



Sphinx Scannermodul

Handbuch Version 5

**Fragebögen einscannen –
Scannen optimieren-
Zusatzfunktionen**

Copyright © sphinx-survey 2000 – 2009. Alle Rechte vorbehalten

Handbuch Scannermodul 14.02.2009

SphinxSurvey

Bachingerstr. 14 85435 Erding

Tel.: 08122 9091 624

Email : kontakt@sphinx-survey.de

Internet : <http://www.sphinx-survey.de>

Inhalt

| | |
|--|----|
| Einführung: Das Einlesen von Fragebögen per Scanner | 5 |
| 1. Grundsätzliches | 5 |
| 2. Einführungstour: Test der Software Sphinx + Scannermodul | 5 |
| 2.1. Testen der Software ohne Scanner..... | 6 |
| 2.2. Testen der Software mit Scanner..... | 7 |
| 3. Benötigte Hard- und Software..... | 9 |
| 3.1. Der Scanner | 9 |
| 3.2. OCR..... | 10 |
| 4. Welche Kreuze werden erkannt..... | 10 |
| Funktionsbeschreibung | 11 |
| 1. Vorbereitung der Dokumente | 11 |
| 1.1. Layout des Fragebogens..... | 11 |
| 1.2. Druck und Vervielfältigung der Fragebögen..... | 11 |
| 2. Mögliche Fehlermeldungen beim Einscannen..... | 11 |
| 2.1 Fehlerarten | 11 |
| 2.2. Speichern..... | 13 |
| 3. Scannereinstellungen..... | 14 |
| 3.1. Einstellungen der Scanmethode | 14 |
| 3.2. Bildbearbeitung..... | 15 |
| 3.3 Behandlung von Flecken und komplett durchgestrichenen Kästchen..... | 16 |
| 4. Weitere Funktionen | 16 |
| 4.1. Speichern..... | 17 |
| 4.2. Seitencodes | 17 |
| 4.3. Spezialränder..... | 19 |
| 4.4. Veränderung der erwarteten Kästchen | 19 |
| 4.5. Identifikation des Antwortenden..... | 19 |
| 4.6. Lesen eines Formulars, das nicht in Sphinx erstellt wurde..... | 20 |
| 4.7. Automatische Erkennung von Texten oder Zahlen | 21 |
| 4.8. Automatische Erkennung von Barcodes..... | 22 |
| 4.9. Lesen eines Datums | 23 |

Scannermodul testen – Funktionsbeschreibung - Optimierungsoptionen

| | | | |
|---|----|--------------------------------------|----|
| Optimierung des Einlesens..... | 24 | 1.3. Nach dem Test..... | 24 |
| 1. Layoutregeln..... | 24 | 2. Empfehlungen..... | 24 |
| 1.1. Während der Gestaltung..... | 24 | 2.1. Druck des Formulars..... | 24 |
| 1.2. Nach der Gestaltung des Fragebogens..... | 24 | 2.2. Reproduktion des Formulars..... | 25 |



EINFÜHRUNG: DAS EINLESEN VON FRAGEBÖGEN PER SCANNER

1. GRUNDSÄTZLICHES

- Ein Befehlsfenster ermöglicht mit einem Klick das Lesen oder Betrachten schon eingeleseener Fragebögen.
- Die Schnelligkeit des Einlesens der Fragebögen hängt von der Schnelligkeit der Scanner und der Kapazität des Blatteinzugs ab. Sie variiert zwischen 100 und 500 Fragebögen pro Stunde für von uns vorgeschlagene Scanner.
- Die Erkennungsquote liegt bei über 99%. Eine spezielle Umgebung ermöglicht die Ergebniskontrolle und die Eingabe offener Fragen.
- Erkennung von Filterführungen, Pflichtantworten, mehrseitigen oder beidseitig bedruckten Fragebögen, Lesen von Barcodes, Erkennung von handgeschriebenen Zahlen oder Texten (mit dem Zusatzmodul OCR).

- Der ganze Prozess (Erarbeiten des Fragebogens, Layout, Ausdruck, Einlesen, Korrekturen, Auswertungen) wird mit einer Software abgedeckt.

Anwendungsbeispiele

Im Rahmen der Qualitätsakkreditierung eines Krankenhauses füllt jeder Patient einen Fragebogen aus. Bisher wurden über 35 000 Fragebögen ausgefüllt, eingelesen und ausgewertet.

Während eines Seminars beantworten die Teilnehmer einen Fragebogen bezüglich ihrer Erwartungen. Während der Einführung werden die Fragebögen eingelesen. Eine Stunde später sind die 300 Fragebögen eingelesen und ausgewertet.

2. EINFÜHRUNGSTOUR: TEST DER SOFTWARE SPHINX + SCANNERMODUL

Damit Sie unsere Lösung testen können, müssen Sie entweder die Demoversion Sphinx Plus² bzw. Sphinx Lexica installiert haben (diese finden Sie auf unserer Homepage), eine zeitlich befristete Vollversion mit Scannermodul bekommen oder die Vollversion mit Scannermodul erworben haben.

Scannermodul testen – Funktionsbeschreibung - Optimierungsoptionen

Eine Beispielumfrage ermöglicht Ihnen, die Funktionen zu testen. Nummerierte Bilder wurden während der Installation auf Ihrer Festplatte installiert.

2.1. TESTEN DER SOFTWARE OHNE SCANNER

Öffnen Sie Sphinx, gehen Sie auf **Eine Umfrage öffnen** und wählen Sie im Verzeichnis **SphinxV5 - Umfragen - Beispiele - Scanner einlesen** die Umfrage **Notfallaufnahmeservice**. Klicken Sie nun einmal auf den Button **Sammlung der Antworten**, anschließend auf **Erfassungsmodul**. Nun gehen Sie auf **SCANNER**, danach auf **Justieren**. Hier können Sie einige Einstellungen vornehmen. Kreuzen Sie die Option **Digitalisierte Bilder benutzen** an und gehen auf **Schließen**. Wenn Sie jetzt auf **Scanner** gehen, werden schon eingescannte Bilder eingelesen und analysiert werden. Die Bilder befinden sich unter **SphinxV5 - Umfragen - Beispiele - Scanner einlesen - Bilder**. Wählen Sie das erste Bild **Notfallaufnahmeservice.tif** per Doppelklick aus. Da die vorab eingescannten Bilder der Reihenfolge nach nummeriert sind, werden sie automatisch nacheinander eingelesen.

Die angekreuzten Kästchen werden automatisch erkannt, die offenen Texte werden wie im ausgefüllten Fragebogen dargestellt.

Im Beispiel sind in der Liste oben links die Beobachtungen in schwarz geschrieben (kein Fehler wurde erkannt) oder in rosa geschrieben (Fehler wurden erkannt).

Klicken Sie in der Liste links oben auf die zweite Beobachtung (in rosa geschrieben) und wählen Sie in dem Dropdownmenü oben in der Mitte **Lesediagnosen (F8)**.

Eine Beschreibung der Fehlermeldungen erscheint:

- Bei den Fragen 5 und 6 wurden zu viele Kästchen angekreuzt.
- Bei der Frage 7 wurden 3 Kästchen angekreuzt, nur 2 waren erlaubt.

Wählen Sie nun im Dropdownmenü oben in der Mitte **Vollständiges Bild der Seite** oder **F7**. Sie sehen nun das eingescannte Bild, über das Sie mögliche Fehler sehen können. Klicken Sie auf **F5**, um auf die **Erkannten oder Erfassten Antworten** zurückzukommen. Klicken Sie auf eine Antwortmodalität (z.B. auf die Antwortmöglichkeiten der Frage 5: Geschlecht), um diese zu verändern und klicken Sie anschließend auf die korrekte Antwort.

Weitere Hinweise bezüglich dieses Bildes:

- Frage 1: Klicken Sie auf F7: das Kästchen 25 wurden komplett ausgefüllt (sprich durchgestrichen), dieses wird nicht als Antwort gewertet (diese Option können Sie ausschalten).

Scannermodul testen – Funktionsbeschreibung - Optimierungsoptionen

- Frage 8: Die Modalität 6 wurde umkringelt und nicht angekreuzt. Sie wird nicht gewertet.
- Frage 9 (Transport): Die Modalität Feuerwehr C wurde sehr nah am Rand angekreuzt, aber dennoch korrekt erkannt.

Wählen Sie die dritte eingelese Seite in der Liste oben links. Die Fragen 7 und 10 sind fehlerhaft, da zu viele Kästchen angekreuzt wurden. Korrigieren Sie diese gegebenenfalls.

Die Frage 1 (Name) ist eine offene Frage und kann auf zwei Arten analysiert werden: (Wenn Sie nicht die OCR-Lösung besitzen)

- Wählen Sie F6, klicken Sie auf diese Frage und tippen Sie den Namen manuell ein.
- Wenn alle Antworten auf diese Frage gleich wären könnte man alle Antworten auf einmal verändern: Gehen Sie auf **Ändern...** dann auf **Zuordnen**. Wählen Sie die Frage 1 aus und geben Sie den Wert ein. Diese Option ist insbesondere sinnvoll, wenn Codes (z.B. Seminarcodes) abgefragt werden, d.h. diese sind bei einer Einleseaktion immer gleich. Sie müssen also pro Seminarbefragung nur einmal den Code eingeben.
- Alternativ können Sie die handgeschriebene Antwort manuell eintragen. Klicken Sie auf die Frage 1. Name. Sie sehen nun die handgeschriebene Antwort, welche Sie abtippen können.

In diesem Beispiel wird ebenfalls das Datum abgefragt. Die Abfrage von drei geschlossenen Fragen im Fragebogen ermöglicht, den Tag, den Monat und das Jahr anzukreuzen und so automatisch das Datum zu berechnen, ohne dass Zahlen oder Buchstaben manuell eingetragen werden müssen, was den Einleseprozeß eventuell verlangsamen könnte.

Nach der Korrektur der Antworten klicken Sie auf **Speichern**, um die 4 eingelese Seiten in der Datenbank von Sphinx abzuspeichern.

Hinweis: In diesem Beispiel wurden extra “Fehler” eingebaut, um Ihnen einige Möglichkeiten der Software zu demonstrieren. Erfahrungsgemäß ist die Nachbearbeitung eingeleseener Seiten zwar notwendig, aber – bei “normal” ausgefüllten Fragebögen – nicht sehr aufwendig.

2.2. TESTEN DER SOFTWARE MIT SCANNER

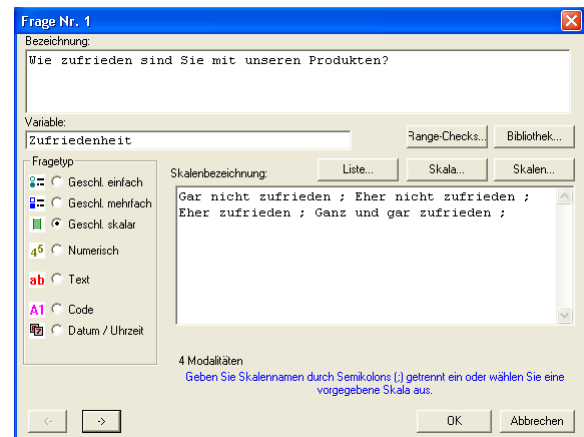
Hinweis: Sie können die Software auch mit Ihrem Flachbrettscanner testen. Achtung: Diese Scanner sind auf Grund ihrer Langsamkeit und dem Fehlen des automatischen Blatteinzugs nicht für die reale Eingabe interessant (die manuelle Eingabe ist hier schneller).

Geeignete Scanner werden später beschrieben. Eine grundlegende Voraussetzung für das Funktionieren Ihres Scanners mit Sphinx ist, dass ein TWAIN-Treiber installiert sein muss.

Scannermodul testen – Funktionsbeschreibung - Optimierungsoptionen

Nachdem Sie Sphinx installiert haben klicken Sie zweimal auf das Icon von Sphinx. Klicken Sie auf **Neue Untersuchung**, danach **Beenden**.

Anschließend klicken Sie auf den Button **Fragebogen erstellen**. Wählen Sie nun in der Menüleiste oben **Erstellen – Neue Frage**.



Unter **Bezeichnung** geben Sie folgenden Wortlaut ein: „*Wie zufrieden sind Sie mit unseren Produkten?*“. Unter **Variable** geben Sie „*Zufriedenheit*“ ein. Wählen Sie unter **Frage typ** den Typ **Geschlossen Skala**. Klicken Sie nun auf **Skalen...** und wählen Sie die Skala **Gar nicht zufrieden; Eher nicht zufrieden; Eher zufrieden...** per Doppelklick aus. Diese Skala erscheint nun im Fenster unter **Modalitäten**. Klicken

Sie auf **Neue Frage** und gehen auf **Bibliothek...**. Klicken Sie zweimal auf **Geschlecht**, so daß diese Frage im Fragefenster erscheint. Gehen Sie wieder auf **Neue Frage** und **Bibliothek...** und wählen per Doppelklick die Frage **Wie alt sind Sie?** aus. Gehen Sie nun auf **OK**. Drücken Sie auf **ESC**, um zum Hauptbildschirm zu gelangen und wählen Sie **Fragebogen layouten**. Speichern Sie die Untersuchung mit beliebigem Namen an einem Ort auf Ihrer Festplatte ab.

Wählen Sie daraufhin **Scanner**. Gehen Sie auf **Weiter**, dann **Fertig stellen**. Die drei Fragen erscheinen jetzt auf dem großen Hauptfenster in der Mitte. Verschieben Sie nun die drei Fragen auf dem Hauptbildschirm so, wie es Ihnen gefällt. Per Doppelklick auf eine Frage können Sie diese formatieren.

Über die Menüleiste **Datei**, anschließend **Drucken** können Sie diesen kleinen Fragebogen 2-3-mal ausdrucken. Füllen Sie die ausgedruckten Fragebögen aus.

Gehen Sie nun auf den Menüpunkt **Stadium** der oberen Menüleiste und wählen Sie **Erfassen**. Die Scanumgebung erscheint. Gehen Sie auf **Justieren**. Entfernen Sie ggf. das Kreuz **Digitalisierte Bilder benutzen**. Gehen Sie auf **Auswählen** und wählen Sie Ihren Scanner aus. Der Scanner muss mit einem TWAIN-Treiber funktionieren. Klicken Sie auf **Konfigurieren** und vergewissern Sie sich, dass die Scannereinstellungen (z.B. Seitengröße A4) korrekt sind. Unter **Min. Wert** und **Max. Wert** können Sie die Werte angeben, ab wann ein Kreuzchen als solches erkannt werden soll. Die Software Sphinx erkennt Graustufen, d.h. z.B. nur

Scannermodul testen – Funktionsbeschreibung - Optimierungsoptionen

wenn ein Kästchen zu 20% ausgefüllt ist, wird es als Kreuz gewertet (ansonsten könnte es auch ein Fleck sein), genauso bis 95% (da falsch ausgefüllte Kästchen oftmals komplett durchgestrichen werden). Diese Werte können Sie ändern.

Schließen Sie nach erfolgreichen Einstellungen den Dialog, legen Sie die ausgefüllten Fragebögen in den Scanner und klicken Sie auf **Scanner...** Wenn die Schrift schwarz ist, ist der Fragebogen korrekt erfasst worden, wenn die Schrift rosa ist, muss er nachbearbeitet werden. Gehen Sie auf einen Fragebogen, bei dem etwas nicht korrekt erkannt wurde (der also mit rosa Schrift geschrieben wurde). In dem Dropdownmenü in der Mitte können Sie sich ansehen, was nicht korrekt erkannt wurde. Wenn Sie im eingescannten Bild auf eine Antwortmöglichkeit klicken, können Sie diese verändern. Bestätigen Sie, dass Sie die Antworten verändern möchten und klicken Sie auf die zu verändernden Antwortmodalitäten. Nachdem Sie die Bilder überprüft haben, können Sie die Datei **Speichern** (Diskettensymbol) und **Beenden** (Türsymbol).

Gehen Sie nun über den Menüpunkt **Stadium** auf **Auswerten/Analysieren**. Wählen **Automatische Erstellung** und **Fertig stellen**. Nun sehen Sie eine Standardauswertung der eingescannten Fragebögen. Hinweise: Details zu möglichen Auswertungen entnehmen Sie bitte dem Handbuch Sphinx.

3. BENÖTIGTE HARD- UND SOFTWARE

Sie benötigen folgende Hard- und Software:

- Die Software von Sphinx Plus oder Sphinx Lexica
- Das Scannermodul von Sphinx
- Einen geeigneten Scanner
- Eventuell eine Schulung
- Eventuell die OCR-Bibliothek von Omnipage zur Erkennung von Handschrift bei offenen Fragen.

3.1. DER SCANNER

Der Scanner muss mindestens 10 Seiten pro Minute lesen können, mit einem automatischen Blatteinzug ausgerüstet sein und kompatibel mit Twain 32 bits sein.

Diese Scanner werden mit einer SCSI-Karte oder USB angeschlossen.

Von unseren Kunden werden häufig Fujitsu-Scanner (FI-Serie) oder Canon-Scanner (DR-Serie) eingesetzt.

Scannermodul testen – Funktionsbeschreibung - Optimierungsoptionen

Fragen Sie unsere Vertriebsmitarbeiter, damit wir Ihnen den für Sie geeigneten Scanner empfehlen und eventuell beschaffen können.

Weitere Informationen finden Sie auch unter:

- Fujitsu: <http://www.fujitsu.de>
- Canon: <http://www.canon.de>

3.2. OCR

Dieses Zusatzmodul ermöglicht zusätzlich das Lesen von Zahlen oder Texten. Dieses OCR-Modul wird von unserem Partner Omnipage entwickelt, welcher Weltmarktführer auf diesem Gebiet ist.

Der Kauf einer solchen Lizenz ist optional.

4. WELCHE KREUZE WERDEN ERKANNT

Die Software erkennt Graustufen, d.h. sie errechnet anhand des Prozentsatzes der Graustufe in einem Kästchen, ob ein Kästchen als ausgefüllt gewertet wird oder nicht (sprich ob es sich um einen Fleck handelt oder ob das Kästchen komplett

durchgestrichen wurde). Die Prozente, ab wann ein Kreuz als solches gewertet wird oder nicht (mehr) können Sie unter **Justieren** einstellen.

Folgende Kreuze werden (nicht) erkannt:



FUNKTIONSBESCHREIBUNG

1. VORBEREITUNG DER DOKUMENTE

1.1. LAYOUT DES FRAGEBOGENS

Bitte fügen Sie eine Anleitung für die Befragten bei, damit diese den Fragebogen korrekt ausfüllen, wie z.B.:



1.2. DRUCK UND VERVIELFÄLTIGUNG DER FRAGEBÖGEN

a) Fotokopie

Der Fragebogen kann ausgedruckt werden (wenn möglich mit Laserdrucker, schwarz/weiß oder farbig) und dann vervielfältigt werden.

Man muss dennoch auf die Qualität der Vervielfältigung achten: Flecken o.ä. müssen auf jeden Fall vermieden werden.

Bitte testen Sie den Fragebogen immer, bevor Sie ihn in großer Menge reproduzieren.

b) PDF

Das Formular wird im PDF-Format abgespeichert und kann so von einer Druckerei gedruckt werden. Die Druckerei kann die Fragebögen in schwarz/weiß drucken oder verschiedene Filme erstellen, um den Fragebogen farbig auszudrucken.

2. MÖGLICHE FEHLERMELDUNGEN BEIM EINSANNEN

2.1 FEHLERARTEN

Grundsätzlich können zwei Fehlerarten beim Einscannen auftreten: Das Einlesen wird angehalten oder es wurden Fehler beim Einlesen im Fragebogen erkannt, welche nachbearbeitet werden müssen.

a) Anhalten des Einlesens Die angezeigte Meldung kann sein:

–« Rand [innen/außen] [rechts/links] fehlt »

Der genannte Eckenrand fehlt vermutlich im Bild oder ist zu nah am Rand des gescannten Bildes. Sehen Sie sich daher das gescannte Bild an. Sie sehen an den Ecken Winkel, die die Lesezone anzeigen. Ebenso muss der eingescannte Fragebogen komplett von einem Rand umgeben sein, also auch unterhalb des Barcodes. Wenn dieser Fehler immer auftritt, müssen Sie die Größe der Scanzone verändern. Gehen Sie hierfür zurück in die Layoutumgebung des Fragebogens. Über den Menüpunkt **Formulare – Eigenschaften des Formulars** können Sie z.B. die Größe der Seite sowie spezielle Ränder einstellen. Insbesondere die letztere Option ist bei Nutzung einiger Drucker notwendig (Rand unten). Bitte achten Sie darauf, dass Fragen nach dem Einfügen von Rändern nicht über diese hinausgehen!

–« Seitenrand deformiert » Die vier Ecken des Formularrandes wurden richtig erkannt, aber die vier Ecken zusammen bilden kein Viereck. Wahrscheinlich wurde der Fragebogen falsch herum in den Scanner gelegt. Klicken Sie über **Justieren** die Option

Autorotation an oder legen Sie den Fragebogen richtig herum in den Scanner.

–« Lesen des Kästchens x der Identifikationsnummer unmöglich »

Die Analyse der Identifikationszone des Formulars (zentrale Zone) in den Seitencodes führte zu einem Problem. Die Nachricht gibt die Nummer an, welche das Kästchen per Voreinstellung hat (Nummer 1 ist das Kästchen ganz rechts).

–« Lesen des Kästchen x der Nummer der Beobachtung unmöglich », « Lesen des Kästchens x der Seitennummer unmöglich ».

Die Analyse einer dieser Zonen in den Seitencodes rief ein Problem hervor. Die Nachricht gibt die voreingestellte Nummer des Kästchens in der Zone an (Nummer 1 ist das Kästchen ganz rechts).

–« Die zuletzt gelesene Seite konnte nicht analysiert werden (erwartete Fragebogennummer: xxx) ». Ein korrekter Identifikator des Formulars wurde erkannt, aber entsprach nicht dem, der erwartet wurde. Überprüfen Sie, ob die Nummer des Formulars unten in der Mitte mit der Nummer, welche erwartet wurde. Gehen Sie eventuelle zurück in die Layoutumgebung des Fragebogens und dort über **Formulare – Eigenschaften des Formulars – Benutzung**. Dort können Sie manuell die Seitennummer des Fragebogens verändern. Bitte beachten Sie: Diese

Seitennummer ist ein Schutz, damit keine falschen Fragebögen eingelesen werden. Ändern Sie die Nummer also nur manuell, wenn Sie sicher sind, dass der Papierfragebogen auch dem Fragebogen in Sphinx entspricht!

–« Lesen der Frage Nr. X unmöglich »

Der Rahmen des Formulars sowie die Seitencodes wurden korrekt erkannt, aber eine Frage konnte nicht analysiert werden. Scannen Sie das Formular nochmals. Wenn die fehlerhafte Frage immernoch die gleiche ist, wurde die Position des erwarteten Kästchens und die Position der tatsächlichen Kästchen verschoben. Sehen Sie sich das gescannte Bild an (im Dropdownmenü oben in der Mitte über **Lesebereiche**). Im Dropdownmenü unten links können Sie die erwarteten und die erkannten Kästchen visualisieren. Ebenso können Sie festlegen, die erwarteten Kästchen zu verändern. Wenn Sie dies tun, klicken Sie auf das zu verändernde blaue Kästchen und schieben Sie es an den gewünschten Ort. Damit Sie auch das richtige Kästchen verschieben, erscheint der Name der Variablen automatisch bei Klick auf das blaue Kästchen.

b) Nachbearbeitung

Die Seitenanalyse geschieht problemlos, aber es wurden Fehler oder nicht berücksichtigte Anweisungen entdeckt. Die Farbe der Beobachtung in der Beobachtungsliste oben links ist daher entweder rosa oder blau.

Wenn z.B. mehrere Kästchen bei einer geschlossenen einfachen Frage angekreuzt wurden oder wenn trotz einer Filterführung Fragen ausgefüllt wurden, erkennt Sphinx dies als Fehler. Sie können die Details der entdeckten Fehler über das Dropdownmenü oben in der Mitte ansehen. Klicken Sie eventuell auf die fehlerhaft erkannte Frage und verändern Sie diese manuell.

Eventuell können Sie auch die minimalen und maximalen Werte über **Justieren...** verändern, um das Einlesen zu optimieren.

2.2. SPEICHERN

Nach dem Einlesen können Sie die Beobachtungen **Speichern** (Diskettensymbol) oder nicht. Wenn Sie speichern, werden die Beobachtungen definitiv in die Datenbank der Umfrage geschrieben.

Es ist auch möglich, einen Stapel von Seiten abzuweisen, z.B. wenn der Blatteinzug des Scanners nicht korrekt funktioniert. In diesem Fall wäre es schwierig herauszufinden, welche Seiten korrekt eingelesen werden und welche nicht.

3. SCANNEREINSTELLUNGEN

Einige Voreinstellungen sind vor dem Scannen notwendig. Diese hängen zum einen von der Konfiguration der benutzten Hardware ab, zum anderen hängen sie von den Besonderheiten des zu lesenden Dokuments ab. Diese Einstellungen werden in Sphinx abgespeichert.

3.1. EINSTELLUNGEN DER SCANMETHODE

a) Scanner auswählen

Wenn mehrere Scanner (oder Scannerpiloten) installiert wurden, können Sie über **Auswählen...** den Scanner festlegen, welcher benutzt wird.

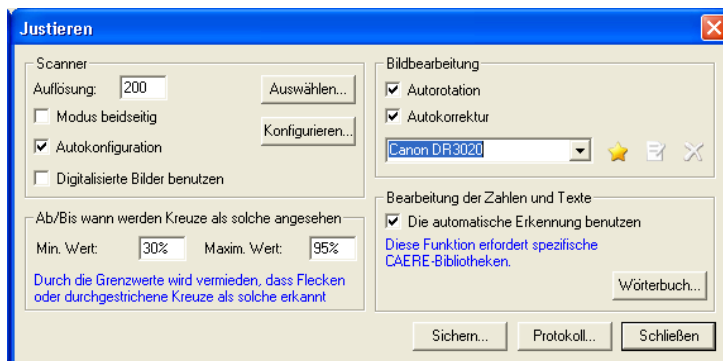
b) Auflösung

Die Auflösung 200 dpi ist ein guter Kompromiß zwischen Schnelligkeit und Auflösung. In bestimmten Fällen kann es notwendig sein, die Auflösung zu erhöhen (z.B. wenn die Ränder der Kästchen sehr dünn und schlecht markiert sind).

c) Modus beidseitig

Einige Scanner ermöglichen, Vorder- und Rückseite einer Seite auf einmal einzulesen. Der Vorteil ist nicht nur Geschwindigkeit, sondern auch eine einfachere Handhabung großer Fragebogenstapel. Zudem werden Papierkosten gespart.

Die von uns vorgeschlagenen Scanner können beidseitig einlesen.



d) Autokonfiguration

Wenn dieses Kästchen angekreuzt ist positioniert die Software automatisch die Scannerregler: Seitengröße des Papiers, Wahl des Einzugs usw. Man kann dieses Kästchen deaktivieren, um einige besondere Scanner zu verwalten.

e) Digitalisierte Bilder benutzen

Man kann diese Funktion benutzen, wenn man schon eingescannte Bilder analysieren möchte (z.B. scan0.tif, scan1.tif, scan2.tif, scan3.tif oder test10.bmp, test11.bmp, ...).

Diese Bilder können im Format (.bmp) oder (.tif) vorliegen.

In diesem Fall müssen Sie den Pfad des ersten Bildes der Serie angeben.

f) Konfigurieren...

Dieser Dialog führt zu Einstellungen, die von dem Hersteller des Scanners oder dessen Piloten kommen. Hier können Sie Dinge wie Kontrast usw. einstellen, um ein klares Bild zu erhalten. Wenn der Fragebogen farbig ausgedruckt wurde, ist es eventuell notwendig, die Behandlung der drei Basisfarben zu programmieren, um nur die nützlichen Elemente zu erhalten (z.B. den Hintergrund ignorieren).

Einige Scanner erfordern besondere Einstellungen, um das adäquate Bild zu erhalten. Im letzten Kapitel des Handbuchs finden Sie Einzelheiten.

3.2. BILDBEARBEITUNG

a) Autorotation

Die Software schlägt per Voreinstellung vor, dass die Seite normal eingelegt wird, d.h. der obere Teil der Seite oben ist. Wenn die benötigten Elemente nicht erkannt werden, kommt eine Fehlermeldung. Wenn Sie die Option **Autorotation** angekreuzt haben, probiert die Software automatisch das Bild neu zu orientieren, wenn dies notwendig ist.

Diese Funktion ist in folgenden Fällen nützlich:

- Im Stapel sind einige Formulare falsch herum und man möchte sich die Arbeit sparen, diese manuell herauszusuchen und umzudrehen.
- Einige Seiten sind oben beschädigt.
- Die Fragebögen wurden im Querformat ausgedruckt und können vom Blatteinzug nicht so eingezogen werden.

b) Verformungen korrigieren, die vom Scanner stammen

Scannermodul testen – Funktionsbeschreibung - Optimierungsoptionen

Die Scanner haben einen Motor, um die Blätter einzuziehen. Bei manchen Scannern ist der Einzug nicht konstant, so dass man die auftretenden Verformungen korrigieren muss.

Eine Zahl von vordefinierten Korrekturen können für die von uns vorgeschlagenen Scanner im Dropdownmenü ausgewählt werden (**Justieren... – Autokorrektur**).

Um seinen Scanner zu eichen, kann man auf das Sternzeichen gehen, um spezifische Korrekturen durchzuführen.

Um die Eichung zu überprüfen genügt es, in die **Lesebereiche** zu gehen (Dropdownmenü oben Mitte in der Scannerumgebung). Die erwarteten Kästchen müssen fast genau auf den gescannten Kästchen sein. Man kann eine Abweichung von $\frac{1}{4}$ der Größe der Kästchen akzeptieren.

3.3 BEHANDLUNG VON FLECKEN UND KOMPLETT DURCHGESTRICHENEN KÄSTCHEN.

Wenn ein anzukreuzendes Kästchen analysiert wird, wird der Prozentsatz der Graustufe berechnet. Die minimalen und maximalen Schwellenwerte werden in Prozent ausgedrückt.

Unter dem minimalen Wert wird das Kästchen als nicht angekreuzt angesehen, über dem maximalen Wert als durchgestrichen gewertet.

Die normalen Werte für die minimale Schwelle liegen bei 15-25%. Dadurch werden Flecken (z.B. durch die Druckerei verursacht) absorbiert.

Die normalen Werte für die maximale Schwelle liegen zwischen 85-100%. Diese Werte sind sinnvoll, da Befragte meist ein Kästchen komplett ausfüllen, wenn Sie Ihre gegebene Antwort ändern möchten.

Hinweise: Bitte beachten Sie, dass es nie vorkommen wird, dass bei Fragebögen, die nicht von geschultem Personal ausgefüllt werden, alle Personen die Antworten gleich ausfüllen. Dennoch sind die genannten Schwellenwerte sehr gute Erfahrungswerte.

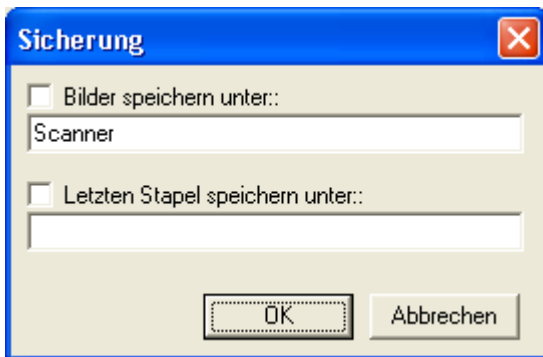
4. WEITERE FUNKTIONEN

Zu diesen weiteren Funktionen gelangen Sie in der Scanumgebung über den Button **Justieren**.

4.1. SPEICHERN

Man kann Sicherheitskopien erstellen lassen, um die Bilder oder einen Fragebogenstapel zu sichern.

a) Speichern von Bildern



Die Bilder werden temporär in einem Verzeichnis abgespeichert, das Sie im Dialog **Sichern...** definieren können. Wenn kein Verzeichnis angegeben ist, wird ein Verzeichnis *Scanner* im Ordner der laufenden Umfrage automatisch angelegt.

Diese Bilder werden konserviert, um die Visualisierung der gescannten Bilder für Überprüfungen und für die manuelle Eingabe von offenen Texten zu ermöglichen.

Bitte beachten Sie, dass eine Seite A4 mit einer Auflösung von 200 dpi ca. 500 Kb benötigt.

a) Speichern des letzten Stapels

Mit dieser Option kann man den letzten gescannten Stapel speichern, und zwar nicht nur den der laufenden Studie, sondern auch den einer anderen Studie, deren Name angegeben sein muss.

b) Protokoll

Das Protokoll dient dazu, alle erfolgten Operationen festzuhalten.

4.2. SEITENCODES

Zu diesen Optionen gelangen Sie in der Layoutumgebung des Fragebogens über die Menüs **Formulare – Eigenschaften des Formulars – Benutzung**.

Wenn der Fragebogen ausgedruckt wird, vergibt das System automatisch eine Nummer. Diese Nummer erscheint mit auf dem Fragebogen (Strichcode unten). Dadurch kann die Software erkennen, ob der gescannte Fragebogen auch dem entspricht, der eingelesen werden sollte.

Scannermodul testen – Funktionsbeschreibung - Optimierungsoptionen

Eigenschaften des Formulars

Benutzung | Größe | Funktionsweise | Plausibilität | Internet

Name des Formulars:
Formulaire scanner n°1

Benutzung: Papier Einlesen per Scanner

Damit die Fragebögen eingescannt werden können, müssen der Rahmen und Strichcode mit ausgedruckt werden..

Spezielle Ränder Links: 0 cm Rechts: 0 cm
Hoch: 0 cm Niedrig: 0 cm

Zusätzliche Randbereiche angeben, falls Ihr Scanner keine ganze Seite einlesen kann..

Benutzung der Seitencodes

Besondere Identifikation der Beobachtungen:

Wenn ausgefüllte Fragebögen gescannt werden, kann die Beobachtung durch ihre Nummer oder eine der numerischen Variablen identifiziert

Manuelle Nummerierung des Fragebogens: 131

OK Abbrechen

Wenn der Fragebogen mehrere Seiten enthält wird ein Strichcode unten rechts angelegt, die die Identifikation der Seitenzahl ermöglicht.

Wenn ein Fragebogen automatisch mit Informationen über den Antwortenden ausgefüllt wird erscheint unten links ein dritter Strichcode. Wählen Sie hierfür in den **Eigenschaften des Formulars** die Option **Besondere Identifikation der Beobachtungen**. Diese Option ist sehr interessant, wenn man personalisierte Fragebögen verteilt, um beispielsweise eine Kundennummer automatisch mit zu erfassen. In diesem Fall muss eine Variable in die Liste der Fragen eingefügt werden mit dem Fragetyp numerisch. Diese Frage füllt man dann manuell aus bzw. importiert die Antworten (beispielsweise eine Liste mit IDs). In der Layoutumgebung des Fragebogens geht man über **Stadium – Betrachten/Verändern** auf das Druckersymbol und druckt unter **Fragebogen** alle Fragebögen einzeln aus, so dass jeder Fragebogen eine individuelle Nummer links hat, welche automatisch eingelesen wird, so dass der Befragte identifiziert werden kann.



Wie oben beschrieben wird bei jedem Ausdruck des Fragebogens automatisch eine Nummer vergeben. Bei neuem Layout bzw. neuem Ausdruck des Fragebogens wird daher abgefragt, was mit der vorigen Konfiguration geschehen soll. Normalerweise wählen Sie vorige Konfiguration ersetzen, drucken die Fragebögen aus und scannen sie dann anschließend ein.

In einigen Fällen kann es aber notwendig sein, die automatisch vergebene Nummer manuell anzupassen. Dies machen Sie in der Layoutumgebung des Fragebogens über **Formulare - Eigenschaften des Formulars – Manuelle Nummerierung des Fragebogens**.

4.3. SPEZIALRÄNDER

Wenn die Umrandung zu nahe am Seitenrand ist, kann es zu Fehlermeldungen kommen. In der Layoutumgebung des Fragebogens können Sie über das Menü **Formulare - Eigenschaften des Formulars** daher **Spezielle Ränder** eingeben.

4.4. VERÄNDERUNG DER ERWARTETEN KÄSTCHEN

Wenn Lesefehler beim Scannen auftauchen, kann dies an Verformungen liegen, die bei der Reproduktion geschehen sind. Man kann daher die Position der erwarteten Kästchen verändern.

Über die Option **Lesebereiche** in der Scannerumgebung (Dropdownmenü oben in der Mitte) wählt man die Funktion **Anzeige der erwarteten Kästchen** im Menü unten links. Kreuzen Sie das Kästchen **die erwarteten Kästchen verändern** an. Mit der Maus können Sie anschließend die Position des Kästchens verändern.

Diese Veränderung wird bis dahin Gültigkeit haben, bis der Fragebogen erneut ausgedruckt wird.

4.5. IDENTIFIKATION DES ANTWORTENDEN

Sie haben die Möglichkeit, die Befragten oder Gruppen von Befragten zu identifizieren und zwar entweder, indem Sie personalisierte Fragebögen ausdrucken, oder indem Sie die Identifikation nach dem Einlesen per Scanner manuell eingeben.

a) Personalisierte Fragebögen zur Identifikation von Befragten oder von Gruppen

Definieren Sie zuerst die Frage (oder Fragen), welche die Befragten oder Gruppen von Befragten (z.B. Teilnehmer des Seminars X) identifizieren.

Diese Daten werden vorab manuell in die Datenbank eingegeben oder aber importiert.

Die zweite Etappe besteht darin, die individualisierten Formulare auszudrucken, nachdem Sie das Layout des Fragebogens gestaltet haben. In der Layoutumgebung des Fragebogens gehen Sie über **Stadium - Betrachten/Verändern** und wählendie Nummern der Fragebögen, die ausgedruckt werden sollen, aus. Diese

individualisierten Fragebögen werden nun hintereinander ausgedruckt.

Wenn eine numerische Frage (z.B. Kundennummer) die Befragten identifiziert, kann diese Nummer als Strichcode am linken unteren Rand des Fragebogens automatisch eingefügt werden. Gehen Sie in diesem Fall in der Layoutumgebung des Fragebogens über **Formulare auf Eigenschaften des Formulars**. Auf den Fragebögen ermöglichen Strichcodes die Identifikation der Beobachtung. Kreuzen Sie die Option **Besondere Identifikation der Beobachtungen** an und wählen Sie die numerische Variable aus. Wenn Sie nun über Stadium – Betrachten/Verändern die Fragebögen hintereinander ausdrucken wird jeder Fragebogen links unten eine individuelle Nummer besitzen, welche aus den vorab eingegebenen oder importierten Daten generiert wird.

b) Identifikationsnummern auf dem Fragebogen ausdrucken

Eine zweite Möglichkeit ist, dass eine Druckerei die Fragebögen automatisch durchnummeriert. Diese gedruckten Nummern können mit Hilfe des OCR-Moduls automatisch mit eingelesen werden.

Diese Nummern sollten nicht zu klein sein, nicht zu nah aneinander stehen und klar gedruckt sein.

Die Umfrage wird später mit der Datenbank fusioniert, so dass der Studienbeauftragte die schon vorhandenen Daten zuordnen kann. Weitere Informationen finden Sie im Abschnitt **OCR-Modul**.

4.6. LESEN EINES FORMULARS, DAS NICHT IN SPHINX ERSTELLT WURDE

Das Einlesen von solchen Formularen ist unter bestimmten Bedingungen möglich, wenn z.B. ein Rand um den Fragebogen gedruckt ist. Wenn dies der Fall ist, können Sie in Sphinx einen Fragebogen erstellen, der dem schon vorhandenen sehr ähnlich ist (entfernen Sie das Kreuz **Benutzung der Seitencodes** in den **Eigenschaften des Formulars**). Starten Sie das Einlesen, hierbei werden mit Sicherheit Fehler angezeigt werden. Durch die manuelle Veränderung der Größe und Position der Kästchen kann man lesbare Dokumente

erhalten.

4.7. AUTOMATISCHE ERKENNUNG VON TEXTEN ODER ZAHLEN

a) Eine Technologie in der Entwicklung

Durch die Integration von Bibliotheken ist das Erkennen von Handschrift, Barcodes, ... möglich. Sphinx benutzt die Bibliothek des Partners OMNIPAGE, welcher Weltmarktführer auf diesem Gebiet ist. Für diese OCR-Lösung muss eine Zusatzlizenz erworben werden.

Leider ermöglichen die derzeitigen Technologien keine perfekte Erkennung von Handschrift: Es werden zwischen 20% und 80% der Objekte erkannt.

Daher ist die OCR nur in bestimmten Fällen sinnvoll:

- Wenn die Fragebögen von geschulten Interviewern ausgefüllt werden.
- Wenn die Elemente einfach sind: z.B. Ziffern oder einfache Worte.
- Wenn die Anzahl der Fragebögen sehr groß ist, kann selbst eine geringe Erkennung sinnvoll sein und zu Zeitersparnis beim Nachbearbeiten führen.
- Wenn bestimmte Regeln beachtet werden: Die Zeichen sollten in Druckschrift groß geschrieben werden.
- Bei gedruckten Zeichen oder Barcodes ist die Texterkennung sehr gut

b) Die OCR Benutzung

Installieren die die OCR Bibliothek. Ein Icon OCR erscheint in der Scanumgebung.

Die Erkennung kann durch die Benutzung eines *Wörterbuches...* (im Menü *Justieren*) verbessert werden.

Diese Datei im Textformat enthält ein Wort pro Zeile. Für jede offene Frage muss bei dem Layouten des Fragebogens **OCR** angekreuzt werden (Doppelklick auf

die offene Frage, Reiter *Funktionsweise*) :

- Wählen Sie im Dropdownmenü rechts daneben, um welche Art von offener Frage es sich handelt und ob der Text geschrieben oder gedruckt wurde.
- In der Textzone *Spezialformat* kann die Erkennung verbessert werden, indem das erwartete Format beschrieben oder das benutzte Wörterbuch angegeben wird. Dieses Feld ist aber nicht obligatorisch. Dieses Spezialformat kann drei Typen haben:

FORMAT= « Beschreibungskette des Formats »

Scannermodul testen – Funktionsbeschreibung - Optimierungsoptionen

#DIC oder **#DIC=** für die Benutzung eines Wörterbuches per Voreinstellung oder eines speziellen Wörterbuches.

#NDIC um die Nutzung eines Wörterbuches für diese Frage zu deaktivieren.

Die Syntax der Beschreibungskette im Format muss folgendermaßen geschrieben sei :

Zeichen

\q Nicht ein Leerzeichen

\d Zahl [0-9].

\w Alphanumerisches Zeichen [0-9] [a-z] [A-Z]

\l Kleinbuchstabe [a-z]

\u Großbuchstabe [A-Z]

\a Buchstabe [a-z] [A-Z]

[x2y] Erlaubte Zeichengruppe: 'x', 'y', '2' wird erwartet

8 Spezielles Zeichen, das Zeichen 8 wird erwartet.

Formate

Eine Postleitzahl: 74601 \d\d\d\d\d

Ein Lieferantencode mit 3 Zeichen und 2 Zahlen : ABC25
\a\a\d\d

Versicherungsnummer:1-52-06-59-581-107 [12]-\d\d-\d\d-\d\d-\d\d\d\d\d

Datum: TT/MM/JJJJ [0-3]\d/[01]\d/[12][90]\d\d

4.8. AUTOMATISCHE ERKENNUNG VON BARCODES

a) Scannerauflösung

Dichte der Barcodes, Auflösung des Scanners

Niedrig (< 5 Zeichen/Punkt) 200 dpi

Hoch (5 bis 8 Zeichen/Punkt) 300 dpi

b) Positionierung einer Barcodezone im Formular

Die Position und Dimension des Barcodes muss in Relation mit der Größe des Barcodes sein.

Bei einem aufgeklebten Etikett kann ein Rahmen, der die Position angibt, hilfreich sein.

c) Eigenschaften der Frage, die einem Barcode assoziiert wird

Klicken Sie zweimal auf die entsprechende Frage beim Layouten des Fragebogens und gehen in den Reiter

Funktionsweise:

- Kreuzen Sie **OCR** an
- Wählen Sie im Dropdownmenü **Strichcodes**.

4.9. LESEN EINES DATUMS

Sie haben mehrere Möglichkeiten:

- Sie wählen anzukreuzende Kästchen: eine Variable mit dem Monatstag, eine mit dem Monat, eine für das Jahr. Anschließend werden diese drei Variablen zu einer

zusammengefasst, um ein Datum oder Zeitraum berechnen zu können.

- Mit dem OCR-Modul: Achtung, die Erkennungsqualität ist nicht perfekt. Wenn es sich um eine essentielle Frage handelt, sollte man dies nicht benutzen.

1. Handschrift: Benutzen Sie am besten drei numerische Variablen, Tag, Monat und Jahr. Diese Methode ist ziemlich sicher.

Nach der Erfassung werden diese drei Variablen zusammengefasst.

2. Vordruck: Man klebt ein vorgedrucktes Etikett in die Erfassungszone der Datumsfrage. Diese Methode ist ziemlich sicher.

OPTIMIERUNG DES EINLESENS

1. LAYOUTREGELN

1.1. WÄHREND DER GESTALTUNG

- Lassen Sie mindestens 1 mm Platz zwischen den Objekten (Achtung Ränder!).
- Benutzen Sie keine Hintergrundfarbe in den Antwortfeldern.
- Keine Schattierungen bei den anzukreuzenden Kästchen.
- Vergewissern Sie sich, dass jede Frage nur einmal auf dem Fragebogen erscheint.
- Vergewissern Sie sich, dass die Fragen nicht übereinander liegen.
- Vergewissern Sie sich, dass kein Objekt ausserhalb des Randes liegt.

1.2. NACH DER GESTALTUNG DES FRAGEBOGENS

- Überprüfen Sie, dass die Seitencodes bei mehrseitigen Fragebögen angezeigt werden.
- Drucken Sie den Fragebogen auf Papier und vergewissern Sie sich, dass der Rand eine geschlossene durchgezogene Linie ist und dass die Marge um die Ränder mindestens 5 mm ist.
- Testen Sie den Fragebogen einige Male an mehreren Personen

1.3. NACH DEM TEST

- Bitte achten Sie auf die Qualität des Ausdrucks und der Reproduktion.

2. EMPFEHLUNGEN

2.1. DRUCK DES FORMULARS

- Benutzen Sie immer den gleichen Rechner und den gleichen Drucker bei der Layoutgestaltung und dem Druck. Bei minimalen Differenzen z.B. bei Druckern kann es zu Fehlern kommen.

2.2. REPRODUKTION DES FORMULARS

- Die Qualität der Reproduktion ist sehr wichtig. Beachten Sie daher bitte die vorher angesprochenen Qualitätskriterien.