

А. А. Крючин,

зам. директора Института проблем регистрации информации НАН Украины,
чл.-кор. НАН Украины

И. В. Горбов,

старший научный сотрудник ИПРИ НАН Украины,
кандидат технических наук

Н. В. Солонина,

младший научный сотрудник ИПРИ НАН Украины

ТЕХНОЛОГИИ ДОЛГОСРОЧНОГО ХРАНЕНИЯ ДАННЫХ В ЭЛЕКТРОННЫХ БИБЛИОТЕКАХ И АРХИВАХ

Рассмотрены проблемы хранения и рационального использования информационных ресурсов электронных библиотек. Проведен анализ объемов информации в электронных библиотеках и определены технологии долгосрочного хранения этих ресурсов.

Ключевые слова: электронные библиотеки, долгосрочное хранение, электронные информационные ресурсы, оптические носители информации.

The problems of storage and rational use of informative resources of electronic libraries are considered. The analysis of information volumes in electronic libraries is conducted and technologies of long-term storage of these resources are determined.

Keywords: electronic libraries, long-term storage, electronic informative resources, optical of data carriers.

Введение

Стремительное развитие электронных цифровых технологий во многом изменило направление деятельности библиотек и радикально поменяло взгляд на библиотеку в современном мире. Все чаще в обществе звучат понятия: электронные информационные ресурсы, цифровые документы, электронные библиотеки. Хранение, развитие и рациональное использование национальных информационных ресурсов является задачей государственного значения. В настоящее время электронная форма позволяет хранить информацию надежно и компактно, распространять ее оперативно и широко.

Насущность создания в библиотеке электронных коллекций обуславливается несколькими факторами, среди которых: стремительное развитие информационно-телекоммуникационных технологий, увеличение

количества документов, существующих только в электронной форме, увеличение потребности граждан в удаленном доступе к информации, все увеличивающаяся привычка, прежде всего научных работников и учащихся, работать только с электронными вариантами документов и, наконец, стремление самих библиотек предоставить своим пользователям информационные услуги повышенного качества. Электронные массивы информации сегодня становятся одним из важнейших компонентов библиотечно-информационной системы [1].

В настоящее время стремительно увеличивается как число электронных библиотек, так и объемы представленной в них информации. Рост объемов информации происходит за счет новых поступлений, а также за счет последовательной работы по оцифровке ранее созданных фондов. Планируется перевести к 2015 году 50 % всех библиотечных фондов России в цифровую форму [2].

Предложено и реализовано несколько структур электронных библиотек, которые успешно функционируют [3]. При любой структуре электронной библиотеки одной из основных проблем остается хранение информации. С ней библиотеки сталкиваются если не в начале пути создания коллекции, то через некоторое время – неизбежно. Необходимо обеспечить не только хранение электронных документов, но и возможность их длительного использования в условиях стремительно развивающихся технологий.

Анализ требований к системам долгосрочного хранения данных

Для определения требований к системам и носителям долгосрочного хранения необходимо, в первую очередь, оценить объем информации, представленной в электронных библиотеках и архивах.

В таблице приведены в качестве примера объемы данных, представленных в некоторых электронных библиотеках и архивах различного уровня.

Следует отметить, что объем данных, представленных в электронном виде, постоянно увеличивается. Например, в Библиотеке Конгресса США ежедневно сканируется от 75 до 200 документов, фонд электронных документов Французской национальной библиотеки ежегодно вырастает на 150 тысяч документов.

При выборе форматов электронных документов необходимо ориентироваться на стандартные форматы, поддерживаемые средствами тех операционных систем, которые сегодня используются в информационных учреждениях и развитие которых можно спрогнозировать в дальнейшем. Так, можно прогнозировать долгий жизненный цикл форматов XML, HTML, а также форматов pdf, tiff – для

Таблица

Объемы данных в электронных библиотеках и архивах

Библиотека	Электронный ресурс	Объем фондов	Примечание
Электронные библиотеки			
Библиотека Конгресса США	10 Тбайт	134x10 ⁶ ед. хранения	[4]
Национальная библиотека Дании	236 Тбайт		[25]
Российская государственная библиотека	162 Тбайт	43,1x10 ⁶ ед. хранения	[6]
Национальная библиотека Беларуси "Электронный архив национальной периодики" (ЭАНП)	300 Гбайт	38,813x10 ³ ед. хранения	[7]
Национальная библиотека Украины им. В.И. Вернадского	прибл.2 Тбайт	15x10 ⁶ ед. хранения	[8]
Электронные архивы			
Архив австралийского сегмента Интернет (PANDORA)	прибл.7 Тбайт	200x10 ⁶ ед. хранения	[8]
Архив Белоруссии	735 Гбайт	735x10 ³ ед. хранения	[5]
Архив ГНПП "Геоинформ Украина"	прибл.160 Гбайт	184x10 ³ ед. хранения	[9]

текстовой информации, avi, mpeg – для мультимедиа, jpeg – для графической. Грамотный подход к выбору форматов хранения информации позволит максимально избежать в будущем проблем поддержки, использования и отображения данных вновь появляющимися программными средствами [1].

Одним из основных технических решений при организации хранения данных является использование иерархической структуры запоминающих устройств.

Ведущие разработчики информационных технологий предлагают для организации хранения данных в электронных библиотеках также технологии:

- ✧ использование многоуровневых сред хранения данных (с дисковыми и ленточными накопителями);
- ✧ технологии переноса и обеспечения мобильности данных;
- ✧ технологии сжатия и дедупликации данных;
- ✧ управление многоуровневыми хранилищами на основе политик с автоматическим переносом и удалением по истечении срока хранения;
- ✧ шифрование данных на дисковых и ленточных накопителях для их защиты;
- ✧ системы хранения без возможности удаления и перезаписи данных и системы WORM (с однократной записью и многократным считыванием);
- ✧ инструменты и решения для управления данными и содержимым, обеспечивающие идентификацию, сбор и классификацию содержимого с автоматическим управлением [10].

Как правило, в электронной библиотеке организуется иерархическая структура запоминающих устройств. При этом данные хранятся на устройствах разных уровней в зависимости от степени важности этих данных и требуемого времени доступа к ним.

На первом (высшем) уровне используются быстродействующие основные запоминающие устройства, включая средства сетей хранения данных (*Storage Area Network, SAN*). Они лучше всего подходят для хранения файлов, к которым пользователи обращаются регулярно. Следующий уровень образуют вторичные накопители. Они могут представлять собой непосредственно подключаемые к серверам запоминающие устройства и системы хранения данных специального назначения. Далее (на третьем уровне) находятся относительно быстродействующие системы резервного копирования, например, записывающие информацию на жесткие диски (*disk-to-disk backup*), а на самом последнем уровне — ленточные накопители и оптические диски [3].

Объем данных большинства электронных библиотек позволяет организовывать хранение всей информации на жестких магнитных дисках. Достигнутый за последние тридцать лет практически экспоненциальный рост плотности записи на жестких магнитных дисках позволяет создавать устройства хранения данных емкостью в десятки терабайт. Устройства на

жестких магнитных дисках позволяют создавать системы хранения данных с высокими объемами записанных данных и скоростями обмена данных (сотни мегабайт/с), малыми временами обращения (десятки миллисекунд). Однако, жёсткие магнитные диски, являющиеся системными и высокоточными электромеханическими устройствами, имеют невысокую надежность хранения информации.

Проведенные компанией Google исследования показали, что к концу четвертого года эксплуатации процент отказа жестких магнитных дисков составляет 24,3 % [11]. В связи с этим актуальной является задача создания носителей для долгосрочного хранения данных, которые по емкости могут уступать магнитным дискам, но существенно превосходить их по срокам хранения данных.

Носители информации для долгосрочного хранения данных

Проблему долгосрочного хранения цифровых документов нужно начинать решать на этапе их создания. Более 65 % стоимости технологического процесса жизненного цикла цифровых документов от создания до сохранности составляет именно обеспечение хранения, и со временем эта доля будет увеличиваться. Нельзя решать проблему хранения отдельно, она должна быть увязана в единый технологический цикл создания – обработки – хранения – использования цифровых документов [12].

Практически во всех иерархических структурах хранение данных предполагает использование оптических носителей для архивного хранения данных. Разработаны архивные хранилища данных на оптических дисках емкостью более 34 Тбайт со сроком хранения более 50 лет [13]. Чаще всего при выборе типов оптических носителей ограничиваются самыми общими рекомендациями типа:

- ✧ необходимо использовать носители известных производителей;
- ✧ применять носители разных типов.

Проводятся многочисленные исследования по созданию специальных носителей с однократной записью для архивного хранения. В носителях с однократной записью предлагается использовать специальные красители с более высокими температурами фазовых переходов и фоточувствительными материалами с необратимыми фазовыми переходами и золотосодержащие, отражающие слои [14].

В компакт-дисках с однократной записью срок службы носителя определяется в основном изменениями характеристик регистрирующей среды (в качестве которой в большинстве случаев используются орга-

нические красители). Регистрирующая среда наиболее чувствительна к повышенной влажности.

По разным оценкам долговечность компакт-дисков может составлять от 30 до 50 лет. Оценки базируются на ускоренных тестах их старения при повышенной температуре и влажности.

Ускоренные испытания проводились с использованием модели Эйринга (Eyring acceleration model), в которой срок службы носителя t определяется на основе законов термодинамики следующим образом [15]:

$$t = AT^d \cdot \exp\left(\frac{E_s}{kT}\right) \cdot \exp\left\{R\left(B + \frac{C}{T}\right)\right\}, \quad (1)$$

где R – относительная влажность; E_s – энергия активации; k – постоянная Больцмана; T – температура, К; A , B , C , d – константы.

При проведении исследований в ограниченном температурном диапазоне уравнение упрощенной модели Эйринга может быть представлено в виде:

$$\ln(t) = \ln(A) + \left(\frac{E_s}{k}\right) \times T^{-1} + B \times R \quad (2)$$

Константы A , B , E_s рассчитываются с использованием линейного регрессионного анализа экспериментальных данных [15].

Приведенные сроки хранения информации на оптических дисках являются достаточно оптимистическими, и для гарантированного хранения данных рекомендуется каждые 3–5 лет производить перезапись данных на аналогичные новые носители [16]. Процесс перезаписи при больших объемах данных является достаточно трудоемким, а главное – при его осуществлении может происходить потеря данных. В связи с этим значительные усилия технологов направлены на разработку специальных оптических носителей для долгосрочного хранения данных (профессиональных оптических носителей).

Первым образцом профессиональных оптических дисков для архивного хранения являются диски типа UDO (ultra density optical – сверхплотный оптический). Такие диски разработала и выпускает компания Plasmon (www.plasmon.com). В этих дисках реализован режим записи WORM (Write-Once/Read-Many – однократная запись/многократное воспроизведение). Повышение гарантированного срока хранения информации достигнуто за счет применения фоточувствительных материалов с необратимым фазовым переходом и специального материала подложки.

UDO-носители могут быть использованы в двух режимах: при реверсивной записи или записи WORM. Процесс записи в режиме WORM

происходит за счет изменения структуры материала регистрирующей среды на молекулярном уровне, он гарантирует невозможность изменения записанной информации на носителях UDO.

Емкость 5,25-дюймового UDO-носителя составляет 30 Гбайт, а гарантированный срок хранения данных превышает 50 лет. Наивысшая для современных оптических носителей плотность записи достигается за счет использования лазеров с длиной волны 405 нм. В 2006 году завершилась разработка UDO-носителя емкостью 60 Гбайт, а в 2008 году – 120 Гбайт. Предусмотрено, что устройства записи/воспроизведения информации для этих носителей будут удовлетворять требованиям обратной совместимости.

Запись и воспроизведение данных с UDO-носителей требует применения специальных устройств, отличающихся от системы записи/воспроизведения данных с компакт-дисков.

Другое, на наш взгляд, более перспективное направление исследований, состоит в усовершенствовании технологии изготовления компакт-дисков, которая позволяла бы существенно увеличить гарантированный срок хранения информации, с одной стороны, а с другой – сохранить возможность чтения их стандартными устройствами воспроизведения данных компакт-дисков.

Для повышения сроков хранения данных на носителях с однократной записью предлагается использовать специальные высокостабильные материалы для регистрирующего слоя, на котором осуществляется перфорационная запись. Материал регистрирующего слоя имеет достаточно высокий коэффициент отражения, что исключает необходимость использования отражающего металлического слоя. Слабая адгезия отражающего металлического слоя к слою органического красителя является одной из основных причин выхода из строя оптических носителей с однократной записью. Отслоение металлического слоя может происходить из-за хранения в условиях повышенной влажности или незначительных механических повреждений отражающего слоя. Структура описанного диска приведена на рис. 1.

Повышение гарантированного срока хранения данных достигается за счет того, что запись данных осуществляется в виде микрорельефной структуры в слое высокостабильного материала. Температура плавления этого материала в 2–3 раза превышает температуру плавления органических красителей, применяемых в качестве регистрирующей среды в компакт-дисках с однократной записью. Фактически предлагается структура носителя информации первых оптических дисков WORM, в которых широко применялась перфорационная запись на тонких

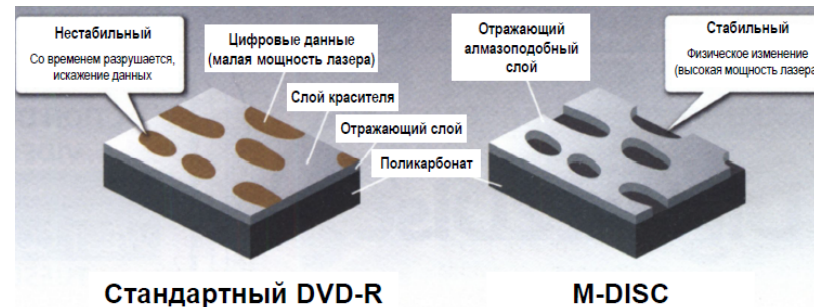


Рис. 1. Структура М-диска и представление информации на таком носителе [17]

стеклообразных халькогенидных полупроводниках [18]. Проведенные нами исследования образцов оптических стеклянных дисков с регистрирующими средами такого типа показали, что после хранения в течение 25–30 лет в неоттапливаемых помещениях форма и размеры пиков сохранились (рис. 2) [19].

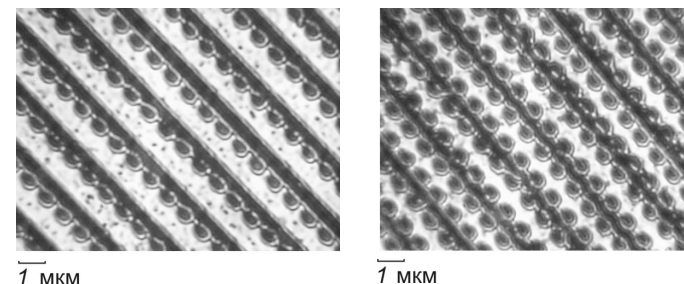


Рис. 2. Внешний вид зоны записи на оптическом носителе с разной плотностью записи

Для записи данных на М-диски необходимо использовать специальные устройства записи, в которых применяются более мощные лазеры [17].

Однако, в таких дисках не решена проблема, связанная с нестабильностью характеристик подложек из поликарбоната. Для существенного повышения сроков службы оптических носителей предлагается использовать носители, в которых применяются подложки из высокостабильных материалов. Одним из вариантов изготовления таких

носителей является применение металлических никелевых подложек. Фактически для долгосрочного хранения информации предлагается использовать никелевые штампы, применяемые при тиражировании компакт-дисков [17, 20, 21]. Копия матрицы, используемой для их издания, передается для архивного хранения [20]. Такая технология долгосрочного хранения данных успешно применяется в ряде стран для хранения мультимедийной информации.

Одной из основных проблем при хранении информации на металлических носителях является защита его поверхности от загрязнения пылевыми частицами. Нами было предложено и реализовано несколько вариантов конструкций металлических носителей для долгосрочного хранения данных, в которых использовались различные методы защиты информационной поверхности от загрязнения пылевыми частицами. Наиболее эффективным оказался метод считывания через дополнительный оптически прозрачный слой толщиной (0,6–1,2) мм [20, 21].

Анализ комплекса характеристик оптических носителей для долгосрочного хранения показал, что наиболее перспективным является создание носителей, в которых данные записываются в форматах стандартных компакт-дисков, а подложка носителя изготавливается из высокостабильного материала. Для дисков, записываемых в форматах CD и DVD, подложка должна быть оптически прозрачной [23]. Такие диски будут считываться на стандартных устройствах воспроизведения компакт-дисков, что избавляет от необходимости использования дорогостоящего специализированного оборудования для воспроизведения.

Метод получения таких носителей основывается на технологии производства дисков-оригиналов, применяемых при изготовлении штампов для тиражирования компакт-дисков. Запись информации осуществляется на станции лазерной записи на подложки, покрытые слоем позитивного фоторезиста. После селективного травления и термической обработки слоя фоторезиста осуществляется ионное или плазмохимическое травление подложки изготавливаемого носителя информации. Для обеспечения значений сигналов считывания, соответствующих уровням сигналов со стандартных компакт-дисков, после удаления остатков фоторезиста ионным пучком производится металлизация подложки носителя информации. Для обеспечения длительных сроков хранения записанной информации необходимо применять металлы, пленки которых имеют высокую адгезию к подложке и являются химически инертными (рекомендуется использовать пленки родия или хрома). Последней технологической операцией является нанесение защитного слоя. Для дисков,

записанных в формате BD, толщина защитного слоя должна соответствовать 0,1 мм и строго контролироваться [23].

Предполагается, что такие носители изготавливаются на специализированных предприятиях по заказу архивных учреждений и библиотек. Предложение изготавливать носители для долгосрочного хранения на специализированных предприятиях расширяет число возможных технологий изготовления носителей. В частности, предлагается изготавливать носители с голографической записью со сроком хранения более 50 лет [24].

При создании систем долгосрочного хранения данных возникает вопрос, могут ли быть прочитаны данные при выходе из строя устройств воспроизведения, которые предназначены для работы с конкретными типами носителей [5].

В связи с тем, что проблема долгосрочного хранения данных в цифровом виде еще не решена, организуется хранение данных на микрофильмах и предлагается осуществить запись данных на специальных кварцевых и сапфировых носителях в аналоговой форме [26,27]. Предполагается, что такая форма представления информации позволит надежнее воспроизводить данные при изменении операционных систем или полной смене устройств считывания данных.

Выводы

1. Объемы данных электронных библиотек быстро увеличиваются и для их надежного и долгосрочного хранения необходима разработка специальных носителей.
2. Для электронных библиотек и архивов разрабатываются и изготавливаются специальные профессиональные оптические носители, характеризующиеся повышенной надежностью и сроками хранения данных.
3. Для изготовления профессиональных оптических дисков, предназначенных для длительного хранения информации, предложено использовать подложки, изготовленные из высокостабильных материалов.
4. Запись информации в виде микрорельефных структур на носители с подложками из высокостабильных материалов предложено производить на станции лазерной записи дисков-оригиналов с последующим плазмохимическим травлением подложек.

Благодарность

Авторы выражают глубокую благодарность заведующему отделом Института проблем регистрации информации НАН Украины, доктору технических наук Д. В. Ландэ за обсуждение материалов статьи, а также сделанные замечания и рекомендации.

Список использованных источников

1. *Моисеева С. А.* Электронные библиотеки: Перспективы и проблемы создания. [Электронный ресурс] / С. А. Моисеева. – Режим доступа: http://www.library.lg.ua/rus/kollegam_public.php. – Загл. с экрана.
2. «Стратегия развития информационного общества в Российской Федерации» от 7 февраля 2008 г. № Пр-212 // РГ. – 2008. – 16 февр. – 16 полоса.
3. *Шухарт С.* Как организовать долговременное хранение данных / С. Шухарт // Сети и Системы связи. – 2003. – № 12. – С. 3–5.
4. *Когаловский М. Р.* Тенденции развития технологий управления информационными ресурсами в электронных библиотеках. [Электронный ресурс] / М. Р. Когаловский // Труды 8^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2006, Суздаль, 2006. – Режим доступа: <http://rcdl.ru/proceedings.php?year=2006>.
5. *Носевич В.* Как сберечь цифровое наследие (2011) / В. Носевич // Архивы и справочества. – 2011. – № 6 (78). – С. 82–92.
6. В России создана крупнейшая в Европе электронная библиотека. [Электронный ресурс]. – Режим доступа: <http://www.grani.ru/news/Internet/m.145533.html>. – Загл. с экрана.
7. *Сакович О. В.* Электронный архив национальной периодики национальной библиотеки Беларуси: современный этап развития. [Электронный ресурс] / О. В. Сакович // Национальная библиотека Беларуси, Минск, Беларусь труды Девятнадцатая Международная Конференция «Крым 2012». [Электронный ресурс]. – Режим доступа: <http://www.gpntb.ru/win/inter-events/crimea2012/disk/113.pdf>.
8. *Копанева В.* Библиотека як центр збереження інформаційних ресурсів Інтернету : монографія / В. Копанева ; ред. О. С. Онищенко ; Нац. б-ка України ім. В. І. Вернадського. – К. : НБУВ, 2009. – 200 с.
9. Фонд геологічних матеріалів [Электронный ресурс] – Режим доступа: http://geoinf.kiev.ua/techn3_7.htm. – Назва з екрана.
10. Сохранение данных в течение нужного времени на самых экономичных носителях для хранения данных [Электронный ресурс]. – Режим доступа: http://www-03.ibm.com/systems/ru/infrastructure/information_retention/. – Загл. с экрана.
11. *Корепанов И.* Как сохранить архив на десятилетия? / И. Корепанов // LAN : журн. сетевых решений. – 2008. – № 3. – С.91–95.
12. *Шварцман М. Е.* К вопросу о долгосрочном сохранении цифровой информации.[Электронный ресурс] / М. Е. Шварцман. – Режим доступа: http://www.benran.ru/Magazin/cgi-bin/Sb_07/pr_07.exe?15!. – Загл. с экрана.
13. Способствовать информационному и технологическому прогрессу Общества и Государства [Электронный ресурс]. – Режим доступа: <http://www.elar.ru/products/>. – Загл. с экрана.
14. Analysis of the plastic substrates, the reflective layers, and the adhesives of today's archival – grade DVDs / Guilin Jiang, Felipe Rivera, Supriya Kanyal, Robert C. Davis, Richard Vanfleet, Barry M. Lunt, Matthew R. Linford // Optical Data Storage – 2010. – 7730. – P. 7730N-1–7730N-8.
15. *Irie M.* Standardized Life Expectancy of High-Speed Recordable Optical Disks / M.Irie, Y.Okino // IEEE Transactions on magnetics. –2007. – Vol. 41, № 2. – P. 864–866.
16. *Устинов В.* Хранение данных на CD- и DVD-дисках: на наш век хватит? / В. Устинов // Broadcasting. Телевидение и радиовещание. – 2006. – № 4. – С. 54–59.
17. *Toppin E.* Setting a new standard in permanent archival storage / E. Toppin // Digital 2Disc. – 2010. – P. 42–44.
18. *Петров В. В.* Надцільний оптичний запис інформації / В. В. Петров [та ін.] ; відп. ред. О. Г. Додонов ; Ін-т проблем реєстрації інформації НАН України. – К. : НАН України, 2009. –282 с. –I SBN 978-966-02-5027-7.
19. *Petrov V. V.* Analysis of properties of optical carriers after long-term storage / V. V. Petrov, A. A. Kryuchin, I. V. Gorbov, I. O. Kossko, S. O. Kostyukevych // Semiconductor Physics. Quantum Electronics & Optoelectronics. – 2009. – Vol.12, № 4. – P. 399–402.
20. Металеві носії для довготермінового зберігання інформації / В. В. Петров, А. А. Крючин, С. М. Шанойло, Л. І. Крючина, І. О. Косско. – К. : ІПРІ НАН України, 2005. – 132 с.
21. *Ефимов Д. А.* Скрижали 21 века – применение металлических носителей для сверхдолгого хранения оцифрованных библиотек [Электронный ресурс] / Д. А. Ефимов // Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса: материалы конф. – Электрон. дан. – М.: ГПНТБ России, 2010. – 1 электрон. опт. диск (CD-ROM). – Систем. требования: IBM PC, Windows 2000 или выше. – Загл. с этикетки диска. – ISBN 978-5-85638-139-8. – № гос. регистрации 0321000673.
22. Пат. 74201 Україна, МПК 7 G 11 В 7/24. Носії для довго-термінового зберігання інформації / Петров В. В., Крючин А. А., Шанойло С. М., Косско І. О. – Опубл. 15.11.2005, бюл. № 11.
23. Оптические диски для долговременного хранения информации / В. В. Петров, В. М. Пузиков, А. А. Крючин, И. В. Горбов // Наносистемы, наноматериалы, нанотехнологии. – 2009. – 7, № 3. – С. 825–832.
24. Компания hVault. Улучшенная программа для архивирования (hVault Advanced Archive Program) // Creative Storage Conference 2012, конф., 03.07.2012, Broadcast [Электронный ресурс]. – Режим доступа: http://www.broadcast.telekritika.ua/show/Hranenie/2158-hvault_utverzhdaet_golograficheskaja_tehnologija_idealno_podhodit_dlia_dolgosrochnogo_hraneniia_dannyh. – Загл. с экрана.
25. *Пеллагеша Н.* Європейська цифрова бібліотека: проект створення / Н. Пеллагеша // Бібл. вісн. – 2008. – № 5. – С. 3–7.
26. Laser-etched quartz will store data for hundreds of millions of years – [Электронный ресурс]. – Mode of access: <http://www.wired.co.uk/news/archive/2012-09/25/hitachi-quartz-data-storage>. – Title from the screen.
27. A Million-Year Hard Disk [Электронный ресурс]. – Режим доступа: <http://news.sciencemag.org/sciencenow/2012/07/a-million-year-hard-disk.html>. – Title from the screen.