



Sicher erkennen durch intelligentes Formulardesign

I.R.I.S. AG

Inhaltsverzeichnis

1. EINLEITUNG	3
2. DIE FORMULAR-PROZESSKETTE	3
3. FUNKTIONSWEISE DER AUTOMATISIERTEN ERKENNUNG	3
3.1 AUSRICHTUNG DES FORMULARS	4
3.2 ERKENNEN DES FORMULARTYPS	4
3.3 SUCHEN DER DATEN AN BEKANNTEN POSITIONEN.....	4
3.4 SUCHEN DER DATEN AN VARIABLEN POSITIONEN.....	4
3.5 AUSLESEN DES DATENFELDES	5
4. BENUTZERFÜHRUNG	5
5. EMPFEHLUNGEN	7

1. Einleitung

Die Automatisierung der Erfassung von Papier-Formularen gehört heute zum Stand der Technik. Dabei gibt es in der Qualität der Gesamtlösung erhebliche Unterschiede. Häufig wird die Leistung einer automatisierten Formularerfassung über den noch verbleibenden manuellen Aufwand gemessen. Dieser manuelle Aufwand hängt zum einen von der verwendeten Extraktions-Technologie ab. Zum anderen spielt die Gestaltung der Formulare eine häufig unterschätzte aber trotzdem wesentliche Rolle bei der Gestaltung eines optimalen automatisierten Erfassungsprozesses.

Dieses Dokument möchte erklären und vorschlagen. Erklärt werden die technischen Hintergründe der Extraktionsansätze. Konkrete Vorschläge zur Gestaltung sollen zur Umsetzung in eigene, für die Automatisierung optimierte Formularlayouts anregen.

2. Die Formular-Prozesskette

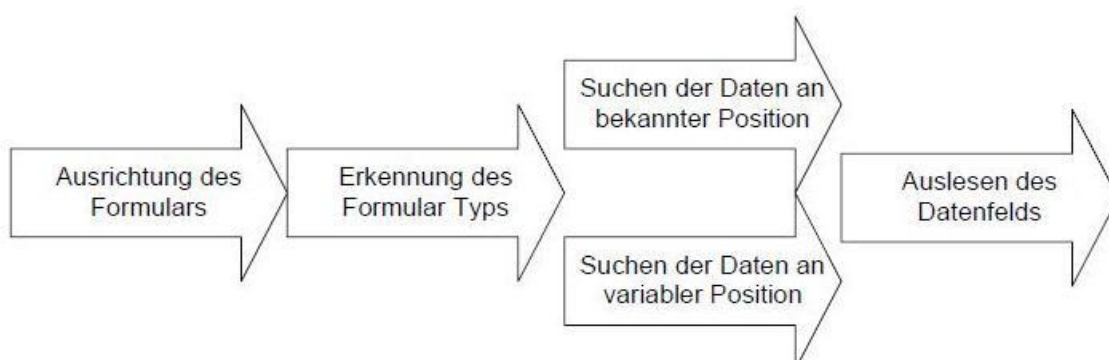
Folgende Schritte führen von der Idee des Formulars zu Daten die in IT Systemen verarbeitet werden:



Jede Veränderung in der Gestaltung hat Einfluss (positiv oder negativ) auf alle Prozess-Schritte. Damit bekommt die Formulgestaltung hohe Bedeutung für das Gesamtprozessergebnis.

3. Funktionsweise der automatisierten Erkennung

Die automatisierte Erkennung von Daten erfolgt nach folgendem Schema:



3.1 Ausrichtung des Formulars

Die Ausrichtung des Formulars sorgt dafür, dass Verdrehungen, Verzerrungen und Verschiebungen korrigiert werden. Diese Effekte können durch den Einzug des Scanners bzw. Fax, Fotokopien oder Toleranzen beim Druck entstehen.

Um die Ausrichtung sicher durchführen zu können, benötigen die heute verfügbaren Technologien Unterstützung, die auf dem Formular aufgebracht sein kann. Diese Unterstützung oder Ankerpunkte sind zum Teil grafische Objekte wie z.B.:



Die Ausrichte-Logik kennt die Sollposition dieser Ankerpunkte. Diese Ankerpunkte werden zunächst auf dem Bild gesucht. Anschließend kann anhand der Ist Position der Ankerpunkte die Ausrichtung erfolgen.

Teilweise erfolgt die Ausrichtung auch über den Formularrand. Voraussetzung ist hier die Verwendung eines Scanners, der einen schwarzen Rand um das Formular erzeugt.

3.2 Erkennen des Formulartyps

Die Erkennung des Formulartyps ist die Voraussetzung für die Verarbeitung gemischter Formular Stapel. Zur Erkennung des Formulartyps können verschiedene Merkmale herangezogen werden. Diese Merkmale können z.B. Barcodes, Textschlüssel, Logos, oder aber auch einfach der Gesamteindruck des Formulars (I.R.I.S. Fingerprint) herangezogen werden.

3.3 Suchen der Daten an bekannten Positionen

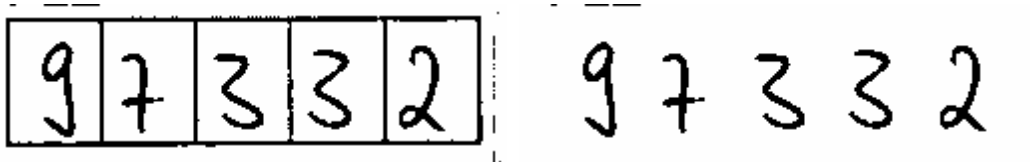
Sobald das Formular ausgerichtet und der Typ bekannt ist, können die Daten aus dem Formular automatisch ausgelesen werden. Dazu verwendet man hinterlegte Information über die zu lesenden Daten. Diese Informationen beinhalten z.B. die exakte Position eines jeden Datenfelds. Weiterhin wird der Datentyp dort beschrieben und abgelegt. Auf Basis dieser Informationen wird jetzt im Suchbereich nach Daten gesucht.

3.4 Suchen der Daten an variablen Positionen

Teilweise variiert die Position bestimmter Informationen. In diesem Fall muss die Information gesucht werden. Dazu müssen Suchkriterien vorliegen, an Hand derer sich die Information finden lässt.

3.5 Auslesen des Datenfeldes

Nachdem die Position der Information bestimmt ist und die Daten gesucht wurden, findet jetzt der Prozess des „Lesens“ statt. Dieser Vorgang basiert auf einzelnen Zeichen.



Aus den Einzelzeichen werden anschließend Worte und Sätze gebildet. Moderne Technologien sind in der Lage, auch vereinzelt zusammenhängende Zeichen zu isolieren und zu lesen. Grundsätzlich kann man aber sagen, dass die Qualität des automatisierten Lesens direkt davon abhängt, wie gut die einzelnen Zeichen zu isolieren sind.

Beispiel für gute Lesbarkeit:

1056768060

Beispiel für schlechte Lesbarkeit:

Ihre Provisionsrechnung

4. Benutzerführung

Wie oben gesehen sollten Formulare so gestaltet sein, dass der Benutzer maximal angeleitet wird, Zeichen einzeln und sauber zu schreiben. Diese Anleitung kann explizit in Form eines Ausfüllhinweises auf dem Formular aufgebracht sein.

Bitte nur Änderungen eintragen (in BLOCKSCHRIFT)!

Die Anleitung kann aber auch implizit durch die Gestaltung der Ausfüllbereiche erfolgen. Folgende implizite Gestaltungsmöglichkeiten werden häufig eingesetzt:

Kästchen oder Boxen:

--	--	--	--	--	--	--	--	--	--

Kämme:

--	--	--	--	--	--	--	--	--	--

Linien:

Diese Gestaltungsmöglichkeiten helfen bei Anleitung Zeichen möglichst isoliert zu schreiben.

Es gibt zwei Arten der Aufbringung dieser Hilfskonstrukte. Zum einen druckt man die Hilfselemente mit einer sogenannten *Blindfarbe*, einer Farbe die der Scanner nicht „sieht“. Damit sind die Hilfslinien auf dem gescannten Bild unsichtbar.



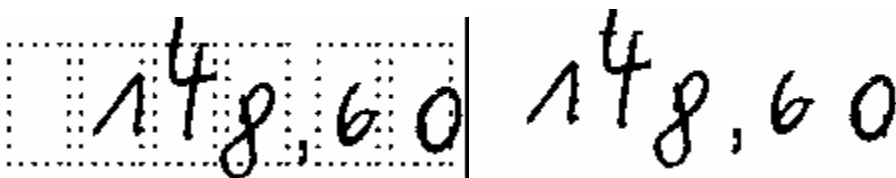
Werden die Elemente mit einer sichtbaren Farbe gedruckt, so hat diese Art der Gestaltung teilweise auch einen erheblichen Nachteil. Wenn der Ausfüller beim Schreiben die Elemente berührt, muss diese Berührung im Nachhinein wieder korrigiert werden. Dieser Prozess ist fehleranfällig und sollte wenn möglich vermieden werden.

Beispiel:



Falls mit sichtbarer Farbe gedruckt wird (z.B. Laserdrucker) sollten die Hilfselemente in der Form aufbracht werden, dass sie nachfolgend einfach und robust eliminiert werden können.

Beispiel:



5. Empfehlungen

Größe von Ausrichtmarken

Ausrichtmarken sollten mindestens 2 mm stark und mindestens 10 mm breit bzw. hoch sein. Der Abstand der Ausrichtmarken vom Seitenrand sollte min 5 mm betragen. Grundsätzlich gilt, dass die Ausrichtung umso effektiver funktioniert je weiter die Marken voneinander entfernt sind.

Größe des Schreibbereichs für 1 Zeichen

Der Schreibbereich für ein Zeichen sollte zwischen 4 – 8 mm breit sein und zwischen 5 – 10 mm hoch sein. I.R.I.S. empfiehlt die Breite von min. 6 mm und die Höhe von min. 8 mm. Weiterhin sollte der Abstand vom Schreibbereich zum nächsten sichtbaren Element min. 2 mm betragen. Der Abstand zwischen den einzelnen Zeichen sollte min. 1 mm betragen.

Gestaltung von Schreibbereichen mit Blindfarbe

Bei der Entscheidung für eine Blindfarbe sollte unbedingt der Scanner-Hersteller kontaktiert werden. Die Scanner Hersteller bieten den Einsatz von unterschiedlichen Lichtquellen an. Mit diesen können verschiedene Farben ausgefiltert werden.

Gestaltung von Schreibbereichen mit sichtbarer Farbe

Wir empfehlen die Verwendung von Punktmustern, da diese robust eliminiert werden können. Dabei sollten die Punkte eine Größe von 1,5 mm und einen Abstand von 0,7 mm haben.

Papiergewicht

Das Papiergewicht entscheidet über das Verhalten beim Scannen. Papier, welches leichter als 70 g/m^2 bzw. schwerer als 110 g/m^2 ist, wird beim Scannen Probleme bereiten. Wir empfehlen 80 g/m^2 .

I.R.I.S. Kontaktinformation:

Heussstraße 23

52078 Aachen

Germany

Tel: +49-(0)241-92035-0

Fax: +49-(0)241-92035-25

E-Mail: info-de@iriscompany.com

© Copyright 2010 I.R.I.S. s.a./n.v./AG

Alle Rechte, auch die des Nachdrucks, der Vervielfältigung oder der Verwertung bzw. Mitteilung des Inhaltes dieses Dokumentes oder von Teilen daraus behalten wir uns vor. Kein Teil darf ohne schriftliche Genehmigung der I.R.I.S. AG in irgendeiner Form reproduziert, an Dritte weitergegeben – insbesondere unter Verwendung elektronischer Systeme – verarbeitet, vervielfältigt, verbreitet oder zu öffentlichen Wiedergaben benutzt werden.

Alle Rechte für alle Länder vorbehalten. I.R.I.S., I.R.I.S.-Produktnamen, I.R.I.S.-Logos und I.R.I.S.-Produktlogos sind Marken von I.R.I.S. Alle anderen aufgeführten Produkte und Namen sind Marken oder eingetragene Marken der entsprechenden Inhaber.