

Data Mining auf Kundendaten am Beispiel verschiedener Alumnivereinigungen

Diplomarbeit im Fach Informatik

Vorgelegt von:

Matthias Linherr
Zürich, Schweiz
Matrikel-Nummer: 98-919-350

Ausgefertigt am
Institut für Informatik der Universität Zürich
Prof. Dr. A. Bernstein



Betreuer: Peter Vorburger

Abgabe der Arbeit: 8. Juli 2007

Zusammenfassung

Ziel dieser Arbeit ist es, eine Plattform für Alumni-Vereine zu implementieren, auf welcher Auswertungen auf Kundendaten vorgenommen werden können. Durch Einsatz von statistischen Methoden und Data-Mining Algorithmen sollen die Mitgliederstruktur und das -verhalten ausgewertet und visualisiert werden. Vier verschiedene Alumniorganisationen nutzen diese Plattform im Rahmen einer web-basierten Applikation, mit welcher sie ihre Datenbestände bewirtschaften, welche als Basis für die in dieser Arbeit beschriebenen Auswertungen dienen.

Abstract

The aim of this thesis is to implement a platform to enable the alumni associations to analyse their member-databases. Using statistical methods and data-mining algorithms, this platform should allow the visualization and appraisal of member behaviour and member structure. Four different alumni organisations utilise this platform in form of a web-based application to maintain their databases. They form the base of the following evaluations.

Inhaltsverzeichnis

1.	Einleitung.....	4
1.1	Alumniorganisationen.....	4
1.2	Das betreuende Unternehmen.....	6
1.3	Alumnitools.....	6
1.4	Aufbau der Diplomarbeit.....	10
2.	Anforderungen.....	11
3.	Voraussetzungen.....	13
3.1	Die Mandanten.....	13
3.1.1	Mandant 1.....	14
3.1.2	Mandant 2.....	17
3.1.3	Mandant 3.....	18
3.1.4	Mandant 4.....	20
3.2	Technisches Umfeld.....	21
4.	Implementierung.....	22
4.1	Einleitung.....	22
4.2	JFreeChart.....	26
4.2.1	Pie-Chart.....	27
4.2.2	Bar-Chart.....	28
4.2.3	Stacked Bar-Chart.....	28
4.2.4	Zeitreihe.....	29
4.3	Das Datacenter.....	30
4.3.1	Erste Säule: Mitglieder.....	30
4.3.2	Zweite Säule: Web-Statistik.....	34
4.3.3	Dritte Säule: Angebote.....	39
4.3.4	Vierte Säule: Inkasso.....	42
4.3.5	Fünfte Säule: Fundraising.....	44
4.3.6	Sechste Säule: Personenkategorien.....	47
4.4	Data-Mining auf Kundendaten.....	49
5.	Resultate.....	51
6.	Ausblick.....	56
7.	Anhang.....	60
7.1	Abbildungsverzeichnis.....	60
7.2	Literaturverzeichnis.....	61
7.3	Inhalt der CD.....	62

1. Einleitung

1.1 Alumniorganisationen

In dieser Arbeit soll ein neues Modul für eine bestehende verteilte Anwendung implementiert werden. Die Applikation „Alumntools“ wird von Alumni-Vereinen zur Bewirtschaftung ihrer Daten und zur Organisation der Vereinstätigkeiten genutzt. Das neue Modul soll den Vereinen die Möglichkeit bieten, Auswertungen auf ihren Datenbeständen und Logfiles durchzuführen, um daraus bestimmte Sachverhalte zu visualisieren. Bei der Implementation sollen nebst statistischen Methoden auch Data-Mining Algorithmen zur Anwendung kommen.

Die Bedeutung des Wortes „Alumni“ hat sich seit seiner Entstehung immer wieder gewandelt. Ursprünglich stammt es vom lateinischen Verb ‚alere‘, was soviel bedeutet wie ‚ernähren‘. Damals wurden verletzte oder ausgediente Soldaten, die vom römischen Reich kostenlos ernährt wurden, als „Alumni“ bezeichnet. Später stand der Begriff für Schüler aus sog. Alumnaten. Alumnate sind aus mittelalterlichen Klosterschulen hervorgegangene Erziehungs- und Bildungseinrichtungen, die für die Unterkunft, die Erziehung, den Unterricht und die Kost ihrer Zöglinge aufkamen. Später wurden in Universitäten jene Studenten als „Alumni“ bezeichnet, die einen Freitisch erhielten, was soviel bedeutet, dass sie nicht selbst für ihr Essen aufzukommen hatten. [Wikipedia, 2007]

Heutzutage werden Ehemaligen-Vereine als Alumniorganisationen bezeichnet. Diese Vereine haben in den letzten Jahren immer mehr an Bedeutung gewonnen. Diesen Umstand verdanken sie einerseits der immer häufigeren Betonung der Notwendigkeit persönlicher Netzwerke, andererseits haben technologische Errungenschaften wie das Internet und die damit verbundenen Technologien effiziente Netzwerke erst möglich gemacht.

Nicht nur Bildungsinstitute, auch grössere Unternehmen folgen diesem Trend und institutionalisieren solche Ehemaligen-Vereine. Je nach Art der Auftraggeber bieten diese Vereine ihren Mitgliedern unterschiedliche Dienstleistungen an. Sind es vor allem Netzwerk-Effekte, die Bildungs-Institute ihren Abgängern für ihre berufliche Laufbahn ermöglichen wollen, dienen sie wirtschaftlichen Betrieben mehr dem Erhalt der persönlichen Kontakte oder der Organisation von

1 Einleitung

gemeinsamen Anlässen, wie Gruppenreisen oder Vorträgen. Daneben existieren auch unabhängige Netzwerke wie zum Beispiel xing.com (ehemals „OpenBC“), denen sich jeder anschliessen kann, der von einem bestehenden Mitglied dazu eingeladen wird. Diese versprechen nicht nur eine vereinfachte Suche nach Informationen und Fähigkeiten, sondern auch nach neuen Mitarbeitern oder neuen Kunden.

In der vorliegenden Arbeit stehen Alumni-Vereinigungen von Bildungs-Instituten im Zentrum der Betrachtung. Diese sollen Beziehungen nicht nur zwischen Ehemaligen, sondern auch zwischen Studenten und Ehemaligen ermöglichen. Im ersten Fall geht es darum, Kontakte zu knüpfen und aufrechtzuerhalten, um auf diese Weise von Informationen und Fähigkeiten anderer profitieren zu können. Im zweiten Fall geht es mehr um Informationen betreffend der Studiumswahl und später um eine Hilfestellung für den Einstieg in die Berufswelt. Von einem grossen Netzwerk profitieren beide Parteien gleichermaßen. Der Wert eines solchen Netzwerks für den Einzelnen steigt überproportional mit der Erhöhung der Zahl seiner Mitglieder, was als „Netzwerkeffekte“ bezeichnet wird. Mit steigendem Wert des Netzwerks steigt auch der Anreiz für Aussenstehende, diesem Netzwerk beizutreten, was wiederum den Wert für die bestehenden Mitglieder steigert und so weiter. Entsprechend ist es die Aufgabe des jeweiligen Vereins, das Angebot und die Dienstleistungen so zu gestalten, dass möglichst viele Ehemalige sich nicht nur für einen Eintritt entscheiden, sondern auch möglichst lange im Kreis der Mitglieder bleiben.

Diese Vereine existieren teilweise bereits seit einigen Jahren. Sie verfügen daher schon oftmals über Datenbanken und setzen für gewisse Vereins-Funktionen bereits bestimmte Applikationen ein. Jedoch verhindern die häufig sehr heterogenen Applikations-Umgebungen und veraltete Datenbank-Technologien eine effiziente Bewirtschaftung der Daten und somit eine angemessene Pflege des Netzwerks und dessen Mitglieder.

Eine Applikation, die speziell für diese Art der Vereins-Organisation entwickelt wurde, um die ebengenannte Heterogenität im Anwendungsbereich zu verhindern, ist die von der IM Strategies AG entwickelte Software „Alumnitools“.

1 Einleitung

1.2 Das betreuende Unternehmen

Die Information Management Strategies (IMS) AG hat ihren Sitz in Wädenswil am Zürchersee und zählt 6 Mitarbeiter. Sie bietet Verwaltungen, Grossfirmen und KMU's massgeschneiderte Dienstleistungen rund um Informationsmanagement und Informationssysteme an. Nebst Beratung in Organisation, Kommunikation und Technologie beinhalten diese Dienstleistungen auch das Erstellen von Analysen und Lösungs-Konzepten für eine reibungsfreie Integration des Informationsmanagements in die bestehenden Geschäftsprozesse. Zu ihren Kunden zählen nebst Verwaltungen und Grossfirmen vor allem auch Hochschulen. Diesen Umstand verdankt die Firma der Anwendung „Alumnitools“, eine In-House entwickelte Lösung zur Unterstützung der operativen Tätigkeit von Alumni-Vereinen. [IMS, 2007]

1.3 Alumnitools

Im Auftrag einer der grössten Hochschulen des Landes entwickelte die IMS AG 2002 eine erste Version der Vereinssoftware. Sie sollte zum einen die Administration des Alumnivereins in der Kunden- und Datenpflege unterstützen, und zum anderen den Mitgliedern eine gemeinsame Plattform bieten zwecks Aufrechterhaltung oder Schaffung von Kontakten. Im Laufe der Zeit wurden zahlreiche Anpassungen und Veränderungen vorgenommen und neue Vereine haben sich dem Kundenkreis angeschlossen. Mittlerweile liegt die Applikation in einer fünften Version vor und wird von 4 unterschiedlichen (Dach-) Organisationen angewendet, welche in Kapitel 3 genauer beschrieben werden.

Die IMS AG agiert als Application Service Provider (ASP). Wie für einen ASP üblich stellt sie ihren Kunden die Applikation, „Alumnitools“, über das Internet zur Verfügung und ist dabei verantwortlich nicht nur für dessen Betrieb und Wartung, sondern auch für die benötigte Infrastruktur. Wie häufig bei verteilten Anwendungen werden die Aufgaben Präsentation, Logik und Datenhaltung voneinander getrennt. Während die Präsentation von einem Web-Browser übernommen wird, sind die Applikation und die Daten auf zwei Servern verteilt. Kundenseitig ist lediglich ein Anschluss ans Internet nötig, um die Applikation nutzen zu können. Jeder Kunde greift dabei auf eine eigene Daten zu. Bei einer

1 Einleitung

Login-Maske muss der Benutzer daher neben seinen Login-Daten auch eine Datenbank angeben, auf welcher er sich einloggen will. Alumnitools besteht aus zwei Teilen:

Der erste Teil ist die sog. „**Innensicht**“, die der Vereins-Administration zur Erledigung der täglichen Aufgaben zur Verfügung steht. Kernelemente der Innensicht sind die „Einfache Suche“ und die „Erweiterte Suche“, womit gezielt bestimmte Mitglieder gefunden werden können, um auf Basis dieser Suchresultate zum Beispiel einen Massenversand durchzuführen oder Rechnungen zu generieren. Daneben können neue Angebote in Form von Mitgliedschaften, Publikationen, Veranstaltungen oder Dienstleistungen erfasst werden, die den Mitgliedern auf dem Online-Portal bereitgestellt werden und die in einem Inkasso-Modul abgerechnet werden können. Nebst der Bewirtschaftung von bestimmten Referenz-Tabellen erlaubt die Innensicht auch die Verarbeitung von VESR-Files, die von Banken zwecks Bestätigung von Zahlungseingängen übermittelt werden.

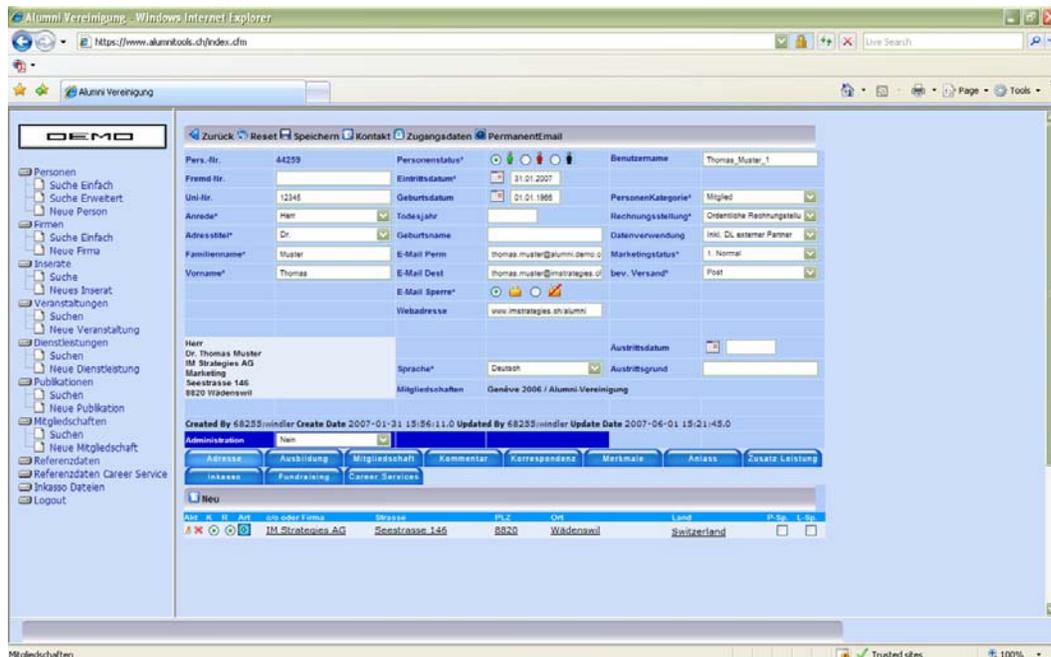


Abb. 1: Alumnitools, die “Innensicht”

1 Einleitung

Der zweite Teil wird als „**Aussensicht**“ bezeichnet, da sie allen Mitgliedern über das Internet frei zugänglich ist. Sie ermöglicht dem User die Pflege seiner persönlichen Daten und die Festlegung der Sichtbarkeitsbereiche jener Daten. Diese beinhalten nebst persönlichen Informationen zur Person und Wohnort auch Informationen über Ausbildungen oder persönliche Interessen. Im sogenannten „who-is-who“ kann gezielt nach anderen Mitgliedern gesucht werden, um mit diesen in Kontakt zu treten. Ist der Kontakt einmal aufgenommen, können persönliche Kontaktlisten geführt werden, wo andere Mitglieder, ihr Einverständnis vorausgesetzt, aufgenommen werden können. Das Register „DL-Nutzung“ informiert über aktuelle Veranstaltungen und angebotene Dienstleistungen, die über die Online-Plattform gebucht werden können.

Bei der Entwicklung dieses Programmteils wurde darauf geachtet, dass der Code ausschliesslich in HTML geschrieben wurde. Durch den Verzicht auf zusätzliche Technologien wie zum Beispiel JavaScript oder Java wird sichergestellt, dass ohne Voreinstellungen ein Zugang Plattform- und Browserunabhängig möglich ist.



Personen-Nummer	40281	E-Mail permanent	<input type="checkbox"/>
Anrede	Herr	E-Mail destination	<input type="text"/>
Kurztitel	-	Im Mitgliederverzeichnis publizieren	<input checked="" type="radio"/> Ja <input type="radio"/> Nein
Vorname	Thomas	Bevorzugter Versandweg	<input type="text" value="Unbekannt"/>
Name	Muster	Datenverwendung	<input type="text" value="Inkl. DL externer Partner"/>
Geburtsname			
Geburtsdatum	(wird nicht publiziert)		
Webadresse	<input type="text" value="www.imstrategies.ch"/>		

Abb. 2: Alumnitools, die “Aussensicht”

Ein zentrales Anliegen bei der Entwicklung von Alumnitools war ein modularer Aufbau. So können Zugriffsberechtigungen nicht nur für ganze Programmteile, sondern auch für einzelne Funktionen über die Datenbank gesteuert werden. Auf

1 Einleitung

diese Weise kann der Funktionsumfang für jeden Kunden individuell bestimmt und auf seine Vereinstätigkeit angepasst werden. Auf Basis des gewählten Funktionsumfangs wird anschliessend der Preis, den der Verein für die Nutzung von Alumnitools zu bezahlen hat, festgelegt. Diese Aufteilung von Zugriffsberechtigungen geschieht nicht nur auf Mandanten-Ebene, sondern auch innerhalb des Vereins. So wird jedem Mitarbeiter eine sogenannte Rolle vergeben. Für jede Rolle sind in der Datenbank unterschiedliche Funktionen freigegeben und unterschiedliche Berechtigungen erfasst. Wird ein neuer Kunde in den Betrieb von Alumnitools aufgenommen, gehört diese Zuteilung der Berechtigungen nebst der Migration der Daten zu den Hauptaufgaben der IMS AG.

Ein anderer wichtiger Punkt, auf den bei der Entwicklung von Alumnitools geachtet wurde, ist die Mandantenfähigkeit. Unter Mandanten werden in dieser Arbeit die Kunden der IMS AG verstanden, also die Alumni-Vereinigungen. Auf diese und das geschäftliche Umfeld wird in Kapitel 3 näher eingegangen. Mandantenfähigkeit bedeutet, dass der Programmier-Aufwand für die Anpassung der Applikation an einen neu hinzugekommenen Mandanten minimal sein sollte. Alle Mandanten greifen auf dieselben Applikations-Files zu. Dabei wird nicht nur, wie oben erwähnt, der für jeden Mandanten unterschiedliche Funktionsumfang durch die Datenbank gesteuert. Auch graphische Elemente werden unterschiedlich angezeigt, indem beispielsweise der Pfad für die Logo-Bilddatei in der Datenbank hinterlegt ist.

1.4 Aufbau der Diplomarbeit

In **Kapitel 2** werden die Anforderungen, die an das neu zu implementierende Modul gestellt werden, beschrieben.

Kapitel 3 beschreibt das Umfeld der Entwicklung. Zu Beginn werden die Kundenbeziehungen der IMS AG erklärt, bevor genauer auf die einzelnen Mandanten eingegangen wird. Im Anschluss daran wird das technische Umfeld beschrieben, in welchem die hier vorliegende Arbeit erstellt wurde.

Kapitel 4 beschreibt die Implementierung. Einleitend werden die wichtigsten Design-Entscheidungen begründet, bevor ein Überblick über die grobe Funktionsweise des neuen Moduls gegeben wird. Das Ende des Kapitels beschäftigt sich mit dem Vorhaben, mittels Data-Mining Algorithmen eine Kundensegmentierung vorzunehmen.

In **Kapitel 5** werden einige Auswertungen aus den unterschiedlichen Registern besprochen.

Kapitel 6 bietet einen Ausblick auf geplante und gedachte Erweiterungen des neuen Moduls.

2. Anforderungen

In dem folgenden Kapitel werden die einzelnen Anforderungen, welche an das hier besprochene Modul gestellt werden, beschrieben. Zum einen sollen die neuen Funktionen zu einem Mehrwert für die Mandanten der IMS AG führen, andererseits gibt es bestimmte Rahmenbedingungen durch die Applikation, in welche das Modul integriert wird. Auf beide wird in den nächsten Zeilen eingegangen.

Es sollte ein neues Modul für die Innensicht geschaffen werden, welches unterschiedliche Auswertungen zulässt und diese in einer adäquaten und ansprechenden Form visualisiert. Dabei sollte die Aufbereitung der erforderlichen Daten möglichst schnell geschehen ohne dass zuviel Rechenleistung dafür benötigt würde. Dies aus dem einfachen Grund, da allzu rechenintensive Datenbank-Abfragen, die während des täglichen Betriebs gemacht werden, die Performance der Server stark beeinträchtigen und die Arbeit mit der Applikation behindern.

Alumni-Vereinigungen haben Interesse daran, die Beschaffenheit ihrer Mitglieder zu kennen und zu wissen, inwiefern diese die angebotenen Dienstleistungen nutzen. Mit diesen Kenntnissen könnten zum Beispiel gezieltere Aussagen gemacht werden über den Erfolg von bestimmten Veranstaltungen, Spenden-Kampagnen oder über das Verhalten der User im Online-Bereich. Zentrales Ziel dieser Arbeit ist die Unterstützung der Ehemaligen-Vereine in ihrem Bestreben, mehr über ihre Mitglieder zu erfahren. Zwar ermöglicht die „Erweiterte Suche“ in der „Innensicht“ von Alumnitools bereits sehr rudimentäre Datenbank-Abfragen, wie zum Beispiel nach allen Neuabsolventen, die männlichen Geschlechts sind. Die Aussagekraft dieser Methode ist jedoch zweifelhaft, da lediglich eine Zahl über die Existenz dieser Personen informiert und andere Aspekte nicht darstellbar sind. Um die Alumni-Vereine optimal zu unterstützen, wurden sie in den Planungs-Prozess einbezogen und nach Auswertungsarten befragt, die für sie relevant sind.

2 Anforderungen

Ein wichtiges Kriterium war die Benutzerfreundlichkeit. Das zu implementierende Modul sollte leicht verständlich und durch jedermann nutzbar sein. Es sollten keine Statistik- oder gar Datamining-Kenntnisse erforderlich sein, um den Informationsgehalt der Auswertungen korrekt deuten zu können.

Zusätzlich zu den eben genannten Eigenschaften sollte auch in diesem Modul die Mandantenfähigkeit aufrechterhalten bleiben. Der Kreis der Kunden vergrößert sich stetig, daher war es ein grosses Anliegen bei der Entwicklung von Alumnitools, dass die Inbetriebnahme einer neuen Datenbank einen minimalen, zusätzlichen Programmier-Aufwand erfordere. Dasselbe gilt auch für Änderungen der Datenbank-Inhalte. Die Auswertungen sollten dynamisch generiert werden, ohne dass in diesem Fall eine Adaption des Codes nötig ist.

Das neue Modul wird für die „Innensicht“ entwickelt und beschränkt sich daher nicht, wie die „Aussensicht“, auf HTML. Die Wahl der Technologie konnte allein durch ihre Erfüllung des Zwecks gerechtfertigt werden, was eine breite Suche nach adäquaten Mitteln zur Visualisierung zuliess. Ausserdem wurde bei der Entwicklung darauf geachtet, dass das sog. „Look&Feel“ der übrigen in der Innensicht eingebundenen Module beibehalten wurde. Es wurde auf diesselben Design-Elemente und -Strukturen zurückgegriffen, die auch in der übrigen Applikation zur Anwendung kommen.

3. Voraussetzungen

In diesem Kapitel werden die Rahmenbedingungen erklärt, unter welchen das hier beschriebene Vorhaben realisiert wurde. Zu Beginn werden die Kunden der IMS AG und damit die eigentlichen User von Alumnitools beschrieben. Diese Informationen sind unverzichtbar für die korrekte Interpretation der Auswertungen, die in späteren Kapiteln gezeigt werden. Im Anschluss daran wird noch das technische Umfeld, in welchem das neue Modul entwickelt wurde und in welches es später integriert wird, beschrieben.

3.1 Die Mandanten

Um Verwirrung zu vermeiden, sollen an dieser Stelle die Geschäftsbeziehungen zwischen der IMS AG und ihren Kunden näher erläutert werden. Die einzelnen Alumni-Vereine werden im Folgenden als Mandanten bezeichnet und sind direkte Kunden der IMS und Nutzer der Software Alumnitools. Häufig handelt es sich um Dachorganisationen, die von mehreren Ehemaligen-Vereinen getragen werden zwecks Erledigung der vereinsübergreifenden Aufgaben. Dazu gehört u.a. die Administration der einzelnen Mitglieder und damit auch die Auswahl einer unterstützenden Software-Lösung.

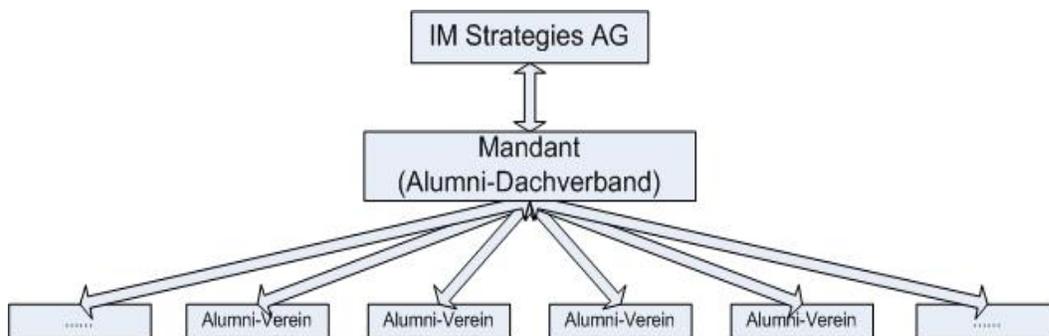


Abb. 3: Kundenbeziehung IM Strategies AG

Wie einleitend erwähnt arbeiten vier verschiedene Mandanten mit Alumnitools. Sie unterscheiden sich untereinander sehr und haben lediglich eines gemein: ihr Vereinszweck ist die Organisation eines Alumninetzwerks für Ehemalige einer

3 Voraussetzungen

bestimmten Bildungseinrichtung. Ansonsten sind sie in ihrer Vereinstätigkeit und ihrem Vereinsgebühren sehr verschieden. Dies zeigt sich nicht nur im Umfang der angebotenen Dienstleistungen, sondern zum Beispiel auch im Aufnahme-Prozess neuer Mitglieder. Da diese und ähnliche Informationen wichtig sind für eine korrekte Interpretation der Auswertungen der vorliegenden Arbeit, werden diese Unterschiede in den folgenden Abschnitten erklärt.

Aus Gründen des Datenschutzes werden die Mandanten anonymisiert und lediglich in ihren Eigenschaften beschrieben, was den Informationsgehalt dieser Arbeit nicht beeinträchtigt. An dieser Stelle soll erwähnt sein, dass es den einzelnen Mandanten lediglich erlaubt sein wird, Auswertungen innerhalb ihrer eigenen Datenbank vorzunehmen, und nicht, wie in dieser Arbeit beschrieben, zwischen den Datenbanken zu wählen. Dies wird bei Inbetriebnahme des hier beschriebenen Moduls mit den in der Applikation üblichen Zugriffsrestriktions-Mechanismen eingeschränkt werden und ist in diesem Umfang nur den Mitarbeitern der IMS AG und bestimmten Personen der Universität Zürich vorbehalten.

Die Mandanten werden aufgrund der Zahl ihrer aktiven Mitglieder geordnet, wobei Mandant 1 die meisten Mitglieder zählt und gleichzeitig den grösstmöglichen Funktionsumfang von Alumnitools aktiviert hat.

3.1.1 Mandant 1

Beim Mandanten 1 handelt es sich um eine technisch naturwissenschaftliche Universität. Insgesamt arbeiten und studieren 18'000 Personen an der Hochschule. Entsprechend gross ist der Wert, der der Ehemaligen-Pflege beigemessen wird. Die Alumni-Organisation, die aus 7 Angestellten besteht und eine Arbeit von 400 Stellenprozent abdeckt, ist mit dieser Aufgabe betraut. Sie ist eine Dachorganisation von rund 35 Ehemaligen-Vereinen, die einzelne Studienrichtungen vertreten.

Normalerweise wird zwischen Mitgliedern, Nichtmitgliedern und Neuabsolventen unterschieden. Bei diesem Mandanten aber wird noch weiter differenziert. Es gibt

3 Voraussetzungen

unterschiedliche Klassen von Mitgliedern, die von der Höhe des Betrags abhängen, den der jeweilige Verein an die Grundleistungen der Geschäftsstelle bezahlt. Zum Beispiel gibt es noch die Nicht-Alumni und daneben noch Kategorien für Firmen, die ebenfalls in der Datenbank gehalten werden, jedoch aufgrund der kleinen Zahl und der damit geringen wirtschaftlichen Relevanz nicht in den Auswertungen berücksichtigt werden.

Personenkategorie	Beschreibung
Neuabsolventen	Universitätsabgänger, die ein Jahr lang eine kostenlose Mitgliedschaft erhalten
Mitglieder	Mitgliedschaft bei einem oder mehreren Vereinen zu einem festen Jahresbeitrag
Nicht-Mitglieder	Absolventen, die keine aktiven Mitglieder sind
Studierende	Im Rahmen eines Mentoring-Moduls können bereits vor dem Abschluss Kontakte geknüpft werden
Mittelschüler	Im Rahmen eines Mentoring-Moduls können bereits vor dem Studium Kontakte geknüpft werden
Partner-natürliche Personen	Mitarbeiter der IMS AG; andere Personen, die in keine der obigen Kategorien fallen

Abb. 4: Beschreibungen der Personenkategorien, Mandant 1

3 Voraussetzungen

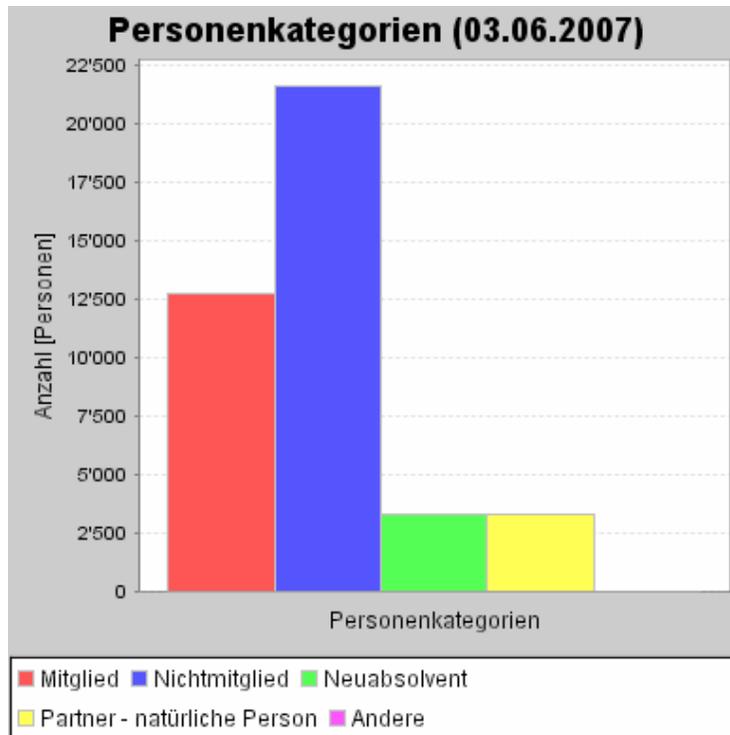


Abb. 5: Verhältnis der Personenkategorien, Mandant 1

Periodisch werden die Daten der Neuabsolventen des Bildungsinstituts importiert, denen im Anschluss an ihr Studium ein kostenloses Jahr in der Alumni-Organisation angeboten wird. Sollten sie sich nach diesem Jahr für eine weiterführende Mitgliedschaft entscheiden, muss ein Jahresbeitrag bezahlt werden, der je nach Verein variiert. Entscheiden sie sich gegen eine solche, verbleiben sie zwar in der Datenbank, dies jedoch unter der Kategorie ‚Nicht-Mitglied‘. Sie profitieren somit nicht von den Privilegien des Vereins, bleiben aber für die Mitglieder auffindbar und werden von Zeit zu Zeit durch Werbung an die Vorteile einer Mitgliedschaft erinnert.

Insgesamt zählt die Datenbank 42'282 aktive Einträge, davon sind, wie in Abb. 5 ersichtlich, 12'778 Mitglieder. Das Verhältnis zwischen den Mitgliedern und den übrigen aktiven Einträgen ist für die Alumni-Vereine von grossem Interesse, da es den Anreiz für Nicht-Mitglieder widerspiegelt, dem Verein beizutreten. Aus diesem Grund wurde in dieser Arbeit grosses Gewicht auf die Zusammensetzung und die Veränderungen der jeweiligen Personenkategorien gelegt, jedoch dazu später mehr.

3 Voraussetzungen

3.1.2 Mandant 2

Hierbei handelt es sich um eine weitere Dach-Organisation von einzelnen Alumni-Vereinen, die im Auftrag einer Universität die Ehemaligen-Pflege übernehmen. Der Datenbestand besteht vorerst noch aus der rechtswissenschaftlichen und der medizinischen Fakultät, wird aber in naher Zukunft um weitere Fakultäten erweitert werden, die nachträglich dem Dachverband beigetreten sind. An dieser Universität waren im Wintersemester 06/07 23'910 Studenten immatrikuliert.

Für die Bewirtschaftung der Daten und der Tagesgeschäfte arbeiten vier Personen in der Geschäftsstelle zu 200 Stellenprozent. Neuabsolventen werden nicht automatisch aufgenommen und das erste Jahr kostenlos als Mitglieder akzeptiert, sondern haben die Wahl zwischen einer Mitgliedschaft mit einem Jahresbeitrag von 40sfr oder einer lebenslangen Mitgliedschaft für 800sfr.

Personenkategorie	Beschreibung
Probemitglied	Neuabsolventen im kostenlosen ersten Jahr
Mitglieder	Absolventen, die eine Mitgliedschaft eingegangen sind
Nicht-Mitglieder	Absolventen, die keine aktiven Mitglieder sind
Kollektivmitglieder	Firmen, die als Mitglieder auftreten
Gönner	(wird nicht verwendet)
Juristische Fakultät	Ausschliesslich Absolventen der juristischen Fakultät, die eine Mitgliedschaft eingegangen sind

Abb. 6: Beschreibungen der Personenkategorien, Mandant 2

3 Voraussetzungen

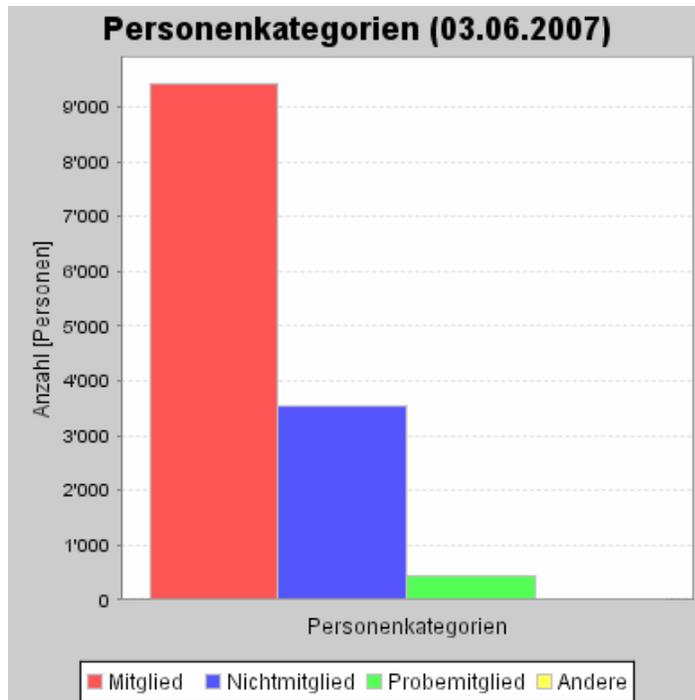


Abb. 7: Verhältnis der Personenkategorien, Mandant 2

3.1.3 Mandant 3

Auch bei diesem Mandanten handelt es sich um eine Dachorganisation von Ehemaligenvereinen, die im Auftrag einer Hochschule die Ehemaligen-Pflege übernimmt. Die Hochschule besteht aus 7 Fakultäten (Medizinische, Wirtschaftswissenschaftliche, Juristische, Philosophisch-historische, Philosophisch-naturwissenschaftliche, Theologische und Psychologische Fakultät) mit insgesamt über 70 Instituten. Die Hochschule existiert bereits über 500 Jahre und gegen 10000 Studenten sind derzeit immatrikuliert. Letztes Jahr haben 2636 Studenten ihren Abschluss gemacht. Der Alumni-Verein existiert erst seit 2 Jahren.

In der Geschäftsstelle der Alumni-Organisation arbeiten 2 Personen ganztags. Hier werden die Neuabsolventen nicht automatisch in den Datenbestand aufgenommen, sondern lediglich auf die Existenz und die Möglichkeiten des Alumni-Vereins aufmerksam gemacht. Das erste Jahr ist kostenlos. Den Mitgliedern wird nebst anderen Dienstleistungen eine webbasierte Alumniplattform zur Verfügung gestellt, wo sie neben der Pflege der persönlichen Daten auch mit anderen Mitgliedern in Kontakt treten können.

3 Voraussetzungen

Personenkategorie	Beschreibung
Neuabsolventen	Universitätsabgänger, die ein Jahr lang eine kostenlose Mitgliedschaft erhalten
Mitglieder	Mitgliedschaft bei einem oder mehreren Vereinen zu einem festen Jahresbeitrag
Nicht-Mitglieder	Absolventen, die keine aktiven Mitglieder sind

Abb. 8: Beschreibungen der Personenkategorien, Mandant 3

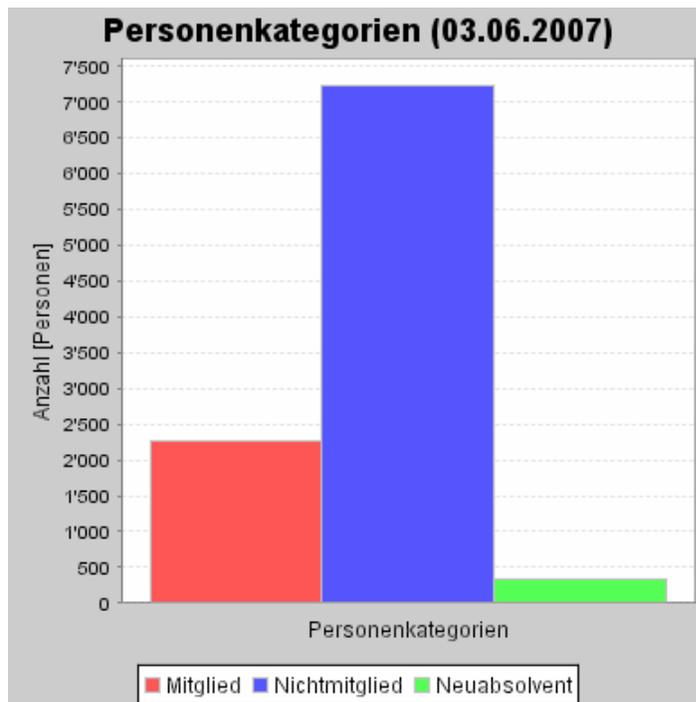


Abb. 9: Verhältnis der Personenkategorien, Mandant 3

3 Voraussetzungen

3.1.4 Mandant 4

Der vierte Mandant bietet Diplom-Kurse für Unternehmensführung an. Die Institution existiert bereits seit über 50 Jahren und hat gegen 4000 Absolventen hervorgebracht. In der Geschäftsstelle arbeitet eine Person zu 50% um die Aufgaben der Datenbewirtschaftung und der Vereinsorganisation zu erfüllen. Neuabsolventen werden auch hier nicht automatisch in die Datenbank aufgenommen, sondern erst bei der Zahlung eines Jahresbeitrags über 60sfr oder einer Dauermitgliedschaft über 600sfr als Mitglied anerkannt. Nicht-Mitglieder werden nicht in der Datenbank gehalten, daher existiert nur eine Personenkategorie, die ‚Mitglieder‘. Das Angebot des Vereins beinhaltet den Zugriff auf die Ehemaligen-Datenbank, vergünstigte Weiterbildungsprogramme, gemeinsame Veranstaltungen zum Zwecke des Networking und Vergünstigungen bei Drittangeboten.

Personenkategorie	Beschreibung
Mitglieder	Mitgliedschaft zu einem festen Jahresbeitrag

Abb. 10: Beschreibungen der Personenkategorien, Mandant 4

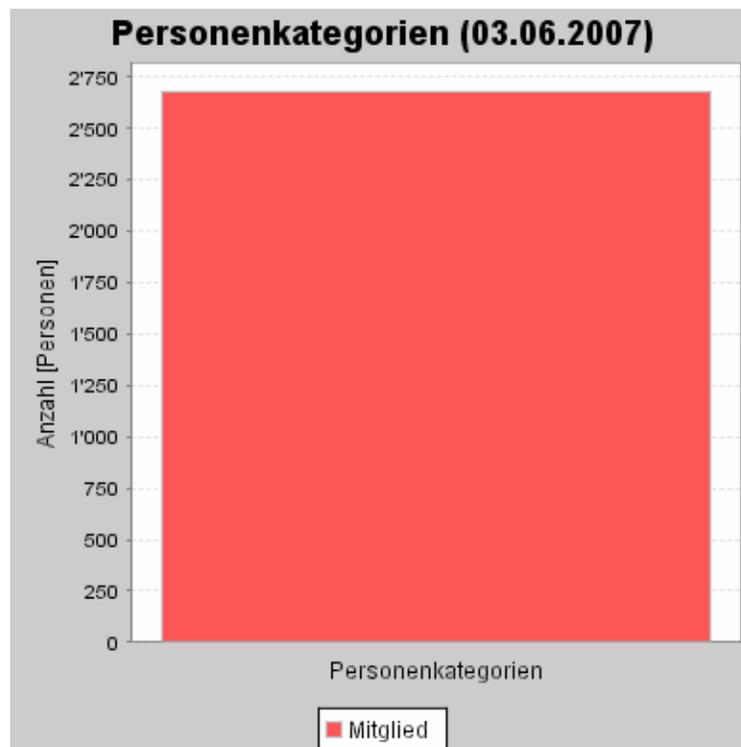


Abb. 11: Verhältnis der Personenkategorien, Mandant 4

3.2 Technisches Umfeld

Zur Entwicklung des neuen Moduls konnte die von der IMS AG angebotene Infrastruktur genutzt werden. Zur Verfügung standen drei verschiedene Server, auf welche per Remote-Desktop zugegriffen werden konnte, was ein ortsunabhängiges Arbeiten ermöglichte. Wie anfangs erwähnt liegen die Anwendungs-Logik und die Daten von Alumnitools getrennt voneinander auf je einem Applikations- bzw. Datenbankserver. Auf beiden Servern ist Windows Server 2003 installiert. Die beiden Live-Server werden mit je einem Intel Pentium 4 Prozessor mit 3GHz Taktleistung und 2MegaByte RAM betrieben. Der dritte Server ist ein Test-Server, auf welchem eine Kopie der Applikation und der Datenbanken gespeichert ist. Auch dieser dritte Server läuft unter Windows Server 2003 und dient mit einer Testumgebung der Weiterentwicklung von Alumnitools. So kann die Applikation weiterentwickelt werden, ohne einerseits die Performance für die Mandanten zu beeinträchtigen, und ohne andererseits den Betrieb durch fehlerhafte Dateien zu gefährden. Der Testserver besitzt eine 2.8GHz Intel Celeron Prozessor und 512MegaByte RAM.

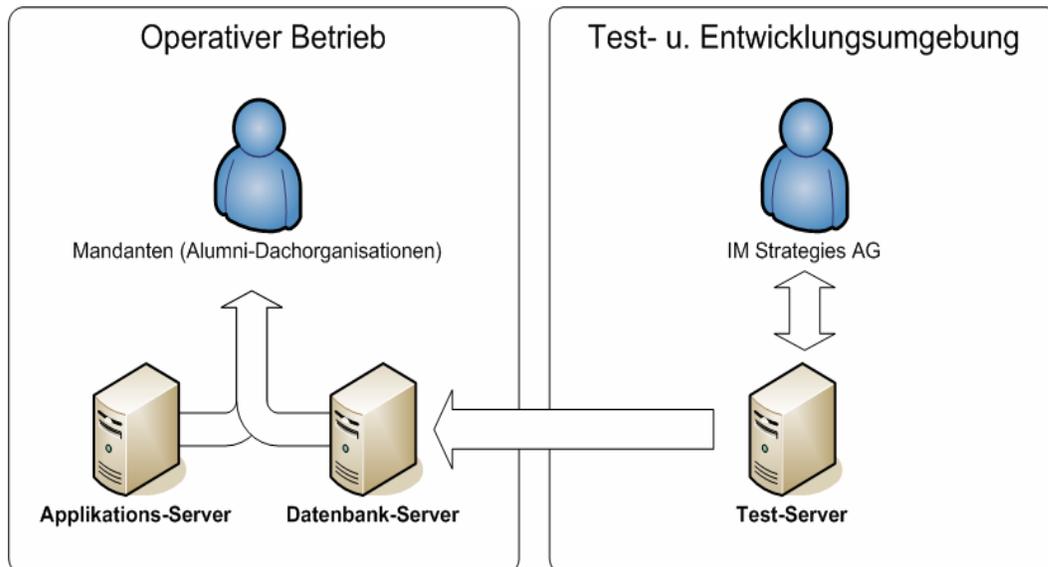


Abb. 12: Technisches Umfeld IM Strategies AG

4. Implementierung

Im folgenden Kapitel wird nun das neugeschaffene Modul, das „Datacenter“, beschrieben. Zuerst werden die wichtigsten Design-Entscheidungen, die für das Datacenter getroffen wurden, begründet. Anschliessend folgt ein grober Überblick über die Funktionsweise der Datenaufbereitung und –visualisierung, bevor näher auf die einzelnen Register eingegangen wird. Am Ende des Kapitels wird noch der Einsatz von Data-Mining Algorithmen diskutiert.

4.1 Einleitung

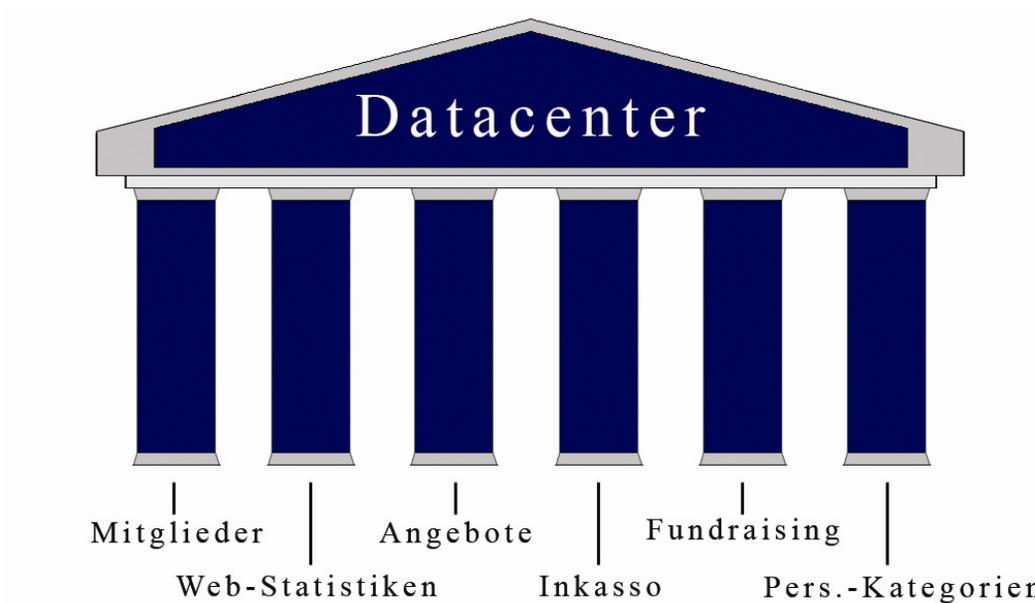


Abb. 13: Die sechs Säulen des Datacenter

Es sollte ein Modul für Alumnitools entwickelt werden, um den Vereins-Administrationen eine bessere Einsicht in ihre Datenbestände zu ermöglichen. Mit Hilfe von Statistiken sollen Informationen über die Beschaffenheit der Vereinsmitglieder abrufbar sein und gleichzeitig graphisch dargestellt werden können. Die Auswahl der Auswertungsarten wurde einerseits in Zusammenarbeit mit der IMS AG getroffen, andererseits wurden direkte Kundenwünsche der

4 Implementierung

Mandanten realisiert. Die einzelnen Register des Datacenters werden im Folgenden als Säulen bezeichnet, da sie sinnbildlich betrachtet alle gleichsam das hier angestrebte Vorhaben stützen. Das neue Modul, das sog. „Datacenter“, besteht aus sechs verschiedenen Säulen, die unterschiedliche Auswertungen zulassen. Die Erste erlaubt Auswertungen auf persönlichen und vereinsinternen Eigenschaften wie Alter, Geschlecht, Personenkategorien oder Dauer der Mitgliedschaft. Die zweite Säule wertet log-Files aus, die über das Verhalten der Online-User Auskunft geben. Die dritte Säule untersucht die Angebot-Nutzer und ergänzend dazu können die bereits aus der ersten Säule bekannten Auswertungstypen als zweite Auswertungs-Dimension herangezogen werden. Auf direkten Kundenwunsch wurden die letzten drei Register implementiert. Diese beinhalten Auswertungs-Möglichkeiten über Zahlungseingänge, seien das Zahlungen für genutzte Dienstleistungen, in Säule 4, oder Spenden, wie in Säule 5. Die sechste und letzte Säule widmet sich der zeitlichen Veränderung der verschiedenen Personenkategorien.

Im Interesse des besseren Verständnis wird zuerst eine grobe Übersicht über die Funktionsweise dieser Auswertungsmethoden gegeben, bevor anschliessend die einzelnen Elemente detaillierter betrachtet werden. Einige Funktionen der einzelnen Säulen sind sich sehr ähnlich oder werden mehrfach benutzt. In diesen Fällen wird auf wiederholende Erklärungen verzichtet und lediglich auf die erneute Anwendung hingewiesen.

Es soll an dieser Stelle noch betont werden, dass in der vorliegenden Arbeit das Datacenter so beschrieben wird, wie es nur den Mitarbeitern der IMS AG und bestimmten Personen der Universität Zürich, die mit dieser Diplomarbeit betraut sind, zur Verfügung steht. Während hier jeweils eine Auswahl des betreffenden Mandanten beschrieben wird, können die Ehemaligen-Vereine aus Datenschutzgründen lediglich Auswertungen der eigenen Datenbank wählen. Dies wird mit entsprechenden Zugriffsrechten, bzw. -restriktionen realisiert, sobald das neue Modul in Betrieb genommen wird.

4 Implementierung

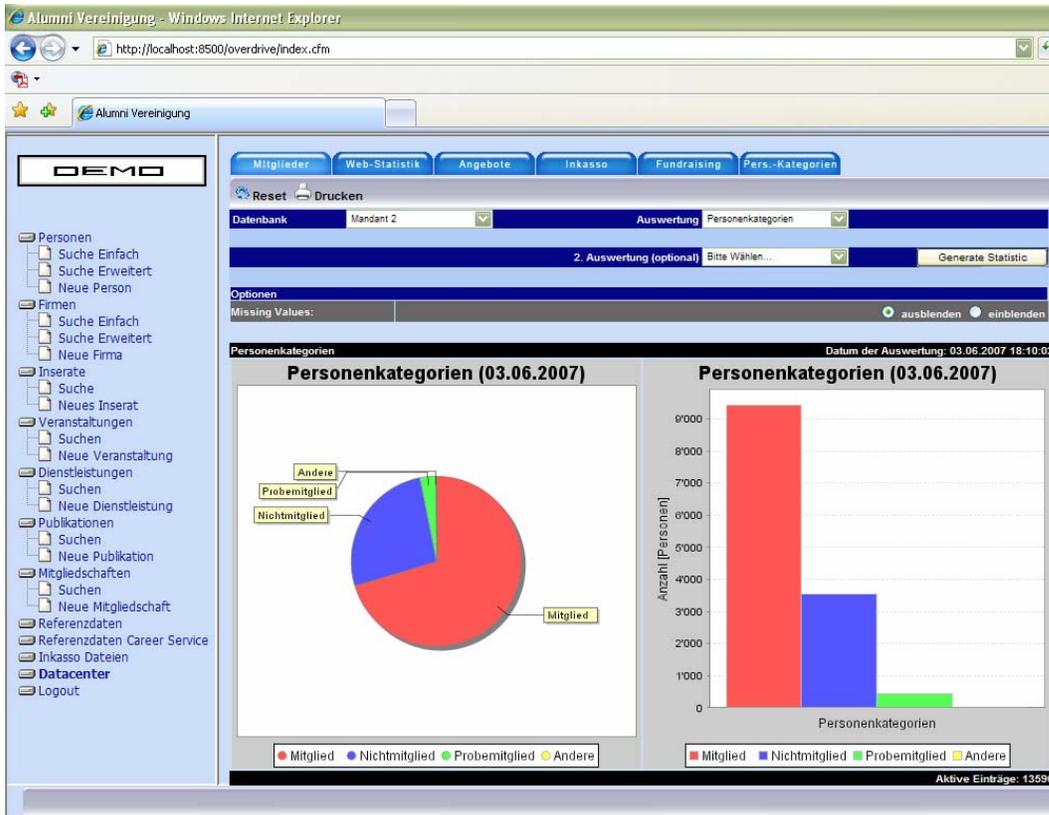


Abb. 14: Das Datacenter in Alumnitools eingebettet

Die Daten werden bei einem entsprechenden Aufruf nicht in real-time aufbereitet, da die ohnehin stark genutzten Datenbanken auf solch immense Abfragen empfindlich reagieren würden. Es wurde eine Lösung angestrebt, welche die Datensammlung und -visualisierung voneinander trennt. So könnten die zeitintensiven Abfragen in der Nacht durchgeführt werden, um so die Mandanten nicht in ihrer täglichen Arbeit zu behindern. Ein weiterer Vorteil ist der damit verbundene, geringere Aufwand für die graphische Aufbereitung der Daten. Dieser Prozess wird nun im Folgenden genauer beschrieben.

Sogenannte „Routinen“ sammeln die nötigen Informationen und speichern diese in Form von Listen in Text-Files, die von den darstellenden Elementen „ausgelesen“ werden (Abb. 15). In welchem Zyklus diese Routinen angestoßen werden um die benötigten Daten aufzubereiten, hängt von dem Bedarf des Kunden ab, wie aktuell die jeweiligen Auswertungen sein sollten. Da sich die darzustellenden Daten nicht in kurzen Abständen ändern, wäre ein täglicher oder sogar wöchentlicher Aufruf der datenaufbereitenden Routinen denkbar.

4 Implementierung

Am ganzen Vorgang beteiligt sind für jede Säule drei Komponenten. Ein erstes Coldfusion(CF)-File, die Routine, stellt in regelmässigen Abständen die Daten zusammen und speichert diese in Text-Files. Wird nun über das GUI eine bestimmte Auswertung angefordert, wird das entsprechende Text-File gelesen und die gesuchten Listen daraus extrahiert. Zur Visualisierung wird nun ein Java-Applet aufgerufen, welches aufgrund dieser Listen Statistik-Graphiken erzeugt. Zur Visualisierung wurde ein Open-Source Tool herangezogen. Es heisst JFreeChart und stellt Klassen zur Verfügung zur Visualisierung verschiedenster Diagramme.

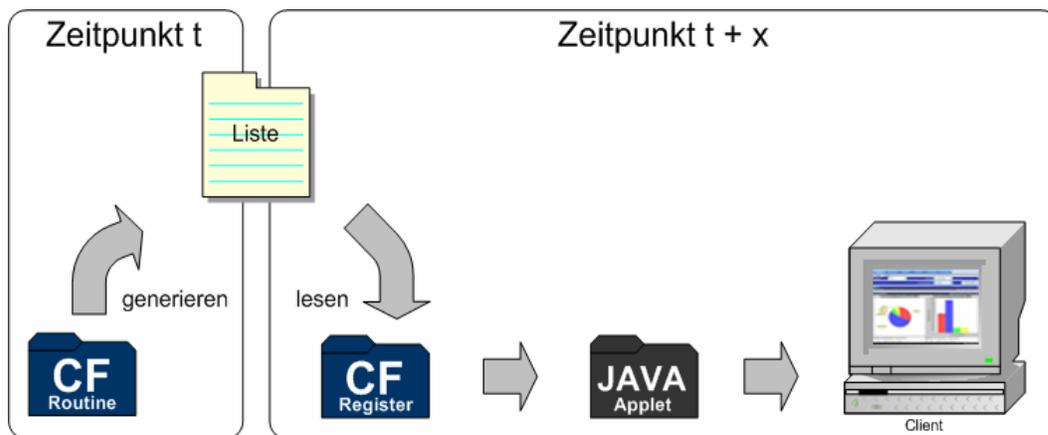


Abb. 15: Am Visualisierungsprozess beteiligte Komponenten

Das Design des Datacenters wurde an jenes der „Innensicht“ angepasst. Auch hier wurde die Idee der Modularität weiterverfolgt. Für jedes Register im Datacenter gibt es eine eigene Routine, um die erforderlichen Daten zu sammeln, ein spezielles Textfile, in welchem die Daten gespeichert werden und ein eigenes Coldfusion-File, welches das GUI für den User bereitstellt. Durch Zugriffsrestriktionen, die in der Datenbank jedes einzelnen Mandanten erfasst werden, können so nur jene Register zugänglich gemacht werden, an denen die Mandanten Interesse haben.

4.2 JFreeChart

Es handelt sich hierbei um ein Open-Source Tool, welches vollständig in Java entwickelt wurde und Entwicklern ermöglichen soll, unterschiedlichste Daten in geeigneter Form darzustellen [JFreeChart, 2007]. Der Source-Code unterliegt der GNU LGPL (Lesser General Public License), die eine uneingeschränkte Nutzung, Erweiterung und/oder Veränderung des Codes erlaubt. [LGPL, 2007]

Sehr geeignet für Alumnitools erschien die Möglichkeit, erzeugte Graphen in unterschiedlichste Formate zu exportieren, u.a. auch in das gängige Bildformat .png. Dieser Umstand wie auch die enorm breite Auswahl an optisch reizvollen Visualisierungstypen waren Hauptgrund für die Wahl dieses Tools. In der vorliegenden Arbeit wurden zur Visualisierung der Auswertungen lediglich vier Typen angewendet.

- Pie-Chart
- Bar-Chart
- Stacked Bar-Chart
- Zeitreihe

Zur Einbindung von JFreeChart in die hier besprochene Coldfusion-Applikation standen keine Mechanismen zur Verfügung, daher wurden einige Klassen zur Daten-Zusammenstellung und -darstellung verändert bzw. ergänzt.

Die grobe Vorgehensweise ist bei allen Graphen ähnlich und wird auf den folgenden Zeilen beschrieben. Für die Visualisierung der aus den Text-Files ausgelesenen Listen braucht es zwei Klassen. Eine erste, die aus den ausgelesenen Listen ein Datenobjekt konstruiert, welches als Input für eine zweite Klasse dient, die ihrerseits aus den Daten einen Graphen generiert. Beide Klassen wurden speziell zu diesem Zweck implementiert und existieren in dieser Form nicht im Applikations-Umfang von JFreeChart. Wird der Befehl zur Visualisierung einer Auswertung gegeben, wird die entsprechende Liste aus dem Text-File gelesen und in eine temporäre Liste gespeichert. Ein Loop durchläuft diese Liste und erzeugt bei jedem Durchgang durch den `cfoject`-Tag ein weiteres Element des Datenobjekts, welches anschliessend als Input für den Graphen dienen soll. Ist der Loop durchlaufen, wird die graphen-generierende Funktion aus der Klasse ‚ChartClient‘ mit eben generierten Datenobjekt als Input-Parameter aufgerufen. Diese Klasse wurde insofern verändert, dass sie nicht nur einen Graphen erstellt

4 Implementierung

und in eine Bild-Datei exportiert, sondern diese an einem vorgegebenen Ort abspeichert. Das Bild enthält im Namen die SessionID des Users, die beim Aufruf der graphen-generierenden Funktion durch Coldfusion als Input-Parameter mitgegeben wird.

Die Bilder, die die erzeugten Graphen zeigen, werden in einem bestimmten Ordner abgespeichert und nach dem Visualisierungstyp und der SessionID des Users benannt. Damit wird eine parallele Nutzung des Statistik-Tools gewährleistet. Es entstehen also maximal 4 Bild-Dateien für jede Session, die momentan noch manuell gelöscht werden müssen.

Im Folgenden wird die Transformation von Listen in Graphen für jeden Visualisierungs-Typ einzeln betrachtet.

4.2.1 Pie-Chart

Wenn in Coldfusion Java-Applets ausgeführt werden sollen, muss zuerst der Konstruktor der jeweiligen Klasse durch die Funktion `init()` aufgerufen werden. Die Daten-Sammlung wird durch eine Funktion der Klasse `'DataElementContainer'` vorgenommen, die Erzeugung des Graphen durch eine Funktion der Klasse `'ChartClient'`.

Ein Loop durchläuft die Listen, welche die Werte und Bezeichnungen für die Graphen enthalten und ruft bei jedem Durchgang die Funktion `'ChartDataElement'` auf. In diesem einfachen Fall bilden zwei Listen den geforderten Input für ein 2-dimensionales Datenobjekt. Nachdem die Listen abgearbeitet sind, wird die Funktion `'drawPie'` der Klasse `ChartClient` aufgerufen. Input-Parameter dieser Funktion sind der absolute Pfad für das zu generierende Bild, der Titel, die Höhe und Breite des Bildes, sowie das in der ersten Klasse erzeugte Datenobjekt.

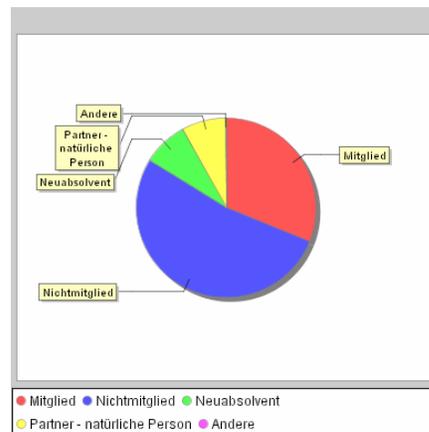


Abb. 16: Beispiel Pie-Chart

4 Implementierung

Ein einfaches Beispiel wäre eine label-list, „Männlich,Weiblich“, und eine value-list mit den Werten „7000,3000“. Damit entsteht ein Pie-Chart, der ein 7:3 Verhältnis zeigt und dieses mit den entsprechenden Labels bezeichnet.

4.2.2 Bar-Chart

Hier werden mit einer Funktion der Klasse ‚BarDataElementContainer‘ die Daten aus den Listen ausgelesen und in ein Datenobjekt transformiert. Dieses Datenobjekt, das durch die Funktion BarDataElement() generiert wird, ist um eine Dimension reicher wie beim Pie-Chart, da es zusätzliche Informationen zu einer zweiten Achse benötigt. Für die Erzeugung des Graphen dient auch hier eine Funktion der Klasse ‚ChartClient‘.

Diese Funktion, drawBar(), verlangt als Input-Parameter den absoluten Pfad der Bild-Datei, den Titel des Graphen, die Achsen-Bezeichnungen, die Bildhöhe und –breite, sowie einen Booleschen Ausdruck über die Aktivierung einer sogenannten Legende. Darin werden die im Graphen dargestellten Kategorien beschrieben.

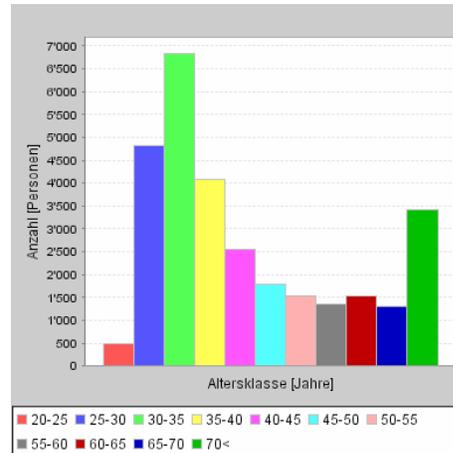


Abb. 17: Beispiel Bar-Chart

4.2.3 Stacked Bar-Chart

Gleich wie beim Barchart heissen die hier benötigten Klassen ‚BarDataElementContainer‘ und ‚ChartClient‘. Das Datenobjekt gleicht jenem für einen Bar-Chart und wird auf dieselbe Art und Weise durch die Funktion BarDataElement() konstruiert. Für die Erzeugung des Graphen ist die Funktion drawStackedBar() zuständig. Als Input-Parameter verlangt sie den absoluten Pfad der Bild-Datei, den Titel des Graphen, die Achsen-Bezeichnungen, das Datenobjekt sowie die Bildhöhe und –breite.

4 Implementierung

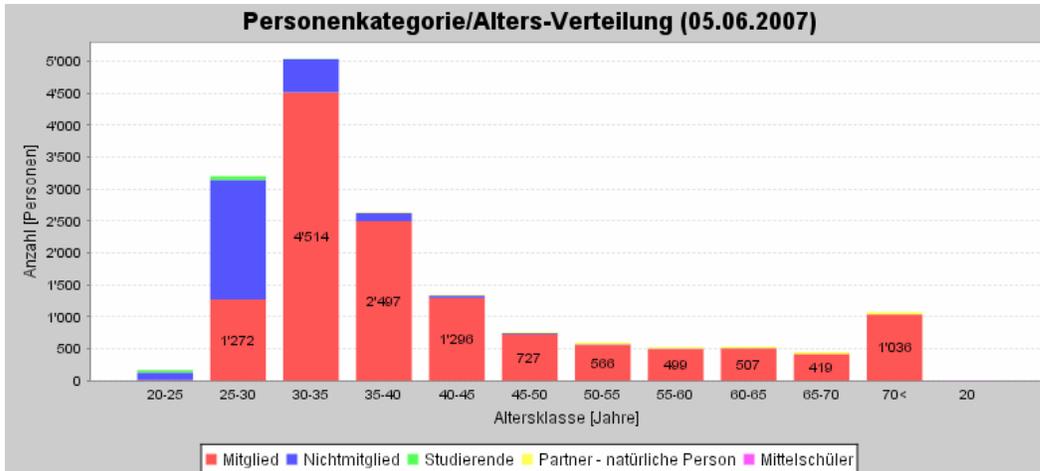


Abb. 18: Beispiel Stacked Bar-Chart

4.2.4 Zeitreihe

Auch hier wird, ähnlich wie beim Pie-Chart, ein 2-dimensionales Datenobjekt benötigt, welches anstelle von Labels Datums-Angaben enthält. Dafür wird die Funktion `TimeSeriesElement()` der Klasse `'DataElementContainer'` in einem Coldfusion-Loop so oft aufgerufen, bis die Listen abgearbeitet sind. Zusammen mit dem absoluten Pfad, dem Titel des Graphen, der Bildhöhe und -breite sowie den Achsenbezeichnungen wird dieses Datenobjekt der Funktion `drawTimeserie()` als Input-Parameter übergeben, welche ihrerseits den Graphen generiert und in eine Bild-Datei exportiert.

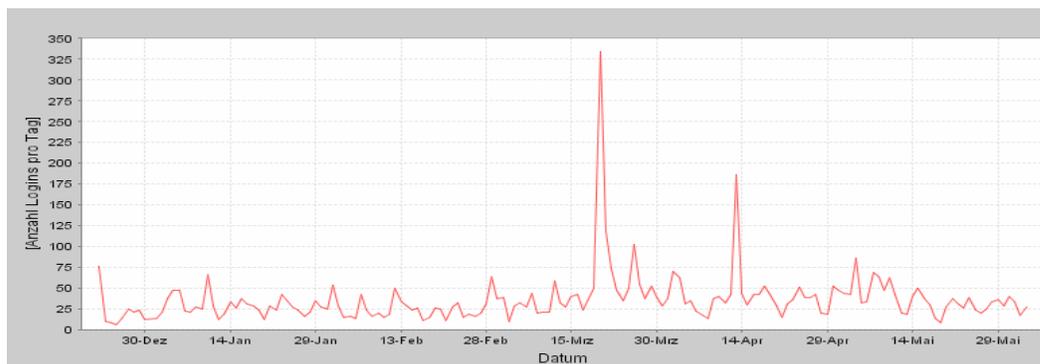


Abb. 19: Beispiel Zeitreihe

4.3 Das Datacenter

4.3.1 Erste Säule: Mitglieder



4.3.1.1 Beschreibung

Hier können 1-dimensionale Graphen dargestellt werden, die zum Beispiel über die Geschlechter-Verteilung informieren. Mögliche Auswertungen sind die folgenden:

- Geschlechterverteilung
- Alters-Verteilung
- Personenkategorien
- Bezüge/keine Bezüge
- Dauer der Mitgliedschaft

Daneben existiert die Option, noch eine zweite Auswertungs-Dimension zu wählen, wenn man zum Beispiel innerhalb der Geschlechter die Aufteilung in Personenkategorien kennen will. Aus 5 möglichen Auswertungen können demnach 20 verschiedene Auswertungs-Kombinationen gewählt werden, da Auswertungen unter ein und denselben Typen ausgeschlossen sind. Bei einfachen Auswertungen werden dem User ein Bar-Chart zusammen mit einem Pie-Chart gezeigt. 2-dimensionale Auswertungen werden durch sog. Stacked Bar-Charts dargestellt.

4.3.1.2 Die Routine

Es wird zu Anfangs eine Liste deklariert, die `mandant_list`, welche alle abzufragenden Datenbanken enthält. Somit wird das Hinzufügen einer neuen Datenbank erheblich erleichtert. Diese Liste wird nun einmal durchlaufen um für jeden Mandanten, (oder hier: Datenbank), ein eigenes Text-File anzulegen. Innerhalb dieses Loops werden Anfangs Abfragen gemacht um einige wenige Metainformationen bereitzustellen, wie z.Bsp. das Datum und die Zeit der Erfassung oder die totale Anzahl Mitglieder in der Datenbank. Anschliessend

4 Implementierung

folgen die Abfragen für die insgesamt 20 verschiedenen Auswertungs-Kombinationen, die einem in diesem ersten Element angeboten werden. Für jede Auswertungs-Kombination werden 3 verschiedene Listen abgespeichert und eindeutig gekennzeichnet.

- `list_value`
- `list_cat`
- `list_serie`

Die Kennzeichnung der Listen geschieht durch vorangestellte und sie abschliessende „Marker“, im Fall der Geschlechterverteilung zum Beispiel „xxxModullxxx“ und „xxxModullendxxx“. In der ersten Liste werden jene Werte abgespeichert, die über die Anzahl einer bestimmten Kategorie Mitglieder informiert. In den anderen Listen werden die zugehörigen Kategorien und Bezeichnungen abgespeichert. Während des gesamten Vorgangs werden die eben generierten Listen in eine Variable geschrieben. Am Ende der Abfragen wird die „newFile“-Variable in ein Textfile mit dem Namen `AlumniTest_values.txt` geschrieben, und die Schleife wiederholt sich für die nächste Datenbank.

4.3.1.3 Visualisierung

Das File, welches aus den soeben beschriebenen Listen Graphen erstellen soll, heisst `statistics.cfm`. Es ist innerhalb eines `<div>`-Tags im Hauptfile, `datacenter.cfm`, eingebettet. Es beinhaltet das GUI für den User und bietet folgende Optionen: Die Auswahl der Datenbank, wobei die eigene Datenbank oder auch alle Datenbanken im Plenum ausgewählt werden können. Daneben wird der Typ der Auswertung gewählt. Die erste Auswertungs-Dimension wird jeweils auf der x-Achse im Bar-Chart gezeigt. Optional kann eine zweite Dimension gewählt werden, die ihrerseits zusätzlich auf der y-Achse abgebildet wird. Wird nun eine bestimmte Auswertungskombination selektiert, wird das Text-File mit dem Namen `AlumniTest_values.txt` (`AlumniTest` steht für den Mandanten) gelesen und nach den jeweiligen Markern gesucht. Die darin eingeschlossenen Listen werden in temporäre Variablen geschrieben und zur Visualisierung an die Java-Applets weitergegeben. Für jede Art von Diagramm wird zuerst ein Applet initiiert, welches den erforderlichen Daten-Container

4 Implementierung

erstellt. Anschliessend wird dieser einer weiteren Java-Funktion weitergegeben, die ihrerseits eine entsprechende Graphik generiert und an einem vorgegebenen Ort abspeichert.

Zwei Funktionen werden unverändert in jedem der in diesem Kapitel besprochenen Register angeboten, daher wird auf eine wiederholende Erwähnung verzichtet. Die erste ist die Reset-Funktion. Bei Betätigung werden in der Eingabe-Maske alle Default-Werte wiederhergestellt und mögliche Auswertungs-Graphiken ausgeblendet, um so schneller zwischen verschiedenen Ansichten wechseln zu können. Die zweite Funktion heisst „Drucken“. Wird sie aufgerufen, wird in einem neuen Browser-Fenster ein PDF geöffnet, welches neben der eigentlichen Graphik auch andere Informationen enthält und nun archiviert oder ausgedruckt werden kann.

Eine hier angebotene Option ist das Ausblenden der sog. Missing Values. Wenn bei bestimmten Auswertungen der Anteil der fehlenden Informationen zu gross ist, kann dies ein Diagramm verzerren und dessen Interpretation erschweren. Missing Values gibt es vor allem dort, wo die zu verarbeitenden Daten den Eintrag NULL zulassen. Bei der Auswertungsvariante „Alters-Verteilung“ wird dies gut ersichtlich, da das Geburtsdatum der Mitglieder kein sog. Mussfeld ist und somit auch leere Einträge akzeptiert.

Zusätzlich zu den Statistiken werden jeweils die Listen, die für die Generierung der Graphiken verwendet wurden, dargestellt, um die exakten Zahlen dem interessierten Benutzer zugänglich zu machen.

4 Implementierung

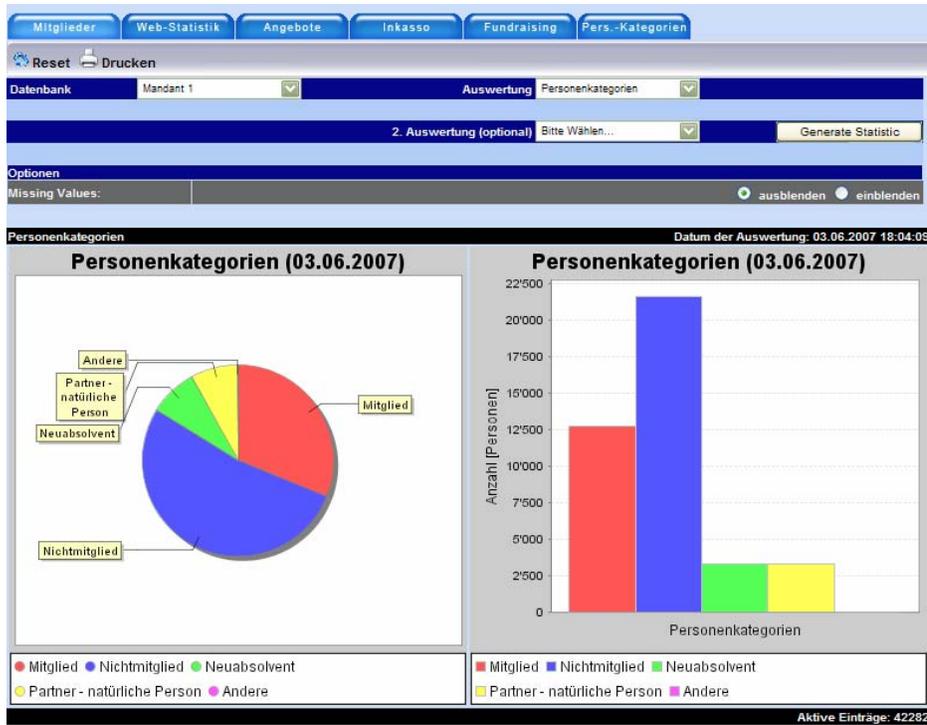


Abb. 20: 1-dimensionale Auswertung, Mandant 1

4 Implementierung

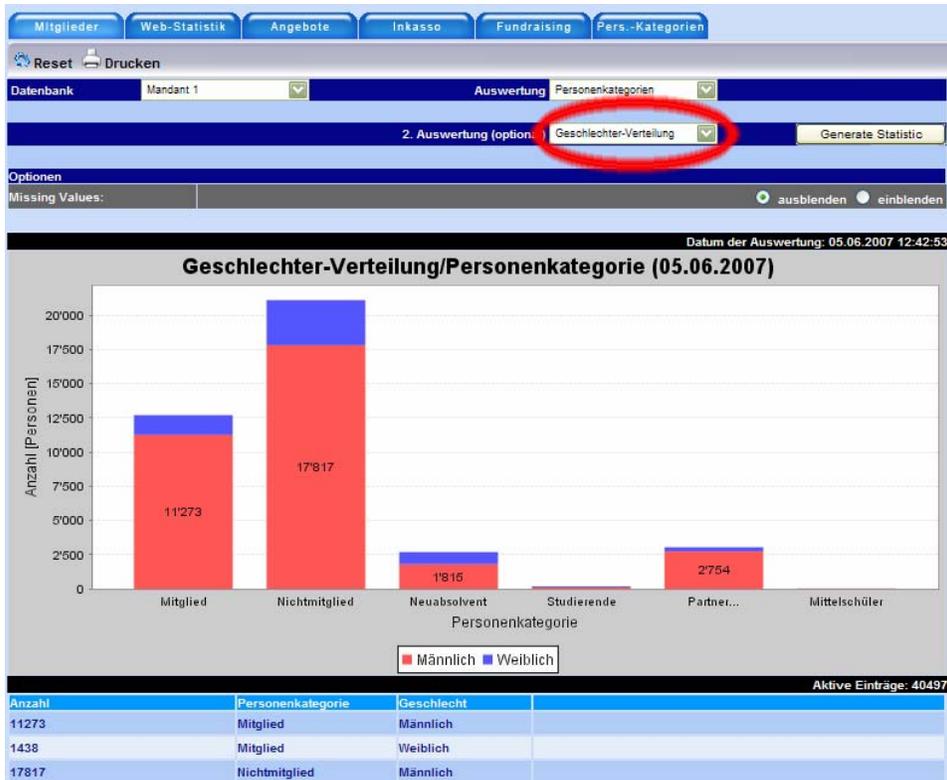


Abb. 21: 2-dimensionale Auswertung, Mandant 1



4.3.2 Zweite Säule: Web-Statistik

4.3.2.1 Beschreibung

Dieses Register soll über die Aktivitäten der Mitglieder im Online-Bereich, der sog. „Aussensicht“, informieren. Als Vorbereitung für diese Aufgabe wurde im Dezember 2006 ein Mechanismus implementiert, der für jeden Besuch auf dem Online-Portal ein log-file anlegt. Die Namen der Logfiles bestehen aus dem Datum und der Session-ID des Users. Ein möglicher Name wäre demnach 31_12_2006_21143486492349.txt. Im Header des Files befinden sich Informationen zum User und der genauen Login-Zeit. Anschliessend werden die einzelnen Schritte, bzw. „Clicks“, des Users aufgezeichnet. Daraus ersichtlich sind das gewählte Register, die genaue Zeit des Aufrufs, mögliche Aktionen wie das Ausführen einer Suche und allfällige Attribute wie in diesem Fall die Suchkriterien. Durch die Auswertung der Logfiles können folgende Aspekte durch eine Zeitreihe dargestellt werden:

4 Implementierung

- Anzahl Besuche pro Tag
- Anzahl Aktionen („Clicks“) pro Session
- Dauer der Session (in Sekunden)
- Häufigkeit der Besuche
- Wahrscheinlichkeit je Besuch (Visualisiert)

Die genannten Logfiles waren ursprünglich gedacht für das Sammeln von Daten innerhalb eines neuen Moduls, dem Mentoring-Modul, welches erst seit Herbst 2006 in Betrieb ist. Dabei können Studenten der Bildungseinrichtung vom Mandanten 1 Ehemalige als Mentoren anfragen und in einer Kontaktliste speichern. Die Absicht war es, Daten zu sammeln bevor und nachdem die Mitglieder über das neue Modul informiert wurden, um so Aussagen machen zu können über die Wirkung einer solchen Ankündigung. Leider blieb eine einheitliche Ankündigung an die Mitglieder aus, was die geplante Analyse erschwert bzw. verunmöglicht. Aus diesem Grund wurden lediglich für Mandant 1 entsprechende log-Files angelegt, die nun in diesem Teil zu einem anderen Zweck ausgewertet werden.

4.3.2.2 *Die Routine*

Für die Zusammenstellung der erforderlichen Daten ist die Datei `logfile_schedule.cfm` verantwortlich. In dieser Routine wird über alle vorhandenen Logfiles ein Loop durchlaufen. Für die Berechnung der Anzahl Besuche pro Tag können lediglich die Filenamen durchgearbeitet werden um so herauszufinden, wieviele Logins es pro Tag gab. Diese Informationen werden wieder in eine Liste geschrieben und in einem Textfile gespeichert. Um die Anzahl Aktionen pro Session herauszufinden, muss das Logfile gelesen und die darin enthaltenen Einträge gezählt werden. Für die Dauer der Session wird lediglich die Differenz zwischen der Login-Zeit und dem Timestamp der letzten Aktion berechnet. Bei der Häufigkeit der Besuche gestaltet sich die Berechnung etwas komplizierter. Für jeden Monat wird eine eigene Liste angelegt, die darüber informiert, wie oft eine Person in diesem Monat sich auf dem Online-Portal angemeldet hat. Die Position des Listenelements gibt Auskunft über die

4 Implementierung

Häufigkeit der Besuche, wobei der Wert des Listenelements die Anzahl Personen bezeichnet, die sich entsprechend häufig angemeldet haben. Würde die Liste folgendermassen beginnen, „47,34,21,.....“, kann herausgelesen werden, das sich in dem entsprechenden Monat 47 Personen einmal, 34 Personen zweimal und 21 Personen dreimal angemeldet haben.

Bei der ‚Wahrscheinlichkeit je Besuch‘ wird ausgewertet, ob ein bestimmtes Register in einem Logfile aufgerufen wurde oder nicht, unabhängig davon, wie oft dies geschah. Das User-Profil wird zu 100% besucht, da jeder User nach dem Login automatisch auf dieses Register geleitet wird.

4.3.2.3 Visualisierung

Die oben beschriebenen Listen werden durch das File `logfile_analysis.cfm` ausgelesen, welches in einem `<div>`-Tag im zweiten Register im Datacenter eingebettet ist. Das GUI zeigt folgende Optionen: Die Art der Auswertung, die Bildbreite und den Zeitraum, in welchem Daten betrachtet werden sollen. Der Vorgang zur Visualisierung ist derselbe wie in dem zuvor besprochenen Fall. Nachdem ein Aufruf erfolgt ist, werden die entsprechenden Listen zwischen den Markern herausgelesen, in eine temporäre Variable geschrieben und an das Java-Applet weitergereicht, um damit den entsprechenden Graphen zu erzeugen.

Für die Visualisierung der online-Nutzung werden Prozent-Angaben in Farbcodes (Hexadezimal) umgerechnet. Jedes Kästchen im GUI (Abb. 23) repräsentiert ein Register in der Aussensicht (Abb. 24), darunter sind die Subregister angehängt, die alle unterschiedlich eingefärbt werden. Wird eine bestimmte Zeitspanne gewählt und die Auswertung angezeigt, werden zuerst die einzelnen Prozent-Zahlen berechnet, um anschliessend aufgrund dieser Liste eine zusätzliche Liste, die sog. `color_list`, mit den entsprechenden Farbcodes zu erstellen. Würde ein bestimmtes Register bei jedem Besuch mindestens einmal besucht werden, die Wahrscheinlichkeit läge also bei 100%, würde die Fläche Rot (Farbcode: `#FF0000`) eingefärbt. Bei einer Wahrscheinlichkeit von 0% würde das entsprechende Feld in der Visualisierung Gelb (Farbcode: `#FFFF00`) erscheinen.

4 Implementierung

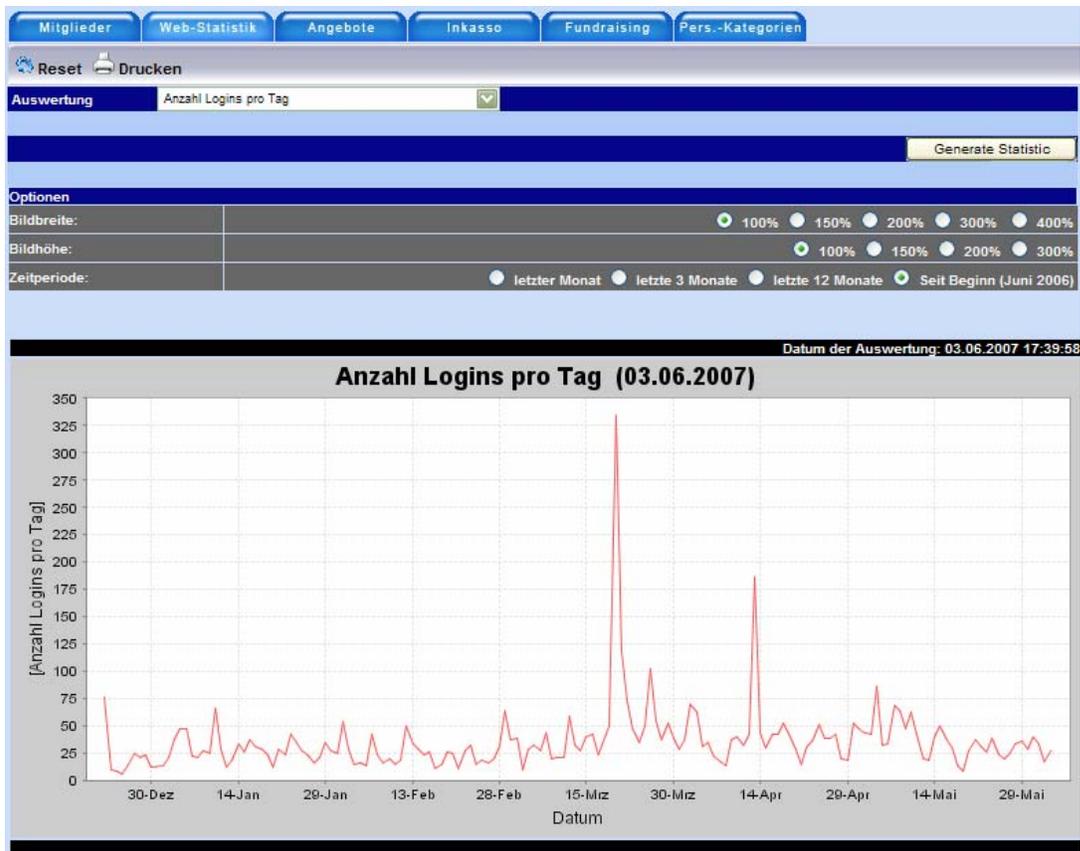


Abb. 22: Auswertungstyp "Anzahl Logins pro Tag", Mandant 1

4 Implementierung

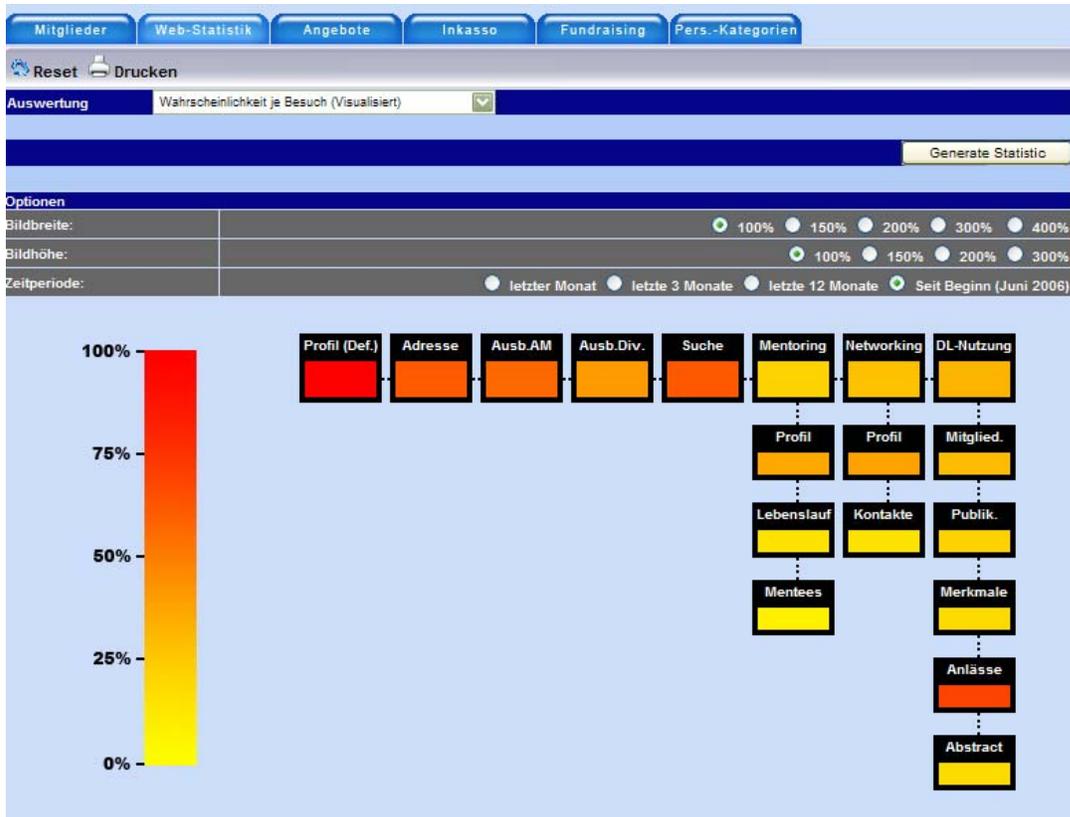


Abb. 23: Auswertungstyp "Wahrscheinlichkeit je Besuch", Mandant 1

Abb. 24: Hauptregister in der "Aussensicht"

4.3.3 Dritte Säule: Angebote



4.3.3.1 Beschreibung

Nebst Spenden gehören die Einnahmen durch angebotene Dienstleistungen zu den wichtigsten Einnahmequellen für die Vereine. Entsprechend gross ist ihrerseits das Interesse, zu wissen, wie und von wem die angebotenen Dienstleistungen genutzt werden.

Es können je nach Mandant unterschiedlich Mitgliedschaften, Veranstaltungen, Publikationen und/oder sonstige Dienstleistungen gebucht werden, abhängig vom jeweiligen Angebot des betreffenden Ehemaligen-Vereins. Über eine erste Auswertungs-Dimension können die unterschiedlichen Dienstleistungsarten gewählt werden. Dargestellt wird die Anzahl Personen, die für die gewählte Dienstleistung in eine bestimmte Ausgabenkategorie fällt. Eine zweite Dimension erlaubt eine weitere Differenzierung jener Personen durch bereits bekannte Kriterien aus den Mitglieder-Statistiken:

- Geschlechterverteilung
- Alters-Verteilung
- Personenkategorien
- Dauer der Mitgliedschaft

In Form eines Bar-Charts, welches auf der y-Achse die Anzahl Personen zeigt und auf der x-Achse die verschiedenen Ausgabenkategorien abbildet, werden die geforderten Informationen visualisiert.

Die am häufigsten genutzten Dienstleistungen sind die Mitgliedschaften, die gegen eine entsprechende Gebühr jedes Jahr erneuert werden müssen. Da dieser Jahreszyklus für Mitgliedschaften auch Einfluss auf die eingehenden Zahlungen der Mitglieder hat, werden die Berechnungen auf Jahresbasis durchgeführt.

4 Implementierung

4.3.3.2 *Die Routine*

In dieser Routine werden die Ausgaben der Mitglieder in Listen gespeichert und in ein Text-File geschrieben. Sie funktioniert folgendermassen: Wenn bereits ein entsprechendes Text-File besteht, werden lediglich die Listen für die Ausgaben des aktuellen Jahres erneuert, jene von den vorangehenden Jahren werden beibehalten. Da sich die Zahlen der vergangenen Jahre in dieser Auswertung kaum mehr verändern, ist eine wiederholte Neuberechnung nicht notwendig. Sollte eine solche Neuberechnung vergangener Jahre dennoch einmal durchgeführt werden, kann lediglich das bereits existierende Text-File gelöscht werden um anschliessend die Routine erneut zu starten. Existiert noch kein entsprechendes Text-File, werden nämlich sämtliche Angebotsbezüge abgearbeitet und es entstehen pro Mandant vier Listen für jedes Jahr seit dem ersten Angebotsbezug, jede für eine der verschiedenen Dienstleistungsarten. Andernfalls werden die Zahlen für das laufende Jahr berechnet und nur diese bestimmte Liste wird durch eine aktuellere ersetzt.

Die Listen werden erneut mit bestimmten Markern umgeben, die Informationen über das Jahr und die Dienstleistungsart enthalten, und damit eine spätere Wiedererkennung ermöglichen.

4.3.3.3 *Visualisierung*

Die oben erwähnten Listen werden mit dem File `statistics_offer.cfm` ausgelesen, welches im dritten Register des Datacenters eingebettet ist. Folgende Optionen sind dargestellt: die Höhe der generierten Bilddatei, um eine Verfeinerung bzw. Einschränkung der Werte auf der y-Achse zu erreichen. Zusätzlich kann der maximale Betrag, der auf der x-Achse abgebildet wird, verändert werden, um mögliche leere Werte in den höheren Ausgabe-Kategorien zu vermeiden.

4 Implementierung

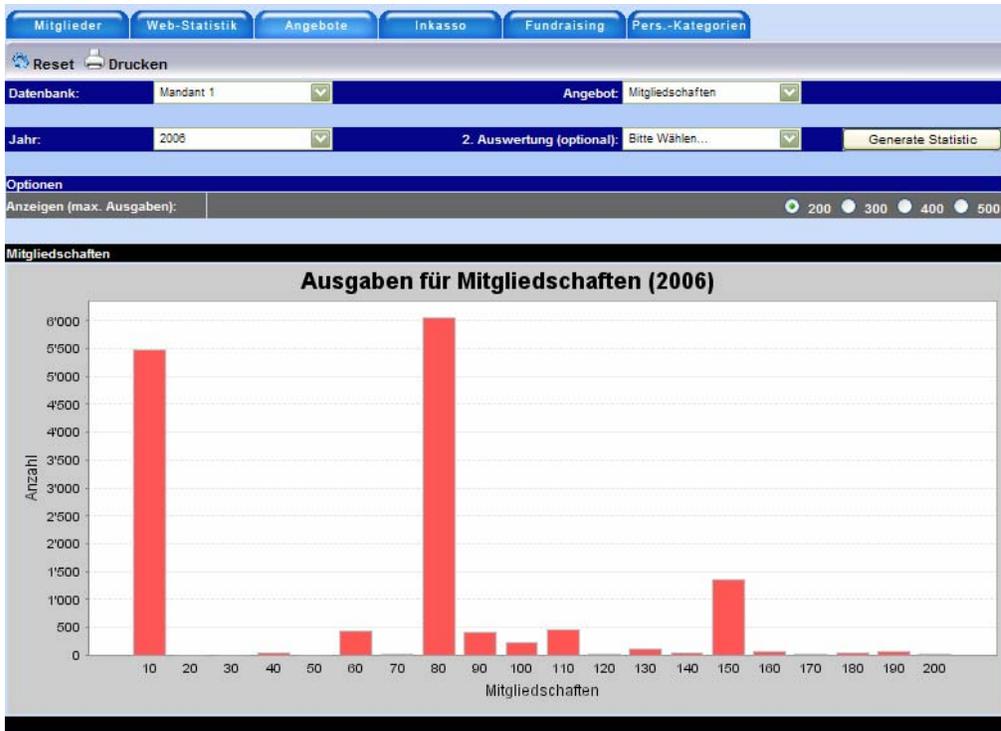


Abb. 25: 1-dimensionale Auswertung, Mandant 1

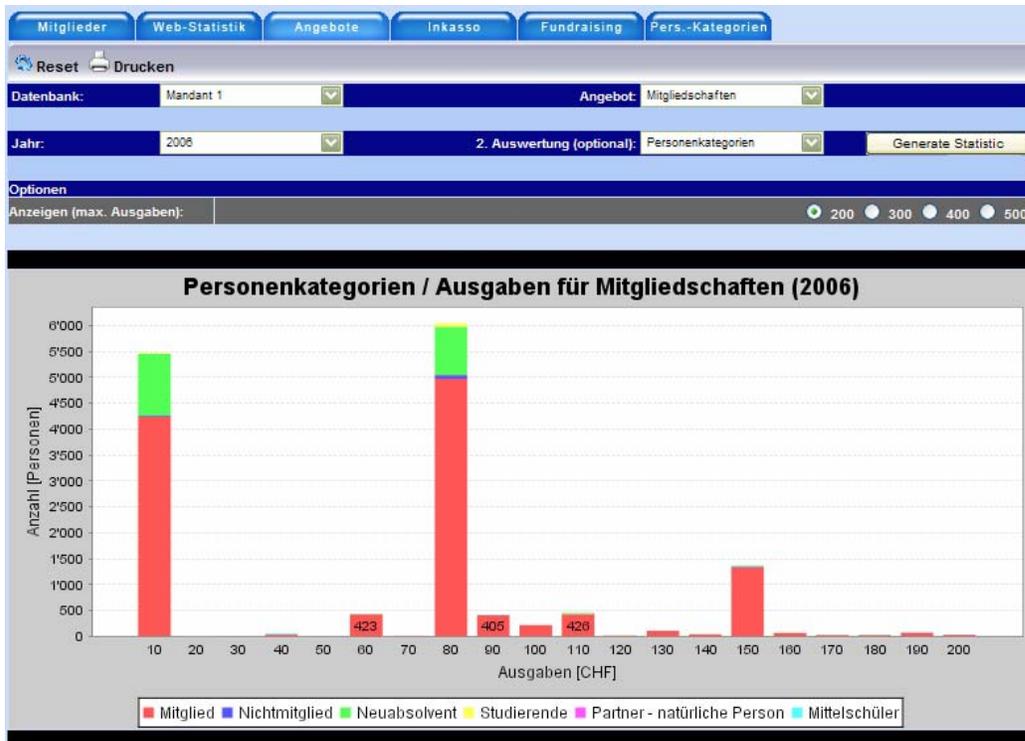


Abb. 26: 2-dimensionale Auswertung, Mandant 1



4.3.4 Vierte Säule: Inkasso

4.3.4.1 Beschreibung

Dieses wie auch das nächste Register wurde auf Kundenwunsch implementiert. Es sollte eine Übersicht schaffen über alle eingegangenen Zahlungsströme, die den Bezug von Dienstleistungen betrafen. Während Mitgliedschaften und Publikationen in einem jährlichen Zyklus abgerechnet werden, erreichen die Zahlungen für bestimmte Anlässe die Vereine unregelmässig, je nach Zeitpunkt der Veranstaltungen. Die Daten werden hier mittels einer Zeitreihe dargestellt.

4.3.4.2 Die Routine

Bei Bezahlung der Rechnungen erhalten die Vereins-Administrationen von den Banken sogenannte VESR-Files, die über den Zahlungseingang informieren und ihn damit bestätigen. Alumnitools bietet den Administrationen die Möglichkeit, einen Upload dieser Files vorzunehmen und sie zu verarbeiten. Eine dafür entwickelte Routine liest diese Files regelmässig aus und speichert sie in der Datenbank. Für jede Zahlung, die mit einer Referenznummer in der Datenbank übereinstimmt, wird ein Eintrag in der Tabelle PersonDebitorenKonto erfasst. Diese Tabelle liefert die hier erforderlichen Informationen. Es werden alle Zahlungseingänge pro Tag zusammengerechnet und zusammen mit dem betreffenden Datum in einer Liste, die wieder mit bestimmten Markern versehen ist, abgespeichert. So entstehen für dieses Register lediglich 2 Listen gleicher Länge, die eine mit den jeweiligen Daten und die andere mit den zugehörigen täglichen Zahlungssummen.

4.3.4.3 Visualisierung

Das File, welches in vierten Register des Datacenters erscheint, extrahiert die wesentlichen Daten aus dem oben beschriebenen Text-File.

Es können nebst dem Mandanten auch der genaue Zeitraum gewählt werden, in welchem der Zahlungs-Verlauf aufgezeigt werden soll. Als Default-Wert wird im ersten Datums-Feld das erste Element der Datumsliste herangezogen, also das

4 Implementierung

Datum der ersten in der Datenbank erfassten Zahlung. Da diese Daten von Mandant zu Mandant variieren, werden für jeden Mandanten eigene Datumsfelder angezeigt. Im zweiten Eingabefeld wird das aktuelle Datum als Defaultwert angezeigt. Werden keine Änderungen an den Default-Werten vorgenommen, wird die gesamte Zeitreihe, von der ersten bis zur letzten erfassten Zahlung, visualisiert.

Da es sich um längere Zeitreihen handelt, wurden mittels variabler Bildhöhe und -breite Optionen implementiert, um eine Verfeinerung bzw. Einschränkung der Werte nicht nur auf der y-Achse, sondern auch auf der x-Achse zu ermöglichen.

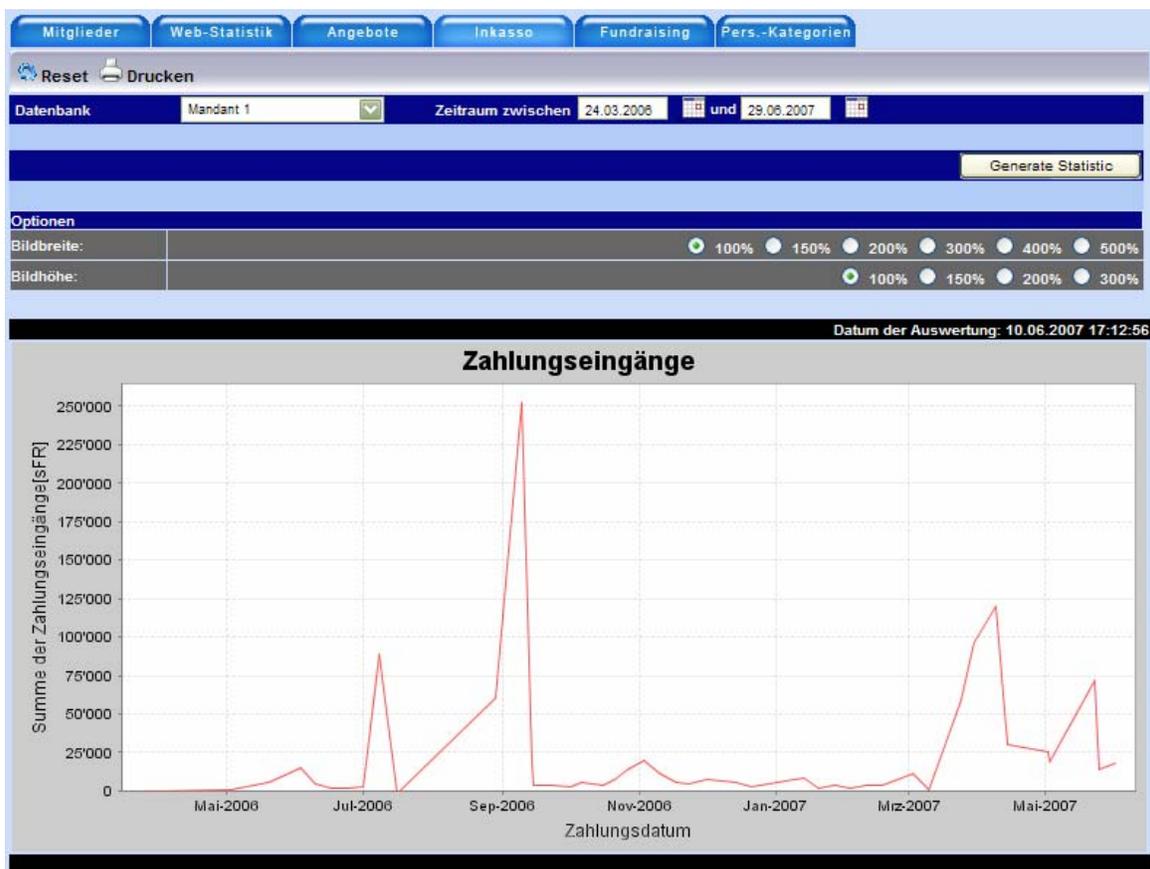


Abb. 27: Zeitreihe von Zahlungseingängen, Mandant 1



4.3.5 Fünfte Säule: Fundraising

4.3.5.1 Beschreibung

Das in Alumnitools vorhandene Modul zur Verwaltung von Spendeneinnahmen wird zur Zeit nur durch Mandanten 1 genutzt, daher berücksichtigen die Auswertungen an dieser Stelle lediglich diese einzelne Datenbank.

Spendeneinnahmen sind für bestimmte Ehemaligen-Vereine unerlässlich und häufig lassen sich Projekte ohne diese Art der Finanzierung kaum realisieren. Nun entsprach es dem Wunsch des Mandanten, eine Übersicht zu haben über die bereits getätigten Spenden auf Basis der unterschiedlichen Kampagnen, um nicht nur genau über den aktuellen Stand informiert zu sein, sondern auch besser über Erfolg oder Misserfolg von Spendenaufrufen urteilen zu können.

Hier werden zwei Arten von Spenden unterschieden. Es gibt neutrale Spenden, die aus gewollt überbezahlten Rechnungen resultieren, oder aber gezielte Spenden, die ein bestimmtes Spenden-Projekt betreffen. Das Ziel dieser Projekte, sogenannter Kampagnen, ist es, einen Mehrwert zu schaffen für den Bildungsort der betreffenden Institution und/oder die Mitglieder des betreffenden Ehemaligen-Vereins.

4.3.5.2 Die Routine

Zuerst wird die Datenbank nach aktuellen Kampagnen abgefragt, um anschliessend für jede Kampagne vier Listen zu generieren: Zwei Listen dienen der Visualisierung einer Zeitreihe. Die erste enthält die genauen Daten, an welchen die Spenden eingegangen sind. Eine zweite Liste speichert die entsprechenden Beträge. Zwei weitere Listen werden für die Darstellung des Bar-Charts benötigt. Die eine Liste enthält die Ausgabe-Kategorien, die andere Liste speichert die Anzahl Personen, die in diese Kategorien fallen. Alle Listen werden zusammen mit Informationen zum Auswertungszeitpunkt in ein Text-File gespeichert und so zur Visualisierung bereitgestellt.

4 Implementierung

4.3.5.3 Visualisierung

Das im fünften Register eingebettete File heisst `fundraising.cfm` und bietet zwei unterschiedliche Arten der Darstellung an. Einerseits kann mit der Option ‚Zeitreihe‘ der zeitliche Verlauf der Spendeneinnahmen visualisiert werden (Abb.29), andererseits bietet die Option ‚Bar-Chart‘ die Möglichkeit, ähnlich dem Register ‚Angebote‘, die Verteilung der Spender auf verschiedene Ausgabe-Kategorien darzustellen (Abb.28). So wird nicht nur ersichtlich, zu welchen Zeitpunkten wieviel gespendet wurde und wie lange die Aktion bereits andauert, sondern auch wie hoch die Anzahl Personen ist, die einen bestimmten Betrag bezahlt haben. Weitere Optionen sind erneut die Bildhöhe und –breite zur Verfeinerung bzw. Einschränkung der Skala auf der jeweiligen Achse. Zusätzlich kann ein maximaler Betrag gewählt werden, der in der Auswertung berücksichtigt wird. Es kommt vor, dass es ausserordentliche Spenden gibt, die sich in ihrer Höhe stark von den übrigen Spenden abheben und so die Interpretation einer Zeitreihe geradezu verunmöglichen. Durch die Auswahl eines geeigneten maximalen Betrags kann dies umgangen werden.

4 Implementierung

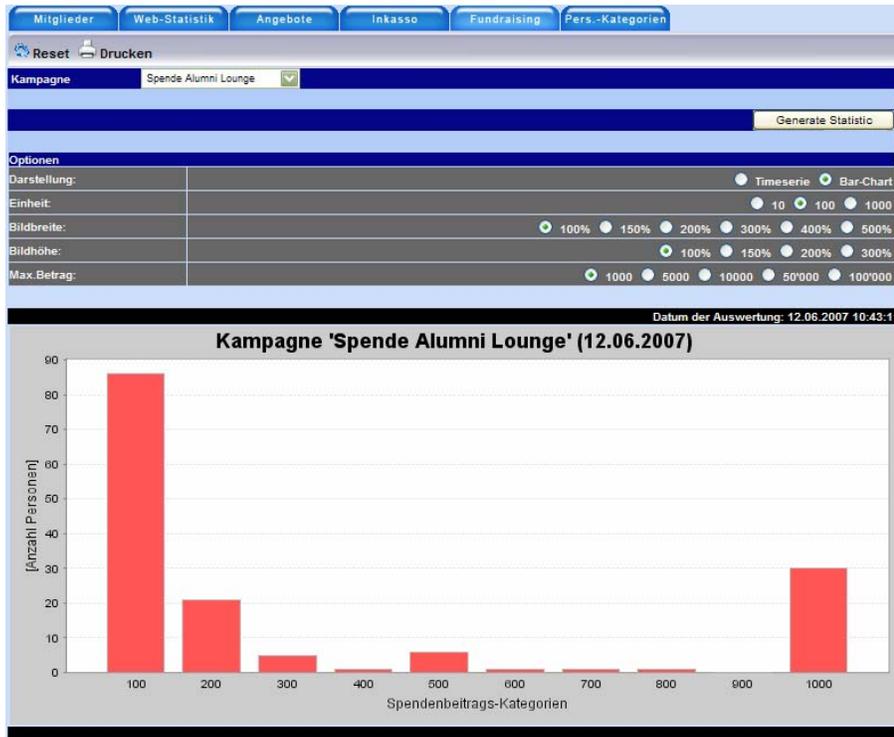


Abb. 28: Bar-Chart von Spendeneingängen, Mandant 1

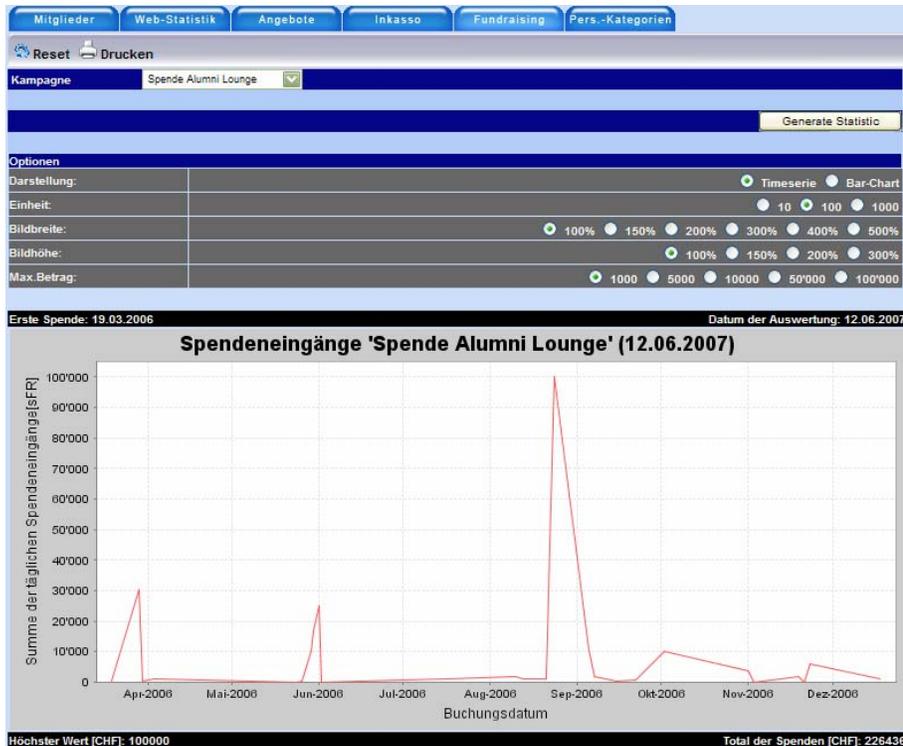


Abb. 29: Zeitreihe von Spendeneingängen, Mandant 1



4.3.6 Sechste Säule: Personenkategorien

4.3.6.1 Beschreibung

Ebenfalls auf Kundenwunsch wurde das sechste und letzte Register realisiert. Hier werden die zahlenmässigen Veränderungen der verschiedenen Personenkategorien für jeden Mandanten zeitlich dargestellt (Abb.30). Das Verhältnis zwischen diesen Personenkategorien ist für die Administrationen von grossem Interesse, da sie aufzeigt, wie gross der Anreiz von Schulabgängern ist, Teil dieses Netzwerks zu werden.

Nachdem ein Student seine Ausbildung abgeschlossen hat, wird er in die Datenbank der jeweiligen Alumni-Organisation aufgenommen. Entscheidet er sich für eine Mitgliedschaft, wird seine Personenkategorie von ‚Neuabsolvent‘ auf ‚Mitglied‘ geändert und er geniesst von dem Moment an alle Privilegien eines Alumnis.

4.3.6.2 Die Routine

Diese Routine muss täglich einmal aufgerufen werden, um die aktuelle Zahl der jeweiligen Personenkategorien zu speichern. Es werden in dem File `categories_values.txt` für jede einzelne Personenkategorie zwei Listen angelegt. Täglich werden beide Listen um ein Element ergänzt. Das aktuelle Datum wird in die erste Liste geschrieben, die absolute Zahl der entsprechenden Kategorie in die zweite Liste. Wurden die Listen ergänzt, werden sie im gleichen Text-File wieder abgespeichert.

4.3.6.3 Visualisierung

Das im letzten Register des Datacenters eingebettete File heisst `categories.cfm`. Die Informationen werden mittels einer Zeitreihe dargestellt. Mögliche Optionen sind erneut die Bildhöhe und –breite, und auch der Zeitraum, in welchem die Veränderungen betrachtet werden sollen. Die Defaultwerte sind ähnlich dem Register Inkasso für den Startzeitpunkt jenes Datum, an welchem die Routine zum ersten mal gestartet wurde, und für das Enddatum der aktuelle Tag.

4 Implementierung



Abb. 30: Zeitreihe der Personenkategorie 'Mitglied', Mandant 1

4.4 Data-Mining auf Kundendaten

Ein Teil des Auftrags dieser Diplomarbeit enthielt die Forderung, anhand verschiedener Data-Mining Algorithmen die Kundendaten so zu analysieren, dass die daraus gewonnenen Resultate für die Mandanten nutzbar gemacht werden könnten. Es war geplant, ein sogenanntes „Targeting-Tool“ zu implementieren, welches dem Kunden beispielsweise erlaubt hätte, einen Spenden-Aufruf oder eine Werbe-Kampagne nur an eine bestimmte Auswahl von Personen zu richten. Die Data-Mining Algorithmen wären dazu eingesetzt worden, jene Mitglieder zu identifizieren, bei denen die Spenden- oder Kauf-Wahrscheinlichkeit am grössten gewesen wäre. Zur Realisierung wäre eine Applikation mit dem Namen „YALE“ („Yet Another Learning Environment“) herangezogen worden. Es handelt sich dabei um eine Open-Source Software, die sich aufgrund ihrer umfassenden Dokumentation und einem vollständig in Java programmierten Code einfach in eine bestehende Anwendung integrieren lässt. [YALE, 2007]

Um das Prinzip zu verdeutlichen wird an dieser Stelle der KDD-Cup'98 (Knowledge-Discovery and Data-Mining) erwähnt, da sich die in dieser Arbeit bearbeitete Aufgabenstellung sinngemäss auf die damalige Problematik abstützt.

Bei diesem alljährlich stattfindenden Wettbewerb wird eine Aufgabenstellung vorgegeben und dazu ein Datensatz, auf welchem die Aufgabe ausgeführt werden soll. Die Teilnehmer müssen dieses Problem mit eigens entwickelten Methoden und Werkzeugen lösen. Gewonnen hat jene Partei, die die höchste Erfolgsquote aufweist in Bezug auf die Fragestellung. In jenem Jahr war die gestellte Aufgabe die folgende: Der Datensatz bestand aus 3.5 Millionen Personen, denen ein Brief mit einer Spenden-Anfrage geschickt wurde. Zusätzlich im Datensatz enthalten waren die tatsächlich eingegangenen Spenden. Die Aufgabe bestand nun darin, diese Informationen so zu analysieren, dass die daraus gewonnenen Erkenntnisse auf die gesamte Datenbank, total 13 Millionen Einträge, angewendet werden könnten, um so die wahrscheinlichsten Spender zu eruieren.

Nach reichlicher Überlegung und in Absprache mit dem Betreuer dieser Arbeit wurde jedoch dieses Vorhaben nicht durchgeführt. Im Zentrum dieser Entscheidung lag die Einsicht, dass durch die Realisierung dieses „Targeting-

4 Implementierung

Tools“ kein nennenswerter Mehrwert für die Mandanten geschaffen worden wäre. Unterschiedliche Überlegungen haben zu diesem Schluss geführt.

Erstens liefern Analysen dieser Art Ergebnisse, die für den Benutzer nur schwierig zu interpretieren sind. Die Berechnungen, die im Hintergrund angestellt werden, sind für den Laien kaum nachvollziehbar und es besteht die Gefahr, dass die falschen Schlüsse aus den Resultaten gezogen werden.

Zweitens wirkt sich ein Versand bei diesen vergleichsweise geringen Datenbeständen finanziell nicht derartig stark aus wie in der oben erwähnten Problemstellung. Es besteht die Gefahr von Falsch-Positiven Resultaten, was bedeutet, dass mögliche Spender nicht als solche erkannt werden und deshalb fälschlicherweise von dem Spenden-Aufruf ausgeschlossen sind.

Drittens handelt es sich bei diesem Mandanten um einen Verein mit zahlenden Mitgliedern. Somit hat jedes Mitglied dasselbe Recht, von allen angebotenen Dienstleistungen zu erfahren und diese auch nutzen zu können. Eine gezielte Werbe-Kampagne würde zur Diskriminierung zahlreicher Mitglieder führen und ist somit nicht erstrebenswert.

5. Resultate

Das Ziel dieses Kapitels ist es, den Nutzen des hier implementierten Moduls zu verdeutlichen, indem Auswertungen verwendet werden um über bestimmte Sachverhalte Auskunft zu geben. Die hier beschriebenen Auswertungen wurden aus Daten generiert, welche im Juni 2007 durch die früher beschriebenen Routinen bereitgestellt wurden. Sie stellen eine Momentaufnahme dar und sind zeitlich von begrenzter Gültigkeit.

Im ersten Register, „Mitglieder“, kann die Geschlechterverteilung innerhalb einzelner Alterskategorien betrachtet werden. Dort zeigt sich eine Zunahme des weiblichen Anteils (im Bild: Blau) der Mitglieder des Mandanten 1 über die vergangenen Jahre. Das Bildungsinstitut bietet auf seiner Homepage Statistiken der letzten Jahrzehnte an, die diese Resultate bestätigen. Dort wird beispielsweise ersichtlich, dass im Jahr 1950 3251 Studenten immatrikuliert waren. Davon waren lediglich 128 Frauen, das sind 3.94%. Im Jahr 1975 waren von 7109 Studenten bereits 665 Frauen, ein Anteil von 9.35%. Weitere 25 Jahre später liegt der Anteil der weiblichen Studentenschaft bei 26,31%.

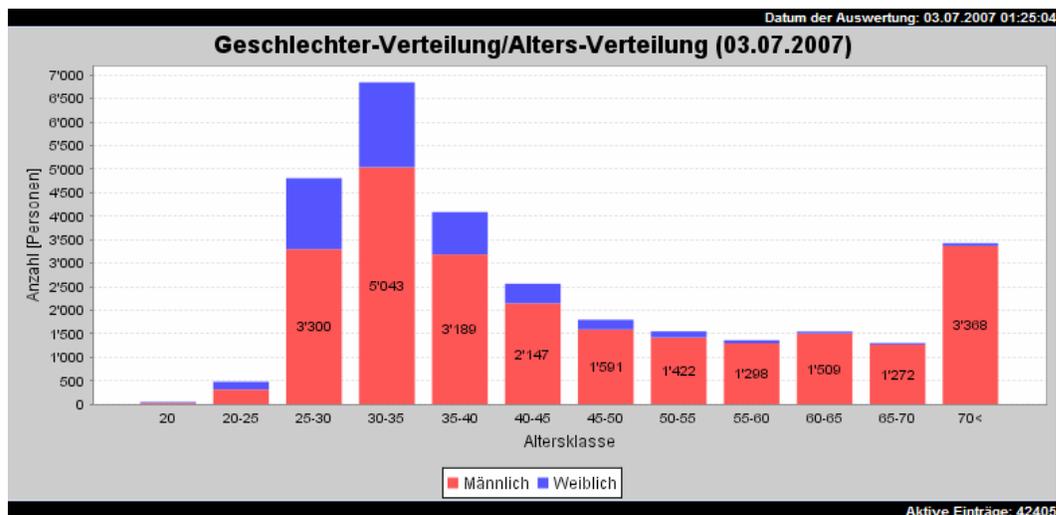


Abb. 31: Auswertungstyp “Geschlechter- /Alters-Verteilung”, Mandant 1

5 Resultate

Dasselbe Phänomen zeigt die Auswertung beim Mandanten 3:

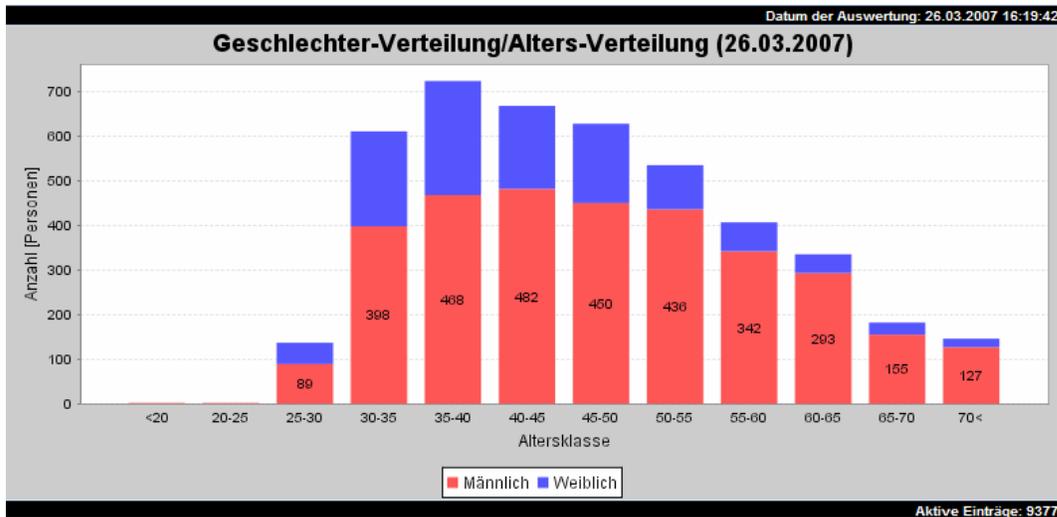


Abb. 32: Auswertungstyp "Geschlechter- /Alters-Verteilung", Mandant 3

Eine weitere Auffälligkeit tritt beim Mandanten 1 auf. Wird der Auswertungstyp „Dauer der Mitgliedschaft“ gewählt, erscheint die Graphik, die in Abb. 26 gezeigt ist. Diese besagt, dass es überdurchschnittlich viele Mitglieder gibt, die seit 4-5 Jahren in diesem Verein sind (Bild: in Pink), genau 48,3%. Nachfragen haben ergeben, dass vor dieser Zeit zwei Vereine bestanden haben, die als Dachorganisation von Alumni-Vereinen fungierten. Deren Datenbestände wurden aufgrund eines Beschlusses der Hochschul-Leitung 2002 zusammengelegt, wobei fälschlicherweise für das Eintrittsdatum das Migrations-Datum erfasst wurde.

5 Resultate

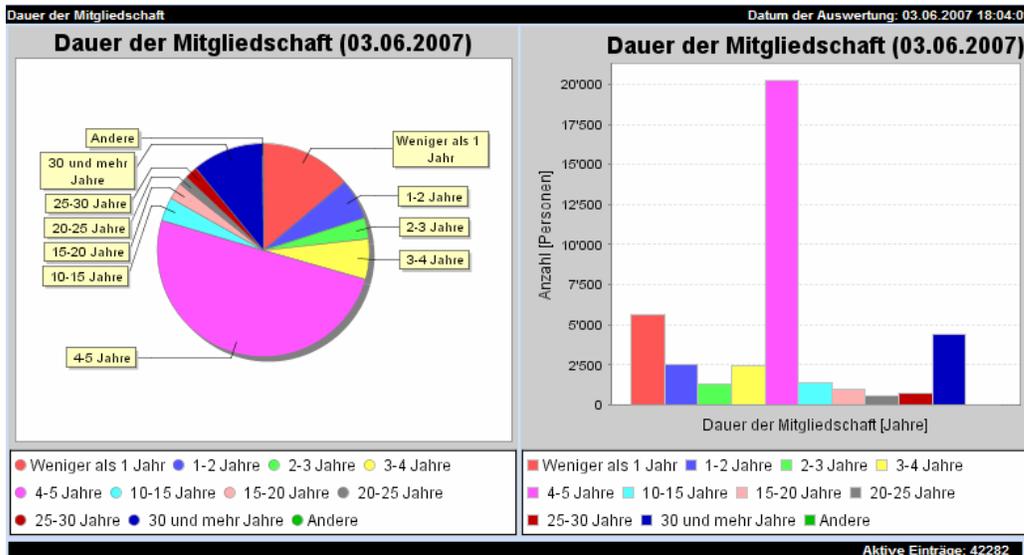


Abb. 33: Auswertungstyp “Dauer der Mitgliedschaft”, Mandant 1

Dieser Umstand muss für die korrekte Interpretation der Auswertungen für Mandanten 1 berücksichtigt werden, sobald diese den Auswertungstyp „Dauer der Mitgliedschaft“ miteinbeziehen.

Vergleicht man die einzelnen Mandanten miteinander, fällt etwas anderes auf. Wenn man für jeden Mandanten die Altersverteilung 1-dimensional anzeigen lässt, zeigen die Balken im Bar-Chart bei den ersten drei Mandanten eine abfallende Kurve, wobei diese bei Mandanten 4 ansteigt. Es gibt also bei den ersten 3 Mandanten vor allem viele 30-35 jährige Mitglieder, in den höheren Kategorien nimmt die Zahl wieder ab. Da es sich bei diesen Mandanten um Hochschulen handelt, die in den meisten Fällen nach einer abgeschlossenen Matur besucht werden, ist es verständlich, dass es in den jüngeren Kategorien weniger Vertreter gibt. Ab 35 Jahren nehmen die Mitglieder zahlenmässig wieder ab, was mehrere Gründe haben kann. Zum einen gab es früher weniger Studenten, andererseits verlassen viele Mitglieder den Verein nach den ersten Jahren wieder. Dazu kommt, dass vor wenigen Jahren das Interesse, Teil eines solchen Netzwerks zu sein, noch nicht so gross war, wie es heute der Fall ist. Dieser Verlauf ist bei den ersten 3 Mandanten zu beobachten, nicht aber bei Mandanten 4. Zur Erinnerung: Bei diesem Mandanten handelt es sich um ein Institut, welches Diplom-Kurse in Unternehmensführung anbietet. Die Studenten des Mandanten 4 haben also bereits einige Jahre in der Berufswelt verbracht und sind bei Eintritt in die Vereinigung deutlich älter als bei den übrigen Mandanten.

5 Resultate

Im Register „Web-Statistik“, welches Auswertungen auf gesammelten Log-Files zulässt und lediglich für Mandanten 1 anwendbar ist, kann wie bereits erwähnt die Wahrscheinlichkeit der Nutzung einzelner Register und Subregister der Online-Plattform, der sog. „Aussensicht“, visualisiert werden. In dieser Visualisierung wird das Online-Portal schematisch dargestellt mit den Registern als kleine Kästchen, und daran angehängt die jeweiligen Subregister. Diese werden je nach Wahrscheinlichkeit, mit der sie besucht werden, unterschiedlich eingefärbt. Rot bedeutet 100%, Gelb 0%. Das einzige Register, welches eine Wahrscheinlichkeit von 100% aufweist, ist das Profil. Das ist so, weil der User nach dem Login default-mässig auf diese Seite geleitet wird. Interessant ist die Tatsache, dass das meist-aufgerufene Subregister nicht etwa die Suche nach anderen Mitgliedern ist, sondern die angebotenen Anlässe (Abb.34).



Abb. 34: Auswertungstyp „Wahrscheinlichkeit je Besuch“, Mandant 1

Wird eine derart überdurchschnittliche Nutzung eines bestimmten (Sub-)Registers erkannt, könnten die entsprechenden Programmeile bei der Verbesserung und Weiterentwicklung der Applikation verstärkt berücksichtigt werden. Erweiterungen in diesen Teilen der Applikation betreffen mehr Mitglieder als in weniger häufig besuchten. Somit könnten die hier visualisierten Resultate Einfluss haben auf weitere Anstrengungen, die unternommen werden, um die Online-Plattform zu verbessern und zu erweitern um damit die Mitglieder-Zufriedenheit zu steigern.

5 Resultate

6. Ausblick

Die Applikation als ganzes wie auch einzelne Programmteile werden durch die Mitarbeiter der IMS AG fortwährend verbessert und erweitert. So auch dieses Modul, das „Datacenter“. In diesem Kapitel werden einige Ideen aufgegriffen, die bereits verhandelt wurden und in naher Zukunft auch implementiert werden, deren Umsetzung aber der zeitliche Rahmen dieser Diplomarbeit nicht zulässt.

Die erste Idee beinhaltet eine Erweiterung der in jeder Säule angebotenen Druckfunktion um die Möglichkeit, die generierten PDF's in einem Archiv abzuspeichern. Im Archiv können Auswertungen, die dem User wichtig erscheinen oder die er aus anderen Gründen festhalten will, abgespeichert und wieder gelöscht werden.

Neben jedem Drucker Icon würde eine weitere Funktion erscheinen mit dem Label „Ins Archiv hinzufügen“. Wird diese betätigt, wird ähnlich wie bei der Druckfunktion ein PDF erstellt, es wird aber nicht angezeigt, sondern lediglich in einem bestimmten Ordner abgespeichert. Es folgt eine Bestätigung, dass die Auswertung dem Archiv hinzugefügt wurde. Über eine weitere Funktion mit dem Label „Archiv“ gelangt man zu einer Übersicht aller auf diese Art und Weise abgespeicherten PDFs. Diese sind in einer sortierbaren Liste dargestellt und können per Mouse-Click auf den betreffenden Eintrag geöffnet werden.

Wenn der Benutzer eine Graphik ins Archiv aufnehmen will, wird diese im PDF-Format in einem Ordner abgespeichert, der nach seiner UserID benannt ist. Für jeden Mandanten werden diese Ordner in anderen Verzeichnissen abgespeichert, um so auch hier die Mandantenfähigkeit sicherzustellen. Um das Archiv anzuzeigen, wird nun also im entsprechenden Verzeichnis nach dem Ordner gesucht, dessen Bezeichnung der UserID des Benutzers entspricht, um dessen Inhalt auszulesen. Im Archiv werden das Datum und die Art der Auswertung angezeigt. Diese Informationen werden im Namen der jeweiligen PDF's untergebracht. Ein möglicher Name wäre zum Beispiel `age_27_06_2007_20_27_52.pdf`, was auf eine Auswertung vom 27.06.2007 des Auswertungstyps Altersverteilung hinweisen würde.

Die folgende Erweiterung wird auf Kundenwunsch implementiert werden. Es geht darum, das neugeschaffene Modul noch weiter in die vorhandene Applikation zu integrieren. In Alumnitools spielen die „Einfache Suche“ und die „Erweiterte Suche“ nach Mitgliedern eine zentrale Rolle. Wurde eine Suche ausgeführt und eine Auswahl unter den Resultaten getroffen, stehen unzählige Funktionen zur Verfügung wie beispielsweise das Mutieren der persönlichen Daten, das Erstellen neuer Zugangsdaten für das Online-Portal oder das Generieren eines Massen-Versands u.a.

Nun soll die Möglichkeit geschaffen werden, die einzelnen Vertreter der Kategorien, die in den Auswertungen gezeigt werden, so in die „Einfache Suche“ oder die „Erweiterte Suche“ zu exportieren, dass die genannten Funktionen auf sie anwendbar sind.

Nebst den gezeigten Graphiken werden, wie bereits erwähnt, die für die Visualisierung der Auswertungen benötigten Listen in einer Tabelle angezeigt (Abb. 35). Dies aus dem Grund, da es oftmals schwierig ist, die exakten Zahlen aus den Graphiken herauszulesen. Um nun eine Auswahl der Kategorien für den Export zuzulassen, könnten Checkboxes am jeweiligen Zeilenanfang jener Tabelle eingefügt werden (Abb. 36). Über ein Button „Export“ könnte die einmal getroffene Auswahl an die „Einfache Suche“ weitergegeben werden, wo die gefragten Funktionen zur Verfügung stehen (Abb.37). Zwei Ansätze werden zur Lösung dieses Problems in Betracht gezogen. Die Routinen, die die erforderlichen Listen für die Visualisierung regelmässig in der Nacht generieren und in ein Textfile speichern, müssten so modifiziert werden, dass sie für jede in der Auswertung enthaltene

Kategorie eine weitere Liste mit den jeweiligen PersonenNummern erstellt. Diese würde dann bei aktiverter Checkbox der „Einfachen Suche“ weitergegeben und dort in der SQL-Abfrage verwendet. Vorteile wären ein geringer

6 Ausblick

Altersklasse [Jahre]

Mitglied
 Nichtmitglied
 Studierende
 Partner - natürliche Person
 Mittelschüler

Aktive Einträge: 40497

Anzahl	Altersklasse [Jahre]	Personenkategorie
0	20	Mitglied
10	20-25	Mitglied
1272	25-30	Mitglied
4514	30-35	Mitglied
2497	35-40	Mitglied
1296	40-45	Mitglied
727	45-50	Mitglied

Abb. 35: Detail-Liste der vorliegenden Version

Altersklasse [Jahre]

Mitglied
 Nichtmitglied
 Studierende
 Partner - natürliche Person
 Mittelschüler

Aktive Einträge: 40497

★ Exportieren

<input type="checkbox"/>	Anzahl	Altersklasse [Jahre]	Personenkategorie
<input checked="" type="checkbox"/>	0	20	Mitglied
<input checked="" type="checkbox"/>	10	20-25	Mitglied
<input checked="" type="checkbox"/>	1272	25-30	Mitglied
<input type="checkbox"/>	4514	30-35	Mitglied

Abb. 36: Detail-Liste nach der geplanten Erweiterung

DEMO

Reset

Personen-Nr. Name Vorname Volltextsuche/Person

Drucken
 Invers
 Kontakt
 Export Excel
 Zugangsdaten
 PermanentEmail
 Rechnungen

Ihre Ergebnisse Treffer pro Seite anzeigen

<input type="checkbox"/>	Aktionen	Pers.-Nr.	Pers.-Kat.	Vorname	Name	E-mail	Korr.-Adresse
<input type="checkbox"/>		40281	Nichtmitglied	Peter	windler	windler@imstrategie.ch	Kilchbergstrasse 222, 9000 St. G.
<input type="checkbox"/>		44256	Mitglied	Thomas	Meier	windler@imstrategie.ch	Chemin de Primerose 22a, 1007
<input type="checkbox"/>		44257	Mitglied	Nicole	Huber	landtwing@imstrategie.ch	Bahnhofstrasse 241, 8000 Zürich
<input type="checkbox"/>		44259	Mitglied	Thomas	Muster	thomas.muster@imstrategie.ch	Seestrasse 146, 8820 Wädenswil

Abb. 37: Exportierte Kategorien in der „Einfachen Suche“ nach der geplanten Erweiterung

Programmieraufwand und eine weniger komplexe SQL-Abfrage in der „Einfachen Suche“. Wird aber berücksichtigt, dass zum Beispiel bei Mandanten 1, bei der 2-dimensionalen Auswertung nach Altersverteilung und Personenkategorie, insgesamt 60 verschiedene Kategorien existieren, für welche je eine Liste mit Personennummern abgespeichert werden muss, scheint die Lösung nicht optimal.

In einem anderen Ansatz würden die Routinen demnach so modifiziert, dass für jede Kategorie lediglich jener Teil des WHERE-Statements, welches ihn beschreibt, abgespeichert wird. Werden nun die ausgewählten Kategorien in die

6 Ausblick

„Einfache Suche“ exportiert, müssen lediglich die jeweiligen WHERE-Statements übergeben werden, welche anschliessend OR-verknüpft in der Query verarbeitet werden. Auf diese Weise müssen nicht Listen mit Tausenden von Elementen zwischen den Files übergeben werden, sondern lediglich einzelne, kurze Strings.

7. Anhang

7.1 Abbildungsverzeichnis

Abb. 1: Alumnitools, die “Innensicht”	7
Abb. 2: Alumnitools, die “Aussensicht”	8
Abb. 3: Kundenbeziehung IM Strategies AG	13
Abb. 4: Beschreibungen der Personenkategorien, Mandant 1	15
Abb. 5: Verhältnis der Personenkategorien, Mandant 1	16
Abb. 6: Beschreibungen der Personenkategorien, Mandant 2	17
Abb. 7: Verhältnis der Personenkategorien, Mandant 2	18
Abb. 8: Beschreibungen der Personenkategorien, Mandant 3	19
Abb. 9: Verhältnis der Personenkategorien, Mandant 3	19
Abb. 10: Beschreibungen der Personenkategorien, Mandant 4	20
Abb. 11: Verhältnis der Personenkategorien, Mandant 4	20
Abb. 12: Technisches Umfeld IM Strategies AG	21
Abb. 14: Das Datacenter in Alumnitools eingebettet	24
Abb. 15: Am Visualisierungsprozess beteiligte Komponenten	25
Abb. 18: Beispiel Stacked Bar-Chart	29
Abb. 19: Beispiel Zeitreihe	29
Abb. 20: 1-dimensionale Auswertung, Mandant 1	33
Abb. 21: 2-dimensionale Auswertung, Mandant 1	34
Abb. 22: Auswertungstyp “Anzahl Logins pro Tag“, Mandant 1	37
Abb. 23: Auswertungstyp “Wahrscheinlichkeit je Besuch“, Mandant 1	38
Abb. 24: Hauptregister in der “Aussensicht”	38
Abb. 25: 1-dimensionale Auswertung, Mandant 1	41
Abb. 26: 2-dimensionale Auswertung, Mandant 1	41
Abb. 27: Zeitreihe von Zahlungseingängen, Mandant 1	43
Abb. 28: Bar-Chart von Spendeneingängen, Mandant 1	46
Abb. 29: Zeitreihe von Spendeneingängen, Mandant 1	46
Abb. 30: Zeitreihe der Personenkategorie ‘Mitglied’, Mandant 1	48
Abb. 31: Auswertungstyp “Geschlechter- /Alters-Verteilung“, Mandant 1	51
Abb. 32: Auswertungstyp “Geschlechter- /Alters-Verteilung“, Mandant 3	52
Abb. 33: Auswertungstyp “Dauer der Mitgliedschaft“, Mandant 1	53
Abb. 34: Auswertungstyp “Wahrscheinlichkeit je Besuch“, Mandant 1	54
Abb. 35: Detail-Liste der vorliegenden Version	58
Abb. 36: Detail-Liste nach der geplanten Erweiterung	58
Abb. 37: Exportierte Kategorien in der “Einfachen Suche”	58

7.2 Literaturverzeichnis

- [IMS, 2007] IM Strategies AG (2007).
www.imstrategies.ch
(besucht am 01.07.2007)
- [JFreeChart, 2007] JFreeChart (2007)
<http://www.jfree.org/jfreechart/>
(besucht am 12.01.2007)
- [LGPL, 2007] GNU Lesser General Public License
<http://www.gnu.org/licenses/lgpl.html>
(besucht am 12.01.2007)
- [Wikipedia, 2007] Wikipedia (2007).
<http://www.wikipedia.de/>
(besucht am 12.01.2007)
- [YALE, 2007] YALE (2007)
<http://sourceforge.net/projects/yale>
(besucht am 14.01.2007)

7.3 Inhalt der CD

Auf der beigelegten CD befindet sich:

- Diplomarbeit.pdf Diplomarbeit als PDF
- Zusammenfassung.pdf Zusammenfassung auf Deutsch
- Abstract.pdf Zusammenfassung auf Englisch

Ordner:

- SourceCode Entwickeltes Modul
- Bilder Alle in der Diplomarbeit verwendete Bilder
- JFreeChart Applikations-Files und Dokumentation