

**ДАГАЕВ Михаил Владимирович** – главный специалист отдела поддержки электронных библиотек Российской государственной библиотеки  
e-mail: [costandy@list.ru](mailto:costandy@list.ru)

## **ОСОБЕННОСТИ СЖАТИЯ ИЗОБРАЖЕНИЙ «ТЕКСТ С ИЛЛЮСТРАЦИЯМИ» ПРИ СОЗДАНИИ БОЛЬШИХ ЭЛЕКТРОННЫХ КОЛЛЕКЦИЙ**

В последние 10-12 лет заметно вырос объем оцифровываемых изданий — как в России, так и во всем мире. До конца 90-х книги, периодика, архивы и тому подобные материалы обычно сканировались, сохранялись в архивах оцифрованных изображений, размещались в открытых для общего доступа сетевых хранилищах как результат отдельных, не всегда связанных друг с другом, проектов. Сейчас же все чаще и чаще подобное происходит как результат стратегических, долгосрочных инициатив, исходящих и от частных структур (Google Books), и от правительств (библиотека имени Б.Ельцина у нас, проект Europeana в ЕС, проект WRLD в США).

Объем оцифрованных изданий, с прежних нескольких десятков и сотен тысяч экземпляров, увеличился уже до миллионов. В результате чего создателям электронных библиотек постоянно приходится искать ответ на вопрос: «В каком виде желательно хранить оцифрованные изображения?». Так, чтобы при этом были бы соблюдены два вроде бы взаимоисключающих требования — как можно меньшие объемы хранения и как можно более точное соответствие сохраняемого изображения оригиналу.

Когда же речь идет не только о самом хранении, но и об использовании, то добавляется еще и третье требование — возможность удобной работы с оцифрованными изображениями в условиях Интернет-доступа.

Если придать этим общим формулировкам более конкретный вид, то получается примерно следующее.

### ***Объемы хранения***

Средненормальный объем страницы формата А5, отсканированной в полутонах серого<sup>1</sup> может колебаться от 2-3 до 10-12 мегабайт (в зависимости от используемого разрешения сканирования). Средненормальный объем страницы этого же формата, отсканированной в цвете, в режиме Truecolor (24 бита на пиксел изображения) уже будет колебаться от 8-10 до 30-40 мегабайт (тоже в зависимости от используемого разрешения сканирования).

При этом сказанное касается только изданий, отпечатанных в массовых полиграфических форматах, где размер одной страницы примерно соответствует формату А5 (148x210 мм). Если же иметь в виду весь спектр печатной продукции, то не так уж и редко встречается, что на одну страницу может приходиться 80-100 мегабайт и более.

В пересчете на книгу со средним листажом в 300-400 страниц получим следующие ориентировочные цифры. Для сканирования в Grayscale занимаемый объем будет от одного до 4-5 гигабайт на одно издание. Для цветного сканирования диапазон окажется от 2-3 до 10-15 гигабайт. А для изданий с большим листажом и/или отпечатанных в нестандартных форматах вполне нормальными значениями будет несколько десятков гигабайт хранимого объема.

### ***Соответствие оригиналу***

При работе с оцифрованным изображением на уровне общего просмотра материала необходимо, чтобы соблюдались следующие условия:

- возможность читать текстовую часть страницы без затруднений (и тем более без ошибок). То есть литеры текста должны иметь относительно четкий, малоразмытый вид, легко отделяться взглядом от фона страницы. Аналогично для линий штриховых рисунков, чертежей, диаграмм и т.д.

- иллюстративная часть (рисунки, фотографии, цветная заливка страницы или ее части) должна иметь вид, максимально близкий к печатному оригиналу. То есть оцифрованное изображение по отношению к оригиналу должно иметь минимум отличий в цветовой палитре (в идеале не иметь их вообще) и в характере перехода оттенков друг в друга.

Если оцифрованное изображение требуется не для общего ознакомления с изданием, а для профессиональной деятельности (историки, архивисты, реставраторы), то к перечисленному добавляется еще и

---

<sup>1</sup> Этот тип изображений также обозначается как Grayscale.

условие максимальной сохранности *всего* внешнего вида каждой из страниц, вплоть до фактуры бумаги, мелких деталей шрифта, изношенности издания, повреждений, рукописных пометок (когда и если таковые присутствуют на странице) и так далее.

### ***Интернет-доступ***

В качестве рабочей модели принимаем работу с книгой/журналом в режиме «постраничный просмотр с перелистыванием». При этом просматриваемое издание не копируется целиком на компьютер пользователя (чтобы на нем же и просматриваться), а находится, там где размещена используемая электронная библиотека, т.е. на одном из Интернет-серверов.

Наблюдениями в области Интернет-юзабилити определено, что нормально-ожидаемый срок открытия в браузере просматриваемого изображения составляет не более 5-6 секунд. Если пользователь очень заинтересован в просматриваемом материале, то эту цифру можно увеличить до 8-10 секунд, максимум. Далее же человек обычно начинает искать необходимые ему данные в более быстродоступном виде.

Если исходить из скорости доступа через модемное подключение (что в России, за пределами столиц, все еще достаточно частое явление), то это будет не более 56 Кбит/сек (7 Кбайт/сек). Если исходить из скорости доступа через подключение по локальной сети или через ADSL-доступ, то это будут заметно большие цифры. Однако случаи прямого (или почти прямого) подключения пользователя к необходимому сайту с оцифрованными изданиями в общем виде явление не такое уж частое. Обычно связь проходит через цепочку маршрутизаторов, на которых часто возникает эффект «бутылочного горлышка», существенно снижающий конечную скорость доступа к данным.

С учетом сказанного средненормальную скорость модемного доступа в первом приближении можно принять за 200-300 Кбит/сек (25-35 Кбайт/сек).

Соответственно порядок максимально допустимого размера для одной просматриваемой страницы/изображения составит 40-60 Кбайт (при модемном доступе) и 150-300 Кбайт (при доступе через другие источники подключения к Сети).

Причем не надо забывать, что это именно *максимально* допустимые значения. Те цифры, которыми желательно оперировать при реальном планировании объемов выкладываемых на сайте изданий, можно спокойно уменьшить раза в полтора, а то и в два.

Теперь рассмотрим, насколько вписываются в эти требования основные используемые форматы сжатия.

### **Форматы GIF, PNG, JPEG.**

Три перечисленных формата сжатия графики являются наиболее распространенными в интернете. Их также называют форматами *первой волны*.

Время их создания — конец 80-х, начало 90-х. Основными отличительными особенностями можно считать:

- а) ориентация преимущественно на обработку фотографий и computer-generated изображений как наиболее преобладающих во времена их создания разновидностей графики;
- б) разработка в расчете на изображения, просматриваемые каждое по отдельности, а не как связная структура уровня книги или журнала;
- в) только изображения этих форматов можно открывать в Web-браузерах без дополнительных плагинов, лишь за счет возможностей самих браузеров.

Как нетрудно заметить из нижеследующего изложения, первые два из них, GIF и PNG, для создания размещаемых в интернете изданий малопригодны. Данные по ним приводятся скорее для полноты обзора.

### **GIF.**

По своей сути является форматом сжатия без потерь. Поэтому сколь-либо заметный выигрыш в объеме сжатого изображения можно получить только двумя способами. Или за счет сужения цветовой палитры конечного изображения с 16 миллионов оттенков Truecolor до используемых в GIF 256 оттенков. Или за счет сжатия больших пространств, залитых одним и тем же цветом.

Первый способ может дать кратность сжатия не более чем в три раза (уменьшение пространства для записи данных о цвете с трех байт на пиксел до одного байта).

Второй на отсканированных изображениях тоже не дает таких больших кратностей, как на computer-generated графике, потому что выходная продукция сканеров всегда содержит заметный объем флюктуационных шумов. В результате чего достаточно большие участки изображения, для человеческого

глаза совершенно однородные, с точки зрения процесса сжатия однородными не являются. И поэтому упаковываются достаточно плохо.

Нормально-ожидаемым коэффициентом сжатия<sup>1</sup> для изображений, отсканированных в TrueColor, будут значения не более 15-20. Что, соответственно, даст объемы сжатых файлов в несколько Мб и более на одно изображение. Для Grayscale-графики КС будет не более 5-6 раз, что даст объемы GIF-файлов, сравнимые с цветными изображениями.

С учетом того, что файлы с цветными изображениями обычно имеют значения объемов в десятки Мб, для сжатия многостраничных изданий, размещаемых в Интернет, формат GIF подходит достаточно мало.

### **PNG.**

Как и GIF является форматом сжатия без потерь. При этом он использует практически те же алгоритмы обработки оцифрованных изображений, что и GIF, только в немного модифицированном варианте. А, значит, возможности по КС для изображений, отсканированных в цвете или полутонах серого, у него мало чем будут отличаться от GIF. Если же учесть, что PNG не сокращает палитру сжимаемого изображения до 256 оттенков как GIF, а стремится сохранить ее в исходном виде, то ожидаемые КС будут еще меньше, чем у GIF.

Соответственно, для обработки выкладываемых в интернете изданий он пригоден в еще меньшей степени чем GIF.

### **JPEG.**

В отличие от двух предыдущих это формат сжатия с потерями данных. И, значит, получаемое изображение будет иметь тот или иной объем отличий от оригинала. Следовательно, основной рабочей задачей становится не столько получить изображение, сжатое до требуемого значения, сколько подобрать необходимый баланс между выходным объемом и получаемым при этом качеством сжатого изображения.

При оценке применимости JPEG в той или иной области, важно помнить, что изначально он создавался для сжатия изображений полутонового типа, т.е. таких, которые состоят из большого объема плавных переходов одного цвета в другой. Для рассматриваемых же в статье изображений «текст с иллюстрациями» характерна другая особенность — наличие заметного количества резких переходов «цвет-цвет» на границе отдельных литер текста и общего фона страницы.

Поэтому при JPEG-сжатии изображений рассматриваемого типа возможно возникновение следующих проблем:

1. резкое искажение первоначальной цветовой палитры, вплоть до полной ее неузнаваемости;
2. появление на изображении бесформенных цветных пятен с однородной заливкой;
3. появление каймы вокруг литер текста и других границ с резким переходом «цвет-цвет».

#### *Цветной JPEG*

1. С точки зрения размеров файла (а значит и скорости открытия его в браузере), наиболее востребованными будут КС, равные 100 и более. Однако в этой области начинают проявляться все три вышеописанных разновидности искажений.

2. Искажения третьего типа (кайма вокруг границы с резким переходом цветов) становятся близки к нулю в районе КС 50-60. Чтобы при этом получить приемлемые объемы JPEG-файлов, надо, чтобы размер оригинала был заметно меньше А5, а разрешение сканирования — не более 300 dpi.

3. Искажения цветов перестают быть заметными при беглом взгляде, начиная со КС 30-35. В этом случае размер практически любого цветного изображения после сжатия будет превышать рабочий максимум для размещения в интернете.

#### *Grayscale JPEG*

Для этой разновидности JPEG несомненным плюсом можно считать то, что исходные изображения имеют заметно меньший размер, чем их Color-аналоги. Обратной стороной такого преимущества будет то, что сжимаются они заметно хуже. При одинаковых установках JPEG-сжатия разница в КС между Grayscale и Color-изображениями может достигать 1.5-2, в пользу последних.

Таким образом, максимальный КС, которого возможно достигнуть при получении Grayscale, на средненормальной странице из книги или журнала, это 60-70. Следовательно, JPEG-изображения, по своим размерам пригодные для размещения в интернете, можно получить из оригиналов, не превышающих 10-12 Мб - это если считать по максимально возможному сжатию и пренебречь возникающими при этом искажениями. Или же 4-5 Мб, если выбирать такие КС, которые не дают искажений, заметных при беглом взгляде.

---

<sup>1</sup> В дальнейшем — КС.

Ситуацию несколько облегчает то, что искажение «кайма вокруг границы», наиболее мешающее нормальному восприятию текста, здесь проявляется при заметно меньшем КС, чем в Color JPEG. Даже на предельно возможных КС общее восприятие вида литер текста ухудшается до уровня трудночитаемости далеко не во всяком случае. А при КС 25-30 этот тип искажений практически сходит на нет и становится незаметным даже при просмотре с увеличением.

Искажение типа «бесформенные цветные пятна с однородной заливкой» полностью исчезают примерно в этом же диапазоне КС, но перестают мешать восприятию уже в при его значении 40-45.

Таким образом, видно, что применимость формата JPEG для сжатия цветных и Grayscale изображений достаточно ограничена.

Можно работать только с такими вариантами, когда требуемые объемы JPEG-файлов реально получить без заметного ухудшения качества оригинала. В этом случае надо использовать изображения с форматом страницы, близким к А5 и разрешением сканирования около 300 dpi.

Или же можно сознательно пожертвовать качеством получаемого изображения ради получения приемлемых объемов JPEG-файлов. Тогда диапазон обрабатываемых форматов увеличивается до А4 (если разрешение не более 300 dpi) или же остается близким к А5, но с разрешениями сканирования от 400 до 600 dpi.

#### Форматы сжатия изображений второго поколения.

Форматы второго поколения, конца 90-х, начала 2000-х, как правило, создавались или на основе одной из реализаций wavelet-алгоритмов, или на основе технологии MRC (Mixed Raster Content, растровое содержимое смешанного типа).

#### **JPEG 2000 как характерный пример использования wavelet-алгоритмов**

Wavelet-сжатие (достаточно известным образцом которого можно считать формат JPEG 2000), образно формулируя, выигрывает у своего предшественника, обычного JPEG, за счет грубой силы. Благодаря используемому в нем математическому аппарату, JPEG 2000, требуется заметно меньший объем информации (сравнительно с JPEG), чтобы восстановить исходное изображение из сжатого. А, значит, существенно уменьшается количество бит, приходящееся в сжатом файле на аналогичный объем данных.

Общая характеристика процесса сжатия в JPEG 2000 выглядит так.

#### *Цветные изображения*

При КС 30-40 разница между оригиналом и сжатым изображением практически незаметна даже при масштабе просмотра 1:1.

При КС в районе 100 возникает некоторый объем нарушений внешнего вида (аналоги искажений первого и третьего типа для обычного JPEG), но он почти не влияет на общую читаемость текста на странице и мало влияет на цветовую палитру. Заметные для беглого взгляда изменения возникают в основном у рисунков с фотореалистическим качеством и фотографий профессионального уровня.

При КС в районе нескольких сотен возникают заметные искажения того же типа, что уже описывались для JPEG (бесформенные цветные пятна с однородной заливкой, кайма вокруг литер текста).

#### *Grayscale изображения*

Аналогично цветным изображениям при КС 30-40 разница между оригиналом и сжатым изображением практически незаметна.

При КС около 100 какая-то заметная разница отмечается только на небольшой части изображений, там где на странице текст набран мелким кеглем или же имеются мелкодетализированные штриховые рисунки.

При КС в районе нескольких сотен аналогично цветным изображениям также возникают заметные искажения «бесформенные цветные пятна с однородной заливкой», «кайма вокруг литер текста».

Основным недостатком wavelet-сжатия является большая ресурсоемкость как процесса сжатия, так и процесса открытия сжатого изображения в программах просмотра и в браузерах.

Требуемые для самого сжатия ресурсы могут заметно отличаться в зависимости от используемых программ. Или, если точнее, от того, в какой степени разработчик сумел оптимизировать процесс обработки изображения во время сжатия. Однако, если оценивать ресурсоемкость по сравнительной скорости открытия JPEG2000 и JPEG-изображений в программах просмотра (при условии одного и того же исходного оригинала и одного и того же КС), то разница будет составлять 3-4 и более раз (в зависимости от объема оригинала и используемой реализации JPEG2000-формата).

Еще одним недостатком, затрудняющим более широкое использование формата JPEG 2000, является его заметно более сложный (сравнительно с обычным JPEG) математический аппарат. В результате, для качественной реализации формата в программном коде требуются трудозатраты, доступные больше профессиональным командам, чем программистам-одиночкам.

### **Форматы, основанные на технологии MRC**

Форматы такого типа используют принцип обработки, радикально отличный как от форматов первой волны, так и от wavelet-сжатия. Если во всех остальных форматах изображение обрабатывается как единое целое, вне зависимости от содержания, то при использовании технологии MRC оно перед собственно сжатием проходит через предварительную операцию, называемую «разделение на слои».

Суть ее сводится к тому, что в изображении выделяются структурные элементы трех типов: текстовая составляющая (литеры текста, штриховые рисунки, чертежи, диаграммы и т.д.), полутоновая составляющая (все области, в которых имеются плавные переходы от одного цвета к другому) и области, залитые одним и тем же цветом, с резким переходом между самой областью и окружающим пространством.

В дальнейшем эти составляющие (или, как их еще называют, «слои») обрабатываются алгоритмами сжатия, используемыми в данном формате, независимо друг от друга. И соединяются в одно результирующее изображение только в момент вывода на экран/печать.

Среди MRC-форматов на данный момент наиболее распространены DjVu и LDF (Luratech Document Format). Поскольку конструктивно они достаточно схожи, то в дальнейшем их можно рассматривать как один формат, но в различной реализации.

Первое, наиболее бросающееся в глаза отличие MRC-форматов от их «коллег» — это порядок КС для сжатых изображения с относительно небольшими искажениями. Если для JPEG рабочей нормой такого КС будет несколько десятков, для JPEG 2000 — цифры порядка 100, то для DjVu и LDF вполне ожидаемыми значениями КС могут быть от 300 до 600, а в случае хорошо сжимаемых изображений даже близкие к 1000.

Вторая особенность — целевая ориентация форматов. GIF, PNG, JPEG, JPEG 2000 создавались преимущественно для обработки отдельных, структурно между собой не связанных, изображений. Разумеется, их можно применить для любой цели, в том числе и для создания сжатых книг/журналов, но в процессе повседневной работы достаточно быстро становится ясным, что к подобной работе они приспособлены не лучшим образом.

Что же касается MRC-форматов, то здесь картина полностью обратная. И DjVu, и LDF способны достаточно качественно сжимать отдельные изображения, но наиболее ярко их возможности проявляются при обработке многостраничных изданий (книг, газет, журналов). Здесь и высокие значения КС, и возможность создания многостраничных файлов по принципу «один файл - одна книга», и возможность создавать удобную структуру навигации по таким файлам, и некоторые другие полезные моменты.

К сожалению, в данном случае затруднительно представить работу процесса сжатия по уже применявшейся ранее схеме — «диапазон значений КС — наблюдаемые при этом искажения».

Это связано с тем, что если в предыдущих случаях пользователь имеет в своем распоряжении только одну регулируемую настройку — задаваемую величину сжатия, то в MRC-форматах количество регулируемых параметров может быть три-четыре и более. Плюс к тому влияние данного конкретного параметра на КС может быть не прямым, а опосредованным и/или зависящим от значений других параметров. Подобная ситуация делает малопродуктивной обычную схему анализа «изменение отслеживаемого параметра — получаемое при этом значение функции».

В зависимости от содержимого сжимаемого изображения и от заданных настроек сжатия КС для Color изображений может колебаться от 50-100 до 900-1600. Аналогичные величины для Grayscale изображений составляют от 20-40 до 300-500. Надо отметить, что даже при максимально возможных КС сжатые изображения сохраняют необходимый минимум читаемости по отношению к текстовой части. От искажений сжатия по большей части страдают рисунки и иллюстрации.

Наиболее выигрышными, с точки зрения возможности максимального высокого сжатия, изображениями можно считать те, где фон страницы — белый или близкий к нему цвет. Именно поэтому Grayscale изображения сжимаются заметно хуже, чем Color. Программа сжатия воспринимает серый цвет фоновой части страницы как разновидность цветной заливки (при том, что цветной фон, близкий к белому, часто воспринимается именно как белый).

С точки же зрения основного содержимого, наиболее высокие КС дают страницы, где минимум рисунков (или того, что программа сжатия может воспринять, как рисунок) и максимум текста.

Проблемы, возникающие при работе с MRC-форматами, являются обратной стороной их преимуществ. Традиционно таких проблем наблюдается, как минимум, две — ресурсоемкость процесса и качество разделения изображения на отдельные слои.

Технология, требующая сначала разделить исходное изображение на слои, потом обработать каждый из них по отдельности и, наконец, каждый раз при выводе изображения собирать его из полученных слоев, не может не иметь заметную ресурсоемкость как при создании сжатого изображения, так и при его просмотре.

Разумеется, создатели форматов сделали все возможное, чтобы облегчить и ускорить работу с изображениями на всех этапах. Также надо учитывать, что рабочие возможности компьютеров растут от года к году.

Но и в DjVu, и в LDF, и в других менее известных MRC-форматах обычно используется какой-нибудь из wavelet-алгоритмов (ими традиционно сжимается полутоновой слой). Ресурсоемкость wavelet-сжатия достаточно велика, причем велика как на этапе самого сжатия, так и на этапе вывода на экран/печать. Поэтому в ближайшие времена вряд ли стоит ожидать, что скорость вывода на экран DjVu или LDF-изображения сравняется со скоростью вывода более привычного JPEG.

Вторая проблема — качество разделения на слои — вытекает из того, что одним из ключевых критериев для определения принадлежности данных к тому или иному слою принимается вид перехода между соседствующими цветами. Или это резкие переходы между литерами текста и фоном страницы, между цветами, которыми залиты отдельные участки плаката/карты, между линиями диаграмм и окружающим фоном. Или это плавные переходы характерные для фотографии, рисунка, фоновой части страницы.

Однако при не очень качественной оцифровке в тех местах, где в оригинале наблюдаются резкие переходы, могут возникнуть плавно-мягкие и наоборот. Фон страницы может оказаться перетемнен и, как следствие, снизится контрастная разница между текстом и фоном. Палитра полутоновой части может оказаться искаженной, в результате чего возникнут резкие переходы оттенков, там где в оригинале они были плавными.

Как следствие, алгоритм отделения одних частей изображения от других не всегда срабатывает, как требуется, и в сжатом изображении возникают искажения. Часть строки с текстом оказывается отнесенной к полутоновой части. Темные области художественной фотографии, наоборот, воспринимаются как участки с текстом. Одна половина сложной диаграммы может оказаться классифицированной как слой «текст», а другая — как слой «полутоновое изображение». И так далее, и тому подобное.

К сожалению, на данный момент неизвестно о достаточно качественно работающих алгоритмах разделения исходного изображения на слои. Таких, которые бы позволяли избежать вышеперечисленных ошибок. Поэтому появление в изображениях, сжатых с использованием технологии MRC, искажений, связанных с некорректным разделением на отдельные слои исходного изображения на слои, пока являются неизбежной платой за высокие значения КС.

#### Описание технической части.

Образцами изображений для анализа служили страницы из:

- современного технического журнала (отсканированы в цвете и в Grayscale, формат А4, разрешение 600);
- технического журнала начала 20-го века (отсканированы в Grayscale, формат близкий к А4, разрешение 600);
- книги современного издания (отсканированы в цвете и Grayscale, формат А5, разрешение 300);
- книги издания 30-х годов (отсканированы в Grayscale, формат близкий к А5, разрешение 600);
- книги по истории архитектуры конца 19 века (отсканированы в цвете, формат А2, разрешение 300 и в Grayscale, формат А4, разрешение 600);
- нотного альбома (отсканированы в Grayscale, формат А4, разрешение 300);
- книги 17-го века на церковнославянском (отсканированы в цвете, формат близкий к А4, разрешение 300);
- книги 18-го века (отсканированы в Grayscale, формат близкий к А2, разрешение 300).

Такой достаточно широкий диапазон видов изданий и времени их печати был выбран с целью более полно отобразить возможный разброс в качестве бумаги и типографской печати, имевший место в разные времена. Поскольку от этих двух параметров, как правило, более всего зависит качество получаемых оцифрованных изображений.

Сжатие выполнялось в программах XnView, версия 1.96 (для всех обсуждаемых форматов, кроме DjVu) и DjVu Document Express Enterprise Editor, версия 5.1 (для формата DjVu).

**Литература:**

1. Patrick Haffner, Léon Bottou, Paul Howard, Yann Le Cun. *DjVu : Analyzing and Compressing Scanned Documents for Internet Distribution // Proceedings of the International Conference on Document Analysis and Recognition. ICDAR. - 1999. - pp. 625-628.*
2. Léon Bottou, Patrick Haffner, Paul G. Howard, Patrice Simard, Yoshua Bengio, Yann Le Cun. *High Quality Document Image Compression with DjVu // Journal of Electronic Imaging. - 1988. - vol.7. - №3. - p. 410-425.*
3. Сэломон, Д. *Сжатие данных, изображений и звука. - Москва: Техносфера, 2004. - 368 с. ISBN 5-94836-027-X.*
4. Миано, Дж. *Форматы и алгоритмы сжатия изображений в действии. - М.: Издательство «Триумф», 2003. — 336 с.: илл.*
5. Земсков, А.И., Шрайберг, Я.Л. *Электронная информация и электронные ресурсы: публикации и документы, фонды и библиотеки. - М.: Издательство «ФАИР», 2007. — 528 с.: илл.*