

Service Point



Document, Print and
Information Management

Ontwerp scanvriendelijke documenten

Richtlijnen scanning

Service Point Belgium

Auteur	Johan Rambags
Datum	30-06-2009
Versie	1.0

Service Point Belgium NV
Industrieterrein Kolmen 1119
Groenmolenstraat 3
B-3570 Alken

T +32 (0)11 59 95 95
F +32 (0)11 31 57 01
info@servicepointbe.com
www.servicepointbe.com

BTW BE 0436270564
RPR Tongeren
ING 335-0362080-25
FORTIS 235-0391710-36



Inhoudsopgave

1	Inleiding	2
2	Algemene richtlijnen	2
2.1	Lettertype	2
2.2	Uitfiltering van kleuren	2
2.3	Kwaliteit van de scans	3
3	Automatische data-herkenning.....	3
3.1	Het belang van een goed formulierontwerp	3
3.2	Elementen op het formulier	3
3.2.1	Identificatie van het formulier	3
3.2.2	Ankerpunten.....	4
3.2.3	Symbolen of tekst	4
3.2.4	Kruispunten van lijnen.....	4
3.3	Data op het formulier	4
3.3.1	OCR	4
3.3.2	ICR	5
3.3.3	OMR.....	5
3.3.4	Barcode.....	5
4	Bijlage: drop out kleuren.....	6
4.1	Filterkleur rood.....	6
4.1.1	Pantone geel.....	6
4.1.2	Pantone oranje.....	6
4.1.3	Pantone rood/roze/perzik.....	6
4.1.4	Pantone fuchsia/purper.....	6
4.2	Filterkleur groen.....	7
4.2.1	Pantone geel.....	7
4.2.2	Pantone oranje.....	7
4.2.3	Pantone groen/blauw	7
4.2.4	Pantone geel/groen.....	7
4.3	Filterkleur blauw.....	7
4.3.1	Pantone blauw	7



1 Inleiding

De resultaten van de herkenning van documenten hangen in grote mate af van de opzet van die documenten. Zaken als lettertype, kleurgebruik en layout kunnen de resultaten van automatische herkenning van data aanzienlijk verbeteren, indien juist toegepast. In dit document worden algemene richtlijnen beschreven waarop gelet dient te worden bij het ontwerpen van een document.

2 Algemene richtlijnen

2.1 Lettertype

Om tot goede resultaten bij de herkenning te komen is het van belang dat de tekst duidelijk leesbaar voor de software is. Daarom is bij het kiezen van het lettertype de afstand tussen de letters belangrijk. Karakters die te dicht op elkaar staan zorgen voor foute herkenning. Het is daarom aan te raden om één van de volgende lettertypes te gebruiken:

- arial
- century gothic
- courier new
- lucida console
- times new roman
- verdana

Als een ander lettertype wordt gekozen, kies dan voor een type met voldoende afstand tussen de letters en kies zeker geen "fantasie"-lettertype. Probeer in de opmaak van het formulier ook het aantal lettertypes ook zoveel mogelijk te beperken, het beste is om maar één lettertype te gebruiken.

Ook de grootte van de karakters is belangrijk. Voor een goede herkenning is een minimale karakterhoogte van 10 punten noodzakelijk en ook hier geldt om zoveel mogelijk dezelfde grootte toe te passen.

Verdere aandachtspunten zijn:

- probeer de tekst zoveel mogelijk in hoofdletters te drukken.
- gebruik één kleur voor de te herkennen tekst, bij voorkeur zwart.
- gebruik geen cursieve, onderlijnde en vetgedrukte karakters.

2.2 Uitfiltering van kleuren

De betere productiescanners ondersteunen het uitfilteren van één kleur. De meest gebruikte filterkleur is rood, maar ook groen en blauw komen wel voor. Zie bijlage 1 voor een overzicht van de kleuren die gebruikt kunnen worden als filterkleur.

Filterkleuren worden gebruikt om kaders en vakken te drukken die de persoon die de gegevens invult ertoe aan te zetten om in die kaders te schrijven. Bij het scannen valt de filterkleur weg en zijn alleen nog de ingevulde gegevens nog zichtbaar (en dus niet de kaders die resultaten verstoren of karakters onduidelijk weergeven), waardoor de resultaten van de herkenning beter zijn. Een van de bekendste voorbeelden van het gebruik van filterkleuren is het overschrijvingsformulier.

Voorbeeld van OMR veld met groene filterkleur:

Geslacht

M V

Een punt waarop wel gelet dient te worden is dat als een filterkleur wordt gebruikt, dat ook alle informatie in die kleur wegvalt. Met andere woorden: als een rode filterkleur wordt gebruikt, laat dan de gegevens invullen met zwarte of blauwe inkt. Als de gegevens namelijk ook in het rood zijn ingevuld, worden die eveneens weggefilterd.

2.3 Kwaliteit van de scans

Dit punt is slechts van toepassing indien digitale beelden aangeleverd worden in plaats van fysieke documenten. Aangezien de scan het beginpunt is van de herkenning, is het zeer belangrijk om kwalitatief goede scans aan te maken. Indien mogelijk wordt er eerst een kwaliteitsverbetering uitgevoerd voordat de herkenning gebeurt. Dit kan onder meer inhouden dat het document recht wordt gezet, dat ruis wordt verwijderd en dergelijke.

Voor herkenning wordt de volgende minimale resoluties aangeraden:

- 300 DPI voor zwart-wit beelden.
- 150 DPI voor kleurenbeelden.

3 Automatische data-herkenning

3.1 Het belang van een goed formulierontwerp

De software die de automatische herkenning uitvoert op de gescande formulieren, kan dit niet zonder configuratie. Op het formulier dienen diverse elementen geplaatst te worden waaraan de software het type document kan herkennen en op basis waarvan de software kan vaststellen waar de zones met data zich bevinden. Het juist plaatsen van deze elementen zorgt voor maximale resultaten van de automatische herkenning.

Een goed formulierontwerp is overigens niet alleen van belang voor de herkenningsoftware. Het is zeker zo belangrijk dat de persoon die het formulier zal invullen dit op een eenvoudige manier kan doen en dat deze persoon zo veel mogelijk gestuurd wordt in de manier waarop het formulier ingevuld dient te worden. Een voorbeeld: voor een geboortedatum worden twee vakken voorzien voor de geboortedag, twee vakken voor een geboortedatum en vier vakken voor het geboortjaar.

3.2 Elementen op het formulier

3.2.1 Identificatie van het formulier

Door op een formulier een unieke identificatie te plaatsen, kan de herkenningsoftware makkelijker bepalen welk formulier (en dus welke data) er herkend moet worden. Als er meerdere types formulieren tegelijkertijd worden gebruikt en ook worden verwerkt, is identificatie een verplichting.

Er zijn verschillende mogelijkheden om een identificatie te plaatsen op een formulier. Bij voorkeur wordt hiervoor een barcode gebruikt, maar het is ook mogelijk om een gedrukte tekst te plaatsen op het formulier. De belangrijkste voorwaarde is dat de waarde die gebruikt wordt uniek is binnen een formulier en ook binnen alle



types formulieren die gelijktijdig verwerkt worden. Een tweede vereiste is dat de identificatie per type document op dezelfde plaats op het formulier staat. Verdere richtlijnen voor barcodes en gedrukte teksten zijn hieronder terug te vinden in het onderdeel "Data op het formulier".

3.2.2 Ankerpunten

Nadat de herkenningsoftware heeft vastgesteld welk formulier herkend moet worden, worden de ankerpunten gezocht. Ankerpunten zijn vaste tekens op een vaste plaats op een formulier die door de herkenningsoftware worden gebruikt om te berekenen waar zich de datazones bevinden.

Er zijn opnieuw verschillende mogelijkheden om ankerpunten op te nemen in een formulier. Een ankerpunt kan een symbool of tekst zijn of het kan een kruispunt van lijnen zijn. Bij voorkeur wordt een combinatie van deze twee mogelijkheden gebruikt. Per formulier dienen minimaal drie ankerpunten gebruikt te worden en deze punten mogen niet op één lijn liggen.

3.2.3 Symbolen of tekst

Een tekst die als ankerpunt wordt gebruikt dient te voldoen aan de volgende voorwaarden:

- voldoen aan de richtlijnen voor lettertypen.
- het is aangeraden om de karakters te kiezen uit de volgende reeks: A B H J K M N O P R S T X Y Z of om cijfers te gebruiken.
- 3 mm witruimte rondom de tekst voorzien.

3.2.4 Kruispunten van lijnen

De lijnen die gebruikt worden moeten aan een aantal voorwaarden voldoen:

- 10 mm lengte.
- 1 mm lijndikte.
- 10 mm vrij van de papierrand.

Voorbeeld van ankerpunten, linker- en rechter bovenhoek:



Het is niet noodzakelijk om aparte lijnen of hoeken te voorzien op een blad. Als er bijvoorbeeld al een kader is voorzien op een formulier is het mogelijk om dat te gebruiken voor de ankerpunten, zolang het natuurlijk voldoet aan de bovengenoemde voorwaarden.

3.3 Data op het formulier

3.3.1 OCR

Voorgedrukte tekst kan worden herkend. Om goede resultaten te behalen, dient met het volgende rekening gehouden te worden:

- voldoen aan de algemene richtlijnen voor lettertypen.
- 3 mm witruimte rondom de tekst voorzien.
- gebruik zoveel mogelijk hoofdletters en cijfers en probeer leestekens zoveel mogelijk te voorkomen.



4 Bijlage: drop out kleuren

De onderstaande kleuren worden aangeraden in combinatie met respectievelijke rode, groene en blauwe scanlampen. UCAD raadt echter aan om eerst een aantal testen uit te voeren voordat een formulier naar de drukker wordt gestuurd.

4.1 Filterkleur rood

4.1.1 Pantone geel

- 100	- 106	- 109	- 115
- 101	- 107	- 113	- 116
- 102	- 108	- 114	- Pantone Geel *

4.1.2 Pantone oranje

- 120	- 123 *	- 129	- 135
- 121	- 127	- 130 *	- 136
- 122	- 128	- 134	- 137 *

4.1.3 Pantone rood/roze/perzik

- 1345	- 141	- 148	- 155
- 1355	- 142	- 149	- 156
- 1365 *	- 143 *	- 150 *	- 157
- 1555	- 162	- 1625	- 169
- 1565	- 163	- 1635	- 170
- 1575 *	- 164 *	- 1645 *	- 171 *
- 176	- 1765	- 182	- 189
- 177	- 1775	- 183	- 190
- 178 *	- 1785 *	- 184 *	- 191 *
- 196	- 204	- 217	- 223
- 197	- 210	- 218	- 224
- 198 *	- 211 *	- 219 *	- 225
- 487 *	- 488	- 489	-

4.1.4 Pantone fuchsia/purper

- 230	- 236	- 243	- 250
- 231	- 237	- 244	- 251



Service Point

Document, Print and
Information Management

- 232 * - 238 * - 245 * - 252 *

* deze kleuren geven acceptabele resultaten indien gebruikt als "tint"

4.2 Filterkleur groen

4.2.1 Pantone geel

- 100 - 106 - 109 * - 114
- 101 - 107 - 113 - Pantone Geel *
- 102 - 108

4.2.2 Pantone oranje

- 120 - 127 - 134 - 1345
- 121 * - 128 *

4.2.3 Pantone groen/blauw

- 372 - 379 - 386 - 389 *
- 373 * - 380 * - 387 - 565 *
- 374 * - 381 * - 388 - 566 *

- 572 * - 586 * - 580 * - 5807 *
- 573 * - 5865 * - 587 - 331 *
- 585 * - 5875 * - 5803 * - 332 *

4.2.4 Pantone geel/groen

- 393 - 395 - 3935 - 3955
- 394 - 396 - 3945 - 3965

* deze kleuren geven acceptabele resultaten indien gebruikt als "tint"

4.3 Filterkleur blauw

4.3.1 Pantone blauw

- 250 - 283 - 2975 - 310
- 263 - 290 - 304 - 3105
- 2635 - 297 - 305 - 317
- 277