

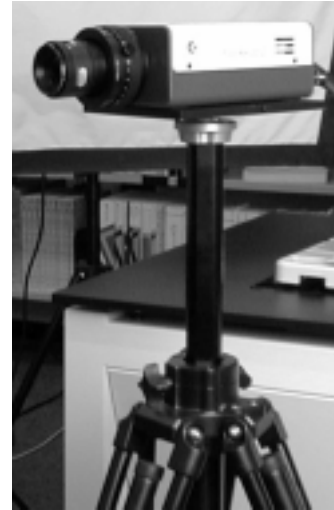
Inhalt

<i>Digitale Photographie</i>	2
<i>FAX-Server</i>	6
<i>Statistiksoftware – ein Überblick</i>	8
<i>TUSTEP</i>	12
<i>Wie wichtig ist ein gutes Paßwort?</i>	13
<i>OCR (Optical Character Recognition)</i>	19
<i>Externe Zugänge</i>	21
<i>Was ist MBone?</i>	24
<i>Universitätsweite Suchmaschine</i>	28
<i>Fujitsu VPP300/2</i>	31

Digitale Photographie

Jürgen Albert und Werner Wegstein

Für Lehre und Forschung an der Universität sind bildliche Darstellungen in allen Fakultäten von zentraler Bedeutung, ob als Abbild einer mittelalterlichen Handschrift oder eines Gemäldes, ob als Aufnahme räumlicher Objekte oder zur Veranschaulichung mathematischer Kurven, Diagramme oder Tabellen.



Standen traditionell am Anfang nur hochauflösende Schwarzweiß-Filme, später auch Farbfilme für die Photographie zur Verfügung, so hat sich mit der rasanten Entwicklung der elektronischen Datenverarbeitung die Digitaltechnik auch längst in die Bildverarbeitung ausgedehnt: in der Feinheit der Bildauflösung und Farbtreue der Abbildung können hochwertige Flachbettscanner heute bereits Archivqualität erreichen, wobei der Vorteil der Digitalisierung nicht zuletzt auch darin zu sehen ist, daß die Bildqualität nach der Speicherung sich nicht mehr verschlechtert - im Gegensatz zu Filmaufnahmen, die herstellungsbedingt selbst bei sachgemäßer Lagerung nach Jahren oder Jahrzehnten dann doch verblassen oder sich farblich verändern. Die Bild-Digitalisierung durch Scannen hat daher schon seit langem in der Universität ihren Einzug gehalten.

Freilich sind nicht alle Objekte in gleicher Weise für die Digitalisierung durch Scannen geeignet: mittelalterliche Handschriften oder neuzeitliche Gemälde können einfach nicht auf einen Flachbettscanner gepresst werden und

dreidimensionale Objekte scheiden ohnehin aus. Hier beginnt der Aufgabenbereich der digitalen Photographie, die ohne photochemische Zwischenstufen unmittelbar ein digitales Bild erzeugt. Museen und Kunstsammlungen, aber auch Mittelalterphilologen scheinen zu den ersten Anwendern zu gehören, die die Fortschritte in der digitalen Bildverarbeitung für konservatorische Zwecke zu nutzen begannen. In internationalen Berichten über die Digitalisierung von Primärquellen (so Peter Robinson, *The Digitization of Primary Textual Sources*, Office for Humanities Communication 4, Oxford 1993) werden verschiedene Kamerasysteme vorgestellt, die für Arbeiten in Archivqualität geeignet sind. Unter diesen nimmt die Kontron ProgRes 3012, was Handhabung und Leistungen angeht, eine führende Rolle ein. Sie wurde deshalb auch für das von der Europäischen Union im Rahmen von ESPRIT II geförderte Projekt „Visual Art System for Archiving and Retrieval of Images“ (VASARI-Projekt) als Basiskomponente ausgewählt. Der auf der Basis der Kontron-Kamera entwickelte VASARI-Scanner wurde für das Folgeprojekt „Methodology for Art Reproduction in Colour“ (MARC) in internatio-

Abbildung 1: Aufnahme mit digitaler Kamera ProgRes 3012



nenalen Kooperationen weiterentwickelt. Der jüngste Bericht von David Saunders (High-quality imaging at the National Gallery: Origins, implementation and applications, in: Computers and the Humanities 31, 1997/98, 153-167) dokumentiert eindrucksvoll Leistungsfähigkeit und Ausbaumöglichkeiten des Systems.

Es ist daher sehr zu begrüßen, dass jetzt im Rechenzentrum der Universität ein Kontron-Kamerasystem angeschafft werden konnte, damit die Lehr- und Forschungseinrichtungen der Universität Erfahrungen in der Handhabung einer hochwertigen digitalen Kamera und in der Erstellung und Bearbeitung digitaler Bilder sammeln können.

Mit dem neuen System werden bereits unter anderem folgende Projekte durchgeführt:

- Digitalisierung medizinisch-biologischer Präparatbilder (Prof. J. Albert, Inst. für Informatik)
- Archivierung von Frühdrucken bzw. von seltenen Drucken des 18. Jahrhunderts (PD W. Wegstein, Inst. f. Deutsche Philologie)
- Reproduktion historischer Aufnahmen, die auf Mikrofiche vorliegen (Prof. H. Beinlich, Institut f. Ägyptologie)
- Einscannen histologischer Präparate für Forschung und Lehre (Prof. H. Heinsen, Morphologische Hirnforschung, Psychiatrische Klinik)

Produktinformation digitale Still Life Kamera ProgRes 3012

Die Flexibilität bei den Vorlagen (Aufsicht, Durchlicht und reale Objekte), der hervorragende Bedienungskomfort und die hohe Datenqualität machen die ProgRes 3012 zum idealen Werkzeug in der Fotografie, Druckvorstufe, im Reprobereich sowie in der Archivierung, Mikroskopie und Bildanalyse.

Aufgrund ihrer einzigartigen Technologie bietet die ProgRes 3012 folgende Möglichkeiten:

- Das farbige Livebild gibt jederzeit die vollständige Kontrolle über die Bildkomposition.
- Erstaunlich kurze Aufnahmezeiten liefern beeindruckende Ergebnisse in Sekunden.
- Die Bilder der ProgRes 3012 sind frei von Farbsäumen und Moiré-Effekten.
- Durch die variable Auflösung kann die Bildgröße den jeweiligen Erfordernissen angepaßt werden.
- Präzise Wiedergabe feinsten Farbverläufe und Detailzeichnung in Lichtern und Schatten sind das Ergebnis ausgereifter Technologie.

Technische Daten	
Typ	Stand Alone-Kamera
nur für unbewegte Motive?	ja
Offline-Einsatz möglich?	nein
CCD-Sensor: Abmessungen	8,5 x 6,4 mm
CCD-Sensor: Pixel	760 x 580
Farberfassung	Microscanning
max. Auflösung der Aufnahme (h x v)	4.608 x 3.480
max. Dateigröße der Aufnahmen	48 MB
Zeit pro Aufnahme in max. Auflösung	35 Sek.
Farbtiefe (Bit)	3 x 12
Autofokus	nein
verwendbare Lichtarten	Halogen 3400 K, HMI
Belichtungszeiten	4 verschiedene
Blenden	objektivabhängig
Sucher	farbiges Livebild auf Computermonitor
Realtime-Preview auf Monitor	ja
Fernsteuerung vom Rechner	komplett computergesteuert außer Fokus und Blende
mitgelieferte Software	ProgRes-Plugin (Photoshop)
Hersteller	Jenoptik Laser Optik Systeme GmbH, Digitale Kameras

Die Technologie

Die Kamera arbeitet mit einem hochwertigen CCD-Flächensensor mit RGB-Filtern auf dem Chip.

Ausgereifte Technologie garantiert exzellente Farbtreue und sehr gute Bildqualität. Durch das PAD-Verfahren (Piezo controlled Aperture Displacement) ist eine hohe Bildauflösung möglich: Informationen, die auf nicht lichtempfindliche Elemente des CCD-Sensors treffen, gehen normalerweise verloren. Mit dem „Microscanning“ ist dies nicht der Fall - ein zweidimensionales Verschieben des Chips garantiert eine lückenlose Erfassung des Bildes. Je kleiner die Verschiebeschritte sind, desto höher ist die Bildauflösung. Jede Bildposition wird dabei viermal abgetastet, von einem roten, einem blauen und zwei grünen Pixeln. Dadurch wird eine absolut moiré- und farbsaumfreie Abtastung der Originale gewährleistet. Das generierte Analogsignal wird innerhalb der Kamera sofort in digitale Daten umgewandelt.

Auflösung	Aufnahmezeit	Dateigröße
512 x 387	1.3 Sek.	0.60 MB
748 x 580	1.9 Sek.	1.24 MB
1024 x 774	5.1 Sek.	2.26 MB
1536 x 1160	3.8 Sek.	5.10 MB
1928 x 1450	6.4 Sek.	8.00 MB
3072 x 2320	15.4 Sek.	20.39 MB
3856 x 2900	25.6 Sek.	32.00 MB
4608 x 3480	34.6 Sek.	45.88 MB

Die hochwertige Kameraelektronik bürgt für hervorragende Detailzeichnung in Lichtern und Schatten, ohne Blooming.

Sensor:

2/3" CCD-Flächensensor, RGB Farbmaske, Interline Transfer, 8.5 x 6.4 mm, PAD modifiziert

Objektive:

- Lametar (f=25 mm, Blende 2,8-11)
- Rodagon (f=50 mm, Blende 2,8-16)
- Zusätzliche Tuben und Zwischenringe

FAX-Server

Hartmut Plehn

Der FAX-Server erlaubt es, im Zusammenspiel mit der Telefonanlage der Universität FAX-Nachrichten am PC-Arbeitsplatz als Dateien zu empfangen oder von Software-Anwendungen (z.B. Textverarbeitungsprogrammen) aus direkt zu verschicken.

Was muß man tun, um den FAX-Server nutzen zu können?

Grundvoraussetzung für den Zugang zum FAX-Server ist, daß man Benutzer des Rechenzentrums ist. Mit der vom Rechenzentrum erteilten Login-Kennung ist eine Nutzungsberechtigung für den FAX-Server bei der Zentralverwaltung der Universität (Referat VI/2) zu beantragen. Das Antragsformular ist vollständig mit Login-Kennung des Rechenzentrums und Anordnungsstellen-Nr. (6-stellig) auszufüllen. Man erhält ein Paßwort für den FAX-Server und die zugeordnete FAX-Nummer. FAX-Nachrichten, die an diese FAX-Nummer geschickt werden, werden von der Telefonanlage an den FAX-Server weitergeleitet und können nur vom Benutzer selbst gele-

sen werden. Als FAX-Benutzerkennung, mit der man sich zusammen mit seinem Paßwort dem FAX-Server gegenüber authentifiziert, wird die übliche Login-Kennung ohne Novell-Kontext verwendet (z.B. abcd123). Nutzungsberechtigte können die FAX-Software (siehe Abbildung1) am mit den RZ-Novellservern verbundenen PC-Arbeitsplatz unter den Betriebssystemen Windows 3.1, Windows 95 und Windows NT (nicht Apple und nicht Unix) durch Aufruf des Programmes

`g:\windows\prog\ccs\winprog\ccswin.exe`

installieren.

Wie benutzt man den FAX-Server?

Eingegangene FAX-Nachrichten kann man mit dem „CCS Mail Office“-Programm, das während des Installationsvorgangs auf dem betreffenden Arbeitsplatz eingerichtet wurde, lesen oder ausdrucken. Zum Abschicken von FAX-Nachrichten kann man in allen gängigen Windows-Anwendungen den Drucker „CCS-

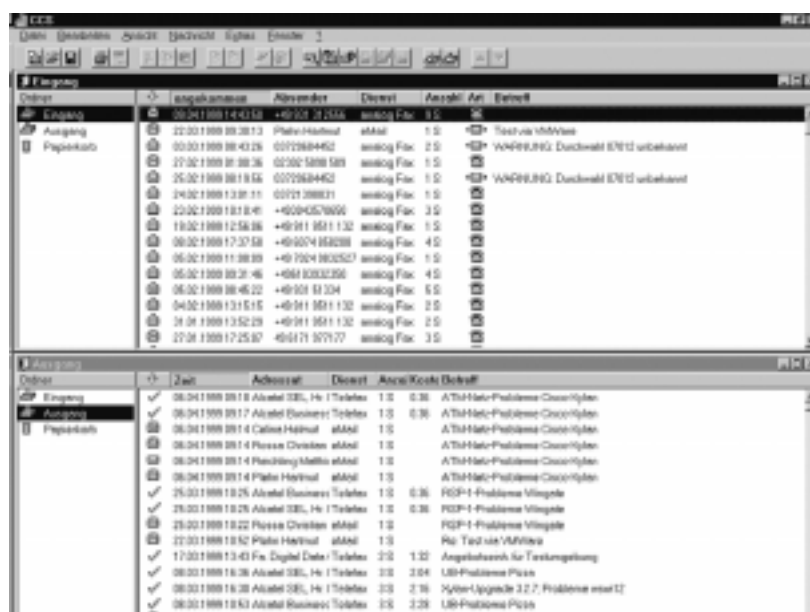
FAX“ verwenden. Der Inhalt der Datei wird dann (statt tatsächlich an einem Drucker ausgedruckt zu werden) an die FAX-Nummer geschickt, die man beim „Ausdrucken“ angibt. Sowohl beim Empfangen von FAX-Nachrichten per „CCS Mail Office“ als auch beim Abschicken per „CCS-FAX“-Drucker muß

man sich mit seiner Nutzerkennung und einem Paßwort dem FAX-Server gegenüber authentifizieren.

Bei FAX-Nachrichten ins öffentliche Telefonnetz ist auf jeden Fall eine „0“ vorzuzahlen. Bei internen Nachrichten zu FAX-Geräten, die im universitätsweiten Telefonnetz angeschlossen sind, ist die vierstellige Nummer und gegebenenfalls davor die Querverbindungsnummer anzugeben. Die eigene vierstellige FAX-Nummer bekommt man vom Rechenzentrum mitgeteilt. Vom öffentlichen Telefonnetz aus muß die „888“ für die Universität und gegebenenfalls die Ortsnetzkennzahl für

Würzburg vorgewählt werden. Wenn an andere Personen innerhalb der Universität, die auch am FAX-Dienst via FAX-Server teilnehmen, eine Nachricht geschickt werden soll, wählt man im Eingabefenster „Dienste“ statt „Telefax“ einfach „An“ und sucht den gewünschten Benutzer, indem man das Feld „Name“ mit der rechten Maustaste anklickt und „Adressieren“ auswählt. Danach selektiert man den/die Benutzer, an die eine FAX-Nachricht geschickt werden soll, mit „Hinzufügen“ und „Ok“. Auf diese Art und Weise wird die Nachricht innerhalb der FAX-Server-Datenbank zugestellt. Telefonanlage und -leitungen werden nicht belastet.

Abbildung 1: Beispiel für die Anwendung „CCS Mail Office“ mit den Ordnern für bisher eingegangene bzw. gesendete Fax-Nachrichten.



Dokumentation

Das Handbuch zum FAX-Server befindet sich auf dem ftp-Server des Rechenzentrums

<ftp://ftp.rz.uni-wuerzburg.de/pub/local/faxserver/handbuch.exe>

als selbstextrahierende Datei. Nach Aufruf dieser Datei werden die einzelnen Kapitel des Handbuchs im Word-Format ausgepackt.

Statistiksoftware – ein Überblick

Alois Spahn

In allen Bereichen der Universität fallen bei Forschungsvorhaben, Studien- und Diplomarbeiten Daten an. Zumeist werden umfangreiche und daher zunächst unübersichtliche Datenmengen gesammelt, die erst nach Veranschaulichung durch geeignete Grafiken und nach Verdichten und Berechnen von Kennzahlen wertvolle Informationen liefern. Nicht selten ist die Beschaffung von Daten aufwendig, teuer oder stellt einen großen Aufwand dar, so daß nur wenige Daten vorliegen. Auch in diesen Fällen sollen jedoch Trends, Unterschiede und Zusammenhänge aufgedeckt werden.

Dies ist die Domäne der Statistikpakete, die heute in der Regel eine grafische Oberfläche besitzen und daher ohne großen Lernaufwand von jedermann benutzbar sind und so eine weite Verbreitung gefunden haben. Allerdings ist die heutige Generation der Statistiksoftware nichts weiter als eine mächtige Rechenhilfe, die die Entscheidung, ob ein Verfahren sinnvoll

ist, und ob dessen Voraussetzungen erfüllt sind, dem Benutzer überläßt, und damit auch keine Fehler im Vorfeld der Auswertungen wie bei der Konzeption einer Studie und der Datenerhebung wiedergutmachen kann. Eine gründliche Versuchsplanung kann keine Software abnehmen, es sollte vielmehr rechtzeitig der Rat eines Fachmannes eingeholt werden.

Excel, StarCalc und andere Tabellenkalkulationsprogramme

Da Office-Pakete zur Grundausstattung eines PCs gehören, werden Tabellenkalkulationsprogramme wie Excel oder StarCalc gerne für Auswertungen benutzt. Sie erlauben die bequeme Dateneingabe, formelartige Berechnungen und die Erstellung von Grafiken. Bzgl. statistischer Funktionen sind sie jedoch schnell mit ihrem Latein am Ende. Sollen mehr Verfahren als nur Mittelwerte, Korrelationen oder t-Test durchgeführt werden, empfiehlt es sich, gleich auf eine spezialisierte, für statistische Auswertungen konzipierte Software zurückzu-

greifen. Bei Benutzung eines Statistikpaketes kann auch davon ausgegangen werden, daß für die Berechnungen solche Algorithmen verwendet werden, die im Rahmen der möglichen numerischen Genauigkeit korrekte Ergebnisse liefern. Dabei darf sich aber niemand darüber hinwegtäuschen, daß jede Software von einiger Komplexität auch Fehler, "Bugs" aufweist. Die Seriosität eines Herstellers zeigt sich meines Erachtens auch daran, inwieweit Listen der aufgetretenen Fehler veröffentlicht werden.

SPSS für Windows

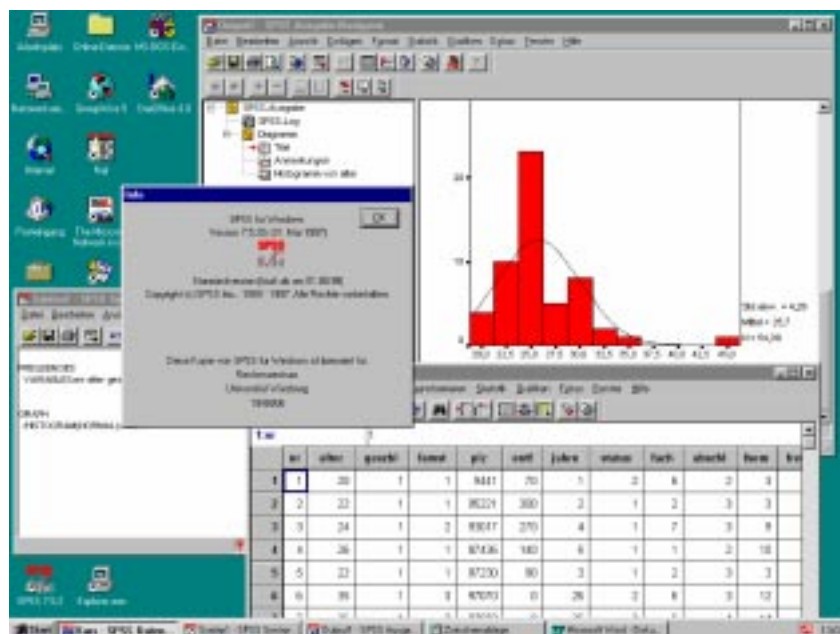
Der "Renner" unter den Statistikpaketen an der Universität Würzburg und wohl auch an den meisten deutschen Universitäten ist SPSS für Windows. Dabei ist SPSS eine Software aus den 60er Jahren, die jedoch zuletzt stark geliftet und mit neuen schicken Kleidern, sprich einer intuitiv zu bedienenden grafischen Oberfläche versehen wurde. Damit hat auch der sporadische Nutzer ohne große Einarbeitung gleich ein Erfolgserlebnis. Gleichzeitig steht den Experten immer noch die alte Programmiersprache zur Verfügung, um auch Feinheiten des Systems nutzen oder nicht vorhandene Verfahren und Koeffizienten selbst programmieren zu können.

Die Datenerfassung erfolgt wie bei Excel in Form einer Rechtecktafel. Zur Veranschaulichung der statistischen Ergebnisse bietet SPSS eine Reihe verschiedener Grafiktypen, die über einen Editor an das persönliche ästhetische Empfinden angepaßt werden können.

Geradezu inflationär ist jedoch die Tendenz des Herstellers, neue Verfahren als neue Module gegen zusätzliches Entgelt zu verkaufen. Zeitreihenverfahren, exakte Tests, Strukturgleichungsmodelle, Analyse kategorialer Variablen oder von fehlenden Werten sind u.a. jeweils eigene Module. Allein für die klassischen Statistikverfahren werden schon drei Module benötigt, und zwar das Base System, Professional und Advanced Statistics.

Verwirrend ist weiterhin die momentane Versionsvielfalt. Aktuell ist Version 8; Version 9 liegt jedoch schon als englischsprachige Version vor, allerdings nur unter Windows 95 und NT. Für die Plattformen Windows 3.1, MacIntosh und Unix wird noch Version 6 angeboten, die Weiterentwicklung dafür ist jedoch zur Zeit eingestellt. SPSS konzentriert sich ganz auf den gewinnträchtigen Markt der neuen Windows-Systeme.

Abbildung 1: Bildschirmauszug der Benutzeroberfläche von SPSS



Am Rechenzentrum steht Ihnen SPSS in den jeweiligen Versionen auf den Unix-Computerservern und allen PCs, die an den dezentralen Novellservern angeschlossen sind, zur Verfügung. Dabei umfaßt die Unix-Lizenz mit Trends, Categories und Tables zusätzliche Module. Kurse und Beratung zur Software werden selbstverständlich auch angeboten. Für eigene Installationen kann auf eine Campuslizenz zurückgegriffen werden. Gegen einen

jährlichen Mietpreis kann die Software in Instituten, aber auch zu Hause von Mitarbeitern und Studenten der Universität Würzburg für Zwecke aus Lehre und Forschung genutzt werden.

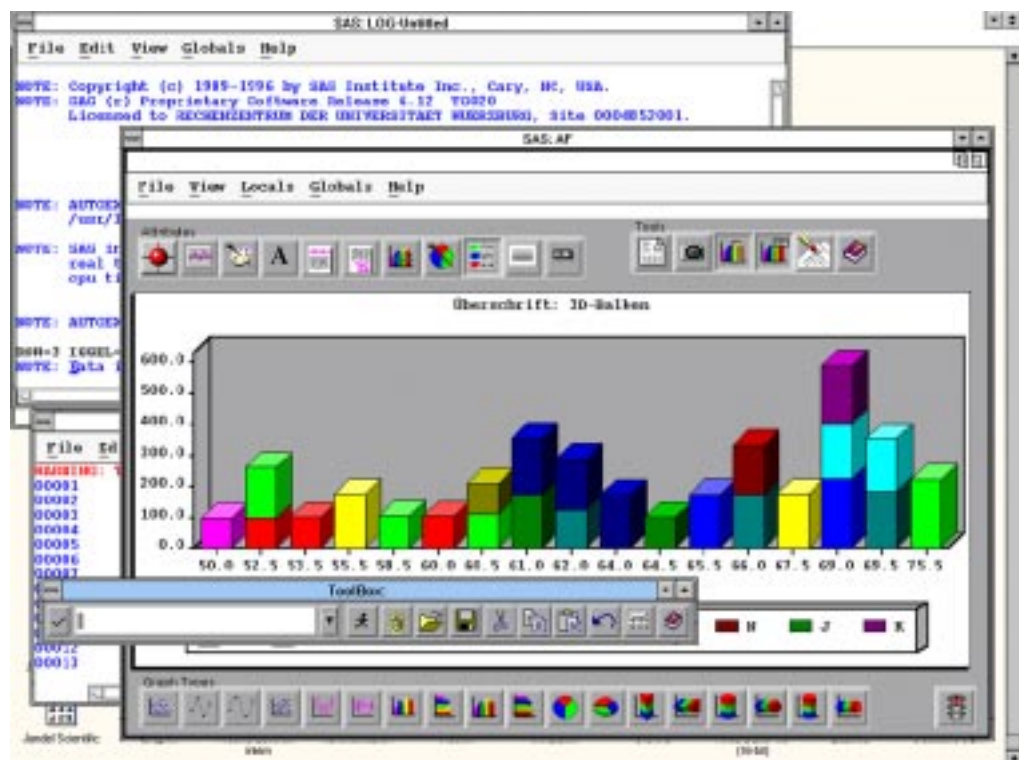
Insgesamt haben Sie mit SPSS für Windows eine zuverlässige und bedienungsfreundliche Software und somit sicherlich eine gute Wahl getroffen.

Das SAS System

Marktführer in Industrie und Wirtschaft in der Sparte „Information Delivery“ ist das SAS System. Ausschlaggebend dafür war, daß sich SAS recht früh schon aus der Marktnische der

statistischen Auswertesysteme löste und weitere Funktionalitäten in die Software integrierte, ein Trend, den SPSS in den letzten Jahren nachvollzieht.

Abbildung 2: Bildschirmauszug der Benutzeroberfläche von SAS



Dazu zählt u.a. die Anwendungsentwicklung, d. h. der Programmierer kann individuelle grafische Oberflächen erstellen, mit denen dann etwa die Firmenleitung arbeitet und so, auf einen einfachen Mausklick hin, Umsatzzahlen in Tabellenform oder als Grafik abrufen kann (Schlagworte Executive Information System, Decision Support System). SAS Institute bietet selbst auch Module an, die mit dieser Anwendungsentwicklung erstellt sind, und die die Funktionalitäten des SAS Systems für bestimmte Wirtschaftszweige wie etwa der pharmazeutischen Industrie oder der Finanzwelt bündeln und maßschneidern. Diese sind dann mit Maus und Menüs zu benutzen. Ansonsten wird man kaum ohne die zugrundeliegende Programmiersprache auskommen. Diese ist jedoch auch eine der Stärken des SAS Systems, erlaubt sie doch Daten, die in verschiedensten Formaten und Datenbanken auf den verschiedensten Rechnern örtlich verteilt

vorliegen, zu sammeln, zu ordnen und zu verdichten (Schlagworte Data Mining, Data Warehouse).

Das SAS System ist aus Kostengründen nur auf einer Unix-Workstation von HP (Hostname wrx13) lizenziert. An Modulen ist das Basis System, STAT für die statistischen Auswertungen, dessen Umfang dem von mehreren SPSS-Modulen entspricht, ETS für Zeitreihenanalysen, GRAPH für Grafiken, FSP zur Datenerfassung und ASSIST für das menügeführte Arbeiten mit dem SAS System vorhanden. Eine Ausdehnung der Installation auf die PC-Welt scheitert zur Zeit noch an der Lizenzpolitik von SAS, die kein "concurrent use" kennt. Jeder an einem Netzwerk angeschlossene PC, der also theoretisch SAS nutzen könnte, muß auch eine Lizenz besitzen, was das Ganze unbezahlbar macht.

Statistica

Eine gute Alternative zu den marktbeherrschenden Statistikpaketen stellt Statistica dar. Die Software ist erfreulicherweise nicht in viele Module aufgeteilt und umfaßt dennoch ein breites Spektrum an Auswertungsmöglichkeiten einschließlich Zeitreihenverfahren, Korrespondenzanalyse und der Modellierung von Strukturgleichungen. Lediglich Verfahren der statistischen Qualitätskontrolle, Prozeßanalyse und Versuchsplanung sind in einem

eigenen Modul ausgelagert. Ein weiterer Pluspunkt sind die im Vergleich zu SPSS deutlich besseren grafischen Möglichkeiten. Die Software ist ebenfalls leicht mit Maus und Menüs zu benutzen und liegt als Campuslizenz für die Windows-Plattformen und MacIntosh vor. Sie kann über die dezentralen Novellserver benutzt oder auf jährlicher Mietbasis für eigene Installationen bezogen werden.

S-Plus

Als Spezialsystem für den Experten muß S-Plus bezeichnet werden. Die Software glänzt zum Teil durch ausgefallene Verfahren, die über eine objekt- und vektororientierte Sprache zu bedienen sind, weist vielfältige mathe-

matische, statistische und grafische Funktionen auf, und ist besonders für den Methodenentwickler interessant. S-Plus ist auf den Unix-Computeservern im Rechenzentrum installiert.

TUSTEP

Wilhelm Ott, Tübingen

Das Tübinger System von Textverarbeitungs-Programmen TUSTEP wurde mit dem Ziel entwickelt, Probleme der wissenschaftlichen Textdaten-Verarbeitung mit knapperen und problemnäheren Anweisungen lösen zu können, als dies auf herkömmliche Weise, z.B. mit Programmiersprachen, möglich wäre.

Zum genannten Zweck wurden für die wichtigsten Grundoperationen der Textdaten-Verarbeitung Programme bereitgestellt, deren Leistung der Benutzer über Parameter spezifizieren und die er in vielfältiger Weise für die Lösung verschiedenster Aufgabenstellungen kombinieren kann.

Selbstverständlich gehören auch die für die Dokumenten-Erstellung notwendigen Funktionen: Eingabe, Korrektur, Formatieren, Drucken von Texten zum Leistungsangebot, da diese in allen Wissenschaftsbereichen zum Zweck der Dokumentation und Publikationsvorbereitung benötigt werden. TUSTEP wurde aber als Werkzeug vor allem für diejenigen Wissenschaften entwickelt, in denen Texte Objekte der Forschung sind: Philologien, Sprachwissenschaften, Literaturwissenschaften, historische Wissenschaften, Bibliothekswesen; Wissenschaften also, in denen nicht nur neue Texte als Produkt der eigenen wissenschaftlichen Arbeit erstellt und publiziert werden sollen, sondern in denen schon existierende, überlieferte, schriftlich fixierte oder zu fixierende Texte (einschließlich literarischer Texte und historischer Quellen) durch kritische Neuedition gesichert, sprachlich und stilistisch analysiert, inhaltlich erschlossen, bibliographisch erfaßt werden müssen. Dem tragen Grundoperationen der Textdaten-Verarbeitung Rechnung, die mit folgenden Schlagwörtern grob charakterisiert werden können: Vergleichen von verschiedenen Textfassungen; Korrigieren nicht nur interaktiv im Editor, sondern auch anhand

vorbereiteter Korrekturanweisungen; Zerlegen von Texten in Elemente (z. B. Wortformen); Sortieren von Textelementen oder von längeren Texteinheiten nach einer Vielzahl von Alphabeten und anderer Kriterien; Register erstellen durch Zusammenfassen sortierter Textelemente; Rechnen mit Zahlenwerten, die bereits im Text enthalten sind (z. B. Kalenderdaten) oder aus ihm gewonnen werden können (z. B. die Zahl der Wörter in einem Satz), und Ausgeben in verschiedenen Formaten, einschließlich solcher, die von anderen Systemen (z. B. SPSS zur statistischen Auswertung) benötigt werden.

Aufgaben, die mit TUSTEP bearbeitet werden, reichen vom Schreiben einer Seminararbeit bis hin zum Erstellen von umfangreichen Bibliographien, Lexika, Indizes, Konkordanzen, Wörterbüchern, Editionen und natürlich auch Monographien, jeweils einschließlich der automatischen Herstellung der Druckvorlagen für diese Werke in der vom Buchdruck gewohnten Qualität.

Eine für alle Rechner identische Benutzer-Oberfläche konnte nur dadurch erreicht werden, daß auf die Ausnutzung spezieller Eigenschaften einzelner Rechner und Betriebssysteme verzichtet wurde.

Auszug aus der Einleitung zum TUSTEP-Handbuch mit freundlicher Genehmigung des Herausgebers.

Wie wichtig ist ein gutes Paßwort?

Peter Dieterich

Das Anmelden an einem Rechner mit Benutzername und Paßwort gehört zu den täglichen Routinetätigkeiten. Gerne würden viele darauf verzichten und wählen häufig zumindest ein sehr einfaches Paßwort. Das Rechenzentrum kontrolliert auf den Unix-Systemen regelmäßig die Güte der Paßwörter und fordert die Benutzer gegebenenfalls auf, ein schwaches Paßwort zu beseitigen. Dieser Artikel gibt einige Hintergrundinformationen zu dieser Thematik sowie praktische Tips zum Umgang mit Paßwörtern aus Unix-Systemen. Die in dieser Hinsicht komplexere PC-Seite wird in einem späteren Artikel behandelt.

Die Problematik

Fast jeder hat es schon unzählige Male gemacht, das Einloggen an einem Rechner mit Hilfe von Benutzername und Paßwort. Doch was passiert eigentlich dabei und was sollte beachtet werden?

Zunächst die "trockene" Definition:¹

"Das Paßwort ist eine individuell gewählte, geheimzuhaltende Zeichenkette für die Authentisierung eines Nutzers gegenüber einem Rechner; es wird auch Zugangscode genannt."

Das heißt, wer im Besitz einer Benutzerkennung (*username*) und des dazugehörigen Paßwortes ist, kann sich Zugang zum entsprechenden Rechner verschaffen und dort mit allen Rechten des Benutzers arbeiten, dem die Kennung gehört. Dies ermöglicht unter anderem

1. den Zugriff auf alle persönlichen Daten,
2. das Versenden von Mails unter dieser Kennung,
3. die Nutzung von Ressourcen (z. B. Rechenzeit und Druckermöglichkeiten), was auch beträchtliche Kosten verursachen kann,
4. den Mißbrauch des Accounts als Sprungbrett, z. B. um an anderer Stelle in ein Computersystem einzubrechen.

Aus diesen Gründen sollte das Paßwort auf alle Fälle geheim gehalten werden, selbst wenn die persönlichen Daten nicht besonders wichtig erscheinen. Ein durch ein schlechtes Paßwort unzureichend geschützter Zugang stellt zudem eine nicht zu unterschätzende Gefahr für die Systeme und die Daten anderer Benutzer dar.

¹ Bericht "Hochschulnetze in Bayern, Zugang, Nutzung, Schutz vor Mißbrauch" (1997), siehe <http://www.rz.uni-wuerzburg.de/netzbericht/>.

Ein paar technische Details und Hintergründe

Beim Einloggen in einen Rechner wird das Paßwort abgefragt. Der Rechner muß die eingegebene Zeichenfolge mit seiner eigenen Information vergleichen und ermöglicht bei Übereinstimmung den Zugang.

Das Paßwort ist jedoch auf dem Rechner nicht im Klartext, sondern nur in verschlüsselter Form verfügbar. Zum Vergleich wird deshalb das vom Benutzer angeforderte Klartextpaßwort nach einem vorgegebenen Verfahren verschlüsselt und mit einem entsprechenden Eintrag in der sogenannten Paßwortdatei verglichen.

Als Verschlüsselungsverfahren kommt bei Unix-Systemen² (siehe Abbildung 1) das DES-Verfahren zum Einsatz. Zu diesem ist keine Umkehrung bekannt, d. h. es kann nicht aus dem verschlüsselten Paßwort auf das Klartextpaßwort geschlossen werden.

Auch einfaches Ausprobieren aller Möglichkeiten - ein sogenannter *brute force* Angriff - geht in der Regel über bestehende Rechenkapazitäten weit hinaus. Wie in Tabelle 1 dargestellt, ist dabei jedoch zu beachten, daß die benötigte Zeit stark von der Menge der zur Verfügung stehenden Zeichen sowie der Länge des Paßwortes abhängt. Werden beispielsweise nur Kleinbuchstaben und eine Paßwortlänge von 6 Zeichen verwendet, so können alle möglichen Paßwörter auf einem leistungsfähigen Rechner in nur 4 Stunden untersucht werden. Hingegen steigt der Aufwand bei 62 möglichen Zeichen und einer Paßwortlänge von acht derartig, daß die heutigen Rechenkapazitäten dafür nicht ausreichen.

² Die Darstellung in diesem Bericht bezieht sich in der Regel auf Unix-Systeme, wobei die Paßwortproblematik auch für andere Betriebssysteme relevant ist. Bei diesen sind - im Gegensatz zu Unix - die Hintergründe und mögliche Abhilfen komplizierter.

Abbildung 1: Der Verschlüsselungsmechanismus auf Unix-Systemen. Aus dem Klartext-Benutzerpaßwort wird durch die crypt-Routine (und einem zufällig beigemischten Anteil, dem sogenannten salt) das verschlüsselte Paßwort generiert.

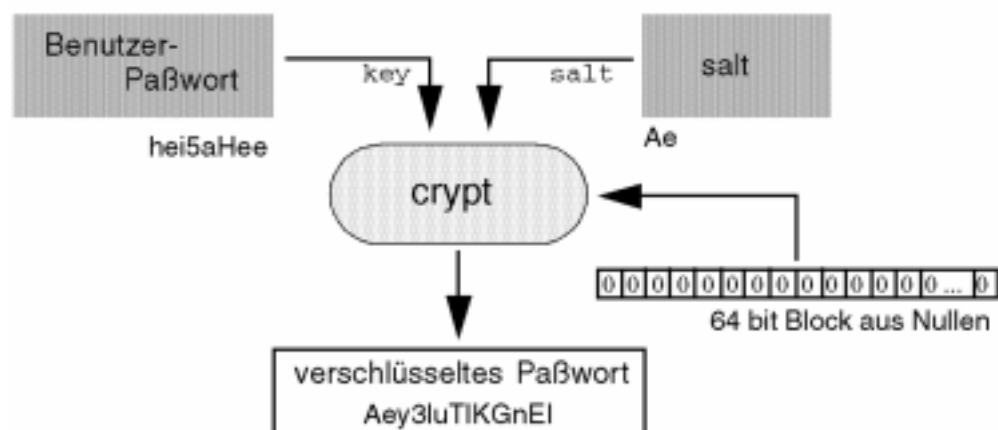


Tabelle 1: Einige Beispiele für den Rechenzeitbedarf von Crack-Läufen auf einer der schnellsten Maschinen (`wrzz17`) des Rechenzentrums. Eine optimierte `crypt`-Routine bearbeitet etwa 20000 Versuche pro Sekunde. Je nach Anzahl der Zeichen, die für das Klartextpaßwort verwendet werden können, ergeben sich die folgenden Zeiten für das Durchlaufen aller Möglichkeiten.

26 Zeichen = Alphabet in Groß- oder Kleinbuchstaben
 52 Zeichen = Alphabet in Groß- und Kleinbuchstaben
 62 Zeichen = Alphabet und 10 Ziffern

Anzahl der Zeichen	Paßwortlänge	Anzahl der Möglichkeiten	Maximale Zeit
128	8	$128^8 = 72057594037927936$	114246 Jahre
	7	$128^7 = 562949953421312$	892 Jahre
	6	$128^6 = 4398046511104$	7 Jahre
95	8	$95^8 = 6634204312890625$	10518 Jahre
	7	$95^7 = 69833729609375$	110 Jahre
	6	$95^6 = 735091890625$	1 Jahr
62	8	$62^8 = 218340105584896$	346 Jahre
	7	$62^7 = 3521614606208$	6 Jahre
	6	$62^6 = 56800235584$	33 Tage
52	8	$52^8 = 53459728531456$	85 Jahre
	7	$52^7 = 1028071702528$	2 Jahre
	6	$52^6 = 19770609664$	11 Tage
26	8	$26^8 = 208827064576$	121 Tage
	7	$26^7 = 8031810176$	5 Tage
	6	$26^6 = 308915776$	4 Stunden

Während ein *brute force* Versuch bei geeigneter Paßwortlänge und Zeichenanzahl in der Regel erfolglos bleibt, stellt bei Unix-Systemen die Paßwortdatei ein besonderes Problem dar. Den Inhalt dieser Datei können meist³ alle Benutzer lesen. Sie finden dort zwar nicht die Klartextpaßwörter aller anderen Benutzer, doch könnten sie versuchen, schwache Paßwörter zu erraten.

Hierzu gibt es auch fertige Programme, das bekannteste heißt Crack. Diese nehmen eine große Anzahl von Wörtern, beispielsweise aus

³ Der Einsatz eines sogenannten *shadow password Files* unterbindet dies. Dies ist jedoch im Zusammenspiel mit dem NIS-Mechanismus, der bei Clustern zur Zentralisierung der Verwaltung fast immer zum Einsatz kommt, meist nur sehr aufwendig (NIS oder DCE/DFS) oder überhaupt nicht möglich.

Lexika, verschlüsseln diese und vergleichen das Resultat mit den verschlüsselten Einträgen im Paßwortfile. Zudem erkennt Crack einfache Veränderungen von bekannten Wörtern, beispielsweise das Anhängen einer Zahl oder das Rückwärtsschreiben eines Wortes. Einige dieser Varianten, die Crack registriert, sind in Abbildung 2 wiedergegeben.

Da Benutzer häufig nur einfache Paßwörter wählen, verlaufen Crack-Tests meist "sehr erfolgreich". In Abbildung 3 sind einige Erfahrungen der Rechenzentrumsläufe festgehalten. Beim letzten Test hatten etwa 10% der Benutzer ein schlechtes Paßwort. Diese wurden aufgefordert, ihr Paßwort möglichst bald zu verändern.

Abbildung 2: Was kann Crack?

- Crack verschlüsselt Wörter und Kombinationen von Wörtern und vergleicht diese mit den kodierten Einträgen in der Paßwortdatei.
- Besonders schnell werden Benutzername = Paßwort bzw. die Verwendung von persönlichen Informationen im Paßwortfile erkannt.
- Crack kennt Namen und Wörter aus verschiedensten Lexika und in vielen Sprachen. Auch Sammlungen von Zitaten, Liedtexten etc. können berücksichtigt werden.
- Crack erkennt Umbenennungen, z. B.
i, l, t nach 1 E nach 3, o nach 0 (Null),
A nach 4, g nach 9, Z nach 7.
- Weiterhin wird das Anhängen von Ziffern und Rückwärtsschreiben erkannt.

Abbildung 3: Typische Resultate eines Crack-Laufs im Rechenzentrum

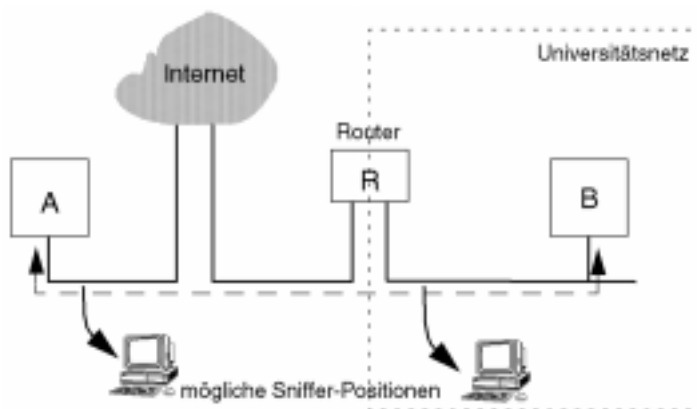
- Viele Vornamen
jennifer, nicole, isabella, christel, jasper, michael
- Viele sonstige Eigennamen
tibet, bristol, tiepolo
- Aus der Tätigkeit abgeleitete Bezeichnungen
geologie, zoologie, hacker
- Prominenz und Exoten
jesuschr, Kanzler, ghandi, axolotl, Tupaia, ramses, grendel

Weitere Gefahren

Auch wenn ein sehr gutes Paßwort gewählt wird, kann dieses unter Umständen leicht in die Hände dritter geraten. Bei jedem Login an einem Rechner, der über das Netz ausgeführt wird, geht in der Regel das Paßwort im Klartext über die Netzleitungen. Befindet sich ein sogenannter *Sniffer* dazwischen (siehe Abbildung 4), so kann dieser Benutzername und Paßwort mitprotokollieren.

Hat ein Angreifer über diesen Weg erst einmal Zugang zu einem Account gefunden, so stehen ihm weitere Möglichkeiten offen, beispielsweise mit Crack-Läufen auch andere Accounts zu knacken oder sich gar Systemprivilegien zu verschaffen.

Abbildung 4: Gefahren, die von Sniffnern im Netz ausgehen können, falls eine Kommunikation zwischen den Rechnern A und B stattfindet.



Derartige Gefahren sind besonders relevant, wenn man sich über das weltweite Internet beispielsweise aus USA am heimischen Rechner einloggt. Andererseits bieten Unix-Systeme kaum Alternativen, die eine verschlüsselte Übertragung des Paßwortes bewirken und somit einen Schutz gegen Sniffer darstellen.

Eine sehr gute Ausnahme ist das *public domain* Programm *Secure Shell*⁴. Hier kommen kryptographische Elemente zum Einsatz, so daß die Kommunikation, d. h. insbesondere auch die Paßwort-Authentisierung, verschlüsselt verläuft. Ein weiterer Vorteil der *Secure Shell* ist deren Verfügbarkeit auf sehr

vielen Betriebssystem-Plattformen wie Unix, Windows, MacOS und OS/2. Informationen über die Funktionalität der *Secure Shell* und deren lokale Installation am RZUW finden sich auf dem RZ-WWW-Server⁵.

Kann man derartige Tools nicht nutzen, so sollte man sich nicht allzu leichtfertig über ein unbekanntes Netz einloggen. Sinnvoll ist auf alle Fälle, sich einen speziellen heimischen Rechner auszuwählen. Für diesen kann man ein besonderes Paßwort wählen und ihn auch genauer überwachen, z. B. häufiger nachsehen, wer wann und von wo eingeloggt war.

⁴ Siehe <http://www.ssh.org/>

⁵ Siehe <http://www.rz.uni-wuerzburg.de/security/ssh/>

Ein paar Schutzmaßnahmen

Ein gutes Paßwort zu finden, das man sich auch gut merken und schnell tippen kann, ist nicht ganz leicht. Deshalb abschließend noch ein paar Hinweise, wie Sie Ihr Paßwort ein wenig besser schützen können und was Sie im Umgang mit dem Paßwort beachten sollten:

1. Das Benutzerpaßwort ist personenbezogen und darf nicht weitergegeben werden. Dies ist auch in der Benutzungsordnung des Rechenzentrums festgelegt.
2. Es sollte ein "gutes" Paßwort gewählt werden, auf keinen Fall Username = Paßwort, einfache Namen, Urlaubsziele oder irgendwo abgedruckte Wörter.
3. Ein Paßwort sollte Sonderzeichen und Kombinationen aus Zeichen und Ziffern enthalten, wobei die Paßwortlänge maximal sein sollte (8 bei Unix-Systemen, vergleiche auch Tabelle 1). Vorsicht ist dabei jedoch geboten, da unterschiedliche Tastaturbelegungen eine Falle darstellen können.
4. Das Paßwort sollte nicht aufgeschrieben werden.
5. Auf den Computerservern des Rechenzentrums gibt es das Programm `checkpw`, das ein mögliches Paßwort auf dessen Güte hin überprüfen kann.
6. Vorsicht ist bei *remote login* über ein unsicheres Netz geboten. Falls möglich sollten Produkte wie die *Secure Shell* verwendet werden.

Fazit

Die Wahl eines guten Paßwortes und der vorsichtige Umgang damit ist sowohl für den persönlichen Schutz der Daten und Ressourcen als auch für alle anderen Benutzer sehr wichtig.

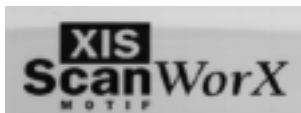
Unix-Systeme haben, was Paßwörter betrifft, diverse Defizite, so daß der Anwender einige in diesem Artikel beschriebene Dinge beachten sollte.

Literaturhinweise

- [1] Practical UNIX Security, Simson Garfinkel und Gene Spafford, 1996 (2. Auflage), O'Reilly & Assoc.
- [2] Firewalls and Internet Security, Repelling the Wily Hacker, William R. Cheswick und Steven M. Bellovin, 1994, Addison Wesley
- [3] Diverse Informationen auf dem WWW-Server von DFN-CERT, <http://www.cert.dfn.de/>

OCR (Optical Character Recognition)

Peter Ruff



Mit ScanWorX können alle Arten gedruckter und maschinengeschriebener Vorlagen automatisch erfaßt werden: Bücher, Zeitschriften, Zeitungen, Briefe, Berichte, Handbücher... Die Texterfassung mit ScanWorX ist genau, schnell und flexibel. Bereits bei geringen Mengen ist eine hohe Wirtschaftlichkeit im Vergleich zur Handerfassung gegeben.

Nach dem Scannen liegt jede Seite als elektronisches Bild vor und wird durch die Erkennungssoftware (ICR Intelligent Character Recognition) analysiert. Die einzelnen Zeichen werden separiert und erkannt. Die auf Kurzweil-Technologie basierende ICR/OCR-Software trainiert sich selbst während des Einlesevorganges automatisch auf die jeweilige Vorlage. Unterschiedliche Software-Module für die jeweils vorkommenden Standarderkennungen und Erkennungsschwierigkeiten spielen dabei zusammen. So gibt es z. B. eine Routine für die Erkennung echter typographischer Ligaturen wie fi, ff, fl etc., zudem

auch eine andere Routine für die Erkennung beliebiger aneinanderstoßender Buchstaben. Das automatische Training reicht jedoch nicht für alle Problemfälle, wie z. B. ungünstig zerbrochene, komplex aneinanderstoßende oder sonstwie zerstörte Zeichen oder Zeichenfolgen: Auch hier liest ScanWorX jedoch weite Bereiche automatisch ein. Dazu kommen mathematische oder kommerzielle Sonderzeichen, Zeichen aus nicht-lateinischen Schriften wie Griechisch, Kyrillisch, Fraktur, usw. Trifft die Software auf solche zerstörten oder unbekanntem Zeichen, werden diese einschließlich des Erkennungsvorschlages angezeigt. Der Anwender bestätigt, korrigiert oder verschlüsselt als Sonderzeichen. Dadurch ist das Zeichen nicht nur berichtet, sondern fließt gleichzeitig in das Training ein, so daß es im folgenden automatisch erkannt wird. Damit verbindet dieses System die Vorteile der automatischen Texterkennung mit den Vorteilen der freien Trainierbarkeit.

Leistungsmerkmale

- automatische ICR-Zeichenerkennung für lateinische Schriften
- integrierte Lexika zur Erkennungsoptimierung: Deutsch, Englisch, Französisch, Italienisch, Spanisch, Niederländisch, Schwedisch, Finnisch, Dänisch, Norwegisch
- Spezialfilter für Matrixdruck und Fax
- Schriftgröße von 6 bis 24 Pica-Punkt
- automatische oder manuelle Spalten- bzw. Layout- Erkennung
- Texterkennung aus TIFF-Dateien; Serienfunktion

- Vorder-/Rückseiten-Verarbeitung
- Dokumentenseparierung
- Markierungen (Schriftgröße, kursiv, fett, Hoch-/Tiefstellung, Unterstreichung, Block-/Flattersatz, Absätze, Einzüge, Tabs, Spaltenbeginn usw.)
- Hoch- und Querformat-Vorlagen
- Textkonvertierung in ASCII, Interleaf, FrameMaker, WordPerfect 5.x, XDOC

ScanWorX kann mittels floating-Lizenzen an den HP-Compute-Servern benutzt werden. Außerdem steht ein spezieller OCR-Arbeitsplatz (Workstation HP Apollo) mit angeschlossenem Scanner (Fujitsu, 400 dpi s/w, DIN A3, Stapelinzug) zur Verfügung.



TextBridge-Pro liefert die Möglichkeit, ein Bild zu lesen und zu erkennen und davon Text auszugeben. Die Bilder können vom lokal angeschlossenen Scanner stammen oder als irgendwo erzeugte TIFF-files vorliegen. Erkannter Text kann in eine Vielzahl von Ausgabeformaten konvertiert werden und damit leicht in einem bevorzugten Textverarbeitungsprogramm weiterverarbeitet werden. Die Benutzeroberfläche von TextBridge-Pro ist einfach im Gebrauch, und die leistungsfähige Dokumenterkennungstechnologie von Xerox Imaging Systems liefert eine Reihe von Funktionen und Möglichkeiten. TextBridge-Pro liest nicht nur alle Arten von Dokumenten, sondern rekonstruiert Ihr Original so, als hätten Sie es in Ihrer Textverarbeitung selbst erstellt, mehrspaltig, mit Tabellen und Bildern, Grafiken ... Durch einen zuschaltbaren Trainingsmodus und ein erweiterbares Benutzerlexikon können auch problematische Dokumente (mit blassen oder ganz dunklen, verschmutzten oder teilweise zerstörten Buchstaben) verarbeitet werden. Mit dem zusätzlich vorhandenen Instant Access OCR-Modul brauchen Sie Ihre Textverarbeitung nicht zu verlassen, nur um eben mal schnell ein Dokument einzuscannen, das Sie in Ihrer aktuellen Arbeit verwenden wollen. Ein zusätzlich vorhandener leistungsfähiger Korrektureditor Proofreader zeigt Ihnen

eventuelle Erkennungsunsicherheiten an, blendet das Originalbild zur optimalen Kontrolle ein und läßt Sie Korrekturen gezielt an den als unsicher erkannten Stellen durchführen, ohne das ganze Dokument prüfen zu müssen. TextBridge-Pro liefert ausgezeichnete Ergebnisse bei der Zeichenerkennung und Formaterhaltung für eine breite Palette von Dokumenten und Vorlagen:

- Dokumente mit Punktgrößen von 6 bis 72 Punkten in nahezu jedem Schriftbild
- Dokumente, die mit Schreibmaschinen, Fotodrucksetzern, maschinellen Druckern, Tintenstrahl- und Laserdruckern erstellt wurden
- Fotokopierte, minderwertige oder verschmutzte Dokumente
- Dokumente mit ein- oder mehrspaltigem Layout
- Dokumente, die Halbtöne erhalten
- Faxe

TextBridge-Pro steht an jedem PC im Novell-Netz zur Verfügung. Im Rechenzentrum sind einige Arbeitsplätze mit angeschlossenem Scanner vorhanden.

Externe Zugänge

Christian Rossa, Peter Tscherner

Der Wunsch, nicht nur vom Arbeitsplatz innerhalb der Universität aus Zugriff auf die DV-Ressourcen (Email, Steuerung von Jobs, ...) zu haben, sondern auch von zu Hause oder von unterwegs, besteht schon seit langem bei vielen Wissenschaftlern. Nun liegt es bereits mehr als 7 Jahre zurück, seit in Würzburg die Ära der Wählzugänge begonnen hat.

Angefangen hat es mit drei analogen Dialogzugängen, die über einen Terminalserver an einen Host (VAX 8810) angeschlossen waren. Die maximale „Übertragungsgeschwindigkeit“ betrug damals 9.6 Kbit/s. Die Zahl der an diesem Dienst interessierten Benutzer wuchs zusehends, so daß sich auch bald Klagen über besetzte Leitungen häuften. Die im Betrieb gemachten Erfahrungen zeigten aber deutlich, daß diese Lösung nicht für den Einsatz von einer großen Anzahl von Anschlüssen geeignet war, da sie unzuverlässig arbeitete und zudem einen hohen Betriebsaufwand mit sich brachte.

Begünstigt durch die Installation von ISDN-fähigen Nebenstellenanlagen in weiten Bereichen der Universität und durch die größere Verbreitung von ISDN in den Privathaushalten konnte bereits im Juli 1995 ein Wählzugang mit 30 digitalen Kanälen realisiert werden. Zum Einsatz kam dabei die ISDN-Router-Lösung netGW der Firma netCS. Dieser ISDN-Router war auf der einen Seite über eine Primärmultiplex-Schnittstelle (S_{2M}) mit der HICOM-Anlage am Hubland verbunden und auf der anderen Seite mit einer Ethernet-Schnittstelle am Hochschulnetz. Er konnte gleichzeitig bis zu 30 digitale Kanäle (ISDN)

mit jeweils 64 Kbit/s bedienen und war sowohl aus dem universitätsinternen wie auch aus dem öffentlichen Telefonnetz erreichbar. Er unterstützte jedoch lediglich das TCP/IP-Protokoll und die statische Adreßvergabe.

Nach der Inbetriebnahme wuchs die Anzahl der interessierten Benutzer sehr schnell, obwohl auch diese Lösung nicht das komplette Bedarfsspektrum abdecken konnte. Die wesentlichen Nachteile waren die feste Zuordnung von IP-Adressen zu Telefonnummern, was mit einem relativ hohen administrativen Aufwand verbunden war, und die Einschränkung der Nutzung ausschließlich auf digitale Anschlüsse (ISDN). Die Entwicklung zeigte nämlich sehr bald, daß trotz der Wachstumsraten bei ISDN der überwiegende Anteil der Benutzer aus Kostengründen bei ihren analogen Anschlüssen bleiben wollte.

Eine Reihe von Firmen auf dem Gebiet der Kommunikationstechnik haben sehr wohl die zunehmende Bedeutung des „Remote-Access“-Bereiches erkannt und entsprechende Entwicklungsarbeiten vorangetrieben. Neben den Universitäten waren in erster Linie auch ISPs (Internet Service Provider) an Lösungen

interessiert, die es erlauben, in möglichst kompakten Komponenten eine sehr große Anzahl von analogen und digitalen Kanälen nebeneinander zu betreiben.

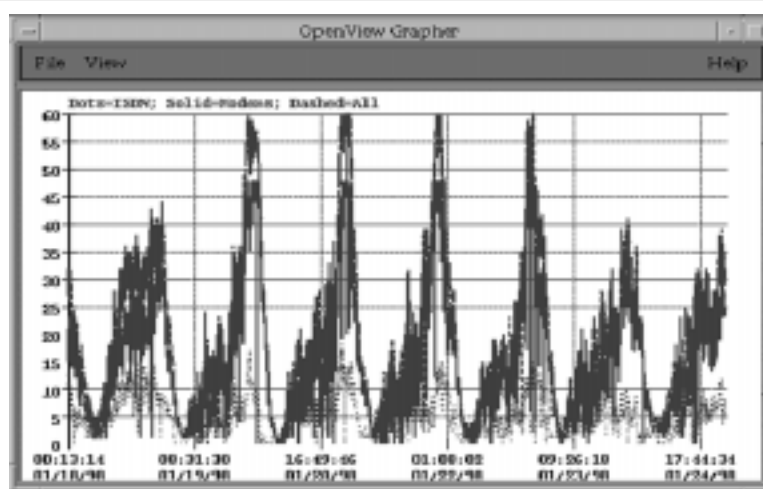
So konnte bereits im April 1996 die 1. Ausbaustufe eines Remote-Access-Routers für den Bereich der Universität in Betrieb genommen werden. Der MAX 4000 der Firma Ascend wurde über eine S_{2M} -Schnittstelle an das öffentliche Telefonnetz angeschlossen. Über diese S_{2M} -Schnittstelle können bis zu 30 Verbindungen parallel aufgebaut werden. Von den insgesamt 30 digitalen Kanälen konnten bis zu 16 analog genutzt werden. Unterstützt wurde das TCP/IP-Protokoll über das weitverbreitete PPP (Point-to-Point-Protokoll) unter Verwendung einer dynamischen Adreßvergabe. In dieser Ausbaustufe betrug die maximale „Übertragungsgeschwindigkeit“ nach der Norm V.34 für analoge Verbindungen 28.8 Kbit/s. Bei den Clients wurden unterstützt:

- Windows 3.x, Windows 95, Windows NT
- Linux
- MacOS
- (OS/2)

Nachdem der Zugang zum Hochschulnetz auch für die Studierenden geöffnet wurde, stieg die Anzahl der Benutzer bereits im Sommersemester 1996 explosionsartig an. Um dem damit verbundenen stark gestiegenen Bedarf Rechnung zu tragen und auch aus Gründen der Ausfallsicherheit wurde bereits im Herbst 1996 hochgerüstet. So wurde ein zweiter baugleicher MAX 4000 über eine separate S_{2M} -Schnittstelle mit einer eigenen Telefonnummer in Betrieb genommen. Darüber hinaus wurden die Modemboards des ersten Access-Servers hochgerüstet, so daß dann 2 Ascend-Router mit jeweils 30 digitalen Kanälen zur Verfügung standen, von denen jeweils 24 analog genutzt werden konnten.

Im Frühjahr und im Herbst 1997 folgten weitere Hochrüstungen der beiden vorhandenen Access-Server, so daß seit Ende des Jahres 1997 an den beiden Ascend-Routern jeweils 60 digitale Kanäle mit jeweils maximal 48 analogen zur Verfügung stehen. Die maximale „Übertragungsgeschwindigkeit“ betrug damals 33.600 bit/s (nach V.34+) bzw. 56.000 bit/s (nach K56flex). Schon der Sommer 1997 zeigte aber, daß auch dieser Ausbau den vorhandenen Bedarf nach Wählzugängen immer noch nicht abdecken konnte.

Abbildung 1: Die folgende Grafik zeigt die typische Auslastung eines der beiden Access-Router mit 60 Kanälen in einer Januarwoche 1998:



In den Spitzenzeiten konnten die Verbindungswünsche nicht immer zufriedenstellend erfüllt werden. Ein weiterer Ausbau wäre wieder mit hohen Investitionskosten und zusätzlichen laufenden Kosten verbunden gewesen. Die Ankündigung der Telekom AG im Frühjahr 1998 im Rahmen eines Kooperationsprojektes Studierenden und Mitarbeitern der Hochschulen den Zugang zu den Hochschulnetzen vom häuslichen PC zu erleichtern, kam daher wie gerufen. Verhandlungen mit der Telekom AG führten Ende Oktober 1998 zu einem Kooperationsvertrag zwischen der Universität Würzburg und der Telekom über die Bereitstel-

lung von Access-Routern und dazugehörigen S_{2M} -Schnittstellen an das öffentliche Telefonnetz. Mitte Januar 1999 konnte die Anfangskonfiguration des von der Telekom bereitgestellten Equipments in Betrieb genommen werden. Sie besteht aus einem MAX TNT der Firma Ascend, der mit 120 digitalen Kanälen (entspricht $4 * S_{2M}$ -Schnittstelle) am öffentlichen Telefonnetz angeschlossen ist, von denen bis zu 96 analog genutzt werden können. Der Kooperationsvertrag sieht vor, daß der Ausbau der Anschlußkapazitäten bedarfsabhängig erfolgen soll.

Tabelle 1: Die folgende Tabelle gibt einen Überblick über den aktuellen Stand der Ausstattung der Wählzugänge im Rechenzentrum.

Produkt (Hersteller)	MAX-TNT (Ascend)	1. MAX4000 (Ascend)	2. MAX4000 (Ascend)
Anzahl aller Kanäle	120	60	60
Davon digital	120	60	60
Davon analog	96	48	48
Zugangsart	PPP	PPP	PPP
Internet-Adreßvergabe	dynamisch	dynamisch	dynamisch
Unterstützte analoge Protokolle	bis V.34+, K56flex, V.90	bis V.34+	bis V.34+, K56flex
Unterstützte Betriebssysteme	Windows 3.x, Windows 95 Windows NT Linux, MacOS, (OS/2)	Windows 3.x, Windows 95 Windows NT Linux, MacOS, (OS/2)	Windows 3.x, Windows 95 Windows NT Linux, MacOS, (OS/2)
Telefonnummer	7052755	7057300	7057400

Eine *Zugangsberechtigung für Universitätsangehörige* kann in der Beratung des Rechenzentrums (Raum 1U02, Tel.: 888-5085) auf Antrag kostenfrei erteilt werden.

Eine *Zugangsberechtigung für Studenten* kann in der studentischen Beratung des Rechenzentrums (Raum 1U05, Tel.: 888-5836) auf Antrag

erteilt werden. Nähere Informationen, technische Details und Installationshilfen mit zugehörigen Softwarepaketen finden Sie auf der WWW-Seite:

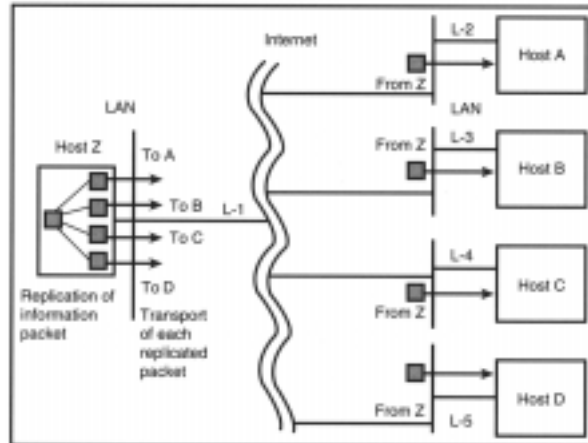
<http://www1.rz.uni-wuerzburg.de/benstud/Zugang.htm>

Was ist MBone?

Holger Fahner, Hartmut Plehn

Die Evolution des Internets von einem einfachen Netz, das hauptsächlich für File-Transfers genutzt wurde, hin zu einem voll integrierten Dienstenetz ist in den vergangenen Jahren mit rapider Geschwindigkeit fortgeschritten. Parallel zu den Entwicklungen im Bereich User-Server-Kommunikation im World Wide Web hat sich mittlerweile aber auch gerade der Bereich der User-User-Kommunikation mit verschiedenen Echtzeitanwendungen auf dem Internet zu einer Massenapplication etabliert. Applikationen, die z. B. das Telefonieren über das Internet oder das Abspielen von Musik aus entfernten Datenbanken ermöglichen, erfreuen sich steigender Beliebtheit.

Abbildung 1: Unicast-Adressierung, aus V. Kumar „M Bone“



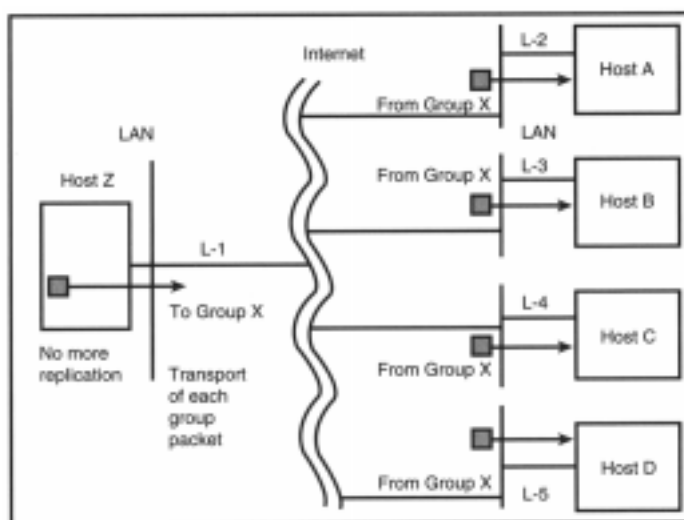
Diese Anwendungen basieren jedoch auf einfachen Punkt-zu-Punkt Verbindungen. Greifen aber viele Teilnehmer gleichzeitig auf dieselben Ressourcen zu, führt dies zu einer Überlastung von Servern, Netzen oder Routern. Die synchrone Kommunikation (also Audio- oder Videokonferenzen), bei der mehrere User beteiligt sind, wird dabei mei-

stens über einen Art Reflektor realisiert, der jedoch auch bald an seine Grenzen stößt. Diese Verfahren lassen sich deshalb nicht unbegrenzt erweitern. Um dieses Problem zu lösen, bietet das Internet-Protokoll einen besonderen Mechanismus zur Gruppenkommunikation, nämlich die Multicastadressierung.

Eine Multicastadresse bezeichnet keinen einzelnen Internetknoten oder gar eine Person, sondern eine Gruppe von Rechnern. Ein Sender schickt dabei nicht an jeden Teilnehmer

eine Kopie seiner Daten (siehe Abbildung 1), sondern sendet nur einmal an die Multicastadresse (siehe Abbildung 2).

Abbildung 2: Multicast-Adressierung, aus V. Kumar „Mbone“



Was perfekt z. B. in einem Ethernet-LAN funktioniert, bei dem alle Gruppenteilnehmer an demselben Übertragungsmedium angeschlossen sind, wird zum Problem, wenn Interessenten außerhalb des LANs an solch einer Session teilnehmen möchten. Deshalb wurden Routingverfahren entwickelt, mit deren Hilfe es möglich ist, voneinander entfernte multicastfähige LANs zu koppeln. Was zunächst sehr experimentell anging, entwickelte sich rasch zu einem weltumspannenden multicastfähigen Netz, dem Mbone (Multicast Backbone).

Diese Technologie wurde 1992 zum ersten Mal zur Übertragung der IETF Konferenz (in Bild und Ton) in größerem Maße genutzt. Der Mbone ist ein Overlay-Netzwerk, welches auf der existierenden Internet-Infrastruktur aufsetzt. Multicastpakete werden dabei von jedem Sender über eine Art virtuelle Baumstruktur zu den Gruppenteilnehmern geschickt. Die Daten werden nur an den Knoten zu denjenigen Ästen

hin vervielfacht, an denen Empfänger vorhanden sind. Somit ist gewährleistet, daß kein unnötiger mehrfacher Verkehr den gleichen Netzabschnitt durchzieht. Ein Videobild einer Konferenz-Übertragung wird dadurch z. B. nur einmal aus den USA nach Europa versendet. Hier wird es erneut weiter verteilt und erreicht jedes Land nur einmal (vorausgesetzt es sind aktive Teilnehmer vorhanden). Das Paket an jeden Zuschauer einzeln zu verschicken, hätte mit Sicherheit den Zusammenbruch der internationalen Leitungen zur Folge.

Was 1992 als Experiment begann, hat sich bis heute zu einem weltweiten komplexen Netzwerk von Multicastroutern entwickelt. Denn nur über eine Struktur wie den Mbone ist es möglich, neue Dienste wie Mehrparteien-Konferenzen, Seminare oder Lehrveranstaltungen effizient zu übertragen. Doch ist der Mbone nicht nur für Video- und Audioübertragungen interessant. Auch die gezielte

Verteilung von großen Datenmengen (z. B. Newsgroups) ist mittels Multicast mit wesentlich geringerer Netzbelastung zu realisieren.

Der MBone ist gerade dabei, seinen Experimentierstatus zu verlassen und zu einem echten Service zu werden. Das Multicasting wird vermehrt von echten Routern übernommen und

nicht mehr nur von entsprechend konfigurierten UNIX-Workstations. Ebenso erkennen auch die Internetprovider und Netzbetreiber das Potential des Multicastnetzes, und in zwei bis drei Jahren wird der MBone mit Sicherheit einen zentralen Service in der Internetlandschaft darstellen.

MBone an der Universität Würzburg

Voraussetzung für die Nutzung von MBone in einem Fachbereich der Universität ist, daß der Router, über den der Fachbereich an das Hochschulnetz angeschlossen ist, MBone unterstützt. Dies ist derzeit aufgrund der Hard- und Software-Ausstattung nicht bei allen Routern der Fall. Wenden Sie sich an Ihren Netzbeauftragten, wenn Sie MBone in Ihrem Fachbereich nutzen wollen und nicht wissen, ob Ihr Router MBone zuläßt.

Die Netzwerkkarte Ihres Rechners muß Multicastadressen verwenden können. Außerdem erfordern die Multimedia-Anwendungen über MBone im Regelfall mindestens eine Möglichkeit zur Audioausgabe und gegebenenfalls Mikrofon und Videokameras.

Wenn alle Anforderungen an das Netzwerk und Ihre Hardware erfüllt sind, können Sie MBone durch Einspielen spezieller Software ohne größeren Konfigurationsaufwand verwenden. Informationen über MBone finden Sie unter <http://www.mbone.de>.

Dort findet man Verknüpfungen zum MBone-FAQ, weiterführender Dokumentation und zu aktuellen Software-Versionen für MBone. Eine umfassende Sammlung von Software für MBone findet man z.B. im Multimedia Conferencing Applications Archive (<http://www.video.ja.net/mice/>).

Im Rahmen des Kooperationsprojektes ReLaTe (Remote Language Teaching) zum Thema Videokonferenzen wurde MBone im ersten Halbjahr 1998 länderübergreifend zwischen der University of Exeter (Dr. Gary Stringer) und der Universität Würzburg (Dr. Werner Wegstein, Peter Ruff) systematisch getestet.

Zunächst wurden auf den Endgeräten die benötigten Hardwarevoraussetzungen geschaffen und verschiedene Betriebssysteme (Windows 95/NT, Linux, Digital Unix) auf ihre Eignung für Multimedia-Videokonferenzen untersucht. Bei den darauffolgenden Tests der Netzinfrastruktur (MBone-Fähigkeit, Bandbreite) zwischen Exeter und Würzburg wurden Routing-Probleme aufgedeckt, die aber in Zusammenarbeit mit dem „Operations Desk“ des „UK Academic & Research Network“ (JANET) schnell behoben wurden.

In einem zweiten Schritt wurde die MBone-Infrastruktur mit einigen Studenten für den Einsatz im Fremdsprachenunterricht erfolgreich getestet. Weitere Anwendungen im Umfeld Teleteaching, z.B. ein Unix-Kurs für Fortgeschrittene, sind geplant.

Teile dieses Artikels stammen aus den Mitteilungen des Rechenzentrums der Universität Stuttgart und werden mit freundlicher Genehmigung des Autors übernommen.

Abbildung 1: Dr. Lewis (Bild Mitte unten) von der University of Exeter hält ein Englisch-Seminar zum Thema „English for special purposes: Computers and information processing“ über das Internet. Die Teilnehmer werden von Prof. Wegstein (Mitte oben) und Herrn Ruff (Bild links) betreut.

Die einzelnen Video-Bilder verursachen Datenströme von etwa 50 - 150 KB/s. Durch Einsatz der Multicast-Technologie wird jedes Videobild nicht zu jedem Teilnehmer extra übertragen, sondern zwischen beteiligten Bereichen des Internet nur einmal. So belastet beispielsweise das Videobild von Dr. Buckett aus Exeter die internationale Datenverbindung zwischen England und Deutschland nur mit einmalig ca. 130 KB/s, obwohl es gleichwertig an mehreren Arbeitsplätzen in Würzburg empfangen wird.



Universitätsweite Suchmaschine

Hartmut Plehn

Das Informationsangebot aller Bereiche der Universität Würzburg im World-Wide Web (WWW) umfaßt im Februar 1999 über 50000 Dokumente auf knapp 130 verschiedenen WWW-Servern. Das grundlegende Problem bei der Nutzung des Internet, aus einem Überfluß von Informationen das Gesuchte herauszufiltern, stellt sich somit auch den Benutzern, die momentan alleine monatlich etwa 80000 mal auf die Homepage der Universität zugreifen. Durch den Betrieb einer lokalen WWW-Suchmaschine möchte das Rechenzentrum diese Benutzer bei der Suche nach Informationen, die auf WWW-Servern der Universität angeboten werden, unterstützen.



Allgemeines

Die Suchmaschine folgt von der Homepage der Universität aus allen WWW-Verknüpfungen (HyperLinks) zu anderen Dokumenten auf

Rechnern innerhalb der Domain uni-wuerzburg.de und übernimmt deren Inhalte in eine Datenbank.

Unter der Adresse <http://www.uni-wuerzburg.de/htdig/search.html> oder durch Anklicken von „Schlagwortsuche“ auf der Homepage der Universität kann diese Datenbank nach Begriffen durchsucht werden. Als Ergebnis erhält man eine Liste aller WWW-Dokumente, in denen die gesuchten Worte vorkommen.

Die Datenbank mit den Inhalten aller WWW-Seiten der Universität wird immer in der Nacht von Samstag auf Sonntag neu erzeugt. Dokumente, bei denen die WWW-Adresse darauf hindeutet, daß die enthaltenen Informationen dynamisch erzeugt werden (Fragezeichen, *.cgi oder /cgi-bin/*), werden dabei nicht berücksichtigt.

Suchmaschine

Als Suchmaschine dient die Software **ht://Dig**, die an der San Diego State University speziell für den lokalen Einsatz an einer Universität entwickelt wurde. Sie ist nicht als Ersatz für professionelle, Internet-umfassende Suchmaschinen wie Lycos, Infoseek, Webcrawler oder AltaVista gedacht.

Suchbegriffe können logisch verknüpft werden, aus technischen Gründen ist aber aufgrund der verwendeten Datenbank eine **Teilwortsuche** leider **nicht möglich**.

Angabe von wichtigen Schlagworten

ht://Dig versucht, die bei einer Schlagwortsuche gefundenen WWW-Seiten in der Reihenfolge ihrer „Wichtigkeit“ auszugeben. Dabei werden z. B. das Erstellungsdatum der Seite und der Kontext des Schlagwortes innerhalb der Seite (Überschrift, Position usw.) berücksichtigt.

Es ist möglich, über die Angabe von Schlagworten im Kopf einer HTML-Seite zu erzwingen, daß diese Seite bei der Suche nach einem der angegebenen Schlagworte eine besonders hohe Priorität erhält.

Die Schlagworte werden dabei über ein sogenanntes „HTML-Meta-Tag“ namens „htdig-keywords“ spezifiziert, z. B. würde die Zeile

```
<META NAME="htdig-keywords"
CONTENT="Vorlesungsverzeichnis Vorlesungen">
```

im Header einer Seite (zwischen <HEAD> und </HEAD>) dazu führen, daß diese Seite bei der Suche nach „Vorlesungsverzeichnis“ oder „Vorlesungen“ sehr weit oben genannt wird.

Dagegen kann das HTML-Meta-Tag „htdig-noindex“ dazu verwendet werden, eine bestimmte Seite bei der Erstellung der Datenbank ignorieren zu lassen. Beispiel:

```
<META NAME="htdig-noindex">
```

Meldung von zu aktualisierenden Seiten

Die Suchmaschine **ht://Dig** bietet außerdem die Möglichkeit, eine Meldung an den Autor einer Seite abzusetzen, falls ein bestimmtes Datum überschritten ist. Dies kann für Seiten, die immer wieder aktualisiert oder nach einer bestimmten Zeit entfernt werden müssen, sehr hilfreich sein.

Über die HTML-Meta-Tags „htdig-email“, „htdig-email-subject“ und „htdig-notification-date“ wird gesteuert, an wen mit welchem Betreff nach welchem Datum eine Nachricht geschickt wird. Beispiel (diese Zeilen müssen sich zwischen <HEAD> und </HEAD> befinden):

```
<META NAME="htdig-email"
CONTENT="webmaster@rz.uni-wuerzburg.de">
```

```
<META NAME="htdig-email-subject"
CONTENT="WWW-Seite muss aktualisiert werden!">
```

```
<META NAME="htdig-notification-date"
CONTENT="2/25/2000">
```

Hier würde bei jeder Aktualisierung der htdig-Datenbank ab dem 26. Februar 2000 eine Mail an webmaster@rz.uni-wuerzburg.de mit der Subject-Zeile „WWW-Seite muss aktualisiert werden!“ und natürlich der URL der betreffenden Seite geschickt werden.

Liste von falschen Links

Die Suchmaschine **ht://Dig** liefert eine Liste aller nicht verfügbaren Hypertext-Verknüpfungen innerhalb der Domäne uni-wuerzburg.de. Diese Liste ist im WWW unter <http://www.uni-wuerzburg.de/htdig/rundig.out> verfügbar, um

Betreibern von dezentralen WWW-Servern und sonstigen Anbietern von Informationen im WWW eine Hilfestellung bei der Pflege der eigenen Hypertext-Links zu bieten.

Fujitsu VPP300/2

Peter Dieterich

Vor etwa einem Jahr wurde im Rechenzentrum ein neuer Hochleistungsrechner als Ablösung für den sechs Jahre alten Vektorrechner CRAY Y-MP EL installiert. Das neue System, ein mit zwei Prozessoren ausgestatteter Vektorparallelrechner der Firma Fujitsu, stellt den momentan leistungsfähigsten Rechner des Rechenzentrums dar. Dieser Artikel gibt einen kurzen Überblick zu den Entwicklungen im Bereich "Highperformance Computing" und zeigt die Besonderheiten des neuen Vektorparallelsystems auf.



Einleitung

Numerische Simulationen nehmen bei der Untersuchung von vielen modernen Forschungsproblemen eine zentrale Rolle ein. Häufig können experimentelle Resultate erst mit Hilfe von aufwendigen Modellen und darauf aufbauenden Computersimulationen interpretiert werden. Zudem ermöglichen Computersimulationen schnellere Entwicklungsprozesse, beispielsweise in der Autoindustrie oder beim Entwurf von Chemieanlagen.

Um möglichst realitätsnahe Aussagen aus den Simulationen zu gewinnen, sind immer komplexere Modelle erforderlich, deren numerische Untersuchung eine immer höhere Rechenleistung bedingt. Beispiele hierfür sind Modelle stark wechselwirkender klassischer und quantenmechanischer Systeme, die zur Beschreibung neuartiger Materialien wie der Hochtemperatursupraleiter untersucht werden. Auch der Bereich der Strömungssimulationen, insbesondere das Studium der Turbulenz sowie

die zusätzliche Kopplung von Strömungen an Verbrennungsprozesse, stellt extreme Anforderungen an die Hochleistungsrechner, da hier nichtlineare Phänomene auf weit auseinanderliegenden Zeitskalen in großen Systemen untersucht werden müssen. Ähnliche Anforderungen ergeben sich bei der Simulation von Ökosystemen oder bei der Untersuchung von molekular- und zellbiologischen Vorgängen.

Auch wenn heute ein großer Anteil des Rechenzeitbedarfs bereits durch leistungsfähige Workstations und Personal Computer abgedeckt wird, kann bei vielen Forschungsprojekten nicht auf Hochleistungsrechner in Form von Vektor- und Parallelsystemen verzichtet werden. Bei den oben genannten Beispielen sind sowohl die Anforderungen an die Rechenleistung als auch den Hauptspeicher des Rechners so hoch, daß diese nur von speziellen Hochleistungssystemen erbracht werden können.

In Bayern stellt das Leibniz-Rechenzentrum (LRZ)¹ in München entsprechende Hochleistungskapazitäten für die bayerischen Forschungseinrichtungen zur Verfügung. Ergänzt werden diese Möglichkeiten durch Hochleistungsrechner, die in den Rechenzentren einiger bayerischer Hochschulen betrieben werden.

Der neue Vektorparallelrechner Fujitsu² VPP300/2 des Rechenzentrums der Universität Würzburg wurde in Abstimmung mit den am Leibniz-Rechenzentrum vorhandenen Systeme-

men beantragt. Der am LRZ leistungsfähigste Rechner Fujitsu VPP700/52 mit 52 Vektorprozessoren ist vollständig kompatibel zum System des Rechenzentrums der Universität Würzburg. Somit kann der lokale Rechner sowohl als Entwicklungsumgebung vor Ort als auch für kleinere bis mittlere Produktionsläufe verwendet werden.

¹ Siehe <http://www.lrz-muenchen.de/services/compute/>

² Vertrieb, Wartung und technische Unterstützung der Fujitsu-Systeme erfolgen durch die Firma Siemens.

Architektur von Höchstleistungsrechnern

Höchstleistungsrechner zeichnen sich sowohl durch eine sehr hohe Rechenleistung als auch durch große Hauptspeicherkapazitäten sowie schnelle Plattensysteme aus. Während der Prozessor die Rechenschritte ausführt, müssen die entsprechenden Daten im Hauptspeicher bereitstehen. Besonders bei Parallelrechnern kommt einer guten "Kommunikation" zwischen Prozessoren und (oftmals verteiltem) Hauptspeicher eine besondere Bedeutung zu. Man unterscheidet grob vereinfachend zwei Fälle (dazwischen existieren Vermischungen der beiden Extreme):

1. Jeder Prozessor ist mit einem eigenen Hauptspeicher ausgestattet.
2. Die Prozessoren nutzen einen gemeinsamen Speicher. Dies wird auch als SMP-Architektur (shared memory processing) bezeichnet.

Im wissenschaftlichen Simulationsbereich ist für den Anwender eines der wichtigsten Leistungskriterien die Anzahl der möglichen Rechenoperationen, die das System beispielsweise pro Sekunde durchführen kann. Dieser Wert wird meist als MegaFlops/s angegeben.

Ein Wert von 500 MegaFlops/s bedeutet, daß pro Sekunde 500 Millionen Rechenschritte ausgeführt werden können.

Neben der sogenannten Peakperformance, d.h. der theoretisch erreichbaren Spitzenleistung, geben Benchmarkresultate die typischen Werte wieder, die in Standardapplikationen erreicht werden. Die in diesem Zusammenhang bekanntesten Größen sind die Linpack³- und SPEC⁴-Benchmarks. Während der etwas ältere Linpack-Benchmark typische Probleme der linearen Algebra behandelt, besteht die SPEC-Benchmark-Suite aus einer größeren Anzahl typischer moderner Anwendungen. Diverse Benchmarkresultate sind unter

<http://performance.netlib.org/performance/>

zusammengestellt und können einfach über eine Webschnittstelle durchsucht werden.

³ Siehe <http://www.netlib.org/> und <ftp://ftp.netlib.org/benchmark/>

⁴ Siehe <http://www.spec.org/>

Tabelle 1: Vergleich von Prozessorleistungen (einige Leistungsdaten sind nicht verfügbar).

Hersteller	Modell	Typ	Takt [MHz]	Peak [MFlops/s]	Linpack [MFlops/s]		Spec95	
					n=100	n=1000	int	fp
Fujitsu	VPP-300/700	vektor	143	2200	203	1936	-	-
NEC	SX-4	vektor	125	2000	578	1944	-	-
DEC/ Compaq	Alpha-625	skalar	612	1224	287	764	18.4	20.8
	Alpha DS20	skalar	500	-	-	-	27.7	58.7
SGI	Mips R10000	skalar	195	390	114	334	11.0	17.0
	Mips R10000	skalar	250	500	-	409	13.6	20.3
Sun	UltraSparcII	skalar	336	672	154	461	-	-
	Ultra60	skalar	296	-	-	-	13.2	18.4
IBM	RS6000/397	skalar	160	640	315	532	8.62	26.6
	RS6000/F50	skalar	332	1328	-	517	14.4	12.6
	RS6000/43P	skalar	200	-	-	-	13.1	30.1
HP	C240	skalar	236	944	197	667	17.3	25.4
	C360	skalar	367	-	-	-	26.0	28.1
Intel	Pentium Pro	skalar	200	200	38	-	8.20	6.21
	Pentium Pro	skalar	333	-	-	-	12.8	9.14
	Pentium II	skalar	450	-	-	-	17.2	12.7

Einige Resultate für sehr unterschiedliche Prozessorarchitekturen sind in Tabelle 1 gegenübergestellt. Wie die Tabelle zeigt, erreichen die sogenannten Vektorprozessoren die höchsten MFlop-Leistungswerte, obwohl deren Taktfrequenz (d. h. die Anzahl der pro Sekunde ausgeführten Schritte) deutlich niedriger liegt als die der meisten skalaren Prozessoren. Der Unterschied erklärt sich aus der Architektur der Vektorprozessoren: Diese können pro Takt mehrere Rechenoperationen ausführen - ähnlich dem Addieren von zwei Vektoren -, wohingegen die sogenannten skalaren Prozessoren meist nur ein oder zwei Operationen pro Takt bearbeiten können.

Werden extrem hohe Rechenleistungen gefordert, so ist dies nicht mehr mit einem Prozessor alleine zu realisieren. Vielmehr werden viele Prozessoren über schnelle Kommunikationswege miteinander gekoppelt. Die resultierenden Architekturen unterscheiden sich im Typ der eingesetzten Prozessoren (skalar oder vektor) sowie in der Art des Memory-Zugriffs (shared, distributed, virtuell shared oder Kombinationen davon).

In Tabelle 2 sind einige der heute aktuellen Architekturen miteinander verglichen. Während massiv parallele Realisierungen, welche aus einer hohen Anzahl von Prozessoren

Tabelle 2: Übersicht einiger installierter Hochleistungsrechner und deren Linpack-Leistung Rmax in GigaFlops/s (1 GFlop/s = 1000 MFlops/s)
Quelle: TOP500-Liste, Stand Juni 1999

Modell	Installationsform und Ort	Rmax [GFlops/s]	TOP500 Position
Intel ASCI Red	Sandia National Labs, Albuquerque USA, 9472 Prozessoren, 3154 GFlops/s Peak, massiv parallel	2121	1
SGI ASCI Blue Mountain	Los Alamos National Laboratory USA, 6144 Prozessoren, 3072 GFlops/s Peak, massiv parallel	1608	2
CRAY T3E1200 SGI	Government USA, 1084 Prozessoren, 1300 GFlops/s Peak, massiv parallel	892	3
CRAY T3E1200 SGI	Deutscher Wetterdienst in Offenbach, 540 Prozessoren, 648 GFlops/s Peak, massiv parallel	448	15
CRAY T3E SGI	Max-Planck-Gesellschaft MPI/IPP Garching, 812 Prozessoren, 487 GFlops/s Peak, massiv parallel	355	19
CRAY T3E SGI	HWW/Universität Stuttgart, 540 Prozessoren, 486 GFlops/s Peak, massiv parallel	341	22
Fujitsu VPP700	ECMWF Reading, UK, Europäischer Wetterdienst, 116 Prozessoren, 255 GFlops/s Peak, vektorparallel	213	40
Fujitsu VPP700	LRZ München, 52 Prozessoren, 114 GFlops/s Peak, vektorparallel	106	66
IBM SP PC604	BASF Ludwigshafen, 256 Prozessoren, 170 GFlops/s Peak, massiv parallel	79	83

(typischerweise einige hundert) mit lokalem Hauptspeicher bestehen (beispielsweise das Intel ASCI Red Projekt)⁵, momentan die bei weitem höchste Peakleistung erreichen, vertrauen Anwender häufig noch auf die leichter und effizienter zu handhabenden Vektorrechnerarchitekturen. Diese setzen sich aus einer moderateren Anzahl (etwa kleiner als hundert) von leistungsfähigen Vektorprozessoren zusammen, wobei teilweise ein globaler Hauptspeicherzugriff möglich ist.

Die Spitze der TOP500-Liste wird klar von den massiv parallelen Systemen dominiert, wobei Vektorrechner aufgrund deren hoher Effektivität (reale Leistung im Vergleich zur Peak-

Performance) auch weiterhin eine wichtige Rolle spielen. Prinzipiell vertraut man in Japan noch stärker den Vektorparallelrechnern, wohingegen in Amerika massiv parallele Architekturen mit Standardprozessoren favorisiert werden. Europa nimmt hier eine Mittelstellung ein.

Eine genauere Analyse der Trends kann mit Hilfe der TOP500-Liste durchgeführt werden. Diese wird zweimal pro Jahr veröffentlicht⁶ und nennt die 500 weltweit stärksten Rechner.

⁵ <http://www.sandia.gov/ASCI/>

⁶ Siehe <http://www.top500.org/>

Hard- und Software des Fujitsu-VPP300/2

Bei der Fujitsu VPP300-Serie handelt es sich um Rechner mit starken Vektorprozessoren, die (im Gegensatz zu den früheren CRAY-Vektorsystemen) in preiswerter CMOS-Technologie gefertigt sind. Die Kommunikation zwischen den Prozessoren erfolgt über ein schnelles Matrixnetzwerk (Crossbar); vergleiche Abbildung 1. Der VPP300 kann bis auf 16 Prozessorelemente (PE) mit einer Leistung von 2.2 GFlops/s und 2 GB Memory pro Prozessor ausgebaut werden. Die Maschine des Rechenzentrums besitzt 2 PEs. Der Prozessor PE0 übernimmt dabei zentrale Betriebssystemaufgaben sowie Ein- und Ausgabefunktionen (I/O). Auf allen anderen PEs läuft eine reduzierte Version des Betriebssystems. Auch können weitere PEs I/O-Aufgaben übernehmen.

Den etwas genaueren Aufbau des VPP300 zeigt Abbildung 1. Eine Einheit des VPPs besteht dabei aus einem PE, dem Hauptspeicher (MEM) und einer sogenannten DTU (data transfer unit), welche die Kommunikation mit den anderen PEs über den Crossbar übernimmt. Dieser ist vor allem bei Systemen mit

einer großen Anzahl von PEs wichtig, da die einzelnen PEs über den Crossbar unabhängig miteinander kommunizieren können. Die Bandbreite beträgt dabei in einer Richtung 570 MB/s pro Prozessor.

Das Prozessorelement besteht aus zwei Einheiten: einer Vektoreinheit (VU) und einem Skalarprozessor (SU). Die Skalareinheit erreicht eine Spitzenleistung von 275 MFlops/s. Die wesentliche Leistung aber liefert die Vektoreinheit mit 2200 MFlops/s. Voraussetzung zum Ausnutzen der Vektorleistung ist jedoch ein Programm, das gut vektorisierbar ist und so die Vektorarchitektur ausnutzen kann.

In der Regel ermöglichen heute die herstellereigenen Compiler eine recht gute automatische Vektorisierung, wobei der Benutzer einige grundlegende Richtlinien, beispielsweise zur Organisation von Schleifen, beachten sollte. Zudem ist die Verwendung der mitgelieferten Programmbibliotheken für numerische Standardverfahren zu empfehlen.

Abbildung 1: Architektur des Fujitsu VPP/300 und VPP/700

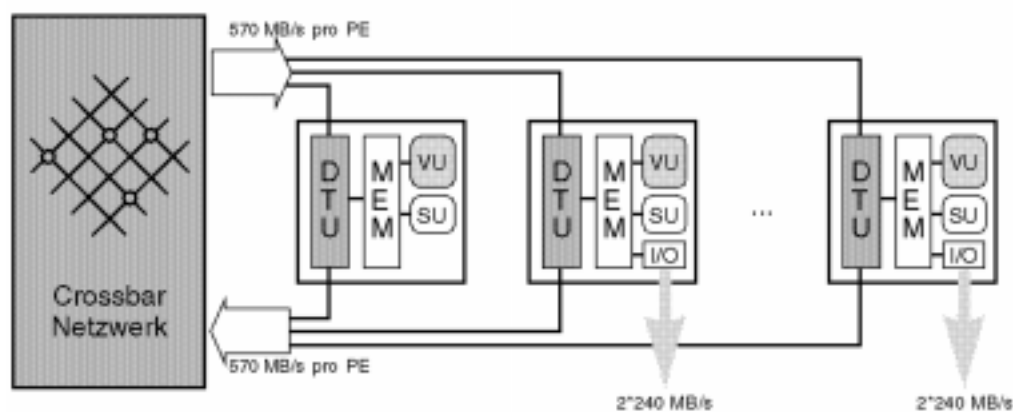
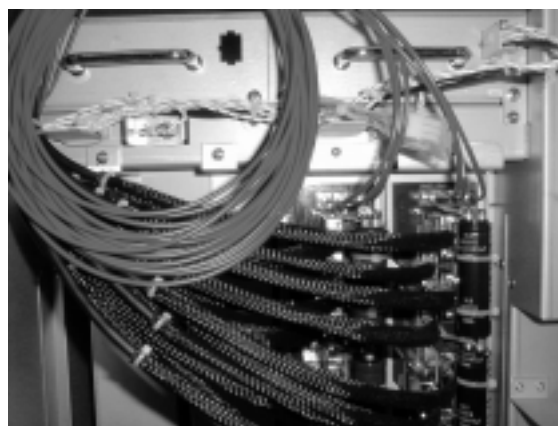
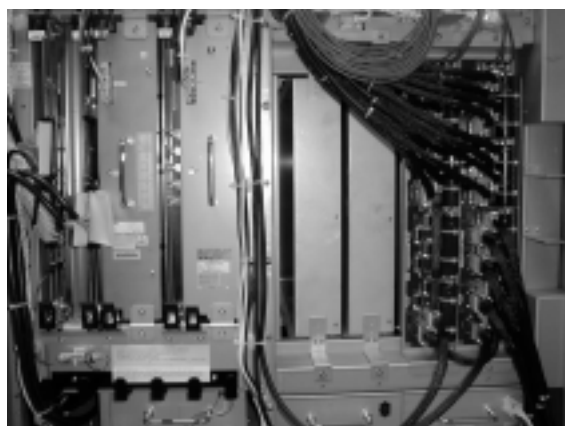


Abbildung 2: Inside VPP300 - klassische, nicht-spektakuläre Großrechner-technologie



Als Software stehen auf dem Würzburger VPP die typischen Compiler (Fortran 77 und 90, C sowie C++) zur Verfügung. Ergänzt wird dies durch numerische Bibliotheken wie die Fujitsu Scientific Library, NAG sowie einige speziellere Anwendungssoftware, beispielsweise Gaussian 94/98. Der Rechenbetrieb wird ähnlich wie bei der CRAY-EL über ein Batchsystem (Network Queuing System - NQS) abgewickelt.

Diverse Tips und Hinweise zur Nutzung des neuen Rechners und insbesondere zu Fragen der Vektorisierung wurden in einem Kurs Ende April 1998 gegeben, der von der Firma Siemens im Rechenzentrum durchgeführt wurde. Zudem sollte der Kurs Anwendern helfen, die bereits Erfahrungen auf anderen Vektor- und Parallelarchitekturen hatten, auf den Fujitsu VPP umzusteigen.

Weitere Informationen zur Nutzung

Weitere Informationen zur VPP-Serie und viele Links zu Anwendungen und zur Programmierung finden sich auf dem LRZ-WWW-Server:

<http://www.lrz-muenchen.de/services/compute/vpp/>

Die komplette VPP-Dokumentation und aktuelle Hinweise zur Nutzung des Würzburger

Rechners finden sich auf dem WWW-Server des Rechenzentrums:

<http://www.rz.uni-wuerzburg.de/vpp/>

Ein Account für den Rechner kann bei der Beratung des Rechenzentrums beantragt werden.

Literaturhinweise

[1] High Performance Computing, Kevin Dowd, 1998 (2. Auflage), O'Reilly & Assoc.

[2] Computer Architecture: A Quantitative Approach, John L. Hennessy und David A. Patterson, 1996 (2. Auflage), Morgan Kaufmann