



OpenOffice.org

***Der Datenpilot in
OpenOffice.org Calc***

Vortrag von Stefan Weigel

Copyright

Copyright und Lizenzen

Dieses Dokument unterliegt dem Copyright ©2008 seiner Autoren und Beitragenden, wie sie im entsprechenden Abschnitt aufgelistet sind. Alle Markennamen innerhalb dieses Dokuments gehören ihren legitimen Besitzern.

Sie können es unter der Voraussetzung verbreiten und/oder modifizieren, dass Sie sich mit den Bedingungen einer der (oder aller) folgenden Lizenzen einverstanden erklären:

- GNU General Public License (GPL), Version 2 oder später (<http://www.gnu.org/licenses/gpl.html>).
- Creative Commons Attribution License (CCAL), Version 2.0 oder später (<http://creativecommons.org/licenses/by/2.0/>).
- Public Documentation License (PDL), Version 1 oder später:
Public Documentation License Notice
The contents of this Documentation are subject to the Public Documentation License Version 1.0 (the "License"); you may only use this Documentation if you comply with the terms of this License. A copy of the License is available at <http://www.openoffice.org/licenses/PDL.html>.

Der Titel der Originaldokumentation ist „[OpenOffice.org, Der Datenpilot in OpenOffice.org Calc, Vortrag von Stefan Weigel](#)“.

Der ursprüngliche Autor der Originaldokumentation ist Stefan Weigel. Er kann unter `sweigel@openoffice.org` kontaktiert werden.

Veröffentlichung und Softwareversion

Dieses Dokument wurde am 11. April 2004 erstmals veröffentlicht. Die Beschreibung bezieht sich auf Version 2.4 von OpenOffice.org

Inhaltsverzeichnis

Copyright	i
Copyright und Lizenzen	i
Veröffentlichung und Softwareversion	i
Abstract	1
Über OpenOffice.org	1
Fallbeispiel Umsatzübersicht	2
Praktische Probleme / Fragestellungen	2
Lösung	3
Vorteile	4
Fallbeispiel Arbeitszeiterfassung	7
Praktische Probleme / Fragestellungen	8
Lösung	9
Unterschied / Vorteile	11
Fallbeispiel Häufigkeiten und Verteilung	13
Lösung mit einer Matrixformel	15
Lösung mit dem Datenpiloten	15
Vorteile	17
Verschenktes Potenzial	20
Aufruf	20
Neue Features	21
Seit OpenOffice.org 2.3 (September 2007)	21
Aktuell in OpenOffice.org 2.4 (März 2008)	21
Zu erwarten in OpenOffice.org 3.0 (Sept. 2008)	21
Offene Wünsche	21

Abstract

Im Support für Office-Software tauchen regelmäßig Anwenderfragen auf, die aus ungeschickten Lösungsansätzen für alltägliche Aufgaben in der Tabellenkalkulation resultieren. In Unkenntnis einfacher und gleichermaßen leistungsfähiger Konzepte und Tools werden oft komplexe Spreadsheets mit monströsen Formelkonstruktionen gebildet. Solider, einfacher und flexibler ist meist eine Lösung mit dem Datenpiloten (auch als „Pivot-Tabelle“ bekannt). Zu unrecht wird der Datenpilot oft als ein Werkzeug für fortgeschrittene Benutzer angesehen. Im Gegenteil: Der Datenpilot setzt kaum Vorkenntnisse voraus und wüssten auch schon Anfänger davon, bliebe ihnen so manche vermeidbare Arbeitsstunde erspart.

Der Vortrag demonstriert dies anhand von Fallbeispielen. Vorhandene, neue und angekündigte, aber auch noch fehlende Features des Datenpiloten von OpenOffice.org Calc werden vorgeführt und diskutiert. Es ist auch ein Aufruf an Trainer und Autoren, den oft kümmerlich behandelten Datenpiloten in Kursen und Büchern entsprechend zu kommunizieren.

Über OpenOffice.org

Das Projekt OpenOffice.org ist eine internationale Gemeinschaft von Freiwilligen und Sponsoren. OpenOffice.org entwickelt die gleichnamige führende Open Source Office Suite. Das Programm enthält eine Textverarbeitung, eine Tabellenkalkulation, ein Präsentationsprogramm, ein Zeichenmodul, ein Datenbankmodul und weitere Komponenten; es benutzt das nach ISO/IEC 26300 standardisierte OpenDocument-Dateiformat (ODF) als Standardformat und unterstützt eine Vielzahl weiterer Dokumentenformate. OpenOffice.org läuft auf allen wichtigen Plattformen, einschließlich Windows, Linux, Solaris, Mac OS X, und es ist in über 80 Sprachen erhältlich. OpenOffice.org ist für private und kommerzielle Nutzung lizenzkostenfrei unter der Open Source-Lizenz LGPL verfügbar.

Fallbeispiel Umsatzübersicht

Ein typisches Einführungsbeispiel in Kursen und Büchern für Anfänger der Tabellenkalkulation ist eine einfache Umsatzübersicht.

Anhand des Beispiels werden die Anwender mit der grundsätzlichen Bedienung der Oberfläche und dem Eingeben von Texten und Zahlen in die Zellen vertraut gemacht. Nützliche Hilfen wie „AutoFüllen“ und „Drag&Drop“ können schon an dieser Stelle demonstriert werden. Vor allem aber, geht es um die Darstellung der rechnerischen Verknüpfung von Zellen über Formeln, wie beispielsweise das Addieren mit dem Plus-Operator oder der Formel SUMME.

	A	B	C	D	E	F	G
1							
2		Erlebnis-Reisen					
3							
4			Jan	Feb	Mrz	Apr	Summe
5		Golf	1000	1100	1200	1300	4600
6		Tennis	2000	2200	2400	2600	9200
7		Segeln	3300	3300	3300	3300	13200
8		Summe	=SUMME(C5:C7)		6900	7200	27000
9							

Abbildung 1: gängiges Anfängerbeispiel

Diese kleine Übung mag für den ersten Umgang mit dem Programm geeignet sein, es behandelt aber nur einen kleinen Teil der damit in der Büropraxis verbundenen Arbeitsaufgabe. Um eine solche Umsatzübersicht zu erstellen, braucht man freilich auch das zugrunde liegende Datenmaterial. Dies liegt normalerweise in Form von einzelnen Geschäftsvorfällen vor.

Das heißt, bevor man sich der Tabellenkalkulation als Werkzeug für die Umsatzübersicht bedient, muss man für obiges Beispiel zuerst zahlreiche Einzelumsätze aus Listen oder Belegsammlungen zu zwölf Teilsummen addieren, die dann in die Zellen C5 bis F7 eingetippt werden.

Praktische Probleme / Fragestellungen

1. Zur Darstellung der im Laufe der Zeit hinzu kommenden Werte (Mai, Juni, ...) muss jeweils eine Spalte hinzugefügt werden. Das heißt, im Zuge der Anwendung muss laufend die Struktur des Kalkulationsblattes modifiziert werden. Das ist nicht nur vom theoretischen Arbeitskonzept her unschön, sondern wirft auch praktische Fragen auf, zum Beispiel: Wie verhalten sich die Bezüge in den Summenformeln, wenn Spalten oder Zeilen eingefügt werden?
2. Das Layout, nach dem die Zeitachse waagrecht nach rechts verläuft, könnte sich nach Ergänzung weiterer Monate als ungünstig erweisen. Besser wäre aus Platzgründen vielleicht eine Ausrichtung mit einer senkrecht nach unten verlaufenden Zeitachse. Wie lässt sich die Tabelle entsprechend transponieren? Muss dazu alles neu eingegeben werden?
3. Nehmen wir an, das Management verlangt unerwartet und fallweise eine zusätzliche Aufgliederung nach Vertriebsregionen und/oder eine Einschränkung der ausgewiesenen Umsätze auf die Teilbeträge eines bestimmten Vertriebsrepräsentanten. Nun muss man also aus den einzelnen Geschäftsvorfällen erneut ma-

nuell Teilsummen bilden und mehrere Tabellen in unterschiedlichen Varianten herstellen.

4. Ist es überhaupt akzeptabel und realistisch, dass man so eine Übersicht erstellt, indem die einzutippenden Werte durch manuelle Addition der Einzelumsätze ermittelt? Das ist doch unendlich viel Arbeit und extrem fehleranfällig!

Lösung

Der eigentlich wesentliche Inhalt der Arbeitsaufgabe aus unserem Beispiel ist das Addieren der Umsätze je Monat und Sparte. Um dies mit zu erfassen, holen wir uns das zugrunde liegende Datenmaterial in OpenOffice.org Calc herein. Entweder erfassen wir die Einzelumsätze von Haus aus in Calc oder wir importieren eine entsprechende Liste zum Beispiel aus dem Buchhaltungssystem.

In jedem Fall gehen wir von einer fortlaufenden Tabelle in Calc aus, die schlicht alle Geschäftsvorfälle in primitiver Form protokolliert:

	A	B	C	D	E
1	Datum	Umsatz	Sparte	Region	Mitarbeiter
2	11.01.08	2.236 €	Tennis	Ost	Brigitte
3	01.01.08	3.210 €	Golf	West	Ute
4	08.03.08	1.236 €	Golf	West	Brigitte
5	25.05.08	646 €	Segeln	West	Kurt
6	26.01.08	1.191 €	Golf	West	Fritz
7	09.06.08	4.638 €	Tennis	Süd	Kurt
8	07.03.08	2.373 €	Golf	Nord	Ute
9	08.06.08	3.045 €	Golf	Nord	Kurt
10	22.04.08	4.496 €	Segeln	Ost	Hans
11	07.04.08	2.418 €	Golf	Ost	Fritz
12	18.03.08	1.712 €	Tennis	Süd	Kurt
13	28.06.08	4.071 €	Segeln	Nord	Brigitte
14	24.05.08	953 €	Tennis	Nord	Hans
15	31.01.08	3.452 €	Golf	West	Hans
16	15.01.08	1.064 €	Golf	Ost	Ute
17	15.04.08	3.487 €	Tennis	Nord	Kurt
18	31.01.08	4.169 €	Golf	West	Kurt
19	06.05.08	826 €	Segeln	West	Fritz
20	15.04.08	4.108 €	Tennis	West	Ute
21	02.03.08	876 €	Tennis	West	Hans
22	22.02.08	3.065 €	Tennis	West	Ute
23	01.04.08	2.634 €	Golf	Ost	Fritz
24	13.03.08	3.446 €	Segeln	West	Brigitte
25	12.04.08	4.606 €	Tennis	Nord	Hans
26	04.01.08	4.697 €	Segeln	Nord	Hans
27	24.03.08	2.554 €	Golf	West	Fritz
28	30.06.08	4.280 €	Golf	Ost	Fritz
29	29.03.08	4.537 €	Tennis	Ost	Hans
30	09.05.08	2.573 €	Golf	Ost	Fritz
31	20.01.08	3.452 €	Golf	Süd	Hans
32	17.06.08	4.234 €	Tennis	Nord	Kurt
33	03.01.08	2.051 €	Golf	Ost	Hans
34	21.01.08	1.661 €	Tennis	Ost	Kurt
35	24.02.08	470 €	Segeln	West	Brigitte

Abbildung 2: Datenbasis in Calc

Die Anwendung des Datenpiloten erfordert nun genau 12 Mausklicks und liefert in wenigen Sekunden die folgende Auswertung:

	A	B	C	D	E	F	G	H	I
1	Filter								
2									
3	Summe - Ur	Datum							
4	Sparte	Jan	Feb	Mrz	Apr	Mai	Jun	Gesamt Ergebnis	
5	Golf	18.230 €	10.637 €	27.138 €	28.981 €	28.460 €	34.569 €	148.015 €	
6	Segeln	27.418 €	12.794 €	31.901 €	24.272 €	22.049 €	35.021 €	153.455 €	
7	Tennis	26.734 €	12.428 €	21.630 €	33.823 €	7.926 €	17.586 €	120.127 €	
8	Gesamt Er	72.382 €	35.859 €	80.669 €	87.076 €	58.435 €	87.176 €	421.597 €	
9									
10									

Abbildung 3: sekundenschnelle Auswertung mit dem Datenpiloten

Im Ergebnis erkennt man wieder das Anfängerbeispiel von oben. Es ist kinderleicht herzustellen, ohne weitere Grundkenntnisse in der Tabellenkalkulation. (Keine einzige Formel muss erstellt werden.)

Vorteile

1. Kein manuelles Zusammenaddieren der Einzelwerte im Vorfeld erforderlich. Kein manuelles Eintragen der Teilbeträge. Weniger Arbeit. Weniger Fehler.

Anmerkung

Für manch einen Zuhörer mag sich der Vortrag schon bis hierher gelohnt haben. Man muss sich klar machen, wie viel Arbeit der Datenpilot ersparen kann, vor allem für Leute, die regelmäßig solche Umsatzaufstellungen u. ä. anfertigen müssen.

Weitere, nicht minder nützliche Leistungen des Datenpiloten sind:

2. Das Layout ist flexibel: Monate waagrecht und Sparten senkrecht oder umgekehrt: 2 Mausklicks
3. Zusätzliche Unterscheidungskriterien sofort verfügbar
4. Vielfältige Auswertemöglichkeiten, zum Beispiel: Anzahl oder Mittelwert statt Summe, kumulierte Werte, Reihenvergleich und mehr.

	A	B	C
1	Filter		
2			
3	Datum	Sparte	
4	Jan	Golf	18.230 €
5		Segeln	27.418 €
6		Tennis	26.734 €
7	Feb	Golf	10.637 €
8		Segeln	12.794 €
9		Tennis	12.428 €
10	Mrz	Golf	27.138 €
11		Segeln	31.901 €
12		Tennis	21.630 €
13	Apr	Golf	28.981 €
14		Segeln	24.272 €
15		Tennis	33.823 €
16	Mai	Golf	28.460 €
17		Segeln	22.049 €
18		Tennis	7.926 €
19	Jun	Golf	34.569 €
20		Segeln	35.021 €
21		Tennis	17.586 €
22	Gesamt Ergebnis		421.597 €
23			

Abbildung 4: Layoutvariante mit buchstäblich einem Mausklick

	A	B	C	D	E	F
1	Filter					
2						
3	Summe - Ur	Sparte				
4	Datum	Golf	Segeln	Tennis	Gesamt Ergebnis	
5	Jan	18.230 €	27.418 €	26.734 €	72.382 €	
6	Feb	10.637 €	12.794 €	12.428 €	35.859 €	
7	Mrz	27.138 €	31.901 €	21.630 €	80.669 €	
8	Apr	28.981 €	24.272 €	33.823 €	87.076 €	
9	Mai	28.460 €	22.049 €	7.926 €	58.435 €	
10	Jun	34.569 €	35.021 €	17.586 €	87.176 €	
11	Gesamt Er	148.015 €	153.455 €	120.127 €	421.597 €	
12						
13						

Abbildung 5: Transponiertes Layout des ursprünglichen Beispiels

Fallbeispiel Umsatzübersicht

	A	B	C	D	E	F	G
1	Filter						
2							
3	Summe - Umsatz		Sparte				
4	Region	Datum	Golf	Segeln	Tennis	Gesamt Ergebnis	
5	Nord	Jan		6.385 €	20.314 €	26.699 €	
6		Feb	7.154 €	2.982 €		10.136 €	
7		Mrz	2.488 €	5.638 €	3.365 €	11.491 €	
8		Apr		9.684 €	7.071 €	16.755 €	
9		Mai	2.764 €	3.656 €		6.420 €	
10		Jun	5.417 €	2.153 €	4.757 €	12.327 €	
11	Ost	Jan	6.888 €	652 €		7.540 €	
12		Feb	3.483 €	4.117 €	3.100 €	10.700 €	
13		Mrz	10.797 €	4.132 €	7.307 €	22.236 €	
14		Apr	15.096 €		2.927 €	18.023 €	
15		Mai	10.370 €	8.575 €	4.042 €	22.987 €	
16		Jun	10.360 €	5.478 €	8.397 €	24.235 €	
17	Süd	Jan	6.384 €	12.676 €	4.398 €	23.458 €	
18		Feb			9.328 €	9.328 €	
19		Mrz	4.909 €	14.309 €	7.186 €	26.404 €	
20		Apr	8.316 €	7.746 €	15.732 €	31.794 €	
21		Mai	13.101 €	5.261 €		18.362 €	
22		Jun	7.061 €	8.749 €	4.432 €	20.242 €	
23	West	Jan	4.958 €	7.705 €	2.022 €	14.685 €	
24		Feb		5.695 €		5.695 €	
25		Mrz	8.944 €	7.822 €	3.772 €	20.538 €	
26		Apr	5.569 €	6.842 €	8.093 €	20.504 €	
27		Mai	2.225 €	4.557 €	3.884 €	10.666 €	
28		Jun	11.731 €	18.641 €		30.372 €	
29	Gesamt Ergebnis		148.015 €	153.455 €	120.127 €	421.597 €	
30							
31							

Abbildung 6: Zusätzliche Aufschlüsselung nach Region, nachträglich hinzugefügt

	A	B	C	D	E	F
1	Filter					
2	Mitarbeiter	Hans				
3		- alle -				
4	Summe - Umsatz	Brigitte				
5	Datum	Fritz	Segeln	Tennis	Gesamt Ergebnis	
6	Jan	Hans	4.974 €	2.668 €	16.915 €	
7	Feb	Kurt		384 €	4.543 €	
8	Mrz	Ute	3.592 €	3.592 €	12.879 €	
9	Apr		13.355 €	7.956 €	28.352 €	
10	Mai		297 €		4.105 €	
11	Jun		8.318 €	266 €	18.555 €	
12	Gesamt Ergebnis		30.536 €	14.866 €	85.349 €	
13						
14						
15						

Abbildung 7: Selektion der Teilsummen für einzelne Mitarbeiter

Fallbeispiel Arbeitszeiterfassung

Dieses Beispiel stammt aus der Beratungspraxis und taucht in verschiedenen Variationen regelmäßig im Usersupport auf. Der Anwender steht vor der Aufgabe, für einen oder mehrere Mitarbeiter eine Arbeitszeiterfassung zu realisieren.

Eine typische Herangehensweise ist das Erzeugen von je einem Tabellenblatt pro Monat und einem Summenblatt für das gesamte Jahr. Pro Mitarbeiter und Jahr wird stets eine eigene Datei erzeugt.

The screenshot shows a spreadsheet titled 'Arbeitszeit - OpenOffice.org Calc'. The active window displays a monthly work time recording for Erika Mustermann in January 2008. The spreadsheet is structured as follows:

Arbeitszeiterfassung für Erika Mustermann					
Januar 2008					
1					
2					
3	1				
4	Datum	kommt	geht	Pause	Stunden
5	01.01.08	07:45	15:30	00:30	7,25
6	02.01.08	08:15	18:30	01:00	9,25
7	03.01.08	08:00	15:15	00:30	6,75
8	04.01.08	07:45	19:30	01:00	10,75
9	05.01.08				
10	06.01.08				
11	07.01.08	08:30	15:45	00:30	6,75
12	08.01.08	08:45	15:15	00:30	6,00
13	09.01.08	08:00	13:45	00:00	5,75
14	10.01.08	08:45	19:15	01:00	9,50
15	11.01.08	09:00	17:30	00:30	8,00
16	12.01.08				
17	13.01.08				
18	14.01.08	09:00	18:15	00:30	8,75
19	15.01.08	07:15	16:15	00:30	8,50
20	16.01.08	08:15	18:45	01:00	9,50
21	17.01.08	07:00	15:45	00:30	8,25

The spreadsheet interface shows the formula bar with the formula `=SUMME(F5:F35)` and the status bar indicating the total hours for the month: **Summe=187,50**.

Abbildung 8: Monatsblatt für Zeiterfassung eines Mitarbeiters

	A	B	C	D	E	F	G
1							
2		Arbeitszeiterfassung für Erika Mustermann					
3							
4		2008					
5							
6		Januar	187,50				
7		Februar	183,50				
8		März	177,75				
9		April	169,50				
10		Mai	161,50				
11		Juni	0,00				
12		Juli	0,00				
13		August					
14		September					
15		Oktober					
16		November					
17		Dezember					
18		Summe	879,75				
19							
20							
21							

Abbildung 9: Jahresübersicht für Zeiterfassung eines Mitarbeiters

Praktische Probleme / Fragestellungen

1. Es ist sehr mühselig die Datei zur Arbeitszeiterfassung zu erstellen: 12 Tabellenblätter, die aus einer Rohvorlage kopiert werden und Monat für Monat anzupassen sind. Dazu ein Summenblatt mit Bezügen quer auf die anderen Tabellenblätter. Nicht selten suchen Anwender nach einer Makroprogrammierung, die das arbeitsintensive Erzeugen solcher Dateien automatisieren soll.
2. Die dargestellte Datei enthält die Daten eines einzelnen Mitarbeiters. Wie kann man die Daten aller Mitarbeiter zusammenführen, sodass eine Zusammenfassung der Arbeitszeit auf Abteilungsebene und für das Gesamtunternehmen möglich wird?
3. Wie lassen sich Mitarbeiter und/oder Abteilungen untereinander vergleichen?
4. Die dargestellte Datei enthält die Daten für ein Geschäftsjahr. Wie ist ein Vergleich zu Zeiträumen aus früheren Jahren möglich?

Lösung

Um die Aufgabe mit dem Datenpiloten zu lösen, erfassen wir alle Daten in einer einzigen, für alle Zeit fortlaufenden Tabelle in Calc. Dies geschieht entweder manuell oder durch einen Import von einer elektronischen Stempeluhr.

In einfachen Fällen führt jeder Mitarbeiter seine Zeiterfassung selbst manuell. Sollen Auswertungen gemacht werden, die sich auf mehrere Mitarbeiter, Abteilungen oder das Unternehmen insgesamt beziehen, so sind die Tabelleninhalte der einzelnen Mitarbeiter durch Kopieren in einer großen Gesamttabelle zusammenzuführen.

	A	B	C	D	E
1	Datum	Name	kommt	geht	Stunden
2	01.01.2008	Brigitte	10:15	16:15	6,00
3	01.01.2008	Fritz	11:00	18:45	7,00
4	01.01.2008	Hans	8:15	19:00	10,00
5	01.01.2008	Kurt	11:15	14:00	2,75
6	01.01.2008	Ute	11:00	13:45	2,75
7	02.01.2008	Brigitte	12:00	18:45	6,00
8	02.01.2008	Fritz	7:45	13:45	6,00
9	02.01.2008	Hans	8:30	17:30	8,25
10	02.01.2008	Kurt	10:45	18:30	7,00
11	02.01.2008	Ute	9:15	16:00	6,00
12	03.01.2008	Brigitte	9:00	18:30	8,75
13	03.01.2008	Fritz	9:00	14:30	5,50
14	03.01.2008	Hans	7:30	18:00	9,75
15	03.01.2008	Kurt	7:30	17:45	9,50
16	03.01.2008	Ute	8:30	18:30	9,25
17	06.01.2008	Brigitte	9:30	17:30	7,25
18	06.01.2008	Fritz	11:00	14:30	3,50
19	06.01.2008	Hans	9:30	16:15	6,00
20	06.01.2008	Kurt	10:45	15:45	5,00
21	06.01.2008	Ute	10:30	15:00	4,50
22	07.01.2008	Brigitte	10:30	15:15	4,75
23	07.01.2008	Fritz	7:45	17:30	9,00
24	07.01.2008	Hans	9:45	15:45	6,00
25	07.01.2008	Kurt	8:00	13:45	5,75
26	07.01.2008	Ute	8:00	14:30	5,75
27	08.01.2008	Brigitte	8:00	16:00	7,25
28	08.01.2008	Fritz	8:15	13:00	4,75
29	08.01.2008	Hans	7:30	16:00	7,75
30	08.01.2008	Kurt	10:00	14:45	4,75
31	08.01.2008	Ute	7:15	15:15	7,25
32	09.01.2008	Brigitte	9:15	16:45	6,75
33	09.01.2008	Fritz	10:30	13:00	2,50

Abbildung 10: Datenbasis in Calc

Die Anwendung des Datenpiloten erfordert nun genau 12 Mausklicks und liefert in wenigen Sekunden die folgende Auswertung:

Fallbeispiel Arbeitszeiterfassung

	A	B	C	D	E	F	G
1	Filter						
2							
3	Summe - ▶	Name					
4	Datum	Brigitte	Fritz	Hans	Kurt	Ute	Gesamt Ergebnis
5	01.01.08	6,00	7,00	10,00	2,75	2,75	28,50
6	02.01.08	6,00	6,00	8,25	7,00	6,00	33,25
7	03.01.08	8,75	5,50	9,75	9,50	9,25	42,75
8	06.01.08	7,25	3,50	6,00	5,00	4,50	26,25
9	07.01.08	4,75	9,00	6,00	5,75	5,75	31,25
10	08.01.08	7,25	4,75	7,75	4,75	7,25	31,75
11	09.01.08	6,75	2,50	5,50	6,00	5,00	25,75
12	10.01.08	7,00	8,75	3,75	10,50	3,50	33,50

Abbildung 11: sekundenschnelle Auswertung mit den Datenpiloten

Das Ergebnis leistet wesentlich mehr, als die Lösung mit klassischer, formelbasierter Tabellenkalkulation.

	A	B	C	D	E	F	G
1	Filter						
2							
3	Summe - ▶	Name					
4	Datum	Brigitte	Fritz	Hans	Kurt	Ute	Gesamt Ergebnis
5	Jan	150,75	137,00	166,50	141,75	137,00	733,00
6	Feb	131,50	139,25	112,25	119,75	116,75	619,50
7	Mrz	140,25	154,50	135,75	134,75	141,00	706,25
8	Apr	140,50	128,50	131,75	131,25	144,00	676,00
9	Mai	126,00	134,75	116,00	121,25	140,00	638,00
10	Jun	127,75	128,00	118,25	120,25	121,25	615,50
11	Gesamt B	816,75	822,00	780,50	769,00	800,00	3988,25

Abbildung 12: Monatssummen

	A	B	C	D	E	F	G
1	Filter						
2							
3	Summe - ▶	Name					
4	Months	Brigitte	Fritz	Hans	Kurt	Ute	Gesamt Ergebnis
5	Jan	20,57%	18,69%	22,71%	19,34%	18,69%	100,00%
6	Feb	21,23%	22,48%	18,12%	19,33%	18,85%	100,00%
7	Mrz	19,86%	21,88%	19,22%	19,08%	19,96%	100,00%
8	Apr	20,78%	19,01%	19,49%	19,42%	21,30%	100,00%
9	Mai	19,75%	21,12%	18,18%	19,00%	21,94%	100,00%
10	Jun	20,76%	20,80%	19,21%	19,54%	19,70%	100,00%
11	Gesamt B	20,48%	20,61%	19,57%	19,28%	20,06%	100,00%

Abbildung 13: Prozentuale Auswertung

3	Summe - Stunden	Name				
4	Months	Datum	Brigitte	Fritz	Hans	Kurt
5	Jan	01.Jan	6,00	7,00	10,00	2,
6		02.Jan	12,00	13,00	18,25	9,
7		03.Jan	20,75	18,50	28,00	19,
8		06.Jan	28,00	22,00	34,00	24,
9		07.Jan	32,75	31,00	40,00	30,
10		08.Jan	40,00	35,75	47,75	34,
11		09.Jan	46,75	38,25	53,25	40,
12		10.Jan	53,75	47,00	57,00	51,
13		13.Jan	59,75	54,00	64,50	54,
14		14.Jan	65,25	60,75	71,75	58,
15		15.Jan	70,00	68,50	75,75	67,
16		16.Jan	80,25	73,00	83,75	73,
17		17.Jan	85,00	80,00	92,50	78,
18		20.Jan	93,75	85,75	103,25	85,
19		21.Jan	103,75	91,75	112,50	91,
20		22.Jan	112,25	99,75	122,25	93,
21		23.Jan	117,75	105,75	127,75	101,
22		24.Jan	125,00	108,75	133,25	109,
23		27.Jan	130,50	113,00	140,25	113,
24		28.Jan	133,00	118,50	145,50	118,
25		29.Jan	139,25	126,75	150,75	129,
26		30.Jan	144,00	134,75	157,50	132,
27		31.Jan	150,75	137,00	166,50	141,
28	Feb	03.Feb	5,00	7,00	5,25	6,
29		04.Feb	12,25	12,00	11,00	12,
30		05.Feb	23,00	21,50	20,75	18,
31		06.Feb	32,00	32,00	27,50	26,

Abbildung 14: Der Datenpilot zeigt auf Wunsch kumulierte Werte.

	A	B	C	D	E	F
1	Filter					
2						
3	Summe -	Name				
4	Months	Brigitte	Fritz	Hans	Kurt	Ute
5	Jan		-13,75	15,75	-9,00	-13,75
6	Feb		7,75	-19,25	-11,75	-14,75
7	Mrz		14,25	-4,50	-5,50	0,75
8	Apr		-12,00	-8,75	-9,25	3,50
9	Mai		8,75	-10,00	-4,75	14,00
10	Jun		0,25	-9,50	-7,50	-6,50
11	Gesamt Ergebnis		5,25	-36,25	-47,75	-16,75

Abbildung 15: Absoluter Vergleich

Unterschied / Vorteile

- Bei der klassischen Vorgehensweise werden die Dateien zur Datenerfassung bereits in der für die Ergebnisbetrachtung gewünschten Struktur angelegt. Die Daten werden in dieser Struktur erfasst. Damit ist man auf diese eine Struktur festgelegt.

- Der Datenpilot hat einen datenbankartigen Ansatz. Die zugrunde liegenden Daten werden in einer Art Normalform, das heißt in einer primitiven, aber universellen Struktur, also in einer einfachen Tabelle erfasst, die alle Informationen enthält. Quasi erst im Moment des Betrachtens wird die gewünschte Ergebnisstruktur und die zu betrachtende Teilmenge der Daten festgelegt.

Fallbeispiel Häufigkeiten und Verteilung

Zur Darstellung der Häufigkeit von Ereignissen verfügt OpenOffice.org Calc über die Tabellenfunktion HÄUFIGKEIT. Sie muss in einer Matrixformel angewendet werden. Dies ist ein Thema für Fortgeschrittene. Alternativ lässt sich die Aufgabe auch mit dem Datenpiloten lösen, was keine besonderen Tabellenkalkulationskenntnisse erfordert.

Im Beispiel geht es um die Anzahl der E-Mails, die auf der deutschen Support-Mailingliste für OpenOffice.org (users@de.openoffice.org) eintreffen. Man möchte wissen, wie sich die Aktivität auf der Mailingliste über den Tag verteilt.

Basis der Untersuchung ist eine aus dem Mailboxfile des Mailclients Thunderbird mit dem Kommando

```
grep 'Delivery-Date:' Inbox
```

erstellte Textdatei. Sie enthält für jede in einem Zeitraum von mehr als zwei Jahren über die Mailingliste gesendete Mail eine Zeile mit Datum und Uhrzeit:

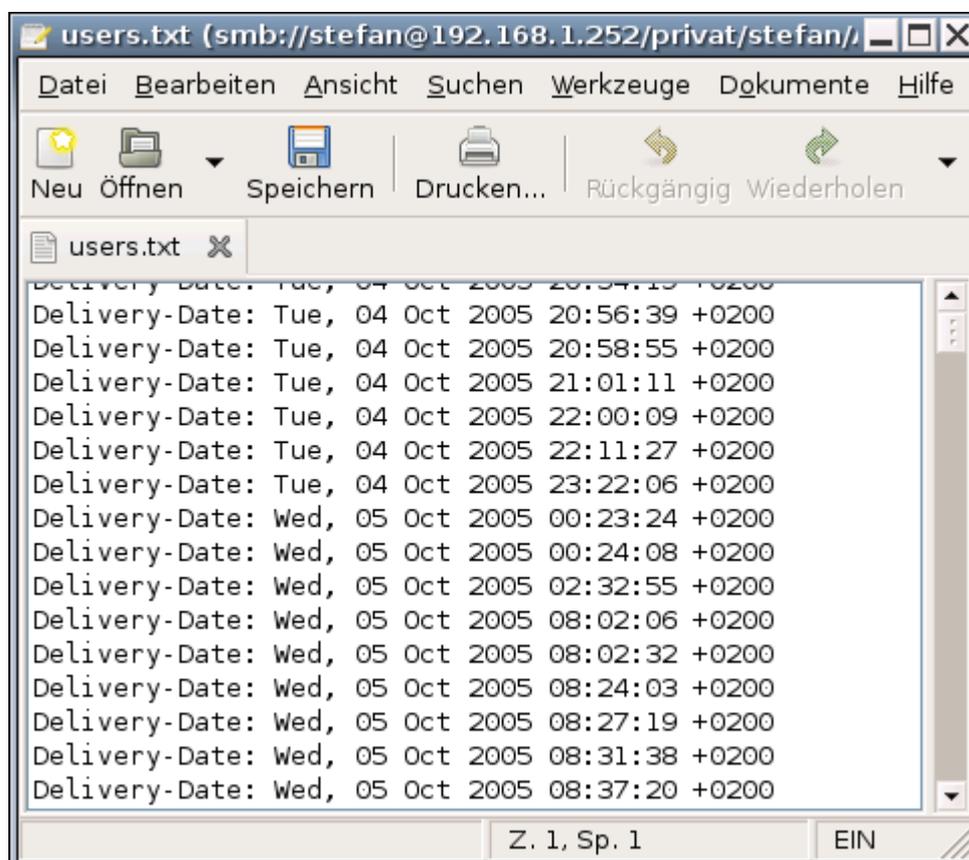


Abbildung 16: Rohdaten für Häufigkeit der Nachrichten

Die Daten werden in eine Tabelle in OpenOffice.org Calc importiert:

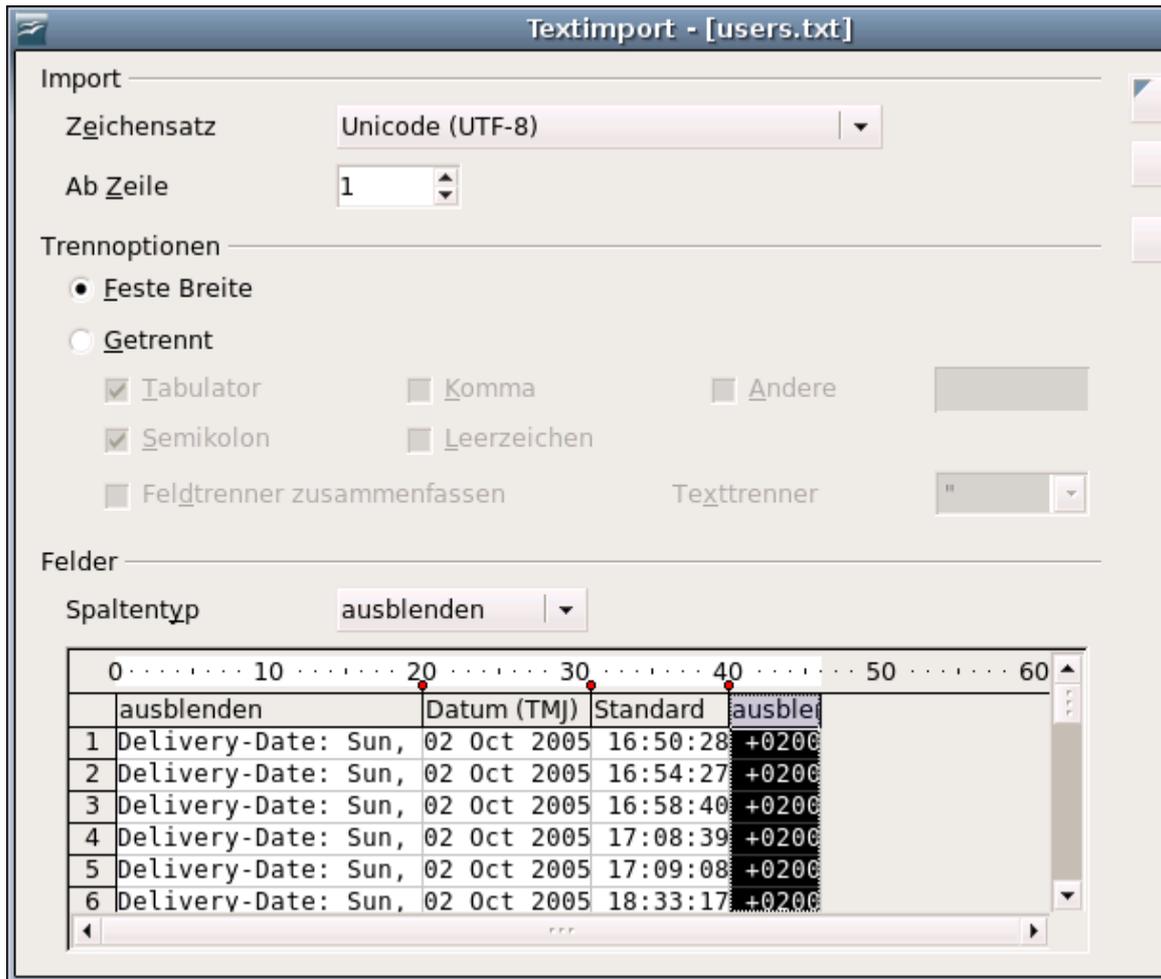


Abbildung 17: Import-Einstellungen

In OpenOffice.org Calc stehen die Rohdaten in drei Spalten zur Verfügung. Die erste Spalte enthält das aus der Textdatei importierte Datum, die zweite Spalte enthält die Uhrzeit. In der dritten Spalte werden Datum und Uhrzeit zu einer kompletten Zeitpunktangabe addiert:

	A	B	C	D
1	Datum	Zeit	DatZeit	
2	02.10.05	16:50:28	02.10.05 16:50	
3	02.10.05	16:54:27	02.10.05 16:54	
4	02.10.05	16:58:40	02.10.05 16:58	
5	02.10.05	17:08:39	02.10.05 17:08	
6	02.10.05	17:09:08	02.10.05 17:09	
7	02.10.05	18:33:17	02.10.05 18:33	
8	02.10.05	19:05:42	02.10.05 19:05	
9	02.10.05	19:24:12	02.10.05 19:24	
10	02.10.05	19:37:24	02.10.05 19:37	
11	02.10.05	20:12:39	02.10.05 20:12	

Abbildung 18: Rohdaten als Liste in Calc

Lösung mit einer Matrixformel

Um die Häufigkeitsverteilung zu berechnen, werden 24 Klassen für jede Stunde rund um die Uhr gebildet. Daneben wird die Häufigkeit der E-Mails mit Hilfe der Funktion HÄUFIGKEIT angegeben:

={HÄUFIGKEIT(B2:B38096;E2:E25)}		
D	E	F
	01:00	922
	02:00	466
	03:00	204
	04:00	123
	05:00	97
	06:00	110
	07:00	271
	08:00	773
	09:00	1538
	10:00	1944
	11:00	2214
	12:00	2418
	13:00	2382
	14:00	2327
	15:00	2428
	16:00	2524
	17:00	2397
	18:00	2274
	19:00	2260
	20:00	2366
	21:00	2266
	22:00	2131
	23:00	2088
	00:00	1572

Abbildung 19: Funktion HÄUFIGKEIT in einer Matrixformel

Das erste Argument der Funktion ist der Zellbereich mit den Uhrzeiten aller gut 38.000 empfangenen E-Mails. Das zweite Argument ist der Zellbereich E2:E25, der die Häufigkeitsklassen beschreibt. Zur Eingabe wird zunächst der Zellbereich F2:F25 markiert, dann die Formel eingetragen und mit der Tastenkombination SHIFT+STRG+ENTER abgeschlossen. Die Matrixformel erscheint danach in der Eingabezeile eingefasst in geschweiften Klammern.

Diese Technik ist in der Regel nur fortgeschrittenen Benutzern vertraut.

Lösung mit dem Datenpiloten

Mit dem Datenpiloten ist das selbe Ergebnis einfacher und schneller zu erreichen. Die Lösung eignet sich auch für wenig erfahrene Benutzer. Ausgehend von den Rohdaten (Abbildung 18) braucht es nur ein paar Mausklicks für die folgende Auswertung:

Fallbeispiel Häufigkeiten und Verteilung

	A	B
1	Filter	
2		
3	Zeit	
4	00	921
5	01	467
6	02	204
7	03	123
8	04	97
9	05	110
10	06	271
11	07	773
12	08	1537
13	09	1945
14	10	2212
15	11	2420
16	12	2382
17	13	2327
18	14	2427
19	15	2525
20	16	2397
21	17	2274
22	18	2259
23	19	2367
24	20	2266
25	21	2130
26	22	2089
27	23	1572
28	Gesamt Erg	38095

Abbildung 20: Häufigkeitsverteilung mit dem Datenpiloten

Vorteile

Der Datenpilot ist gegenüber der Lösung mit Formeln nicht nur einfacher und schneller, sondern auch vielseitiger. Auf Wunsch zeigt der Datenpilot statt der absoluten die relativen Häufigkeiten an:

	A	B
1	Filter	
2		
3	Zeit	
4	00	2,42%
5	01	1,23%
6	02	0,54%
7	03	0,32%
8	04	0,25%
9	05	0,29%
10	06	0,71%
11	07	2,03%
12	08	4,03%
13	09	5,11%
14	10	5,81%
15	11	6,35%
16	12	6,25%
17	13	6,11%
18	14	6,37%
19	15	6,63%
20	16	6,29%
21	17	5,97%
22	18	5,93%
23	19	6,21%
24	20	5,95%
25	21	5,59%
26	22	5,48%
27	23	4,13%
28	Gesamt Erg	100,00%

Abbildung 21: Relative Häufigkeit

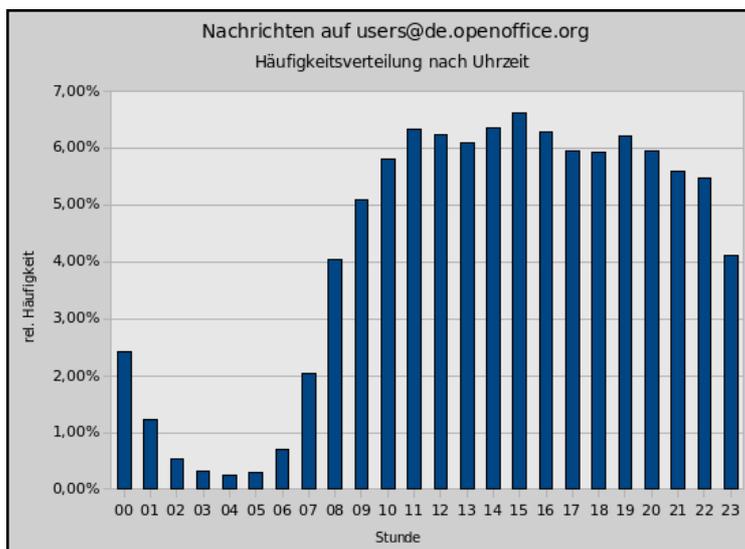


Abbildung 22: Diagramm auf Basis des Datenpiloten

Fallbeispiel Häufigkeiten und Verteilung

Eine andere Auswertung zeigt die Entwicklung der Anzahl der Nachrichten auf der Mailingliste von Ende 2005 bis Anfang 2008:

	A	B	C
1	Filter		
2			
3	Years	Datum	
4	2005	Okt	1339
5		Nov	1973
6		Dez	1784
7	2006	Jan	2130
8		Feb	1715
9		Mrz	2063
10		Apr	1860
11		Mai	1633
12		Jun	1249
13		Jul	1473
14		Aug	1707
15		Sep	1340
16		Okt	1554
17		Nov	1352
18		Dez	1118
19	2007	Jan	1259
20		Feb	1236
21		Mrz	1256
22		Apr	1214
23		Mai	1023
24		Jun	714
25		Jul	978
26		Aug	932
27		Sep	999
28		Okt	1090
29		Nov	1006
30		Dez	697
31	2008	Jan	869
32		Feb	532
33	Gesamt Ergebnis		38095

Abbildung 23: Anzahl der Postings pro Monat

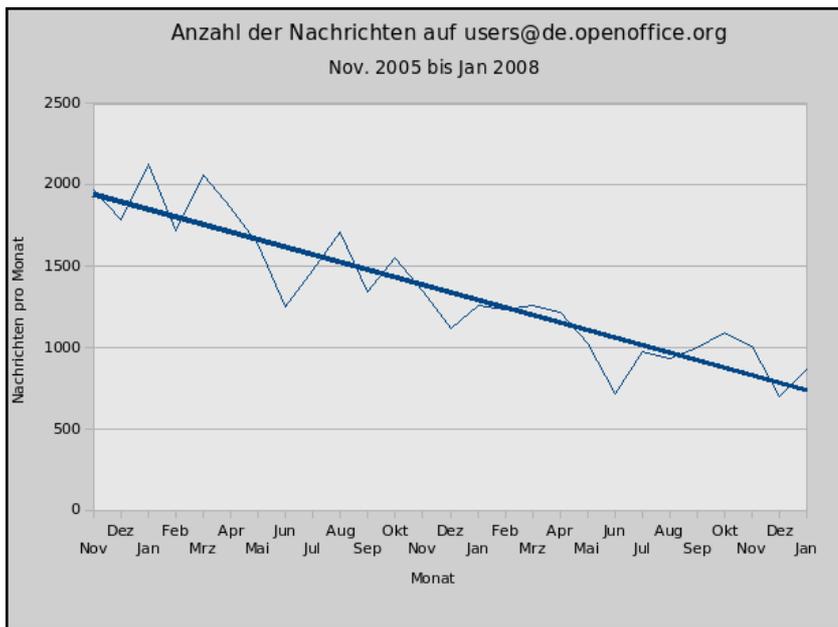


Abbildung 24: Diagramm auf Basis des Datenpiloten

Anmerkung	<p>Trotz steigender Nutzerzahlen ist die Anzahl der Postings auf der Supportmailingliste konstant rückläufig. Hierfür werden folgende Erklärungen vermutet:</p> <ol style="list-style-type: none"> 1. Mit dem Versionsprung von 1.x nach 2.x gab es Ende 2005 besonders viele Änderungen im Programm. 2. Rückgang von Flamewars. 3. Verbesserte Verfügbarkeit freier Dokumentation, FAQs, Wiki, HowTo. 4. Veränderung des Nutzertyps. Die heute neu zu OpenOffice.org hinzu stoßenden Anwender gehören immer weniger zum Typ der Mailinglistenleser, sondern mehr zum Typ der Webforenteilnehmer. Webforen sind ebenso verfügbar.
-----------	---

Verschenktes Potenzial

Die Kenntnis und Anwendung des Datenpiloten bringt in unzähligen Anwendungsfällen eine immense Arbeitersparnis und Leistungssteigerung im Vergleich zu klassischen Vorgehensweisen in der Tabellenkalkulation.

Die typischerweise mit dem Datenpiloten gut lösbaren Aufgaben gehören zu den häufigsten Anwendungen von Tabellenkalkulationsprogrammen. Denn es ist eine alltägliche Beobachtung, dass Anwender OpenOffice.org Calc verwenden, um listenförmig vorliegende Daten zu speichern.

Dennoch ist das Tool unter den Anwendern vergleichsweise wenig bekannt. Kurse und Bücher behandeln den Datenpiloten meist stiefmütterlich. Wenn überhaupt, erfolgt eine Besprechung im Rahmen der Programmfunktionen für Fortgeschrittene.

Dabei ist der Datenpilot ohne weiteres auch für Anfänger geeignet. Die Anwendung erfordert praktisch keine sonstigen Kenntnisse aus der klassischen Tabellenkalkulation.

Aufruf

- Zeigt den Leuten die Möglichkeiten des Datenpiloten!
- Zeigt es Ihnen frühzeitig am Anfang ihrer Begegnung mit OpenOffice.org Calc!
(Warum sollen sie sich erst plagen und lange Zeit später, als „Fortgeschrittene“ erfahren, wie viel Arbeit sie sich von Beginn an hätten sparen können?)

Anmerkung

Die Tatsache, dass der Datenpilot unter den Anwendern wenig bekannt ist und wenig bekannt gemacht wird, trifft auch für das als „Pivot-Tabelle“ bezeichnete Pendant in Microsoft Excel zu. Der Aufruf bezieht sich daher generell auf alle Tabellenkalkulationsprogramme mit entsprechender Funktionalität.

Neue Features

OpenOffice.org wird abgesehen von Bugfixes in der Regel in einem halbjährlichen Release-Zyklus aktualisiert. So lässt sich praktisch laufend über Neuerungen und Verbesserungen berichten.

Hier ein paar Beispiele für Neuerungen im Zusammenhang mit dem Datenpiloten:

Seit OpenOffice.org 2.3 (September 2007)

- Funktion PIVOTDATENZUORDNEN

Die Funktion PIVOTDATENZUORDNEN gibt einen Ergebniswert aus einer Datenpilottabelle zurück, um diesen in Formeln außerhalb des Datenpiloten weiterverwenden zu können. Der Wert wird durch Feldnamen und Elementnamen ausgewählt, sodass der Bezug gültig bleibt, falls sich der Aufbau der Datenpilottabelle ändert.

Aktuell in OpenOffice.org 2.4 (März 2008)

- Drilldown

Ein Doppelklick auf eine Zelle mit einem Ergebniswert im Datenpiloten erzeugt ein neues Tabellenblatt, das genau die Datensätze aus der zugrunde liegenden Datenmenge erhält, deren Zusammenfassung zu genau dem angeklickten Wert geführt hat.

- Manuelle Sortierung

Die Zeilen und Spalten des Datenpiloten sind primär alphanumerisch aufsteigend entsprechend der Kategorien sortiert. Es besteht nun die Möglichkeit, diese Reihenfolge mittels Drag&Drop zu verändern. Der Datenpilot behält auch nach einer Aktualisierung diese Ordnung bei.

Zu erwarten in OpenOffice.org 3.0 (Sept. 2008)

- Zellbereich für Datenquelle änderbar

Der Dialog für den Datenpiloten enthält nun auch ein Feld zum Editieren des Zellbereichs, auf dem die Auswertung basiert, sodass dieser auch nachträglich geändert werden kann.

Offene Wünsche

Issue-Tracker: http://de.openoffice.org/issues/show_bug.cgi?id=xxxxx

- 84899 datapilot : Select and display the same value several times
- 84898 datapilot : Select and display different calculated fields f
- 72304 DataPilot : sorting by dataField
- 70135 DataPilot gives percent results in an unreasonable number format
- 32310 DataPilot: Show pages