

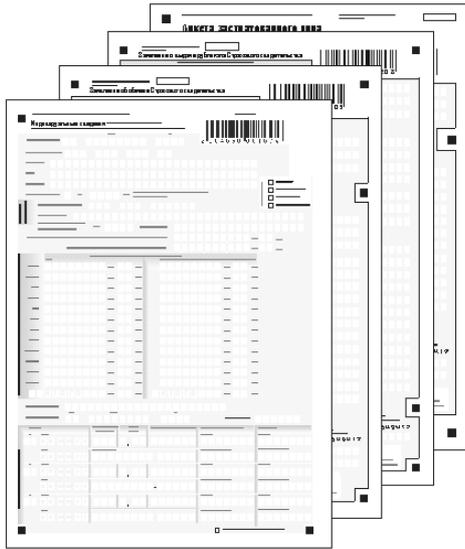
ABBYY FlexiCapture
Система потокового ввода данных

Создание машиночитаемых форм

Содержание

Что такое форма?	3
Машиночитаемая форма	4
Заполнение формы	4
Элементы машиночитаемой формы	4
<i>Текст</i>	4
<i>Поле для ввода</i>	4
<i>Поле метки</i>	5
<i>Группа меток</i>	5
<i>Репер</i>	5
<i>Разделитель</i>	5
<i>Идентификатор</i>	5
<i>Картинка</i>	6
<i>Таблица</i>	6
<i>Группа элементов</i>	6
Виды машиночитаемых форм	7
Цветные фоновые формы	7
<i>Подбор цвета</i>	8
Черно-белые формы с растровым фоном	8
Черно-белые формы с растровыми границами	8
Черно-белые линейчатые формы	8
Выбор вида формы	9
Требования к машиночитаемым формам	11
Фон формы	11
Реперы	11
Поле метки	12
Геометрические параметры разметки текста на формах	12
Расположение элементов формы	12
Тиражирование форм	13
Заполнение форм от руки	13
Рекомендуемые цвета для разработки цветных фоновых форм	14

Что такое форма?



Анкеты, карточки социального страхования, опросные листы, бюллетени, гарантийные талоны и т.д. – все это формы, используемые для сбора разного рода информации.

Что отличает формы от обычных документов?

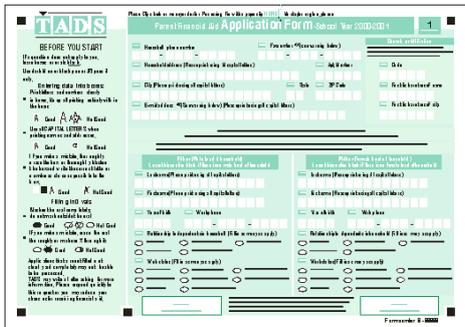
1. В отличие от обычных документов форма всегда имеет фиксированный набор полей.
2. В каждом поле формы может находиться информация строго определенного содержания: например, в поле "Фамилия" может находиться только фамилия (если, конечно, заполнивший форму не вписал туда по ошибке адрес), в поле Дата – дата и т.д.

Использование форм обычно подразумевает получение информации от большого числа респондентов с последующей компьютерной обработкой полученных данных.

Ручной ввод информации в компьютер очень трудоемок. Более того, такая форма работы ведет к большому количеству ошибок (опечаток), на исправление которых уходит много времени. Поэтому эффективнее автоматизировать процесс ввода форм.

Процесс автоматизированного ввода форм состоит из следующих стадий:

1. Настройка программы автоматизированного ввода на обработку формы данного образца (описание формы, выделение на образце формы данных для распознавания).
2. Создание изображения формы (сканирование).
3. Обработка изображений программой автоматизированного ввода (распознавание и проверка данных).
4. Экспорт данных во внешнюю информационную систему.



Наиболее просто и эффективно производится автоматическая обработка форм, удовлетворяющих специальным требованиям. Данное руководство рассказывает о том, какие требования предъявляются к дизайну форм, которые могут быть наиболее успешно обработаны программой автоматизированного ввода.

Машиночитаемая форма

При обработке форм программа автоматизированного ввода сталкивается с рядом задач, от решения которых зависит качество распознавания данных из формы.

Выделим две из них:

1. Определение положения элементов на изображении формы.
2. Отделение внесенной в форму информации (данных) от разметки текста, границ полей, фона, пояснительных надписей и т.д.

Чтобы облегчить программе решение этих задач, создаются специальные машиночитаемые формы.

Для решения первой задачи машиночитаемым формам выдвигается **требование точного соответствия формы образцу**. Это значит, что у всех форм одного образца расположение элементов формы должно быть одинаковым (совпадать на просвет).

Для решения второй задачи к машиночитаемым формам предъявляется **требование специального дизайна формы**: он должен позволять корректно отделить собственно данные от элементов формы: разметки текста, границ полей, фона, пояснительных надписей и др.

Заполнение формы

Форма может быть заполнена следующими способами:

- От руки.
- На матричном принтере.
- На пишущей машинке.
- Типографским способом. К данному способу относится также заполнение на принтерах (не матричных), которые обеспечивают разрешение не менее 300 точек на дюйм.
- Заполнение представляет собой комбинацию любых способов заполнения, перечисленных в предыдущих пунктах.

Элементы машиночитаемой формы

Рассмотрим подробнее элементы машиночитаемой формы.

Текст

Текст – это элемент машиночитаемой формы, представляющий собой описательный текст (название формы, название полей, пояснения к полям и т.д.)

Поле для ввода

Поле для ввода – это элемент машиночитаемой формы, предназначенный для ввода текста при заполнении формы. Для удобства ввода внутри поля для ввода текст может быть по-разному размечен. Поле для ввода обычно сопровождается описательным текстом, поясняющим смысл вводимого текста.

Имя: 

Существует несколько типов разметки текста:

Текст по линии	Тип разметки текста, при котором ограничителем текста является линия снизу.	
Буквы в рамках	Тип разметки текста, при которой ограничителем каждого символа текста является рамка; рамки не изолированы друг от друга.	
Буквы в изолированных рамках	Тип разметки текста, при которой ограничителем каждого символа текста является рамка; рамки изолированы друг от друга.	
Буквы в гребенке	Тип разметки текста, при которой ограничителем текста является гребенка.	
Текст в рамке	Тип разметки текста, при котором ограничителем текста является рамка.	

Текст в рамке с гребенкой	Тип разметки текста, при котором ограничителем текста является рамка с нижней границей в виде гребенки.	
---------------------------	---	---

Поле метки

Поле метки – это элемент машиночитаемой формы (чаще всего квадрат), в который при заполнении формы вносится метка (галочка, крест, сплошная закрашка и т.д.).

Поле метки обычно сопровождается описательным текстом, поясняющим смысл метки.

Подписаться на рассылку новостей

Поле метки может иметь ограничители следующих типов:

Квадрат	Круг	Подчеркивание
		

Возможны и другие типы ограничителей поля метки: многоугольник, овал и т.д.

Группа меток

Группа меток – это элемент машиночитаемой формы, объединяющий несколько находящихся вблизи друг от друга полей меток, из которых при заполнении формы будет заполнено только одно. Группа меток обычно сопровождается описательным текстом, поясняющим смысл группы.

Выберите время года, к которому относится месяц май:

Зима Лето

Весна Осень

Репер

Репер – это элемент машиночитаемой формы, необходимый для определения принадлежности формы к определенному образцу форм и ориентации формы во время автоматизированного ввода.

Первое, что определяется при распознавании данных из формы, это принадлежность формы к тому или другому образцу. Как уже говорилось, машиночитаемые формы создаются с учетом требования автоматизированного ввода, одним из которых является требование точного соответствия формы образцу. Количество и расположение реперов является основной характеристикой принадлежности формы к тому или другому образцу. Настроенная на работу с формами данного образца программа автоматизированного ввода ориентирует форму, т.е. пытается расположить её так, чтобы реперы совпали с реперами образца. Если реперы совпали, то программа может определить, к какому образцу относится данная форма и где находятся данные для распознавания.

Стандартные реперы:

Черный квадрат	Уголок	Крест	Черный прямоугольник
			

Присутствие стандартных реперов на машиночитаемой форме крайне желательно, хотя в некоторых случаях в качестве репера может использоваться разделитель, описательный текст или штрих-код. Эти варианты реперов считаются приемлемыми, но по возможности их следует избегать, т.к. они могут привести к ухудшению распознавания формы.

Разделитель

Разделитель – это элемент машиночитаемой формы, представляющий собой вертикальную или горизонтальную черную линию. Разделитель используется для визуального разделения формы на несколько смысловых частей.

Идентификатор

Идентификатор – это элемент машиночитаемой формы, однозначно характеризующий форму данного образца.

Если при распознавании невозможно однозначно определить по реперам принадлежность формы к образцу (т.е. реперы у нескольких форм различных образцов совпадают), то для точного определения используется идентификатор.



Рекомендуется в качестве идентификатора использовать штрих-код.

Штрих-код – это набор штриховых знаков, кодирующих информацию посредством чередующихся полос черного и белого цвета различной ширины.

Возможно использование и других элементов в качестве идентификаторов.

Картинка

Картинка – это элемент машиночитаемой формы, не подлежащий распознаванию и предназначенный для экспорта в графическом формате.

Любое графическое изображение, место для печати или подписи – все это картинки.

Таблица

Таблица содержит столбцы с однотипными элементами.

Столбцы таблицы могут содержать элементы следующих типов:

- текст;
- поле ввода;
- дата;
- число;
- метка.

Таблица может содержать строку с заголовками столбцов.

Группа элементов

Группа элементов – элемент машиночитаемой формы, предназначенный для объединения в группу нескольких логически связанных элементов.

Виды машиночитаемых форм

Как уже говорилось, машиночитаемые формы создаются с учетом требования автоматизированного ввода, одним из которых является требование к дизайну формы: дизайн формы должен позволять корректно отделить внесенную в форму информацию от разметки текста, границ полей, фона, пояснительных надписей и т.д.

Существует два различных подхода для реализации этого требования:

1. Дизайн формы создается таким образом, чтобы при сканировании формы и последующей очистке изображения разметка текста, границы полей и фон исчезали.
2. Дизайн формы создается таким образом, чтобы на изображении формы после сканирования и очистки изображения были четко выделены разметка текста и границы полей.

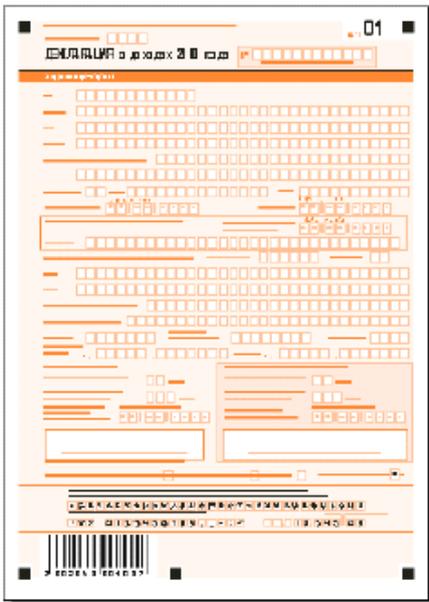
При использовании первого подхода на изображении формы после сканирования и очистки остаются только реперы, описательный текст и информация, которую ввели при заполнении формы. Настроенная на работу с такими формами программа автоматизированного ввода будет искать на изображении данные, не ограниченные никакой разметкой, что сузит задачи программы и повысит качество распознавания.

Для реализации этого подхода используются формы следующих видов: цветные фоновые формы, черно-белые формы с растровым фоном, черно-белые формы с растровыми границами, и черно-белые линейчатые формы.

При использовании второго подхода на изображении формы после сканирования остаются реперы, описательный текст, информация, которую ввели при заполнении формы, а также разметка текста и границы полей. Настроенная на работу с такими формами программа автоматизированного ввода будет искать на изображении именно ограниченные разметкой данные. Отсутствие разметки приведет к ухудшению качества распознавания.

Для реализации этого подхода используются **черно-белые линейчатые формы**.

Цветные фоновые формы



Цветные фоновые формы – это формы с цветным фоном, на котором белыми прямоугольниками выделены знакоместа (а также поля меток и др. поля), в которые при заполнении формы вносится информация. Например, Серые фоновые формы. Фон этих форм выполнен серым цветом.

Частным случаем цветных фоновых форм являются форма, на которых фоном выделены только ограничители полей для ввода информации.

Знакоместо – это прямоугольник, предназначенный для написания ровно одного символа (буквы, цифры и т.д.)

Цвет фона имеет большое значение, т.к. при сканировании фон таких форм должен "пропадать". Обычно фон имеет светлый "розово-оранжевый" или "зеленый" оттенок. Существуют цвета, рекомендуемые для использования в качестве фона для фоновых форм (подробнее см. Рекомендуемые цвета для разработки цветных фоновых форм).

В идеальном случае после сканирования и очистки изображения с цветной фоновой формы должны "пропадать" все элементы, за исключением реперов и информации, внесенной во время заполнения формы.

Пояснительная информация может оставаться на форме (если она выполнена в черном цвете); но можно добиться того, что и этот элемент будет "пропадать" при сканировании. Для этого пояснительную информацию необходимо заполнить в том же цвете, что и фон формы, но с большей насыщенностью.

Выбирая между черным цветом и цветом фона для пояснительной информации, учитывайте следующее:

- Размер файла изображения формы с пояснительной информацией, выполненной в черном цвете (или любом другом, который останется после сканирования на форме), больше, чем размер файла изображения формы с пояснительной информацией, выполненной в цвете фона (или любом другом, который исчезнет после сканирования на форме).
- Размещая на форме пояснительную информацию, выполненную в черном цвете, необходимо отступить от других элементов на расстояние, не меньше 1,5–2 мм. В случае пояснительной информации, выполненной в цвете фона, можно не так пристально отслеживать расстояние между пояснительной информацией и другими элементами формы.

Форма с пояснительной информацией, выполненной в черном цвете, является более читаемой. Рекомендуется использовать именно черный цвет для пояснительной информации, если на форме отсутствуют реперы.

Совет. Для настройки шаблона форм с пояснительной информацией, выполненной в цвете фона, сканируйте форму в цвете или оттенках серого, иначе границы элементов формы и пояснительная информация не будут видны.

Рассмотрим приемы обработки фоновых форм. Технологической основой "пропадания" фона при сканировании является отсечение цвета фона с помощью цветной лампы или цветного фильтра. Цвет лампы подбирается в зависимости от того, какой оттенок имеет цвет фона "розово-оранжевый" или "зеленый".

Итак, сканирование производится:

- На цветном сканере с программной фильтрацией красного или зеленого цвета в зависимости от оттенка фона формы.
- На монохромном сканере с "красной лампой" (Red-Lamp) или "зеленой лампой" (Green-Lamp) соответственно (аппаратная фильтрация цвета).
- На монохромном сканере с использованием белой лампы и красного или зеленого фильтра соответственно. Качество фильтрации, в отличие от первого и второго вариантов, хуже – может остаться фон формы или исчезнуть часть значимой информации.
- На монохромном сканере с использованием белой лампы и без фильтра.

При сканировании с использованием белой лампы возможно отсечение различных цветов светлых оттенков без фильтрации. Исчезновение фона достигается за счет слабой насыщенности цвета фона. В этом случае подбор цвета формы и его насыщенность должны производиться опытным путем для каждого конкретного сканера.

 **Замечание.** У многих цветных сканеров, кроме того, есть режим программной фильтрации синего цвета. Но не рекомендуется выполнять формы в "синих" оттенках, поскольку нельзя гарантировать, что все формы будут заполнены только черной, а не темно-синей ручкой. В случае сканирования форм с фильтрацией синего цвета данные, написанные синей ручкой, "исчезнут".

Подбор цвета

Если выбирать между зелеными и красными оттенками форм, то надо иметь в виду, что красно-оранжевые оттенки предпочтительнее, так как они являются максимально контрастными к синему цвету, что улучшает качество полученных при сканировании изображений и распознавания текста, написанного синей ручкой.

В разделе «**Рекомендуемые цвета для разработки цветных фоновых форм**» приводится список рекомендуемых цветов, которые "пропадают" при сканировании форм практически любым сканером. Для каждого конкретного сканера можно также получить у производителя или дилера список цветов в стандарте Pantone (или любом другом), которые можно использовать с данным сканером.

Черно-белые формы с растровым фоном

Черно-белые формы с растровым фоном – это формы с растровым фоном (растр – мелкие черные точки), на котором белыми прямоугольниками выделены знакоместа (а также поля меток и др.), в которые при заполнении формы вносится информация. Фон должен состоять из отдельных точек. Рекомендуемый размер точек растра не более 0,1 мм. Расстояние между точками растра должно быть около 1 мм. Размер точек и расстояние между точками имеют большое значение, т.к. они должны восприниматься программой распознавания как мусор и удаляться с изображения при очистке. Также точки растра не должны "слипаться" при сканировании.

Рассмотрим приемы обработки черно-белых форм с растровым фоном.

Растровый фон не исчезает непосредственно при сканировании. Но при очистке изображения точки растра фильтруются как мусор, и в идеальной ситуации изображение уже не содержит точек растра после очистки.

Черно-белые формы с растровыми границами

Черно-белые формы с растровыми границами – это формы без фона, на которых границы знакомест, полей меток и др. элементов, в которые при заполнении формы вносится информация, выполнены растром (мелкими черными точками). Размер точки растра 0,39 – 0,5 пункта. Рекомендуемый размер точки растра 0,39 пункта. Расстояние между точками растра должно быть в пять раз больше размера самой точки.

Размер точек и расстояние между точками имеют большое значение, т.к. при очистке изображения точки растра должны восприниматься программой распознавания как мусор и удаляться с изображения. Также точки растра не должны "слипаться" при сканировании.

Рассмотрим приемы обработки черно-белых форм с растровыми границами.

Растровые границы не исчезают непосредственно при сканировании. Но при очистке изображения точки растра фильтруются как мусор, и в идеальной ситуации после очистки изображение уже не содержит точек растра.

Черно-белые линейчатые формы

Черно-белые линейчатые формы – это формы без фона (как правило), на которых границы знакомест, полей меток и др. элементов, в которые при заполнении формы вносится информация, выполнены черными линиями.

Границы полей черно-белой линейчатой формы остаются на изображении после сканирования и очистки.

 **Замечание.** Формы этого вида особенно чувствительны к качеству заполнения. Под качеством заполнения подразумевается следующее: при заполнении формы текст помещается внутри ограничителей, и становится крайне важно, чтобы внесенный текст не заезжал на ограничители, а в идеале, и не касался их. В случае касания или "наползания" символов на ограничитель программа распознавания будет воспринимать символы и ограничитель как единое целое и качество распознавания ухудшится.

Выбор вида формы

Кроме учета особенностей дизайна и обработки форм необходимо учитывать следующие характеристики видов форм: сложность изготовления, тиражирование, стоимость печати, размер изображения, режим сканирования, время обработки изображения, качество распознавания, расположение пояснительной информации, привлекательность, качество заполнения. Для удобства сравнения видов форм, мы собрали их характеристики в сводную таблицу:

Критерий	Цветная фоновая форма		Черно-белая растровая форма		Черно-белая линейчатая форма	
	Недостаток	Преимущество	Недостаток	Преимущество	Недостаток	Преимущество
Сложность изготовления		Простота изготовления в любом редакторе.	Ограниченные возможности по созданию форм в текстовых редакторах. Важность размера точек растра.			Простота изготовления в графических редакторах.
Тиражирование	Сложность тиражирования в условиях офиса.			Возможность тиражирования в условиях офиса.		Возможность тиражирования в условиях офиса.
Стоимость печати	Выше стоимость изготовления при типографском способе тиражирования по сравнению с черно-белой формой.			Ниже стоимость изготовления при типографском способе тиражирования по сравнению с цветной фоновой формой.		Ниже стоимость изготовления при типографском способе тиражирования (по сравнению с цветной фоновой формой).
Размер изображения		Меньший размер отсканированного изображения.	Большой размер отсканированного изображения.		Большой размер отсканированного изображения.	
Режим сканирования	Сканирование только в определенных режимах.		Меньшая свобода в подборе яркости и контрастности.		Меньшая свобода в подборе яркости и контрастности.	
Время обработки изображения		Меньше среднее время автоматической обработки.	Больше среднее время обработки за счет необходимости очистки "мусора" на изображении.		Большая вероятность ошибок распознавания при касании или "наползании" символов на ограничители.	
Качество распознавания		Очень высокое качество распознавания.		Высокое качество распознавания.	Сильно зависит от аккуратности заполнения.	При правильном заполнении высокое качество распознавания.

Критерий	Цветная фоновая форма		Черно-белая растровая форма		Черно-белая линейчатая форма	
	Недостаток	Преимущество	Недостаток	Преимущество	Недостаток	Преимущество
Расположение пояснительной информации		Более гибкое расположение пояснительной информации (вплоть до размещения непосредственно внутри полей для ввода).	Наличие жестких требований по размещению пояснительной информации и других элементов формы.		Наличие жестких требований по размещению пояснительной информации и других элементов формы.	
Привлекательность и читаемость	Выбор низкой контрастности фона и цвета пояснительной информации может снизить читаемость формы.	Наиболее привлекательный вид формы.	Менее привлекательные формы. Не очень удобны для зрительного восприятия.		Менее привлекательные формы.	Хорошая читаемость.
Качество заполнения		Менее чувствительны к выходу текста за рамки полей.	При заполнении формы символы должны быть написаны четко и крупно, чтобы они не были удалены вместе с мусором.	Менее чувствительны к выходу текста за рамки полей.	Очень чувствительны к выходу текста за рамки полей.	

Требования к машиночитаемым формам

Познакомившись с элементами машиночитаемых форм, их видами и типами разметки текста, попробуем систематизировать и сформулировать требования, при соблюдении которых созданная форма будет машиночитаемой.

Фон формы

1. По возможности используйте **цветную фоновую форму** или **форму с растровыми ограничителями полей ввода**.
2. При проектировании цветной фоновой формы используйте цвет, фильтруемый при сканировании. Перечень рекомендованных цветов приведен в разделе Рекомендуются цвета для разработки цветных фоновых форм.
3. На цветной фоновой форме пояснительную информацию рекомендуется выполнять в том же цвете, что и фон формы (кроме случаев, когда по каким-либо причинам Вы не можете разместить на форме стандартные реперы и вынуждены использовать черный статический текст в качестве реперов). Если цвета фона и пояснительной информации совпадают, то пояснительная информация может быть размещена внутри полей ввода.
4. Не используйте оттенки синего для фона тех форм, которые будут заполняться от руки. Такой фон может слиться с синей ручкой, что отрицательно скажется на качестве распознавания.

Реперы

Рекомендуется размещать на форме стандартные **реперы** (черные квадраты, уголки, кресты) в количестве 5 штук: четыре – по углам формы (при соединении центров этих реперов должен получаться прямоугольник) и один – на одной из сторон этого прямоугольника.

1. Если по каким-то причинам на форме невозможно разместить стандартные реперы, то рекомендуются следующие комбинации элементов:
 - а) не менее 2-х горизонтальных **разделителей** и 2-х вертикальных разделителей;
 - б) не менее 4-х элементов **Текст** и элемент **Идентификатор**;
 - в) возможны другие комбинации этих элементов.
2. Расстояние от репера до ближайшего элемента формы должно быть не менее 3 мм.
3. Расстояние от репера до края листа должно быть не менее 8 мм.
4. Реперы должны быть выполнены в темном цвете, не исчезающем при сканировании. Лучше всего для этого подходит черный цвет.
5. На форме должны быть реперы, расположенные так, чтобы все заполняемые поля находились внутри ограниченной ими области.

Помимо вышеперечисленных общих требований ко всем реперам, для каждого типа реперов существуют специальные требования:

Черные квадраты

На формах одного типа все квадраты должны иметь одинаковый размер: от 4x4 мм до 8x8 мм. Рекомендуемый размер 5x5 мм. Прямоугольники не допускаются.

Уголки и кресты

1. Длина стороны квадрата, описанного вокруг реперов, должна находиться в интервале от 4 мм до 8 мм. Рекомендуемый размер – 5 мм.
2. Размеры реперов одного вида должны быть одинаковыми внутри одной формы.
3. Стороны реперов должны быть равными.
4. Толщина линий реперов должны быть одинаковыми внутри одной формы и находиться в интервале от 0,3 мм до 1 мм. Рекомендуемая толщина – 0,5 мм.

Элемент Текст (в случае использования элемента в качестве репера)

Должен использоваться шрифт размером не менее 7 пунктов. Для заголовков и крупных надписей рекомендуется использовать шрифт размером не менее 14 пунктов.

Элемент Разделитель (в случае использования элемента в качестве репера)

Толщина линии должна быть не менее 0,8 пункта. Рекомендуемая толщина 0,8–1,5 пункта.

Штрих-код (в случае использования элемента в качестве репера)

1. Ширина штрих-кода (расстояние между крайней левой и крайней правой полосами штрих-кода) – 47–50 мм.
2. Высота штрих-кода (без учета цифр) – 12–15 мм.
3. Ориентация штрих-кода (направление полос) должна совпадать с ориентацией формы.
4. Рекомендуется использовать штрих-код типа EAN 13.
5. Рекомендованное расстояние от штрих-кода до любых элементов формы должно быть не менее 10 мм.

Поле метки

1. Рекомендуемый размер **поля метки** – 3,5х3,5 мм; 4х4 мм; 4,5х4,5 мм; 5х5 мм (в случае ограничителя поля метки типа Круг приведенный размер – это размер квадрата, в который вписан ограничитель)
2. Рекомендуемая толщина линии ограничителя поля метки – 0,4 мм (при любом размере поля метки).

Геометрические параметры разметки текста на формах

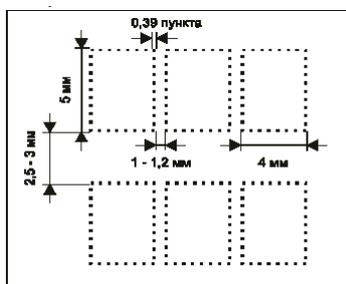
Точки растра:

1. В случае использования **черно-белой формы с растровыми границами** размер точек растра должен составлять 0,39 пункта.
2. Оптимальное расстояние между точками растра – в пять раз больше размера самой точки.

⚠️ Внимание! Обычно угроза "склеивания" точек возникает по углам **знакоместа**, поэтому особое внимание следует уделять именно угловым точкам.

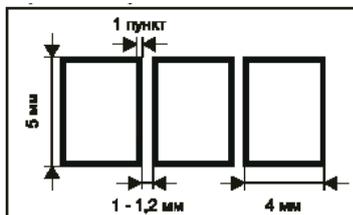
💡 Совет. Если редактор не позволяет задать расстояние между точками растра, необходимо выбрать такой стиль линии, чтобы ее точки были расположены на достаточном расстоянии друг от друга и не "склеивались" на отсканированном изображении.

Размер знакоместа:



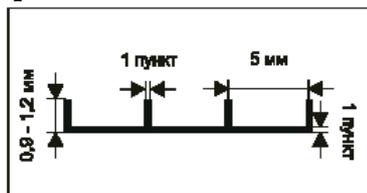
1. Рекомендуемый размер знакоместа – 4х5 мм, меньший размер знакоместа использовать не рекомендуется, т.к. поля с таким размером знакоместа требуют более аккуратного заполнения, что не всегда удастся обеспечить. Также не рекомендуется использовать знакоместа очень большого размера: для получения хорошего результата при распознавании такие поля требуется заполнять непривычно большими буквами. Размер знакоместа должен примерно соответствовать общепринятым нормам письма.
2. Рекомендуемое расстояние между строчками: не менее 2,5–3 мм. Расстояние между знакоместами по горизонтали: не менее 1–1,2 мм.

Толщина линий разметки текста:



Если границы представлены черными линиями, что соответствует типам разметки Текст в рамке, Буквы в изолированных рамках или Буквы в рамках, – толщина линий должна составлять 1 пункт.

Гребенки:



1. Рекомендуемый шаг гребенки в пределах одного поля формы: 5 мм.
2. Рекомендуемое расстояние между строчками: не менее 7,5–8 мм.
3. Рекомендуемая высота «гребенки»: 0,9–1,2 мм.
4. Рекомендуемая толщина линий гребенки: 1 пункт.
5. В случае выполнения границы в виде «гребенки в рамке» минимальное расстояние между рамками должно быть 2,5–3 мм.

Расположение элементов формы

1. Рекомендуется оставлять по краям формы отступы. Это требование вызвано тем, что элементы, размещенные слишком близко к краям формы, могут пропасть при сканировании, что приведет к потере части информации.
2. Рекомендуемое расстояние между элементами формы: не менее 2 мм.
3. Рекомендуемое расстояние между **полем метки** и поясняющим текстом: не менее 2/3 размера поля метки.
4. Пояснительная информация должна размещаться на форме так, чтобы она не пересекалась с полями для ввода информации.
5. Размещая на форме пояснительную информацию, выполненную в черном цвете, необходимо отступить от других элементов не менее 1,5–2 мм. Если пояснительная информация выполнена в цвете фона, можно не отслеживать расстояние между пояснительной информацией и другими элементами формы.

Тиражирование форм

При изготовлении форм необходимо выполнять следующие требования:

1. Формы могут быть изготовлены типографским способом или распечатаны на принтере. **Цветные фоновые формы** рекомендуется изготавливать в типографии, в противном случае необходимо тестирование цвета формы на сканере. Необходимо проводить тестирование пробного тиража, полученного из типографии.
2. Все экземпляры одного типа должны быть отпечатаны с одного образца. Они должны "совпадать на просвет".
3. Поля формы (расстояние от края листа до ближайшего объекта) должны быть не менее 8 мм, рекомендуется делать поля не менее 12 мм.
4. Допустимое отклонение по линейным размерам элементов на форме не более 0.15%, что составляет 0,5 мм при формате А4.
5. При печати формы нужно использовать лазерные принтеры с разрешением не менее 600 точек на дюйм.
6. Желательно распечатывать формы на одном и том же принтере или на разных принтерах одной и той же модели.

Совет. Не рекомендуется множить машиночитаемые формы с использованием копировальной техники. В этом случае происходит изменение размеров и смещение элементов формы: утолщение рамок или растра ограничителей полей ввода, изменение насыщенности цветов, что ухудшает качество распознавания.

Заполнение форм от руки

Цвет и вид чернил

Формы лучше всего заполнять черной перьевой, капиллярной, гелевой или шариковой ручкой. Темно-синяя или фиолетовая ручка тоже дает хорошие результаты (особенно на цветных фоновых формах). Фломастер приводит к более низкому качеству распознавания из-за заливания мелких деталей. Хуже всего распознаются карандаш и светлая цветная ручка.

Размер и написание букв

Для улучшения качества распознавания формы лучше всего заполнять отдельными прописными буквами. Размер буквы должен быть таким, чтобы буква занимала все пространство знакоместа, но не касалась его границ.

Если знакоместа явно выделены на форме, т.е. используются следующие типы разметки текста: буквы в рамках, буквы в изолированных рамках, то проще управлять заполняющим форму, т.к. требуемый размер букв явно указан знакоместом и явное наличие знакомест мешает написать буквы слитно.

Сложнее ситуация при использовании разметки текста следующих типов: текст в рамке с гребенкой или буквы в гребенке. Здесь задается ширина знакоместа, но из-за отсутствия полных боковых ограничителей ничто не мешает заполняющему форму написать буквы слитно. И если в первом случае высота знакоместа указана явно, то во втором случае может возникнуть ситуация, когда заполняющий форму впишет слишком маленькие или слишком высокие буквы.

В разметках текста "текст в рамке" или "текст по линии" вообще не содержится информации о ширине знакоместа. Поэтому ничто не мешает заполняющему форму написать буквы слитно, что приведет к значительному ухудшению качества распознавания. В разметке текста "текст по линии" также отсутствует информация о высоте знакоместа, поэтому этот тип разметки не рекомендуется применять для форм, требующих высокого качества распознавания данных.

Помимо размера букв и отдельности написания, важным требованием является требование отсутствия "наползания" символов на ограничители. Это требование не так актуально для цветной фоновой или черно-белой растровой формы, т.к. ограничители в этих формах пропадут после сканирования и очистки изображения. А для черно-белой линейчатой формы это требование очень актуально, т.к. нарушение этого требования влечет за собой значительное ухудшение качества распознавания.

Образец заполнения

Чтобы обратить внимание респондента на перечисленные особенности заполнения формы, на форме рекомендуется указывать образец заполнения. Обычно он содержит рекомендации по заполнению данной формы и образцы букв, цифр и символов, необходимых для заполнения.

ВНИМАНИЕ! Необходимо заполнять шариковой, капиллярной или гелевой ручкой **ЧЕРНЫМИ** или **ТЕМНО-СИНИМИ** чернилами **ЗАГЛАВНЫМИ ПЕЧАТНЫМИ БУКВАМИ** по следующим образцам (Ответы на вопросы отмечайте или):
А Б В Г Д Е Ж З И Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я | 1 2 3 4 5 6 7 8 9 0

Рекомендуемые цвета для разработки цветных фоновых форм

Таблицы содержат примеры цветов фона (с номерами соответствующих пантонов), "исчезающих" при сканировании

- любым сканером;

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			

- цветным сканером с фильтрацией красного или монохромным сканером с красной лампой/красным фильтром.

100U	101U	102U	1905U	190U	1915U	488U	489U	495U
106U	107U	108U	191U	1923U	192U	496U	5035U	503U
109U	113U	114U	196U	197U	198U	509U	510U	5175U
115U	116U	1205U	199U	203U	204U	517U	5245U	530U
120U	1215U	121U	205U	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	2375U	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 02U
165U	169U	170U	461U	4685U	468U	Process Yellow U	Process Magenta U	Purple U
1787U	1788U	178U	473U	474U	4755U	Red 03U	Reddened Red U	Reddened Red U
182U	183U	184U	475U	481U	482U	Warm Gray 1U	Warm Red U	Yellow U
185U	1895U	189U	485U	486U	487U			