

УДК 778.14.072.2

## ОСНОВНЫЕ НАПРАВЛЕНИЯ СОВРЕМЕННЫХ ЗАРУБЕЖНЫХ ИССЛЕДОВАНИЙ ПО ПРОБЛЕМЕ СОХРАНЕНИЯ ЦИФРОВОЙ ИНФОРМАЦИИ НА МИКРОФИЛЬМАХ

Гаврилин А.П., Завалишин П.Е.

Научно-исследовательский институт репрографии, Тула, e-mail: fgupniir@yandex.ru

Приводится описание и анализ основных направлений современных зарубежных исследований по проблеме сохранения цифровой информации на микрофильмах по материалам публикаций ведущих европейских специалистов. Основной проблемой является то, что существующие системы хранения цифровой информации в силу быстрого устаревания программно-аппаратных платформ и носителей данных не в состоянии обеспечить долговременное (более 100 лет) сохранение электронной информации и ее адекватное воспроизведение будущими поколениями технических средств. В этой связи в ведущих странах мира единственным на сегодняшний день носителем, пригодным для долгосрочного, надежного и стабильного хранения цифровой информации, признан микрофильм, срок службы которого достигает 500 лет. При современном уровне развития микрографической техники существует возможность сохранения цифровой информации на микрофильме на битовом уровне (в виде потоков двоичных данных) – т.е. в том виде, в котором цифровая информация существует непосредственно, с последующим ее считыванием сканирующими устройствами. В статье рассмотрены различные технические подходы, позволяющие долговременно сохранять любую цифровую информацию на микрофильмах при помощи двухмерных растровых штрих-кодов. Применение данных технологий позволит значительно приблизиться к решению проблемы долгосрочного архивного сохранения электронных документов.

**Ключевые слова:** долговременное хранение, электронный документ, микрофильм, двухмерный штрих-код

## MAIN AREAS OF MODERN FOREIGN INVESTIGATION ON THE CONSERVATION OF DIGITAL DATA ON MICROFILM

Gavrilin A.P., Zavalishin P.E.

Scientific research Institute of reprography, Tula, e-mail: fgupniir@yandex.ru

Description and analysis of main areas of modern foreign investigations on conservation of digital data on microfilm, based on the issues by the leading European professionals. The key problem is the inability of current digital information storage systems to ensure long-term (over 100 years) storage of information and its proper reproduction in the future due to the quick obsolescence of firmware and media. In this connection, the leading countries consider microfilm with its life expectancy over 500 years to be the only media suitable for long-term, reliable and stable storage. The current microfilming technology development allows storing digital information on microfilm on bit level (as binary data streams): it's the ingenious aspect of digital information that can be read by scanning devices. The paper reports on the technical approaches to the long-term storage of any digital information on microfilm by using 2D bitmap barcodes. The use of these techniques allows advancing in finding a solution for the long-term digital data archiving.

**Keywords:** long-term storage, digital records, microfilm, 2D barcode

В современном мире новых информационных технологий все большее количество документов создается, обрабатывается, используется и хранится в электронном виде. Электронный документооборот стремительно вытесняет традиционные бумажные носители информации в сфере деятельности органов государственной власти, архивных и библиотечных учреждений, промышленных предприятий и других государственных и общественных организаций. Преимущества электронного документооборота общеизвестны и не подлежат сомнению, однако при этом возникает ряд проблем, важнейшей из которых является обеспечение долговременного архивного сохранения ключевой электронной документации для ее страхового восстановления в случае возникновения непредвиденных обстоятельств.

Существующие системы хранения цифровой информации в силу быстрого устаревания

программно-аппаратных платформ и носителей данных не в состоянии обеспечить долговременное (более 100 лет) сохранение электронной информации и ее адекватное воспроизведение будущими поколениями технических средств. Для выхода из этой ситуации было опробовано несколько решений, основанных, по большей части, на регулярной миграции данных как с устаревших носителей на новые, так и из одного формата в другой. Однако, особенно в последнем случае, опыт показал, что в процессе миграции электронный документ подвергается изменениям, некоторые из которых могут оказаться нежелательными ввиду его дальнейшего использования, особенно если эти изменения будут накапливаться с каждой новой миграцией. Поэтому многие инициативы и проекты рекомендуют всегда поддерживать цифровую версию объекта в оригинальном формате, чтобы при необходимости иметь возмож-

ность реконструкции и доступа, например, с помощью эмуляции исходной программной среды.

Кроме того, для большинства систем хранения необходима поддержка стабильной резервной копии всех данных (включая версии миграций) помимо актуальной версии для текущего использования. К сожалению, большинство технологий хранения цифровых данных не предназначены для этих целей: для RAID-массивов жестких дисков необходима постоянная работа, и их нужно заменять каждые несколько лет. Данные на ленточных носителях и оптических дисках необходимо периодически переписывать, чтобы информация оставалась доступной. Более того, при текущем уровне прогресса долговечность ленточных и оптических носителей превосходит срок службы и технической поддержки соответствующих считывающих устройств, из-за чего многие ленты и диски могут стать нечитаемыми, если данные с них не были вовремя перенесены на новые типы носителей.

Актуальность проблемы долгосрочного сохранения цифровой информации имеет большое значение и остро осознается всем мировым сообществом. Это заставляет исследователей разных стран искать, обосновывать и практически апробировать другие, более надежные и экономичные стратегии сохранения важнейшей электронной информации с помощью стабильных технологий долговременного хранения, не требующих постоянного обновления и поддержки. В этой связи в ведущих странах мира единственным на сегодняшний день носителем, пригодным для долгосрочного, надежного и стабильного хранения цифровой информации признан микрофильм, срок службы которого достигает 500 лет. При этом сохранение электронной информации на микрофильмах невозможно без использования специальных устройств записи цифровой информации на микропленку – СОМ-систем. На современном рынке микрографического оборудования данные устройства широко представлены в различных модификациях. Хорошо зарекомендовали себя у потребителей СОМ-системы производства фирм Microbox, Zeutschel, SMA (Германия), Kodak (США), Fuji (Япония) и др.

Российская наука не стоит в стороне от указанных проблем. Так, в нашей стране Федеральным государственным унитарным предприятием «Научно-исследовательский институт репрографии» на протяжении последних лет в интересах национальной безопасности государства теоретически обосновываются, нормативно-методиче-

ски закрепляются и практически внедряются современные гибридные электронно-микрографические технологии создания, сохранения и использования единого российского страхового фонда документации, позволяющие интегрировать традиционные (микрографические) и современные (электронные) способы создания страховых фондов документации различного назначения. Данные технологии позволяют долгосрочно сохранять на микрофильме определенные виды цифровой информации, в частности текстовую, фотографическую и чертежно-графическую документацию, созданную как путем оцифровки бумажных оригиналов, так и непосредственно в ЭВМ [4, 5]. Исследования, проводимые в данной области, опираются на твердую государственную поддержку, высокую научную квалификацию сотрудников института, передовой зарубежный опыт и парк современного электронно-микрографического оборудования (СОМ-системы, сканеры микроформ), позволяющего проводить различные эксперименты, отрабатывать технологические схемы и моделировать цепочки взаимодействия новых устройств в условиях функционирования системы СФД. При этом сотрудниками института осуществляется регулярный мониторинг зарубежной информации по проблеме исследований, осуществляются ее сбор, накопление и анализ.

По результатам анализа последних зарубежных публикаций по теме сохранности цифровой информации на микрофильмах можно констатировать, что зарубежные исследования проблемы динамично развиваются и находятся в стадии поиска оптимальных решений и подходов. Из работ, появившихся на Западе в последние два года, наибольший интерес представляют следующие: «Исследование цветного микрофильма как носителя для долгосрочного хранения цифровых данных» [1], «Модель коммуникационного канала микрофильма» [2] и «Долгосрочное хранение цифровых данных на микрофильме» [3].

Общей идеей данных публикаций является то, что при современном уровне развития микрографической техники существует возможность сохранения цифровой информации на микрофильме на битовом уровне (в виде потоков двоичных данных) – т.е. в том виде, в котором цифровая информация существует непосредственно, с последующим ее считыванием сканирующими устройствами. Любой цифровой документ состоит из набора двоичных данных – битовой информации. Эта битовая информация может быть особым образом закодирована

в виде двухмерного штрих-кода, состоящего из информационных точек и представлен в виде двухмерного растрового изображения, которое при помощи СОМ-системы сохраняется на микрофильме. При необходимости восстановления информации штрих-кодовые данные считываются с микрофильма сканирующим устройством, а затем декодируются, в результате чего происходит восстановление оригинального электронного документа.

Авторы первых двух статей предлагают использовать для этих целей следующий подход. Хранение должно осуществляться гибридным способом, т.е. на микрофильм записываются как само аналоговое изображение, так и его цифровой код. По своей природе микрофильм позволяет считывать информацию и человеку, и машине, поэтому он может использоваться как гибридный носитель, сочетая аналоговую и цифровую информацию.

В качестве конкретного носителя авторами предлагается цветной микрофильм производства Ipfochrome Micrographic. Для хранения данных на цветной пленке есть свои основания, главное из которых заключается в том, что при хранении можно использовать все три цветовых слоя, благодаря чему увеличится объем сохраняемых

данных. Двухмерный штрих-код, в который преобразовываются оригинальные документы – это растровое изображение, в котором каждая растровая точка представляет собой состояние. Одна растровая точка служит бинарным описанием состояния (максимальная или минимальная оптическая плотность) или описанием состояния более высокого порядка (несколько уровней плотности). По данным экспериментальных исследований, в которых для записи цветного микрофильма использовалась цветная лазерная СОМ-система нового поколения Archive Laser Recorder, была достигнута достаточно высокая плотность записи информации. Так, при размере точек 15 мкм на шестисотметровом рулоне цветной пленки 35 мм можно сохранить 22 гигабайта данных. При размере точки 12 мкм – 38 гигабайт. При 9 мкм – примерно 70 гигабайт на одном рулоне. Кажется, что такой объем не составляет конкуренции таким носителям, как например, жесткий диск. Но не стоит забывать, что при хранении цифровой информации вместимость не всегда является определяющим фактором, особенно по сравнению с долговечностью и стабильностью. Схема записи изображения на цветной микрофильм приведена на рис. 1.

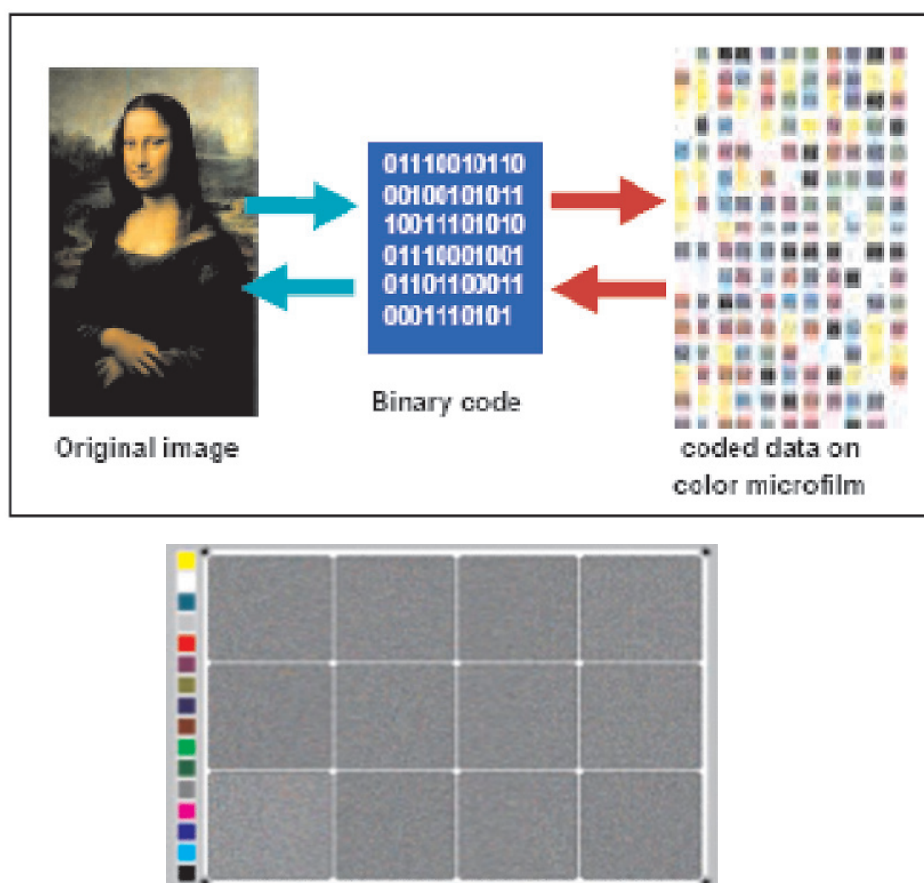


Рис. 1. Запись изображения на цветной микрофильм в бинарном штрих-кодовом виде

По мнению авторов, оптимизация параметров экспозиции и настроек считывания, подбор экспонирующего оборудования и типа пленки позволит достигнуть в будущем хороших результатов. Кроме того, в настоящее время разрабатываются оптимизированная и более совершенная система обработки сигналов и коды коррекции ошибок для хранения цифровых данных на цветном микрофильме. В конечном итоге, делается вывод, что необходимы масштабные практические испытания новой технологии.

Однако данному способу присущи определенные недостатки. В частности, итоговые характеристики цветных слоев нельзя рассматривать как независимые. Оптические свойства каждого из слоев могут различаться. Причиной данного явления является спектральное наложение применяемых красителей, ведущее к взаимодействию, подобному так называемым «перекрестным помехам» в системах коммуникации, что приводит к увеличению количества ошибок при обратном считывании информации с микрофильма. Применение цвета в такой системе добавляет сложности, так как со временем пленке свойственно менять цвет.

Кроме того, такая технология является достаточно затратной, так как для записи требуется цветная пленка, цветной лазер (цветные СОМ-устройства) и химико-фотографическая обработка цветной пленки. Все это чрезвычайно дорого. Сканирующее оборудование, необходимое для считывания цветной пленки, также является более сложным и дорогим, чем аналогичное оборудование для черно-белых материалов. Соответственно, если цвет решающего значения не имеет, рациональнее использовать черно-белый микрофильм.

Именно такой, более простой и экономичный подход представлен в статье С. Шильке (Франкфуртский университет, Германия) и А. Раубера (Венский технологический университет, Австрия) [3]. Акцент сделан на использовании стандартных технологий и широко распространенного оборудования для записи и считывания микрофильмов, а не на специально сконструированные исследовательские модели.

В качестве носителя используется обычный черно-белый микрофильм, а исходные электронные документы (их бинарные данные) кодируются с помощью двумерного черно-белого графического штрих-кода. Затем эти данные трансформируются в изображение и сохраняются (экспонируются) на микропленку. При воспроизведении бинарных данных микрофильм сканируется, а изображение декодируется с помощью расшифровки отсканированного штрих-

кода. В результате снова получается поток бинарных данных, из которых восстанавливается исходный электронный документ.

При экспериментальных исследованиях в данном случае применялось записывающее СОМ-устройство Kodak Digital Archive Writer (DAW) i9610 и сканер для микрофильмов Kodak 3000 DSV (также называемый Minolta ES-7000). Также использовался черно-белый микрофильм 16 мм длиной 30,5 м, который обрабатывался с помощью проявочной машины Kodak Prostar.

Для создания графических штрих-кодов использовалась специальная программа, позволяющая преобразовывать любой цифровой документ в черно-белое двумерное растровое штрих-кодовое изображение, пригодное для записи на микрофильм через СОМ-систему. Данная программа сохраняет с высоким разрешением исходный документ в виде одного или нескольких (если исходный файл большого объема) черно-белых штрих-кодовых изображений форматов А4, А3 или А2 по выбору. Штрих-кодовое растровое изображение (рис. 2) записывается на микрофильм, хранится на нем, а затем при необходимости считывается с помощью обычного сканера микрофильмов.

Далее считанное штрих-кодовое изображение с использованием той же программы декодируется и приобретает исходный вид. Необходимо отметить, что для повышения надежности кодирования/декодирования указанная программа снабжена встроенным механизмом коррекции ошибок Рида-Соломона, аналогичным тому, который применяется для коррекции ошибок записи/чтения в системе «считывающий привод/оптический диск».

В результате эксперимента было установлено, что черно-белые штрих-коды позволяют добиться относительно высокой плотности записи информации. Этот способ в сочетании с эффективной системой коррекции ошибок декодирования на выходе позволяет также более точно воспроизводить данные.

Выяснилось, что на одном 16-мм микрофильме длиной 30,5 м в штрих-кодах можно сохранить 7200 изображений формата А4 или 45,32 Мб информации (на 35 мм микрофильме соответственно в 2 раза больше). По расчетам авторов, в данном случае стоимость хранения 1 гигабайта составит 0,22 евро. В долгосрочной перспективе хранения эта цена представляется наиболее оптимальной по сравнению с другими системами, особенно сравнивая ее со стоимостью миграции каждые 5–7 лет, необходимой для других форматов, и стоимостью их техни-



ческой поддержки. Так, например, хранение на современных жестких дисках обходится 0,1–0,3 доллара за 1 гигабайт, но эти технологии требуют значительных затрат в процессе, так как большое количество дисков должно постоянно функционировать, чтобы поддерживать систему в рабочем состоянии. Это требует значительных затрат на электроэнергию, инфраструктуру и техобслуживание на протяжении относительно короткого срока службы. К тому же в отличие от других носителей, таких как жесткие диски, флеш-карты, CD или DVD диски, технологии считывания микрофильма очень просты и универсальны. Тогда как для воспроизведения данных с популярных электронных носителей необходимы спе-

циализированные интерфейсы и сложные технологии (оптические диски с лазерной технологией, высокоточное расположение считывающих устройств для магнитных носителей, контролирующие программы и оборудование и т.д.), для считывания данных с микрофильма необходимы только простые оптические устройства. Это выгодно отличает данный носитель от IT-систем. Если найти в будущем устаревший привод для DVD или лент или USB-порт, совместимый с новыми компьютерными системами, будет очень сложно, то для микрофильма будет достаточно любого современного оптического устройства для формирования изображения – будь то сканер, камера или другой аппарат.

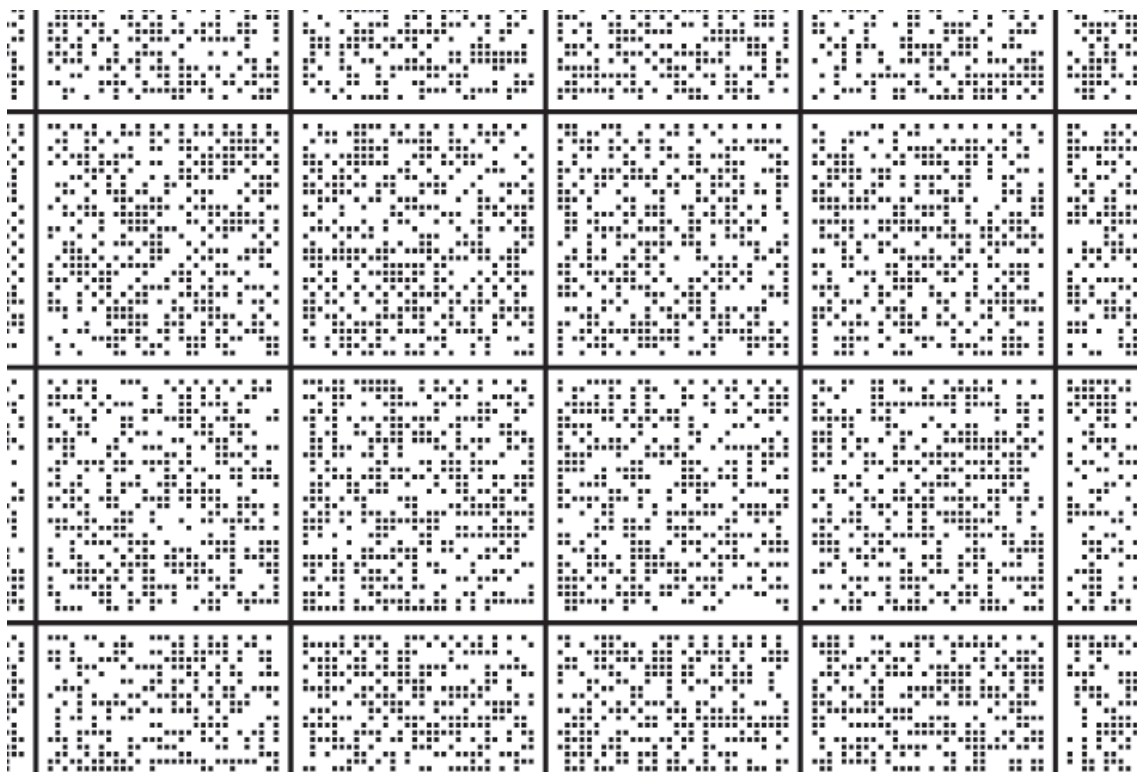


Рис. 2. Увеличенный фрагмент черно-белого штрих-кодowego изображения, полученного в результате работы программы кодирования и пригодного для записи на микрофильм

Подводя итог, можно сказать, что рассмотренные выше технологии сохранения цифровых данных на микрофильме открывают новую эру в истории микрографии. Их значение заключается в том, что впервые появилась теоретически обоснованная и технологически реализуемая возможность долгосрочно сохранять на микрофильме любую цифровую информацию и документацию. При этом тип электронного документа не имеет значения, так как все цифровые файлы состоят из набора двоичных данных и соответственно могут быть

представлены в виде двухмерных графических штрих-кодов. Помимо уже осуществляемого сохранения цифровой цветной и черно-белой чертежно-графической, текстовой и фотографической документации, применение данного метода открывает казавшиеся ранее невозможными перспективы сохранения на микрофильмах цифровой аудиовизуальной документации, программных продуктов, трехмерной документации CAD-приложений и др., т.е. любого типа цифровых данных. Таким образом, данный способ сохранения позволяет значительно

приблизиться к решению проблемы долгосрочного страхового сохранения электронных документов.

Разумеется, рассмотренные способы сохранения цифровых данных еще далеки от совершенства. Требуются длительные исследования и эксперименты по подбору параметров записи, отработке режимов, синхронизации оборудования, оптимизации настроек элементов системы, технико-экономические расчеты и т.д. Однако первые шаги уже сделаны и дальнейшие исследования возможности применения данного перспективного метода обязательно будут продолжены как за рубежом, так и в нашей стране.

#### Список литературы

1. Christoph Voges, Volker Margner, Tim Fingscheidt. Investigations on Color Microfilm as a Medium for Long-Term Storage of Digital Data. URL: [http://www.imaging.org/ist/publications/reporter/articles/REP25\\_3\\_ARCH2010\\_VOGES.pdf](http://www.imaging.org/ist/publications/reporter/articles/REP25_3_ARCH2010_VOGES.pdf) (дата обращения 18.08.2011).

2. Towards a Channel Model for Microfilm / A. Amir, F. Muller, P. Fornaro, R. Geschwind, J. Rosenthal, L. Rosenthaler. – URL: <http://www.peviar.ch> (дата обращения 18.08.2011).

3. Schilke S.W., Rauber A. Long-term archiving of digital data on microfilm // Int. J. Electronic Governance. – 2010. – Vol. 3, № 3. – P. 237–253.

4. Гаврилин А.П. Применение электронно-микрографических технологий для сохранения и оперативного доступа к документированной информации // Успехи современного естествознания. – 2008 – № 11. – С. 62–64.

5. Гаврилин А.П. Применение электронно-микрографических технологий для обработки и сохранения документации // Фундаментальные исследования. – 2006 – № 6. – С. 51–53.

#### Рецензенты:

Котов В.В., д.т.н., доцент, профессор кафедры «Робототехника и автоматизация производства» ГОУ ВПО «Тульский государственный университет», г. Тула;

Ларкин Е.В., д.т.н., профессор, зав. кафедрой «Робототехника и автоматизация производства» ГОУ ВПО «Тульский государственный университет», г. Тула.

Работа поступила в редакцию 11.11.2011.