Suche in der ganzen Home-Datenbank: Es wird im ganzen Hyper-G Server, in dem man eingeloggt ist, gesucht.
- Suche in aktivierten Collections: Mit Hilfe des Clients werden in der hierarchischen Struktur der Datenbank verschiedene Collections (und deren dazugehörigen Unteräste) für die Suchanforderung 'aktiviert' (=ausgewählt).
- Verteilte Suche über verschiedene Server hinweg: Zusätzlich zu lokalen Collections, kann man auch Collections, die auf andere Server verweisen für die Suchanfrage aktivieren. Besonders interessant ist in diesem Hinblick, daß es jederzeit möglich ist, eine Liste **aller** aktiven Hyper-G Servern zu erhalten. Dies gestatten eine einfache Auswahl (durch Selektierung) der Server, in denen gesucht werden soll.

Mehrsprachigkeit

Ein weiterer Aspekt universell einsetzbarer Informationssysteme ist die Unterstützung von Mehrsprachigkeit. So kann in Hyper-G jedem Dokument als Eigenschaft eine bestimmte Sprache, oder eine Reihe von Sprachen ('multi-lingual') zugeordnet werden. Der Benutzer kann nun am Client diejenige Sprache einstellen, in der er am Liebsten seine Dokumente angezeigt haben möchte. Sind sie allerdings in der von ihm vorgewählten Sprache nicht verfügbar, wird die nächstbeste (auch konfigurierbar) vom System angeboten. Zur Zeit werden Englisch, Deutsch, Französisch, Italienisch, Spanisch und *Steirisch(!)* unterstützt. In einer weiteren Ausbaustufe wird vorgeschlagen, daß sich auch das gesamte Benutzerinterface des Clients an die gewählten Sprache anpaßt (zur Zeit nur in Harmony teilweise implementiert).

Benutzerverwaltung und Zugriffsrechte

Hyper-G stellt den Anspruch auch große Datenmengen verwalten zu können. Dafür ist es unbedingt notwendig, daß eine geregelte Möglichkeit besteht, daß mehrere Benutzer neue Dokumente in das Informationssystem einfügen können. Dies wiederum erfordert einen Mechanismus zur Verwaltung von Benutzern und deren Schreib- und Leserechte. Hyper-G bietet überdies sogar die Möglichkeit, mehrere Benutzer zu Gruppen zusammenzufassen und so die Verwaltung zu vereinfachen. Damit aber nicht für jedermann ein eigener Benutzerstatus eingerichtet werden muß, gibt es auch anonymen Benutzer, die allerdings aus Sicherheitsgründen keine Schreibrechte besitzen.

Als weitere Eigenschaft können für jedes Dokument mehrere Schreib-, Lese- und Löschrchte vergeben werden.

Amadeus

Amadeus ist der erste Hyper-G Client für Microsoft Windows. Er zeichnet sich durch folgende Highlights aus:

- echte 32 Bit-Applikation (voll Windows 95 tauglich), läuft unter Windows 3.1x Mithilfe von Win32s
- komfortables Installationsprogramm
- üblicher Windows-Komfort wie:
 - MDI,
 - dockable-Toolbar mit Toolbar-Tips,
 - context-sensitive Popup-Menüs,

- ausgiebige Online-Hilfe etc.
- zahlreiche eingebaute Dokument-Viewer für:
 - Texte (HTF und HTML)
 - Raster-Bilder (GIF, JPEG, TIFF, PCX, BMP, ...)
 - Filme (MPEG)
 - PostScript-Dokumente
 - 3D-Szenen für Virtual Reality (zur Zeit noch ein externer Viewer nur für Windows NT 3.5x)
- eingebauter Konverter von RTF nach HTF. Dadurch kann man sehr leicht mit einem herkömmlichen Textverarbeitungsprogramm einen RTF-Text erstellen und ihn mittels Konverter in die Datenbank einfügen.
- umfangreiches 40-seitiges Handbuch

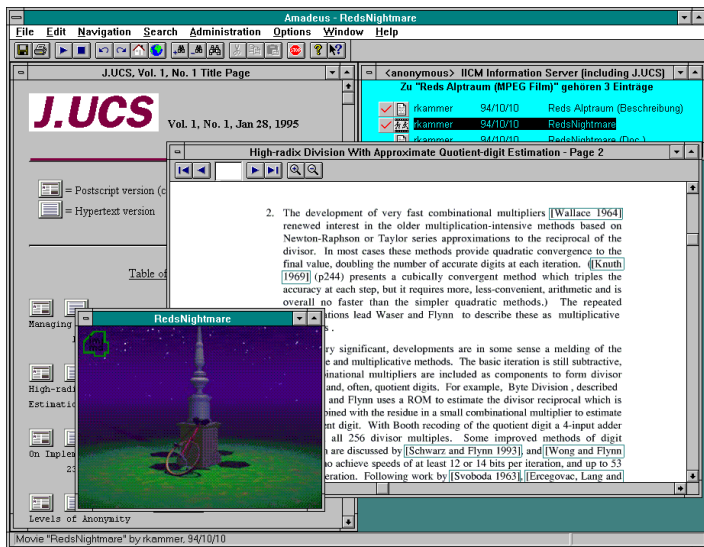


Bild 4: Amadeus - der Windows Hyper-G Client

Ausblick:

Amadeus wird in einer der nächsten Versionen bereits eine lokale Variante des Hyper-G Informationssystems (mit voller Funktionalität wie Suche etc.) zusätzlich eingebunden haben. Dadurch ist es möglich, Teile der Datenbank lokal abzuspeichern und zum Beispiel auf CD-ROM auszuliefern. Ein weiteres Einsatzgebiet wäre der umgekehrte Weg: Erstellen von Teilen der Datenbank Off-Line in Form von Tele-Arbeit und anschließendes Einfügen der Daten ins Informationssystem über Modem etc..

Außerdem wird bis Herbst ein HTML 3.0 fähiger Textviewer implementiert sein, um völlige Kompatibilität mit WWW zu garantieren. Zusätzlich dazu werden weitere Internet-Dienste (ähnlich NetScape) wie FTP, News, Email etc. direkt in Amadeus eingebaut.

Harmony

Harmony ist ein Hyper-G Client für UNIX/X11 und hat gegenüber Amadeus einen Entwicklungsvorsprung von ca. einem Jahr. Aus diesem Grund sind hier bereits wesentlich mehr Hyper-G Konzepte verwirklicht

als in Amadeus. Harmony ist praktisch für fast alle wichtigen UNIX-Plattformen portiert. Seine hervorstechendsten Merkmale sind:

- Multilinguale Bedieneroberfläche
- Baumartige Darstellungsmöglichkeit der hierarchischen Datenbankstruktur
- 3D-Information-Landscape (auf Silicon Graphics): dieser neuartige Zugang zu den Daten läßt die Struktur der Datenbank förmlich 'erleben'. Man fliegt über die einzelnen Objekte, kreist über interessante Collections etc.. Diesem neuen Gebiet wird eine besonderer Forschungsschwerpunkt beigemessen.
- Local Map: diese erlaubt einen grafischen Überblick über die Position eines Objektes in der Datenbank in Relation zu einer gewählten Umgebung. So kann man zum Beispiel graphisch abfragen, welche Links auf dieses Dokument zeigen, und welche von ihm ausgehen. Oder man läßt anzeigen, welche Kinder und welche Eltern dieses Objekt besitzt etc.
- Mitgelieferte linkfähige Viewer:
 - Text (HTF und HTML)
 - Bilder (GIF, JPEG, TIFF)
 - Filme (MPEG)
 - Musik (AU)
 - PostScript Dokumente
 - 3D-Szenen

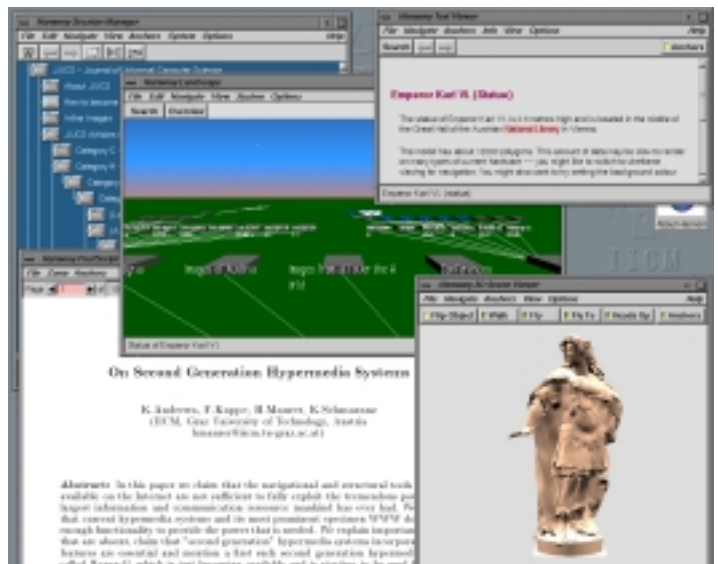


Bild 5: Harmony - der UNIX/X11 Hyper-G Client