

ABBYY FlexiCapture : création d'un formulaire lisible par machine

Par ABBYY

Date de publication : 6 décembre 2019

TOUT PUBLIC

ABBYY FlexiCapture est une plate-forme LAD/RAD de traitement intelligent des documents, conçue pour répondre aux besoins de l'entreprise numérique complexe d'aujourd'hui.

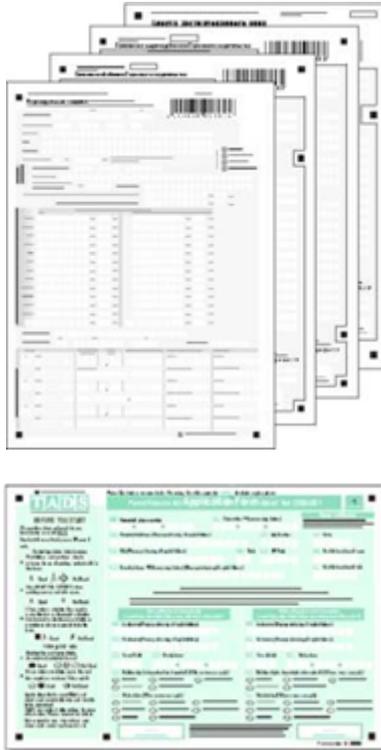
Dans ce tutoriel, vous allez apprendre comment créer un formulaire lisible par une machine. **Commentez.**

I - Qu'est-ce qu'un formulaire ?.....	3
II - Formulaires lisibles par machine.....	4
II-A - Méthodes de remplissage de formulaires.....	4
II-B - Éléments de formulaires lisibles par machine.....	4
II-B-1 - Texte.....	4
II-B-2 - Champ d'entrée.....	5
II-B-3 - Champ de coche.....	5
II-B-4 - Groupe de coches.....	5
II-B-5 - Marque de référence.....	6
II-B-6 - Séparateur de ligne.....	6
II-B-7 - Identificateur de formulaire.....	6
II-B-8 - Image.....	7
II-B-9 - Tableau.....	7
II-B-10 - Groupe d'éléments.....	7
III - Types de formulaires lisibles par machine.....	7
III-A - Formulaires dropout.....	8
III-A-1 - Choix de la couleur appropriée.....	9
III-B - Formulaires en noir et blanc avec des arrière-plans raster.....	9
III-C - Formulaires en noir et blanc avec des bordures raster.....	10
III-D - Formulaires linéaires en noir et blanc.....	10
III-E - Sélection du type de formulaire approprié.....	10
IV - Conditions générales requises pour les formulaires lisibles par machine.....	13
IV-A - Conditions d'arrière-plan de formulaire.....	13
IV-B - Conditions de marque de référence.....	13
IV-B-1 - Conditions pour les carrés noirs.....	13
IV-B-2 - Conditions pour les croix et les angles.....	14
IV-B-3 - Condition pour le texte (si utilisé comme marques de référence).....	14
IV-B-4 - Condition pour les séparateurs de ligne (si utilisés comme marques de référence).....	14
IV-B-5 - Condition pour les codes-barres (si utilisés comme marques de référence).....	14
IV-C - Conditions de champ de coche.....	14
IV-D - Conditions de marquage de texte.....	14
IV-D-1 - Taille des points raster.....	14
IV-D-2 - Taille des cellules de caractères.....	15
IV-D-3 - Épaisseur des lignes.....	15
IV-D-4 - Peignes.....	15
IV-E - Conditions de positionnement d'éléments.....	16
IV-F - Conditions de qualité d'impression.....	16
IV-G - Conditions de remplissage de formulaire.....	16
IV-G-1 - Stylos et encre.....	16
IV-G-2 - Taille et style des lettres.....	16
IV-G-3 - Échantillon.....	17
IV-H - Couleurs recommandées pour les formulaires dropout.....	17
V - Remerciements Developpez.com.....	18

I - Qu'est-ce qu'un formulaire ?

Les questionnaires, formulaires de sécurité sociale, bulletins de vote, formulaires de garantie sont des types de formulaires différents destinés à la collecte d'informations de types différents.

En quoi les formulaires diffèrent-ils des autres types de documents ?



- 1 Un formulaire comporte toujours un nombre défini de champs.
- 2 Chaque champ contient uniquement un type particulier d'informations. Par exemple, le champ « Nom » contient uniquement des noms (lorsqu'ils sont remplis correctement) et un champ « Date » contient uniquement des dates.

Les formulaires sont utilisés lorsque des informations provenant d'un grand nombre de personnes interrogées doivent être regroupées. La collecte manuelle des informations est un processus long et fastidieux dans lequel les typos et les erreurs sont quasiment inévitables ; c'est pourquoi les formulaires lisibles par machine sont destinés à automatiser ce processus.

Le traitement automatisé de formulaires se décompose en quatre étapes :

- 1 configuration de l'application de traitement de formulaires (en créant un modèle et en spécifiant les champs à reconnaître) ;
- 2 acquisition d'images du formulaire (numérisation) ;
- 3 traitement des images du formulaire (reconnaissance des images et validation des données extraites) ;
- 4 exportation des données extraites vers un système externe d'information.

Tous les formulaires utilisés pour regrouper des informations ne sont pas adaptés au traitement automatique. Pour être lisible par machine, un formulaire doit répondre aux conditions décrites dans ce guide.

II - Formulaires lisibles par machine

Pour pouvoir lire les informations contenues sur les formulaires, une application de lecture de formulaire doit effectuer les opérations suivantes :

- 1 déterminer l'emplacement des éléments du formulaire ;
- 2 séparer le contenu des champs des bordures de champ, du marquage de texte, de l'arrière-plan, du texte explicatif, etc.

Les formulaires lisibles par machine permettent au programme d'effectuer ces tâches.

Afin que la première tâche soit effectuée correctement, **les formulaires doivent correspondre au motif ou au modèle de formulaire**, c'est-à-dire que l'emplacement de tous les éléments du formulaire doit être identique sur tous les formulaires de type identique.

Afin que la première tâche soit effectuée correctement, **les formulaires doivent être conçus en prévision de leur traitement automatisé**, c'est-à-dire que le programme doit être capable de distinguer facilement les données à capturer des éléments de formulaire non reconnaissables, tels que les bordures de champ, le marquage de texte, les arrière-plans ou le texte explicatif.

II-A - Méthodes de remplissage de formulaires

Un formulaire peut être rempli par l'une des méthodes suivantes :

- à la main ;
- avec une imprimante matricielle ;
- avec une machine à écrire ;
- dans une imprimerie, ou au moyen d'une imprimante à jet d'encre ou laser avec une résolution supérieure à 300 ppp ;
- par combinaison de plusieurs de ces méthodes.

II-B - Éléments de formulaires lisibles par machine

Les éléments suivants peuvent être présents sur un formulaire :

- texte ;
- champ d'entrée ;
- champ de coche
- groupe de coches ;
- marque de référence ;
- séparateur de ligne ;
- identificateur de formulaire ;
- image ;
- tableau ;
- groupe d'éléments.

II-B-1 - Texte

Le **texte** est un élément de formulaire lisible par machine qui contient du texte descriptif : titre du formulaire, noms des champs, explications, etc.

II-B-2 - Champ d'entrée

Un **champ d'entrée** est un élément de formulaire lisible par machine dont le texte a été saisi par la personne qui a complété le formulaire. Pour faciliter la saisie de texte, les champs d'entrée doivent contenir du marquage de texte particulier. Les champs d'entrée sont généralement accompagnés de texte décrivant ou expliquant la nature des données à saisir.

Nom:

Les types de marquage de texte possibles sont répertoriés dans le tableau ci-dessous :

Texte sur une ligne	Le texte est saisi sur une ligne.	ABCDE
Lettres encadrées	Les lettres sont saisies dans des cadres contigus.	ABCD
Lettres dans des cadres séparés	Les lettres sont saisies dans des cadres séparés.	A B C D
Lettres séparées type peigne	Les lettres sont saisies dans un peigne.	A B C D
Texte encadré	Le texte est saisi dans une cadre.	ABCDE
Texte encadré avec lettres séparées type peigne	Le texte est saisi dans un cadre type peigne.	A B C D

II-B-3 - Champ de coche

Un **champ de coche** est un élément de formulaire lisible par machine (généralement un carré appelé « case à cocher »), dans lequel une marque doit être insérée par la personne qui complète le formulaire.

Un champ de coche est généralement accompagné par un texte qui éclaircit la signification de cette case à cocher.

Souscrire aux nouvelles

Le tableau suivant affiche des formes possibles de champs de coche :

Carré	Cercle	Souligné
☐	◯	—

D'autres formes sont également possibles, telles que les polygones, les ovales, etc.

II-B-4 - Groupe de coches

Un groupe de coches est un élément de formulaire lisible par machine qui regroupe des champs de coche parmi lesquels un seul doit être sélectionné. Les groupes de coches sont généralement accompagnés de texte explicatif décrivant la nature du choix à effectuer.

A quelle saison appartient le mois de mai ?

- l'hiver l'été
 le printemps l'automne

II-B-5 - Marque de référence

Une **marque de référence** est un élément de formulaire lisible par la machine et utilisé pour faciliter l'association d'images de formulaire aux modèles et déterminer l'orientation des formulaires au cours du traitement automatisé.

Avant que le programme ne commence la reconnaissance du contenu du formulaire, il doit associer le formulaire à son modèle. Tous les formulaires lisibles par machine du même type doivent être conformes à un modèle, c'est-à-dire que l'emplacement de tous les éléments de formulaire doit être exactement identique sur tous les formulaires. Le programme tentera d'associer le formulaire à son modèle, en se basant sur le numéro et l'emplacement des marques de référence sur le formulaire. Si le programme parvient à associer les marques de référence du formulaire aux marques de référence du modèle, il saura exactement où chercher les champs contenant les données à reconnaître.

Les marques de référence standard sont répertoriées dans le tableau ci-dessous.

Carré noir	Angle	Croix	Marque de temporisation
■	└	+	▬

Pour optimiser les résultats de la reconnaissance, il est recommandé que le formulaire comporte des marques de référence standard, mais, dans certains cas, des séparateurs de ligne, du texte explicatif ou des codes-barres peuvent également être utilisés comme marques de référence. Cependant, les marques de référence standard sont plus fiables et doivent constituer votre premier choix.

II-B-6 - Séparateur de ligne

Un séparateur de ligne est un élément de formulaire lisible par machine en forme de ligne noire horizontale ou verticale.

Les séparateurs de ligne sont utilisés pour fractionner les formulaires en divers segments logiques.

II-B-7 - Identificateur de formulaire

Un **identificateur de formulaire** est un élément de formulaire lisible par la machine et qui est utilisé pour sélectionner le modèle approprié lorsque plusieurs modèles possèdent des marques de référence disposées de la même manière.

Il est recommandé d'utiliser des codes-barres en tant qu'identificateurs de formulaire.

Un code-barres est une suite de lignes verticales en noir et blanc de largeurs différentes, il est utilisé pour le codage d'informations.



Certains autres éléments du formulaire peuvent également être utilisés en tant qu'identificateurs de formulaire.

II-B-8 - Image

Une **image** est un élément de formulaire lisible par machine qui doit être reconnu et exporté en tant que fichier image.

Par exemple, la partie du formulaire réservée au timbre ou à une signature constitue une image.

II-B-9 - Tableau

Un **tableau** contient des colonnes avec des éléments de même type.

Les colonnes d'un tableau peuvent contenir des éléments des types suivants :

- texte ;
- champ d'entrée ;
- date ;
- nombre ;
- coche.

Un tableau peut contenir une ligne avec des en-têtes de colonnes.

II-B-10 - Groupe d'éléments

Groupe d'éléments - élément d'un formulaire lisible par machine qui permet de combiner plusieurs éléments liés de manière logique en un groupe.

III - Types de formulaires lisibles par machine

Les formulaires lisibles par machine sont conçus en prévision de leur traitement automatisé, ce qui signifie que le programme doit être capable de différencier facilement les données à capturer des éléments de formulaires non reconnaissables, tels que les bordures de champ, le marquage de texte, les arrière-plans ou le texte explicatif.

Ceci peut s'effectuer des deux manières suivantes :

- 1 le formulaire doit être conçu de telle sorte que les bordures de champ, l'arrière-plan et le marquage de texte disparaissent après la numérisation ;
- 2 le formulaire doit être conçu de telle sorte que les bordures de champ, l'arrière-plan et le marquage de texte ressortent clairement après la numérisation.

Si vous optez pour la première approche, seuls les marques de référence, le texte explicatif et les données saisies subsisteront après la numérisation. Ainsi, le programme n'aura pas besoin de séparer les données des éléments de formulaire, ce qui rend la reconnaissance plus simple et plus fiable.

Dans le premier cas, les types de formulaires suivants sont utilisés : formulaires dropout en couleur, formulaires en noir et blanc avec un arrière-plan raster, formulaires en noir et blanc avec des bordures raster et formulaires linéaires en noir et blanc.

Si vous optez pour la seconde approche, les marques de référence, le texte explicatif, les données saisies, ainsi que le marquage de texte et les bordures de champ subsisteront dans le formulaire après la numérisation. Le programme recherchera les informations autour des bordures de champ et du marquage de texte. L'absence de ces éléments de formulaire aura un effet néfaste sur la qualité de la reconnaissance.

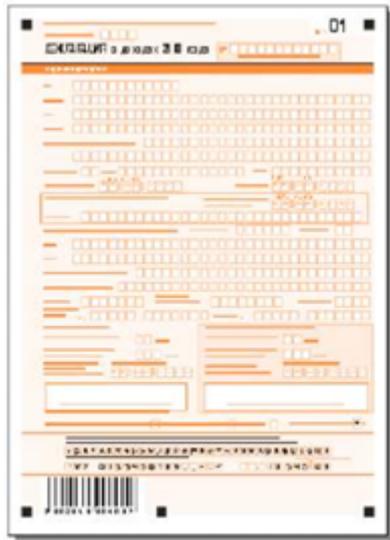
Dans le deuxième cas, les formulaires linéaires en noir et blanc sont utilisés.

III-A - Formulaires dropout

Les formulaires dropout ont un arrière-plan en couleur claire avec des rectangles blancs servant de cellules de caractères, des champs de coche, etc. Une cellule de caractère est un petit rectangle dans lequel un seul caractère doit être saisi (ex. : une lettre ou un chiffre).

Les formulaires dropout en couleur incluent les formulaires sur lesquels seules les bordures d'éléments sont colorées.

La couleur de l'arrière-plan est très importante, car elle doit disparaître au cours de la numérisation. L'arrière-plan est généralement rouge-orange ou vert clair. Pour obtenir la liste des couleurs recommandées, reportez-vous à la rubrique « **Couleurs recommandées pour les formulaires dropout** ».



Idéalement, tous les éléments de formulaire, sauf les marques de référence, doivent disparaître au cours de la numérisation, et seul doit subsister le contenu des champs devant être reconnu dans l'image de formulaire.

Les informations explicatives peuvent être soit conservées (lors d'une impression en noir), soit filtrées avec l'arrière-plan. Pour faire disparaître les informations explicatives au cours de la numérisation, celles-ci doivent être imprimées de la même couleur que l'arrière-plan du formulaire, mais avec un niveau de saturation plus élevé.

Lorsque vous décidez de procéder à l'impression des informations explicatives en noir ou sur un arrière-plan en couleur, les éléments suivants doivent être pris en considération :

- la taille du fichier image sera plus grande si les informations explicatives sont conservées ;
- l'espace entre les informations explicatives et les autres éléments du formulaire imprimé en noir et blanc doit être d'au moins 1,5 à 2 mm. L'espace entre le texte explicatif et le champ de coche correspondant doit représenter au moins les deux tiers de la taille de champ de coche. Si le texte explicatif est imprimé sur un arrière-plan en couleur, les espaces entre le texte et les autres éléments de formulaire ne sont pas essentiels.

Les formulaires contenant des informations explicatives en noir sont plus lisibles. Il est recommandé d'imprimer des textes explicatifs en noir, notamment si le formulaire ne contient pas de marques de référence.



Conseil :

Lors de la création d'un modèle pour un formulaire comprenant des informations explicatives imprimées, assurez-vous de numériser le formulaire en couleur ou en niveau de gris pour que les informations explicatives ou les éléments du formulaire soient clairement visibles.

Les formulaires dropout doivent être numérisés en utilisant une lampe ou un filtre de couleur. La couleur de la lampe est sélectionnée selon la couleur de l'arrière-plan.

Les formulaires avec des arrière-plans rouge ou vert doivent être numérisés soit :

- avec un scanner utilisant un logiciel de filtrage couleur ;
- avec un scanner noir et blanc utilisant une lampe rouge et verte (filtrage de couleur matériel) ;
- avec un scanner en noir et blanc équipé d'une lampe blanche et d'un filtre vert ou rouge (la qualité du filtrage sera moins bonne dans ce cas-là, car il se peut que l'arrière-plan ne disparaisse pas complètement ou que le contenu des champs ait été accidentellement supprimé) ;
- avec un scanner noir et blanc utilisant une lampe blanche sans filtre.

Lors de la numérisation avec une lampe blanche, certains arrière-plans plus clairs peuvent être supprimés sans filtrage : un niveau peu élevé de saturation de l'arrière-plan garantira la disparition de l'arrière-plan au cours de la numérisation. Dans ce cas, la couleur et la saturation de l'arrière-plan doivent être sélectionnées par approximation pour chaque scanner.



Remarque : de nombreux scanners en couleur possèdent également un logiciel de filtrage, utilisé pour filtrer et supprimer les couleurs bleues. Cependant, il n'est pas recommandé d'utiliser des teintes de bleu dans vos formulaires, car il est probable que les formulaires seront complétés non seulement à l'encre noire, mais aussi à l'encre bleue. Dans ce cas, le contenu de champ saisi à l'encre bleue disparaîtra,

III-A-1 - Choix de la couleur appropriée

Les couleurs rouge-orangé sont préférables au vert, car elles représentent le plus grand contraste possible par rapport au bleu, ce qui permet d'améliorer la qualité de la numérisation et de la reconnaissance si les formulaires sont complétés à l'encre bleue.

Pour en savoir davantage sur les couleurs recommandées pour le traitement de formulaires (c'est-à-dire celles qui sont le moins susceptibles de disparaître lors de la numérisation avec pratiquement tous les scanners), reportez-vous à la rubrique « **Couleurs recommandées pour les formulaires dropout** ». Vous pouvez également obtenir une liste des couleurs « dropout » pour un type particulier de scanner (au format Pantone ou autre) auprès du fabricant/revendeur de votre scanner.

III-B - Formulaires en noir et blanc avec des arrière-plans raster

Sur ce type de formulaires, les champs sont de simples espaces vierges (généralement des rectangles) sur un arrière-plan raster. L'arrière-plan est constitué de points individuels. La taille recommandée des points ne doit pas dépasser 0,1 mm, tous les points étant distants d'environ 1 mm les uns des autres. La taille des points et la distance qui les sépare sont primordiales, étant donné que l'arrière-plan raster est supprimé lors du nettoyage et non pas au cours de la numérisation : le programme considère les points comme des parasites et les supprime. Par conséquent, il est déconseillé de coller les points sur les images numérisées.

III-C - Formulaires en noir et blanc avec des bordures raster

Ces formulaires n'ont généralement pas d'arrière-plan. Toutes les bordures de champ (bordures de cellules de caractères, champs de coche et autres éléments de formulaire) sont composées de lignes raster, c'est-à-dire de suites de petits points noirs. La taille des points raster est généralement de 0,39 à 0,5 pt. La taille recommandée est de 0,39 pt, avec une espace entre chaque point représentant cinq fois la largeur des points. La taille des points et la distance qui les sépare sont primordiales, étant donné que l'arrière-plan raster est supprimé lors du nettoyage et non pas au cours de la numérisation : le programme considère les points comme des parasites et les supprime. Par conséquent, il est déconseillé de coller les points sur les images numérisées.

III-D - Formulaires linéaires en noir et blanc

Ces formulaires n'ont généralement pas d'arrière-plan. Toutes les bordures de champ (cellules de caractères et autres éléments de formulaire) sont composées de lignes noires solides. Les bordures de champ ne disparaissent pas lors de la numérisation ou du nettoyage.

Remarque : les formulaires de ce type doivent être complétés avec minutie. Si les formulaires ne sont pas correctement remplis, il se peut que les lettres empiètent sur les bordures de champ, réduisant ainsi la qualité de la reconnaissance. Ceci est dû au fait que chaque fois que le contenu du champ empiète sur les bordures de champ, il est extrêmement difficile pour l'application de différencier les bordures du texte. Idéalement, aucune partie du texte ne doit empiéter sur les bordures de champ.

III-E - Sélection du type de formulaire approprié

Outre les conditions requises en termes de conception et de traitement pour chaque type de formulaire, d'autres critères doivent également être pris en considération lors de la sélection du type de formulaire : les coûts et les méthodes d'impression, le nombre de copies requises, la taille de l'image en sortie, les modes de numérisation, la précision du degré de reconnaissance requis, la vitesse de traitement de formulaire, l'aspect général des formulaires, la netteté des formulaires, etc. Le tableau ci-dessous présente les avantages et les inconvénients de chaque type de formulaire.

Critère	Formulaire Dropout		Formulaire Raster en noir et blanc		Formulaire linéaire en noir et blanc	
	Inconvéni	Avantage	Inconvéni	Avantage	Inconvéni	Avantage
Complexité de la conception		Conception facile grâce aux outils de conception graphique.	Les éditeurs de texte offrent des outils de création pour ce type de formulaire. La taille des points raster est essentielle			Conception facile grâce aux outils de création graphique.
Impression	Difficile d'obtenir des formulaires de qualité			Facile à imprimer chez soi.		Facile à imprimer chez soi.

	lors d'impression en grande quantité.					
Coût de l'impression	Si vous avez recours à des services professionnels d'impression le coût de l'impression est supérieur à celui de formulaires en noir et blanc.			Si vous avez recours à des services d'impression professionnels le coût de l'impression est moins élevé que pour les formulaires dropout.		Si vous avez recours à des services d'impression professionnels, le coût de l'impression est moins élevé que pour les formulaires dropout.
Taille de l'image		La taille des fichiers image est inférieure.	La taille des fichiers image est supérieure		La taille des fichiers image est supérieure	
Mode de numérisation	Seuls des modes de numérisation particuliers (filtrage des couleurs) doivent être utilisés.		Étendue limitée des paramètres de luminosité et de contraste.		Étendue limitée des paramètres de luminosité et de contraste.	
Vitesse de traitement		Vitesse de traitement plus rapide.	Durée moyenne de traitement plus longue, car les images doivent être nettoyées.		Des erreurs sont susceptibles de se produire si les caractères sont collés ou empiètent sur les bordures de champ.	
Qualité de la reconnaissance		Qualité de la reconnaissance très élevée.		Qualité de la reconnaissance élevée.	La qualité de la reconnaissance dépend largement de l'état de «propreté» du	Qualité de la reconnaissance élevée si le formulaire a été rempli correctement.

					formulaire rempli.	
Emplacement des informations explicatives		Des informations explicatives peuvent être ajoutées n'importe où, y compris à l'intérieur des champs (du moment que le formulaire est imprimé avec la même couleur que l'arrière-plan du formulaire)	Les informations explicatives et autres éléments de formulaire doivent être disposés avec précaution		Les informations explicatives et autres éléments de formulaire doivent être disposés avec précaution.	
Aspect et lisibilité du formulaire	Des paramètres de contraste peu élevés et des informations explicatives peuvent avoir un effet négatif sur la lisibilité du formulaire.	Satisfaisant au niveau esthétique.	Aspect moins attrayant. Lisibilité réduite.		Aspect moins attrayant.	Lisibilité satisfaisante.
Aspect final		Peu sensible à l'empiètement de texte sur les bordures de champ.	Des caractères de petite taille doivent être écrits avec précaution et doivent être suffisamment grands pour ne pas être	Peu sensible à l'empiètement de texte sur les bordures de champ.	Très sensible à l'empiètement de texte sur les bordures de champ.	

			considérés comme des «parasites» et être par conséquen supprimés.			
--	--	--	---	--	--	--

IV - Conditions générales requises pour les formulaires lisibles par machine

Examinons à présent les conditions qu'un formulaire doit remplir pour pouvoir être lisible par la machine.

IV-A - Conditions d'arrière-plan de formulaire

- 1 Utilisez les formulaires dropout avec des bordures de champ raster si possible.
- 2 Lors de la création d'un formulaire dropout, utilisez une couleur qui disparaît au cours de la numération. Reportez-vous à la liste des couleurs recommandées dans la rubrique « **Couleurs recommandées pour les formulaires dropout** ».
- 3 Assurez-vous que les informations explicatives sont suffisamment éloignées du contenu du champ. Dans le cas de formulaires dropout, il est recommandé que toutes les informations soient imprimées de la même couleur que l'arrière-plan du formulaire (sauf si le texte explicatif est utilisé comme marque de référence au lieu de ou conjointement aux carrés noirs). Si vous imprimez des informations dont la couleur est la même que celle de l'arrière-plan, vous pouvez les placer n'importe où sur le formulaire, y compris dans les champs de formulaire.
- 4 N'utilisez pas de couleurs bleues pour les formulaires qui sont remplis manuellement. L'arrière-plan bleu risque de se confondre avec l'encre bleue susceptible d'être utilisée pour remplir les formulaires, ce qui aura un effet négatif sur la qualité de la reconnaissance.

IV-B - Conditions de marque de référence

- 1 Il est préférable d'utiliser des marques de référence standard pour créer des formulaires lisibles par la machine (ex. : carrés noirs, coins, croix). Le nombre de marques de référence recommandé est cinq : quatre marques de référence pour chaque angle du formulaire (ils doivent former un rectangle si des lignes imaginaires sont tracées entre chaque paire d'angles) et un carré sur l'un des côtés du rectangle imaginaire.
- 2 Si, pour une raison quelconque, aucune marque de référence n'est utilisée, nous vous recommandons d'utiliser la combinaison d'éléments suivante comme marques de référence :
 - a au moins deux séparateurs de ligne verticaux et deux séparateurs de ligne horizontaux ;
 - b au moins quatre éléments texte et un élément identificateur de formulaire ;
 - c d'autres combinaisons sont également possibles.
- 3 La distance entre le bord de la marque de référence et l'élément de formulaire le plus proche doit être d'au moins 3 mm.
- 4 La distance entre une marque de référence et le bord de la page doit être d'au moins 8 mm.
- 5 Les marques de référence doivent être imprimées de couleur sombre (en noir de préférence), afin de ne pas disparaître au cours de la numérisation.

Outre les conditions requises répertoriées ci-dessus, chaque type de marque de référence doit remplir ses propres conditions, décrites dans les chapitres ci-après.

IV-B-1 - Conditions pour les carrés noirs

Les carrés noirs sur un même type de formulaire doivent tous être de la même taille et d'une taille comprise entre 4 x 4 mm et 8 x 8 mm. La taille recommandée est de 5 x 5 mm. Les rectangles ne sont pas autorisés.

IV-B-2 - Conditions pour les croix et les angles

- 1 Le taille d'un carré imaginaire qui peut être tracé autour d'une marque de référence de ce type doit être comprise entre 4 x 4 et 8 x 8 mm. La taille recommandée est de 5 x 5 mm.
- 2 La taille des marques de référence d'un type doit être partout identique au sein d'un formulaire.
- 3 Tous les côtés d'une marque de référence doivent être de taille égale.
- 4 L'épaisseur de la ligne utilisée pour tracer les marques de référence doit être identique et comprise entre 0,3 et 1 mm. La valeur recommandée est de 0,5 mm.

IV-B-3 - Condition pour le texte (si utilisé comme marques de référence)

La taille de la police doit être d'au moins 7 pt (pour les en-têtes, d'au moins 14 pt).

IV-B-4 - Condition pour les séparateurs de ligne (si utilisés comme marques de référence)

L'épaisseur des lignes doit être d'au moins 1 pt. La valeur recommandée est de 1 à 1,5 pt.

IV-B-5 - Condition pour les codes-barres (si utilisés comme marques de référence)

- 1 La largeur des codes-barres (distance entre la première et la dernière barre) doit être d'au moins 47 à 50 mm.
- 2 La hauteur doit être d'au moins 12 à 15 mm (chiffres non inclus).
- 3 L'orientation du code-barres (sens des barres) doit être identique à l'orientation de la page (il est recommandé de numériser les pages dans la direction des barres).
- 4 Le format de code-barres EAN 13 est recommandé.
- 5 La distance recommandée entre un code-barres et tout autre élément doit être d'au moins 10 mm.

IV-C - Conditions de champ de coche

- 1 Les tailles de champ recommandées sont 3,5 x 3,5 mm, 4 x 4 mm, 4,5 x 4,5 mm ou 5 x 5 mm (si vous utilisez des champs de coches circulaires, vous devez les ajuster dans des carrés de la taille recommandée).
- 2 La largeur de ligne recommandée pour les bordures de champ de coche est de 0,4 mm (quelle que soit la taille du champ).

IV-D - Conditions de marquage de texte

IV-D-1 - Taille des points raster

- 1 Les tailles de champ recommandées sont 3,5 x 3,5 mm, 4 x 4 mm, 4,5 x 4,5 mm ou 5 x 5 mm (si vous utilisez des champs de coches circulaires, vous devez les ajuster dans des carrés de la taille recommandée).
- 2 La largeur de ligne recommandée pour les bordures de champ de coche est de 0,4 mm (quelle que soit la taille du champ).



Attention : il est fort probable que les points soient collés dans les angles des cellules de caractères.

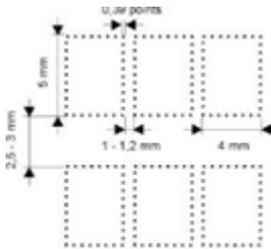


Conseil :

si votre logiciel ne vous permet pas de spécifier la distance entre chaque point raster, vous devez sélectionner un style de ligne qui comporte des points suffisamment séparés les

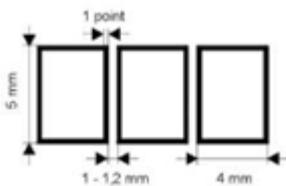
uns des autres. Ceci pour empêcher qu'ils ne soient collés les uns aux autres sur l'image numérisée et pour qu'ils disparaissent de l'image après son nettoyage, étant donné que les gros points ne sont pas interprétés par le système OCR comme des parasites et ne peuvent pas être supprimés de l'image au cours du nettoyage.

IV-D-2 - Taille des cellules de caractères



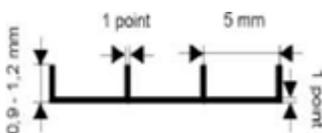
- 1 La taille des cellules de caractères recommandée est de 4 x 5 mm ; l'utilisation de tailles plus petites requiert un degré de précision plus élevé de la part de la personne remplissant le formulaire, ce qui est difficile à obtenir. Il est également déconseillé de séparer les caractères par des espaces trop importants, car ceci encourage l'utilisation de lettres anormalement grandes. Idéalement, la taille des cellules de caractères doit refléter la taille moyenne des lettres.
- 2 La distance entre les lignes de champ doit être d'au moins 2,5 à 3 mm et la distance entre deux cellules de caractères adjacentes sur la même ligne doit être d'au moins 1 à 1,2 mm.

IV-D-3 - Épaisseur des lignes



Si les bordures sont des lignes noires (« texte encadré », « lettres dans des cadres séparés » ou « lettres dans des cadres »), l'épaisseur de la ligne doit être de 1 pt.

IV-D-4 - Peignes



- 1 La distance recommandée entre les entailles dans un seul champ de formulaire doit être de 5 mm.
- 2 Dans le cas d'une entaille de peigne simple, la distance entre les deux lignes à entaille doit être d'au moins 7,5 à 8 mm.
- 3 La hauteur de l'entaille doit être d'environ 0,9 à 1,2 mm.
- 4 L'épaisseur des entailles et du cadre doit être de 1 pt.
- 5 Si un cadre avec des entailles de peigne est utilisé, la distance minimale entre les cadres doit être de 2,5 à 3 mm.

IV-E - Conditions de positionnement d'éléments

- 1 Assurez-vous de définir des marges suffisantes pour votre formulaire. Si ce n'est pas le cas, il est possible que les éléments de formulaire qui se trouvent trop près des bords disparaissent lors de la numérisation, entraînant ainsi une perte de données.
- 2 L'espace entre les éléments de formulaires voisins doit être d'au moins 2 mm.
- 3 L'espace entre un champ de coche et le texte explicatif doit représenter au moins les deux tiers de la taille du champ de coche.
- 4 Assurez-vous que le texte explicatif n'empiète pas sur les bordures de champ d'entrée.
- 5 Si les informations explicatives sont imprimées en encre noire, assurez-vous qu'un espace d'au moins 1,5 à 2 mm sépare les informations explicatives des autres éléments. La distance n'est pas cruciale pour les informations explicatives imprimées dans la même encre que l'arrière-plan.

IV-F - Conditions de qualité d'impression

Lors de l'impression de formulaires vierges, assurez-vous de respecter les conditions suivantes :

- 1 les formulaires doivent être imprimés soit de façon professionnelle, soit sur imprimante. Il est recommandé de recourir à des services professionnels pour les formulaires en couleur. Si ce n'est pas possible, testez les couleurs du formulaire sur un scanner avant de lancer l'impression ;
- 2 toutes les copies du formulaire doivent être imprimées avec la même source de document, de même que l'emplacement des champs sur chaque formulaire doit être identique ;
- 3 la taille de la marge minimale (c'est-à-dire la distance entre le bord de la page et l'élément de formulaire le plus proche) doit être d'au moins 8 mm (des marges de 12 mm sont recommandées) ;
- 4 la déviation linéaire autorisée des éléments de formulaire est de 0,15 % seulement (soit 0,5 mm pour une page de format A4) ;
- 5 si vous imprimez des formulaires avec une imprimante, ne les imprimez pas en utilisant une résolution inférieure à 600 ppp.

Utilisez toujours la même imprimante pour imprimer des formulaires provenant du même lot. Si ce n'est pas possible, assurez-vous au moins que les imprimantes utilisées sont du même modèle.

IV-G - Conditions de remplissage de formulaire

IV-G-1 - Stylos et encre

Il est préférable que les formulaires soient propres, remplis en majuscules, avec un crayon à bille ou à gel noir (le bleu foncé et le violet sont également tolérés). Les crayons à pointe souple de type crayon-feutre ne sont pas recommandés, car les caractères risquent d'être trop épais, entraînant ainsi des problèmes de reconnaissance. La qualité de la reconnaissance est également réduite si les formulaires sont remplis à l'aide d'un crayon à papier ou avec une encre claire.

IV-G-2 - Taille et style des lettres

Pour améliorer les résultats de la reconnaissance, les formulaires doivent être remplis avec des lettres majuscules. Les lettres saisies doivent occuper toute la surface des cellules de caractères, mais ne doivent pas empiéter sur les bordures de cellule.

Si les cellules de caractères sont signalées visuellement dans le formulaire (c'est-à-dire si les types de marquage de texte suivants sont utilisés : lettres dans des cadres, lettres dans des cadres séparés), les risques que la personne qui remplit le formulaire colle les lettres sont réduits.

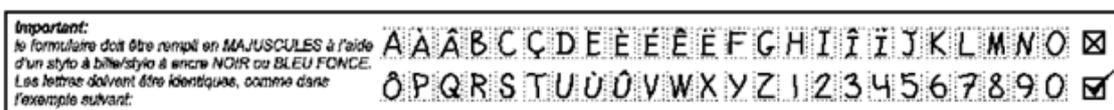
Les types suivants de marquage de texte posent des problèmes à la reconnaissance : texte encadré avec lettres séparées type peigne, lettres séparées type peigne. Seule la largeur de la cellule de caractère est prédéfinie et l'absence de bordure rend le collage des lettres possible. Dans le cas d'un peigne sans cadre, les lettres saisies peuvent être trop petites ou trop grandes.

Si le texte encadré ou le texte sur une ligne est utilisé, il n'existe aucune restriction évidente quant à la largeur des cellules de caractères. Les lettres peuvent être jointes, ce qui aura un effet néfaste sur la qualité de la reconnaissance. Dans le cas d'un texte sur une ligne, il n'y a pas de restriction quant à la hauteur des caractères. Ainsi, ce type de marquage n'est pas recommandé pour les formulaires qui requièrent un degré élevé de précision de la reconnaissance.

Une autre condition importante pour une reconnaissance réussie est que les lettres ne doivent pas empiéter sur les bordures de champ. Ceci n'est pas essentiel pour les formulaires dropout ou raster en noir et blanc, car les bordures de champ disparaîtront au cours de la numérisation et du nettoyage. Cependant, si les lettres empiètent sur les bordures de formulaires linéaires noir et blanc, ceci peut réduire considérablement la qualité de la reconnaissance.

IV-G-3 - Échantillon

Pour favoriser un remplissage correct de votre formulaire, nous vous recommandons d'y inclure une remarque du type de celle ci-dessous :



IV-H - Couleurs recommandées pour les formulaires dropout

Le tableau ci-après contient des échantillons de couleurs dropout et le numéro Pantone leur correspondant. Ces couleurs d'arrière-plan disparaissent si les formulaires sont numérisés :

● via un scanner.

Process Yellow U	100U	101U	388U	3935U	393U
102U	106U	107U	3945U	394U	3955U
108U	109U	113U	395U	3965U	396U
114U	115U	116U	461U	475U	503U
1205U	120U	1215U	586U	587U	600U
121U	1225U	122U	601U	602U	603U
123U	127U	1345U	607U	608U	609U
134U	1355U	135U	614U	691U	705U
1485U	148U	149U	706U	712U	713U
1555U	162U	169U	714U	719U	Cool Gray 1U
372U	386U	387U			

● via un scanner couleur doté du logiciel de filtrage des couleurs rouges ou via un scanner noir et blanc avec une lampe rouge ou un filtre rouge.

100U	101U	102U	1905U	190U	1915U	488U	489U	495U
106U	107U	108U	191U	1925U	193U	496U	5035U	503U
109U	113U	114U	196U	197U	198U	509U	510U	5175U
115U	118U	1205U	199U	203U	204U	517U	5245U	530U
120U	1215U	121U	206U	206U	210U	5315U	531U	585U
1225U	122U	1235U	211U	212U	213U	586U	587U	600U
123U	127U	128U	217U	218U	219U	601U	602U	603U
129U	130U	1345U	223U	224U	225U	604U	607U	608U
134U	1355U	135U	226U	230U	231U	609U	614U	656U
1365U	136U	1375U	232U	2365U	236U	663U	664U	670U
137U	141U	142U	2378U	237U	2385U	671U	672U	673U
143U	144U	1485U	238U	2395U	239U	677U	678U	679U
148U	1495U	149U	243U	244U	245U	684U	685U	691U
1505U	150U	151U	246U	250U	251U	692U	698U	699U
1555U	155U	1565U	252U	2562U	256U	700U	705U	706U
156U	1575U	157U	2572U	263U	2706U	707U	708U	709U
1585U	158U	1625U	379U	386U	3935U	712U	713U	714U
162U	1635U	163U	393U	3945U	394U	715U	719U	720U
1645U	164U	1655U	395U	459U	460U	726U	Cool Gray 1U	Orange 127U
169U	169U	170U	461U	4685U	468U	Process Yellow U	Process Yellow U	Process U
1787U	1788U	178U	473U	474U	4755U	Warm Gray 1U	Process Gray U	Process Gray U
182U	183U	184U	475U	481U	482U	Warm Gray 1U	Warm Gray U	Yellow U
185U	1895U	189U	485U	486U	487U			

V - Remerciements Developpez.com

Developpez.com remercie ABBYY pour l'autorisation de publication de ce tutoriel. Tous nos remerciements aussi à **Guillaume SIGUI** pour la mise au gabarit et **escartefigue** pour la relecture orthographique.