

# Anwendung von Datenreduktionsverfahren in mobiler Umgebung

Studienarbeit

Universität Rostock, Fachbereich Informatik



vorgelegt von

Stefan Audersch

geboren am 08. April 1976 in Rostock

und

Christian Dethloff

geboren am 10. März 1976 in Waren (Müritz)

Betreuer: Prof. Dr. Andreas Heuer  
Dipl. Ing. Astrid Lubinski



# Aufgabenstellung

## **Stefan Audersch**

Nutzer sind zunehmend mobil und nicht an einen Arbeitsplatz und ein Gerät gebunden. Erstes Ziel jedes mobilen Nutzers ist die Realisierung von Kommunikation. Eine besonders mobile Umgebung findet man bei der Nutzung von mobilen Geräten, die jedoch spezielle Probleme mit sich bringen, die nicht den Nutzern überlassen bleiben sollten. So sind mobile Ressourcen begrenzt. Dies betrifft insbesondere die Bildschirmgröße, den Speicher und die Energieressourcen. Darüber hinaus stellen umfangreiche Datenübertragungen per Luftschnittstelle immer noch einen bedeutenden Kostenfaktor dar. In dieser Studienarbeit soll eine Applikation auf einem mobilen Gerät implementiert werden. Anfragen werden lokal interpretiert und greifen serverseitig auf vorbereitete Informationen in einer Datenbank, also bereits nach unterschiedlichen Kriterien reduzierbare, zu. Die Ergebnisse erfahren dann eine weitere Abstraktion bzw. Wichtung und können vom Nutzer beliebig detailliert werden. Folgende Aufgaben sind in dieser Studienarbeit konkret durchzuführen:

- Implementierung einer Client-Server-Architektur für eine mobile Umgebung und
- Zugriff auf in einer Datenbank verwalteten Informationen.

## **Christian Dethloff**

Mobile Geräte beziehungsweise mobile Links sind immer noch ressourcenbegrenzt. Eine Möglichkeit, das Ungleichgewicht zwischen den von lokal verfügbaren oder angefragten Daten benötigten und den verfügbaren Ressourcen zu mildern, besteht darin, den Umfang der Daten mit möglichst geringem Semantikverlust zu reduzieren. Die Reduktion kann sowohl die Objektgröße als auch die Objektanzahl betreffen. Aus der Literatur sind eine Vielzahl von Datenreduktionstechniken bekannt, die für unterschiedliche Zwecke wie statistische Zwecke, Suchmaschinen im WWW, Data Mining, oder auch Schätzungen bei der DB-Optimierung entwickelt und verwendet werden. In dieser Studienarbeit soll die Nutzbarkeit existierender Verfahren für den mobilen Einsatz bewertet werden. Von den als geeignet erkannten Datenreduktionsverfahren soll eines ausgewählt und auf mobile Anforderungen angepasst werden. Eine Testimplementierung zur Reduktion auf einem mobilen Gerät unter Nutzung einer Funkverbindung soll die Nutzbarkeit des Konzeptes zeigen.

# Inhaltsverzeichnis

Aufgabenstellung.....	3
Inhaltsverzeichnis.....	4
1. Einleitung.....	6
1.1. Motivation .....	6
1.2. Zielstellung und Vorgehensweise.....	6
1.3. Anwendungsszenario – elektronisches Börseninformationssystem <i>eBIS</i> .....	6
1.3.1 Der Neue Markt – ein Segment der Deutschen Börse.....	7
1.4. Gliederung.....	8
2. Konzepte zur Implementierung eines Börseninformationssystems für die mobile Umgebung .....	9
2.1. Anwendungsspezifische Konzepte.....	9
2.2. Technische Konzepte.....	12
3. Datenreduktion .....	17
3.1. Histogramme.....	17
3.2. Clustering-Techniken.....	24
3.3. Dimensionsreduktion für Data Mining.....	27
4. Evaluation von Hard- und Software .....	31
4.1. Das mobile Endgerät .....	31
4.1.1 Der Palm.....	33
4.2. Kommunikation zwischen Server und Client .....	34
4.3. PalmOS und Java .....	35
4.3.1 Die Java 2 Micro Edition .....	36
4.4. Datenbank auf dem Palm .....	40
4.5. Java 2 auf dem Server .....	42
4.6. Datenbank auf dem Server.....	43
5. Die Teilkomponenten des Systems.....	44
5.1. Wrapper .....	44
5.2. Vorreduktion.....	48
5.3. DB auf dem Server + DB-Schema .....	61
5.4. Anfragebearbeitung und profilabhängige Reduktion.....	64

5.5. Anfragebearbeitung auf dem Client – Aktualisierung der Börseninformationen .....	70
5.6. Kommunikationsschnittstellen auf dem Server und Client .....	73
5.7. GUI auf dem Palm.....	75
5.8. Profilverwaltung auf dem Palm .....	79
5.9. Informationscache auf dem Palm .....	82
6. Ergebnisse der Datenreduktion .....	85
7. Zusammenfassung .....	87
Literaturverzeichnis.....	89
Abbildungsverzeichnis.....	91
Anhang A.....	93
Anhang B.....	96
Anhang C.....	97

# 1. Einleitung

## 1.1. Motivation

Galten vor einiger Zeit mobile Geräte noch als Exoten, so sind sie heutzutage kaum noch aus dem Bereich der Informationstechnik wegzudenken. Gerade der Personal Digital Assistant (PDA) trägt durch seine rasche Entwicklung maßgeblich hierzu bei. Waren damals Nutzer noch an einen Arbeitsplatz und ein stationäres Gerät gebunden, so ist heutzutage der Wunsch nach Mobilität entscheidend gewachsen. Nicht nur an dem fest installierten PC, sondern auch „mobil unterwegs“ möchte man mit den neusten Informationen versorgt werden, hierbei jedoch nicht auf Bedienungskomfort und Informationsqualität verzichten. Erster Schritt in diese Richtung ist die Realisierung von Kommunikation. Doch schon hier treten erste Probleme auf, die besonderer Rücksicht bedürfen. Oft steht nur eine Datenübertragung per Luftschnittstelle zur Verfügung, die bei dem derzeitigen GSM-Standard (Global Mobile System) lediglich Übertragungsgeschwindigkeiten bis 9600 Bit/s erlaubt. Die hieraus resultierenden langen Kommunikationszeiten stellen für den Nutzer einen besonderen Kostenfaktor dar. Eine Verbesserung versprechen zwar Neuerungen, wie GPRS (General Packet Radio Service), HSCSD (High Speed Circuit Switched Data) oder UMTS (Universal Mobil Telecommunications System), doch nicht nur die Übertragungsgeschwindigkeit stellt ein besonderes Problem der Mobilien Nutzung dar. Der Nutzer möchte kleine, leichte mobile Geräte verwenden. Diese, auch „Westentaschen PC“ genannten Geräte, sind vor allen Dingen PDA's, wie beispielsweise Palm, Pocket PC's mit Windows CE oder die auf dem EPOC-Betriebssystem basierenden Geräte von Psion. Diese für den mobilen Einsatz ausgelegten Geräte sind jedoch in ihren Ressourcen begrenzt. Dies betrifft insbesondere die Bildschirmgröße, den Speicher und die Energieressourcen. Bei der Beachtung der genannten Beschränkung ist die logische Schlussfolgerung zur Realisierung eines mobilen Informationssystems, die ressourcenoptimierte Implementierung einer Applikation für das mobile Endgerät sowie eine anwendungsspezifische, serverseitige Reduktion der darzustellenden Informationen.

## 1.2. Zielstellung und Vorgehensweise

Im Rahmen dieser Studienarbeit werden unterschiedliche, existierende Datenreduktionsverfahren hinsichtlich ihrer Tauglichkeit für ein speziell ausgewähltes, mobiles Anwendungsszenario geprüft. Hierzu wird zunächst ein Szenario gewählt und für das daraus resultierende System Anforderungen spezifiziert. Umfangreiche Analysen werden Aufschluss über die gewählten Parameter und die gestellten Anforderungen geben. Mit diesen Kenntnissen kann eine geeignete Datenreduktionsmethode gewählt und entsprechend den Systemvoraussetzungen angepasst werden. Eine prototypische Implementierung gibt letztendlich Aufschluss über die gesammelten Ergebnisse.

## 1.3. Anwendungsszenario – elektronisches Börseninformationssystem *eBIS*

Für die Studienarbeit wurde als Anwendungsszenario ein elektronisches Börseninformationssystem (eBIS) für die mobile Umgebung gewählt. Börsen-

informationen liegen in ausreichender Menge und frei zugänglich im Internet vor. Sie bestehen aus numerischen und nicht-numerischen Werten und eignen sich daher für die Anwendung von Datenreduktionsverfahren.

### **1.3.1 Der Neue Markt – ein Segment der Deutschen Börse**

Seit 1997 ist die Börse auch für Kleinanleger interessant. Die Privatisierung der Deutschen Telekom begann mit einer Platzierung, sprich der Ausgabe, von Aktien aus dem Staatsbesitz im zweistelligen Prozentbereich. Ziel des Telekommunikationsunternehmens war es, den "kleinen Mann auf der Straße" mit Werbekampagnen zu erreichen, um ihn somit zum Kauf von Aktien zu motivieren; mit Erfolg. Viele deutsche Bürger wurden über Nacht zu Aktionären. Der Aktienboom nahm seinen weiteren Lauf. Am 10. März 1997 wurde der Neue Markt, ein neues Segment der Deutschen Börse AG, in Frankfurt gegründet. Er bietet wachstumsstarken und innovativen Unternehmen neue Chancen der Eigenkapitalaufnahme und zeichnet sich durch eine hohe Transparenz im marktwirtschaftlichen Sinne aus. Die Unternehmen bieten diese durch Veröffentlichung von Unternehmensberichten, wie beispielsweise Emissionsprospekte, Kooperations-, Fusions- und Quartalsberichte. Ein weiteres Merkmal des Neuen Marktes ist die aktive Vermarktung von Seiten der Deutschen Börse, durch Liquidität und einer europaweiten Vernetzung. Emittenten des Neuen Marktes sind kleine, mittlere, junge und innovative Wachstumsunternehmen, die sich durch ein hohes Umsatz- und Gewinnwachstum pro Jahr im zweistelligen, manchmal sogar dreistelligen Prozentbereich auszeichnen sowie in zukunftsweisenden, aber auch in traditionellen Branchen bei Produkt-, Prozess- oder Service-Innovationen tätig sind. Sie haben meist eine internationale Ausrichtung und besitzen eine hohe Bereitschaft zur aktiven Investor-Relations-Politik.

Um die derzeit 341 Unternehmen innerhalb des Neuen Marktes übersichtlicher zu gestalten, wurden diese Anfang Juni 2000 von der Deutschen Börse AG in 10 Branchen unterteilt. Die Partitionierung erfolgte in Bereiche, die stellvertretend für die enorme innovative Entwicklung der Gesellschaft in Bezug auf das Informationszeitalter stehen: Biotechnology, Financial Service, Industrials & Industrial Services, Internet, IT Services, Media & Entertainment, MedTech & Health Care, Software, Technology, Telecommunications. Für die Aktionäre wurde somit eine bessere Übersicht geschaffen, da Aktien aus einer Branche viele Ansatzpunkte zum Vergleichen liefern und es leichter fällt, die interessantesten herauszufiltern.

Gerade Aktien aus dem Internet- und Biotechnologiesektor beschäftigten in den letzten Monaten nicht nur die Großaktionäre, sondern weckten auch bei vielen Aktionären, die für gewöhnlich nur einen geringen Anteil am Aktienkapital einer Gesellschaft halten, reges Interesse.

Von Beginn an gehörte der Kleinanleger neben institutionellen Anlegern wie Banken, Investmenthäuser, Versicherungen usw. zu der Investorenzielgruppe der Neuen Marktwerte. Das lag einerseits daran, dass die Fondmanager ihren Fokus zum größten Teil auf andere Segmente des Börsenmarktes, beispielsweise DAX, MDAX oder EURO-STOXX richteten. Andererseits wollten die Großaktionäre abwarten, ob die Unternehmen den hohen Wachstumsansprüchen gerecht werden. Die Unternehmen des Neuen Marktes standen vor der prekären Aufgabe, durch gezielte Öffentlichkeitsarbeit Kleinaktionäre zu erreichen und somit für ihre ausgegebenen Aktien zu werben. In Form von Ad-Hoc-News berichten sie über anstehende Akquisitionen, Quartals-, Halbjahres-, Neunmonats- und Jahreszahlen. Diese Informationen werden über verschiedene Medien, wie z.B. einschlägige Börsenzeitungen und das Internet,

veröffentlicht. Gerade das Internet bietet eine ideale Voraussetzung zur schnellen und günstigen Verteilung von Börseninformationen.

Homepages, wie <http://www.aktiencheck.de>, [www.goingpublic-online.de](http://www.goingpublic-online.de) oder <http://www.stockworld.de> versuchen dem Kleinanleger die Welt der Börse so transparent wie möglich zu gestalten. Das erreichen sie unter anderem durch Veröffentlichung gering zeitverzögerter Kurslisten und der Publizierung vieler Analystenmeinungen, in denen Kauf-, Halte- oder Verkaufsempfehlung gegeben werden. Auf der Homepage von aktiencheck.de AG werden beispielsweise durchschnittlich 20 Analysen pro Tag präsentiert. Noch vor einigen Jahren ist der Kleinanleger nur durch den Besuch des Börsenexperten seiner Hausbank in den Genuss einer Analystenmeinung zu einem Unternehmen gekommen.

Bei all der Transparenz des Neuen Marktes ist der Aktionär zeitlich nicht mehr in der Lage, alle wichtigen Informationen zu erfassen. Die extrem hohe Menge an Börsennachrichten, die pro Tag anfallen, als auch die Intention der Entwicklung eines Börseninformationssystems für die mobile Umgebung, erfordern den Einsatz von Datenreduktion.

## **1.4. Gliederung**

Folgende Arbeit zeigt den Aufbau eines Börseninformationssystems in einer mobilen Umgebung. Im Kapitel 2 werden zunächst die zugrunde liegenden Anwendungsdaten analysiert und verschiedene technische Konzepte zur Umsetzung einer Client-Server-Architektur diskutiert. Kapitel 3 stellt unterschiedliche Datenreduktionsverfahren vor und prüft deren Einsatz innerhalb des gewählten Anwendungsszenarios. Für eine prototypische Implementierung werden im darauf folgenden Kapitel verschiedene Hard- und Software evaluiert. Darin wird unter anderem das mobile Endgerät und die dafür verwendete Entwicklungsumgebung festgelegt. Die Teilkomponenten des Börseninformationssystems, deren Aufgaben und Umsetzung werden in Kapitel 5 beschrieben. Abschließend fasst Kapitel 6 die Ergebnisse zusammen, führt eine Nachbetrachtung durch und gibt Ausblicke für die Lösung aufgetretener Probleme.

## 2. Konzepte zur Implementierung eines Börseninformationssystems für die mobile Umgebung

Ziel der Studienarbeit ist die prototypische Implementierung eines Börseninformationssystems für die mobile Umgebung unter Verwendung geeigneter Datenreduktionstechniken. Um diesem gerecht zu werden, ist es zunächst notwendig, das gewählte Anwendungsszenario genau zu untersuchen. Hierzu gehören auf der einen Seite anwendungsspezifische Gesichtspunkte, wie beispielsweise die detaillierte Analyse der zu verwendenden Anwendungsdaten und deren Herkunft, als auch andererseits technische Konzepte, bei denen unter Berücksichtigung der geforderten Systemumgebung die gewählte Struktur des Systemaufbaus skizziert wird.

### 2.1. Anwendungsspezifische Konzepte

Eine Aktie ist ein Wertpapier, welches ein Anteilsrecht an einer AG festlegt und verbrieft. Sie sind an der Börse handelbar und werden eindeutig durch einen Schlüssel gekennzeichnet. Solche identifizierenden Schlüssel sind beispielsweise die Wertpapierkennnummer (WPKN), die International Security Identification Number (ISIN) und das Kürzel (bestehend aus 3 bis 4 Zeichen).

Für den Handel mit den Aktien informieren sich die Investoren über die einzelnen Wertpapierentwicklungen. Dieses geschieht in der Regel nach unterschiedlichen Gesichtspunkten die sich in verschiedenen Informationskategorien widerspiegeln. Für den Aktienhandel haben sich dabei folgende Informationskategorien etabliert:

- Analysen
- Ad-Hoc-News
- Tops und Flops
- Schwankungen
- Neuemissionen

Diese Kategorien stellen deshalb auch für das entwickelte Börseninformationssystem die Basis dar.

In den folgenden Abschnitten werden die fünf Kategorien sowie der daraus resultierende Informationsgewinn für einen Aktionär näher erläutert:

**Analysen** sind verfasste Texte und dienen der Bewertung von Aktien. Sie sind Hilfsmittel für jeden Aktionär, der sich mit den Märkten, national oder auch international, beschäftigt. Grundsätzlich existieren drei Bewertungen, die charakteristisch für eine Analyse sind: Kaufen, Verkaufen und Halten. Die Verfasser, auch Analysten genannt, haben einen wesentlichen Einfluss auf die Kursentwicklungen der Aktien. Sie arbeiten für große Banken und Investmenthäuser, die wiederum Aktien in großen Mengen kaufen. Ein Investor, der sich für eine Aktie interessiert und diese kaufen will, kann nicht ignorieren, dass beispielsweise die Deutsche Bank dieses Papier als Verkaufsposition sieht.

Eine Analyse setzt sich grundsätzlich aus einer Überschrift (Subject), dem Erscheinungsdatum, dem Analysten und dem Text zusammen, wie in dem folgenden Beispiel dargestellt.

*Subject: Rösch interessant*

*Datum: 04.12.00*

Analyst: EURO am Sonntag

Die Experten von EURO am Sonntag sehen in der Rösch-Aktie (WKN 529140) ein interessantes Investment.

Die Produkt-Pipeline sei prall gefüllt. Bis 2003 solle "Nimos" auf den Markt kommen, ein Pflaster zur Blutzucker-Messung für Diabetiker. Der Clou daran sei, dass ein Chip die Werte kontinuierlich kontrolliere. Einstiche zur Blutentnahme seien nicht nötig. Die Tochterfirma MDC aus Neu-Ulm habe digitale Röntgen-Sensoren für Zahnarztpraxen entwickelt, mit denen die Strahlenbelastung der Patienten um bis zu 80 Prozent reduziert werden könne, so EURO am Sonntag.

Eine Vielzahl von Anbietern stellen im World Wide Web Analysen strukturiert zur Verfügung. Diese sind zum Beispiel: [www.comdirect.de](http://www.comdirect.de), [www.aktiencheck.de](http://www.aktiencheck.de) oder [www.ariva.de](http://www.ariva.de). Die für das Börseninformationssystem benötigten Analysen stammen von [www.aktiencheck.de](http://www.aktiencheck.de).

Daneben spielen **Ad-Hoc-News** ebenfalls eine wesentliche Rolle, Aktien einzuschätzen. Diese News enthalten in der Regel Quartals-, Halbjahres-, Neun-Monats- und Jahreszahlen. Aber auch über Kooperationen, Fusionen und Konkursanträge wird berichtet. Eine Ad-Hoc-News stellt die fundamentalen Gegebenheiten und Zukunftsaussichten der Unternehmen dar. Im Gegensatz zu Analysen werden diese Informationen jedoch nicht von unternehmensfremden Institutionen verfasst, sondern von den Unternehmen selbst. Die News auf den Internetseiten von [www.aktiencheck.de](http://www.aktiencheck.de) stellt die Informationsquelle für diese Kategorie dar.

Die nächste Kategorie umfasst die **Tops** beziehungsweise **Flops**. Hat ein Aktionär die Absicht eine Aktie zu kaufen, ist es für ihn wichtig, sich über die Kursentwicklung des Papiers in einem bestimmten Zeitraum zu informieren. Gehört diese Aktie vielleicht zu den Gewinnern oder Verlierern hinsichtlich der letzten Zeit? Tops sind Wertpapiere, welche durch eine positive Kursperformance ausgezeichnet sind, deren Kurse also in letzter Zeit gestiegen sind. Flops hingegen kennzeichnen Aktien mit fallenden Kursen. Ein Anleger muss sich fragen, welche Strategie er verfolgt. Ist er kurzfristig, mittelfristig oder eher langfristig an einem Wertpapier interessiert. Aus diesem Grunde gibt es für die Kategorie Tops und Flops unterschiedliche Betrachtungszeiträume: ein Tag, eine Woche, ein Monat, drei Monate, sechs Monate und zwölf Monate.

Der Anbieter [www.goingpublic-online.de](http://www.goingpublic-online.de) stellt beispielsweise Tops und Flops auf seinen Webseiten tabellarisch dar.

Kurse: Tops/Flops (Neuer Markt)					
Top des Tages			Flop des Tages		
Unternehmen	WKN	Aktuell	Vortag		% Diff.
REFUGIUM HOLDIN...	700710	2,01	1,57		+28,03
FORIS AG AKTIEN...	577580	5,60	4,70		+19,15
INFOGENIE EUROP...	530990	2,20	1,90		+15,79
GES.F.NETW.TRAL...	587980	2,88	2,55		+12,94
PSB AG F.PROGRA...	696760	17,00	15,50		+9,68
EIAY AG AKTIEN	549570	1,88	1,55		+8,39

Abbildung 2-1: Übersicht Tops des Anbieter [www.goingpublic-online.de](http://www.goingpublic-online.de)

Kurse: Tops/Flops (Neuer Markt)						
Top des Tages			Flop des Tages			
Unternehmen	WKN	Aktuell	Vortag			% Diff
CAMELOT TELE.CO...	501089	1,29	2,20			-41,36
SCM MICROSYSTEM...	909247	10,60	17,30			-38,73
MUSICMUSICMUSIC...	925475	0,62	0,70			-11,43
LD MEDIA AG AK...	622860	6,15	6,90			-10,87
MET@BOX AG AKTI...	692120	1,61	1,80			-10,56
VALOR COMPUTERI...	928731	2,64	2,90			-9,07

Abbildung 2-2: Übersicht Flops des Anbieter [www.goingpublic-online.de](http://www.goingpublic-online.de)

Die Tops- und Flopsinformationen für das entwickelte Börseninformationssystem werden jedoch nicht aus den in Abbildung 2-1 und Abbildung 2-2 dargestellten Übersichten gesammelt, sondern aus den Kurslisten des gleichen Anbieters generiert. Die Abbildung 2-3 zeigt die Darstellung einer Kursliste.

Kurse: Neuer Markt										
A-C	D-I	J-P	Q-Z	A-Z	% Sortierung					
Unternehmen	Akt.	Diff.	% Diff.	Hoch	Tief	Handels- vol. in Tsd.	Datum	Zeit		
3U TELEKOMMUNIK...	0,00	0,00	0,00	0,00	0,00	0,00	27.04.	00:00		
4MBO INTERN. EL...	18,00	-1,00	-5,26	18,10	18,00	1.232,00	27.04.	10:44		
AAP IMPLANTATE...	12,45	+0,35	+2,89	12,45	12,20	1.398,00	27.04.	11:36		
ABIT AG AKTIEN...	6,35	+0,10	+1,60	6,40	6,25	2.890,00	27.04.	11:10		
AC-SERVICE AG N...	3,95	+0,05	+1,28	3,95	3,85	350,00	27.04.	10:58		
ACG AG F.CHIPK...	25,80	-1,80	-6,62	26,75	25,45	67.723,00	27.04.	11:32		
AD PEPPER MEDIA...	2,70	-0,05	-1,82	3,10	2,70	14.540,00	27.04.	11:31		
ARCON TELEMETRY...	0,55	0,10	+1,04	0,65	0,40	3.800,00	27.04.	11:22		

Abbildung 2-3: Kursliste beim Anbieter [www.goingpublic-online.de](http://www.goingpublic-online.de)

Wie auch Tops und Flops basiert die Berechnung von **Schwankungen** der Aktien auf der Kursliste von [www.goingpublic-online.de](http://www.goingpublic-online.de). Unter einer Schwankung versteht man einen Prozentsatz, der Auskunft über den Kursanstieg und -verfall eines Wertpapiers in einem bestimmten Zeitraum gibt. Dieser ist der Quotient aus der Kursdifferenz des höchsten und niedrigsten Kurses sowie dem niedrigsten Kurs als Divisor in Bezug auf den Betrachtungszeitraum.

Aktie	Tageshoch	Tagestief	Schwankungen
A	20 €	10 €	50 %
B	100 €	98 €	2 %

Abbildung 2-4: Schwankungen

Je nachdem ob ein Investor eine spekulative oder konservative Anlagestrategie verfolgt, wählt er anhand der Schwankungen interessante Wertpapiere aus. Der spekulative, eher risikobewusste Aktionär bevorzugt Aktien mit hoher Schwankung, die im Börsenjargon auch sehr volatil genannt werden. Dagegen favorisiert der konservative Investor Aktien mit geringer Schwankung, also Aktien die nicht unter hohen Kursschwankungen leiden.

Die Tabelle in Abbildung 2-4 zeigt die berechneten Schwankungen für die Aktien A und B. Die Aktie A würde man als volatile, B als nicht volatiles Papier einstufen. Ein Aktionär kann nun selbst entscheiden, ab welchem Prozentsatz er eine Aktie als sehr schwankungsintensiv bezeichnet.

Analog zu den Tops und Flops werden auch hier verschiedene Betrachtungszeiträume zu Grunde gelegt.

Die letzte Kategorie, die hier erwähnt werden soll, sind die **Neuemissionen**. Heutzutage streben immer mehr Firmen den Gang an die Börse an. Sie erhalten vom Markt frisches Kapital, mit dem der Aufbau des Unternehmens vorangetrieben werden kann. Die damit verbundene Ausgabe der Aktien des Unternehmens nennt man Neuemission.

Charakteristische Werte für neu emittierte Aktien ist der Emissionskurs und der erste Kurs an der Börse. Der Emissionskurs spiegelt den Preis wider, zu dem die Aktien während der Zeichnungsfrist von dem Unternehmen an die Aktionäre ausgegeben werden. Die Aktien der Firma werden innerhalb der Zeichnungsfrist noch nicht an der Börse gehandelt. Der erste Preis ist hingegen der tatsächlich erste Kurs, den Aktionäre für die Papiere an der Börse zahlen müssen. Gerade die Spekulation auf einen höheren Kurs nach der Emission gegenüber dem Emissionskurs regt zum Zeichnen dieser Papiere an.

Die für diese Kategorie notwendigen Informationen leiten sich aus den Fundamentaldaten der Aktien und Kurslisten ab. Die Fundamentaldaten sowie auch die Kurslisten stellt der Anbieter [www.goingpublic-online.de](http://www.goingpublic-online.de) im World Wide Web zur Verfügung.

## 2.2. Technische Konzepte

Wie bereits im Kapitel 2.1 Anwendungsspezifische Konzepte ab Seite 9 angesprochen, ist es notwendig, die Börseninformationen von verschiedenen Informationsanbietern aus dem World Wide Web zu sammeln. Die gesammelten Daten können nicht in dieser Form an die mobilen Endgeräte weitergeleitet werden, da die Geräte nur über geringe Ressourcen verfügen. Dieses betrifft speziell deren meist nicht mehr als 32 MByte umfassenden Speicher und deren Prozessor mit in der Regel Taktfrequenzen von ca. 20 MHz. Weiterhin muss berücksichtigt werden, dass die mobilen Geräte nur zeitweise mit dem Netz verbunden sind. Ziel des Systems soll es sein, dem Endanwender jede - für ihn wichtige Information - über das aktuelle Geschehen am Neuen Markt auf seinem PDA zu präsentieren. Der allgemeine Aufbau des Systems setzt sich daher folgendermaßen zusammen:

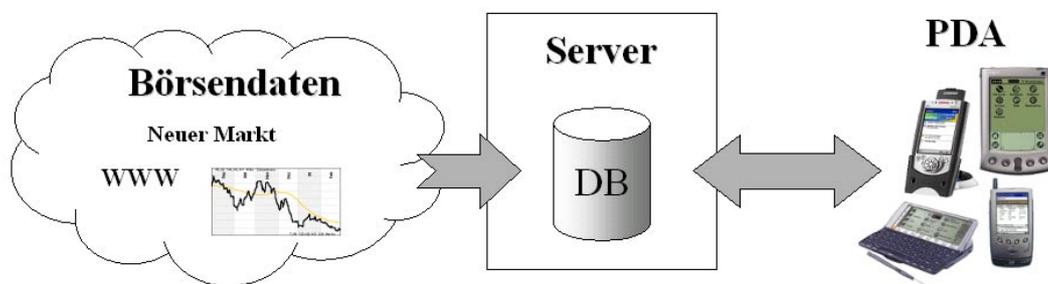


Abbildung 2-5: allgemeiner Systemaufbau des Börseninformationssystems

Der Applikationsserver sowie die Endanwendung auf dem PDA bilden hierbei das Kernstück des entwickelten Systems. Die Aufgaben des Servers sind:

### 1. Börseninformationen sammeln

Mit Hilfe eines Wrappers sammelt der Server die notwendigen Daten von den verschiedenen Anbietern. Der Wrapper ist hierzu an die unterschiedlichen Webseiten angepasst.

### 2. Informationen archivieren

Die gesammelten Informationen werden für eine bestimmte Zeit in einer Datenbank gespeichert. Hierfür gibt es zwei Gründe: Zum einen ist der Endanwender nicht nur an gerade aktuellen Informationen des Neuen Markt interessiert, sondern möchte auch über Ereignisse, die eventuell einige Zeit zurückliegen (Hinweis: Das mobile Endgerät ist nicht jederzeit online) aber

dennoch wichtig sind, informiert werden. Zum anderen werden die historischen Daten sowohl zur Berechnung von Informationen als auch zur Bewertung einer Information herangezogen.

### **3. Vorhandene Daten geeignet reduzieren**

Die gesammelten Börseninformationen werden geeignet reduziert, um sie entsprechend an die Ressourcen des PDA anzupassen.

### **4. Bereitstellung der reduzierten Daten für den PDA, auf dem die Endanwenderapplikation läuft, mit Hilfe einer geeigneten Schnittstelle**

Der Server stellt eine Kommunikationsschnittstelle für den Personal Digital Assistant bereit. Über diese baut der PDA eine Verbindung mit dem Server auf und lädt die Informationen herunter.

Die Client-Applikation auf dem PDA stellt die Benutzerschnittstelle für den Endanwender dar. Deren Aufgaben sind:

#### **1. Bereitstellung einer Kommunikationsschnittstelle zum Server**

Über die Schnittstelle baut der PDA eine Verbindung zum Server auf, um mit diesem zu kommunizieren.

#### **2. Darstellung der Informationen**

In einem geeigneten GUI werden die vom Server zur Verfügung gestellten Börseninformationen entsprechend ihrer unterschiedlichen Kategorien dem Benutzer übersichtlich und verständlich dargestellt.

Durch den allgemeinen Aufbau des Systems ist die Arbeitsweise und die daraus resultierende Herangehensweise zur Implementierung des Servers sowie der Endanwenderapplikation nicht genau vorgegeben. So existieren unterschiedliche Ansatzpunkte zur Zusammenarbeit von Server und Client als auch zur Realisierung der Datenreduktion. Diese werden im folgenden Abschnitt näher dargestellt.

Die Kernfrage zur Realisierung des Börseninformationssystems ist die Frage, wie der Server mit der Applikation auf dem Personal Digital Assistant zusammenarbeitet. Hierzu gibt es unterschiedliche Ansatzpunkte:

#### **Server – Browser auf dem PDA**

Der Server hat hierbei nicht nur die oben geforderten Aufgaben, sondern stellt dem Client die vollständige, reduzierte Informationsmenge zur Verfügung. Der Endanwender kann mit Hilfe des PDA und einem darauf installierten Browser durch das Informationsangebot navigieren. Denkbar ist beispielsweise die Nutzung von WAP (Wireless Applikation Protokoll). Der Server erzeugt aus den gesammelten Börsendaten WML-Seiten (Wireless Markup Language), die auf einem WAP-Server zur Verfügung gestellt werden. Ein WAP-Browser auf dem Personal Digital Assistant ermöglicht das Navigieren und die Darstellung der Informationen. Nachteilig an diesem Ansatz ist, dass der Benutzer nicht die Möglichkeit besitzt, die Börseninformationen im Offline-Betrieb zu lesen. Dies kann nur unter der Verwendung von clientseitigen, intelligenten Cache-Mechanismen bewerkstelligt werden. Weiterhin ist der Nutzer gezwungen, sich selbständig die für ihn wichtigen Informationen herauszusuchen, welches nicht das Ziel dieser Arbeit sein sollte.

#### **Server mit Profilverwaltung – Browser auf dem PDA**

Eine Verbesserung ergibt sich bei der zusätzlichen Verwendung von Profilen. Für jeden Nutzer wird ein Profil angelegt. In diesem kann der Nutzer Prioritäten für das Informationsangebot setzen und erhält bei einer Verbindung zum Server nur die von ihm priorisierten Börseninformationen. Der Ansatz zieht unweigerlich die

Integration einer Profilverwaltung am Server als auch die Notwendigkeit der Autorisierung des Benutzers am Server nach sich. Die Möglichkeit zur Darstellung der Informationen im Offline-Betrieb ist in diesem Ansatz ebenfalls nicht verwirklicht.

### **Server – Endanwenderapplikation mit Profilverwaltung und Informationscache**

Die Aufgaben der Clientsoftware sind die Kommunikation mit dem Server, die Darstellung der Börsendaten im Online- als auch im Offline-Betrieb, die Profilverwaltung sowie das Zwischenspeichern der Informationen für den Offline-Modus. Bei der Kommunikation erhält der Server die Profilinformationen, reduziert die Börseninformationen profilbezogen und sendet diese an den Client. Dem Nutzer ist somit die Möglichkeit gegeben seine Anforderungen an die Informationen innerhalb der Profilverwaltung auf dem PDA zu definieren, welche letztendlich den Reduktionsprozess beeinflussen.

Die Implementierung des Börseninformationssystems richtet sich nach dem dritten Ansatz, da diese Alternative den höchsten Bedienkomfort für den Endanwender bietet. Hierzu ist die Endanwenderapplikation den Anforderungen des Börseninformationssystems angepasst und implementiert.

Die Darstellung der Börseninformationen auf dem PDA macht eine vorherige Reduktion dieser Daten auf dem Server notwendig. Nach Erhalt des Nutzerprofils vom Client, zur Aktualisierung der Börseninformationen, werden die Daten profilorientiert, serverseitig reduziert. Um brauchbare Antwortzeiten des Servers zu erhalten, ist es erforderlich, den Reduktionsablauf zu optimieren. Folgende Reduktionsabläufe sind denkbar:

#### **Daten bei Abruf reduzieren**

Die gesammelten Daten werden erst bei Anfrage eines Clients am Server profilorientiert reduziert. Durch die teilweise großen Datenmengen und hohen Reduktionsverarbeitungszeiten ist eine akzeptable Antwortzeit des Servers nicht garantiert.

#### **Teilung der Reduktion in Vorreduktion und profilabhängige Reduktion**

Die Reduktion der Daten beruht auf eine Reduktionsberechnung. Diese kann beispielsweise eine Signifikanzberechnung sein. Mit Hilfe von ermittelten Signifikanzen für bestimmte Attribute können relevante Informationen aus der gesamten Datenmenge herausgefiltert werden.

Die Reduktionsberechnung teilt sich bei diesem Ansatz in zwei Schritte. Im ersten Schritt (profilunabhängige Reduktion beziehungsweise Vorreduktion) werden alle möglichen profilunabhängigen Zwischenergebnisse der Berechnung ermittelt, so dass der Rechenaufwand für den zweiten Schritt (profilabhängige Reduktion) minimal ist. Die Vorreduktion wird vom Server zu bestimmten Zeitpunkten automatisch durchgeführt und die dabei entstandenen Zwischenergebnisse gespeichert. Stellt ein Client eine Anfrage an den Server, um die Börseninformationen zu aktualisieren, schickt dieser zunächst das Benutzerprofil an den Server. Mit dem empfangenen Profil führt der Server die profilabhängige Reduktion durch und ermittelt so die relevanten Börseninformationen.

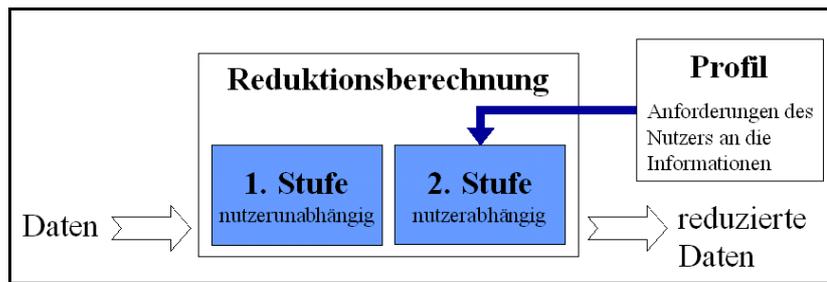


Abbildung 2-6: Teilung der Reduktionsberechnung

Durch die Teilung der Reduktion in Vorreduktion und profilabhängige Reduktion liefert der Server geeignete Antwortzeiten. Dieser optimierte Ansatz ist Bestandteil der Referenzimplementierung des Börseninformationssystems.

Nicht nur für den Ablauf der Reduktion, sondern auch für die Reduktionsmethode existieren unterschiedliche Verfahren. Die verschiedenen Methoden unterscheiden sich hinsichtlich ihrer Voraussetzungen, ihrer Arbeitsweise und der Qualität der Resultate.

Für die Referenzimplementierung wird zur Reduktion eine Klassifizierungsmethode verwendet. Mit Hilfe von errechneten Signifikanzen werden die Informationen zwei Klassen zugeordnet. Eine Klasse steht hierbei für interessante und die andere für uninteressante Informationen.

Zur Lösung des zu Grunde liegenden Klassifizierungsproblems mit zwei Klassen kann als Alternative zur Signifikanzberechnung theoretisch auch ein Lernalgorithmus verwendet werden. Der Lernalgorithmus, beispielsweise durch ein Neuronales Netz oder Entscheidungsbaum realisiert, wird vor der Nutzung des Systems vom Endanwender innerhalb einer Lernphase konfiguriert. Problematisch bei diesem Lösungsansatz ist jedoch der erforderliche Erstaufwand für den Anwender, die Genauigkeit der Klassifizierung sowie die umfangreiche Implementierung beziehungsweise die Anpassung des Lernalgorithmus.

Einen genaueren Überblick zur verwendeten und anderen Reduktionsmethoden gibt das Kapitel 3. Datenreduktion ab Seite 17.

Trotz einer an die individuellen Bedürfnisse des Endanwenders angepasste Reduktion der Börseninformationen kann nicht immer hundertprozentig gewährleistet werden, dass eine nach der Reduktion ausgewählte Information auch dem Fokus des Nutzers entspricht. Ähnlich wie es bei den meisten IMAP-E-Mail-Clients realisiert wird, ist es hilfreich, die Information in mehreren Stufen zu empfangen. Bei der Kommunikation zwischen IMAP-Server und Client werden in der Regel zunächst nur die Header der E-Mails an den Client übertragen. Dieser stellt die Informationen, wie Subject, Adresse des Verfassers und Größe der E-Mail dar. Anschließend kann der Nutzer entscheiden, ob die gesamte E-Mail gegebenenfalls empfangen werden soll.

Der geschilderte Ansatz ist ebenfalls Bestandteil der Referenzimplementierung. In der ersten Stufe werden nur die Header der Börseninformationen (Informationsheader), bestehend aus Subject, ID und Informationsbewertung, an das mobile Gerät gesendet. Ist der Anwender an einer Information näher interessiert, so können die dazugehörigen Daten vom Server nachgeladen werden. Die Integration der Methode in die Implementierung des Börseninformationssystems gewährleistet eine weitere optimalere Nutzung der nur begrenzt vorhandenen mobilen Ressourcen.

Eine eigene Vorgabe für die Referenzimplementierung der Studienarbeit ist es, XML (eXtended Markup Language) zur Kommunikation zwischen Server und Client zu verwenden. Des Weiteren soll für die Entwicklung der Serversoftware als auch der Endanwenderapplikation auf dem PDA als mobiles Endgerät Java eingesetzt werden.

Der aus allen Ansätzen und Vorgaben resultierende Systemaufbau stellt sich folgendermaßen dar:

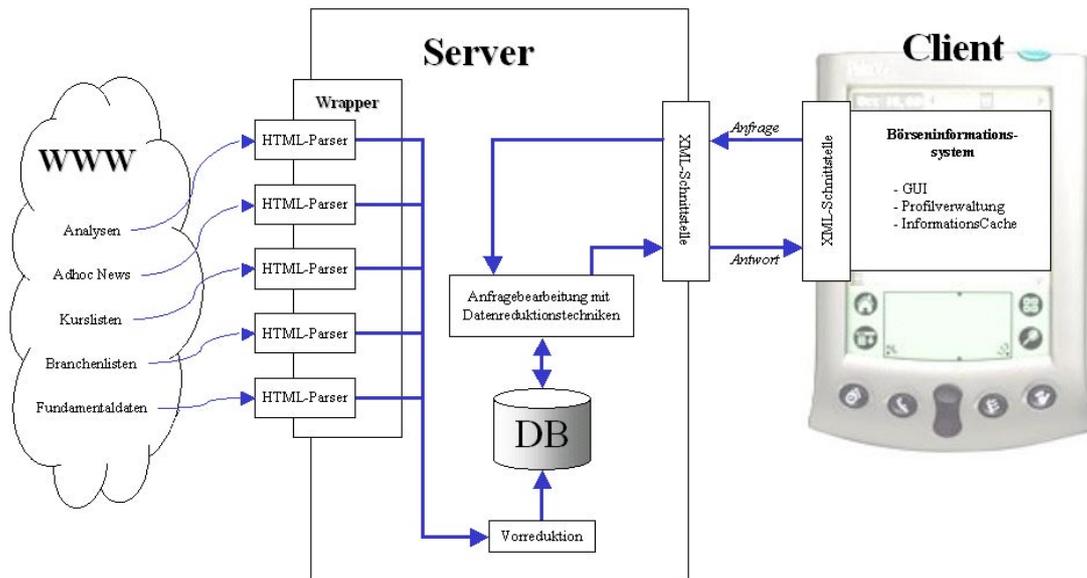


Abbildung 2-7: Aufbau des Börseninformationssystems

### 3. Datenreduktion

Das Börseninformationssystem eBIS, welches im Rahmen der Studienarbeit entstanden ist, verarbeitet eine große Menge von Börsendaten. Der Wrapper auf der Serverseite holt in zeitlichen Abständen die entsprechenden Informationen von verschiedenen Webseiten. Um diese Informationen dem Nutzer kompakt zur Verfügung stellen zu können, müssen die Daten reduziert werden. Die mobile Umgebung, wie sie im Systemaufbau in Kapitel 2.2 Technische Konzepte ab Seite 12 skizziert wurde, und die geringen Ressourcen auf dem mobilen Client machen eine Extraktion und Reduktion der Börseninformationen nach individuellen Bedürfnissen des Nutzers notwendig. Ein Aspekt, der berücksichtigt werden muss, ist die Gewährleistung einer geringen Antwortzeit des Servers.

Im Rahmen der Studienarbeit wurden drei verschiedene Datenreduktionsverfahren untersucht: Histogramme, Clustering Techniken und Verfahren der Dimensionsreduktion für Data Mining. Diese werden im einzelnen nachfolgend näher beschrieben.

#### 3.1. Histogramme

Das klassische Histogramm bezieht sich auf ein nicht-negatives, numerisches Attribut X in einer Relation R und schätzt die Daten von X derart ab, dass Gruppierungen der Attributwerte gebildet werden. Die Gruppierungen nennt man auch Blöcke. In diesem Fall spricht man von einem eindimensionalen Histogramm. Mit der Zeit hat man unterschiedliche Verfahren entwickelt, um die Blöcke zu bilden. Die Summen der Attributwertehäufigkeiten sind zum Beispiel im „Equi-depth“-Histogramm in jedem Block gleich. Bei einem anderen Histogramm, "Equi-width", bildet der Wertebereich des Attributes X die Grundlage. Bei diesem sind die Abstände zwischen dem niedrigsten und höchsten Attributwert in jedem Block gleich.

Ein Histogramm kann aus mehreren Attributen und seinen Werten gebildet werden. Solche nennt man mehrdimensionale Histogramme. Zu diesem Typus wurden in der Vergangenheit viele interessante Aspekte veröffentlicht, wie zum Beispiel in [MD88] erläutert. Häufig findet ein Histogramm seinen Einsatz im Anfrageoptimierer eines Datenbankmanagementsystems, wie DB2, Informix, Ingres, Sybase und MS SQL-Server. Er dient dort dem Abschätzen von Anfrageergebnisgrößen. Anhand dieser Approximation kann der kostengünstigste Anfrageplan ausgewählt werden.

Es stellt sich nun die Frage, wie die Daten mit einem Histogramm reduziert werden. In der Wissenschaft haben sich verschiedene Methoden durchgesetzt. Allgemein beschreibt man diesen Vorgang als Abschätzung der Attributwerte und deren Häufigkeiten in jedem Block. Alle Histogramme, die in der Vergangenheit entwickelt wurden, approximieren die Attributwertehäufigkeiten, indem das arithmetische Mittel gebildet wird. Die Abbildung 3-1 verdeutlicht diesen Vorgang.

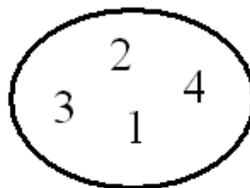


Abbildung 3-1: Häufigkeit der Attributwerte

In diesem Block befinden sich vier Attributwerte mit den jeweiligen Häufigkeiten 1,2,3 und 4. Das arithmetische Mittel ergibt 2.5.

Für die Approximation der Attributwerte innerhalb der Blöcke existieren verschiedene Verfahren. Eines beruht auf [SAC79]. Charakteristisch ist die Annahme, dass alle möglichen Attributwerte von dem Attribut X, die in dem Bereich eines Blockes liegen, in der Datenbank gespeichert vorliegen. Dem Attribut wird beispielsweise der Wertebereich der natürlichen Zahlen zugewiesen. Angenommen, es existiert ein Block, wie in der Abbildung 3-2 dargestellt. Dieser Block enthält die Attributwerte, deren Häufigkeiten in dem oberen Block dargestellt wurden.

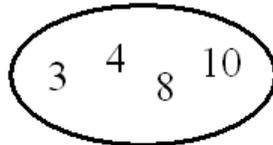


Abbildung 3-2: Attributwerte in dem Block

Der Bereich dieses Blockes besitzt die untere und obere Grenze 1 beziehungsweise 10. Bei diesem Verfahren wird nun von einer vollständigen Existenz der Attributwerte 1 – 10 in der Datenbank ausgegangen, obwohl aktuell nur die Werte 3,4,8 und 10 gespeichert sind. Diese Annahme führt zu einem Reduzieren auf die Zahlen 3 und 10. Wenn man nun die beiden oberen Blöcke in Abbildung 3-1 und Abbildung 3-2 betrachtet, wird folgende Reduktion vorgenommen:

Attributwert	Häufigkeit des Wertes
3	1
4	2
8	3
10	4

gespeicherte Werte: 2.5, 3, 10

Abbildung 3-3: Reduktionsergebnis der ersten Annahme

Diese gespeicherten Werte repräsentieren die Attributwerte 3, 4, 8 und 10 mit ihren Häufigkeiten.

Ein anderes Verfahren zum Abschätzen der Attributwerte ist ausführlich in [PC84] dargestellt. Es wird auch von einer Annahme ausgegangen. In jedem Block befindet sich nur ein Attributwert, der sich auf den niedrigsten Wert der aktuell vorliegenden bezieht. Die Attributwerte aus dem Block, welcher in der unteren Abbildung 3-4 dargestellt wird, sollen approximiert werden.

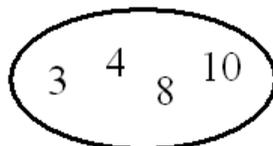


Abbildung 3-4: Attributwerte in dem Block

Die Zahl 3 ist der kleinste Wert, und es wird nur von der alleinigen Existenz in dem Block ausgegangen und demzufolge gespeichert. Betrachtet man die Abbildung 3-1 und Abbildung 3-4 erhält man folgendes Reduktionsergebnis:

Attributwert	Häufigkeit des Wertes
3	1
4	2
8	3
10	4

→ gespeicherte Werte: 2,5, 3

Abbildung 3-5: Reduktionsergebnis der zweiten Annahme

Es wurden noch mehr Verfahren entwickelt, um die Approximation der Attributwerte durchzuführen, die hier nicht vorgestellt werden. Deshalb sei auf [PIH96] verwiesen, die einen Überblick geben.

Mit Hilfe der vorgestellten Approximationsverfahren der Attributwerte und ihrer Häufigkeiten sollen im nächsten Abschnitt die Nachteile der Histogramme aufgezeigt werden, die die Integration dieses Reduktionsverfahrens in das Börseninformationssystem erschweren.

Im Kapitel 2.1 **Anwendungsspezifische Konzepte** ab Seite 9 wurden die speziellen Daten, die Analysen, Ad-Hoc-News, TopsFlops, Schwankungen und Neuemissionen, erwähnt. Auf die kann der Anwender, je nach seinen Präferenzen, zugreifen. An den Beispielen der Analysen und der TopsFlops soll gezeigt werden, weshalb sich die Histogramme als Datenreduktionsverfahren für dieses System nicht eignen.

Die Aufgabe besteht bei der Bildung der reduzierten Analyse darin, das Verhältnis zwischen Verkaufs-, Halte- und Kaufempfehlungen eines Unternehmens in einem bestimmten Zeitraum abzuschätzen und zu einer Bewertung des Unternehmens zu kommen. Überwiegen die positiven Analystenmeinungen, liegt der Schluss nahe, dass die Aktie einen Kauf darstellt. Beispielsweise kann man die Gesamtmenge der Analysen der letzten 20 Werkstage betrachten. Es ist nun möglich ein eindimensionales Histogramm aufzubauen. Die Definition bezieht sich auf ein Attribut mit numerischen Werten, auf denen das Histogramm aufgebaut wird. Dieser Umstand zwingt zu einer Konvertierung der Attributwerte. Die Empfehlungsarten halten, kaufen und verkaufen werden auf die Zahlen 0, 1 beziehungsweise -1 abgebildet. Die Unternehmen müssen ebenfalls auf numerische Zahlen abgebildet werden. Sie werden einfach fortlaufend nummeriert, und zwar auf die Zahlen 1 bis 341.

Es besteht nun die Möglichkeit, auf zwei verschiedene Arten das Histogramm aufzubauen. Das erste Histogramm besteht aus drei Blöcken, die sich auf die Analystenurteile verkaufen, halten und kaufen beziehen. Dabei lässt man die Analysen der letzten 20 Werkstage in den jeweiligen Block fallen, unabhängig vom Unternehmen. Statistische Untersuchungen im Rahmen der Studienarbeit haben ein Aufkommen von ca. 900 Analysen im betrachteten Zeitraum ergeben. Bei diesem Histogramm werden die einzelnen Unternehmen nicht mehr betrachtet, sondern nur ihre Einschätzungen. Wenn vom Unternehmen Intershop zwanzig Analysen in den letzten 20 Werktagen vorliegen, und es zu einer Verteilung von 10 Verkaufs-, 5 Halte- und 5 Kaufempfehlungen gekommen ist, dann fallen in den Verkaufsbereich 10, in den Haltebereich 5 und in den Kaufbereich 5 Werte. Attributwerte von Intershop fallen mit den Werten anderer Unternehmen in den gleichen Block. Die Unternehmen sind untereinander nicht mehr vergleichbar. Die Reduktion, die diese Art von Histogramm zulässt, ist gewaltig. Sämtliche Analysen der 341 Unternehmen, 900 Analysen, werden insgesamt auf sechs Werte reduziert. Das sind die Zahlen -1, 0 und 1 und ihre Häufigkeiten. Die Tabelle in Abbildung 3-6 veranschaulicht diese Vorgehensweise.

<i>Blöcke</i>	<i>Attribut- werte</i>	<i>Häufigkeit der Attributwerte</i>
I	-1	300
II	0	200
III	1	400



gespeicherte Werte: **300, 200, 400**  
sowie 3 Analysen, je eine Verkaufs-,  
Halte- und Kaufempfehlung

Abbildung 3-6: Reduktionsergebnis der Analysen (I)

900 Werte werden auf sechs Werte reduziert. Das entspricht einer Reduktionsquote von 99,4%. Ausgehend von der oberen Annahme von der alleinigen Existenz eines Attributwertes im Block wird in jedem der drei Blöcke ein Wert, die  $-1$ ,  $0$  beziehungsweise  $1$ , abgespeichert. Die Reduktionsquote ist zu hoch. Bei diesem Histogramm kann dem Nutzer nicht mehr sehr viel mitgeteilt werden. Dem Anwender kann nur eine statistische Auswertung über das Analyseaufkommen und die Verteilung von Verkaufs-, Halte- und Kaufempfehlungen gegeben werden. Der Anwender ist aber an anderen Tatsachen interessiert. Zum Beispiel wird er sich sehr häufig die Fragen stellen: Welches Unternehmen ist interessant, wird also sehr häufig von Analysten erwähnt? Wie sieht dahingehend die Einschätzung aus? Das Histogramm muss also modifiziert werden.

Die Analysen müssen je Unternehmen untersucht werden. Der Wertebereich für die Attributwerte wird bei der Einteilung  $-1$  bis  $1$  belassen. Die Abbildung 3-7 stellt beispielhaft 10 Aktien mit der jeweiligen Menge an Analysen, aufgeteilt in Kauf-, Verkaufs- und Halteempfehlungen, dar.

<i>Unternehmen</i>	<i>Anzahl Kauf Bewertung = 1</i>	<i>Anzahl Halt Bewertung = 0</i>	<i>Anzahl Verkauf Bewertung = -1</i>
ACG	3	0	0
Medigene	0	0	1
Aixtron	1	2	1
SZ Testsysteme	1	0	0
Gauss Interprise	2	0	0
Articon	1	0	0
D.Logistics	5	0	2
ADVA	1	1	0
Linos	2	0	0
Comdirect	0	0	1

Abbildung 3-7: Aktien mit Analysen

Die Vorgehensweise ist jetzt folgende: Zuerst müssen die Zahlen in der Tabelle auf den Wertebereich  $-1$  bis  $1$  abgebildet werden. Dann wird ein Histogramm erstellt und die zehn Werte in die Blöcke abgebildet. Für jedes Unternehmen werden die Anzahl der Kauf-, Verkaufs- und Halteempfehlungen mit der jeweiligen Bewertung der Kategorie multipliziert. Die drei Zahlen, die man dann erhält, werden anschließend addiert. Die Abbildung 3-8 enthält für jede Aktie den berechneten Wert. Außerdem enthält die Abbildung noch die normierten Werte. Das Ziel ist die Abbildung der berechneten Werte in der zweiten Spalte auf den Bereich  $-1$  bis  $1$ . Um dies zu erreichen, dividiert man die positiven Zahlen durch die größte positive Zahl, in diesem Fall die  $3$ . Analog dazu, dividiert man die negativen Zahlen durch den Betrag der kleinsten negativen Zahl, hier die  $-1$ . Die  $0$  dividiert man durch die größte positive Zahl. Zum Beispiel wurde für

SZ Testsysteme der Wert 1 in der zweiten Spalte berechnet. Die Division durch 3 führt zu dem Ergebnis 1/3.

<i>Unternehmen</i>	$(\text{Kauf} \cdot 1) +$ $(\text{Halt} \cdot 0) +$ $(\text{Verkauf} \cdot -1)$	<i>normierte Werte</i>
ACG	3	1
Medigene	-1	-1
Aixtron	0	0
SZ Testsysteme	1	1/3
Gauss Interprise	2	2/3
Articon	1	1/3
D.Logistics	3	1
ADVA	1	1/3
Linus	2	2/3
Comdirect	-1	-1

Abbildung 3-8: Aktien mit den normierten Analysewerten

Nun stellt sich die Frage nach der Anzahl der Blöcke und deren Grenzen. Es bietet sich an, für jede der drei Analystenkategorien Verkaufen, Halten und Kaufen einen Block zu bilden. Interessante Aktien sind solche, die sehr häufig analysiert werden, und bei denen gleichzeitig das Verhältnis von Kauf- und Verkaufsempfehlungen gegenüber den Halteempfehlungen überwiegt. Wenn eine Aktie zehnmal in einem bestimmten Zeitraum erwähnt wird, aber alle Einschätzungen definieren diesen Wert als Halteposition, ist dies für den Nutzer uninteressant. Deshalb wird ein Block definiert, der sämtliche Aktien enthält, deren normierter Wert um den Nullpunkt liegt. Zwei mögliche Konstellationen führen zu so einem Wert. Die Halteempfehlungen überwiegen den Kauf- und Verkaufsanalysen oder Kauf- und Verkaufsmeldungen heben sich gegenseitig auf. Im zweiten Fall gehen die Analystenmeinungen zu einer Aktie weit auseinander. Solch unausgewogenes Verhältnis ist für den Nutzer nicht wünschenswert. Die Abbildung 3-9 verdeutlicht diese Sichtweisen.

<i>Szenario</i>	<i>Unternehmen</i>	<i>Kauf</i>	<i>Verkauf</i>	<i>Halt</i>	<i>normierter Wert</i>
1	Kinowelt	0	0	5	0
2	Kinowelt	3	3	0	0

Abbildung 3-9: Szenarien für die Darstellung der uninteressanten Aktien

Außerdem sind noch zwei Blöcke notwendig, in denen normierte Werte von Aktien liegen, die nahe der -1 und 1 notieren. Diese Werte sind für den Nutzer interessant, da die Aktien, bei einem normierten Wert nahe der -1, mehrfach zum Verkaufen und bei einem Wert um die 1 mehrmals zum Kaufen eingeschätzt wurden. Die Blockgrenzen liegen bei -1, -0.5, 0.5 und 1. Aufgrund dieser Tatsachen ist es möglich, das Histogramm aufzubauen, wie in der Abbildung 3-10 dargestellt.

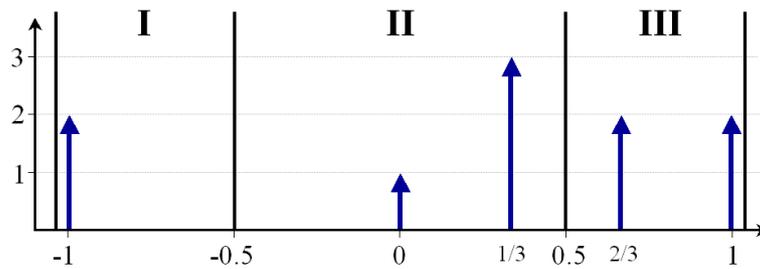


Abbildung 3-10: Histogramm für die Analysen

Im nächsten Schritt muss die Reduktion durchgeführt werden. Pro Block werden die Werte ermittelt, die ihn repräsentieren und abgespeichert werden. Die Tabelle in Abbildung 3-11 enthält die abgeschätzten Werte.

<i>Annahme</i>	<i>Block</i>	<i>approximierte Häufigkeiten</i>	<i>approximierte Attributwerte</i>
2	I	2	-1
2	II	2	0
2	III	2	2/3
<hr/>			
1	I	2	k.A.
1	II	2	0, 1/3
1	III	2	2/3, 1

Abbildung 3-11: reduzierte Analysen nach den zwei Annahmen

In jedem Block werden die Häufigkeiten per Mittelwertsberechnung ermittelt. Bei der Approximation der Attributwerte kann man, wie oben bereits vorgestellt von zwei Annahmen ausgehen, und somit die zu speichernden Werte ermitteln. Die zweite Annahme führt im ersten Block des Histogramms in Abbildung 3-10 zu einem Speichern der -1, in den Blöcken zwei und drei zum Speichern der Werte 0 und 2/3.

Charakteristisch bei der ersten Annahme, als 1 in der Abbildung in der Spalte Annahme markiert, ist, dass sich alle möglichen Attributwerte, die in den Bereich eines Blockes fallen, in der Datenbank vorliegen. Hierbei wird der niedrigste und höchste Wert in dem Block gespeichert. Nachteilig wirkt sich diese Annahme auf den ersten Block aus, der nur einen Wert enthält.

Die Werte in der Abbildung 3-11 zeigen den Nachteil der Histogramme für die Datenreduktion im Rahmen der Studienarbeit. Es gehen zu viele Informationen verloren. Die Annahmen führen nur zu einem Betrachten der niedrigsten und höchsten Werte in jedem Block und den approximierten Häufigkeiten. Unabhängig von dem Datenaufkommen in einem Block wird nur der höchste und niedrigste oder sogar nur der niedrigste Wert betrachtet. In Hinblick auf die mobile Umgebung ist die hohe Reduktionsquote sehr gut, weil dadurch eine schnelle Übertragung der reduzierten Daten möglich ist. Die reduzierten Daten sollen aber auch aussagekräftig sein. Nach der zweiten Annahme werden im dritten Block zwei Werte abgespeichert, die 2 und die 2/3. Das bedeutet, dass es zwei Unternehmen gibt, für die ein normierter Wert von 2/3 berechnet wurde. Ursprünglich waren es vier, zwei mit je 2/3 und wiederum zwei mit je 1. Die Häufigkeiten werden nach der Mittelwertsberechnung approximiert. Dadurch fehlt nach der Reduktion der Bezug der reduzierten Attributwerte zu den Unternehmen. Deshalb können nur zufällig zwei von den ursprünglich vier Unternehmen ausgewählt werden, die die reduzierten Werte 2/3 erhalten. Das ist zu ungenau.

Ein anderes Kriterium soll noch betrachtet werden, welches die Nachteile der Histogramme für dieses Projekt aufzeigt. Auf vielen Homepages im World Wide Web, die Inhalte zum Thema Börse liefern, spielen die Kurse der Aktien eine bedeutende Rolle. Interessant sind zum Beispiel die Kursentwicklungen auf den aktuellen Tag betrachtet. Dabei werden Klassifikationen, wie die Top5 oder Flop5, erstellt. Darunter fallen Aktien, die die beste beziehungsweise schlechteste prozentuale Kursentwicklung auf den aktuellen Tag gesehen, gezeigt haben. Die Abbildung 3-12 zeigt 10 Unternehmen mit den entsprechenden prozentualen Angaben.

<i>Unternehmen</i>	<i>proz. Veränderung zum Vortag in %</i>	<i>normierte Werte</i>
ACG	50	1
Medigene	40	4/5
Aixtron	30	3/5
SZ Testsysteme	20	2/5
Gauss Interprise	10	1/5
Articon	0	0
D.Logistics	-10	-1/4
ADVA	-20	-2/4
Linos	-30	-3/4
Comdirect	-40	-1

Abbildung 3-12: Aktien mit den normierten Kurswerten

Gleichzeitig sind in dieser Abbildung die normierten Werte aufgeführt. Sie berechnen sich analog zu den normierten Werten im Abschnitt zu den Analysen. Alle positiven Werte werden durch die größte positive Zahl dividiert, einschließlich der 0. Die negativen Zahlen werden durch den Betrag des kleinsten negativen Wertes dividiert. Ergebnis sind die Zahlen in der Spalte, die sich wieder im gewünschten Bereich von -1 bis 1 befinden. Nun kann das Histogramm aufgebaut werden. Da die Kategorien Tops und Flops mit Hilfe des Histogramms gebildet werden sollen, bietet es sich an, drei Blöcke zu wählen. Einen jeweils für die Tops und für die Flops und den dritten für die restlichen Werte. Die Werte -1, -0.3, 0.3 beziehungsweise 1 stellen die Blockgrenzen dar. Aus diesen Angaben ergibt sich das Histogramm in Abbildung 3-13.

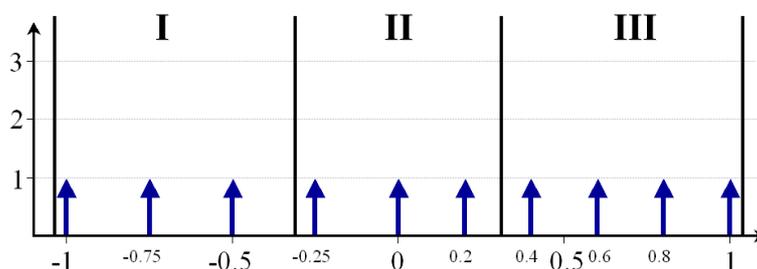


Abbildung 3-13: Histogramm für die Kurse

In dem nun folgenden Schritt werden die Approximationen durchgeführt, die die gespeicherten Daten für jeden Block darstellen und ihn somit repräsentieren. Abgeschätzt werden die Attributwerte und deren Häufigkeiten, die pro Block gemittelt werden. Das ergibt eine 1 für jeden Block. Die Abbildung 3-14 zeigt die errechneten

Werte. Es werden außerdem die Ergebnisse der beiden Verfahren für die Abschätzung der Attributwerte dargestellt. Bei der zweiten Methode, in der Abbildung 3-14 als 2 in der Spalte „Annahme“ markiert, wird von einem einzigen Attributwert in einem Block ausgegangen. Die Zahlen -1, -0.25 und 0.4 werden für die Blöcke I,II beziehungsweise III gespeichert. Die erste Methode von [SAC79] speichert den niedrigsten und höchsten Wert pro Block.

<i>Annahme</i>	<i>Block</i>	<i>approximierte Häufigkeiten</i>	<i>approximierte Attributwerte</i>
2	I	1	-1
2	II	1	-0.25
2	III	1	0.4
<hr/>			
1	I	1	-1, -0.5
1	II	1	-0.25, 0.2
1	III	1	0.4, 1

Abbildung 3-14: reduzierte Kurse nach den zwei Annahmen

Mit der zweiten Methode wird eine Reduzierungsquote von 40% und mit der ersten eine Reduzierung von 10% erreicht.

Diese Art der Datenreduktion mittels der Histogramme ist für die Kategorien der Tops und Flops nicht anwendbar. Normalerweise werden bei dieser Klassifikation mindestens jeweils 5 Werte angegeben. Die Abbildung 3-14 zeigt jedoch maximal 2 Attributwerte, die als repräsentative Objekte für die Blöcke abgespeichert werden. Es geht die Flexibilität verloren. Wenn man dem Nutzer die Möglichkeit gibt, und das ist bei diesem Projekt sehr sinnvoll, per Profil die Datenreduktion zu beeinflussen, sind die Histogramme schwer zu integrieren. Unabhängig von der Anzahl der Attributwerte in den Blöcken vorliegen, maximal können nur zwei pro Block abgespeichert werden. Das ist unbefriedigend.

### 3.2. Clustering-Techniken

In den letzten 30 Jahren wurde diese Technik sehr intensiv im Bereich der Statistik untersucht. Grundsätzlich wird dabei nach Clustern in den zugrundeliegenden Daten gesucht. Unter Clustern versteht man dabei ein Sammeln von Daten, die einander gleich sind. Diese Gleichheit wird durch eine Distanzfunktion ausgedrückt. Für jedes Paar von Objekten  $p_1, p_2$  ist die Distanz  $D(p_1, p_2)$  bekannt. Zusätzlich gibt es eine Funktion, die die Qualität des Clusters bestimmt. Weiterführende Erläuterungen zu den Distanz- und Qualitätsfunktionen kann [BD97] entnommen werden.

Die fundamentale Frage lautet nun: „Wie viele Cluster existieren in der gegebenen Datenmenge?“. Um diese Frage zu beantworten, wurden verschiedene Algorithmen entwickelt, die sich in zwei Gruppen unterteilen lassen. Zum Einen die statistischen Methoden, die hier eine untergeordnete Rolle spielen. Dem interessierten Leser sei [BD97] zu empfehlen. Zum anderen die Datenbankmethoden. Zu dieser Technik gehören zum Beispiel CLARANS, beschrieben in [NH94] und der Algorithmus DBSCAN, welcher in [EKXS96] erklärt wurde.

Das Verfahren PAM, welches 1990 von L. Kaufmann und P.J. Rousseeuw entwickelt wurde, soll jetzt näher vorgestellt werden. Um  $k$  Cluster zu finden, bestimmt der

Algorithmus repräsentative Objekte für jedes Cluster. Diese Elemente werden Medoide genannt. Jedes nicht-repräsentative Objekte, welches dem Medoiden ähnlich ist, wird mit ihm gruppiert. Die Eingabeobjekte werden zufällig aus der Gesamtmenge ausgewählt. In jedem Algorithmusschritt werden die repräsentativen Objekte neu bestimmt, in dem die Medoide mit den nicht-repräsentativen Objekten ausgetauscht werden, wenn sich dadurch die Qualität des Clusters verbessert. Es kommt nur zu einem Austauschen zwischen Medoid und nicht-repräsentativen Objekt, wenn die Kosten angemessen sind. Bei der Berechnung der Kosten werden vier Fälle unterschieden, die im einzelnen in [NH94] nachgelesen werden können. Wenn kein Tauschen von Objekten mehr möglich ist, wird der Algorithmus beendet. Das vollständige Verfahren kann ebenfalls [NH94] entnommen werden.

Der nächste Abschnitt soll die Gründe erklären, weshalb die Clustering-Techniken für dieses Projekt ungeeignet sind. Wie auch schon bei den Histogrammen im vorhergehenden Kapitel werden die Analysen und TopsFlops betrachtet. Wenn man die Analysen in den letzten 20 Werktagen betrachtet, werden diese pro Unternehmen zusammengefasst. Anschließend kann eine Clustermethode repräsentative Objekte, in diesem Fall Aktien, bestimmen. Bei allen Algorithmen werden Distanzfunktionen verwendet, um die repräsentativen Objekte zu bestimmen. Es ist daher notwendig, die Analysen der Unternehmen in numerische Werte umzuwandeln. Eine Konvertierung könnte zum Beispiel die Zahlen 0, 1 und -1 ergeben. Diese Zahlen entsprechen den Analystenurteilen Halten, Kaufen und Verkaufen. Die Abbildung 3-15 zeigt Aktien mit Empfehlungen und den entsprechenden Bewertungen.

<i>Unternehmen</i>	<i>Art der Analyse</i>	<i>Bewertung der Analyse</i>	<i>Summe</i>
A	Kaufen	1	2
A	Kaufen	1	
B	Kaufen	1	1
C	Halten	0	0
C	Halten	0	
D	Verkaufen	-1	-2
D	Verkaufen	-1	
E	Verkaufen	-1	-1

Abbildung 3-15: Übersicht über Analysen und deren Bewertung je Unternehmen

Die Aufgabe besteht nun darin, Cluster aufzustellen. Verwendet wird das Verfahren des statischen Clusters. Es sollen 3 Cluster gebildet werden. Das wird festgelegt. Im Unterschied dazu existiert das dynamische Clustern, wobei die zu reduzierende Datenmenge gescannt wird, und anschließend die Anzahl der Cluster vom Algorithmus bestimmt wird. Um nun das statische Clustern effizient durchzuführen, wird vorab eine Normalisierung der Zahlen in der Spalte Summe der Abbildung 3-15 durchgeführt. Dabei werden die Werte auf Zahlen zwischen -1 und 1 abgebildet. Die Normalisierung wird folgendermaßen durchgeführt: Die positiven Zahlen werden durch die größte positive Zahl dividiert. Im Gegensatz dazu werden die negativen Zahlen durch den Betrag der kleinsten negativen Zahl dividiert. Das Resultat wird in der Abbildung 3-16

dargestellt. Die Zahl 0 wird übernommen. Die größte positive Zahl stellt das Objekt A mit 2 dar. Die kleinste negative Zahl beträgt -2 des Objektes D.

<i>Unternehmen</i>	<i>Summe</i>	<i>Normalisierung</i>
A	2	1.0
B	2	0.5
C	0	0.0
D	-2	-1.0
E	-1	-0.5

Abbildung 3-16: Normierung der Werte

Cluster werden nun um die Zahlen -1, 1 und 0 gebildet, wobei Zahlen größer gleich 0.5 zum Cluster mit dem repräsentativen Objekt 1 und Zahlen kleiner gleich -0.5 zum Cluster mit dem repräsentativen Objekt -1 gehören. Wendet man diese Regeln auf die normierten Werte an, gehören zu dem Cluster 1 die Objekte A und B, zu dem Cluster 2 das Objekt C und zu dem dritten Cluster die Objekte D und E. Die Abbildung 3-17 stellt dies graphisch dar. Die Objekte A, C und D wurden jeweils als repräsentative Objekte in ihrem Cluster festgelegt.

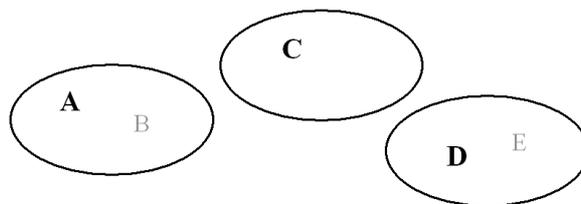


Abbildung 3-17: Darstellung der Cluster mit den jeweiligen repräsentativen Objekten

Wie bei allen Clustering-Techniken üblich, wird mit den repräsentativen Objekten weitergearbeitet, um die Datenreduktion zu realisieren. Während der Reduktion von Werten gehen Informationen verloren. Im ersten Cluster wird für die weitergehende Betrachtung das Objekt B nicht mehr berücksichtigt. Das Objekt A repräsentiert das gesamte Cluster. Dieser Verlust an Informationen der restlichen Objekte in dem Cluster ist für das Projekt zu hoch. Auf eine Anfrage erhält der Anwender nur die Informationen des repräsentativen Objektes. Wünschenswert wäre dagegen die Rückgabe von mehreren Objekten. Es muss ein guter Kompromiss zwischen schneller Übertragung und ausreichender Größe der Reduktionsdaten gefunden werden. Die schnelle Übertragung ist wegen des Einsatzes eines mobilen Clients erforderlich.

Benutzer, die interessante Aktien, die normierte Werte nahe der 1 oder -1 besitzen, angezeigt bekommen möchten, wären nicht zufrieden, wenn sie nur das repräsentative Objekt übermittelt bekämen. Der Endanwender ist darüber hinaus an Aktien interessiert, für die eben diese normierten Werte nahe der 1 und -1 berechnet wurden.

Die Vorgehensweise bei der Erstellung der Cluster lässt sich ohne Probleme auf die TopsFlops anwenden. Die Kategorie TopsFlops fasst mehrere Aktien zusammen, die die beste oder schlechteste Kursperformance, bezogen auf einen bestimmten Zeitraum, gezeigt haben. Streng nach den Algorithmen der Clustering Techniken wäre es aber nur möglich, das repräsentative Objekt als Ergebnis einer Anfrage zurückzugeben.

Die Abbildung 3-18 verdeutlicht diesen Sachverhalt. Eine Menge mit fünfzehn Objekten wird untersucht. Es werden die gleichen Normierungsregeln analog zu den Analysen verwendet. Die zehn Objekte werden in drei Cluster eingeteilt, wobei die

fettgedruckten Buchstaben in der Spalte „Unternehmen“ die repräsentativen Objekte darstellen. Nun werden die Kriterien Tops und Flops auf die Tabelle angewendet. Der Zeitraum spielt bei diesem Beispiel keine Rolle. Prozentual am höchsten stiegen die Aktien des ersten Clusters, wobei aber nur das repräsentative Objekt A als Ergebnis zurückgegeben wird. Dies ist aber in den meisten Fällen nicht wünschenswert, da meistens mehr Aktien in der Kategorie Tops aufgeführt werden. Deshalb eignen sich unter diesem Gesichtspunkt nicht die Clusteringverfahren. Besser wäre die Einführung eines Schwellenwertes in Bezug auf die normalisierten Werte innerhalb des Clusters. Ein Überschreiten dieses Grenzwertes führt zu einer Aufnahme in die Antwortmenge.

<i>Unternehmen</i>	<i>Kurs- Performance in %</i>	<i>Normalisierung</i>	<i>Clusternummer</i>
<b>A</b>	50	1	1
B	40	0.8	1
C	30	0.6	1
<b>D</b>	20	0.4	2
E	10	0.3	2
F	0	0	2
G	-10	-0.2	2
H	-20	-0.4	2
<b>I</b>	-30	-0.6	3
J	-40	-0.8	3

Abbildung 3-18: Clusterbildung der Unternehmen, bezogen auf die jeweilige Kursperformance

Denkbar wäre zum Beispiel im ersten Cluster ein Wert von 0.7, wobei dann die Objekte A und B zurückgegeben würden. Die Mächtigkeit der Ergebnismenge beträgt hierbei zwei. In der Praxis wählt man üblicherweise zwischen fünf und fünfzehn Objekte.

Das Kriterium Flops bezieht sich auf das dritte Cluster. Bei der Betrachtung wird analog vorgegangen, allerdings werden hierbei die Aktien mit der schlechtesten Kursperformance herausgefiltert.

In diesem Abschnitt wurden die Clustering Techniken in ihren Grundzügen vorgestellt. Gezeigt wurden anhand von zwei wichtigen Kriterien die Nachteile des Clusters, so dass ein Einsatz dieser Algorithmen innerhalb des Projektes ausgeschlossen wird.

### 3.3. Dimensionsreduktion für Data Mining

Die Aufgaben des Data Mining bestehen darin, Daten zu analysieren und in ihnen nach Mustern zu suchen, um somit neue und interessante Informationen zu gewinnen. Hierzu dienen Vorhersagemethoden. Bei der für die Mustererkennung zu Grunde liegenden Datensätzen handelt es sich meist um sehr große Mengen, die so nicht verarbeitet werden können. Um Muster in den Daten zu finden, bietet sich eine vorherige Reduktion an.. Die Aufgabe der Datenreduktion liegt in der Abschätzung welche der Daten für das Data Mining benötigt werden und welche nicht. Ziel ist es also, die Daten auf das Notwendigste zu reduzieren, wobei die reduzierte Datenmenge die für die Mustererkennung notwendigen Informationen noch beinhaltet.

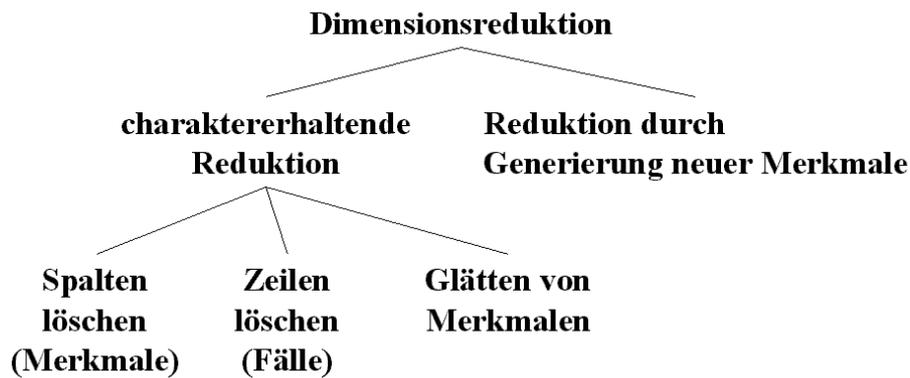


Abbildung 3-19: Formen der Dimensionsreduktion

Die allgemeine Herangehensweise, Daten für Vorhersagemethoden zu vereinfachen, ist die Dimensionsreduktion. Dabei unterscheidet man grundsätzlich zwischen zwei Arten: die charaktererhaltende Reduktion und die Reduktion durch Generierung neuer Merkmale (Spalten in der Relation). Bei der Reduktion durch Generierung neuer Merkmale werden aus der Grunddatenmenge neue Merkmale generiert, mit denen im weiteren Arbeitsprozess gearbeitet wird. Auf dieses Verfahren soll hier nicht näher eingegangen werden.

Ein anderer Ansatz ist die charaktererhaltende Reduktion, wobei Attributwerte aus der Grunddatenmenge ausgeblendet werden. Auf diese Weise wird eine Reduktion durchgeführt. Hierbei unterscheidet man zwischen drei Möglichkeiten: Spalten (Merkmale) löschen, Zeilen löschen oder das Glätten von Merkmalen. Im Gegensatz zur Reduktion durch Generierung neuer Merkmale bleiben hier die Attributeigenschaften erhalten, da hier Attribute bzw. Datensätze vollständig gelöscht oder Werte geglättet werden. Es bleibt somit der Charakter der originalen Tabelle bewahrt.

Bei der Merkmalselektion werden die Spalten überprüft und ihr Vorhersagepotential für den Data-Mining-Prozess bestimmt. Spalten können gelöscht werden, wenn sie nur über ein geringes Vorhersagepotential verfügen oder sie bezüglich anderer Spalten redundant sind. Ziel ist es, die Spalten zu selektieren, mit denen die gleichen Muster zu finden sind, wie mit der gesamten Spaltenmenge. Für die Suche nach vielversprechenden Teilmengen existieren verschiedene Herangehensweisen:

### **Merkmalauswahl mit Durchschnitt und Varianz**

Die Merkmalauswahl mit Durchschnitt und Varianz lässt sich im allgemeinen als Klassifizierungsproblem mit zwei Klassen betrachten. Hierbei werden jeweils zwei Spalten miteinander verglichen. Für den Vergleich wird eine einfach zu berechnende Abstandsmessung verwendet. Der Abstand wird mittels Durchschnitt, normalisiert mit der Varianz der Spalten, berechnet. Ähneln sich die Spalten, ist der Abstand gering und das Interesse an einem Merkmal nimmt ab. Hingegen nimmt das Interesse zu, wenn der Abstand der Durchschnitte groß beziehungsweise gar nicht vergleichbar ist. Es gibt zwei verschiedene Verfahren, wie die Merkmale miteinander verglichen werden. Bei der unabhängigen Merkmalselektion werden zwei Spalten miteinander verglichen ohne Beachtung des Verhältnisses zu anderen Merkmalen. Hingegen werden bei der optimalen abstands-basierten Merkmalselektion die Merkmale gemeinschaftlich untersucht. Eine vorgegebene Signifikanz wird genutzt, um zu entscheiden welche Merkmale selektiert werden.

## Heuristische Merkmalselektion

Bei der heuristischen Merkmalselektion unterteilt man die Spalten in zwei Mengen: ausgewählte und verbleibende. In einer interaktiven Prozedur wird im ersten Schritt für jedes Merkmal aus der *verbleibenden Merkmalmenge* der zusätzliche Gewinn berechnet. Das Merkmal mit dem maximalen Gewinn wird zur *ausgewählten Merkmalmenge* hinzugefügt. Im zweiten Schritt werden alle Merkmale in der *ausgewählten Merkmalmenge* individuell untersucht und gegebenenfalls entfernt. Es werden keine Merkmale mehr hinzugefügt oder entfernt, wenn dadurch der Gewinn nicht steigt. Zur Berechnung des Gewinns werden Abstandsmessungen beziehungsweise Signifikanzberechnungen verwendet.

## Merkmalauswahl mit Entscheidungsbäumen

Die Merkmalauswahl mit Entscheidungsbäumen ist in die Vorhersagemethode des Data-Mining-Prozesses integriert und mit der Suche nach Mustern koordiniert. Der Entscheidungsbaum kann genutzt werden, um Merkmale für andere Methoden, wie z.B. Neuronale Netze, zu selektieren. Der Entscheidungsbaum enthält die Merkmale, die nicht gelöscht werden. Die Bewertung der Merkmale geschieht im Kontext zu den bereits ausgewählten.

Die Selektion von Spalten reicht in der Regel für das Data Mining nicht aus. Zusätzlich wird eine Wertereduktion durch das Löschen von Zeilen vorgenommen. Die Idee hierbei ist, die Distinct-Werte der Tabelle zu erfassen und die redundanten Werte zu eliminieren. Das Problem daran ist, dass meist nicht genügend Tupel entfernt werden, da die Anzahl der gleichen Werte gewöhnlich sehr gering ist. Mit verschiedenen Ansätzen wird versucht die Wertereduktion effektiver zu gestalten:

### Wertereduktion durch Glätten

Der Ansatz ähnelt dem der Histogramme. Hierbei wird ebenfalls der Wertebereich eines Attributes in Blöcke eingeteilt und die darin enthaltenen Werte approximiert. Dieses kann durch die Bildung eines Mittelwertes oder durch die Blockgrenzen durchgeführt werden. Jeder Block wird dann nur noch durch repräsentative Attributwerte beschrieben.

### Wertereduktion durch Sampling

Die Wertemenge wird durch eine zufällige Auswahl von Werten reduziert. Es muss gewährleistet sein, dass die Inputmenge für Vorhersagemethoden verlässliche Tupel enthält. So ist für eine Klassifizierung mindestens ein Tupel für jede Klasse notwendig. Problem bei der Wertereduktion durch Sampling ist also die Auswahl der Beispiele und die Bestimmung der Anzahl.

Die unterschiedlichen Reduktionsmethoden für Data Mining können natürlich nicht in dieser Form für das Reduktionsproblem des Börseninformationssystem übernommen werden. Ansätze wie die Klassifikation mit zwei Klassen, die Verwendung einer Abstandsfunktion, Normierung von Abständen und vorgegebene Signifikanzen können jedoch genutzt werden.

Die Auswahl von interessanten Börseninformationen kann als Klassifizierung mit zwei Klassen aufgefasst werden. Die Klassen stehen hierbei für interessante beziehungsweise uninteressante Informationen. Durch eine geeignete Abstandsfunktion können die Börseninformationen den Klassen zugeordnet werden. Für die Abstandsfunktion wird

eine Berechnung gewählt deren Ergebnis ein numerischer Wert ist (Signifikanz). Dieser kann mit einer vorgegebenen Signifikanz verglichen werden. Ist die berechnete Signifikanz der Information größer als die vorgegebene, so ist die Information interessant und wird der entsprechenden Klasse zugeordnet. Analog werden alle Informationen mit kleinerer Signifikanz der anderen Klasse zugewiesen. Da sich die Börseninformationen in unterschiedliche Kategorien gliedern, sind sie nicht direkt miteinander vergleichbar. Zum Vergleich müssen die Signifikanzen der verschiedenen Kategorien normiert werden.

Im Börseninformationssystem ist die Datenreduktion in zwei Stufen geteilt: Vorreduktion und profilabhängige Reduktion. Die Berechnung und Normierung der Signifikanzen erfolgt innerhalb der Vorreduktion. Bei der profilabhängigen Reduktion werden eine benutzerspezifische Gewichtung der berechneten Signifikanzen und die Klassifizierung der Börseninformationen durchgeführt.

## 4. Evaluation von Hard- und Software

Bevor in Kapitel 5 Die Teilkomponenten des Systems ab Seite 44 auf den konkreten Aufbau und die Funktionalität der im System vorhandenen Komponenten eingegangen werden soll, wird in diesem Kapitel die für deren Implementierung genutzte Hard- und Software genauer untersucht. Der Hauptanteil der Evaluation lief bereits im Vorfeld der Implementierung ab und unterstützte unter anderem die Entscheidung, welche Hard- und Software im Börseninformationssystem Verwendung findet.

### 4.1. Das mobile Endgerät

Längst helfen Kleinstcomputer, wie PDA's, nicht nur bei der Prozentberechnung oder Verwaltung von Adressen und Terminen, sondern sie ermöglichen dem Anwender auch die Nutzung komplexerer Applikationen sowie die Kommunikation mit der Außenwelt. Im Gegensatz zum Notebook oder Subnotebook zeichnen sie sich durch ihre kleine und leichte Bauform aus. Gerade die Tragbarkeit (Portabilität) dieser Geräte ermöglichen die mobile Nutzung, entsprechend des gewählten Anwendungsszenarios. Hinsichtlich Ergonomie, Kompatibilität und Flexibilität müssen bei den PDA's jedoch Einschränkungen in Kauf genommen werden. Je kleiner die Geräte um so schwieriger gestaltet sich die Ein- und Ausgabe. Mit Hilfe eines Stiftes kann auf dem Touchscreen agiert werden und nur selten steht eine Tastatur mit zumeist sehr kleinen Tasten zur Verfügung. Das Display zeigt erfahrungsgemäß nur einen Ausschnitt dessen, was zum Anzeigen einer Information notwendig ist, so dass eine gute Übersicht fehlt. Auch die Kompatibilität stellt ein Problem dar. PDA's verfügen über ein anderes Betriebssystem als der Desktop-PC, so dass einem Datenaustausch immer eine Konvertierung vorausgeht. Bezüglich Flexibilität bieten die kleinen Geräte nicht die Softwarevielfalt, wie man es von herkömmlichen PCs gewohnt ist. Zudem ist auch nicht jede Software dazu geeignet, auf einem PDA zu laufen. Die genannten Einschränkungen sind bei der Auswahl des Endgerätes und der Implementierung der Anwendungssoftware berücksichtigt.

Für die Implementierung standen der Psion Serie 5 MX von Psion PLC, der Compaq 810c von Compaq Inc. und der Palm Vx von Palm Inc. zur Auswahl.



Abbildung 4-1: Psion Serie 5 MX

Der Psion Serie 5 MX verfügt über ein Bildschirm in halber VGA-Auflösung (640 x 240 Bildpunkten) mit 16 darstellbaren Graustufen. Dieser dient nicht nur zur Darstellung von Informationen sondern fungiert auch als Touch-Pad. Zur weiteren Eingabe besitzt der PDA eine Tastatur mit großen Tasten, das vom Schreibgefühl schon fast an ein Subnotebook erinnert. Das Gerät arbeitet mit dem eigens entwickelten

EPOC-Betriebssystem in einer 32-Bit-Version und einem RISC-Prozessor vom Typ ARM 7100 bei 18,4 MHz. Zur Speicherung von Applikationen und Daten sowie zum Arbeiten ist der Psion mit 8 MByte RAM ausgestattet. Zur Stromversorgung dienen zwei Mignonzellen, die laut Herstellerangaben einen Betrieb bis zu 30 Stunden gewährleisten. Inklusive der Batterien wiegt der PDA ca. 350 Gramm.



Abbildung 4-2: Compaq 810c

Die Bauform des Compaq 810c ist vergleichbar mit der des Psion. Der Compaq ist ebenfalls mit einem 640 x 240 großen Touchscreen, der jedoch 256 Farben darzustellen vermag, und einer Tastatur ausgestattet. Der Compaq arbeitet mit einem 75 MHz RISC-Prozessor (Philips PR31700 MIPS), 16 MByte RAM und benutzt das Microsoft Windows CE 2.0 Betriebssystem. Als Energiezufuhr dient ein wiederaufladbares NiMH-Akku oder ein externer Netzadapter. Weitere Ausstattungsmerkmale des Compaq sind ein integriertes 33,6 Kbps Modem, eine serielle Schnittstelle, ein PC-Card-Einschub, ein Infrarot-Port (IrDA konform), ein Lautsprecher und ein Mikrofon. Mit einem Gewicht von ungefähr 470 Gramm inklusive der Batterien ist der Compaq das schwerste der drei mobilen Endgeräte.



Abbildung 4-3: Palm Vx

Das Design des Palm Vx hebt sich deutlich von dem des Psion und Compaq ab. Der PalmPilot verfügt als einziger nicht über eine Tastatur. Alle Eingaben werden mit Hilfe eines Stiftes auf dem Touchscreen vorgenommen. Das aktive FSTN-Display hat eine Auflösung von 160 x 160 Pixel und kann 16 Graustufen darstellen. Intern arbeitet der Palm mit einem 20 MHz getakteten Dragonball EZ Prozessor (Motorola MC68EZ328) von Motorola und 8 MByte RAM. Der PDA ist mit dem Palm OS Version 3.5 Betriebssystem ausgestattet. Zur Stromversorgung nutzt der Palm ein integriertes Lithium-Ionen Akku, welches mit der mitgelieferten Dockingstation aufgeladen werden kann. Zur Kommunikation mit anderen Geräten (z.B. Synchronisation mit einem Desktop-PC) ist eine serielle Schnittstelle (spezielles Kabel oder Dockingstation

notwendig) und eine Infrarotschnittstelle (IrDA konform) vorhanden. Der Palm ist das leichteste der drei PDA's mit einem Gewicht von 113 Gramm.

Alle drei Geräte laufen mit einem anderen Betriebssystem. So galt es nicht nur eine Auswahl des mobilen Endgerätes hinsichtlich der in den Geräten integrierten Hardware, sondern auch in Hinblick auf Flexibilität und Kompatibilität des Betriebssystems vorzunehmen. Eine kleine und leichte Bauform der Geräte sowie die Eignung zur Ausführung von Applikationen, die in Java implementiert sind, waren weitere Auswahlkriterien.

Für das im Rahmen der Studienarbeit entwickelte Börseninformationssystem fiel die Wahl auf den Palm Vx als mobiles Endgerät.

#### 4.1.1 Der Palm

Lange Zeit sah es so aus, als ob mit den Geräten unterhalb des Notebooks kein Geld zu verdienen sei. So verkaufte 1994 der Pilot-Entwickler Jeff Hawkins nach einem Jahr erfolgloser Suche nach Geldgebern für seine portable Entwicklung sein Unternehmen Palm Computing an den Modemhersteller U.S. Robotics. Dieses ging wiederum 1997 an 3Com. Im März 2000 wurde Palm vom Unternehmen 3Com ausgegliedert. Das neu entstandene Unternehmen Palm Incorporation (Palm Inc.) wurde an die amerikanische Wachstumsbörse NASDAQ platziert. Derzeit ist Palm Inc. mit insgesamt ca. 10,9 Millionen verkauften Geräten Marktführer im Kleincomputersegment. Alleine im 2. Quartal 2000 verkaufte die Firma aus Santa Carla 2,1 Millionen PalmPilots, 45% mehr als im Quartal zuvor (1,5 Millionen) [InVi00].

Inzwischen wird der PalmPilot auch von anderen Firmen vertrieben, wie z.B. der WorkPad von IBM oder der Visor von Handspring.



Abbildung 4-4: PalmPilots

Die Produktpalette von Palm Inc. umfasst verschiedene Versionen des PalmPilots. Hierzu gehören derzeit die Palm-III-Serie (IIIc, IIIx, IIIe, IIIx), die Ver-Serie (V, Vx), der Palm m100 für das Niedrigpreissegment sowie die in Deutschland noch nicht erhältlichen Palm VII und VIIx. Ausgestattet sind die Palm-Geräte mit 2 bis 8 MByte, einem Motorola Dragonball Prozessor mit einer Taktrate von 16 bis 20 MHz und eignen sich somit mehr oder weniger für Verwendung von Java-Applikationen. Die Geräte arbeiten mit dem Betriebssystem Palm OS in der Version 3.5. Dieses kennt lediglich zwei unterschiedliche Dateitypen. Bei Dateien mit der Endung *.prc* handelt es sich um Programmdateien, die ausführbaren Programmcode enthalten. Dateien, die der Speicherung von Daten dienen, haben die Dateiendung *.pdb* und werden Kataloge genannt.

Die Displays der Geräte haben eine Auflösung von 160 x 160 Pixel und können 16 Graustufen oder 256 Farben beim IIIc darstellen. Die Zeicheneingabe erfolgt beim PalmPilot mit einem Stift und dem Programm Graffiti, welches fest im Betriebssystem

integriert ist und Buchstaben in einer vorgegebenen Art und Weise auf dem Eingabefeld erwartet. Zur Synchronisation mit einem Desktop-PC verfügen die Geräte über eine Infrarot- sowie serielle Schnittstelle. Der Infrarotport des Palm ermöglicht dem PDA die direkte Kommunikation untereinander und somit auch den Austausch von Daten zwischen diesen. Außer dem Palm m100 vermögen alle Geräte eine TCP/IP-Verbindung zur Kommunikation mit der Außenwelt aufzubauen. Die Verbindung erfolgt mit Hilfe eines geeigneten Mobiltelefons oder einem im Netz angemeldeten PC. Diese kommunizieren über ein Kabel oder Infrarot mit dem PalmPilot.

Zum Testen entwickelter Palm-Programme steht dem Programmierer der Palm-OS-Emulator zur Verfügung. Dieser auch POSE genannte Emulator ermöglicht die Ausführung von ROM's der verschiedenen Palm-Geräte. Ist im Emulator erst einmal ein ROM eines Palm-Gerätes geladen und registriert kann mit dessen Hilfe ohne Einschränkung die entwickelte Software getestet werden, ohne die privaten Adressen und Termine auf dem Palm in Gefahr zu bringen. Ebenfalls wie der originale Palm erlaubt der Emulator den Aufbau einer TCP/IP-Verbindung, indem die Verbindung auf die des PC umgeleitet wird.

## **4.2. Kommunikation zwischen Server und Client**

Zur Aktualisierung der Börseninformationen auf dem mobilen Endgerät muss sich der PDA mit dem Server in Verbindung setzen. Hierzu wird eine TCP/IP-Verbindung zwischen den beiden aufgebaut. Der Palm bietet verschiedene Möglichkeiten zur Herstellung einer solchen Verbindung an. Er verfügt über eine serielle und eine Infrarotschnittstelle, über die mit einem Mobilfunkgerät oder einem anderen Rechner kommuniziert werden kann. Mit deren Hilfe ist ein Kommunikationsaufbau zu dem bereitgestellten Server möglich.

Folgende Verbindungskonfigurationen wurden während der Studienarbeit genutzt:

### **Palm – Infrarot – separater Rechner – Server**

Der Palm verbindet sich über die vorhandenen Infrarotschnittstelle mit einem separaten Rechner, der ebenfalls über eine Infrarotschnittstelle verfügt. Hierzu werden am Palm und Rechner lediglich die Kommunikationsgeschwindigkeit abgeglichen. Der separate Rechner leitet dann die Verbindungsanforderungen an seine eigene Netzwerkverbindung weiter. Bei Microsoft Windows Betriebssystemen wird dieses mit Hilfe der RAS-Verbindungsverwaltung, die hierfür konfiguriert werden muss, bewerkstelligt. Die Nutzung des RAS-Servers stellt unter Microsoft Windows eine enorme Herausforderung dar und gewährleistet keinen stabilen Ablauf. Mit Hilfe der Software Mocha W32 PPP von MochaSoft [Mocha] kann dieses Problem jedoch umgangen werden. Mocha W32 PPP ist speziell eine Software, die für die Kommunikation von PalmOS und Windows CE Geräten mit dem World Wide Web entwickelt wurde. Hierzu wird lediglich die Übertragungsgeschwindigkeit und der Kommunikationsport bestimmt. Die Software lauscht an einer seriellen Schnittstelle, auf die auch ein Infrarotport eines Rechners gemappt werden kann, und leitet alle Anforderungen an die Netzwerkumgebung des Rechners weiter. Über diesen erfolgt die Verbindung zum Applikationsserver, die dann auch dem mobilen Endgerät zur Verfügung steht.<sup>1</sup>

---

<sup>1</sup> Hinweis: Verwendet der HotSync Manager der Palm-Synchronisationssoftware auf dem PC dieselbe Infrarotschnittstelle wie Mocha W32 PPP, so muss dieser während der Kommunikation deaktiviert werden.

### **Palm – serielle Schnittstelle (Dockingstation) – separater Rechner – Server**

Diese Verbindungskonfiguration ähnelt der vorab dargestellten. Jedoch wird hierbei anstelle der Nutzung der Infrarotschnittstellen eine Verbindung über die serielle Schnittstelle des Palms und des separaten Rechners aufgebaut. Zur Umleitung der Netzwerkanforderungen des Endgerätes von der seriellen Schnittstelle des Desktop-PC an deren Netzwerkumgebung kommt ebenso Mocha W32 PPP zum Einsatz. Auch bei dieser Kommunikationsmethode ist die Abstimmung der Kommunikationsgeschwindigkeiten der Palm-Netzwerkeinstellung, der seriellen Schnittstelle und der Mocha-Software aufeinander notwendig.<sup>2</sup>

### **Palm – Infrarot – Mobilfunkgerät mit Softmodem – Einwahlknoten – Server**

Diese Konfiguration entspricht durchweg den Vorstellungen der Anwendungsnutzung in einer mobilen Umgebung und ist somit besonders für diese geeignet. Der Nutzer ist lediglich von der Netzabdeckung seines Mobilfunkanbieters abhängig.

Für die Evaluation der Konfiguration wurde ein S25 der Firma Siemens mit integriertem Softmodem und Infrarotschnittstelle verwendet.

Für den Verbindungsaufbau werden am Palm die Parameter des Providers (Rufnummer, Benutzerkennung, Kennwort etc.), die Verbindungsmethode *IrCOMM an Modem* und eine Übertragungsgeschwindigkeit von 9600 Bit/s (entsprechend dem derzeitigen GSM-Standard) eingestellt. Die Bereitstellung der Verbindung erfolgt dann durch die Aktivierung am PDA. Hierzu muss das Mobilfunkgerät mit der IR-Schnittstelle erreichbar für den Palm platziert werden.

### **PalmOS-Emulator – Server**

Zum Testen der entwickelten Applikation kann ebenso der PalmOS Emulator verwendet werden. Er erlaubt es, alle Netzwerkanfragen an die Netzwerkumgebung des Desktop-PC's weiterzuleiten. Hierzu wird die Funktion *Redirect NetLib calls to host TCP/IP* (im Menüpunkt *Settings | Properties*) aktiviert.

## **4.3. PalmOS und Java**

Eine eigene Vorgabe für die Referenzimplementierung der Studienarbeit ist, sowohl für die Entwicklung auf dem Server als auch auf dem Palm als mobiles Endgerät, Java einzusetzen. Zwar empfiehlt 3COM für die Implementierung eigener Software die kommerzielle C++-Entwicklungsumgebung von Metrowerks, doch seit längerer Zeit stehen auch andere frei zugängliche Umgebungen zu Verfügung. Durch die ersten Implementierungen von Java Virtual Machines öffnet sich der Palm nun auch für Java. Anders als beim herkömmlichen PC mit Betriebssystemen wie Microsoft Windows, Mac OS oder Linux stehen den entwickelten Applikationen auf einem PDA jedoch viel geringere Ressourcen zur Verfügung. Der gigabyteverwöhnte Softwareentwickler findet auf den mobilen Geräten in der Regel nur 2 bis 32 MByte Speicher vor, die nicht nur als RAM fungieren, sondern ebenfalls als Festplattenersatz dienen. Die Taktfrequenzen der Prozessoren befinden sich ebenso nicht im Gigahertzbereich, sondern belaufen sich bei ca. 20MHz. Eine Javaumgebung wie die Virtual Maschine der Java Standard Edition von Sun ist also nicht geeignet für den Palm.

---

<sup>2</sup> Hinweis: Verwendet der HotSync Manager der Palm-Synchronisationssoftware auf dem PC dieselbe serielle Schnittstelle wie Mocha W32 PPP, so muss dieser während der Kommunikation deaktiviert werden.

Derzeit existieren folgende Virtual Machines:

- Kilobyte Virtual Maschine (KVM) von Sun
- Waba VM von Wabasoft
- Jbed VM von Esmertec AG
- VisalAge Micro Edition (VAME) von IBM
- MicrochaiVM von Hewlett-Packard
- KadaVM von Kada Systems

Für die Referenzimplementierung des Börseninformationssystems kommt die Kilobyte Virtual Maschine von Sun Microsystems zum Einsatz. Die KVM ist Bestandteil der Java 2 Micro Edition und stellt in der verwendeten Version eine Weiterentwicklung des Spotless-Projektes von SunLabs dar.

### 4.3.1 Die Java 2 Micro Edition

Auf der JavaOne99, der größten Java-Entwickler-Konferenz, wurde erstmals die Neugruppierung der Java Technologie von Sun in drei Bereiche vorgestellt. Jeder dieser Bereiche ist auf ein spezielles Marktsegment ausgerichtet.

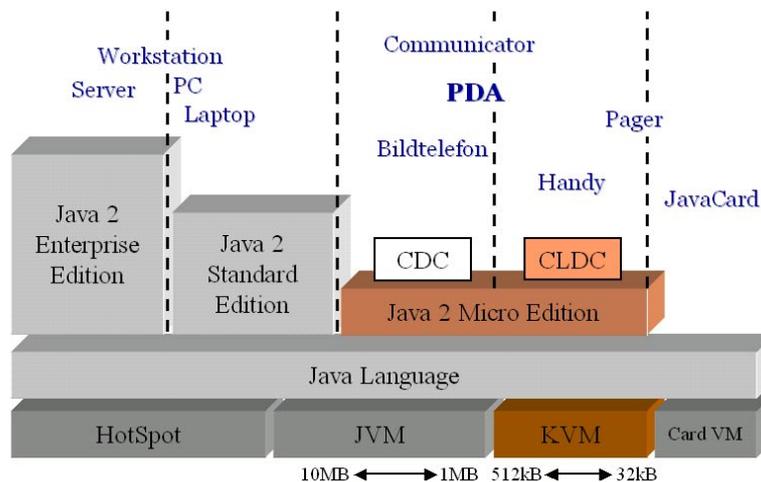


Abbildung 4-5: Java 2 Technologie

- **Java 2 Enterprise Edition (J2EE)**  
Die J2EE ist besonders für die Unternehmen ausgelegt, die ihren Kunden, Lieferanten und Mitarbeitern solide, komplette und skalierbare Internetbusiness-Server-Lösungen zur Verfügung stellen wollen
- **Java 2 Standard Edition (J2SE)**  
Die J2SE bezieht sich mehr auf den derzeit etablierten Desktopmarkt.
- **Java 2 Micro Edition (J2ME)**  
Mit der J2ME wendet sich Sun besonders an:
  - Hersteller von eingebetteten Systemen, die eine Vielzahl von unterschiedlichen Informationssystemen entwickeln,
  - Serviceprovider, die ihre Inhalte über diese Geräte verteilen,
  - Informationsanbieter, die ihre Informationen für kleine, ressourcenbegrenzte Geräte anbieten und deren Konsumenten.

Innerhalb der Editionen werden die Java Virtual Machine sowie spezielle Bibliotheken und API's für die entsprechenden Produkte definiert. Mit der Java 2 Micro Edition ist nun auch eine Java-Schnittstelle für den Gebrauch auf ressourcenbegrenzten Geräten vorhanden. Da die Java-Standardbibliotheken für diese Geräte vollkommen überdimensioniert sind, wurden für die Micro Edition sinnvolle Untermengen oder ganz neue API's spezifiziert. Mittelpunkt der J2ME ist die Virtual Machine. Diese wurde speziell auf die Restriktionen von eingebetteten Systemen, Mobilfunkgeräten, Organizer und Personal Digital Assistants zugeschnitten. Für sie ist nur ein Speicherbereich von einigen Kilobyte notwendig und wird von Sun als Kilobyte Virtual Machine (KVM) bezeichnet.

#### 4.3.1.1 Der Aufbau der Java 2 Micro Edition

Die Zielgeräte der Java 2 Micro Edition haben viele Eigenschaften gemeinsam, unterscheiden sich aber auch in Leistung und Funktionalität. Die Architektur der J2ME ist modular und skalierbar. Dieses sichert eine flexible Entwicklung und die Unterstützung der unterschiedlichsten Arten von Geräten. Die Java 2 Micro Edition unterteilt sich hierzu in unterschiedliche Schichten. Die unterste Schicht bildet das Betriebssystem, wie zum Beispiel PalmOS, Windows oder Sun.

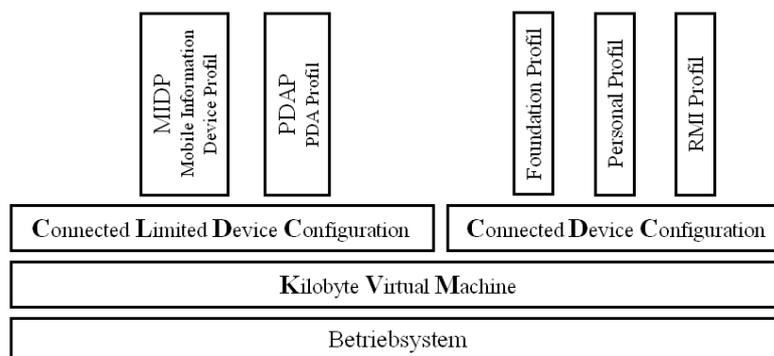


Abbildung 4-6: Aufbau der Java 2 Micro Edition

Auf der Betriebssystemebene setzt eine plattformunabhängige Implementierung der KVM auf. Die Merkmale und Funktionalitäten der KVM werden grundsätzlich durch die Konfiguration und ein darauf aufbauendes Profil beschrieben.

Die **Konfigurationen** beziehen sich auf Klassen von Geräten mit ähnlicher Ausstattung und Leistung. Hier werden die Grundfunktionen (low-level Klassen für Ein- und Ausgabe sowie der Sprachumfang) der KVM beschrieben.

Um eine Zersplitterung der J2ME vorzubeugen, wird nur eine geringe Anzahl von Konfigurationen entwickelt. Gegenwärtig existieren zwei Konfigurationen:

- Die **Connected Limited Device Configuration (CLDC)** setzt Geräte mit mindestens 128-512 KByte Speicher und eine Netzwerkverbindung mit 9600 Bit/s voraus. Sie wurde speziell für die Geräteklassen Mobiltelefone, Pager und PDA's spezifiziert.
- Die **Connected Device Configuration (CDC)** ist für etwas leistungsfähigere Geräte bestimmt, wie beispielsweise SetTop-Boxen, Bildtelefone und Spielkonsolen mit mindestens 512 KByte ROM, 256 KByte RAM und einer schnellen Netzwerkverbindung. Die CDC befindet sich derzeit noch in der Spezifikationsphase.

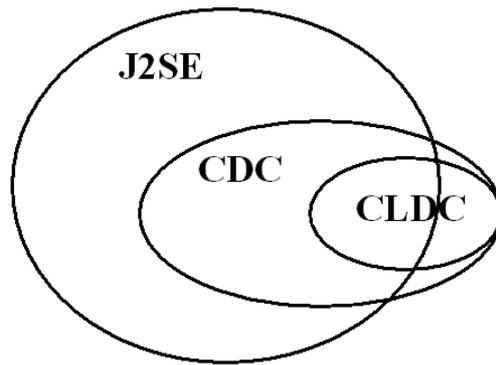


Abbildung 4-7: Beziehungen zwischen der J2SE und den Konfigurationen der J2ME

Um eine Abwärtskompatibilität zwischen den Konfigurationen zu gewährleisten, ist die CLDC auf einer Teilmenge der CDC aufgebaut.

Auf der Konfiguration setzt das **Profil** auf, das jeweils nur für eine bestimmte Klasse von Zielgeräten definiert ist. Im Profil werden unter anderem die Klassen für die Benutzerschnittstellen oder die persistente Datenhaltung spezifiziert.

Derzeit existieren für die CLDC das **Mobile Information Device Profile (MIDP)** und das **Personal Digital Assistant Profile (PDAP)**, die sich jedoch noch in unterschiedlichen Stufen des Spezifikationsprozesses befinden.

#### 4.3.1.2 Die KVM Version 1.0

Für die im Rahmen der Studienarbeit entwickelte prototypische Implementierung des Börseninformationssystems wird die Referenzimplementierung der CLDC-KVM in der Version 1.0 verwendet. Die Implementierung ist für die Desktop-Betriebssysteme Windows und Solaris sowie für PalmOS vorhanden. Der verwendeten KVM ist jedoch eine Sonderstellung innerhalb der J2ME einzuräumen, da sie eine Weiterentwicklung des Spotless-Projektes von SunLabs ist und die hierbei ursprünglichen GUI- und Datenbankklassen zu Demonstrationszwecken in die Referenzimplementierung übernommen wurden. Da der Palm unter das noch nicht fertiggestellte PDA-Profil fällt, enthält die Referenzimplementierung der CLDC-KVM eine Palm-spezifische Klassenbibliothek, die kJava-API.

Um Java-Applikationen auf dem Palm laufen zu lassen, ist es zunächst notwendig, die Java Virtual Machine auf dem mobilen Gerät zu installieren. Hierzu werden die *kvm.prc* sowie die *kvmulti.prc* auf den Palm übertragen. Zur Generierung der Klassendateien aus dem Java-Quellcode, die letztendlich in der KVM ausgeführt werden sollen, sind im allgemeinen folgende Schritte notwendig:

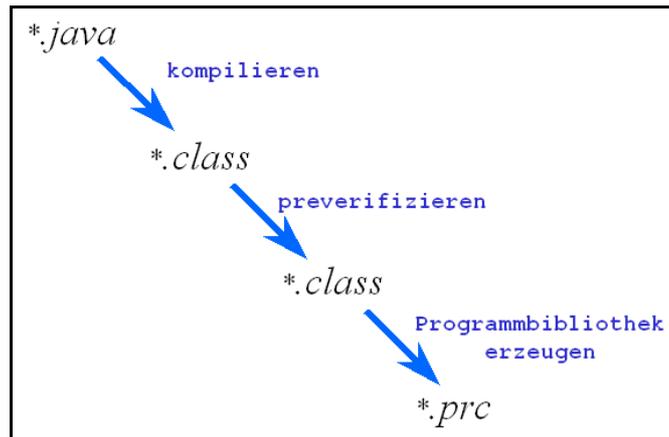


Abbildung 4-8: Erzeugen einer Java-Applikation für den Palm

1. Der Java-Quellcode wird zunächst mit dem JDK 1.2 kompiliert. Beim Aufruf des Compilers muss der Klassenpfad der CLDC-API mit angegeben werden.
2. Zum Ausführen der Klassendateien in der KVM ist es vorab notwendig diese zu prüfen. Dieser **Preverify**-Schritt fügt Hinweise in die Klassendatei ein, die dem KVM-Classloader die Prüfung der Klassen erleichtert und diese vor allen Dingen beschleunigen soll. Die Arbeit des Preverifiers schließt die Überprüfung gültiger Datentypen ein. Wurde die Prüfung erfolgreich abgeschlossen, kann die Klassendatei problemlos von der KVM geladen werden.
3. Letztendlich ist es notwendig, die entstandenen *class*-Dateien in ein für den Palm lesbares Format zu überführen. Hierzu werden alle notwendigen Klassendateien durch den Aufruf der Applikation *palm.database.MakePalmApp* in einer *prc*-Datei vereint, die dann auf dem Palm übertragen und dort ausgeführt werden kann. Bei diesem Schritt ist es des Weiteren möglich, der Applikation ein Icon zuzuweisen.

Die KVM stellt zur Entwicklung von Programmen vereinfachte Versionen der Pakete *java.lang*, *java.io* und *java.util* zur Verfügung. Für den Zugriff auf den persistenten Speicher sowie zur Erstellung von Benutzerschnittstellen ist das Paket *com.sun.kjava* vorhanden. Dieses ist jedoch nur auf die Verwendung einfacher GUI-Elemente mit absoluter Positionierung auf dem Display beschränkt. Weiterhin erschweren die sehr begrenzte Unterstützung für die Ereignisbehandlung sowie das Fehlen wichtiger Bestandteile, wie etwa Layout-Manager, Gruppierungselemente oder Menüs die Implementierung. Abhilfe schafft dabei eine Bibliothek eines Drittanbieters, die kAWT-Bibliothek in der Version 0.98 des kAWT-Projektes. Diese setzt auf die kJava-Klassen auf und behebt viele der beschriebenen Beschränkungen, indem eingeschränkte AWT-Klassen zur Verfügung gestellt werden. Die kAWT-Bibliothek ist aufwärtskompatibel zum Desktop-AWT, wodurch sich alle kAWT-Programme ohne Modifikation oder Emulator auf dem Desktop ausführen und testen lassen.

Zum Aufbau einer Socket-Verbindung wird bei der Micro Edition das Generic Connection Framework in dem Paket *javax.microedition.io* verwendet. Dieses stellt die Klasse *Connector* und verschiedene Schnittstellen für die einzelnen Verbindungsarten zur Verfügung.

Ziel bei der Implementierung der Studienarbeit ist es, die Börseninformationen zwischen Client und Server mit Hilfe von XML zu übertragen. Ein XML-Parser für den

Client, wie der *Oracle-XML-Parser*, der beim Server Verwendung findet, wäre dabei weitgehend überdimensioniert. Voraussetzung für den clientseitigen XML-Parser ist es, nur geringe Ressourcen zu verwenden und zugleich unter der Palm-KVM lauffähig zu sein. Der SAX-XML-Parser *TinyXML V0.7* von T. Gibara besteht aus lediglich sieben kleinen Klassendateien, die den Funktionsumfang des benötigten XML-Parsers erfüllen und entspricht den geforderten Eigenschaften. Mit einer geringfügigen Änderung, zum Zugriff auf Strings mit einer Länge über 256 Bytes, ist *TinyXML* in die Implementierung eingegliedert.<sup>3</sup>

#### 4.4. Datenbank auf dem Palm

Zur Darstellung der Börsendaten auch im Offline-Betrieb ist es notwendig, Daten auf dem Palm zu speichern. Ebenso die Profileinstellungen sollen nicht jedes Mal vom Benutzer erneut eingegeben werden und müssen demzufolge persistent sein. Zur Speicherung dieser Informationen bieten sich relationale oder objektorientierte Datenbanken an. Mittlerweile haben sowohl Sybase, Oracle als auch IBM eine relationale Datenbank für die Palm-Plattform herausgebracht, jedoch bietet keine dieser Datenbanken eine Schnittstelle zu Java mit Hilfe von JDBC (Java Database Connectivity) oder anderen Java fähigen APIs an (siehe [Fa00]).

Ein anderer Weg ist die Nutzung einzelner Dateien im Dateisystem zur Datenablage. Das Palm Betriebssystem kennt zwei Dateitypen: Kataloge mit der Endung *.pdb* zum Speichern von Daten und Programmdateien mit der Endung *.prc* in denen ausführbarer Code enthalten ist. Die Palm-Programme haben keinen direkten Zugriff auf das Dateisystem. Sie können lediglich Kataloge öffnen, lesen und schreiben. Innerhalb der Kataloge, auch Datenbanken genannt, können Daten in Record-Struktur gespeichert werden.

Die Klasse *com.sun.kjava.Database* der KVM bietet die Möglichkeit mit den Katalogen zu arbeiten. Den Zugriff mit Java auf einen Katalog zeigt Abbildung 4-9. Darin wird zunächst ein Katalog erzeugt, ein Wert gespeichert und gleich wieder ausgelesen.

---

<sup>3</sup> Die Referenzimplementierung der CLDC-KVM erlaubt lediglich das Arbeiten mit Strings bis zu einer Länge von 256 Bytes. Für Strings mit einer Länge über 256 Bytes muss ein Byte-Array verwendet werden.

```

privat final String dbName      = "Test DB";
privat final String dbCreator   = "DB_1";
privat final String dbType     = "DB_1";

public void writeData(String data)
{
    // Erzeugen des Katalogs
    if (!Database.create(0,
                        dbName,
                        decodeIdString(dbCreator),
                        decodeIdString(dbType),
                        false)) return;

    // Öffnen des Katalogs
    Database db = new Database (decodeIdString(dbType),
                              decodeIdString(dbCreator),
                              Database.READWRITE);

    // Schreiben der Daten in den Katalog
    db.addRecord(data.getBytes());

    // Lesen der Daten aus dem Katalog
    String newData = new String(db.getRecord(0));

    System.out.println(newData);
}

/**
 * Dekodiert einen 4-Zeichen-String in ein 32-bit-Integer.
 * @param str 4-Zeichen-String
 * @return 32-bit-Integer des Strings
 */
public static final int decodeIdString(String str)
{
    int id = 0;
    for (int i = 0; i < 4; i++)
        id = 256 * id + (int) str.charAt(i);
    return id;
}

```

Abbildung 4-9: Zugriff mit Java auf den Palm-Katalog

Die Palm-Datenbank dient zur Ablage von Records, die in Form eines Byte-Arrays abgelegt werden. Zur Speicherung von Datentypen wie Integer, Strings und anderen ist es daher notwendig, diese in Byte-Arrays zu konvertieren. Die Methode *int16uToByte* aus Abbildung 4-10 zeigt die Umwandlung eines 16 Bit vorzeichenlosen Integerwertes in ein Byte-Array und die Methode *byteToInt16u* analog für die Gegenrichtung.

```

/**
 * Konvertiert ein 16 Bit unsigned Integer in
 * ein ByteArray.
 * @param i der zu konvertierende IntegerWert
 * @return ByteArray des Integerwertes
 */
public static byte[] int16uToByte(int i)
{
    byte[] image = new byte [2];

    image[0] = (byte) ((i >> 8) & 0xFF);
    image[1] = (byte) ( i          & 0xFF);

    return image;
}

/**
 * Konvertiert ein ByteArray in ein 16 Bit
 * unsigned Integer.
 * @param m das zu konvertierende ByteArray
 * @return IntegerWert des ByteArrays
 */
public static int byteToInt16u(byte[] m)
{
    int tmpVal = m[0];
    if (tmpVal < 0) tmpVal += 256;
    int val = tmpVal;
    tmpVal = m[1];
    if (tmpVal < 0) tmpVal += 256;
    val = val * 256 + tmpVal;

    return val;
}

```

Abbildung 4-10: Konvertierung der Integerwerte

Der Einsatz von Katalogen als Persistenzmechanismus zieht hinsichtlich des direkten Zugriffs auf die Records Probleme mit sich. Ein Record lässt sich über dessen Index in der Palm-Datenbank ansprechen. Jedoch bezieht sich dieser nicht direkt auf das entsprechende Record, sondern nur auf deren Nummerierung innerhalb des Kataloges. Wird im Katalog ein Record gelöscht, so ändert sich demzufolge der Index aller nachfolgenden Einträge.

Hinsichtlich des Funktionsumfangs einer „normalen Datenbank“ bietet der Palm-Katalog lediglich die Eigenschaft Daten darin zu speichern. Diese Beeinträchtigungen erschweren die Nutzung der Persistierung von Daten und fielen daher in der Implementierung der Endanwenderapplikation auf dem Palm unter besondere Berücksichtigung. So ist beispielsweise eine Breitensuche zum Zugriff auf bestimmte Informationen innerhalb eines gesamten Kataloges teilweise unausweichlich.

## 4.5. Java 2 auf dem Server

Wie bereits im Kapitel 2.2 Technische Konzepte ab Seite 12 erwähnt, ist eine eigene Vorgabe, die Implementierungen des Server in Java zu realisieren. Hierzu wird das JDK 1.2.2 der Java 2 Standard Edition verwendet.

Die Standard-Java-Klassen sind durch ein paar Java-API's erweitert worden, die in verschiedenen Komponenten des Servers Anwendung finden.

Eine Erweiterung stellt der ORO Matcher in der Version 1.1 von D.F. Savarese dar. Es handelt sich hierbei um einen Perl-Parser für Java, dessen Funktionalität im Wrapper des Börseninformationsservers Verwendung findet. Durch reguläre Ausdrücke können so die Börseninformationen von den Seiten der WWW-Anbieter gesammelt werden. Weiterführende Informationen finden sich hierzu im Kapitel 5.1 Wrapper ab Seite 44.

Mit Hilfe der Mail-API von Sun Microsystems ist der Wrapper um eine Benachrichtigungskomponente erweitert. Dessen Aufgabe ist es, den Administrator des Börseninformationsservers bei Fehlfunktionen, über die Ursachen und Schwere des Fehlers zu informieren. Deren Gründe werden in Kapitel 5.1 Wrapper ab Seite 44 ausführlicher behandelt.

Weiterhin wird der Oracle-XML-Parser von Oracle eingesetzt. Der DOM-Parser, ist Bestandteil der Kommunikationsschnittstelle am Server. Die Anfragen des Clients werden durch deren Einsatz geparkt und einer Bearbeitungsroutine zugeordnet. Ausführliche Informationen hierzu befinden sich im Kapitel 5.6 Kommunikationsschnittstellen auf dem Server und Client ab Seite 73.

## **4.6. Datenbank auf dem Server**

Für den Einsatz des Applikationsserver des Börseninformationssystem ist es notwendig, Daten zu speichern. Hierzu wird ein Datenbankmanagementsystem verwendet. Zur Auswahl standen dabei:

- Informix
- Oracle
- DB2
- MySQL Version 3.23
- Borland Interbase 6.0

Für das Börseninformationssystem kommt Borland Interbase in der Version 6.0 zum Einsatz. Es sei hierbei zu erwähnen, dass der Einsatz der anderen Systeme zum gleichen Ergebnis geführt hätte.

## 5. Die Teilkomponenten des Systems

Im Kapitel 2.2 Technische Konzepte ab Seite 12 wurde bereits der Systemaufbau des Börseninformationssystems skizziert. Die Client-Server-Architektur gliedert sich in folgende Teilkomponenten:

Server

- Wrapper
- Vorreduktion
- Datenbank
- Anfragebearbeitung und profilabhängige Reduktion

Client

- Anfragebearbeitung
- Graphisches User Interface
- Profilverwaltung
- Informationscache

Server und Client

- Kommunikationsschnittstelle

In den folgenden Abschnitten werden die Teilkomponenten näher beschrieben. Es werden die Aufgaben, die Realisierung und die Probleme der einzelnen Komponenten dabei genauer erläutert.

### 5.1. Wrapper

Der Wrapper bildet die Kernkomponente zum Sammeln der Börseninformation am Server. In regelmäßigen Abständen baut dieser automatisch eine HTTP-Verbindung zu den verschiedenen Anbietern auf und lädt die Informationen von den entsprechenden Seiten herunter. Zum Herausfiltern der benötigten Informationen aus den Webseiten ist in dem Wrapper der ORO-Matcher, ein Perl-Parser für Java, integriert.

Wie in dem Kapitel 2.1 Anwendungsspezifische Konzepte ab Seite 9 beschrieben, werden Analysen, Ad-Hoc-News, Fundamentaldaten und Kurslisten für das Börseninformationssystem benötigt. Der Wrapper besteht aus mehreren HTML-Parsern, die nach bestimmten zeitlichen Festlegungen die Daten von den Seiten der Anbieter mittels regulärer Ausdrücke herausfiltern. Analysen werden wochentags zwischen Börsenbeginn und -schluss alle vier Stunden von [www.aktiencheck.de](http://www.aktiencheck.de) geparkt. Ebenfalls, aber alle acht Stunden erfolgt die Sammlung der Ad-Hoc-News von den Seiten [www.aktiencheck.de](http://www.aktiencheck.de). Die Kurse der Aktien stellt [www.goingpublic-online.de](http://www.goingpublic-online.de) auf seiner Homepage zur Verfügung. Die Kursliste wird alle 60 Minuten nach dem Börsenstart geparkt. Der Anbieter [www.goingpublic-online.de](http://www.goingpublic-online.de) stellt die Fundamentaldaten der Aktien am Neuen Markt auf seinen Webseiten dar. Diese werden täglich vor dem Börsenstart herausgefiltert. Die gesammelten Daten werden in der Datenbank auf dem Server gespeichert. Einen Überblick zu deren Speicherung und dem dazu verwendeten Datenbankschema gibt das Kapitel 5.3 DB auf dem Server + DB-Schema ab Seite 61.

In dem folgenden Abschnitt wird die Datensammlung beispielhaft an den Fundamentaldaten gezeigt. Es wird die Funktionsweise des HTML-Parser vorgestellt,

der die Elemente der Fundamentaldaten einer Aktie herausfiltert. Zu den Fundamentaldaten der Aktien gehören folgende Attribute:

- Name
- WPKN
- Kürzel
- Branche
- Beginn der Börsennotierung
- Emissionspreis
- Der erste festgestellte Börsenpreis

Die Seite mit der URL <http://www.goingpublic-online.de/review/neuermarkt/-neuermarkt.hbs?abc=z> bietet eine Tabelle mit allen am Neuen Markt erhältlichen Aktien. Die Abbildung 5-1 zeigt einen Ausschnitt dieser Tabelle.

Unternehmen	Akt.	Diff.	% Diff.	Hoch	Tief	Handelsvol. in Tsd.	Datum	Zeit
3U TELEKOMMUNIK...	0,00	0,00	0,00	0,00	0,00	0,00	27.04.	00:00
AMBO INTERN. EL...	19,00	-1,00	-5,26	19,10	18,00	1.232,00	27.04.	10:44
AAP IMPLANTATE	12,46	+0,35	+2,89	12,45	12,20	1.398,00	27.04.	11:36
ABIT AG AKTIEN	6,35	+0,10	+1,60	6,40	6,25	2.890,00	27.04.	11:10
AC-SERVICE AG N...	3,95	+0,05	+1,28	3,95	3,85	350,00	27.04.	10:58
ACG AG F. CHIPK...	25,80	-1,80	-6,52	26,75	25,46	67.723,00	27.04.	11:32
AD PEPPER MEDIA...	2,70	-0,05	-1,82	3,10	2,70	14.540,00	27.04.	11:31
ADCON TELEMETRY...	9,55	-0,10	-1,04	9,65	9,40	3.890,00	27.04.	11:22
ADLINK INTERNET...	1,95	+0,05	+2,63	1,97	1,91	7.637,00	27.04.	11:23
ADDRI AG AKTIEN...	2,29	-0,01	-0,43	2,29	2,29	1.000,00	27.04.	10:14
ADS SYSTEM AG A...	5,20	-0,05	-0,95	5,20	5,20	811,00	27.04.	10:46
ADVA AG OPTICAL...	10,93	-0,77	-6,58	11,30	10,75	67.368,00	27.04.	11:26
ADVANCED MEDIEN...	0,83	+0,01	+1,22	0,83	0,81	3.325,00	27.04.	10:40

Abbildung 5-1: Browserdarstellung der Aktienliste

Der HTML-Quellcode der Tabelle erlaubt einen Zugriff auf die Fundamentaldaten der einzelnen Aktien. Abbildung 5-2 zeigt einen Teil des HTML-Quellcodes der Tabelle. Die markierten Stellen heben die für das Börseninformationssystem wichtigen Daten hervor. Hierbei handelt es sich um zwei Links zu weiteren Elementen der Fundamentaldaten und dem Namen der Aktie. Innerhalb der Links ist die WPKN enthalten, in diesem Beispiel 516790.

```

...
<tr align=right class=w>
  <td align=left nowrap>
    <a href="/factsheet/factsheet.hbs?wkn=516790">
      3U TELEKOMMUNIK...</a>
  </td>
  <td>
    <a href="/factsheet/aktkursdaten.hbs?wkn=516790">
...

```

Abbildung 5-2: HTML-Quellcode für ersten Scanvorgang

Mit Hilfe von regulären Ausdrücken werden die markierten Informationen aus dem HTML-Quellcode herausgefiltert. Der zum Parsen dieser Informationen benötigte reguläre Ausdruck ist in Abbildung 5-3 dargestellt. Die markierten Zeichen entsprechen den Rückgabewerten des Parsers.

```

...
<tr align=right class=.>
  <td align=left nowrap>
    <a href="(.*).wkn=(\d{6})">(.*)</a>
  </td>
  <td>
    <a href="(.*).wkn=\d{6}">
  </td>
...

```

Abbildung 5-3: regulärer Ausdruck des ersten Scanvorgangs

Der Parser scannt die gesamte HTML-Seite und gibt pro gefundener HTML-Sequenz, also pro Aktie, 4 Elemente zurück. Im ersten Scanvorgang konnten die Namen und WPKN der Aktien extrahiert werden. Die restlichen Elemente erhält man durch das Parsen von zwei weiteren Seiten pro Aktie. Die Verweise zu diesen Seiten werden aus den gesammelten Informationen im ersten Scanvorgang gebildet:

- <http://www.goingpublic-online.de/factsheet/factsheet.hbs?wkn=516790> und
- <http://www.goingpublic-online.de/factsheet/aktkursdaten.hbs?wkn=516790>

Die Abbildung 5-4 zeigt einen Ausschnitt aus dem HTML-Quellcode, den man beim Verfolgen des ersten Links erhält. Aus der Seite können die Elemente Kürzel und Branche einer Aktie gewonnen werden. Durch den regulären Ausdruck in Abbildung 5-5 werden diese beiden herausgefiltert.

```

...
  <td class="grau" valign="top">Reutersk&uuml;rzel /
                                B&ouml;rsek&uuml;rzel</td>
  <td class="grau">UUUG.F&nbsp;/&nbsp;/UUU</td>
</tr>
<tr>
  <td class="weiss" valign="top">Marktsegment</td>
  <td class="weiss">NM</td>
</tr>
<tr>
  <td class="grau" valign="top">Branche / Subbranche</td>
  <td class="grau">Telecomm. / Provider</td>
...

```

Abbildung 5-4: HTML-Quellcode für den zweiten Scanvorgang

```

...
  <td[\n| ]class="grau" valign="top">Reutersk.uuml.rzel /
                                B.ouml;rsek.uuml;rzel</td>\n
  <td class="grau">.*nbsp;..nbsp;(.*)</td>
</tr>
<tr>
  <td class="weiss"\nvalign="top">Marktsegment</td>
  <td class="weiss">NM</td>
</tr>
<tr>
  <td\nclass="grau" valign="top">Branche / Subbranche</td>
  <td class="grau">(.*) /.*\n</td>
...

```

Abbildung 5-5: regulärer Ausdruck des zweiten Scanvorgangs

Der zweite Link verweist auf eine HTML-Seite. Diese enthält die restlichen Elemente:

- Beginn der Börsennotierung
- Emissionspreis
- der erste festgestellte Börsenpreis

Der Quelltext und der dazugehörige reguläre Ausdruck ist in Abbildung 5-6 und Abbildung 5-7 dargestellt.

```
...
    <td class="weiss">Emissionspreis ( 25.11.1999)</td>
    <td class="weiss">30,00 Euro</td>
</tr>
<tr>
    <td class="grau">Performance zum Emissionspreis</td>
    <td class="grau"> -76,67 % </td>
</tr>
<tr>
    <td class="weiss">Erstnotiz</td>
    <td class="weiss"> 30,50 Euro</td>
</tr>
<tr>
    <td class="grau">
...

```

Abbildung 5-6: HTML-Quellcode für den dritten Scanvorgang

```
...
    <td[\n| ]class=".*">Emissionspreis . (\d+).(\d+).(\d+).*</td>
    <td[\n| ]class="weiss">\n*(.*)</td>
</tr>
<tr>
    <td class="grau">Performance zum Emissionspreis</td>\n
    <td class="grau">.*</td>
</tr>
<tr>
    <td class="weiss">Erstnotiz</td>\n
    <td class="weiss">{.*}</td>
</tr>
<tr>
    <td class="grau">
...

```

Abbildung 5-7: regulärer Ausdruck des dritten Scanvorgangs

Nach drei Scanvorgängen sind die Fundamentaldaten der Beispielaktie vollständig herausgefiltert. Die Tabelle in Abbildung 5-8 zeigt abschließend die herausgefilterten Fundamentaldaten der Beispielaktie.

<i>Bezeichnung</i>	<i>Wert</i>
Name	3U TELEKOMMUNIK...
WPKN	5167090
Kürzel	UUU
Branche	Telecomm.
Beginn der Börsennotierung	25.11.1999
Emissionspreis	30,0 €
Der erste festgestellt Börsenpreis	30,5 €

Abbildung 5-8: Fundamentaldaten der Beispielaktie

Es hat sich herausgestellt, dass das Parsen der Fundamentaldaten im Gegensatz zu den Kurslisten, Analysen und Ad-Hoc-News am zeitintensivsten ist. Zu Beginn wird die Tabelle mit den zur Zeit 341 Aktien geparkt. Anschließend werden pro Aktie jeweils 2 Scanvorgänge durchgeführt, in denen die restlichen Daten gewonnen werden. Damit ergeben sich insgesamt 683 Scanvorgänge, die der Wrapper zum Sammeln Fundamentaldaten ausführen muss.

### Probleme

Der Wrapper ist an die Struktur der Anbieterseiten und den unmittelbaren Kontext der zu findenden Informationen auf diesen Seiten angepasst. Treten Strukturveränderungen auf den Webseiten auf, so kann der Wrapper die benötigten Informationen beziehungsweise die Verweise zu diesen nicht mehr eindeutig finden. Die regulären Ausdrücke bieten zwar ein mächtiges Werkzeug an, die kleine Veränderungen der Webseiten vernachlässigbar machen, doch bei umfangreicheren Strukturveränderungen bleibt eine manuelle Anpassung des Wrappers unausweichlich.

Ein direkter Zugriff auf die benötigten Börseninformationen, wie zum Beispiel die Direktanbindung an die Datenbanken der Informationsanbieter, ohne den Umweg über HTML-Seiten und dem Informationsparsen aus diesen Seiten, würde eine deutliche Verbesserung darstellen. Somit wäre eine sichere Arbeitsweise der Komponente zum Sammeln der Börseninformationen am Server gewährleistet.

## 5.2. Vorreduktion

Im Kapitel 3 Datenreduktion ab Seite 17 wurden drei Datenreduktionsmethoden hinsichtlich ihrer Tauglichkeit für das Börseninformationssystem untersucht. Zum Einsatz kommt ein Reduktionsverfahren, beruhend auf verschiedenen Ansätzen der Dimensionsreduktion aus dem Bereich des Data Mining. Dem Verfahren liegt eine Reduktionsberechnung zu Grunde. Im Hinblick auf die für die Berechnung benötigte Zeit, wurde diese in Vor- und profilabhängige Reduktion geteilt. In diesem Kapitel wird die Vorreduktion und deren Anpassung an die Anforderungen des Börseninformationssystems dargestellt.

Die Aufgaben der Vorreduktion sind:

1. Aus den Informationen (Fundamentaldaten, Kurslisten, Ad-Hoc-News und Analysen), die durch den Wrapper gewonnen werden und in der Datenbank abgelegt sind, für den Nutzer relevante Daten zu gewinnen. Diese werden in die folgenden fünf Kategorien untergliedert:
  - Analysen
  - News

- Tops und Flops
  - Schwankungen
  - Neuemissionen
2. Die gewonnenen Informationen werden hinsichtlich ihrer Relevanz (Wichtigkeit) bewertet. Diese wird mit Hilfe einer Signifikanz dargestellt.

Der folgende Abschnitt ist in fünf Teile gegliedert, die nach den in Punkt 1 vorgestellten Kategorien benannt sind. Hierbei wird erläutert, wie Berechnungen auf den vorhandenen Börsendaten durchgeführt werden, um daraus Informationen entsprechend den Kategorien zu gewinnen. Weiterhin wird darauf eingegangen, wie den gewonnenen Informationen Signifikanzen zugeordnet werden.

## Die Analysen

Um einen besseren Überblick über ein Wertpapier zu haben, werden von Analysten Einschätzungen in Form von Analysen verfasst. In der heutigen Zeit existieren jedoch immer mehr Analysten, die diese Bewertungen vornehmen. So werden beispielsweise um die 900 Analysen über Aktien am Neuen Markt innerhalb von 20 Werktagen veröffentlicht. Bei dieser Menge von Informationen fällt es schwer, einen Überblick zu behalten. Ziel bei der Vorreduktion der Analysen ist es deshalb, eine ausgezeichnete Menge von Analysen zu bilden. Zur Lösung dieser Aufgabe ist die Vorreduktion in folgende Schritte gegliedert:

1. Kategorisierung der Analysen
2. zeitliche Bewertung der Analysen
3. Gruppierung der Analysen nach ihrer WPKN
4. Normierung
5. Erstellung der vorreduzierten Analyse

Die vom Wrapper gesammelten Analysen bestehen aus den Attributen WPKN, Datum, Überschrift (Subject), Analyst und Inhalt. Einige dieser Attribute fließen in die Vorreduktionsberechnung mit ein.

Die in Abbildung 5-9 gezeigte Tabelle sollen die Daten einiger Analysen darstellen und werden im weiteren Verlauf als durchgehendes Beispiel verwendet.

<i>Datum</i>	<i>WPKN</i>	<i>Subject</i>	<i>Analyst</i>	<i>Text</i>
28.02.2001	909247	SCM Microsystems untergewichten	Independent Research	Die Analysten von ...
01.03.2001	909247	SCM Microsystems weniger interessant	Neuer Markt Inside Online	Ein Bericht von ...
01.03.2001	880064	Bipop Carire kaufen	Credit Suisse First Boston	Die Experten von ...

Abbildung 5-9: Beispielanalysen

Für die Reduktionsberechnung werden lediglich die Analysen der letzten 20 Werktage berücksichtigt. Es könnten auch weiter zurückliegende bewertet werden, jedoch ist eine Analyse, die beispielsweise zwei Monate zurückliegt, nicht mehr von Bedeutung. Diese Analysteneinschätzung wurde entweder vom Markt gänzlich ignoriert oder ist im Kurs der Aktie bereits eingeflossen.

Im ersten Schritt werden die Analysen den Kategorien Verkaufen, Halten und Kaufen zugeteilt. Zu einer Analyse gehört unter anderem eine Überschrift (Subject). Diese wird nach bestimmten Schlüsselwörtern untersucht. Die Analysten verwenden beispielsweise

für eine zum Kaufen eingestufte Aktie Schlüsselwörter, wie Outperformer, Upgrade, Buy oder Strong Buy.

<i><b>Kaufen</b></i>	kaufen, strong buy, akkumulieren, Outperformer, interessant, Trading Buy, Depotbeimischung, über Erwartungen, Upgrade, spekulativ, Standardinvestment, Basisinvestment, Aufwärtspotenzial, Übergewichten, Buy, überdurchschnittlich, unterbewertet, ...
<i><b>Verkaufen</b></i>	verkaufen, unter den Prognosen weiter reduzieren, reduzieren, meiden, downgrade, überbewertet, untergewichtet, kein kauf, underperformer, trading sell, unter druck, nicht kaufen, weiter abwärts, weniger interessant, hoch bewertet, zu teuer, ...

Abbildung 5-10: Schlüsselwörter zum Katalogisieren der Analysen

Die Suche nach Schlüsselwörtern in den Subjects der Analysen beschränkt sich nur auf die Kategorien Kaufen und Verkaufen. Für die restlichen, nicht zugeordneten Analysen, wird eine Halteempfehlung angenommen.

Ziel der Kategorisierung ist es, den Analysen einen numerischen Wert zuzuweisen, der für die weitere Berechnung verwendet werden kann. Dabei werden den Analysen, wie in der Abbildung 5-11 dargestellt, eine 1, 0 oder -1 zugewiesen.

<i>Empfehlungsart</i>	<i>Bewertung</i>
Verkaufen	-1
Halten	0
Kaufen	1

Abbildung 5-11: Bewertungszuordnung bei Analysen

Die Bewertung der Beispielanalysen sieht dementsprechend folgendermaßen aus.

<i>Datum</i>	<i>WPKN</i>	<i>Subject</i>	<i>Analyst</i>	<i>Inhalt</i>	<i>Bewertung</i>
28.02.2001	909247	SCM Microsystems untergewichten	Independent Research	...	<b>-1</b>
01.03.2001	909247	SCM Microsystems weniger interessant	Neuer Markt Inside Online	...	<b>-1</b>
01.03.2001	880064	Bipop Carire kaufen	Credit Suisse First Boston	...	<b>1</b>

Abbildung 5-12: Bewertung der Beispielanalysen

Im zweiten Schritt wird das Alter einer Analyse berücksichtigt. Eine Kaufempfehlung, deren Veröffentlichung 7 Tage zurückliegt, hat für den Nutzer weniger Gewicht, als eine vor 2 Tagen veröffentlichte. Dasselbe gilt für Halte- und Verkaufsempfehlungen. In Abbildung 5-13 wird gezeigt, mit welchem Faktor die zuvor ermittelte Bewertung multipliziert wird.

<i>Alter der Analyse in Werkstage</i>	<i>Faktor</i>
0	1
1 – 3	0,8
4 – 7	0,5
8 – 13	0,25
14 – 20	0,1
> 20	0

Abbildung 5-13: Altersbewertung bei Analysen

Nach der Betrachtung des Alters einer Analyse folgt die Gruppierung dieser nach den WPKN. Ziel hierbei ist es nicht einzelne Analysen zu betrachten, sondern den Fokus auf die Gesamtheit aller Analysen eines Wertpapiers zu legen. Erfahrungsgemäß veröffentlichen verschiedene Analysten zu einer Aktie unterschiedliche Einschätzungen und Meinungen. Um sich einen umfassenden Eindruck über ein Wertpapier bilden zu können, studiert der Aktionär in der Regel die verschiedenen dazugehörigen Analysen und verschafft sich so ein Gesamtbild. Dieser Vorgang wird in der Berechnung der Datenreduktionsmethode nachgebildet. Dabei werden die zuvor gewonnenen Bewertungen der Analysen einer Aktie summiert. Die Abbildung 5-14 zeigt das Ergebnis dieser Addition für das Beispiel.

<i>WPKN</i>	<i>Summe der Bewertungen<sup>4</sup></i>
909247	<b>-1,8</b>
880064	<b>1</b>

Abbildung 5-14: Gruppierung der Beispielanalysen

Gewöhnlich liegt das ermittelte Bewertungsmaß für die Analysen eines Wertpapiers um den Nullpunkt. Es gibt jedoch durchaus Unternehmen, deren Analysen nach der bisherigen Berechnung mit über 5 bewertet werden und im Gegensatz dazu auch Aktien mit negativen Werten bis zu -3. In der Regel ist festzustellen, dass weniger negative als positive Werte für die Analysen erreicht werden. Dieser Umstand ist auf das Verhalten der Analysten zurückzuführen, die im allgemeinen häufiger Kauf- als Verkaufsempfehlungen veröffentlichen. Für eine gesamte Bewertung sind die oben genannten Probleme störend. Außerdem ist es notwendig für einen Vergleich mit den Bewertungsmaßen, resultierend aus den Vorreduktionen von News, Top, Flops, Schwankungen und Neuemissionen, auf einen gemeinsamen Wertebereich abzubilden. Erreicht wird dieses durch die Normierung der Bewertung. Bei diesem Vorgang werden alle positiven Werte durch die größte positive Summe und analog alle negativen durch den Betrag der kleinsten negativen dividiert. Das daraus entstandene Ergebnis ist die Signifikanz für die Analysen eines Wertpapiers und liegt zwischen den Werten -1 und 1. Bei einer gegebenen größten positiven Zahl von 6,5 und einer kleinsten negativen von -2,3 ergeben sich für die Beispieldaten die in Abbildung 5-15 dargestellten Signifikanzen.

---

<sup>4</sup> Für die Altersbewertung der Beispielanalysen ist der 01.03.2001 als aktueller Tag zu Grunde gelegt.

<i>WPKN</i>	<i>Signifikanz</i>
909247	<b>-0,783</b>
880064	<b>0,154</b>

Abbildung 5-15: Signifikanzen der Beispielanalyse

Der Nutzer des Systems ist an Aktien interessiert, die besonders positiv bzw. negativ von den Analysten eingeschätzt wurden. Dieser Fakt spiegelt sich in der ermittelten Signifikanz für die Analysen wider.

Im letzten Schritt der Reduktionsberechnung wird für jede Aktie eine vorreduzierte Analyse erstellt. Die zu erstellenden Analysen setzen sich aus folgenden Komponenten zusammen:

- ID
- Sig
- WPKN
- Subject<sup>5</sup>
- Inhalt<sup>5</sup>
- Analyst<sup>5</sup>
- Datum<sup>5</sup>
- Kurs<sup>6</sup>
- Datum des aktuellen Kurs<sup>6</sup>

Die ID zur eindeutigen Identifikation der Information besteht aus 23 Zeichen. Die erste Stelle, einem „A“, dient zur Kennzeichnung der Kategorie Analyse. Anschließend folgt das Kürzel des Wertpapiers. Die nächsten Zeichen setzen sich aus dem Erstellungszeitpunkt der vorreduzierten Analyse zusammen. Das Datum und die laufende Nummer der letzten veröffentlichten Analyse auf den Webseiten des Anbieters [www.aktiencheck.de](http://www.aktiencheck.de) wird den Zeichen 15 bis 22 zugeordnet.

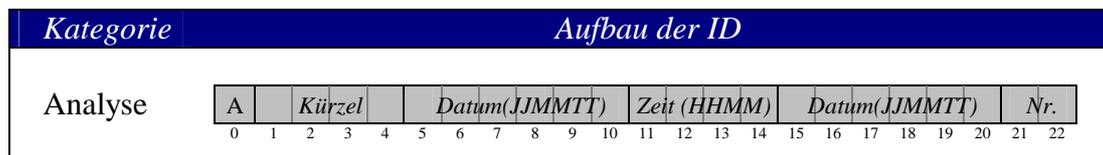


Abbildung 5-16: Aufbau der ID einer vorreduzierten Analyse

Abbildung 5-17 stellt die aus der Vorreduktion resultierende Beispielanalyse dar.

<i>ID</i>	ASMY_010301180001030101
<i>Sig</i>	-0,783
<i>WPKN</i>	909247
<i>Subject</i>	SCM Microsystems weniger interessant
<i>Inhalt</i>	Einem Bericht von NMI Online zufolge, hat ...
<i>Analyst</i>	Neuer Markt Inside Online
<i>Datum</i>	01.03.2001
<i>Kurs</i>	17,3 €
<i>Datum des aktuellen Kurses</i>	01.03.2001 17:59

Abbildung 5-17: vorreduzierte Beispielanalyse

<sup>5</sup> Die Daten beziehen sich auf die letzte veröffentlichte Analyse für das entsprechende Wertpapier.

<sup>6</sup> Die Daten beziehen sich auf den aktuellen Kurs der Aktie zum Erstellungszeitpunkt der vorreduzierten Analyse.

## Die Ad-Hoc-News

Ad-Hoc-News enthalten fundamentale Informationen über Unternehmen. Das können Veröffentlichungen von Geschäftszahlen über einen bestimmten Zeitraum (3 Monate, 6 Monate ...) sein, aber auch Fusionen und Kooperationen zwischen Firmen sind Themen von Ad-Hoc-News. Im Gegensatz zu den Analysen, bei denen Meinungen über ein Unternehmen veröffentlicht werden, besteht eine News aus Fakten, die die wirtschaftliche Lage der Firmen darlegen.

Wie auch bei den Analysen ist es für die Aktionäre in der heutigen Zeit schwer geworden, den Überblick über die Menge der veröffentlichten Ad-Hoc-News zu behalten. Durchschnittlich werden 20 News pro Werktag über Firmen, die am Neuen Markt gelistet sind, publiziert. Die Aufgabe der Vorreduktion in dieser Kategorie ist es, eine Menge von relevanten News aufzubauen. Hierzu wird in folgenden Schritten vorgegangen:

1. zeitliche Bewertung der Ad-Hoc-News
2. Gruppierung der News nach ihrer WPKN
3. Normierung
4. Erstellung der vorreduzierten News

Die zum Beginn der Vorreduktion vorliegenden Ad-Hoc-News setzen sich aus den Elementen WPKN, Datum, Thema (Subject), Quelle und Inhalt zusammen. Die Abbildung 5-18 zeigt Beispiele, die der Wrapper während seines Parsvorganges herausgefiltert hat.

<i>Datum</i>	<i>WPKN</i>	<i>Subject</i>	<i>Inhalt</i>
20.11.2000	691030	Plambeck neue Übernahme	Die Plambeck Neue Energien ...
21.11.2000	691030	Plambeck Übernahme	Nach Angaben von Neuer Markt ...
22.11.2000	691090	PC-Ware steigert Halbjahresumsatz	PC-Ware Information Technologies ...

Abbildung 5-18: Beispiel-Ad-Hoc-News

Diese Daten werden in diesem Abschnitt beispielhaft vorreduziert.

In den vier oben genannten Schritten sind Ähnlichkeiten zu der Vorgehensweise bei den Analysen festzustellen. Anfangs wird ebenfalls eine Bewertung vorgenommen, die sich bei den News aber nur auf den zeitlichen Aspekt beschränkt. Im Unterschied zu den Analysen kann eine Einteilung der Ad-Hoc-News nach inhaltlichen Aspekten nicht vorgenommen werden. Es fehlen eindeutige Schlüsselwörter im Subject, die eine Kategorisierung erlauben würden.

Nur Ad-Hoc-News, die in den letzten vier Werktagen veröffentlicht wurden, sind ausschlaggebend. Ältere werden bei der Vorreduktion nicht beachtet, da sich der Informationsgewinn für den Nutzer als zu gering erweist. Den Ad-Hoc-News werden entsprechend ihres Alters, wie in Abbildung 5-19 dargestellt, numerische Werte zugewiesen.

<i>Alter der Ad-Hoc-News in Werktagen</i>	<i>Bewertung</i>
0	1
1	0,8
2	0,6
3	0,4
> 3	0

Abbildung 5-19: Bewertung der Ad-Hoc-News

Die Abbildung 5-20 zeigt die Bewertung an den Beispielnews aus der Abbildung 5-18.

<i>Datum</i>	<i>WPKN</i>	<i>Subject</i>	<i>Inhalt</i>	<i>Bewertung<sup>7</sup></i>
20.11.2000	691030	...	...	<b>0,6</b>
21.11.2000	691030	...	...	<b>0,8</b>
22.11.2000	691090	...	...	<b>1</b>

Abbildung 5-20: Bewertung der Beispiel-Ad-Hoc-News

Analog zu den Analysen ist der Nutzer an relevanten Informationen zu einer Aktie interessiert, und nicht nach einzelnen News. Deshalb werden im zweiten Schritt die News nach ihrer WPKN gruppiert und die Bewertungen je Gruppierung addiert. Dieses ist für die Beispieldaten in der Abbildung 5-21 demonstriert.

<i>WPKN</i>	<i>Summe der Bewertungen</i>
691030	<b>1,4</b>
691090	<b>1</b>

Abbildung 5-21: Gruppierung der Beispiel-Ad-Hoc-News

In der Regel schwanken die gewonnenen Summen pro WPKN stark. Nahe der Null liegende Werte, sowie Zahlen größer gleich zehn sind keine Seltenheit. Analog zu den Analysen werden im dritten Schritt die gewonnenen Bewertungsmaße der News unter der Verwendung der größten Summe normiert. Das Beispielergebnis bei einer maximalen Bewertung von 5,7 zeigt die Tabelle in der Abbildung 5-22.

<i>WPKN</i>	<i>Signifikanz</i>
691030	<b>0.246</b>
691090	<b>0.175</b>

Abbildung 5-22: Signifikanzen der Beispiel-Ad-Hoc-News

Der letzte Schritt umfasst das Aufbauen der vorreduzierten News. Sie besteht aus folgenden Daten:

- ID
- Signifikanz
- WPKN
- Subject der letzten News für die WPKN
- Inhalt der letzten News für die WPKN
- Datum
- Aktueller Kurs
- Datum des aktuellen Kurses

Die ID einer vorreduzierten News ist im Aufbau fast identisch zu der einer Analyse. Ein Unterschied besteht lediglich in dem ersten Zeichen, einem „N“, welches eine News kennzeichnet. Die Abbildung 5-23 zeigt den Aufbau der ID.

<i>Kategorie</i>	<i>Aufbau der ID</i>																						
News	N	Kürzel	Datum(JJMMTT)	Zeit(HHMM)	Datum(JJMMTT)	Nr.																	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22

Abbildung 5-23: Aufbau der ID einer vorreduzierten News

<sup>7</sup> Für die Bewertung der Beispiel-Adhoc-News ist der 22.11.2000 als aktueller Tag zu Grunde gelegt.

Die restlichen Elemente der vorreduzierten News sind vom Wrapper geparste Daten, die in der Datenbank vorliegen. Das sind zum einen Kursinformationen, wie beispielsweise der aktuelle Kurs, und zum anderen Daten der zuletzt veröffentlichten News. Die vorreduzierten Beispielnews ist in der Abbildung 5-24 dargestellt.

<i>ID</i>	NPCW_001122180000112201
<i>Sig</i>	0.175
<i>WPKN</i>	691090
<i>Subject</i>	PC-Ware steigert Halbjahresumsatz
<i>Inhalt</i>	PC-Ware Information Technologies AG ...
<i>Datum</i>	22.11.2000
<i>Kurs</i>	17,2 €
<i>Datum des aktuellen Kurs</i>	22.11.2000 17:59

Abbildung 5-24: vorreduzierte Beispielnews

## Die TopsFlops

Börseninformationssysteme bieten Informationen über die Kursentwicklung von Aktien an. Hierzu gehören in der Regel Listen über Tops und Flops. Diese zeigen, welche Aktien sich positiv beziehungsweise negativ hinsichtlich unterschiedlicher Betrachtungszeiträume am Markt entwickelt haben.

Für die Generierung von Tops- oder Flopsinformationen (TopsFlops) im System werden Daten aus der Kursliste verwendet, wie zum Beispiel der aktuelle Kurs und Schlusskurse, entsprechend der unterschiedlichen Betrachtungszeiträume.

Die Bildung vorreduzierter Tops beziehungsweise Flops teilt sich in fünf Schritten:

1. Sammeln der relevanten Daten
2. Berechnung der prozentualen Kursveränderungen mit unterschiedlichen Ansätzen
3. Mittelwertbestimmung
4. Normierung
5. Erstellung der vorreduzierten Top oder Flop

Im ersten Schritt werden alle Informationen, die für die Berechnung relevant sind zusammengetragen. Da die Berechnung für unterschiedliche Zeiträume durchgeführt wird, ist es notwendig alle entsprechenden Schlusskurse, Höchst- und Tiefstkurse zu ermitteln. Grundlage für die Berechnung sind:

- aktueller Kurs
- Schlusskurs<sup>8</sup>
- Höchstkurs<sup>9</sup>
- Tiefstkurs<sup>9</sup>

Die Berechnung wird am Beispiel einer Aktie durchgeführt, jedoch nur für den Betrachtungszeitraum von einem Tag. Die Abbildung 5-25 zeigt die Daten einer Beispielaktie.

---

<sup>8</sup> letzter Kurs vor einem Tag, einer Woche, einem Monat, drei Monaten, sechs Monaten und zwölf Monaten

<sup>9</sup> innerhalb des Zeitraumes von einem Tag, einer Woche, einem Monat, drei Monaten, sechs Monaten und zwölf Monaten

<i>Datum &amp; Zeit</i>	<i>WPKN</i>	<i>Tageshoch</i> <i>in €</i>	<i>Tagestief</i> <i>in €</i>	<i>aktueller Kurs</i> <i>in €</i>
28.02.2001 19:30	500770	22	18	21,5
...	...	...	...	...
01.03.2001 17:37	500770	26	18,2	23

Abbildung 5-25: Daten einer Aktie für den Betrachtungszeitraum 1 Tag

Nachdem die relevanten Daten gesammelt wurden, können im zweiten Schritt die prozentualen Veränderungen ermittelt werden. Diese ergeben sich aus der Gegenüberstellung des aktuellen Kurses zu dem Schlusskurs, dem höchsten und dem niedrigsten Kurs. Von den drei resultierenden Werten wird der Durchschnitt berechnet. Abbildung 5-26 stellt die Werte für die Beispielaktie dar.

<i>gegenüber dem</i>	<i>prozentuale Kursänderung</i>
Schlusskurs	7,0 %
Höchstkurs	-11,5 %
Tiefstkurs	27,8 %
∅	7,77 %

Abbildung 5-26: prozentuale Kursänderungen

Aktien mit positivem Durchschnitt entsprechen einem Top und analog negative einem Flop.

Zum Vergleich mit den Signifikanzen der Informationen anderer Kategorien werden die gemittelten Werte normiert. Die Normierung wird für jeden Zeitraum separat durchgeführt und bildet die Werte auf den Zahlenbereich -1 bis 1 ab. Die daraus resultierenden Werte ergeben die Signifikanzen der Top- beziehungsweise Flopinformationen. Für die Beispieldaten soll die Signifikanz 0,86 sein.

Im nächsten Schritt werden die vorreduzierten TopsFlops erstellt. Sie besteht aus den Elementen:

- ID
- WPKN
- Signifikanz
- aktueller Kurs
- Datum und Zeit des aktuellen Kurses
- höchster Kurs
- Datum des höchsten Kurses
- niedrigster Kurs
- Datum des niedrigsten Kurses
- Schlusskurs
- Datum des Schlusskurses
- prozentuale Veränderung aktueller Kurs – Schlusskurs
- prozentuale Veränderung aktueller Kurs – höchster Kurs
- prozentuale Veränderung aktueller Kurs – niedrigster Kurs

Mit der ID wird der Top beziehungsweise Flop eindeutig identifiziert. Deren Aufbau ist in Abbildung 5-27 gezeigt. Das erste Zeichen bestimmt den Betrachtungszeitraum und die Art der vorreduzierten Information.

Kategorie	Aufbau der ID																														
TopsFlops	<table border="1" style="margin: auto;"> <tr> <td style="width: 15px; height: 15px; border: 1px solid black; border-radius: 50%;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> <td style="width: 15px; height: 15px; border: 1px solid black;"></td> </tr> <tr> <td style="text-align: center;">0</td> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> <td style="text-align: center;">6</td> <td style="text-align: center;">7</td> <td style="text-align: center;">8</td> <td style="text-align: center;">9</td> <td style="text-align: center;">10</td> <td style="text-align: center;">11</td> <td style="text-align: center;">12</td> <td style="text-align: center;">13</td> <td style="text-align: center;">14</td> </tr> </table>																0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14																	
	<table border="1" style="margin: auto;"> <tr> <td style="width: 100px; text-align: center;"><i>Kürzel</i></td> <td style="width: 100px; text-align: center;"><i>Datum(JJMMTT)</i></td> <td style="width: 100px; text-align: center;"><i>Zeit (HHMM)</i></td> </tr> </table>	<i>Kürzel</i>	<i>Datum(JJMMTT)</i>	<i>Zeit (HHMM)</i>																											
<i>Kürzel</i>	<i>Datum(JJMMTT)</i>	<i>Zeit (HHMM)</i>																													
	<p>Tops: <math>\circ \in \{B, C, D, E, F, G\}</math>  Flops: <math>\circ \in \{O, P, Q, R, S, T\}</math></p> <p>entsprechend dem Betrachtungszeitraum (1 Tag, 1 Woche, 1 Monat,  3 Monate, 6 Monate, 1 Jahr)</p>																														

Abbildung 5-27: Aufbau der ID bei TopsFlops

Die nächsten vier Zeichen entsprechen dem Kürzel des zugehörigen Wertpapiers. Zeichen 5 bis 14 bezieht sich auf das Datum und die Uhrzeit des aktuellen Kurses. In Abbildung 5-28 wird der vorreduzierte TopFlop der Beispielaktie angegeben.

<i>ID</i>	BACG_0103011737
<i>WPKN</i>	500770
<i>aktueller Kurs</i>	23 €
<i>Zeit des aktuellen Kurses</i>	01.03.2001 17:37
<i>Schlusskurs</i>	21,5 €
<i>Datum des Schlusskurses</i>	28.02.2001
<i>Höchstkurs</i>	26 €
<i>Datum des Höchstkurses</i>	01.03.2001
<i>Tiefstkurs</i>	18 €
<i>Datum des Tiefstkurses</i>	28.02.2001
<i>proz.Ver. akt. Kurs - Schlusskurs</i>	7,0 %
<i>proz.Ver. akt. Kurs - Höchstkurs</i>	-11,5 %
<i>proz.Ver. akt. Kurs - Tiefstkurs</i>	27,8 %
<i>Sig</i>	0,86

Abbildung 5-28: vorreduzierter TopFlop

## Die Schwankungen

Bei der Anlageentscheidung achten Aktionäre sehr oft auf die Schwankungsbreite der Aktien, mit deren Hilfe diese in riskante beziehungsweise konservative Wertpapiere eingestuft werden können. Für die Vorreduktion der Schwankungen werden die Schwankungsbreiten der Aktien berechnet. Wie auch bei den TopsFlops stammen die dafür notwendigen Daten aus der Kursliste und beziehen sich auf unterschiedliche Betrachtungszeiträume (1 Tag, 1 Woche, 1 Monat, 3 Monate, 6 Monate und 1 Jahr). Das Berechnen einer vorreduzierten Schwankung gliedert sich in folgende Schritte und wird separat für jeden Betrachtungszeitraum durchgeführt:

1. Sammeln von Daten für die Berechnung der Schwankungsbreite
2. Berechnung der Schwankungsbreite
3. Normierungen
4. Erstellen der vorreduzierten Schwankung

Im ersten Schritt werden alle Daten gesammelt, die für die Berechnung der Schwankungsbreite notwendig sind. Diese sind:

- niedrigster Kurs<sup>10</sup>

- höchster Kurs<sup>10</sup>

Die Abbildung 5-29 zeigt die Daten für eine Beispielaktie, wobei der Zeitraum von einem Tag gewählt wurde.

<i>niedrigster Kurs</i>	26 €
<i>höchster Kurs</i>	18 €

Abbildung 5-29: Kurse der Beispielaktie

Im zweiten Schritt wird die Schwankungsbreite der Aktien nach folgender Form bestimmt:

$$\text{Schwankungsbreite} = \frac{\text{höchster Kurs} - \text{niedrigster Kurs}}{\text{niedrigster Kurs}}$$

Zum Vergleich der Schwankungen mit anderen vorreduzierten Daten werden die Schwankungsbreiten normiert. Hierzu werden diese durch die größte dividiert. Das normierte Ergebnis sind Signifikanzen im Bereich zwischen 0 und 1.

Abbildung 5-30 zeigt für die Beispielaktie die berechnete Schwankungsbreite und die daraus resultierende Signifikanz bei einer angenommenen maximalen Schwankungsbreite von 61,1%.

<i>Schwankungsbreite</i>	44,4 %
<i>Signifikanz</i>	0,73

Abbildung 5-30: Schwankungsbreite und Signifikanz der Beispielaktie

Im letzten Schritt werden die vorreduzierten Schwankungen erzeugt. Diese setzen sich aus folgenden Elementen zusammen:

- ID
- WPKN
- Signifikanz
- aktueller Kurs
- Datum des aktuellen Kurses
- höchster Kurs
- Datum des höchsten Kurses
- niedrigster Kurs
- Datum des niedrigsten Kurses
- Schwankungsbreite zwischen dem höchsten und niedrigsten Kurs

Die ID dient der eindeutigen Identifizierung der vorreduzierten Schwankung. Deren Aufbau ist in der Abbildung 5-31 dargestellt.

---

<sup>10</sup> innerhalb des Zeitraumes von einem Tag, einer Woche, einem Monat, drei Monaten, sechs Monaten und zwölf Monaten

Kategorie	Aufbau der ID																															
Schwankungen	<table border="1" style="margin: auto;"> <tr> <td style="width: 20px; text-align: center;">○</td> <td style="width: 20px;"></td> <td style="width: 20px; text-align: center;"><i>Kürzel</i></td> <td style="width: 20px;"></td> <td style="width: 20px;"></td> <td style="width: 20px;"></td> <td style="width: 20px; text-align: center;"><i>Datum(JJMMTT)</i></td> <td style="width: 20px;"></td> <td style="width: 20px; text-align: center;"><i>Zeit (HHMM)</i></td> </tr> <tr> <td style="text-align: center;">0</td> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> <td style="text-align: center;">6</td> <td style="text-align: center;">7</td> <td style="text-align: center;">8</td> <td style="text-align: center;">9</td> <td style="text-align: center;">10</td> <td style="text-align: center;">11</td> <td style="text-align: center;">12</td> <td style="text-align: center;">13</td> <td style="text-align: center;">14</td> </tr> </table> <p style="text-align: center;">○ ∈ {U, V, W, X, Y, Z}</p> <p style="text-align: center;">entsprechend dem Betrachtungszeitraum (1Tag, 1 Woche, 1 Monat, 3 Monate, 6 Monate, 1 Jahr)</p>	○		<i>Kürzel</i>				<i>Datum(JJMMTT)</i>									<i>Zeit (HHMM)</i>	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
○		<i>Kürzel</i>				<i>Datum(JJMMTT)</i>									<i>Zeit (HHMM)</i>																	
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14																		

Abbildung 5-31: Aufbau der ID bei Schwankungen

Das erste Zeichen in der ID kennzeichnet den Zeitraum, über den die vorreduzierte Schwankung für eine Aktie generiert wurde. Das Kürzel, zur Kennzeichnung der Aktie, für die die vorreduzierte Schwankung gebildet wurde, belegt die Zeichen 1 bis 4. Die letzten Zeichen beziehen sich auf den aktuellen Kurs der Aktie und stellen dessen Datum und Zeit dar.

In der Abbildung 5-32 ist die generierte vorreduzierte Schwankung für die Beispielaktie abgebildet.

<i>ID</i>	BACG_0103011737
<i>WPKN</i>	500770
<i>aktueller Kurs</i>	23 €
<i>Zeit des aktuellen Kurses</i>	01.03.2001 17:37
<i>höchster Kurs</i>	26 €
<i>Datum des höchsten Kurses</i>	01.03.2001
<i>tiefster Kurs</i>	18 €
<i>Datum des tiefsten Kurses</i>	28.02.2001
<i>Schwankungsbreite</i>	44,4 %
<i>Sig</i>	0,73

Abbildung 5-32: vorreduzierte Schwankung

## Die Neuemissionen

Wird ein Unternehmen an der Börse neu notiert, so wird die damit verbundene Ausgabe der Aktien des Unternehmens Neuemission genannt. Dabei spielen zwei Kursangaben eine besondere Rolle: der Emissionskurs und der erste Kurs. Der Aktionär möchte sich in der Regel darüber informieren, wie sich die Aktien der neuen Unternehmen entwickeln, das heißt, er betrachtet die prozentuale Veränderung des aktuellen Kurses jeweils zum Emissionskurs und zum ersten Kurs. Die für die Berechnung notwendigen Daten sind Bestand der Fundamentaldaten der Aktien und der aktuellen Kursliste, die vom Wrapper gesammelt werden. Abbildung 5-33 zeigt die Fundamentaldaten einer Aktie, die in diesem Abschnitt als Beispiel dienen soll, um daran die Berechnungen, die im Bezug der Kategorie Neuemissionen anfallen, zu zeigen.

<i>Name</i>	Dr. Hoenle AG
<i>WPKN</i>	515710
<i>Kürzel</i>	HNL
<i>Branche</i>	Technology
<i>Begin der Börsennotierung</i>	24.01.2001
<i>Emissionskurs</i>	12,0 €
<i>erster Börsenpreis</i>	16,5 €

Abbildung 5-33: Fundamentaldaten der Beispielaktie

Für die Erstellung vorreduzierter Neuemissionen werden nur die Aktien betrachtet, deren Ausgabezeitpunkt nicht länger als drei Monate zurückliegt. Nach der Auswahl der entsprechenden Aktien werden mit den folgenden Schritten die vorreduzierten Neuemissionen gebildet:

1. Sammeln der relevanten Daten
2. Berechnung der prozentualen Kursveränderungen mit unterschiedlichen Ansätzen
3. Mittelwertbestimmung
4. Normierung
5. Erstellung der vorreduzierten Neuemissionen

Im ersten Schritt werden die relevanten Daten zusammengestellt, die beim Berechnen der prozentualen Veränderungen notwendig sind. Diese sind der Emissionskurs, der erste Kurs und der aktuelle Kurs der Aktie. In der Abbildung 5-34 sind diese Daten für die Beispielaktie dargestellt.

<i>aktueller Kurs</i>	13,98 €
<i>Emissionskurs</i>	12,0 €
<i>erster Börsenpreis</i>	16,5 €

Abbildung 5-34: Kurse der Beispielaktie

Mit den vorliegenden Daten können die prozentualen Veränderungen des aktuellen Kurses bezüglich des Emissionskurses und des ersten Kurses berechnet werden. Anschließend wird der Mittelwert der beiden Prozentangaben ermittelt. Abbildung 5-35 zeigt die berechneten Werte für die Beispielaktie.

<i>gegenüber dem</i>	<i>prozentuale Kursänderung</i>
Emissionskurs	16,5 %
ersten Börsenpreis	-15,3 %
Ø	0,6 %

Abbildung 5-35: prozentuale Kursänderungen

Die Normierung des Mittelwertes ist notwendig, um sie mit den Signifikanzen der vorreduzierten Börseninformationen anderer Kategorien (Analysen, Ad-Hoc-News, TopsFlops und Schwankungen) vergleichbar zu machen. Dabei werden die Mittelwerte aller Neuemissionen, die positiv als auch negativ sind, auf die reellen Zahlen zwischen -1 und 1 abgebildet. Die normierten Werte stellen die Signifikanzen dar. Bei einem angenommenen maximalsten positiven Mittelwert von 21,3 % beträgt die Signifikanz der Beispielaktie 0,028.

Im letzten Schritt wird die vorreduzierte Neuemission gebildet, die aus folgenden Teilen besteht:

- ID
- WPKN
- Signifikanz
- aktueller Kurs
- Emissionskurs
- erster Kurs
- prozentuale Veränderung aktueller Kurs – Emissionskurs (Ak\_Emiss)
- prozentuale Veränderung aktueller Kurs – erster Kurs (Ak\_Erst)

Die ID besitzt im ersten Zeichen ein „I“ zur Identifikation der Kategorie Neuemissionen. Es folgt das Kürzel der Aktie und zum Schluss die Kurszeit, bestehend aus Datum und Uhrzeit des aktuellen Kurses. Die Graphik in der Abbildung 5-36 zeigt den allgemeinen Aufbau der ID einer Neuemission.

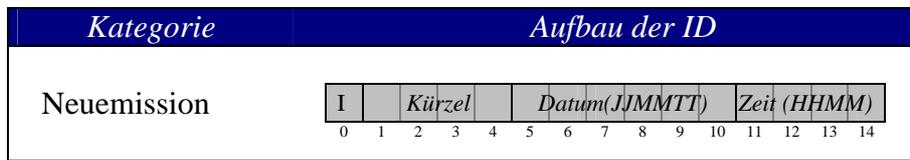


Abbildung 5-36: Aufbau der ID bei Neuemissionen

In der Abbildung 5-37 ist letztendlich die vorreduzierte Neuemission der Beispielaktie dargestellt.

<i>ID</i>	IHNL_0103041723
<i>WPKN</i>	515710
<i>aktuellerKurs</i>	13,98 €
<i>Emmissionskurs</i>	12,0 €
<i>erster Börsenpreis</i>	16,5 €
<i>proz.Ver. akt. Kurs - Emmissionskurs</i>	16,5 %
<i>proz.Ver. akt. Kurs - erster Börsenpreis</i>	-15,3 %
<i>Sig</i>	0,028

Abbildung 5-37: vorreduzierte Neuemission

### 5.3. DB auf dem Server + DB-Schema

Zur Speicherung der Börseninformationen sowie der Ergebnisse aus der profilunabhängigen Reduktion kommt das Datenbankmanagementsystem Borland Interbase Version 6.0 zum Einsatz. Die Kommunikation zwischen Datenbank und dem Applikationsserver wird über eine JDBC-Schnittstelle realisiert. Hierzu stellt der Borland Interbase InterClient 1.6 einen JDBC-Treiber zur Verfügung. Mit diesem Treiber und JDBC 2.0 von Sun Microsystems, das im JDK 1.2.2 enthalten ist, wird eine Kommunikationsverbindung zur Datenbank hergestellt. Die genaue Vorgehensweise kann dem Quellcode entnommen werden.

Der folgende Abschnitt erläutert das Schema der Datenbank, in dem die geparteten Börseninformationen des Wrappers als auch die vorreduzierten Daten verwaltet werden. Zur Speicherung der aus dem World Wide Web gewonnenen Börseninformationen werden 4 Tabellen in der Datenbank benötigt. Die erste Tabelle dient zur Speicherung der Fundamentaldaten. Hier wird unter anderem die Wertpapierkennnummer (WPKN), der Aktienname, die Branche sowie Informationen zur Platzierung des Papiers am Neuen Markt erfasst. In den anderen drei Tabellen werden die aktuellen Kurse, die Analysen und die Ad-Hoc-News mit ihren entsprechenden Attributen hinterlegt. Abbildung 5-38 zeigt die dazugehörigen Relationsschemata. Die darin fett dargestellten Attribute sind der Primärschlüssel der Tabellen.

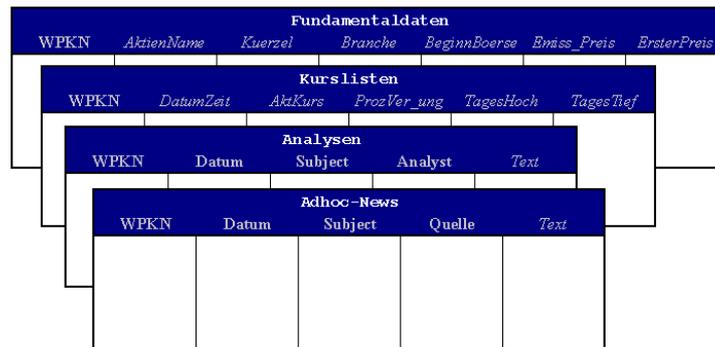


Abbildung 5-38: DB-Schema

Eine für die Anwendung notwendige Anfrage ist es, den aktuellen Kurs der Wertpapiere zu ermitteln. Eine mögliche Anfrage über die Kurslisten-Tabelle ist:

```
SELECT A.WPKN, A.AktKurs
FROM Kursliste A, ( SELECT WPKN AS WPKN, MAX(DatumZeit) AS
MaxDatumZeit
                    FROM Kursliste
                    GROUP BY WPKN) B
WHERE  A.WPKN = B.WPKN
      AND A.DatumZeit = B.MaxDatumZeit
```

Borland Interbase erlaubt jedoch keine *GroupBy*-Klausel innerhalb einer geschachtelten Anfrage. Zur Lösung des Problems wurde eine *View* definiert, mit deren Hilfe der aktuelle Kurs ermittelt werden kann.

```
CREATE VIEW MaxDatumZeitKurs AS
  SELECT WPKN AS WPKN, MAX(DatumZeit) AS MaxDatumZeit
  FROM Kursliste
  GROUP BY WPKN
```

Die Anfrage lautet unter Verwendung des *View* nun:

```
SELECT A.WPKN, A.AktKurs
FROM Kursliste A, MaxDatumZeitKurs B
WHERE  A.WPKN = B.WPKN
      AND A.DatumZeit = B.MaxDatumZeit
```

Die Börseninformationen, die der Endanwender des Systems auf seinem mobilen Endgerät erhält, gliedern sich in die bereits erläuterten fünf Kategorien:

- Analysen
- Ad-Hoc-News
- TopsFlops
- Schwankungen
- Neuemissionen

Die Daten für die genannten Kategorien werden während der Vorreduktion aus den geparsten Informationen, die bereits in den Tabellen *Fundamentaldaten*, *Kurslisten*, *Analysen* und *Ad-Hoc-News* abgelegt sind, erstellt (siehe Kapitel 5.2 Vorreduktion ab Seite 48). Die erzeugten Werte setzen sich aus den spezifischen Informationsdaten, einer ID zur eindeutigen Identifizierung und einer Signifikanz, die das Maß der Wichtigkeit dieser Information darstellt, zusammen. Um eine kurze Antwortzeit des Servers zu garantieren werden die Daten entsprechend ihrer Kategorie in der Datenbank

abgelegt und für die profilabhängige Reduktion verwendet. In Abbildung 5-39 sind die dazugehörigen Relationsschemata dargestellt.

AnalysenVorreduktion							
ID	Sig	WPKN	Subject	Inhalt	Analyst	Datum	aktuellerKurs

NewsVorreduktion							
ID	Sig	WPKN	Subject	Inhalt	Datum	aktuellerKurs	

NeuemissionenVorreduktion							
ID	Sig	WPKN	Emissionskurs	ersterKurs	Ak_Emiss	Ak_Erst	

Abbildung 5-39: Vorreduktionstabellen für Analysen, News und Neuemissionen

Für die Betrachtung von Tops, Flops und Schwankungen werden unterschiedliche Betrachtungszeiträume zugrundegelegt, zwischen denen der Endanwender mit Hilfe seiner Profileinstellungen wählen kann. Aufgrund dieser Tatsache ist es notwendig, für die Kategorien separate Tabellen, wie in Abbildung 5-40 für TopsFlops beispielhaft dargestellt, zu verwenden.

TopsFlopsTag				
ID	SIG	WPKN	aktuellerKurs	...

TopsFlopsWoche				
ID	SIG	WPKN	aktuellerKurs	...

TopsFlopsMonat				
ID	SIG	WPKN	aktuellerKurs	...

TopsFlops3Monate				
ID	SIG	WPKN	aktuellerKurs	...

TopsFlops6Monate				
ID	SIG	WPKN	aktuellerKurs	...

TopsFlopsJahr				
ID	SIG	WPKN	aktuellerKurs	...

Abbildung 5-40: Vorreduktionstabellen für unterschiedliche Betrachtungszeiträume

Der Empfang der Börseninformationen auf dem mobilen Endgerät gliedert sich in zwei Schritte. Im ersten werden zunächst nur die Header der Börseninformationen, resultierend aus der profilabhängigen Reduktion, empfangen. Bei näherem Interesse kann der Nutzer im zweiten Schritt die detaillierten Daten zu einer Information anfordern. In den bisher oben aufgeführten Tabellen werden nur die aktuellen Daten, das Ergebnis des letzten serverseitigen Vorreduktionslaufes, abgespeichert. Da der Zeitpunkt für den zweiten Schritt der Informationsanforderung des Nutzers am PDA unabhängig von den Vorreduktionsläufen ist, ist es notwendig die nach einem neuen Vorreduktionslauf veraltenden Daten nicht zu verwerfen. Hierzu existiert für jede Tabelle eine weitere, in der die Daten gespeichert werden.

NeuemissionenVorreduktion							
ID	Sig	WPKN	Emissionskurs	ersterKurs	Ak_Emiss	Ak_Erst	

NeuemissionenVorreduktionAlt							
ID	Sig	WPKN	Emissionskurs	ersterKurs	Ak_Emiss	Ak_Erst	

Abbildung 5-41: Vorreduktionstabellen zur Historisierung

## 5.4. Anfragebearbeitung und profilabhängige Reduktion

Die Aufgaben der profilabhängigen Reduktion sind die Signifikanzen, der in der Vorreduktion (siehe Kapitel 5.2 Vorreduktion ab Seite 48) gebildeten Börseninformationen, zu wichten und anhand dieser die Informationen für die Übertragung an das mobile Endgerät zu selektieren. Die allgemeine Herangehensweise ist hierbei zunächst, die berechneten Signifikanzen mit Faktoren zu multiplizieren, um so Informationen auf- beziehungsweise abzuwerten. Mit Hilfe von gegebenen Schranken können in einem anderen Schritt die Informationen in interessante und uninteressante eingeteilt werden. Die interessanten ergeben die Anfrageergebnismenge.

Faktoren: $sig = sig \cdot \text{Faktor}$
Schranken: <b>if</b> (Ausdruck $\circ$ Schranke) <b>then</b> Information uninteressant <b>else</b> Information interessant mit $\circ \in \{>, \geq, <, \leq\}$

Abbildung 5-42: Wirkung von Faktoren und Schranken

Die Faktoren und Schranken werden vom Anwender auf dem mobilen Endgerät mit Hilfe des Profils definiert. Im folgenden Abschnitt werden die einzelnen Parameter, deren Werte sich durch die Profileinstellungen ergeben und bei einer Anfrage an den Server übermittelt werden, hinsichtlich ihres Einflusses auf die profilabhängige Reduktion näher erläutert. Die Einstellungsmöglichkeiten des Profils und deren Bedeutung für den Nutzer des Börseninformationssystems wird in Kapitel 5.8 Profilverwaltung auf dem Palm ab Seite 79 gegeben.

Die Parameter unterteilen sich in Faktoren und Schranken und beziehen sich auf einzelne oder mehrere Kategorien. In der folgenden Übersicht werden in Klammern die Bezeichnungen der Parameter innerhalb der Profileinstellung der mobilen Endanwenderapplikation angegeben.

### Faktoren

**buy** (Kaufempfehlungen)

Der Parameter wirkt sich auf die Analysen mit positiven Signifikanzen, also Kaufempfehlungen aus. Die Abbildung 5-43 zeigt den Wertebereich des Parameters und die dazugehörigen Faktoren.

Wert	Faktor
0	0,0
1	0,6
2	0,8
3	1,0
4	1,2
5	1,4

Abbildung 5-43: Parameter buy und die dazugehörigen Faktoren

**sale** (Verkaufsempfehlungen)

Im Gegensatz zu buy wirkt dieser Parameter nur auf Analysen mit negativen Signifikanzen. Die Abbildung der Werte des Parameters ist analog zum buy.

**weights** (Priorität)

Bei dem Parameter handelt es sich um ein 4-Tupel, dessen Werte sich jeweils auf alle Signifikanzen der 4 Informationsgruppen (1. Analysen; 2. News; 3. Tops, Flops, Schwankungen; 4. Neuemissionen) auswirken. Die Zuordnung der Faktoren auf die Werte zeigt die Tabelle in Abbildung 5-44.

<i>Wert</i>	<i>Faktor für die Signifikanzen der jeweiligen Kategorie</i>
0	0,0
1	0,6
2	0,8
3	1,0
4	1,2
5	1,4

Abbildung 5-44: Faktoreuzuordnung des Parameter `weights`

**stocks** (Aktien)

Der Parameter besteht aus 10 Aktiennamen. Die Signifikanzen der vorreduzierten Börseninformationen aller Kategorien, die sich auf diese Aktien beziehen, werden entsprechend der in Abbildung 5-45 dargestellten Tabelle aufgewertet.

<i>Aktie Nr.</i>	<i>Faktor</i>
1 <sup>11</sup>	1,5
2	1,45
3	1,4
4	1,35
5	1,3
6	1,25
7	1,2
8	1,15
9	1,1
10	1,05

Abbildung 5-45: Faktoren beim Parameter `stocks`

**sectors** (Branchen)

Die Signifikanzen aller Kategorien werden entsprechend der Branchensortierung, die dieser Parameter angibt mehr oder weniger aufgewertet.

---

<sup>11</sup> Aktie steht bei der Profileinstellung auf dem Palm ganz oben

<i>Branchen Nr.</i>	<i>Faktor</i>
1 <sup>12</sup>	1,2
2	1,15
3	1,1
4	1,05
5	1,0
6	0,95
7	0,9
8	0,85
9	0,8
10	0,75

Abbildung 5-46: Faktoren beim Parameter `sectors`

## Schranken

### **sig\*** (gesamt)

Der Parameter hat erst am Ende der profilabhängigen Reduktion Einfluss auf diese und sein Wirkungsbereich erstreckt sich dabei über alle Kategorien. Der vom Benutzer vorgegebene Wert für `sig*` liegt zwischen 0 und 100. Für deren Anwendung wird er zuvor durch 100 dividiert. Informationen werden dann nur vom System zur Übertragung ausgewählt, wenn deren Signifikanz größer oder gleich `sig*` ist.

### **maxInfoCount** (maximale Anzahl gesamt)

Wie auch `sig*` findet der Parameter `maxInfoCount` erst zum Abschluss der profilabhängigen Reduktion Anwendung und gilt für alle Kategorien. Im Gegensatz hierzu gibt er jedoch die Anzahl der maximal zu selektierenden Börseninformationen an.

### **maxTops** (maximale Anzahl Tops)

`maxTops` bestimmt die maximale Anzahl der zu selektierenden Informationen mit positiven Signifikanzen der Kategorie `TopsFlops` an.

### **maxFlops** (maximale Anzahl Flops)

Analog zu `maxTops` gibt `maxFlops` jedoch die maximale Anzahl von Informationen mit negativen Signifikanzen, also `Flops` an.

### **maxTopsNewI** (maximale Anzahl der Besten Neuemissionen)

`maxTopsNewI` gibt die maximale Anzahl selektierbarer Informationen mit positiver Signifikanz in der Kategorie `Neuemissionen` an.

### **maxFlopsNewI** (maximale Anzahl der schlechtesten Neuemissionen)

Der Parameter hat die gleiche Bedeutung wie `maxTopsNewI`, bezieht sich allerdings dabei auf die negativen Signifikanzen.

---

<sup>12</sup> Branche steht bei der Profileinstellung auf dem Palm ganz oben

### **minFluctuation** (minimale Volatilität)

Der Wert des Parameters bestimmt die minimale Volatilität der Börseninformationen in der Kategorie Schwankungen, die zur Übertragung ausgewählt werden können.

### **sonstige**

#### **timePeriod** (Betrachtungszeitraum)

Der Parameter `timePeriod` dient zur Auswahl der Vorreduktionstabelle in den Kategorien `TopsFlops` und `Schwankungen`.

Nachdem die Parameter im einzelnen erläutert wurden, beschreibt der nächste Abschnitt, wie und wann die Parameter im einzelnen zum Einsatz kommen.

Bei einer Anfrage des Endanwenders vom mobilen Gerät aus werden die Signifikanzen der Informationen in den Vorreduktionstabellen mit den oben zuvor beschriebenen Faktoren gewichtet und daraus Tupel mit Hilfe der Schranken für die Übertragung ausgewählt. Während dessen werden gelegentlich Normierungen der Signifikanzen vorgenommen. Der Vorgang geschieht für jede Anfrage separat. Es werden hierbei auf die Vorreduktionstabellen lediglich Anfrage- und keine Änderungsoperationen durchgeführt. Das heißt, dass die profilabhängige Reduktion keinen Einfluss auf die Daten in den Vorreduktionstabellen hat.

Im ersten Schritt der Reduktion werden die kategoriebezogenen Parameter auf die jeweilige Vorreduktionstabelle der Kategorien angewendet. Im Anschluss daran findet eine Vereinigung der in den einzelnen Tabellen verbliebenden Informationen zu einer gesamten Datenmenge statt. Diese wird anschließend mit allgemeinen Parametern manipuliert und führt somit zur Ergebnismenge.

Der erste Schritt wird im anschließenden Abschnitt für die einzelnen Kategorien kurz dargestellt:

### **Analysen**

1. Alle positiven Signifikanzen werden in der Form  $sig = sig \cdot \text{Faktor}(\text{buy})$  und alle negativen in der Form  $sig = sig \cdot \text{Faktor}(\text{sale})$  beziehungsweise abgewertet.
2. Die Signifikanzen werden durch die betragsmäßig größte dividiert und somit normiert.
3. Negative Signifikanzen werden mit -1 multipliziert, um diese für die Vereinigung anzupassen.
4. Die Informationen werden mit den Parametern `stocks` und `sectors` gewichtet. Hierbei werden die Signifikanzen der Informationen mit dem entsprechenden Faktor (siehe Abbildung 5-45 und Abbildung 5-46) multipliziert. Das heißt zum Beispiel, dass die Signifikanzen aller Informationen, die zu der am höchsten priorisierten Branche gehören, mit dem Faktor 1,2 aufgewertet werden.

### **News**

1. Durch die Parameter `stocks` und `sectors` werden die Informationen in der Vorreduktionstabelle der Kategorie `News` wie bei den Analysen in Punkt 4 beziehungsweise abgewertet.

## TopsFlops

1. Mit dem Parameter `timePeriod` und mit dem damit im Zusammenhang stehenden Betrachtungszeitraum kann die Vorreduktionstabelle für die TopsFlops ausgewählt werden.
2. Die Signifikanzen werden mit den Parametern `stocks` und `sectors` gewichtet.
3. Mit Hilfe von `maxTops` und `maxFlops` werden die Anzahl der Tops beziehungsweise Flops beschränkt. Es werden die Informationen mit den jeweils betragsmäßig größten Signifikanzen genommen. Alle anderen bleiben für die Weiterverarbeitung außen vor.
4. Die negativen Signifikanzen der Flops werden mit -1 multipliziert.

## Schwankungen

1. Wie bei den TopsFlops dient der Parameter `timePeriod` zur Auswahl der Vorreduktionstabelle.
2. Alle Schwankungen mit einer Volatilität (Schwankungsbreite) kleiner als `minFluctuation` werden ausgeblendet.
3. Im letzten kategorieabhängigen Schritt werden die Parameter `Stocks` und `Sectors` angewendet.

## Neuemissionen

1. Wie bei allen anderen Kategorien findet auch bei den Neuemissionen eine Manipulation der Signifikanzen durch die Parameter `stocks` und `sectors` statt.
2. Die Anzahl der besten und der schlechtesten Neuemissionen wird durch die Parameter `maxTopsNewI` beziehungsweise `maxFlopNewI` beschränkt. Analog zu TopsFlops werden nur die Informationen mit den betragsmäßig größten Signifikanzen für die Weiterverarbeitung ausgewählt.
3. Durch die Multiplikation der Signifikanzen schlechter Neuemissionen mit -1 werden diese für die Vereinigung angepasst.

Nachdem im ersten Schritt der profilabhängigen Reduktion Parameter auf die Vorreduktionstabellen der einzelnen Kategorien angewendet wurden, können die Tabellen vereinigt werden. Vor der Vereinigung werden die Signifikanzen der Börseninformationen in den noch separaten Tabellen mit dem jeweilig zur Kategorie zugehörigen Wert des Parameters `weights` multipliziert. Die vereinigte Tabelle enthält nun Informationen der unterschiedlichen Kategorien mit positiven Signifikanzen, die jeweils größer oder gleich 0 sind. Eine anschließende Normierung bildet alle Signifikanzen auf einen Wertebereich zwischen 0 und 1 ab. Der letzte Teilschritt der Reduktion beinhaltet die Selektion der zu übertragenden Börseninformationen mit Hilfe von `sig*` und `maxInfoCount`. Hierbei werden zunächst alle Informationen selektiert, deren Signifikanz größer oder gleich `sig*` ist. Übersteigt die Anzahl der Elemente in der selektierten Menge `maxInfoCount`, so wird die Menge auf die vorgegebene maximale Anzahl reduziert. Die Informationen mit den geringsten Signifikanzen werden aus der Ergebnismenge entfernt. Die Börseninformationen in der resultierenden Ergebnismenge stellen die interessantesten Informationen für den Nutzer an dem mobilen Endgerät dar und können an diesen gesendet werden. Wie bereits in Kapitel 2.2 Technische Konzepte ab Seite 12 angesprochen, sendet der Server zunächst nur die Informationsheader, die unter

anderem einen eindeutigen Schlüssel zur Information beinhalten. Die detaillierten Daten einer Börseninformation können bei Bedarf nachgeladen werden.

Bei den Profileinstellungen auf dem Palm können einige Parameter aktiviert oder deaktiviert werden. Soll ein Parameter nicht verwendet werden, er ist also deaktiviert, so fällt die Anwendung von diesem auf die Börseninformation aus. Der Einsatz anderer Parameter finden jedoch wie in dem zuvor dargestellten Ablauf statt.

In der abschließenden Abbildung 5-47 wird der zeitliche Ablauf der profilabhängigen Reduktion und die damit verbundene Anwendung von Parametern auf Vorreduktionstabellen bis zur endgültigen Ergebnismenge noch mal verdeutlicht.

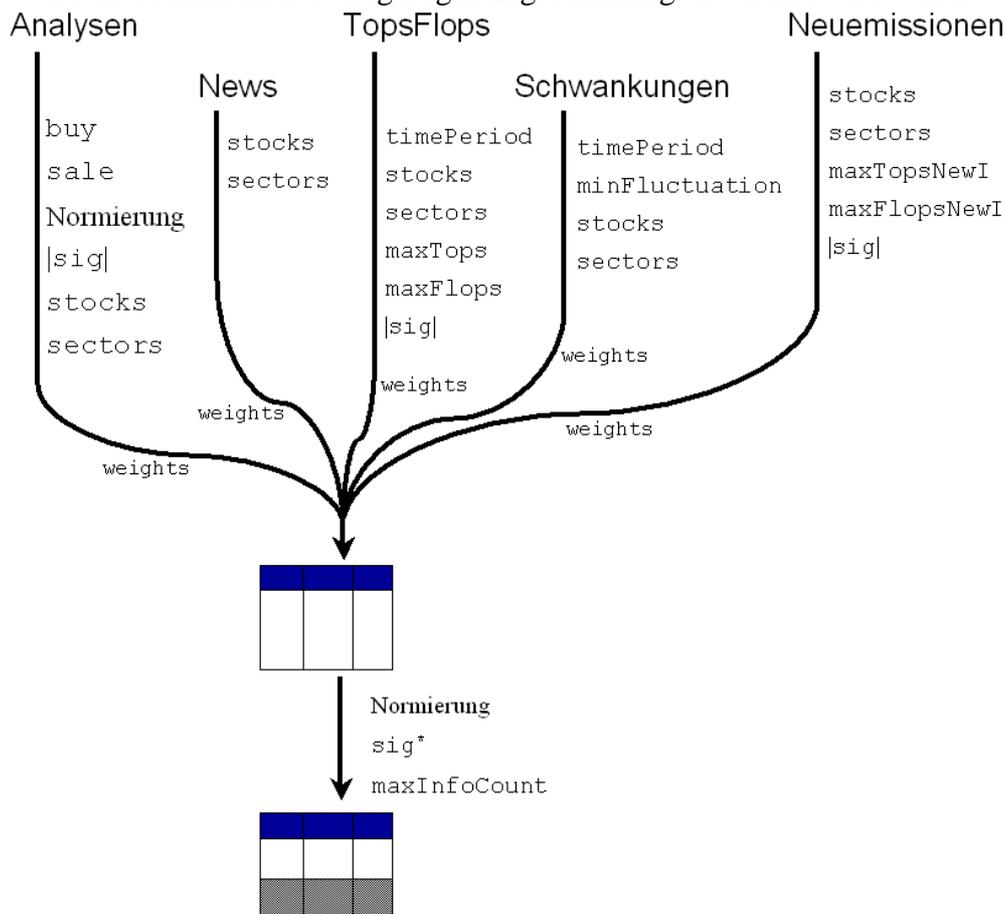


Abbildung 5-47: Ablauf der profilabhängigen Reduktion

In diesem Kapitel wurde die Wirkungsweise der Parameter des Profils auf die vorreduzierten Daten erläutert. Speziell bei Parametern, die zur Auf- beziehungsweise Abwertung von Informationen dienen (buy, sale, stocks, sectors und weights), ist die Abbildung dieser auf die Faktorwerte zunächst frei gewählt. Nicht nur bei der profilabhängigen Reduktion, sondern auch bei der Vorreduktion spielen Bewertungen und Faktoren eine Rolle, zum Beispiel die zeitliche Bewertung von Analysen und Ad-Hoc-News.

Andere Wertezuordnungen haben entsprechende Auswirkungen auf das Reduktionsergebnis. Eine statistische Untersuchung könnte Aufschluss über eine bestmögliche Zuordnung geben. Dieses soll jedoch nicht Umstand der Studienarbeit sein und wird daher hier nicht weiter behandelt.

## 5.5. Anfragebearbeitung auf dem Client – Aktualisierung der Börseninformationen

Die Aktualisierung der Börseninformationen auf dem mobilen Endgerät ist eine der Hauptfunktion der Endanwenderapplikation. Der Vorgang unterteilt sich in zwei wesentliche Schritte: Im ersten Schritt werden zunächst nur die Informationsheader aktualisiert. Ein Header enthält die Überschrift (Subject), die Informationsbewertung sowie die ID einer Börseninformation und beschreibt so die jeweiligen Börseninformationen. Ist der Endanwender an der einen oder anderen Information näher interessiert, so kann dieser im zweiten Schritt die dazugehörigen Börseninformationsdaten vom Server herunterladen. Hierzu muss der Nutzer die Information lediglich im Online-Betrieb auswählen.

Die Informationsheader sind clientseitig in der ID\_DB einem Palm-Katalog abgelegt (siehe Kapitel 5.9 Informationscache auf dem Palm ab Seite 82) und werden durch Betätigung des Buttons „Aktualisieren“ im Hauptfenster der Endanwenderapplikation aktualisiert. Die Abarbeitung des Aktualisierungsvorgang läuft in folgender Reihenfolge ab:

1. Löschen veralteter Informationsheader und falls vorhanden, deren zugehörigen Informationsdaten im Cache.
2. Aufbau einer Kommunikationsverbindung zum Server.
3. Senden der Anfrage an den Applikationsserver. Innerhalb der Anfrage befinden sich die Profileinstellungen. (siehe *SELECT*-Anfrage in Kapitel 5.6 Kommunikationsschnittstellen auf dem Server und Client ab Seite 71)
4. Empfangen der Antwort vom Server.
5. Vergleich der vom Server empfangenen Informationsheader mit denen in der ID\_DB der Endanwenderapplikation.
6. Speichern neuer Informationsheader in der ID\_DB. Bei Analysen oder News eventuell nur Markierung einfügen.
7. Schließen der Kommunikationsverbindung zum Server.

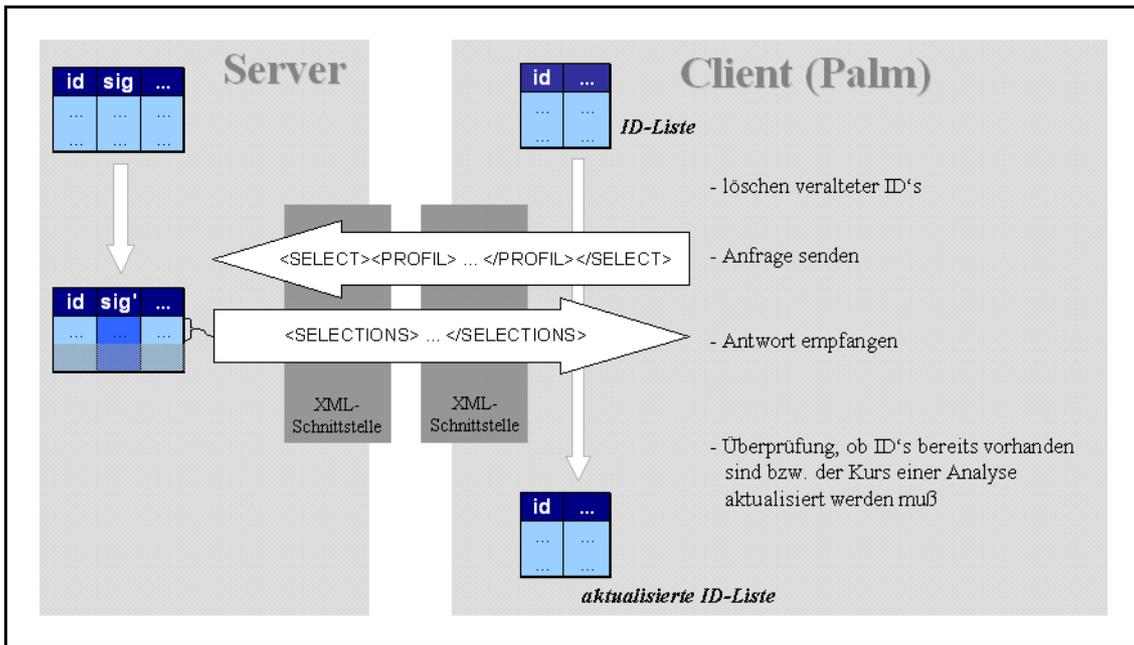


Abbildung 5-48: Aktualisierung der Informationsheader am Beispiel der Analysen

In dem folgenden Abschnitt wird nun der Vorgang zur Aktualisierung der Informationsheader am Beispiel der Analysen näher erläutert. Für das bessere Verständnis soll vorab der Aufbau der ID für Analysen noch mal dargestellt werden:

Kategorie	Aufbau der ID																						
Analyse	A	Kürzel			Datum(JJMMTT)					Zeit(HHMM)				Datum(JJMMTT)					Nr.				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22

d1: = Erstellungszeitpunkt der ID (Byte 5 bis 14)

d2: = Datum und Nummer der Analyse (Byte 15 bis 22)

Abbildung 5-49: Aufbau der ID einer Analyse

Im ersten Schritt der Aktualisierung werden zunächst alle alten Informationsheader in der ID\_DB gelöscht. Das Alter einer Information lässt sich aus deren ID und dem aktuellen Datum ermitteln.

$$\text{Alter der Information} = \text{aktuelles Datum} - d1$$

Abbildung 5-50: Berechnung des Alters einer Information

Überschreitet das Alter eine bestimmte Grenze, so wird der Header der Information sowie die eventuell im Cache der Endanwenderapplikation zugehörigen Informationsdaten gelöscht.

Ist die Bereinigung des Cache durch die Endanwenderapplikation abgeschlossen, so baut dieser eine Kommunikationsverbindung zum Server auf und sendet die *SELECT*-Anfrage an diesen. Die Anfrage enthält die Profileinstellungen des Nutzers und

veranlasst den Server zur Ausführung der profilabhängigen Reduktion. Die vom Client daraufhin empfangenen Informationsheader werden mit den in der ID\_DB gespeicherten Header verglichen.

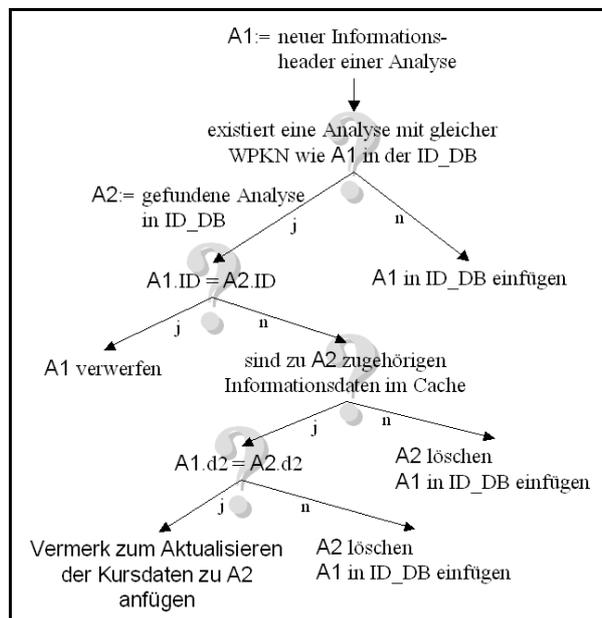


Abbildung 5-51: Vergleich der Informationsheader am Beispiel der Analysen

Entsprechend dem Ergebnis des Vergleiches werden die Header verworfen, eingefügt oder zum Markieren im Cache vorhandener Header verwendet. Markierungen existieren nur für Analysen und News. Sie werden vorgenommen unter der Voraussetzung, dass sich die zu einer Analyse oder News zugehörigen Informationsdaten im Cache der Endanwenderapplikation befinden und sich seit dem letzten durchgeführten Aktualisierungsvorgang lediglich der Kurs jedoch nicht der Analysen- bzw. Newstext geändert hat. Dieser Zustand tritt in der Regel öfter auf und würde ohne besondere Berücksichtigung beim Herunterladen der entsprechenden Analysen bzw. News eine unnötige Mehrfachübertragung des dazugehörigen Textes bedeuten. Sollen die Informationsdaten zu einem markierten Informationsheader heruntergeladen werden, so stellt die Clientsoftware nur eine Anfrage nach neuen Kursdaten des Wertpapiers und verhindert somit die mehrfache Informationsübertragung.

Die Aktualisierung der Informationsheadern bei Tops, Flops, Schwankungen, Neuemissionen und News verläuft analog zum oben erläuterten Aktualisierungsvorgang bei Analysen.

## 5.6. Kommunikationsschnittstellen auf dem Server und Client

Grundlage der Kommunikation zur Aktualisierung der Börseninformationen zwischen Server und Client ist ein Challenge-Response-Verfahren. Dieses oft auch als Pingpong-Kommunikation bezeichnete Verfahren beruht darauf, dass der Client eine Anfrage an

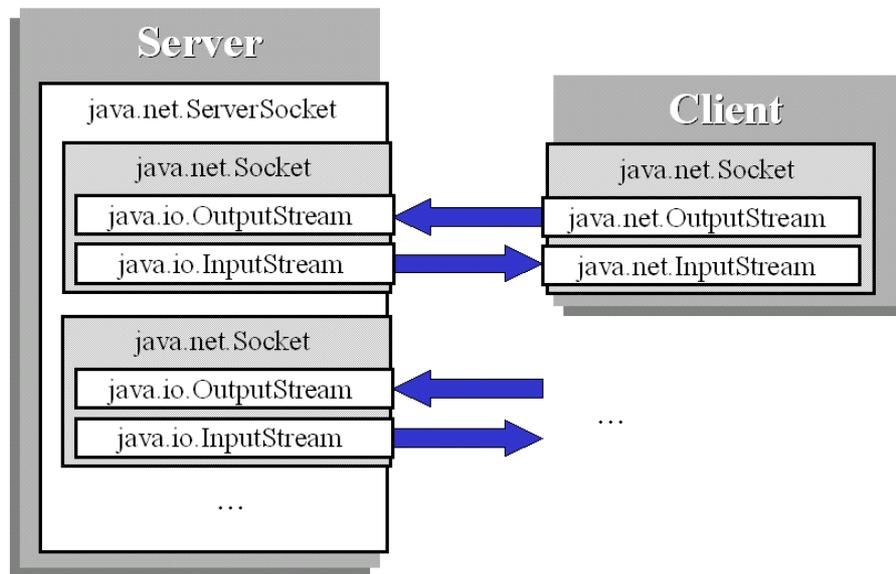


Abbildung 5-52: Kommunikationsverbindung zwischen Server und Client

den Server sendet (Challenge), der diese dann bearbeitet und eine Antwort (Respons) an den Client zurückschickt. Zunächst baut die Clientapplikation auf dem mobilen Endgerät hierzu eine Socketverbindung zum Server auf. Der Server lauscht auf einem bestimmten Port, in der Referenzimplementierung Port 1501, und öffnet jede Socketverbindung in einem separaten Thread, in dem die Anfragen des Clients bearbeitet werden. Die Kommunikation verläuft dabei stets in der angegebenen Reihenfolge:

1. Senden einer Anfrage des Clients an den Server.
2. Analyse und Bearbeitung der Anfrage auf dem Server. (siehe Kapitel 5.4 Anfragebearbeitung und profilabhängige Reduktion ab Seite 63)
3. Senden der erstellten Antwort an den Client.
4. Analyse und Verarbeitung der Antwort auf dem Client. (siehe Kapitel 5.5 Anfragebearbeitung auf dem Client – Aktualisierung der Börseninformationen ab Seite 70)
5. Senden einer weiteren Anfrage oder Schließen der Verbindung.

Die Anfragen des Clients und Antworten des Servers basieren auf XML. Deren Struktur sind in DTD's (Document Type Definition) beschrieben. Zur Analyse der XML-Dokumente kommt server- als auch clientseitig ein XML-Parser zum Einsatz. Beim Server handelt es sich hierbei um den *Oracle-XML-Parser* für Java. Dieser parst zunächst das gesamte Dokument und erstellt hieraus ein DOM-Objekt. Mit Hilfe der Strukturinformation der Anfragen, kann innerhalb des Objektes navigiert werden und so

die Anfrage sowie die darin enthaltenen Parameter einer Bearbeitungsroutine zugeordnet werden.

Die Verwendung eines DOM-Parsers ist am Client ressourcenbedingt nicht möglich. Zur Analyse der Antworten wird daher clientseitig der SAX-XML-Parser *TinyXML* verwendet. Im Gegensatz zum DOM-Parser wird kein Objekt, in dem navigiert werden kann, aufgebaut, sondern zu Beginn eines Tags, zum Ende eines Tags, bei einem Tagattribut usw. bestimmte, zu implementierende Methoden aufgerufen. Mit Kenntnissen über den Aufbau der Antworten ist der Client in der Lage, das vom Server erstellte Dokument zu analysieren und die darin enthaltenen Börseninformationen zu verarbeiten.

Die Struktur der Anfragen ist in der DTD *Client2Server.dtd* (siehe Anhang A DTD's zur Kommunikation zwischen Server und Client ab Seite 93) spezifiziert. Es existieren fünf verschiedenen Arten von Anfragen:

- GETALLSTOCKS* : Die Anfrage wird vom Client zur Aktualisierung der Aktiennamen gesendet. Der Server ist aufgefordert alle Aktiennamen zu senden.
- GETALLSECTORS* : Zur Aktualisierung der Branchennamen sendet der Client diese Anfrage an den Server.
- GETINFORMATION* : Mit Hilfe dieser Anfrage fordert der Client den Server auf, alle Daten einer Börseninformation zugehörig zu einer bestimmten ID zu senden. Die ID wird als Parameter innerhalb der Anfrage mitgesendet.
- GETCOURSE* : In manchen Fällen ist es nicht notwendig, eine gesamte Information zu aktualisieren, sondern nur die Kursdaten zu erneuern. Mit der Anfrage ist der Server aufgefordert, die aktuellen Kursdaten einer Information zu senden. Die ID ist in der Anfrage des Clients enthalten.
- SELECT* : Diese Anfrage ist die wichtigste. Sie beinhaltet die Profilinformationen des Endanwenders und veranlasst den Server die profilabhängige Reduktion mit den gesendeten Informationen durchzuführen. Von den, aus der Reduktion resultierenden, Börseninformationen werden die ID's, Subjects und Informationsbewertungen an den Client zurückgesendet.

In der DTD *Server2Client.dtd* (siehe Anhang A DTD's zur Kommunikation zwischen Server und Client ab Seite 93) ist die Struktur der Antworten, die der Server erstellt und an den Client sendet, beschrieben. Folgende Antworten existieren innerhalb dieser Definition:

- ALLSTOCKS* : Zum Senden aller Aktiennamen an den Client wird diese Antwort verwendet.
- ALLSECTORS* : Diese Antwort dient zum Übermitteln aller Branchennamen.
- SELECTIONS* : Die ID's, Subjects und Informationsbewertungen, die sich aus der profilabhängigen Reduktion ergeben, werden innerhalb dieser Antwort gesendet.
- INFORMATION* : Mit dieser Antwort werden die Daten zu einer bestimmten Börseninformation gesendet. Folgende Informationsarten sind zur Übertragung spezifiziert:
  - NEWS* - Ad-Hoc-News

- ANALYSIS* - Analysen
- TOPFLOP* - Tops oder Flops
- NEWISSUES* - Neuemissionen
- FLUCTUATION* - Schwankungen
- COURSE* - Innerhalb dieses Knotens befinden sich die aktuellen Kursdaten einer Analyse oder Ad-Hoc-News

## Probleme

Bei der Nutzung der implementierten XML-Schnittstelle am Client tritt gelegentlich ein unerwarteter Fehler auf. Nachdem die Kommunikation zwischen Server und Client zunächst problemlos funktioniert, stürzt die Endanwenderapplikation auf dem PDA beim Lesen der Antwort ab. Bei der Verwendung einer Dockingstation und Mocha W32 PPP zum Aufbau einer Kommunikationsverbindung lässt sich feststellen, dass der Austausch von Paketen über die TCP/IP-Verbindung zwischen Server und Client längst abgeschlossen ist, bevor der Fehler auftritt. Das Parsen und somit auch das Analysieren der Antwort auf dem Client ist jedoch zum Fehlerzeitpunkt noch aktiv. In der XML-Schnittstelle findet kein Caching statt und die Zeichen im XML-Dokument entsprechen alle der XML-Norm sowie der Spezifikation. Bei der Veränderung der Sendegeschwindigkeit am Server ist festzustellen, dass der Fehler zu unterschiedlichen Analysezeitpunkten bei gleichen Daten auftritt. Bei einer starken Verringerung der Sendegeschwindigkeit kommt der Fehler gar nicht mehr vor. Die Fehlersituation lässt keinen eindeutigen Rückschluss auf dessen Ursache zu. Das Problem liegt vermutlich bei der Zwischenspeicherung der empfangenen Daten. Dieses könnte an einer fehlerhaften Implementierung der *Connection*- oder *InputStream*-Klasse der KVM 1.0 liegen.

Durch eine geeignete Anpassung der Sendegeschwindigkeit des Servers kann der Fehler vermieden werden. Die Beschränkung mindert jedoch durch die daraus resultierenden langen Übertragungszeiten wesentlich den Bedienungskomfort des gesamten Systems.

## 5.7. GUI auf dem Palm

Das Graphische User Interface auf dem Palm bildet die Bedienungsfläche für den Endbenutzer des Börseninformationssystems. Dieses ermöglicht dem Nutzer die Börseninformationen zu aktualisieren, zu lesen sowie das Profil entsprechend den eigenen Ansprüchen anzupassen.



Abbildung 5-53: Hauptfenster der mobilen Endanwenderapplikation

Beim Start der Applikation öffnet sich das Hauptfenster. Dieses enthält eine Menüleiste, Navigationselemente, eine Überschriftenliste und den Aktualisierungsbutton. Über die

Menüleiste kann die Profilverwaltung (siehe Kapitel 5.8 Profilverwaltung auf dem Palm ab Seite 79) erreicht werden. Innerhalb der Überschriftenliste werden die Überschriften der auf dem PDA vorhandenen Börseninformationen dargestellt. Da das Display des Palms sehr klein ist, werden hier nur maximal 9 Überschriften aufgeführt. Zum Weiterblättern stehen die rechts oben im Hauptfenster angeordneten Navigationselemente zur Verfügung. Die Überschrift setzt sich aus einer Abkürzung für die Art der Information (A: Analyse; N: News; T: Top; F: Flop; S: Schwankung; E: Neuemission) und dem Subject aus dem Header der Information zusammen. Durch Anklicken einer Überschrift öffnet sich ein Fenster mit den Daten der zugehörigen Börseninformation, falls diese bereits im Informationscache vorhanden sind. Sind die notwendigen Daten noch nicht im Informationscache (siehe 5.9 Informationscache auf dem Palm ab Seite 82) vorhanden, so werden diese im Online-Betrieb von der Applikation automatisch vom Server heruntergeladen, gespeichert und dargestellt. Mit der Betätigung des Button „Aktualisieren“ im Hauptfenster startet die Aktualisierung der Informationsheader (siehe Kapitel 5.5 Anfragebearbeitung auf dem Client – Aktualisierung der Börseninformationen ab Seite 70).

Zur Gestaltung von graphischen Oberflächen stehen bei der Kilobyte Virtual Machine von Sun die Klassen des Packages `com.sun.kjava` zur Verfügung. Diese beschränken sich jedoch auf einfache Oberflächenelemente, für deren Anordnung absolute Koordinaten angegeben werden müssen. Für die kJava-Dialogelemente fehlen sowohl Layout-Manager als auch Gruppierungselemente wie Panels. Wie auch die Erstellung von Oberflächen weicht das Eventhandling bei der KVM weitgehend von der Java 2 Standard Edition ab. Werden dort die Events mit Hilfe von Listener überwacht, so müssen bei der KVM die eingehenden Events von der `penDown`-Methode der Hauptklasse, die vom Objekt `Spotless` abgeleitet ist, an die Dialogelemente weitergeleitet werden. Das Listing in Abbildung 5-54 zeigt zur Veranschaulichung der Oberflächenprogrammierung und des Eventhandling die Implementierung der HelloWorld-Applikation für die KVM mit dem kJava-Package.

```

/**
 * HelloWorld mit den KJava-Klassen der KVM
 */

import com.sun.kjava.*;

public class HelloWorld extends Spotlet
{
    Button button;

    public HelloWorld()
    {
        button = new Button("Exit", 100, 100);

        Graphics g = Graphics.getGraphics();
        g.clearScreen();
        g.drawString("HelloWorld", 10, 10);
        button.paint();
    }

    public void penDown(int x, int y)
    {
        if (button.pressed(x, y) System.exit(0);
    }

    public static void main(String args[])
    {
        new HelloWorld().register(NO_EVENT_OPTIONS);
    }
}

```

Abbildung 5-54: HelloWorld mit den kJava-Bibliothek der KVM

Durch die Verwendung der kAWT-Bibliothek für die KVM werden viele der genannten Einschränkungen aufgehoben. Sie erweitert die KVM um eingeschränkte AWT-Klassen wie sie aus der Java 2 Standard Edition bekannt sind. Somit stehen für die Java-Programmierung auf dem Palm Dialogelemente wie beispielsweise Frame, Panel und Menu zur Verfügung. Die kAWT-Bibliothek erweitert weiterhin die KVM-Klassen um Layout-Manager (Card-, Border-, Flow- und GridLayout), durch deren Verwendung die Positionierung der Dialogelemente mit absoluten Koordinaten überflüssig ist, und um Event-Listener zur Vereinfachung des Eventhandling. Die Implementierung der HelloWorld-Applikation unter der Verwendung der kAWT-Klassen ist in Abbildung 5-55 dargestellt.

```

/**
 * HelloWorld mit den kAWT-Klassen
 */

import java.awt.*;
import java.awt.event.*;

public class HelloWorld extends Frame
{
    public HelloWorld()
    {
        super("HelloWorld");

        Button button = new Button("Exit");
        button.addActionListener(new ActionListener()
        {
            public void actionPerformed (ActionEvent e)
            {
                System.exit (0);
            }
        });
        add(BorderLayout.SOUTH, button);
        pack();
    }

    public static void main(String args[])
    {
        new HelloWorld().show();
    }
}

```

Abbildung 5-55: HelloWorld mit der kAWT-Bibliothek

Da sich die Oberflächenprogrammierung und das Eventhandling mit der Unterstützung der kAWT-Klassen wesentlich komfortabler und übersichtlicher gestaltet, kommt für die Implementierung der Endanwendung auf dem Palm die kAWT-Bibliothek zum Einsatz.

## Probleme

Die Erstellung einer in Java implementierten Applikation zur Ausführung auf dem Palm, erfolgt bei der KVM, wie bereits in Kapitel 4.3.1.2 Die KVM Version 1.0 ab Seite 38 geschildert, in drei Schritten:

1. Kompilieren des Java-Quellcode mit dem JDK 1.2
2. Preverifizieren der Klassendateien
3. Erzeugen einer für den Palm lesbaren *prc*-Datei

Wird innerhalb der erstellten Applikation eine Netzwerkverbindung, beispielsweise für die Verwendung einer Socketverbindung benötigt, so muss dieses im 3. Schritt zur Erzeugung der KVM-Applikation angegeben werden. Dieses erfolgt durch die Angabe des Parameters *-networking* beim Aufruf der für diesen Schritt notwendigen *palm.database.MakePalmApp* Programms. Vor dem Start der für den Palm generierten Java-Applikation wird somit zunächst eine Netzwerkverbindung aufgebaut. Eine Kommunikationsverbindung kann demnach nicht erst bei Bedarf von der Applikation heraus initiiert werden. Die Endanwenderapplikation des Börseninformationssystems ist für den Online- als auch für den Offline-Betrieb gedacht. Aus den oben genannten Gründen ist es für deren Realisierung notwendig zwei Applikationen zu schreiben. Die eine dient dem Gebrauch im Online-Betrieb und die andere ausschließliche der Arbeit im Offline-Betrieb.

Ein weiteres Problem ist die Darstellung von langen Texten, wie sie in Analysen oder News durchaus vorkommen können. In der derzeitigen Implementierung der KVM sind String-Objekte auf eine Zeichenkette mit maximal 256 Zeichen begrenzt. Bei der

Verarbeitung größerer Zeichenketten kann ein byte-Array verwendet werden. Die Oberflächenelemente TextBox, TextArea, TextField und andere, zur Darstellung von Texten bei der kJava- bzw. kAWT-Bibliothek, akzeptieren jedoch nur String-Objekte. Reicht die Größe des Textes einer Analyse bzw. News über 256 Zeichen hinaus, so ist die Verwendung der genannten Darstellungselemente nicht mehr möglich. Zur Lösung des Problems wurde ein neue TextArea-Klasse implementiert, welche die Anzeige von längeren Text erlaubt.

## 5.8. Profilverwaltung auf dem Palm

Das Ziel der Referenzimplementierung ist es, dem Anwender interessante und an seine individuellen Bedürfnisse angepasste Börseninformationen auf dem mobilen Endgerät zu präsentieren. Jedoch stellt jeder Nutzer andere Kriterien an die Informationen, die bei der Informationsreduktion am Server zu berücksichtigen sind. Das Profil gibt dem Endanwender die Möglichkeit seine Anforderungen an die Informationen zu definieren und somit den Reduktionsprozess am Server zu beeinflussen. Zur Konfiguration des Profils verfügt die Applikation auf dem Palm über eine Profilverwaltung. Hier kann der Nutzer Festlegungen zur Priorität der unterschiedlichen Informationsarten oder zur Bevorzugung bestimmter Aktien machen. Weiterhin ist in der Profilverwaltung die Möglichkeit gegeben, die Branchen des Neuen Marktes zu wichten und Schranken zu definieren.



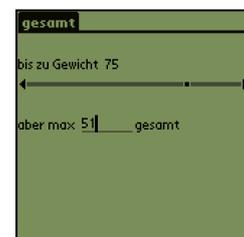
Abbildung 5-56: Profilverwaltung auf dem Palm

Die Profilverwaltung ist über den Menüpunkt Optionen/Einstellungen in der Endanwenderapplikation zu erreichen und gliedert sich in folgende Formulare:

### gesamt

Für jede Information wird serverseitig eine Signifikanz errechnet. Diese wird durch die anderen Profileinstellungen gegebenenfalls noch einmal gewichtet. Entsprechend der Signifikanz ist eine Information interessant oder nicht. Welche der Informationen jedoch bei einer Anfrage an den Client gesendet wird, lässt sich in diesem Fenster einstellen. Bei dem ersten Einstellungselement wird ein Gewicht (analog zu Signifikanz) zwischen 0 und 100 festgelegt, welches eine Information haben muss, um für den Anwender interessant zu sein. 0 steht hierbei für alle Informationen und 100 nur für die wichtigste(n) Information(en).

Um den Palm bei einer geringen Einstellung des oben erwähnten Wertes nicht mit Börseninformationen zu überfluten, kann zusätzlich angegeben werden, wie viele Informationen maximal erwartet werden.



## Tops, Flops, Schwankungen

Die Bestimmung der Tops, Flops und Schwankungen am Server bezieht sich auf unterschiedliche Betrachtungszeiträume. Der Endanwender kann in diesem Formular den zu Grunde liegenden Betrachtungszeitraum (1 Tag, 1 Woche, 1 Monat, 3 Monate, 6 Monate oder 1 Jahr) definieren. Des weiteren lässt sich die maximale Anzahl von Tops sowie Flops bei einer Aktualisierung festlegen. Gewöhnlich sinken und steigen die Kurse ständig. Besonders interessant sind Aktien, bei denen der Abstand zwischen Hoch und Tief sehr groß ist. Im letzten Einstellungselement kann Einfluss darauf genommen werden wie stark die Schwankung einer Aktie sein muss, um eine Information darüber zu erhalten.



## Kaufen –Verkaufen

Entsprechend den Bedürfnissen ist der Anwender des Systems mehr oder weniger an Kauf- beziehungsweise an Verkaufsempfehlungen interessiert. In diesem Fenster kann die Vergabe von Prioritäten für die beiden Empfehlungsarten vorgenommen werden. Die Priorität wird mit Hilfe eines Wertes zwischen 0 und 5 dargestellt, wobei 0 für nicht wichtig und 5 für sehr wichtig steht.



## Aktien

Besitzt der Nutzer des Börseninformationssystems Anteile einer oder mehrerer Aktien, so zeigt er in der Regel besonderes Interesse für diese Papiere. Eine Meldung zu der Aktie soll nicht erst übermittelt werden, wenn die Aktie zu den Top5 gehört oder besonders zum Verkauf eingestuft wurde. In diesem Fenster kann der Anwender 10 bevorzugte Aktien auswählen und entsprechend dem individuellen Interesse nach Wichtigkeit sortieren. Die dazugehörigen Informationen werden bei der Reduktion am Server höher gewichtet und fallen demzufolge öfter in die Auswahl, die bei der Aktualisierung an den PDA gesendet wird.



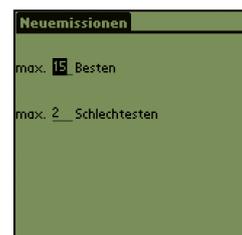
## Branchen

Für eine bessere Übersicht sind die Aktien des Neuen Marktes von der Deutschen Börsen AG in 10 Branchen unterteilt. Ist der Endanwender mehr oder weniger an speziellen Branchen interessiert, so hat er in diesem Fenster analog zur Priorisierung der Aktien die Möglichkeit die Branchen nach Wichtigkeit zu sortieren. Bei der serverseitigen profilabhängigen Reduktion werden die Börseninformationen der dazugehörigen Aktien dann gemäß der Sortierung gewichtet.



## Neuemissionen

Besonders Neuemissionen stehen im außerordentlichen Interesse der Aktionäre. Wie viele Informationen der Endanwender zu den besten oder auch schlechtesten Neuemissionen maximal erhalten möchte, kann in diesem Formular definiert werden.



## Priorität

Die Informationen des Systems unterteilen sich in 4 verschiedene Gruppen: 1. Analysen; 2. News; 3. Tops, Flops, Schwankungen; 4. Neuemissionen. Um zu definieren, welche der Informationsgruppen mehr oder weniger wichtig für den Nutzer des Systems sind, kann diesen eine Priorität von 0 bis 5 zugeordnet werden. Die Zahl 5 ist hierbei gleichbedeutend für sehr wichtig und die 0 für nicht wichtig.



In den oben genannten Formularen kann der Endanwender des Börseninformationssystems seinen Informationsbedarf individuell definieren, um somit auf dem mobilen Endgerät nur die für ihn interessantesten Börseninformationen zu erhalten. Hierzu wird das Profil bei der Aktualisierung der Börseninformationen an die Serverapplikation gesendet und innerhalb der profilabhängigen Reduktion (siehe Kapitel 5.4 Anfragebearbeitung und profilabhängige Reduktion ab Seite 63) verwendet.

Neben den Formularen für die Profileinstellungen sind in der Profilverwaltung der Endanwenderapplikation noch 3 weitere Funktionen enthalten. Dabei handelt es sich um Funktionen zur Aktualisierung der Aktien- und Branchennamen sowie um ein Formular zur Servereinstellung.

### Aktiennamen aktualisieren

Die Anzahl der Aktien am Neuen Markt ist nicht konstant. Fortwährend planen Unternehmen die Wandlung in eine Aktiengesellschaft und somit die Chance zur Eigenkapitalaufnahme. Der Endanwender des Systems kann die Informationen spezieller Aktien durch Definition in den Profileinstellungen bevorzugen. Um auch die Einstellungsmöglichkeit zu geben, die Börseninformationen neuer Aktien zu favorisieren, ist es notwendig die Aktiennamen auf dem PDA zu aktualisieren. Mit dem Menüpunkt „Aktiennamen aktualisieren“ wird dieser Vorgang gestartet und die Namen der Aktien auf dem PDA gespeichert.



### Branchennamen aktualisieren

Analog zu den Aktiennamen können die Branchennamen aktualisiert werden. Die Notwendigkeit dieser Aktion ist zwar seit der Aufteilung des Neuen Marktes in 10 Branchen im Juni 2000 noch nicht eingetreten, kann bei eventuellem Bedarf jedoch ausgeführt werden.

### Servereinstellungen

In diesem Formular können die Attribute, die für die Kommunikation mit dem Applikationsserver des Börseninformationssystems erforderlich sind, modifiziert werden. Hierbei handelt es sich um die IP und den Port unter dem der Börseninformationsserver zu erreichen ist.



Die Profil- und Servereinstellungen sowie die Namen der Aktien und Branchen werden in einem Palmkatalog eBIS-DB 2 gespeichert. Der Aufbau des Kataloges ist im Anhang C Datenstruktur des Palm-Kataloges zur Speicherung des Profils auf Seite 97 dargestellt.

Die Profilverwaltung ermöglicht jedem Nutzer die auf dem PDA präsentierten Börseninformation an seine individuellen Bedürfnisse anzupassen. Durch deren Integration in die Endanwenderapplikation kann die Profilkonfiguration unabhängig vom Server im Online- sowie Offline-Modus durchgeführt werden.

## 5.9. Informationscache auf dem Palm

Die Aufgabe des Informationscache's auf dem mobilen Endgerät besteht darin, den Kommunikationsaufwand zwischen Server und Client zu minimieren und den Betrieb im Offline-Modus zu ermöglichen. Hierzu werden alle vom Server empfangenen Börsendaten in eine Datenbank, den Cachespeicher, geschrieben. Greift der Endbenutzer auf eine bestimmte Information zu, wird diese zunächst in der Datenbank gesucht. Beim Auffinden können die Daten aus dem Cache zur Darstellung verwendet werden. Sind die Daten nicht vorhanden und der PDA ist Online, werden die Börseninformationen vom Server geholt und in die Datenbank eingefügt. Nach dem Ablauf der Geltungszeit einer Information, die von der Informationsart abhängig ist, wird diese aus dem Cache entfernt.

Aus den in Kapitel 4.4 Datenbank auf dem Palm ab Seite 40 geschilderten Gründen werden zur Speicherung von Daten die Palm-Kataloge verwendet. Für die unterschiedlichen Informationsarten existieren folgende Kataloge auf dem Palm:

<i>Bezeichnung</i>	<i>Katalogbezeichnung auf dem Palm</i>	<i>Beschreibung</i>
ID_DB	eBIS-DB 2	In der ID_DB werden alle Daten gespeichert die der Server bei einer <i>SELECT</i> -Anfrage an den Client zurücksendet. Diese umfassen ID, Subject und Informationsbewertung. Zusätzlich ist vermerkt, ob sich die zugehörigen Daten der Information im Cache befinden und wenn wo ungefähr. Bei Analysen und News kann zudem eine weitere ID gespeichert sein, die zur Aktualisierung des dazugehörigen Kurses dient.
ANALYSIS_DB	eBIS-DB 3	Innerhalb der ANALYSIS_DB sind die Attribute ID, ausführliche Überschrift, Analyst, aktueller Kurs, Datum & Zeit des aktuellen Kurses und Text einer Analyse gespeichert.
NEWS_DB	eBIS-DB 4	Die NEWS_DB enthält die gleiche Einträge wie die ANALYSIS_DB jedoch für News. Lediglich das Attribut Analyst entfällt.
TOPSFLOPS_DB	eBIS-DB 5	Die Daten eines Tops oder Flops werden in der TOPSFLOPS_DB hinterlegt. Hierzu gehören unter anderem die Attribute ID, aktueller -, Schluss-, Höchst- und Tiefstkurs sowie verschiedene Performance- als auch die dazugehörigen Zeitangaben.
NEWISSUES_DB	eBIS-DB 6	Neuemissionen werden in der NEWISSUES_DB mit den Attributen ID, aktueller Kurs, Emissionskurs, Erstnotiz, prozentuale Veränderung gegenüber dem Emissionskurs sowie gegenüber der Erstnotiz gespeichert.
FLUCTUATIONS_DB	eBIS-DB 7	In der FLUCTUATIONS_DB werden die Attribute ID, aktueller Kurs, Höchstkurs, Tiefstkurs, prozentuale Schwankungsbreite zwischen Hoch und Tief und die Zeiten der Kurse einer Schwankung hinterlegt.

Abbildung 5-57: Kataloge zum Speichern der Börseninformationen auf dem Palm

Bei üblichen Datenbanken werden zur Speicherung von Daten Datensätze (Tupel), meist bestehend aus mehreren Attributen, in Tabellen abgelegt. Im Gegensatz hierzu stehen bei der Verwendung von Palm-Katalogen nur Byte-Arrays, die einem Index zugewiesen sind, zur Verfügung. Zur Persistierung der Börseninformationen ist es notwendig, die Attribute geeignet an die Datenstruktur der Kataloge anzupassen. Hierfür gibt es drei Vorgehensweisen:

1. Die Attribute eines Datensatzes werden in einem Record zusammengefasst. Dieses wird in ein Byte-Array umgewandelt und im Katalog gespeichert. Bei dieser Vorgehensweise sind alle Attribute einer Börseninformation in einem Byte-Array vorhanden und unter einem Index in dem dazugehörigen Katalog hinterlegt.
2. Jedes Attribut des Datensatzes wird in ein Byte-Array umgewandelt. Diese werden hintereinander in einer bestimmten Reihenfolge im Katalog abgelegt. Im Katalog ergibt somit nicht nur ein Byte-Array, sondern mehrere aufeinanderfolgende Arrays die notwendigen Daten für eine Börseninformation.
3. Die Verfahren 1. und 2. werden miteinander vermischt. Die zu einem Datensatz zugehörigen Byte-Arrays enthalten mehrere oder einzelne Attribute.

Für den Informationscache der Endanwenderapplikation auf dem Palm wird das 3. Verfahren verwendet. Dieses erlaubt den schnellen Zugriff auf ein einzelnes Attribut, wie zum Beispiel die ID, da nicht der gesamte Datensatz ausgelesen und hieraus das erforderliche Attribut extrahiert werden muss. Das Speichern beziehungsweise Aktualisieren eines Attributs verläuft ebenfalls problemlos, denn es ist nicht notwendig das gesamte Record im Katalog zu schreiben.

Bezeichnung	Anzahl der Attribute für ein Tupel ( $\tau$ )
ID_DB	4
ANALYSIS_DB	6
NEWS_DB	5
TOPSFLOPS_DB	10
NEWISSUES_DB	5
FLUCTUATIONS_DB	6

Abbildung 5-58: Anzahl der Attribute eines Tupels

Ein Datensatz in der ID\_DB enthält die Attribute ID, Subject, Informationsbewertung, den Vermerk darüber, ob und wo ungefähr sich die zugehörigen Informationsdaten in dem entsprechenden Katalog befinden. Zusätzlich kann eine Analyse oder News ein Attribut mit einer zusätzlichen ID enthalten. Diese notwendigen Attribute werden in Form eines 4-Tupel in der ID\_DB hinterlegt. Eine genaue Darstellung über den Aufbau der Datenstruktur in der ID\_DB und den anderen Katalogen ist in Anhang B Datenstruktur der Palm-Kataloge ab Seite 96 zu finden.

Die Position einer Börseninformation und der dazugehörigen Attribute richtet sich nach der Attributanzahl eines Tupels in dem jeweiligen Katalog. So befinden sich zum

Beispiel die IDs der ID\_DB, in der ein Eintrag aus einem 4-Tupel besteht, bei Index 0, 4, 8, 12, ... und 4n.

$i = n \cdot t + p$			
i	Index		
n	Nummer des Datensatzes		
t	Anzahl der Attribute	eines	
	Datensatzes		
p	Position des Attributs	im	Datensatz

Abbildung 5-59: Indexberechnung eines Attributs im Palm-Katalog

Die allgemeine Berechnungsvorschrift zur Positionsbestimmung eines Attributes im Katalog ist in **Abbildung 5-59** dargestellt.

Bereits in **Kapitel 4.4 Datenbank auf dem Palm** ab Seite 40 wurde erwähnt, dass die Nutzung von Katalogen auf dem Palm gegenüber normalen Datenbanken wesentliche Einschränkungen mit sich bringen. So stellt der Index in den Palm-Katalogen lediglich die Nummerierung der Records im Katalog dar und bezieht sich nicht direkt auf einen Record. Dieser Umstand erschwert die Abbildung von Beziehungen zwischen verschiedenen Katalogen. Im konkreten Beispiel der Referenzimplementierung äußert sich dieses folgendermaßen: In der ID\_DB werden unter anderem die ID und die Information über die Existenz der zugehörigen Daten im Cache für jede Börseninformation gespeichert. Sind die zugehörigen Daten einer Information vorhanden und werden zur Darstellung benötigt, so ist es unausweichlich innerhalb des entsprechenden Kataloges nach der ID zu suchen. Durch die Speicherung der Position des Datensatzes innerhalb der ID\_DB wäre der Suchvorgang nicht nötig. Werden jedoch zwischenzeitig andere Datensätze innerhalb des entsprechenden Kataloges gelöscht, so ist die Positionsinformation in der ID\_DB eventuell bereits wieder hinfällig und somit die Zuordnung fehlerhaft.

## 6. Ergebnisse der Datenreduktion

Mit der gegebenen Architektur des Börseninformationssystems werden innerhalb unterschiedlicher Arbeitsabläufe Reduktionen vorgenommen. Durch das Zusammenwirken dieser ist die Informationsmenge, die letztendlich an das mobile Endgerät zu senden ist, reduziert.

In den folgenden Abschnitten werden die wesentlichen Reduktionsschritte und deren Effizienz diskutiert:

### Reduktion beim Wrappen

Börseninformationen die vom System aus dem Internet gesammelt werden sind in HTML-Seiten integriert. Betrachtet ein Anwender eine Information mit einem Browser, so muss der gesamte HTML-Code, indem sowohl die Information und deren Layout enthalten sind, geladen werden. Hinzu kommen eventuell dazugehörige Graphiken. Bei der Sammlung von Börseninformationen durch das System filtern die HTML-Parser nur die spezifizierten Daten heraus. Layoutinformationen und Graphiken entfallen. Für eine ausgewählte Analyse ließen sich die in Abbildung 6-1 dargestellten Informationsgrößen, die jeweils zu übertragen waren, ermitteln.

	<i>Information</i>	<i>Datengröße in Byte</i>
zur Darstellung im jeweiligem System übertragende Daten	reine Information der Analyse	
	Header (Subject, Analyst und Datum)	1.192
	Text	32
		$\Sigma$ 1.214
	Webseite auf dem Browser	
	Header (HTML-Seite mit einer Liste von Analysen)	232 + Overhead (min. 281.600) <sup>13</sup>
	Text (HTML-Seite ohne Bilder)	28.832
		$\Sigma$ <b>30.064</b>
	Börseninformationssystem eBIS	
	Header (innerhalb der ID-Liste)	93 + Overhead (min. 57) <sup>13</sup>
Text (Analysentext mit XML)	1.317	
	$\Sigma$ <b>1.410</b>	

Abbildung 6-1: Informationsgrößen einer ausgewählten Analyse

Die Gegenüberstellung der übertragenen Datengrößen der verschiedenen Systeme ergibt einen Reduktionskoeffizienten von etwa 1:21.

### Zusammenfassen von Informationen bei der Vorreduktion

In der Vorreduktion der Analysen werden die gesammelten Informationen zugehörig zu einer Aktie auf die aktuellste Analyse reduziert. Die anderen Analysen dienen nur zur Berechnung der Signifikanz. Analog wird mit den Ad-Hoc-News verfahren. So lassen sich zum Beispiel ca. 950 Analysen, die durchschnittlich in 20 Werktagen anfallen, auf

<sup>13</sup> Die Datengrößen der Header beziehen sich nur auf die speziellen Daten der ausgewählten Analyse. Der Overhead beinhaltet die von der gewählten Analyse unabhängig zu übertragenden Daten.

maximal 341 aktuellste Analysen reduzieren. Die Reduktion für die zu übertragende Datenmenge an das mobile Endgerät hängt dabei stark von der Häufigkeit des Aktualisierungsvorgangs auf dem Palm ab.

### **Profilabhängige Reduktion**

Im Verlauf der profilabhängigen Reduktion wird für ein definiertes Profil bestimmt, welche Informationen für die Übertragung selektiert werden. Die Mächtigkeit der selektierten Informationsmenge ist in der Regel wesentlich kleiner als die der Gesamtmenge und hängt dabei erheblich von den Profileinstellungen des Nutzers ab.

## 7. Zusammenfassung

Die Aufgabe der Studienarbeit ist die Implementierung einer Client-Server-Architektur für die mobile Umgebung. Hierbei wird auf innerhalb einer Datenbank verwaltete Informationen zugegriffen. Geringe Ressourcen des mobilen Endgerätes und eine langsame Kommunikationsverbindung zwischen Server und Client machen eine Reduktion der Daten notwendig. Für ein gewähltes Szenario (Börseninformationssystem) wurden verschiedene Datenreduktionsverfahren, wie Histogramme, Clustering Techniken und Reduktionstechniken aus dem Bereich des Data Mining untersucht. Deren Einsatz ist jedoch in der jeweils gegebenen Form für das Szenario nicht geeignet, da sich die darin verwendeten Reduktionsberechnungen speziell auf numerische Attribute beziehen. Erst die Anpassung von Ansätzen aus der Dimensionsreduktion des Data Mining ergeben eine zweckdienliche Datenreduktionsmethode für das Börseninformationssystem. Die Reduktion teilt sich in zwei Stufen und verwendet eine Abstandsfunktion, auf numerische Attribute und Attributen mit Zeichenketten, zur Bewertung von Börseninformationen. Hierbei bereitet speziell die Erfassung der semantischen Bedeutung von Zeichenketten Probleme, die umfassend nur mit linguistischen Analysen und dafür definierten Ontologien zu lösen ist. Die Bewertung der Informationen dient als Selektionskriterium. In der ersten Reduktionsstufe werden die Daten aufbereitet und in fünf Kategorien eingeteilt. Die zweite Stufe gestattet durch die Verwendung eines benutzerdefinierten Profils eine personalisierte Sicht auf die Daten. Das verwendete Datenreduktionsverfahren ist stark von dem gewählten Anwendungsszenario abhängig und kann ohne grundsätzliche Modifikationen nicht auf allgemeine mobile Anwendungen übertragen werden. Weitere unterstützende Ansätze hinsichtlich der Nutzung mobiler Ressourcen sind der Empfang von Börseninformationen in mehreren Abstraktionsstufen und deren Speicherung auf dem mobilen Endgerät.

Der Umsetzung des Börseninformationssystems in eine prototypische Implementierung ging eine detaillierte Analyse der fachspezifischen und technischen Konzepte voraus. Die Implementierung der Client-Server-Architektur wurde in der plattformunabhängigen Programmiersprache Java vorgenommen. Die mobile Endanwenderapplikation setzt dabei auf die KVM von Sun auf. Die aktuelle Version der KVM ist noch fehlerbehaftet und unterstützt nicht im vollen Umfang die Plattformunabhängigkeit. Weiterhin existieren zu vorhandenen Datenbanksystemen auf dem Palm keine Schnittstellen, wie zum Beispiel JDBC. Der alternative Einsatz von Palm-Katalogen, deren Anbindung durch die KVM unterstützt wird, stellt sich dagegen als unbefriedigend heraus. Eine Lösung des Problems könnte die Verwendung von DB2-Everyplace und eine damit verbundene Implementierung der Endanwenderapplikation mit der VisualAge Micro Edition von IBM darstellen.

Die Wrapperkomponente des Servers sammelt die für das Börseninformationssystem notwendigen Daten. Die dafür benutzten regulären Ausdrücke können die Börseninformationen bei Strukturveränderungen an den Quelldaten nicht mehr erfassen. Als Lösungsansatz sei in diesem Zusammenhang World Wide Web Wrapper Factory (W4F) genannt.

Zur Kommunikation zwischen Applikationsserver und Client ist XML als Datenaustauschformat gewählt. XML erlaubt die Trennung zwischen Daten und Struktur, erweist sich aber innerhalb der mobilen Systemumgebung als unnötiger Kommunikationsoverhead.

Der Einsatz der Referenzimplementierung des Börseninformationssystems zeigt, dass die entwickelten Konzepte eine geeignete Datenreduktion liefern. Der Applikationsserver und die clientseitige Endanwenderapplikation stellen eine passende Systemarchitektur für die mobile Umgebung dar, durch deren Einsatz eine effiziente und vom Nutzer beeinflussbare Reduktion erreicht wird.

## Literaturverzeichnis

- [Au99] S.H. Audersch: Datenreduktion für Data Mining. In: Ausarbeitung zum Hauptseminar, Universität Rostock, 1999.
- [BD97] D. Barbara, W. DuMouchel, ...: The New Jersey Data Reduction Report. In: Data Engineering Vol. 20 No. 4, pages 3-44, 1997.
- [CLDC] Sun Microsystems: Java 2 Platform Micro Edition – Connected Limited Device Configuration Version 1.0.  
<http://www.sun.com/software/communitysource/j2me>, 2000.
- [De99] C. Dethloff: Untersuchung von nichtparametrisierten Datenreduktionsverfahren. In: Ausarbeitung zum Hauptseminar, Universität Rostock, 1999.
- [EKXS96] M. Ester, H.P. Kriegel, J. Sander and X. Xu: A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proc. Second International Conference on Knowledge Discovery and Data Mining, pages 226-231, 1996.
- [EPOC] Symbian Ltd.: EPOC Betriebssystem für Psion Personal Organizer.  
<http://www.symbian.com>, 2000.
- [Ern97] N. Ernst: PDA: Psion contra Windows CE. In c't Band 8, Seite 50, 1997.
- [Fa00] T. Fanghaenel: Vergleich und Bewertung kommerzieller mobiler Datenbanksysteme. In: Diplomarbeit, Friedrich-Schiller-Universität Jena, 09/2000
- [HK00] S. Haustein und M. Kroll: Klein, aber fein: Java-Entwicklung mit CLDC auf Kleinstgeräten. In: Java Spektrum Band 6, Seite 24, 2000.
- [InVi00] Investor-Village: Palm Kursabschlag. In: <http://www.aktiencheck.de>, aktiencheck.de AG, 21.12.2000.
- [kAWT] The kAWT-Project. <http://www.kawt.de/>, 2000.
- [MD88] M. Muralikrishna and D.J. Dewitt: Equi-depth histograms for estimating selectivity factors for multi-dimensional queries. In: Proc. Of ACM SIGMOD Conf., pages 28-36, 1988.
- [Mocha] MochaSoft: Mocha W32 PPP. <http://www.mochasoft.dk>, 2000.
- [NH94] R. Ng and J. Han: Efficient and Effective Clustering Method for Spatial Data Mining. In: Proc. VLDB, pages 144-155, 1994.
- [Orac] Oracle Inc.: Oracle XML Parser 2.0.2.6.0.  
[http://technet.oracle.com/tech/xml/parser\\_java2/](http://technet.oracle.com/tech/xml/parser_java2/), 2000.
- [ORO] D.F. Savarese: OROMatcher. ORO Inc., <http://www.oro-inc.com/>, 2000.
- [Palm] Palm Inc.: Palm Produkte und Palm OS. <http://www.palm.com>, 2000.

- [PC84] G. Piatesky-Shapiro and C. Connell: Accurate estimation of the number of tuples satisfying a condition. In: Proc. Of ACM SIGMOD Conf., Pages 256-276, 1984.
- [Perl] D.F. Savarese: PerlTools. ORO Inc., <http://www.oroinc.com/>, 2000.
- [PIH96] V. Poosala, Y.E. Ioannidis, P.J. Haas and E.J. Shekita: Improved Histograms for Selectivity Estimation of Range Predicates. In: Proc. Of ACM SIGMOD Conf., Pages 294-305, 1996.
- [POSE] Palm Inc.: Palm OS Emulator. <http://www.palm.com>, 2000.
- [Psio] Psion PLC: Psion Personal Organizer. <http://www.pSION.com>, 2000.
- [Rin00] J. Rink: Personal Digital Assistants: Multimedia in der Hand. In: c't Band 24, Seite 144, 2000.
- [Rin01] J. Rink: Mobile Internet überall: Das optimale Zubehör für den Online-Spaß unterwegs. In c't Band 4, Seite 124, 2001.
- [SAC79] P.G. Selinger, M.M. Astrahan, D.D. Chamberlin, R.A. Lorie and T.T. Price: Access path selection in a relational database management system. In: Proc. Of ACM SIGMOD Conf, pages 23-34, 1979.
- [Sch00] T.J. Schult: Farbwahl: 3Com-Organizer Palm IIIe, IIIc und Vx. In: c't Band 5, Seite 94, 2000
- [Sch00b] T.J. Schult: Personal Digital Assistants: Tastenkünstler. In c't Band 7, Seite 110, 2000.
- [Sch97] T.J. Schult: Organizer: Neue Beweglichkeit. In c't Band 15, Seite 138, 1997.
- [SR98] T.J. Schult und J. Rink: Mobile Computing: Überblick: Leichtes Gepäck. In c't Band 23, Seite 173, 1998.
- [Sta00] G. Starke: Gullivers Freunde: Java auf dem Palm-Pilot. In: Java Spektrum Band 3, Seite 18, 2000.
- [Sun00] Sun Microsystems: Java 2 SDK, Standard Edition Documentation. <http://java.sun.com/products/jdk/1.2>, 2000.
- [Tiny] T. Gibara: TinyXML. <http://www.gibaradunn.srac.org/tiny/index.shtml>, 2000.
- [Waba] Wabasoft Inc.: Waba Virtual Maschine für PalmOS und WindowsCE. <http://www.wabasoft.com/>, 2000.
- [WI98] S.M. Weiss and N. Indurkha: Predictive Data Mining: A Partical Guide. Morgan Kaufmann Publishers, San Francisco, CA, 1998.
- [ZRL96] T. Zhang, R. Ramakrishnan and M. Livny: Birch: An Efficient Data Clustering Method for Very Large Database. In: Prod. Of ACM SIGMOD Conf, pages 103-114, 1996.

## Abbildungsverzeichnis

Abbildung 2-1: Übersicht Tops des Anbieter www.goingpublic-online.de .....	10
Abbildung 2-2: Übersicht Flops des Anbieter www.goingpublic-online.de.....	11
Abbildung 2-3: Kursliste beim Anbieter www.goingpublic-online.de.....	11
Abbildung 2-4: Schwankungen.....	11
Abbildung 2-5: allgemeiner Systemaufbau des Börseninformationssystems .....	12
Abbildung 2-6: Teilung der Reduktionsberechnung .....	15
Abbildung 2-7: Aufbau des Börseninformationssystems.....	16
Abbildung 3-1: Häufigkeit der Attributwerte .....	17
Abbildung 3-2: Attributwerte in dem Block.....	18
Abbildung 3-3: Reduktionsergebnis der ersten Annahme.....	18
Abbildung 3-4: Attributwerte in dem Block.....	18
Abbildung 3-5: Reduktionsergebnis der zweiten Annahme .....	19
Abbildung 3-6: Reduktionsergebnis der Analysen (I).....	20
Abbildung 3-7: Aktien mit Analysen .....	20
Abbildung 3-8: Aktien mit den normierten Analysewerten.....	21
Abbildung 3-9: Szenarien für die Darstellung der uninteressanten Aktien .....	21
Abbildung 3-10: Histogramm für die Analysen.....	22
Abbildung 3-11: reduzierte Analysen nach den zwei Annahmen.....	22
Abbildung 3-12: Aktien mit den normierten Kurswerten .....	23
Abbildung 3-13: Histogramm für die Kurse.....	23
Abbildung 3-14: reduzierte Kurse nach den zwei Annahmen.....	24
Abbildung 3-15: Übersicht über Analysen und deren Bewertung je Unternehmen .....	25
Abbildung 3-16: Normierung der Werte.....	26
Abbildung 3-17: Darstellung der Cluster mit den jeweiligen repräsentativen Objekten.	26
Abbildung 3-18: Clusterbildung der Unternehmen, bezogen auf die jeweilige Kursperformance.....	27
Abbildung 3-19: Formen der Dimensionsreduktion .....	28
Abbildung 4-1: Psion Serie 5 MX .....	31
Abbildung 4-2: Compaq 810c .....	32
Abbildung 4-3: Palm Vx .....	32
Abbildung 4-4: PalmPilots .....	33
Abbildung 4-5: Java 2 Technologie.....	36
Abbildung 4-6: Aufbau der Java 2 Micro Edition.....	37
Abbildung 4-7: Beziehungen zwischen der J2SE und den Konfigurationen der J2ME ..	38
Abbildung 4-8: Erzeugen einer Java-Applikation für den Palm .....	39
Abbildung 4-9: Zugriff mit Java auf den Palm-Katalog.....	41
Abbildung 4-10: Konvertierung der Integerwerte .....	42
Abbildung 5-1: Browserdarstellung der Aktienliste .....	45
Abbildung 5-2: HTML-Quellcode für ersten Scanvorgang.....	45
Abbildung 5-3: regulärer Ausdruck des ersten Scanvorgangs .....	46
Abbildung 5-4: HTML-Quellcode für den zweiten Scanvorgang .....	46
Abbildung 5-5: regulärer Ausdruck des zweiten Scanvorgangs.....	46
Abbildung 5-6: HTML-Quellcode für den dritten Scanvorgang.....	47
Abbildung 5-7: regulärer Ausdruck des dritten Scanvorgangs .....	47
Abbildung 5-8: Fundamentaldaten der Beispielaktie.....	48
Abbildung 5-9: Beispielanalysen .....	49
Abbildung 5-10: Schlüsselwörter zum Katalogisieren der Analysen .....	50
Abbildung 5-11: Bewertungszuordnung bei Analysen .....	50
Abbildung 5-12: Bewertung der Beispielanalysen .....	50
Abbildung 5-13: Altersbewertung bei Analysen .....	51
Abbildung 5-14: Gruppierung der Beispielanalysen.....	51

Abbildung 5-15: Signifikanzen der Beispielanalysem .....	52
Abbildung 5-16: Aufbau der ID einer vorreduzierten Analyse.....	52
Abbildung 5-17: vorreduzierte Beispielanalyse .....	52
Abbildung 5-18: Beispiel-Ad-Hoc-News .....	53
Abbildung 5-19: Bewertung der Ad-Hoc-News .....	53
Abbildung 5-20: Bewertung der Beispiel-Ad-Hoc-News .....	54
Abbildung 5-21: Gruppierung der Beispiel-Ad-Hoc-News .....	54
Abbildung 5-22: Signifikanzen der Beispiel-Ad-Hoc-News .....	54
Abbildung 5-23: Aufbau der ID einer vorreduzierten News.....	54
Abbildung 5-24: vorreduzierte Beispielnews .....	55
Abbildung 5-25: Daten einer Aktie für den Betrachtungszeitraum 1 Tag.....	56
Abbildung 5-26: prozentuale Kursänderungen.....	56
Abbildung 5-27: Aufbau der ID bei TopsFlops .....	57
Abbildung 5-28: vorreduzierter TopFlop.....	57
Abbildung 5-29: Kurse der Beispielaktie .....	58
Abbildung 5-30: Schwankungsbreite und Signifikanz der Beispielaktie.....	58
Abbildung 5-31: Aufbau der ID bei Schwankungen.....	59
Abbildung 5-32: vorreduzierte Schwankung.....	59
Abbildung 5-33: Fundamentaldaten der Beispielaktie .....	59
Abbildung 5-34: Kurse der Beispielaktie .....	60
Abbildung 5-35: prozentuale Kursänderungen.....	60
Abbildung 5-36: Aufbau der ID bei Neuemissionen .....	61
Abbildung 5-37: vorreduzierte Neuemission.....	61
Abbildung 5-38: DB-Schema.....	62
Abbildung 5-39: Vorreduktionstabellen für Analysen, News und Neuemissionen.....	63
Abbildung 5-40: Vorreduktionstabellen für unterschiedliche Betrachtungszeiträume....	63
Abbildung 5-41: Vorreduktionstabellen zur Historisierung .....	63
Abbildung 5-42: Wirkung von Faktoren und Schranken .....	64
Abbildung 5-43: Parameter <i>buy</i> und die dazugehörigen Faktoren.....	64
Abbildung 5-44: Faktorenuordnung des Parameter <i>weights</i> .....	65
Abbildung 5-45: Faktoren beim Parameter <i>stocks</i> .....	65
Abbildung 5-46: Faktoren beim Parameter <i>sectors</i> .....	66
Abbildung 5-47: Ablauf der profilabhängigen Reduktion .....	69
Abbildung 5-48: Aktualisierung der Informationsheader am Beispiel der Analysen.....	71
Abbildung 5-49: Aufbau der ID einer Analyse .....	71
Abbildung 5-50: Berechnung des Alters einer Information.....	71
Abbildung 5-51: Vergleich der Informationsheader am Beispiel der Analysen.....	72
Abbildung 5-53: Hauptfenster der mobilen Endanwenderapplikation .....	75
Abbildung 5-54: HelloWorld mit den kJava-Bibliothek der KVM.....	77
Abbildung 5-55: HelloWorld mit der kAWT-Bibliothek.....	78
Abbildung 5-56: Profilverwaltung auf dem Palm .....	79
Abbildung 5-57: Kataloge zum Speichern der Börseninformationen auf dem Palm .....	82
Abbildung 5-58: Anzahl der Attribute eines Tupels .....	83
Abbildung 5-59: Indexberechnung eines Attributs im Palm-Katalog .....	84
Abbildung 6-1: Informationsgrößen einer ausgewählten Analyse.....	85

# Anhang A

## DTD's zur Kommunikation zwischen Server und Client

### Client2Server.dtd

```
<!-- DTD Client2Server -->
<!-- Diese DTD dient zur Kommunikation zwischen Client (Palm) und Server des -->
<!-- Boerseninformationssystems. -->
<!-- In dieser DTD wird die AnfrageStruktur des Client (Palm) an den Server -->
<!-- festgelegt -->

<!ENTITY % timePeriodes "( 1 | 2 | 3 | 4 | 5 | 6 )">
    <!-- Tag, Woche, Monat, 3 Monate , 6 Monate,-->
    <!-- Jahr -->

<!ENTITY % priorities "( 0| 1 | 2 | 3 | 4 | 5 )">
    <!-- 0=deaktiviert 5=hoch -->

<!ELEMENT REQUEST (GETALLSTOCKS | GETALLSECTORS | GETCOURSE | GETINFORMATION | SELECT)>
    <!-- Wurzelknoten -->

<!ELEMENT GETALLSTOCKS EMPTY>
    <!-- Anfrage an den Server, alle Aktienamen -->
    <!-- zurueckzugeben (Rueckgabe: Aktienamen -->
    <!-- aller Aktien) -->

<!ELEMENT GETALLSECTORS EMPTY>
    <!-- Anfrage an den Server, alle Branchennamen -->
    <!-- zurueckzugeben (Rueckgabe: Branchennamen -->
    <!-- aller Branchen) -->

<!ELEMENT GETINFORMATION EMPTY>
    <!-- Anfrage an den Server, die Information -->
    <!-- mit der entsprechenden ID zurueckzugeben -->
    <!-- (Rueckgabe: Information) -->

<!ATTLIST GETINFORMATION
    iid CDATA; #REQUIRED>
    <!-- ID der Information -->

<!ELEMENT GETCOURSE EMPTY>
    <!-- Anfrage an den Server, den Kurse fuer -->
    <!-- die entsprechenden ID zurueckzugeben -->
    <!-- (Rueckgabe: Information) -->

<!ATTLIST GETCOURSE
    iid CDATA; #REQUIRED>
    <!-- ID der Information -->

<!ELEMENT SELECT (STOCKS? , SECTORS?, WEIGHTS?)>
    <!-- Anfrage an den Server, die wichtigsten -->
    <!-- Informationen zurueckzugeben (Rueckgabe: -->
    <!-- IDs und kurze Ueberschriften der -->
    <!-- ausgewaehlten Informationen) -->

<!ATTLIST SELECT
    sig CDATA; #REQUIRED
    <!-- Signifikanz, die eine Information -->
    <!-- erreichen muss, um ausgewaehlt zu werden -->
    mIC CDATA; #REQUIRED
    <!-- maximale Anzahl von Informationen, die -->
    <!-- vom Server ausgewaehlt duerfen -->
    buy CDATA; #REQUIRED
    <!-- Prioritaet fuer Kaufempfehlungen -->
    sal CDATA; #REQUIRED
    <!-- Prioritaet fuer Verkaufsempfehlungen -->
    tPe %timePeriodes; #REQUIRED
    <!-- Betrachtungszeitraum bei Tops, Flops -->
    <!-- und Schwankungen fuer die entsprechenden -->
    <!-- Berechnungen -->
    mTo CDATA; #REQUIRED
    <!-- maximale Anzahl von Tops, die vom Server -->
    <!-- ausgewaehlt duerfen -->
    mFl CDATA; #REQUIRED
    <!-- maximale Anzahl von Flops, die vom Server -->
    <!-- ausgewaehlt duerfen -->
    flu CDATA; #REQUIRED
    <!-- minimale Schwankungsbreite bei -->
    <!-- Schwankungen, die vom Server ausgewaehlt -->
    <!-- werden duerfen -->
```

```

mTN CDATA; #REQUIRED <!-- maximale Anzahl der Besten Neuemmissionen, -->
                        <!-- die vom Server ausgewaehlt werden duerfen -->
mFN CDATA; #REQUIRED> <!-- maximale Anzahl der Schlechtesten -->
                        <!-- Neuemmissionen, die vom Server ausgewaehlt -->
                        <!-- werden duerfen -->

<!ELEMENT STOCKS EMPTY> <!-- hoehere Priorisierung der folgenden 10 -->
                        <!-- Aktien (nach der Reihenfolge sortiert; -->
                        <!-- sl=wichtigste) -->

<!ATTLIST STOCKS
s1 CDATA; #REQUIRED
s2 CDATA; #REQUIRED
s3 CDATA; #REQUIRED
s4 CDATA; #REQUIRED
s5 CDATA; #REQUIRED
s6 CDATA; #REQUIRED
s7 CDATA; #REQUIRED
s8 CDATA; #REQUIRED
s9 CDATA; #REQUIRED
s10 CDATA; #REQUIRED>

<!ELEMENT SECTORS EMPTY> <!-- Priorisierung der folgenden 10 Branchen -->
                        <!-- (nach der Reihenfolge sortiert; -->
                        <!-- sl=wichtigste) -->

<!ATTLIST SECTORS
s1 CDATA; #REQUIRED
s2 CDATA; #REQUIRED
s3 CDATA; #REQUIRED
s4 CDATA; #REQUIRED
s5 CDATA; #REQUIRED
s6 CDATA; #REQUIRED
s7 CDATA; #REQUIRED
s8 CDATA; #REQUIRED
s9 CDATA; #REQUIRED
s10 CDATA; #REQUIRED>

<!ELEMENT WEIGHTS EMPTY> <!-- Wichtigung (Prioritaet) der einzelnen -->
                        <!-- Informationskategorien -->

<!ATTLIST WEIGHTS
ana %priorities; #REQUIRED <!-- Analysen -->
new %priorities; #REQUIRED <!-- News -->
tFF %priorities; #REQUIRED <!-- Tops, Flops und Schwankungen -->
nIs %priorities; #REQUIRED> <!-- Neuemmissionen -->

```

## Server2Client.dtd

```

<!-- DTD Server2Client -->
<!-- Diese DTD dient zur Kommunikation zwischen Client (Palm) und Server des -->
<!-- Boerseninformationssysteme. -->
<!-- In dieser DTD wird die AntwortStruktur des Servers zum Client (Palm) festgelegt -->

<!ELEMENT S (#CDATA)> <!-- Aktien- bzw. BranchenName -->

<!ELEMENT BR EMPTY> <!-- Zeilenumbruch -->

<!ELEMENT RESPONSE (ALLSTOCKS | ALLSECTORS | INFORMATION | SELECTIONS)>
                        <!-- Wurzelknoten -->

<!ELEMENT ALLSTOCKS (S*)> <!-- Rueckgabe alle Aktiennamen -->

<!ELEMENT ALLSECTORS (S*)> <!-- Rueckgabe alle Branchennamen -->

<!ELEMENT INFORMATION (NEWS | ANALYSIS | TOPFLOP | NEWISSUES | FLUCTUATION | COURSE)>
                        <!-- Rueckgabe einer Information -->

<!ELEMENT NEWS (#CDATA+ | BR)> <!-- News (bestehend aus Text mit -->
                        <!-- Zeilenvorschueben) -->

<!ATTLIST NEWS
sub CDATA; #REQUIRED <!-- ausfuehrliche Ueberschrift -->

```

```

        cou CDATA; #REQUIRED          <!-- aktueller Kurs -->
        dat CDATA; #REQUIRED          <!-- Datum & Zeit des aktuellen Kurses -->

<!ELEMENT ANALYSIS (#CDATA+ | BR)>   <!-- Analyse (bestehend aus Text mit -->
<!-- Zeilenvorschueben) -->
<!ATTLIST ANALYSIS
    sub CDATA; #REQUIRED             <!-- ausfuehrliche Ueberschrift -->
    ana CDATA; #REQUIRED             <!-- Analyst -->
    cou CDATA; #REQUIRED             <!-- aktueller Kurs -->
    dat CDATA; #REQUIRED             <!-- Datum & Zeit des aktuellen Kurses -->

<!ELEMENT TOPFLOP EMPTY>            <!-- Top bzw. Flop -->
<!ATTLIST TOPFLOP
    cou CDATA; #REQUIRED             <!-- aktueller Kurs -->
    fCo CDATA; #REQUIRED             <!-- Schlusskurs -->
    hCo CDATA; #REQUIRED             <!-- Hoechstkurs -->
    lCo CDATA; #REQUIRED             <!-- Tiefstkurs -->
    vFC CDATA; #REQUIRED             <!-- Performance zum Schlusskurs -->
    vHC CDATA; #REQUIRED             <!-- Performance zum Hoechstkurs -->
    vLC CDATA; #REQUIRED             <!-- Performance zum Tiefstkurs -->
    dFC CDATA; #REQUIRED             <!-- Datum des Schlusskurses -->
    dHC CDATA; #REQUIRED             <!-- Datum des Hoechstkurses -->
    dLC CDATA; #REQUIRED             <!-- Datum des Tiefstkurses -->

<!ELEMENT NEWISSUES EMPTY>         <!-- Neuemission -->
<!ATTLIST NEWISSUES
    cou CDATA; #REQUIRED             <!-- aktueller Kurs -->
    emC CDATA; #REQUIRED             <!-- Emissionskurs -->
    fiC CDATA; #REQUIRED             <!-- Erstnotiz -->
    vEC CDATA; #REQUIRED             <!-- prozentuale Veraenderung gegeneuber -->
    <!-- dem Emissionskurs -->
    vFC CDATA; #REQUIRED             <!-- prozentuale Veraenderung gegeneuber -->
    <!-- der Erstnotiz -->

<!ELEMENT FLUCTUATION EMPTY>       <!-- Schwankung -->
<!ATTLIST FLUCTUATION
    cou CDATA; #REQUIRED             <!-- aktueller Kurs -->
    hCo CDATA; #REQUIRED             <!-- Hoechstkurs -->
    lCo CDATA; #REQUIRED             <!-- Tiefstkurs -->
    vHL CDATA; #REQUIRED             <!-- prozentuale Schwankungsbreite zwischen -->
    <!-- Hoechst- und Tiefstkurs -->
    dHC CDATA; #REQUIRED             <!-- Datum des Hoechstkurses -->
    dLC CDATA; #REQUIRED             <!-- Datum des Tiefstkurses -->

!ELEMENT COURSE EMPTY>            <!-- Kurs -->
<!ATTLIST COURSE
    cou CDATA; #REQUIRED             <!-- aktueller Kurs -->
    pBD CDATA; #REQUIRED             <!-- Performance zum Tag -->
    dat CDATA; #REQUIRED             <!-- Datum & Zeit des aktuellen Kurses -->

<!ELEMENT SELECTIONS (SUB*)>       <!-- ausgewaehlte Informationen des Servers -->
<!ATTLIST SELECTIONS
    size CDATA; #REQUIRED           <!-- Anzahl der ausgewaehlten Informationen -->

<!ELEMENT SUB (#CDATA)>           <!-- eine ausgewaehlte Information des Servers -->
<!-- (bestehend aus einer kurzen Ueberschrift) -->
<!ATTLIST SUB
    id CDATA; #REQUIRED              <!-- ID der Information -->
    rat CDATA; #REQUIRED             <!-- Bewertung der Information -->

```

## Anhang B

### Datenstruktur der Palm-Kataloge

Die Struktur der Datensätze im Palm-Kataloge ID\_DB setzt sich folgendermaßen zusammen:

ID\_DB

<i>Nr. (p)</i>	<i>Bezeichnung</i>	<i>Datentyp</i>	<i>Datentyp in der DB</i>
0	ID	String	Byte[]
1	Subject	String	Byte[]
2	Informationsbewertungen	int	8 Bit Integer unsigned (Byte)
3	1	Information in DB	boolean
3	2	Aktuellster Index in DB	int
3	3	ID2 (bei News und Analysen)	String

ANALYSIS\_DB

NEWS\_DB

TOPSFLOPS\_DB

NEWISSUES\_DB

FLUCTUATIONS\_DB

## Anhang C

### Datenstruktur des Palm-Katalogs zur Speicherung des Profils

Index	Fenster im GUI	Bezeichnung	Typ	Wertebereich
0	Server	IP 1	int16u	0...999
1	Server	IP 2	int16u	0...999
2	Server	IP 3	int16u	0...999
3	Server	IP 4	int16u	0...999
4	Server	Port	int16u	0...9999
5	gesamt	Gewicht	int8u	5...100
6	gesamt	max	int16u	>0
7	Prioritäten	isAktiv	boolean	
8	Prioritäten	Analysen	int8u	0...5
9	Prioritäten	News	int8u	0...5
10	Prioritäten	Tops, Flops, Schwankungen	int8u	0...5
11	Prioritäten	Neuemissionen	int8u	0...5
12	Tops,Flops, Schwankungen	Betrachtungszeitraum	int8u	0...5
13	Tops,Flops, Schwankungen	max Tops	int16u	>0
14	Tops,Flops, Schwankungen	max Flops	int16u	>0
15	Tops,Flops, Schwankungen	Volatilität	int8u	>0
16	Kaufen - Verkaufen	Kaufempfehlungen	int8u	0...5
17	Kaufen - Verkaufen	Verkaufempfehlungen	int8u	0...5
18	Neuemissionen	max Besten	int16u	>0
19	Neuemissionen	max Schlechtesten	int16u	>0
20	Aktien	isAktiv	boolean	
21	Aktien	Aktie 1	string	
22	Aktien	Aktie 2	string	
23	Aktien	Aktie 3	string	
24	Aktien	Aktie 4	string	
25	Aktien	Aktie 5	string	
26	Aktien	Aktie 6	string	
27	Aktien	Aktie 7	string	
28	Aktien	Aktie 8	string	
29	Aktien	Aktie 9	string	
30	Aktien	Aktie 10	string	
31	Branchen	isAktiv	boolean	
32	Branchen	Anzahl Branchen	int8u	>0 <Anz. Br. m.
33	Branchen	Anzahl Branchen max	int8u	>0
34	Branchen	Branche 1	string	
.	.	.	.	
.	.	.	.	
.	.	.	.	
34 + Anzahl Branchen - 1	Branchen	Branche n	string	
34 + Anzahl Branchen	Branchen	<i>null</i>	<i>leer</i>	
.	.	.	.	
.	.	.	.	
.	.	.	.	
34 + Anzahl Branchen max - 1	Branchen	<i>null</i>	<i>leer</i>	
34 + Anzahl Branchen max	Aktien	AktienName 1	string	
.	.	.	.	
.	.	.	.	
.	.	.	.	
ENDE	Aktien	AktienName n	string	