

УДК 004.4

*С. Н. Горбач¹, Л. А. Макаревич¹, К. Ю. Талецкий¹, В. А. Петрухин¹,
В. И. Шоломова²*

¹Московский физико-технический институт (государственный университет)

²Московский государственный университет имени М.В. Ломоносова

Организация обмена электронными документами для повышения качества образовательного процесса

В статье предлагается способ организации обмена электронными документами в высшем учебном заведении и рассматривается создание электронной системы свободного обмена информацией. В ходе работы были проанализированы возможные способы решения данной задачи, заданы критерии оценки результата, разработана архитектура системы и анализируются результаты внедрения.

Ключевые слова: система электронного документооборота, хранение данных, сетевая архитектура, пользовательский интерфейс, распределённая система общественной печати, веб-сервер, прокси-сервер, виртуальная частная сеть, PDF, PCL, PJJ, PostScript, Linux, IPP, SMB, CUPS.

Введение

За последние двадцать лет способы обмена информацией радикально изменились. Мало кто мог подумать, что на смену многотомным энциклопедиям придут сайты, а их авторами будут уже не сотни человек, а миллионы. Звучат мнения, что само существование печатных изданий под угрозой. Газеты и журналы со столетней историей борются за влияние и аудиторию с миллионами онлайн-дневников, чьи авторы излагают и оценивают любое событие по-своему, создавая доселе невообразимую широту взглядов. Телевидение начнёт проигрывать свои позиции сайтам, на которых ежедневно выкладываются сотни тысяч коротких роликов. Файлообменные сети дают нам возможность за считанные минуты передать в любую точку мира любимый фильм или музыкальное произведение.

Мы видим, что те проекты, которые основывались на принципах общедоступности и открытости, пользуются огромной популярностью. Примерами таких поразительно успешных социальных проектов могут служить Википедия или торренты. Опен-сорс сообщество – ярчайшая демонстрация того, как сложнейшие системы могут быть созданы людьми, объединившимися для решения общей проблемы, а не с целью извлечения быстрой выгоды. Совместное использование и развитие ресурсов приходит на смену персональным приложениям и небольшим коллективам авторов и разработчиков. Образовательный процесс с этой точки зрения не исключение. Помогая учащимся и преподавателям интенсифицировать обмен информацией, можно повысить скорость и качество обучения, а также удовлетворённость учеников самим процессом.

Информационные технологии в образовании, особенно связанные со свободным ПО, в современной России – тема о которой часто пишут в СМИ и рассуждают в министерствах. Но, к сожалению, до сих пор есть слишком мало примеров успешного внедрения и использования на высоком уровне принятия решений. Нередко неудачный опыт связан с низким качеством предлагаемых ИС, что, в частности, может объясняться недобросовестным отношением исполнителей госзаказов. К сожалению, несмотря на очевидную выгоду от автоматизации, информатизации и процессов обучения и внедрения открытых и свободных решений, иногда образовательные учреждения отказываются от внедрений.

Данная работа посвящена нашей попытке предоставить студентам простой, а главное, естественный способ обмена и совместного пользования документами.

Созданная нами система способна существовать как отдельное решение на коммерческой основе, так и быть внедрённой как часть инфраструктуры образовательного учреждения, органично вписываясь в его процессы и учитывая интересы руководства и студентов. Благодаря такому подходу система обладает большой гибкостью в распространении.

1. Свободный обмен информацией в образовании

1.1. Критерии реализации Свободного Хранилища Документов

Основная цель Свободного Хранилища Документов — предоставить студентам и, возможно, преподавателям простой и удобный способ обмена, хранения и структуризации любых документов. Это позволит значительно стимулировать образовательный процесс за счёт значительного ускорения обмена знаниями между студентами и консолидации самых актуальных учебных материалов, отбираемых самими студентами.

Перед построением любой системы необходимо учитывать опыт разработок, решающих схожие задачи, схожие по основным структурным решениям, направленные на схожую целевую аудиторию. Одним из ключевых моментов предварительного анализа является выработка критериев удобства использования («юзабилити») и определение наиболее важных свойств системы с точки зрения пользователя.

С наиболее общей точки зрения нашу Систему можно охарактеризовать как:

- Свободное хранилище документов, ориентированное на активное использование в учебном процессе для хранения и обмена информацией между его участниками и формируемое самими участниками.

По общей концепции такая система близка к современным социальным ресурсам вроде Википедии или YouTube. На примере этих систем можно попытаться проследить общие законы таких ресурсов, без которых, возможно, они бы не стали сверхпопулярными.

Одним из основных факторов, делающих эти ресурсы интересными своей аудитории, является сама аудитория. Она постоянно генерирует содержимое. Механизмы саморегуляции сообщества обеспечивают естественный отбор самой интересной информации и самых популярных авторов. Это постоянно привлекает новых посетителей и создаёт отличную саморекламу.

Для первоначального набора необходимой «критической массы» аудитории всегда принимаются дополнительные меры, но они все малоэффективны, если процесс создания нового содержимого не упрощён до предела, а доступ к информации производится без каких-либо ограничений.

Мы хотели бы выделить общие черты этих систем, на наш взгляд, в значительной мере обусловивших их популярность:

- 1) Свободный просмотр любого содержимого.
- 2) Свободное добавление любого содержимого.
- 3) Всё время пополняющийся и актуальный контент.
- 4) Простота просмотра и добавления новых материалов.
- 5) Покрытие спроса пользователей в широкой области.
- 6) Начальная «критическая масса» содержимого.

На основе этих критериев можно сформулировать общие требования к Свободному Хранилищу Документов для образовательного процесса. Они слово в слово повторяют шесть перечисленных пунктов.

Первые два пункта могут быть удовлетворены за счёт создания простых удобных интерфейсов и выбора бизнес-модели, не подразумевающей непосредственной монетизации

контента. Предпоследнее требование легко удовлетворить, предоставив пользователям самим управлять содержимым и поощряя добавление информации, интересной сообществу.

Последний, но хронологически самый первый критерий требует найти способ быстро набрать «критическую массу» документов, после чего ресурс начнёт развиваться очень активно уже сам по себе. Наиболее быстрый и малозатратный способ это сделать — найти место, где уже пересекаются основные информационные потоки, откуда пользователи смогут наиболее просто и удобно добавлять свой контент в Хранилище.

Если от пользователя будет требоваться самое минимально возможное количество действий, чтобы добавить свой документ в Хранилище, он иногда будет этой функцией пользоваться. Чем больший поток данных и количество пользователей удастся охватить этим процессом, тем быстрее будет происходить это «тунеллирование» документов в Хранилище через «потенциальный барьер» лени пользователей.

При реализации третьего и четвёртого пункта возникают сложности, также требующие анализа потоков информации в образовательном процессе, предпочтений пользователей и специфики задачи.

1.2. Информационные потоки в образовании

В процессе обучения студенты постоянно используют информацию из книг, лекций или семинаров и обмениваются ей друг с другом. Причём многие из них считают, что обмен знаниями друг с другом не только помогает в освоении нового материала, но и является отличным способом структурирования знаний и необходим для глубокого понимания.

На примере образовательного процесса в МФТИ выделим основные источники знаний, используемые студентами:

- Конспекты лекционных курсов и семинарских занятий.
- Книжные библиотеки.
- Электронные книги и пособия.
- Самостоятельные письменные работы.

Все эти разнородные источники информации встречаются в одном месте – при переводе бумажных источников в электронный вид, а электронных – в бумажный.

За последние пять лет распространённость персональных компьютеров в студенческом городке МФТИ выросла в несколько раз: от одного компьютера на 4–5 человек до одного-двух ПК на одного человека. Интенсивность использования любых электронных источников информации выросла многократно с ростом доступности различных интернет-изданий и онлайн-библиотек.

С увеличением интенсивности использования электронных документов необходимость в переводе из бумажного в электронный вид и наоборот только возрастает – бумажные документы значительно удобнее читать, некоторые документы создаются именно в бумажном виде, при сдачах лабораторных, курсовых и дипломных работ используются печатные варианты.

Так мы можем концептуально обозначить «грань» между «мирами» реальных и виртуальных документов. Учащиеся постоянно переводят нужные им документы через эту «грань». Поэтому мы можем сказать, что эти документы

- непосредственно участвуют в образовательном процессе,
- заведомо интересны учащимся,
- постоянно обновляются.

Наконец, их поток со временем только увеличивается. А если оптимизировать и процесс перевода через «грань», это только ещё больше стимулирует процессы перехода.

Отсюда становится понятным, что, возможно, самое удобное место, где документы смогут «туннелировать» в Свободное Хранилище и есть эта «граница» между электронными и бумажными документами. Благодаря интенсивному потоку документов даже при малой начальной вероятности «туннелирования» документа, Хранилище начнёт пополняться и через некоторое время выйдет на режим самоподдержки и перейдёт в стадию экспоненциального роста.

1.3. Основная концепция построения Свободного Хранилища Документов

Основываясь на приведённой модели первоначального пополнения Хранилища, мы разработали Систему, логически состоящую из двух компонент:

- **Распределённая Система Общественной Печати** – та самая «грань», через которую пользователи будут централизованно переводить свои документы
- **Свободное Хранилище Документов** – объединено с системой печати, имеет с ней общий пользовательский интерфейс и реализует принцип минимальности действий при добавлении и просмотре содержимого.

Такое построение системы решает ряд внедренческих и экономических проблем.

Во-первых, система печати сама по себе оказывается очень востребованной, что уже делает всю систему привлекательной как для обучающихся, так и для руководства образовательных учреждений.

Во-вторых, Система может быть легко коммерциализована за счёт платного оказания услуг печати. Это легко принимают пользователи, а всю Систему делает меньше зависимой от конкретных условий в образовательном учреждении, делает её привлекательной с точки зрения инвестиций и позволяет ей быть как минимум самокупающейся.

Далее мы рассмотрим архитектуру и основные концепции, инструменты, программные и организационные решения, применённые в Распределённой Системе Общественной Печати и Свободном Хранилище Документов, функционирующей уже третий год на территории студенческого городка МФТИ.

2. Архитектура

Распределённая система общественной печати – это информационная система, объединяющая программно-аппаратный комплекс (рис. 1) общедоступных принтеров и систему обработки, хранения и анализа информации. С его помощью пользователи могут печатать документы, находить интересующую для себя информацию, максимально простым и удобным образом делиться своими документами.

Для того, чтобы система, основной расчёт которой делается на пользователей, была эффективна, она должна быть в первую очередь удобна в использовании. Именно критерии удобства определили архитектуру системы в том виде, в котором она есть сейчас.

Логически система состоит из трех типов компонентов (рис. 2):

- Веб-сервер обеспечивает работу пользователей с сайтом Системы.
- Сервер БД содержит описание документов системы и пользовательские настройки.
- Обсчитывающие сервера служат для анализа, хранения и отправки их на печать документов.

Сервера объединены между собой в виртуальную сеть с применением технологии VPN. В качестве VPN-сервера и клиентов используется OpenVPN [1].

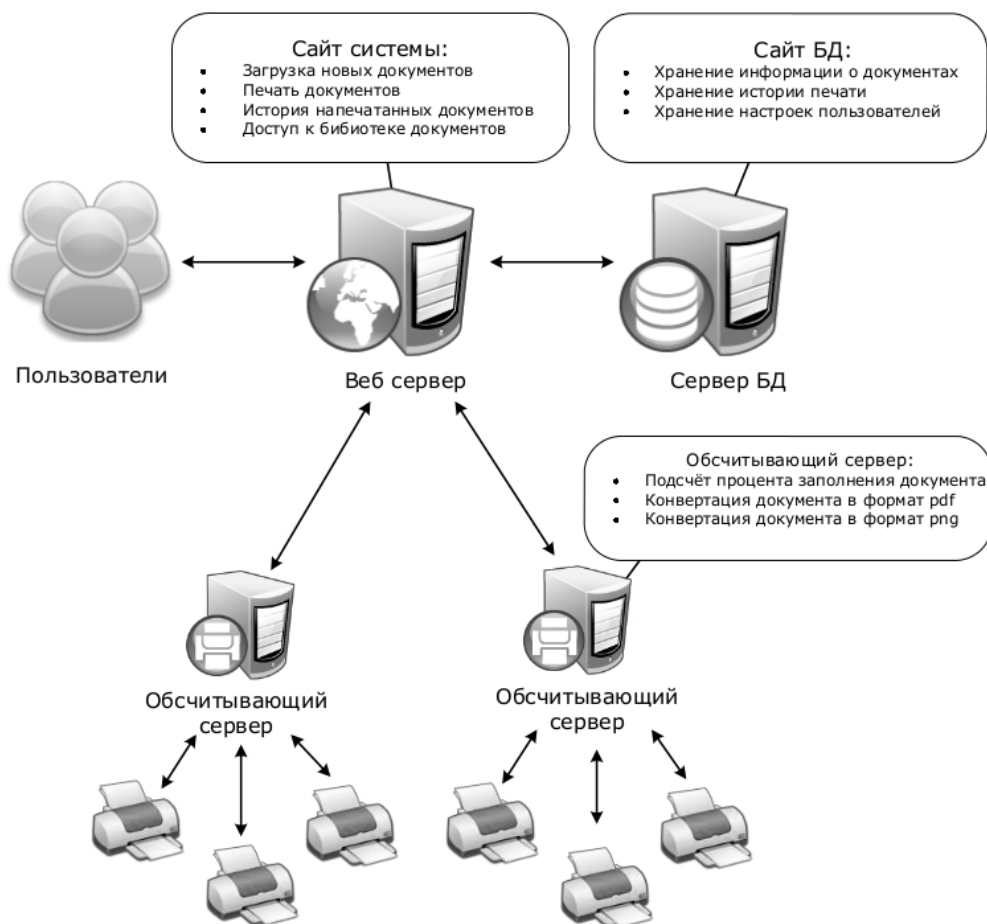


Рис. 1. Схема организации компонентов распределённой системы печати

2.1. Веб-сайт

С точки зрения пользователя центральным узлом системы является сайт. Такое архитектурное решение является не случайным. Именно применение веб-технологий позволяет решить множество проблем и сделать систему существенно более простой и удобной в использовании. Пользователь не должен устанавливать дополнительное программное обеспечение. Веб-браузер имеет любая операционная система. Отсутствие какого бы то ни было дополнительного ПО и требований к клиентскому компьютеру обеспечивает Системе платформенную независимость.

Основным направлением использования сайта является поиск и печать на принтерах Системы разнообразных учебных материалов, документов, статей. Процедура добавления нового документа также является ключевой и не требует никаких сложных дополнительных процедур.

Основной веб-сервер системы – Apache2, для генерации динамического содержимого и обработки запросов применяется PHP5.3 в виде mod_php. Веб-сервер настроен с учётом характера возникающих нагрузок.

2.2. Обсчитывающие сервера системы

Задача обработки документов является одной из ключевых для последующего анализа и организации совместного использования. Кроме того, при отправке исходные документы могут иметь достаточно большие размеры, что значительно затрудняет их передачу через Интернет. Для преодоления этих трудностей используются обсчитывающие сервера, которые располагаются в непосредственной близости как к принтерам, так и к пользователям.

Потоки данных

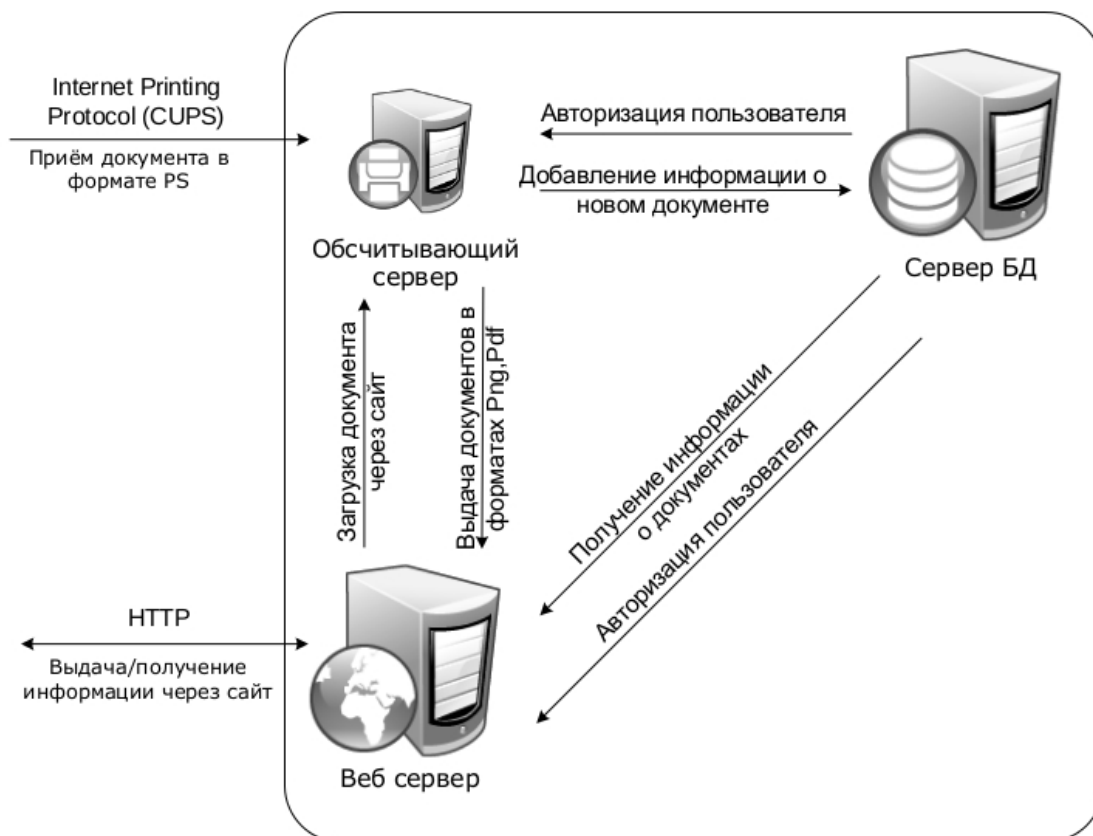


Рис. 2. Потоки данных в распределённой системе печати

Основными функциями обсчитывающих серверов является анализ документов, конвертирование их в необходимые форматы, хранение, а также передача информации о них на центральный сервер.

Обсчитывающие сервера также предоставляют доступ пользователям к веб-сайту Системы, выступая в роли прокси-серверов. В качестве прокси используется `mod_proxy` веб-сервера Apache2 с установленным `mod_rewrite` для обеспечения HRU [2]. Связь с веб-сервером и сервером БД осуществляется через VPN.

Отправка документов возможна либо с помощью сайта Системы, либо с помощью драйвера принтера посредством сетевого протокола `ipp`. Загрузка документов вторым способом предпочтительней, т.к. в этом случае снимается ограничение на изначальный формат документа, и напечатан может быть любой документ, который пользователь может у себя открыть.

При отправке документов через сайт существует значительное количество технических сложностей связанных с обработкой документов. Для того чтобы аналитический сервер имел возможность правильно совершить процедуры конверсий, ему необходимо точно знать, с каким типом документа он работает. Анализ типа документа и конверсия документов из разнообразных форматов, является задачей, которая не решается в общем случае. Отправка документов посредством драйвера принтера не имеет таких ограничений. Система осуществляет конвертирование в документы трёх форматов:

PNG – растровый графический формат, использующий сжатие без потерь [3]. Этот формат необходим системе для генерации собственного предпросмотра, для отображения в интернет-браузере, оценки процента заполнения для учёта тонера при печати заказа.

PDF – кроссплатформенный формат электронных документов, созданный фирмой

Adobe Software [4]. Предназначен в первую очередь для электронного представления полиграфической продукции. В нашем случае используется как основной формат документов в библиотеке.

PCL – язык управления принтером [5]. Инструкции этого языка позволяют управлять режимами печати принтера, задавать различные опциональные возможности, получать статусные сообщения от принтера.

Указанные преобразования производятся из исходного формата **Postscript** [6]. Это высокоуровневый аппаратно-независимый язык описания страниц, применяется как язык разметки полиграфической продукции. Он позволяет получать от пользователей документы в легко интерпретируемом формате вне зависимости от вида ОС, её версии и прикладного ПО, которое сгенерировало заказ на печать. Делает систему независимой от моделей принтеров и их возможностей.

Для приёма заказов по протоколу IPP [7] используется CUPS-сервер [8]. Протокол IPP был выбран как замена протоколу SMB/CIFS [9] из-за плохой переносимости последнего и невозможности обеспечить полную поддержку с помощью свободного ПО. IPP обеспечивает возможность передачи заданий печати через любые сети, поскольку является надстройкой над HTTP/1.1 [10], имеет отличную поддержку в CUPS во всех ОС, где он используется, и в ОС семейства Windows в числе стандартных, включенных по умолчанию компонент, кроме Windows Vista/7, где его требуется включать [11].

Для управления подключёнными принтерами, обеспечения обратной связи с принтерами и передачи заданий печати, используются разработанные нами backend'ы для CUPS.

3. Результаты

В качестве экспериментального внедрения для определения востребованности подобной системы был выбран студенческий городок МФТИ (Московского физико-технического института).

Испытание Системы [12] было естественным образом разбито на два этапа. Вначале при малом количестве пользователей и соответственно объёмах документов, поступающих в систему, основное развитие происходило как развитие сервиса печати. В этот период наблюдается монотонный рост количества пользователей и количества документов, прошедших через систему.

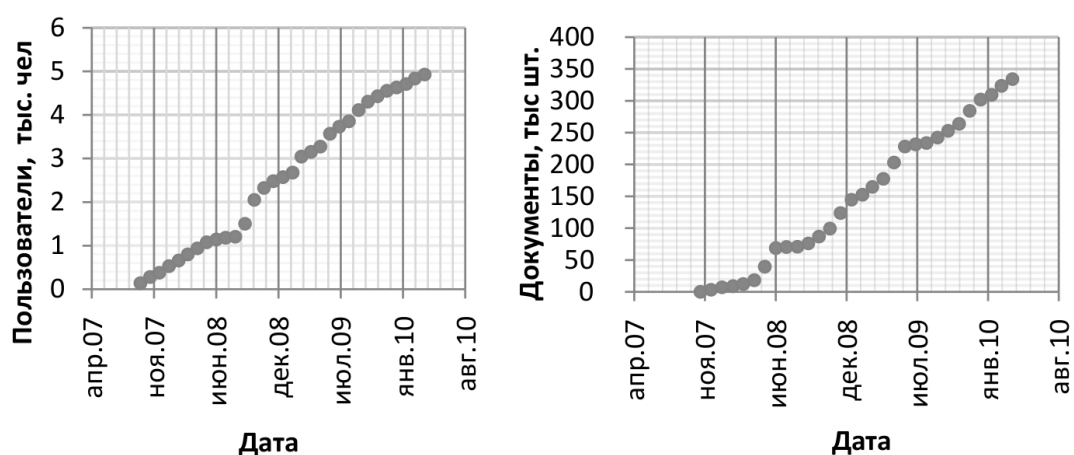


Рис. 3. Рост числа пользователей системы (слева) и числа напечатанных документов (справа)

Через два года после начала использования системы, когда накопилось достаточное количество активных пользователей (рис. 3), появилась возможность развития открытого хранилища. Сразу после своего создания открытая библиотека продемонстрировала быстрый рост количества документов (рис. 4).

Частота добавления документов в хранилище сначала резко выросла (рис. 5), затем постепенно стала снижаться, а частота извлечения документов остаётся на примерно постоянном уровне. Из этого можно сделать вывод, что пользователи быстро наполнили хранилище самыми часто используемыми учебными материалами. Анализ наиболее популярных документов подтвердил эту гипотезу: преимущественно это учебники и конспекты лекций по основным учебным дисциплинам.

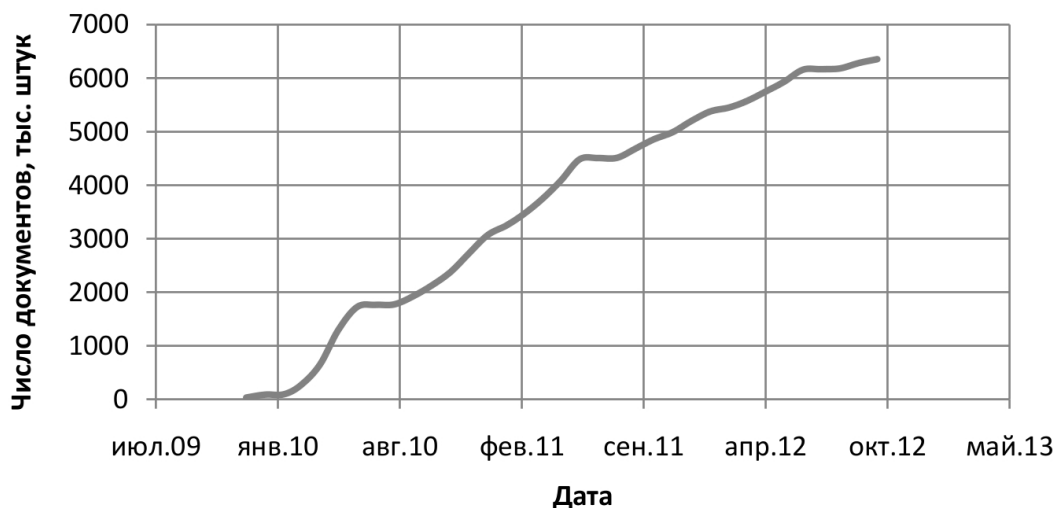


Рис. 4. Динамика числа документов пользователей системы (слева) и числа напечатанных документов (справа).

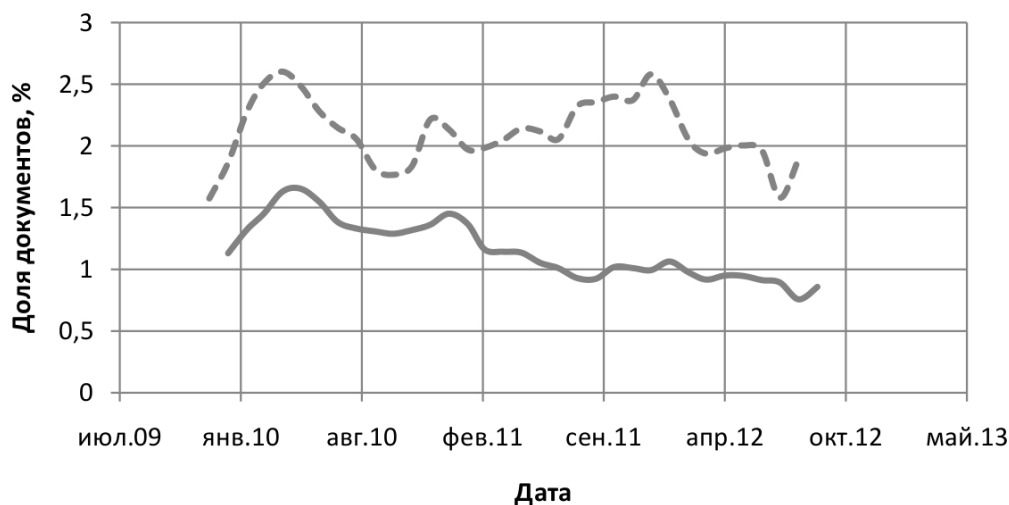


Рис. 5. Доля документов от общего числа напечатанных в системе за день, попавших в библиотеку (сплошная линия) и напечатанных из библиотеки (пунктирная линия).

Литература

1. Brumbaugh S. Deploying a VPN with PKI// Technology books, tech conferences, IT courses, news – O'Reilly media, 2004 — http://www.oreillynet.com/pub/a/security/2004/10/21/vpns_and_pki.html
2. HRU. Wikipedia. —<http://en.wikipedia.org/wiki/HRU>.
3. Portable Network Graphics (PNG) Specification (Second Edition)// World Wide Web Consortium (W3C), 2003 — <http://www.w3.org/TR/PNG/>.

4. PDF Reference and Adobe Extensions to the PDF Specification// Adobe Systems Inc., 2008 — http://www.adobe.com/devnet/pdf/pdf_reference.html.
5. PCL 5 Technical Reference Manual/ HP Corp., 1992.
6. PostScript Language Reference Third Edition/ Adobe Systems Inc., 1999 — ISBN 0-201-37922-8.
7. *Hastings T., Herriot R., deBry R., Isaacson S., Powell P.* Internet Printing Protocol/1.1: Model and Semantics// Internet Engineering Task Force, 2000. — <http://www.ietf.org/rfc/rfc2911.txt>.
8. CUPS. Common Unix Printing System/Apple Inc, 2010 — <http://www.cups.org/>.
9. Server Message Block. Wikipedia. — http://en.wikipedia.org/wiki/Server_Message_Block.
10. Hypertext Transfer Protocol – HTTP/1.1.// World Wide Web Consortium (W3C) — <http://www.w3.org/Protocols/rfc2616/rfc2616.html>.
11. TMG Ltd. Включение Internet Printing Client в ОС Windows Vista / Server 2008 / Физтех-Печать, 2009 — http://old.print.mipt.ru/docs/vista_ipp/.
12. TMG Ltd. Физтех печать/ 2007-2010 — <http://print.mipt.ru>.

Поступила в редакцию 31.10.2012.