# **Pflichtenheft**

# Tabellenauswerter

Christian Dreier

Matr.-Nr.: 732761

# Inhaltsverzeichnis

1	Einleitung	4
	1.1 Dokumentverwaltung	4
	1.1.1 Änderungshistorie	4
	1.1.2 Abkürzungen	4
	1.1.3 Ablage des Dokuments	4
	1.2 Projekthintergrund	
	1.2.1 Projektinitiierung	
	1.2.2 Projekt-Meilensteine	
	1.3 Ziele, Rahmenbedingungen und Prämissen	
	1.3.1 Ziele	
	1.3.2 Organisation	
	1.4 Referenzdokumente	
2		
2	Beschreibung der Ausgangssituation	
	2.2 Auftraggeber	
	2.3 Zielbeschreibung	
3	Produktanforderungen	
_	3.1 Gruppen von Anforderungen	
	3.2 Funktionale Anforderungen	
	3.3 Nichtfunktionale Anforderungen	14
4	Fachliche Konzeption	15
5	Entwicklungs- und Produktionsrahmenbedingungen	27
	5.1 Entwicklungsschritte	27
	5.2 Entwicklungsergebnisse	27
	5.3 Werkzeuge	27
6	Systemschnittstellen	28
	6.1 HW/SW-Konfiguration	29
	6.2 Schnittstellen zu externen Systemen	29
	6.3 Benutzungsschnittstellen	
7	Prozessschnittstellen	
8	Einverständnis-Erklärung	
9	Appendix	32

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		3 / 33

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		4 / 33

# 1 Einleitung

# 1.1 Dokumentverwaltung

# 1.1.1 Änderungshistorie

Version	Datum	Autor	Änderungsgrund
V 0.9	31.05.2011	Christian Dreier	Initiale Version
V 1.0	14.07.11	Christian Dreier	Änderung im GUI,
			Spezifikation des Dateiformats als
			Anhang hinzugefügt

# 1.1.2 Abkürzungen

# 1.1.3 Ablage des Dokuments

Dieses Dokument ist schreibgeschützt und befindet sich im Verzeichnis:

<Projektordner>\meilensteine\m3

Pflichtenheft\_Dreier.pdf

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		5 / 33

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	-Eigenschaft.	14.07.2011 01:30
Pflichtenheft_Dreier.docx		6 / 33

# 1.2 Projekthintergrund

### 1.2.1 Projektinitiierung

Der Ursprüngliche Anlass für dieses Projekt war ein Auftrag, bestimmte Daten von bestimmten Webseiten automatisiert auszulesen und weitere Auswertungen zu ermöglichen. Im Rahmen von dem Fach Projektmanagement (BHT-Berlin, SS11, Dozent: Dipl. Ing. Dirk Hausburg) wird der einfache Prototyp, der zum Zeitpunkt der Projektinitiierung bereits existiert, soweit weiterentwickelt, dass der Auftraggeber seine Aufgaben selbstständig ausführen kann.

# 1.2.2 Projekt-Meilensteine

Das Projekt teilt sich in vier Meilensteine auf.

- M1: Projektauftrag
  Dieser Meilenstein umfasst die Teilaufgabe "Projektvorbereitung" und wurde am 18.04.2011 fertiggestellt.
- M2: Lastenheft
  Dieser Meilenstein umfasst die Teilaufgabe "Anforderungsanalyse" und wurde zusammen mit dem Lastenheft am 11.05.2011 fertiggestellt.
- M3: Pflichtenheft
  Dieser Meilenstein umfasst die Teilaufgabe "Design".
- M4: Projektabschluss
  Dieser Meilenstein umfasst die Teilaufgaben "Umsetzung" und "Projektabschluss" und soll am 22.07.2011 fertiggestellt sein.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		7 / 33

# 1.3 Ziele, Rahmenbedingungen und Prämissen

#### 1.3.1 Ziele

Ziel dieses Projektes ist es ein Programm zu implementieren, welches Daten, die Tabellenförmig in Web-Seiten angezeigt werden, einließt und zur späteren Weiterverarbeitung speichert. Das Programm, welches im Rahmen des Projekts implementiert wird, wird sich zunächst nur mit der Webseite "Asianbookie" [siehe Appendix – Hinweise – Asianbookie] sinnvoll verwenden lassen.

### 1.3.2 Organisation

Das Projektteam besteht nur aus einem Mitglied:

• Christian Dreier, E-Mail: cdreier84@googlemail.com

Zu den Stakeholders gehören:

- Immanuel Hinz (Kunde)
- Dipl. Ing. Dirk Hausburg (Dozent Projektmanagement)

## 1.4 Referenzdokumente

Neben diesem Dokument werden dem 3. Meilenstein folgende Dokumente mitgeliefert:

- Projektstrukturplan (projektstrukturplan.mm, projektstrukturplan.png)
- Kapazitätenplanung (kapazitätenplan.xlsx)
- Risikoanalyse (risikoanalyse.docx)

Alle Dokumente befinden sich beim Projektleiter (Christian Dreier).

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	-Eigenschaft.	14.07.2011 01:30
Pflichtenheft_Dreier.docx		8 / 33

# 2 Beschreibung der Ausgangssituation

# 2.1 Kurzbeschreibung der Ausgangssituation

Im WWW werden viele Informationen in Form von Tabellen dargestellt. Es wäre für verschiedene Zwecke interessant, wenn sich diese Tabellen auf dem lokalen Rechner speichern und die enthaltenen Informationen auswerten lassen würden.

Ein Problem ist, dass weder die allermeisten Webseiten, noch der Browser eine Möglichkeit bieten, die enthaltenen Tabellen z.B. nach Excel zu exportieren. Ein weiteres Problem besteht darin, dass umfangreiche Tabellen sich oft über mehrere Webseiten erstrecken, d.h. wenn man die ganze Tabelle sehen will, muss jede Seite, auf der die Tabelle fortgesetzt wird, geöffnet werden. Das kann bei mehreren Tausend Zeilen, von denen pro Webseite, vielleicht ein halbes Dutzend angezeigt werden, eine sehr umfangreiche Aufgabe sein.

Für die Lösung dieses Problems (zumindest teilweise) wäre ein Programm sinnvoll, was die Tabellen zuverlässig einlesen kann und auch Tabellen, die über mehrere Webseiten aufgespalten wurden selbstständig einließt und zusammensetzt.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	-Eigenschaft.	14.07.2011 01:30
Pflichtenheft_Dreier.docx		9 / 33

# 2.2 Auftraggeber

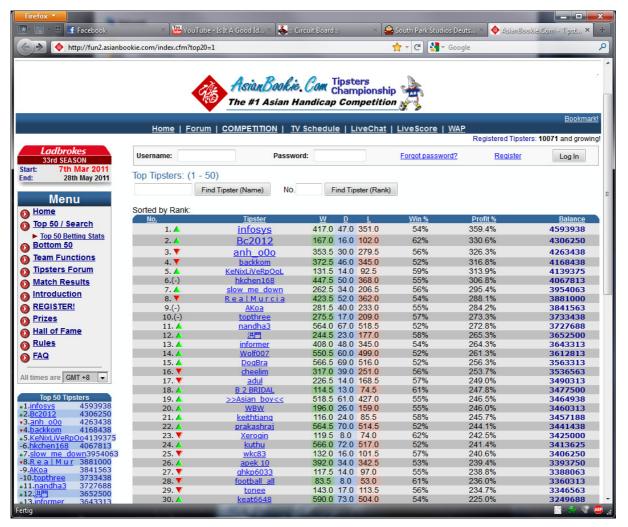


Abbildung 1

Der Auftraggeber (auch als "Kunde" bezeichnet) interessiert sich für die Webseite "Asianbookie" (siehe Abbildung 1). Bei dieser Webseite handelt es sich um eine Gemeinschaft für virtuelle Wetten. Kern der Seite (bzw. das, wofür sich der Kunde interessiert) ist eine Tabelle, die alle Teilnehmer (über 10.000) und einige Daten über ihre (Miss-) Erfolge auflistet. Die besten 50 Teilnehmer werden unter einer festen Internetadresse aufgelistet. Die restlichen Teilnehmer erreicht man, wenn man auf den richtigen Link klickt und sich dann von Teiltabelle zu Teiltabelle hangelt.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	-Eigenschaft.	14.07.2011 01:30
Pflichtenheft_Dreier.docx		10 / 33

# 2.3 Zielbeschreibung

Der Kunde möchte die Daten von *allen* Teilnehmern einmal pro Saison speichern, um sie dann über mehrere Saisons zu beobachten und sein eigenes Wettverhalten mit Hilfe der gewonnenen und ausgewerteten Daten zu optimieren.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	-Eigenschaft.	14.07.2011 01:30
Pflichtenheft_Dreier.docx		11 / 33

# 3 Produktanforderungen

# 3.1 Gruppen von Anforderungen

Das Programm teilt sich im Wesentlichen in folgende Funktionen auf:

- Tabellen von Web-Seiten einlesen
- GUI
- Eingelesene Tabellen speichern
- Mehrere Tabellen einlesen und "richtig Zusammenfügen"
- Weitere Auswertung der gewonnenen Daten

# 3.2 Funktionale Anforderungen

#### Tabellen von Web-Seiten einlesen

Die wahrscheinlich wichtigste Teilaufgabe, die bereits von den Prototypen implementiert wurde ist es überhaupt die Tabellen in den Web-Seiten zu erkennen, einzulesen und darzustellen. Für diese Aufgabe muss das Programm in der Lage sein, den HTML-Code [siehe Appendix – Hinweise – HTML] zu interpretieren, die enthaltenen Tabellen zu erkennen, in interne Datenstrukturen zu übersetzen und diese internen Datenstrukturen darzustellen.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	-Eigenschaft.	14.07.2011 01:30
Pflichtenheft_Dreier.docx		12 / 33

#### **GUI**

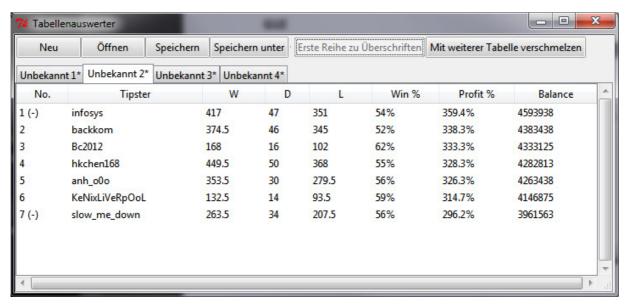


Abbildung 2: Prototyp

Wie man in Abbildung 2 erkennen kann, enthält das Programm bereits eine einfache grafische Oberfläche, über der die Tabellen dargestellt werden. Es steht dem Projektleiter vollkommen frei, wie diese GUI weiter gestaltet, bzw. verändert wird. Der Kunde möchte zunächst nur etwas haben, womit er seine Aufgabe selbstständig ausführen kann.

#### Eingelesene Tabellen speichern

Bisher ist das Programm nicht in der Lage, die dargestellten Tabellen dauerhaft zu speichern. Es gibt zwar fertige Funktionen, mit deren Hilfe sich die internen Datenstrukturen als Dateien speichern lassen, aber ihre Benutzung hat einen großen Nachteil: Wenn an diesen Datenstrukturen Veränderungen vorgenommen werden (z.B. Hinzufügen neuer Eigenschaften und Funktionen), könnte das Programm nicht mehr ohne weiteres die Dateien öffnen, die noch aus der alten Datenstruktur entstanden sind.

Deswegen muss ein s.g. Dateiformat definiert werden. Dazu muss bestimmt werden, welche Informationen überhaupt dauerhaft gespeichert werden müssen. Dann muss definiert werden, wie diese Informationen angeordnet werden. Weiterhin ist zu beachten, dass es möglich sein soll, später zusätzliche Informationen zum Dateiformat hinzuzufügen, ohne, dass die alten Dateien, die noch nicht diese Informationen enthalten, nicht mehr kompatibel mit dem überarbeiteten Programm wären (d.h. das Dateiformat muss aufwärtskompatibel sein).

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		13 / 33

### Mehrere Tabellen einlesen und "richtig Zusammenfügen"

Zur Verdeutlichung, was der Kunde genau wünscht, folgen einige Bilder und eine Erklärung:

	_				
No.	Tipster	W D L	Win %	Profit %	Balance
1. (-)	<u>backkom</u>	263.0 27.0 222.5	54%	336.3%	4362500
2. 🚣	<u>keithtiang</u>	70.0 16.0 40.5	63%	265.2%	3651563
3. 🛕	<u>infosys</u>	296.0 29.0 252.5	54%	263.6%	3636250
4. ₹	>>Asian boy<<	398.0 46.0 314.0	56%	261.7%	3617188
5. 🛕	<u>topthree</u>	249.5 15.0 190.0	57%	253.5%	3535313
6. (-)	<u>Xerogin</u>	95.0 7.0 51.0	65%	248.2%	3481875
7. ▼	<u>MoeMin</u>	104.5 9.0 68.0	61%	246.8%	3468438
			10C 0/	D 514	
No.	Tipster	W D L	Win %	Profit %	Balance
4. ▼	>>Asian_boy<<	398.0 46.0 314.0	56%	261.7%	3617188
5. 🛦	<u>topthree</u>	249.5 15.0 190.0	57%	253.5%	3535313
6. (-)	<u>Xerogin</u>	95.0 7.0 51.0	65%	248.2%	3481875
7. 🔻	<u>MoeMin</u>	104.5 9.0 68.0	61%	246.8%	3468438
8. 🚣	<u>anh_00o</u>	273.5 26.0 214.0	56%	243.2%	3432188
9. 🛕	<u>DogBra</u>	451.0 51.0 415.5	52%	239.4%	3393750
10. ▼	<u>Bc2012</u>	119.0 14.0 80.0	60%	237.9%	3379063
No.	- Tipster	W D L	Win %	Profit %	Balance
7. 🔻	MoeMin	104.5 9.0 68.0	61%	246.8%	3468438
8. 🛕	anh o0o	273.5 26.0 214.0	56%	243.2%	3432188
9. 🛕	DogBra	451.0 51.0 415.5	52%	239.4%	3393750
10. 🔻	Bc2012	119.0 14.0 80.0	60%	237.9%	3379063
11.	nandha3	447.0 50.0 419.0	52%	235.7%	3357188
12. 🗸	lhonglhong	376.0 44.0 310.0	55%	233.6%	3336250
13. 🛕	cecehao	242.5 36.0 208.0	54%	230.5%	3305313
15. 🔼	<u>cecenao</u>	242.5 30.0 200.0	3470	200.076	3303313
No.	Tipster	W D L	Win %	Profit %	Balance
10. ▼	Bc2012	119.0 14.0 80.0	60%	237.9%	3379063
11. 🛕	nandha3	447.0 50.0 419.0	52%	235.7%	3357188
12. 🛕	lhonglhong	376.0 44.0 310.0	55%	233.6%	3336250
13. 🛕	cecehao	242.5 36.0 208.0	54%	230.5%	3305313
14. 🛕	informer	297.5 30.0 250.0	54%	227.2%	3272188
15. ₹	H3r0cw	163.0 19.0 125.5	56%	223.6%	3236250
16. 🛕	getsmart	272.5 35.0 214.5	56%	222.3%	3223125

Das sind vier Tabellen, die sich jeweils auf einer eigenen Web-Seite befinden. Diese zeigen die ersten 16 Benutzer von über 10.000 an. Man gelangt zur jeweils unteren Tabelle (bzw. nächste Web-Seite), indem man auf den jeweils letzten Benutzernamen (Spalte "Tipster") klickt. Das Programm muss selbstständig diese Links folgen, um die jeweilige Tabelle einzulesen. Das soll das Programm machen, bis *alle* Benutzer (bzw. Datensätze) eingelesen wurden. Der Kunde will aber nicht viele einzelne Tabellen haben, sondern nur eine, die alle Datensätze enthält.

Deswegen müssen die eingelesenen Tabellen zur einer zusammengefügt werden. Wenn nun die Tabellen, wie sie eingelesen wurden, einfach hintereinander zusammengefügt werden würden, gäbe es viele doppelte Datensätze, was nicht gewünscht ist. Deswegen

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		14 / 33

muss das Programm dafür sorgen, dass die fertige Tabelle jeden Datensatz nur einmal enthält. In der fertigen Tabelle soll z.B. hinter den 7. Reihe direkt die 8. kommen usw.

#### Weitere Auswertung der gewonnenen Daten

Der Kunde möchte die oben beschriebene Prozedur über mehrere Saisons machen (einmal pro Saison, am Ende). Er hätte dann für jede Saison eine Tabelle. Die weitere Auswertung (so, wie es sich der Kunde zunächst wünscht) besteht darin, die Benutzer, die in den verschiedenen Saisons teilgenommen haben zu erkennen und ihre Daten zusammenzurechnen. Diese Zusammenrechnung besteht z.Z. nur darin die saisonalen Werte in den Spalten "Balance" (die auch negativ sein können), "W", "D" und "L", zu addieren.

# 3.3 Nichtfunktionale Anforderungen

Der Kunde verlangt nur, dass die o.g. Funktionen zuverlässig ausgeführt werden. Weitere Wünsche sind keine vorhanden.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		15 / 33

# 4 Fachliche Konzeption

Die Umsetzung gestaltet sich so, dass, ausgehend von den grundsätzlichen Anforderungen, zunächst die grafische Oberfläche entworfen wird. Daraus wird definiert, welche Funktionen die einzelnen Elemente haben sollen und was diese Funktionen genau machen sollen.

Unabhängig von der grafischen Oberfläche muss für die Speicherung der Daten ein Format definiert werden. Im Rahmen dieses Pflichtenheftes wird zunächst nur definiert, welche Informationen in diesem Dateiformat enthalten sind und wie diese Informationen grundsätzlich strukturiert sind.

Da die eigentliche Aufgabe, die vom Kunden verlangt wurde (mehrere Seiten einlesen und die enthaltenen Tabellen zu einer zusammenfügen), relativ komplex ist, wird diese einem eigenen Abschnitt gewidmet.

#### 4.1 Grafische Oberfläche

Hinweis: Bei den nachfolgenden Abbildungen handelt es um reine Entwürfe, die nicht genau mit dem fertigen Programm übereinstimmen müssen.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		16 / 33

# Allgemeiner Aufbau

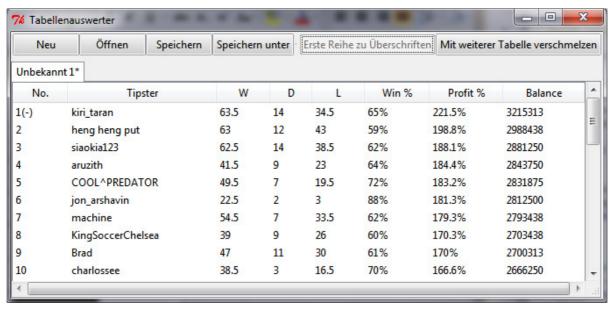


Abbildung 3: Das Hauptfenster mit einer einzelnen Tabelle

Wie in Abbildung 3 zu erkennen ist, spaltet sich das Hauptfenster im Wesentlichen in drei Teile auf. Ganz oben befindet sich die Werkzeugleiste (bzw. ein billiger Nachbau derselben). Direkt nach dem Start der Anwendung wird die Oberfläche auch nur aus dieser Werkzeugleiste bestehen.

Der größte Teil wird von der gerade angezeigten Tabelle eingenommen. Weil nach dem Einlesen von einzelnen Webseiten mehrere Tabellen erkannt worden sein können, befindet sich direkt über der Tabelle noch ein Reiterelement, mit dessen Hilfe man zwischen den verschiedenen Tabellen navigieren kann.

Die Tabelle, wie sie vom Programm angezeigt wird, enthält eine Zeile, welche die Überschriften der Spalten enthält. Wenn bei der eingelesenen Tabelle keine Zeile mit Überschriften erkannt werden konnte, enthält die Zeile generische Einträge (d.h. es wird dort "Spalte 1", "Spalte 2", … stehen). Ein Klick auf der jeweiligen Überschrift sorgt dafür, dass die ganze Tabelle nach den Werten in der jeweiligen Spalte sortiert wird. Dabei muss bei der Implementierung des Programms darauf geachtet werden, dass die Spalten, die Zahlen enthalten nach ihren Wert sortiert werden (z.B. 1, 2, 11, …) und nicht alphabetisch (in diesem Beispiel wäre das dann 1, 11, 2, …).

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		17 / 33

# Werkzeugleiste



Abbildung 4

Abbildung 4 zeigt das Programm, wie es direkt nach dem Start aussehen wird.

#### Neu

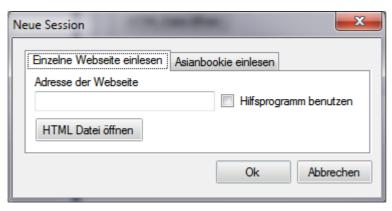


Abbildung 5

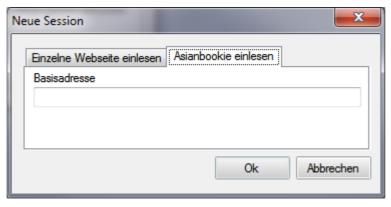


Abbildung 6

Wenn der Benutzer auf den Button "Neu" klickt, erscheint der oben abgebildete Dialog. Dieser Dialog besteht aus einem Reiterelement [siehe Appendix – Hinweise – Reiterelement], mit dessen Hilfe der Benutzer entscheiden kann, ob er nur eine einzelne Webseite einlesen will, oder automatisiert die Tabellen in der Webseite "Asianbookie" einlesen will.

Die Seite "Einzelne Webseite einlesen" enthält eine Textbox [siehe Appendix – Hinweise – Textbox], in dem der Benutzer entweder die vollständige URL [siehe Appendix – Hinweise

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verfassers unzulässig und strafbar.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		18 / 33

– URL] oder den Pfad zu einer lokal gespeicherten Datei eingibt. Wenn der Benutzer auf den Butten "HTML Datei öffnen" klickt, erscheint ein Datei-öffnen-Dialog, mit dessen Hilfe der Benutzer die Datei auswählen kann. Nachdem der Benutzer den Dialog bestätigt hat, wird der Pfad zu der Datei in die darüber liegende Textbox eingefügt. Bei manchen Webseiten, bei denen im Webbrowser Tabellen angezeigt werden, können vom Programm keine erkannt werden. Für diesen Fall gibt es den Checkbox [siehe Appendix – Hinweise – Checkbox] "Hilfsprogramm benutzen". Wenn dieser Checkbox ausgewählt ist, wird zum einlesen ein kleines Hilfsprogramm benutzt, mit dessen Hilfe sich diese Tabellen in vielen Fällen doch erkennen lassen.

Die Seite "Asianbookie einlesen" enthält nur eine Textbox, die schon mit der unter [Appendix – Hinweise – Asianbookie] genannten Adresse vorbelegt ist.

Wenn der Benutzer auf "OK" klickt, wird abhängig von der Seite des Reiterelements, die gerade dargestellt wird, die entsprechende Aktion durchgeführt. Wenn der Benutzer eine einzelne Seite einlesen will, werden folgende Aktionen durchgeführt:

- Wenn der Text in der entsprechenden Textbox mit "http://" anfängt, geht das Programm davon aus, dass die genaue Adresse zu einer Webseite angegeben wurde und versucht diese abzurufen. Wenn der Checkbox "Hilfsprogramm verwenden" gesetzt ist, wird zum Abrufen der Webseite ein Hilfsprogramm verwendet.
  - Wenn es nicht gelingt, die Webseite abzurufen, wird eine Fehlermeldung ausgegeben.
- Wenn der Text in der entsprechenden Textbox nicht mit "http://" anfängt, geht das Programm davon aus, dass es sich um eine lokal gespeicherte HTML-Datei handelt und prüft zunächst, ob die Datei mit ".htm" oder ".html" endet. Wenn das der Fall ist, wird die Datei geöffnet.
- Nachdem die Webseite eingelesen wurde, versucht das Programm die dort enthaltenen Tabellen zu erkennen und darzustellen.
- Die Darstellung erfolgt innerhalb eines Reiterelements, bei dem jeder Reiter genau eine Tabelle (bzw. was als Tabelle erkannt wurde) enthält.

Die Aktionen, die bei Bestätigung von "Asianbookie einlesen" durchgeführt werden, sind etwas umfangreicher. Deswegen wird das unter Punkt 4.3 (Asianbookie einlesen und weitere Auswertung der Daten) behandelt.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		19 / 33

#### Öffnen

Mit Hilfe dieses Buttons kann der Benutzer eine Tabelle öffnen, die vorher mit Hilfe des Programms in eine Datei gespeichert wurde. Der Dialog besteht aus einem Standard Datei-öffnen-Dialog. Als Filter wird die Dateiendung "table" voreingestellt sein, d.h. normalerweise wird dieser Dialog nur die genannten Dateien und Ordner anzeigen (dieses Verhalten ist aus vielen anderen Programmen bekannt).

### Speichern

Speichert die gerade angezeigte Tabelle in eine Datei. Wenn das Programm noch nicht den Dateinamen kennt, erscheint ein Speichern-unter-Dialog (siehe "Speichern unter").

#### Speichern unter

Zeigt immer ein Standard-Speichern-unter-Dialog vor dem Speichern an. Wenn der vom Benutzer angegebene Dateiname nicht mit "table" endet, wird diese Endung angefügt.

#### Erste Reihe zu Überschriften

Bei vielen Webseiten (u.a. Asianbookie) kann das Programm nicht erkennen, ob die erste Reihe der Tabelle Daten enthält oder aus Überschriften besteht. Da so etwas von einem menschlichen Benutzer leicht erkannt werden kann, wird dem Kontextmenü ein entsprechender Menüpunkt hinzugefügt. Wenn der Benutzer darauf klickt, wird die erste Reihe in Überschriften umgeformt und die Darstellung entsprechend angepasst.

#### Mit weiterer Tabelle verschmelzen

Diese Funktion wird in Punkt 4.3 (Asianbookie einlesen und weitere Auswertung der Daten) näher erläutert.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		20 / 33

#### 4.2 Dateiformat

Wie bereits erwähnt, ist für die dauerhafte Speicherung der gewonnenen Daten ein s.g. Dateiformat nötig. Dabei handelt es sich um eine Spezifikation darüber, wie die Daten innerhalb einer Datei angeordnet und kodiert werden. Im Rahmen des Pflichtenheftes werden zunächst nur die Anforderungen und die grundsätzliche Struktur definiert.

# Anforderungen

Beim Entwurf und Implementierung des Dateiformats muss folgendes Beachtet werden:

- Es muss erweiterbar sein, weil damit gerechnet wird, dass das Programm nach Projektende weiterentwickelt wird und dass die gesammelten Daten auch nach Jahren für den Kunden wichtig sein werden. Deswegen muss es möglich sein, dem Format später neue Elemente hinzuzufügen, ohne, dass die alten Dateien inkompatibel mit der neuen Programmversion sein würden.
- Es muss mit beliebigen Datenmengen funktionieren, weil diese Wahrscheinlich sehr groß werden. Pro Saison wird mit über 10.000 Teilnehmern gerechnet und es ist nicht bekannt, wie groß die Schnittmenge zwischen den Saisons ist. Der Kunde möchte auch die Teilnehmer speichern, die nur in einigen (bzw. auch eine) Saison(s) mitmachen. Deswegen muss bei der Implementierung darauf geachtet werden, dass die Größe der Datei, bzw. der enthaltenen Strukturen nicht durch irgendwelche Faktoren begrenzt wird.

Weiterhin wird jede Datei *eine* Tabelle enthalten und die Datei wird die Endung ".table" haben.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		21 / 33

#### Struktur

Wie bereits erwähnt, wird die Datei genau eine Tabelle enthalten. Dabei wird jede Zeile einem Datensatz entsprechen. Jeder Datensatz besteht wiederum aus "Einträgen".

Jeder Eintrag soll folgende Informationen enthalten:

- Die Daten, die vom Programm angezeigt werden. Für ihre Sortierung ist es sinnvoll verschiedene Typen von Daten zu unterscheiden. Die Spalten, die Zahlen enthalten sollen nämlich nach ihrem Wert sortiert werden und nicht alphabetisch.
  - o Dafür soll es folgende Typen geben:
    - Reine Zeichenketten: Diese werden alphabetisch sortiert.
    - Reine Zahlen: Diese werden nach ihrem Wert sortiert.
    - Kombination aus Zeichenkette und Zahl (z.B. "50%", "1. (-)"): Diese werden auch nach ihrem Wert sortiert.
- Optionaler Verweis: In den Webseiten, die das Programm ausliest, enthalten viele Einträge Verweise zu anderen Webseiten. Die enthaltene Adresse soll auch gespeichert werden.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		22 / 33

# 4.3 Asianbookie einlesen und weitere Auswertung der Daten

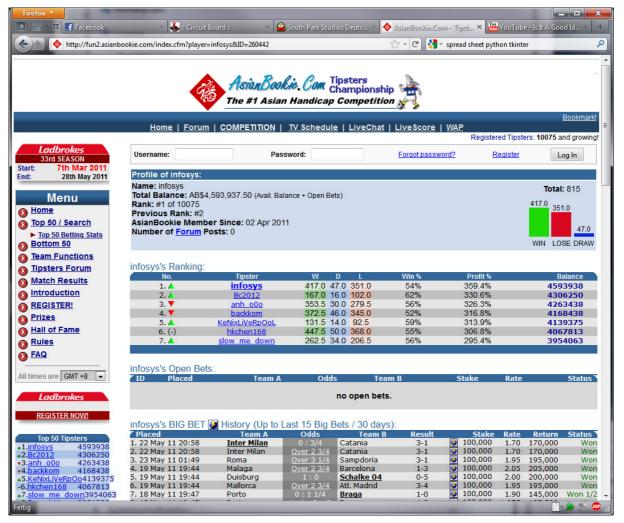


Abbildung 7: Seite, wo das Programm das Einlesen beginnen wird

Wie bereits erwähnt, will der Kunde die Daten von allen Teilnehmern, die bei "Asianbookie" mitmachen speichern und auswerten. Eine Liste der ersten 50 Teilnehmer wird unter der "Basisadresse" [siehe Appendix – Hinweise – Basisadresse] angezeigt. Die Einträge unter der Spalte "Tipster" enthalten Verweise, die jeweils zu einer Seite mit weiteren Informationen führt (Abbildung 7). Das einzige, was für die Aufgabe interessant ist, ist die Tabelle, die in Abbildung 7 unter "infosys's Ranking:" aufgeführt ist. Die dortigen Einträge enthalten wiederum Verweise zu anderen Profilseiten, bei denen die dortigen Tabellen weitere Teilnehmer enthalten. Sinnvollerweise soll das Programm immer den letzten Verweis in der jeweiligen Tabelle folgen.

Um zu verdeutlichen, was gerade gemeint ist, folgt eine Darstellung der jeweiligen Tabelle der ersten vier Seiten, die das Programm aufrufen wird:

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		23 / 33

#### 1. Seite

No.	Tipster	W	D	L	Win %	Profit %	Balance
1.	<u>infosys</u>	417.0	47.0	351.0	54%	359.4%	4593938
2.	Bc2012	167.0	16.0	102.0	62%	330.6%	4306250
3.	anh o0o	353.5	30.0	279.5	56%	326.3%	4263438
4.	<u>backkom</u>	372.5	46.0	345.0	52%	316.8%	4168438
5.	KeNixLiVeRpOoL	131.5	14.0	92.5	59%	313.9%	4139375
6. (-)	hkchen168	447.5	50.0	368.0	55%	306.8%	4067813
7.	slow me down	262.5	34.0	206.5	56%	295.4%	3954063

Diese Tabelle wird u.a. erkannt, wenn der Verweis zu dem erstplatzierten aufgerufen wird. Wie in der Formatierung angedeutet und bereits erwähnt, enthalten die Einträge unter der Spalte "Tipster" Verweise, die zu anderen Profilseiten führen. Wenn der Verweis unter Platz 7 gefolgt wird, wird die Seite aufgerufen, was hier als "2. Seite" bezeichnet ist.

#### 2. Seite

No.	Tipster	W	D	L	Win %	Profit %	Balance
4.	<u>backkom</u>	372.5	46.0	345.0	52%	316.8%	4168438
5.	KeNixLiVeRpOoL	131.5	14.0	92.5	59%	313.9%	4139375
6. (-)	hkchen168	447.5	50.0	368.0	55%	306.8%	4067813
7.	slow me down	262.5	34.0	206.5	56%	295.4%	3954063
8.	Rea Murcia	423.5	52.0	362.0	54%	288.1%	3881000
9. (-)	<u>AKoa</u>	281.5	40.0	233.0	55%	284.2%	3841563
10. (-)	<u>topthree</u>	275.5	17.0	209.0	57%	273.3%	3733438

Wie man hier erkennt enthält die Tabelle auf dieser Seite nur zum Teil neue Einträge. Wenn diese Tabelle an die erste angefügt werden soll, muss das Programm erkennen, welche Zeilen schon in der ersten Tabelle vorhanden sind und nur die neu hinzugekommenen Zeilen anfügen.

Ansonsten wird das Programm wieder den Verweis in der letzten Zeile folgen und zur "3. Seite" gelangen.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	14.07.2011 01:30	
Pflichtenheft_Dreier.docx	24 / 33	

### 3. Seite

No.	Tipster	W	D	L	Win %	Profit %	Balance
7.	slow me down	262.5	34.0	206.5	56%	295.4%	3954063
8.	Rea Murcia	423.5	52.0	362.0	54%	288.1%	3881000
9. (-)	AKoa	281.5	40.0	233.0	55%	284.2%	3841563
10. (-)	topthree	275.5	17.0	209.0	57%	273.3%	3733438
11.	nandha3	564.0	67.0	518.5	52%	272.8%	3727688
12.	<u>??</u>	244.5	23.0	177.0	58%	265.3%	3652500
13.	<u>informer</u>	408.0	48.0	345.0	54%	264.3%	3643313

Hier gilt das gleiche, wie für die 2. Seite. Wieder wird der Verweis in der letzten Zeile gefolgt, um diesmal zur "4. Seite" zu gelangen.

### 4. Seite

No.	Tipster	W	D	L	Win %	Profit %	Balance
10. (-)	topthree	275.5	17.0	209.0	57%	273.3%	3733438
11.	nandha3	564.0	67.0	518.5	52%	272.8%	3727688
12.	<u>??</u>	244.5	23.0	177.0	58%	265.3%	3652500
13.	informer	408.0	48.0	345.0	54%	264.3%	3643313
14.	Wolf007	550.5	60.0	499.0	52%	261.3%	3612813
15.	<u>DogBra</u>	566.5	69.0	516.0	52%	256.3%	3563313
16.	<u>cheelim</u>	317.0	39.0	251.0	56%	253.7%	3536563

Das Prinzip müsste inzwischen klar sein.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	14.07.2011 01:30	
Pflichtenheft_Dreier.docx		25 / 33

# Fertige Tabelle

No.	Tipster	W	D	L	Win %	Profit %	Balance
1.	<u>infosys</u>	417.0	47.0	351.0	54%	359.4%	4593938
2.	Bc2012	167.0	16.0	102.0	62%	330.6%	4306250
3.	anh o0o	353.5	30.0	279.5	56%	326.3%	4263438
4.	<u>backkom</u>	372.5	46.0	345.0	52%	316.8%	4168438
5.	KeNixLiVeRpOoL	131.5	14.0	92.5	59%	313.9%	4139375
6. (-)	hkchen168	447.5	50.0	368.0	55%	306.8%	4067813
7.	slow me down	262.5	34.0	206.5	56%	295.4%	3954063
8.	Rea Murcia	423.5	52.0	362.0	54%	288.1%	3881000
9. (-)	AKoa	281.5	40.0	233.0	55%	284.2%	3841563
10. (-)	topthree	275.5	17.0	209.0	57%	273.3%	3733438
11.	nandha3	564.0	67.0	518.5	52%	272.8%	3727688
12.	<u>??</u>	244.5	23.0	177.0	58%	265.3%	3652500
13.	informer	408.0	48.0	345.0	54%	264.3%	3643313
14.	Wolf007	550.5	60.0	499.0	52%	261.3%	3612813
15.	<u>DogBra</u>	566.5	69.0	516.0	52%	256.3%	3563313
16.	cheelim	317.0	39.0	251.0	56%	253.7%	3536563
						•	

Das ist ein Ausschnitt von der Tabelle, wie sie am Ende aussehen soll. Die fertige Tabelle wird über 10.000 Einträge umfassen (soviele, wie es Teilnehmer gibt).

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	14.07.2011 01:30	
Pflichtenheft_Dreier.docx		26 / 33

### Weitere Auswertung

Die o.g. Prozedur wird einmal pro Saison gemacht. Der Kunde möchte das über mehrere Saisons machen. Ihm reicht es aber nicht für jede Saison jeweils eine Tabelle zu haben, er möchte auch die Tabellen zusammenfügen, um dann bestimmte Werte zusammenzurechnen. Dazu muss folgendes beachtet werden:

- Es muss erkannt werden, welche Teilnehmer neu hinzugekommen sind und welche bereits vorher mitgemacht haben.
  - Neu hinzugekommene Teilnehmer werden einfach hinten eingetragen und die Werte für die Saisons, wo er nicht mitgemacht hat, werden auf 0 gesetzt.
  - Bei Teilnehmern, die bereits in vorherigen Saisons mitgemacht haben, werden die Werte, die in den Spalten "W", "D", "L" und "Balance" stehen, einfach addiert. Dabei muss berücksichtigt werden, dass die Werte auch negativ sein können, d.h. eine Addition mit einem negativen Wert ist eine Subtraktion.

Der Kunde erhofft sich, dass der Wert unter "Balance" nach mehreren Saisons aussagekräftig ist und die Teilnehmer, die dort nach mehreren Saisons einen hohen Wert haben ein Verhalten haben, den man mit Profit nachahmen kann.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	14.07.2011 01:30	
Pflichtenheft_Dreier.docx		27 / 33

# 5 Entwicklungs- und Produktionsrahmenbedingungen

# 5.1 Entwicklungsschritte

Folgendes muss für die Fertigstellung des Projekts durchgeführt werden:

- Spezifizierung des Dateiformats und Implementierung der Speicherfunktion
- GUI implementieren/fertigstellen
  - o Kontextmenü
  - Dialog um neue Session anzulegen
- "Asianbookie einlesen" implementieren
  - o Automatisches einlesen der Webseiten, die in Frage kommen
  - o Zusammenfügen der enthaltenen Tabellen zu einer großen
- "Mit weiterer Tabelle verschmelzen" implementieren

### 5.2 Entwicklungsergebnisse

Nach der Durchführung der o.g. Punkte wird es jeweils einen Zwischenversion des Programms geben, welches die jeweilige Funktion zusätzlich beinhaltet. Nach der Bearbeitung des letzten Punktes wird das Programm fertig gestellt sein.

### 5.3 Werkzeuge

Die Implementierung des Programms erfolgt mit der Entwicklungsumgebung "PyScripter" [siehe Appendix – Hinweise – PyScripter].

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	14.07.2011 01:30	
Pflichtenheft_Dreier.docx	28 / 33	

# 6 Systemschnittstellen

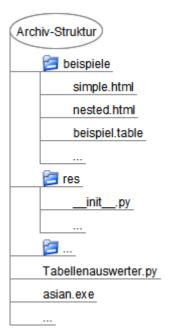


Abbildung 8: Vorschau auf das auszuliefernde ZIP-Archiv

Das fertige Projekt wird als ZIP-Archiv ausgeliefert. Der Benutzer muss dieses Archiv in einen Ordner seiner Wahl entpacken. Abbildung 8 deutet an, wie der Inhalt des ZIP-Archivs aussehen wird. Wichtig für den Benutzer ist vor allem die Datei "Tabellenauswert.py". Ein Doppelklick auf diese Datei startet das Programm (inklusive grafischer Oberfläche) und der Benutzer kann dann seine Aufgaben ausführen. Im Ordner "beispiele" befinden sich einige Dateien, die sich vom Benutzer mit Hilfe des Programms öffnen lassen.

Alle anderen Dateien, die sich im Archiv befinden, werden vom Programm benötigt und sollten nicht vom Benutzer angefasst werden. Weiterhin ist zu beachten, dass die obige Darstellung nur als Andeutung zur Erklärung zu sehen ist und das fertige Archiv noch mehrere weitere Dateien umfassen wird.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	14.07.2011 01:30	
Pflichtenheft_Dreier.docx		29 / 33

# 6.1 HW/SW-Konfiguration

Der Benutzer braucht folgende Software um das Programm auszuführen:

- Python (min. Version 3.2) [siehe Appendix Hinweise Python]
- .Net (min. Version 4.0)
  - o Das führt dazu, dass das Programm nur unter Windows laufen wird.

# 6.2 Schnittstellen zu externen Systemen

Zum Einlesen der Webseiten ist eine funktionierende Internetverbindung nötig.

# 6.3 Benutzungsschnittstellen

Das Programm wird sich nach dem Start mit einer grafischen Oberfläche präsentieren, die sich mit Hilfe der Maus bedienen lässt.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	14.07.2011 01:30	
Pflichtenheft_Dreier.docx		30 / 33

# 7 Prozessschnittstellen

Der Prozess, der durch dieses Programm ermöglicht wird, gestaltet sich folgendermaßen: Zuerst werden jeweils am Ende der Saison die Daten von allen Teilnehmern gespeichert. Die Daten der jeweiligen Saisons werden dann mit Hilfe des Programms zu einer Gesamtliste zusammengeführt. Der Kunde erhofft sich, dass wenn er genug Saisons zusammen hat, dass die Daten, die in der Gesamtliste gesammelt und zusammen gerechnet wurden, aussagekräftig genug sind, um seinen eigenen Verhalten daran zu orientieren.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument	14.07.2011 01:30	
Pflichtenheft_Dreier.docx		31 / 33

# 8 Einverständnis-Erklärung

Die Parteien bestätigen, dass der Inhalt dieses Pflichtenheftes die jeweiligen vertraglichen Verpflichtungen darstellt. Es besteht gemeinsames Einverständnis, dass das Pflichtenheft im Laufe des Projektes einvernehmlich schriftlich geändert werden kann.

Unterschrift:	 Datum:
Name:	
Unterschrift:	 Datum:
Name:	
Unterschrift:	 Datum:
Name:	
Unterschrift:	 Datum:
Name:	

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verfassers unzulässig und strafbar.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		32 / 33

# 9 Appendix

#### Hinweise

#### Asianbookie

Web-Seite, dessen Daten der Kunde auslesen will. Die Daten, die ihn interessieren, sind über den folgenden Link zu erreichen:

http://fun2.asianbookie.com/index.cfm?top20=1

#### Basisadresse

Die Adresse, die unter "Asianbookie" erwähnt wurde.

### Python

Eine interpretierte Programmiersprache, d.h. Programme, die in dieser Sprache geschrieben wurden, funktionieren i.a. nur mit der Laufzeitumgebung. Diese Laufzeitumgebung und weitere Informationen erhält man unter dieser Adresse:

http://www.python.org/

#### PyScripter

Kostenlose Entwicklungsumgebung für die Programmiersprache Python. Kann unter dieser Adresse heruntergeladen werden:

http://code.google.com/p/pyscripter/

#### HTML

Hyper Text Markup Language - Eine s.g. Auszeichnungssprache (d.h. es enthält nur Formatierungsanweisungen und wird deswegen nicht als Programmiersprache bezeichnet), die dafür benutzt wird, Web-Seiten zu gestalten.

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verfassers unzulässig und strafbar.

BHT-Berlin	Christian Dreier	Tabellenauswerter
Fehler! Unbekannter Name für Dokument-Eigenschaft.		14.07.2011 01:30
Pflichtenheft_Dreier.docx		33 / 33

#### URI

Uniform Resource Locator - Anderes Wort für "Internetadresse". Eine vollständige URL besteht immer aus den verwendeten Protokoll (z.B. HTTP) und die vollständige Adresse zu der Ressource.

Ein Beispiel: http://www.beispiel.de/pfad/zur/datei.html

Die meisten Browser haben bestimmte Konventionen implementiert, mit deren Hilfe unvollständige Adressen vervollständigt werden - z.B. aus "www.beispiel.de/bla.html" wird "http://www.beispiel.de/bla.html".

#### **Textbox**

Element in grafischen Oberflächen, mit dessen Hilfe vom Benutzer kleinere Mengen Text eingegeben werden kann.

#### Checkbox

Element in grafischen Oberflächen, mit dessen Hilfe der Benutzer Häkchen setzen kann.

#### Reiterelement

Element in grafischen Oberflächen, mit dessen Hilfe mehrere Elemente in einer gemeinsamen Fläche angeordnet werden können. Es wird immer nur ein Element gleichzeitig angezeigt. Das Wort "Reiterelement" wird hier auch für die Schaltelemente benutzt, mit deren Hilfe der Benutzer zwischen den Elementen wechseln kann.