РУКОПИСЬ.doc: ЗАМЕТКИ О ТЕКСТОЛОГИИ ЦИФРОВЫХ ДОКУМЕНТОВ

АРТЕМ ШЕЛЯ
(Тарту, Тартуский университет)

Когда мы берем в руки карандаш и что-то записываем на листе бумаги, то интуитивно понимаем физические отношения между носителем информации и инструментом письма. Бумага, как правило, достоверно хранит следы контакта с ней: кофейные пятна, записи, слои исправлений, пометки, рисунки. Большинство этих следов доступно любому наблюдателю. Мы можем ничего не понимать в палеографии, не представлять особенности химического состава бумаги или чернил и не разбираться в истории бумажного производства, но это не помешает нам прочесть основную информацию в рукописи. И даже если для текстолога косвенные свидетельства, информация вспомогательных дисциплин и химической экспертизы могут оказаться ключевыми в деле изучения текста, интуитивное представление о физических свойствах рукописи доступно каждому. Мы приблизительно знаем, как можно сохранить бумажную рукопись и — что не менее важно — как ее уничтожить, если мы не хотим, чтобы какая-то информация когда-либо оказалась у третьих лиц.

Несмотря на то, что современные электронные носители по своей сути являются технологиями «надписи» на поверхности (магнитной или оптической)[1] [Kirschenbaum 2008: 93], наши

[1] За исключением полупроводниковой флеш-памяти, стремительно вытесняющей жесткие диски.

интуитивные представления о поведении и свойствах цифровых документов (и вообще цифровой информации), как правило, очень далеки от действительности, которая обычно скрыта от пользователя слоями металла, экрана, графических интерфейсов и заключенных в них метафорических функций («открыть», «вырезать», «вставить», «сохранить», «удалить» и т. д.).

Когда мы «открываем» текстовый файл (например, .doc), он копируется с носителя в оперативную память (RAM), где находится информация, использующаяся компьютером в данный момент. Оттуда программа работы с текстом (например, MS Word) реконструирует документ, следуя указаниям, содержащимся внутри самого файла. Большинство подобных процессов — копирование, проверка копий, реконструкция информации, интерпретация структуры файлов — скрыто от пользователя, как скрыты и следы, оставленные документом, который проходит эти трансформации. Например, части текстовых документов можно восстановить напрямую из RAM даже после закрытия файла [Al-Sharif et al.]. То же относится к «удалению» файлов: если документ удаляется при помощи операционной системы, то он не уничтожается *физически,* хотя его уже нельзя обнаружить в привычных папках. Удаляется только путь к документу из файловой системы, связывающей конкретный поток битов на физическом носителе с ярлыком, названием этого файла. Сама информация (биты) остается нетронутой, однако открывается для дальнейшей перезаписи. Нетрудно представить, что при современных объемах информационных носителей биты, составляющие «удаленный» файл, могут долго оставаться не перезаписанными. Восстановление путей к удаленной информации — это рутинная операция во многих криминалистических программах для исследования жестких дисков.

Было бы легко определить цифровой документ как последовательность нулей и единиц, отрезок двоичного кода, записанный

где-то на электронном носителе. Но также это было бы малопродуктивно, поскольку нули и единицы можно декодировать бесчисленным множеством способов: представить их в шестнадцатеричной или десятеричной системе, прочесть как текст в определенной кодировке, как временную структуру (FILETIME), звуковой поток или изображение. В этом отношении цифровая информация неотделима от способов ее интерпретации и представления и не сводится к материальной цепочке битов в носителе, поэтому исследователи говорят о распределенной [Blanchett] и «формальной» [Kirschenbaum 2008] материальности цифровых данных, которая складывается из работы множества компьютерных систем.

Эти особенности цифровой информации ведут к ряду сложных проблем, связанных с ее долгосрочным (архивным) хранением, формами доступа, правовым статусом (вопрос об идентичности и происхождении документа, конфиденциальности данных)[2]. Для исследователя литературных текстов природа рукописи, созданной на компьютере, также перестает быть очевидной. За аккуратным интерфейсом текстового процессора, послушно воспроизводящим символы в (обычно) правильном порядке, лежат невидимые пользователю системы и структуры данных, которые и позволяют тексту появиться на экране в определенном виде.

Исследования природы цифровых текстов и сопутствующих им программ, в силу технической специфики проблемы, в основном принадлежат области компьютерной криминалистики[3]. Однако начиная с пионерского исследования М. Киршенбаума о природе цифровых данных [Kirschenbaum 2008], научный подход

---

[2] Об общей проблеме сохранения цифровых данных см.: [Kirschenbaum 2013; Owens].

[3] См. специальные работы о криминалистическом подходе к текстовым документам: [Castiglione et al.; Garfinkel & Migletz; Fu et al.]

к сохранению подобной информации распространился в ряде гуманитарных дисциплин, открыв путь и текстологической работе с электронными текстами (см.: [Ries] и библиографию).

Мои краткие и разрозненные заметки посвящены информации, которая может быть скрыта внутри текстовых файлов, но никак не обнаруживать свое присутствие во время обычной работы с документом. Потенциально эта информация может быть значимой для текстологического анализа: прояснять историю текста, хранить следы бытования документа в различных операционных системах и, в некоторых случаях, предоставлять возможность для реконструкции разночтений и правок. Как я постараюсь показать на материале личного архива Е. А. Шварц, достаточно самого простого знакомства с устройством цифровых рукописей, чтобы получить доступ к информации, запись и сохранение которой зачастую находится в области, контролирующейся не пользователем, а программным обеспечением и особенностями текстового формата.

Домашний архив Е. А. Шварц и цифровые рукописи

Елена Шварц (1948–2010) — важнейший поэт ленинградского литературного андерграунда 1970–1980-х годов — начала пользоваться компьютером в конце 1990-х. Ее цифровые документы, созданные и редактировавшиеся на множестве систем в различных версиях текстового процессора MS Word, являются частью личного архива и естественным продолжением его «бумажной» части (автографов и машинописей). В некоторых случаях бумажные копии цифровых документов, сохранившиеся в архиве, можно связать с их оригиналами на компьютере: Шварц распечатывала стихи для чтения и гранки книг на правку.

От хранителя архива мы получили коллекцию файлов, собранных с нескольких жестких дисков уже после смерти поэта. Эти файлы были скопированы на новый носитель и поэтому утратили

всякую связь со своим физическим источником: нам не доступно ни изначальное расположение документов в файловой системе, ни место этой информации в оригинальном потоке битов и структуре цифровой среды. Обычным методом сохранения цифровых данных является создание «образа» носителя — точной побитовой копии, которая отражает некоторые физические особенности оригинала (геометрию файловой системы, точный порядок битов, сохраняющий остаточные данные, в том числе и удаленную информацию [Kirschenbaum 2008: 110; Ries: 391]). Несмотря на отсутствие таких копий источников, в самих файлах хранятся некоторые отпечатки тех систем и программ, в которых создавался и редактировался текст.

В частности, в метаинформацию текстовых форматов .doc и .docx записываются данные о версии программы и авторе документа (имя пользователя операционной системы)[4]. Сравнивая изменения в имени и версии MS Word с датами создания различных файлов, можно приблизительно реконструировать историю того, как менялись компьютерные системы у Шварц:

пользователь	**годы документов**	**версия Word**
AAA Elena Schwarz Schwarz	1998	8.0
Лена Elena	2001	8.0 / 9.0
Elena	2001–2003	8.0
Лена Lena	2004	8.0
USER	2004–2006	10.0

[4] Эта информация легко доступна в современных версиях Windows при просмотре свойств документа. Существует также множество программ с открытым кодом, позволяющим читать метаданные прямо из файла. Я пользовался **olefile** и **ExifTool**.

Home	2006–2008	10.0
XTreme	2009–2010	12.0

Изменение имени пользователя обычно можно ожидать при смене операционной системы, хотя никаких ограничений по изменению этих данных изнутри системы обычно нет. Кроме того, на одной системе могут существовать несколько пользователей. Однако в целом имена и версии, как видно из таблицы, сменяются хронологически и не возвращаются к прежним значениям (особенно с начала 2000-х гг.), что говорит о постепенной миграции Шварц на новые компьютерные системы. При этом видно, что сосуществование нескольких систем/жестких дисков также было обычным делом. Файлы 2001 года распределены между пользователями «Лена» и «Elena». В 2007 г., во время стабильного «Home» с Word 10.0 (Office XP) появляются файлы, созданные пользователем «Lena» на раннем Word 8.0 (Office 97).

Сама по себе эта информация сообщает не слишком многое. Однако полезно понимать границы домашней системы автора и вариативность в программах для работы с текстом. Мы видим, что большинство времени Шварц пользовалась Word 8.0, вышедшим в 1997 г. В 2001, однако, появляются документы, созданные на Word 9.0 (1999 г., Office 2000). Если это не случайный артефакт метаданных, то у Шварц эта версия не задерживается и не вытесняет предыдущую — поэт вновь возвращается к ранней версии Word, с которой останется до 2004 года (видимо, до обновления операционной системы). Кроме того, хронология в таблице сообщает, что в этом архиве мы в основном столкнемся с «непрозрачным» бинарным форматом .doc, а не .docx, основанным на XML-разметке и введенным в употребление в Office 2007 (Word 12.0). Последние два года Шварц работала в Office 2007, однако .docx

файлов в доступной коллекции так мало, что можно предположить, что даже на поздней системе с новым Word 12.0, Шварц предпочитала хранить тексты в привычном формате .doc.

Составной бинарный формат (Compound Binary Format) .doc файлов можно представить как маленький диск, в который «сложены» разные потоки данных, связанных между собой файловой системой. Один из этих потоков — WordDocument — содержит и текст, который мы видим на экране. Однако сам текст занимает лишь маленькую часть файла, который заполнен многочисленными таблицами, определяющими связи между элементами файла, инструкциями по его форматированию для Word и метаданными. Такой файл, в отличие от XML-кода, человек практически не может понять, и предназначен он исключительно для машинной интерпретации. Из-за своей сложной структуры .doc файлы содержат особенно много остаточной информации, которая может оказаться важной для текстологической работы.

Представление об исторических напластованиях систем и программ важно еще и потому, что дает возможность отслеживать вмешательства в авторский массив данных. Один из «погодовых» файлов со стихотворениями, «2003.doc», создан пользователем «Elena», однако последнее изменение сделано неким «ksn» в 2015 г. Так как при каждом сохранении информация в файле о рабочей системе перезаписывается, то и в метаданных осталась версия Word со времени последней правки — это 14.0, вышедшая спустя месяц после смерти поэта. Очевидно, что это вмешательство в документ было совершено кем-то не связанным с домашней системой, без доступа к оригинальным жестким дискам (иначе имя пользователя было бы знакомым). Впрочем, эта загадка решается просто: «ksn» — это мое пользовательское имя на старом ноутбуке, куда были скопированы файлы из архива Шварц. Изменения в резервной копии файла произошли, по-видимому, случайно при просмотре.

Старые пути

Word с включенным режимом автосохранения записывает внутрь .doc файлов временные пути к автоматическим копиям в операционной системе, чтобы при необходимости документ можно было восстановить. Версии программы с 8.0 по 10.0 также записывали имена 10-ти последних авторов, редактировавших текст. Вместе с именем сохранялось и расположение файла в системе. Таким образом, некоторые документы в постоянном обращении накапливали «след» своего присутствия в различных системах и папках. Доступ к подобным артефактам можно получить, открыв .doc файл в HEX-редакторе или простом текстовом редакторе (блокноте), который прочтет весь файл от начала до конца как набор символов. Рассмотрим пару примеров.

Файл «пинд.doc» со стихотворениями 1998–2000 гг. сохраняет следы по крайней мере трех «авторов». Он создан 19 октября 1998 года пользователем «AAA» и в последний раз изменен 11 марта 2001 года как «Elena». Внутри файла обнаруживается и промежуточное место существования документа:

```
Лена#C:\DOS\Автокопия пинд.asd
Лена#C:\Windows\Desktop\пинд.doc
Elena#C:\WINDOWS\TEMP\Автокопия пинд.asd
Elena#C:\Мои документы\пинд.doc
```

Вместе с путями автосохранений (.asd) в файле записано расположение документов и имена сменявшихся пользователей. Файл «пинд.doc» в какой-то момент переместился с рабочего стола «Лена» в «Мои документы» нового пользователя «Elena».

В другом случае внутри документа сохранился путь к диску A:\. Этой буквой Windows традиционно обозначает дисковод гибких дисков — свидетельство о том, что документ был скопирован на дискету, изменен на ней и затем перенесен в новую систему:

Schwarz#C:\DOS\Автокопия Соло на раскаленной трубе.asd#
Schwarz#C:\WINWORD\Геликон\poem\соло\Соло на раскаленной трубе.doc#
Schwarz#A:\соло\Соло на раскаленной трубе.doc#
Elena#C:\WINDOWS\TEMP\Автокопия Соло на раскаленной трубе.asd#
Elena#C:\WINDOWS\Рабочий стол\Соло на раскаленной трубе.doc

Отметим, что впоследствии этот файл с одноименной книгой Шварц 1998 года изменил название, став «Соло.DOC». Это могло произойти из-за технических особенностей файловой системы, иногда упрощающей длинные файлы, однако важнее здесь другое: документ, сменивший заглавие, может сохранить данные о своем прошлом. Кроме того, в «Соло.DOC» отражен небольшой фрагмент личной структуры файлов с жесткого диска двадцатилетней давности, к которому у нас никогда не было прямого доступа. Файл с книгой в 1998 г. был расположен внутри директории «Геликон\», видимо содержавшей основные цифровые рукописи. Обитель муз у Шварц находилась не так далеко от нового инструмента письма (WinWord).

Название (Title)

В метаданных, которые находятся внутри .doc формата, есть особое поле «Title» — название документа, зачастую не совпадающее с его файловым именем (таким как «Соло.doc», «пинд.doc» и т. д.).

При определенных условиях[5] в title записывается *изначальная* первая фраза текстового документа и затем не меняется по мере изменений файла. Это «название» позволяет отслеживать небольшие изменения в составе документов и, в некоторых случаях, обнаруживать разночтения в текстах.

Файл с книгой «Соло на раскаленной трубе», созданный 4 июля 1998 г., начинается с титульного листа, на котором указано имя автора и название сборника. Однако поле title в метаданных регистрирует другой текст — «О несовершенстве органов чувств». Это первая строка стихотворения, с которого, вероятно, начинался файл во время создания документа. Можно предположить, что этот текст и открывает сборник, однако это не так. «Соло на раскаленной трубе» начинается с более резкого манифеста, с «Маленькой оды к безнадежности», тогда как «О несовершенстве...» следует вторым. Состав книги изменился, однако файл сохранил след ранней композиции.

Другой знакомый документ «пинд.doc» открывается стихотворением 1998 г. «Никого кроме Тебя / Больше нету у меня». Название в метаданных, однако, указывает на существенное разночтение: «*Ничего* кроме Тебя». Переход от неодушевленного к одушевленному отрицательному местоимению сильно меняет это обращение к Богу, располагая адресата не среди сущностей,

---

[5] Насколько можно судить, это происходит из-за автоматического заполнения полей метаданных при неопределенности документа. Если сначала запустить Word, в открывшемся пустом документе ввести текст, а после сохранить документ, то программа предложит выбрать директорию сохранения и автоматически задаст его название на основе первой строки текста в файле. Даже если изменить это название, первая строка будет записана в метаданные как Title. Она останется там даже в том случае, если первая строка изменится или вовсе перестанет быть первой строкой. Это не работает, если пустой документ создан заранее в конкретной директории. Тестировалось с Windows 98 SE на Word 8.0.

а среди существ (близких и семьи). Интимная религиозная лирика была одним из ключевых жанров в поэзии Шварц.

Принцип формирования поля title, как мы видим, позволяет уловить лишь фрагменты истории текста при случайном стечении обстоятельств. Иногда метаданные могут указать и на композиционные изменения в стихотворении. Большой документ «2005 1.doc» начинается с январского стихотворения «Сердце то будто тихий дождь идет...». Текст сопровожден автоэпиграфом «Разрослой клубникой / Сердце сладеет...». Однако title файла фиксирует не эпиграф, а сразу первую строку: соответственно, эпиграфа там изначально не было и Шварц внесла его после создания документа (4 февраля 2005 г.). Можно возразить, что Шварц могла скопировать текст из неизвестного источника, пропустив эпиграф. Действительно, во всех документах с этим текстом, созданных после «2005 1.doc», эпиграф присутствует («npo.doc», «poems 2005.doc», «Разрослой клубникой.doc»). В файле «npo.doc» (создан спустя месяц — 13 марта), который начинается с «Сердце то будто тихий дождь идет...», title уже фиксирует первую строку эпиграфа, т. е. сам предваряющий текст появился не позднее этого времени. Однако единственный возможный источник текста для февральского файла находится в первом документе со стихами 2005 года («2005.doc», создан 2 января) и там никакого эпиграфа еще нет. Все это говорит о том, что изначально стихотворение существовало без эпиграфа, который Шварц добавила в текст уже после 4 февраля и не позднее 13 марта. В свою очередь, стихотворение в файле «2005 1.doc» стало источником для всех поздних копий.

Артефакты быстрого сохранения

В Word долгое время была функция «быстрого сохранения» (fast save). Если она включена, то текстовый процессор во время сохранений не перезаписывает документ каждый раз целиком, а заносит слои правок прямо в поток бинарного файла. Это должно уменьшать нагрузку на жесткий диск и ускорять работу с большими документами. Однако структура .doc файла при этом становится чрезвычайно сложной: изменения, внесенные при сохранении, записываются как отдельные потоки данных и хранят информацию о своем положении относительно оригинала и друг друга. Если открывать такие .doc файлы родным редактором, то в нем не отобразится ничего необычного, однако в других программах весь скрытый хаос проявится. Очевидно, что подобные файлы с быстрым сохранением являются важнейшим источником для текстологии цифровых документов.

В архиве Шварц, насколько я могу судить, файлов с артефактами быстрого сохранения немного. В Word 8.0 эта функция вообще была изначально отключена. Однако, по-видимому, Шварц использовала ее в работе с большими документами на ранних системах. Множество данных, оставшихся от быстрого сохранения, обнаружилось в файлах «Лавиния.doc» (создан 27 июля 1998 г.) и «Мусагет.doc» (10 апреля 2003).

Оба этих текста были завершены до того, как Шварц начала работать с компьютером. Книга «Труды и дни Лавинии, монахини из ордена Обрезания Сердца», соединившая все важнейшие поэтические темы Шварц 1970–1980-х гг., датируется 1984 годом и была издана в «Ардисе» в 1987. Сборник поэм «Хомо мусагет» (тексты 1970–1990-х годов) был подготовлен к печати в 1996 г. (сохранилась чистовая машинопись), но остался неизданным.

Здесь мы сталкиваемся со случаем набора объемных и уже сформировавшихся текстов на компьютере. Основная часть артефактов в файлах — это густой слой мелкой правки опечаток. Внутри документов обнаруживаются длинные списки замененных букв и знаков препинания. Иногда записываются замененные строки и части строк: «камнями, на букы<ах>», было исправлено на «камнями, на буквах». Периодически можно восстановить колебание Шварц в пунктуации: быстрое сохранение записывает варианты «птицею в пруду,» и «птицею в пруду –». В окончательном тексте документа стоит многоточие: «Мелькнула птицею в пруду...» (в редакции «Ардиса» есть четвертый вариант: «в пруду, –» [Шварц 1987: 95]). Следов существенных отличий от ранних редакций в документах нет, однако безусловно ситуация набора текста на компьютере служила импульсом к очередному пересмотру текста и появлению новых разночтений.

Идентификаторы изменений (.docx)

Функцию быстрого сохранения, слишком часто приводившую к порче файлов, Microsoft отключили в 2007 году одновременно с тем, как перешли на открытый формат документов, основанный на XML-разметке (Office Open XML). В этом формате, рассчитанным на простоту и совместимость с любыми системами, откладывается намного меньше информационных артефактов. Формат .docx представляет собой обычный .zip архив (расширение можно просто поменять вручную), в котором хранится несколько .xml файлов, описывающих и представляющих документ кодом, который человек может (в принципе) прочесть. В целом Microsoft со временем дали пользователям больший контроль над скрытыми метаданными, а операционные системы стали отображать этой информации больше по сравнению с 1997 годом.

Однако и здесь внутреннее устройство формата оставляет пространство для неожиданной косвенной информации, присутствие которой незаметно для пользователя и может быть чувствительным. Речь идет о т. н. идентификаторах изменений (revision id, RSID), которые в xml-тегах сопутствуют тексту внутри основного архивного файла document.xml [Fu et al. 2011]. Это случайно сгенерированный код, служащий для того, чтобы Word мог отслеживать версии документа (в т. ч. — коллективную правку) и правильно интерпретировать изменения. Идентификаторы создаются в любых условиях, даже с отключенной функцией Track Changes. Текст, записанный одним пользователем в едином форматировании за одну сессию (сохранение), получит одинаковые идентификаторы. Сохранение, форматирование частей и другие манипуляции с текстом вызовут генерацию новых идентификаторов на измененных местах. Кроме того, все RSID документа дополнительно откладываются в файле settings.xml. Даже после удаления изначального текста его идентификаторы останутся. Для криминалистики это значит, что можно установить след от контакта двух документов (если один является копией другого), даже если тексты внутри будут полностью различаться. Для текстологии это значит, что след от контакта автора с рукописью все еще можно обнаружить.

Постараюсь показать это на примере позднего стихотворения Шварц, записанного в один из немногих документов OOXML формата — «Игольчатое море.docx».

00754FD1 Купанье прачки
00722011
00754FD1 Вошедши с плотомойни в реку
00754FD1 Нагая баба говорила-
00754FD1 Какой ты Оредеж холодный-
00754FD1 Как будто молодцу случайному
00754FD1 Или родному человеку.
00754FD1 Какой холодный ты сегодня…

00754FD1 Сказала и погладила рукою
00754FD1 Нагую воду с нависшей от березы тенью
00754FD1 А Оредеж стремительный и темный
00754FD1 Как будто бы чурался ее горячего бесформенного тела
00754FD1 И мимо пролететь хотел
00754FD1 И избегал ее прикосновенья
00722011 Как будто не рекою был а духом
00722011 Или горою льдистой
00722011 Что с отвращеньем будто муху
00722011 В алмазах терпит альпиниста.

00754FD1 1июля

00754FD1 горячее живое
00754FD1 И пышное

Параллельно с текстом стихотворения приведен идентификатор изменений «rsidR», соответствующий каждой отдельной строке[6]. Вся эта система изначально не несет никакой другой функции, кроме идентификации изменений. Она не содержит информации о времени изменений, об их типе, источнике или пользователе. Однако с первого взгляда на идентификаторы мы можем представить как мог складываться этот текст.

Сначала Шварц пишет стихотворение до строки «И избегал ее прикосновенья» и заканчивает его, ставя под текстом дату — 1 июля. Затем, вероятно, сохраняет документ и через какое-то время возвращается к нему. На данном этапе все последующие изменения в тексте появляются с новым значением RSID. Поэт отделяет заглавие от основного текста (идентификатор у пустой строки) и дописывает последние четыре стиха.

---

[6] На самом деле этот идентификатор описывает параграф, но так как стихи отделяются друг от друга, то Word интерпретирует каждый из них как отдельный параграф.

Обратный вариант — четыре стиха пишутся сначала, остальное стихотворение вырастает вокруг — невозможен. Это не отражено в выписке, но «второй» тип rsidR (00722011) на уровне текста также относится к одной финальной букве «ю» в строке «Нагую воду с нависшей от березы тенью». Шварц исправляла опечатку в уже существовавшем тексте (с идентификатором параграфа 00754FD1). Даже без этой детали обратный вариант был бы маловероятен: последний катрен с внезапной перекрестной рифмовкой и точными рифмами формально слишком сильно отличается от остального текста и больше подходит на роль финальной фигуры, дописанной в противовес расшатанному стиху. К тому же, сложно представить, что стихотворение развило свою синтаксическую структуру вложенных сравнений из последних стихов, также начинающихся со сравнения. Впрочем, вне зависимости от этих рассуждений, всегда остается возможность того, что второй слой правки (00722011) скрыл от нас какую-то изначальную концовку, полностью переписанную Шварц.

Обрывки фраз, следующие после даты, наверняка представляют собой перебор эпитетов и относятся к строке «Как будто бы чурался ее горячего бесформенного тела». Эти фрагменты были созданы во время первой итерации текста (00754FD1) и своим присутствием (вместе со слоями RSID) делают цифровую рукопись похожей на черновик. Возможно, так и было: текст стихотворения был не перенесен Шварц в Word с какого-то источника, а создавался сразу в текстовом документе.

Заключение

Архитектура компьютеров и пользовательских приложений создает иллюзию интуитивного понимания цифрового текста, отражающегося на экране монитора. Интерфейсы, среди которых живет человек, с течением времени лишь усиливают действие этой

иллюзии, располагая цифровую информацию все ближе к нашим ожиданиям от устройства мира. Однако эта видимость поддерживается сложнейшими системами, которые на самом деле структурируют, описывают и представляют данные в неконфликтной и понятной форме.

Как я старался показать, даже поверхностное исследование знакомых файлов простыми инструментами предоставляет информацию об операционных системах, которых давно может не существовать, об изменениях и разночтениях, десятилетиями хранящихся внутри текста, который выглядит законченным. Текстологии еще предстоит найти язык для работы с «распределенными» документами, которые предстают как чистовая редакция, содержат в себе следы предыдущих состояний (черновика) и сами себя описывают, следуя особенностям формата и текстового процессора.

Эти заметки, однако, не только о текстологии. Сохранение любой информации о тексте должно находится в области авторского контроля. Все информационные артефакты и косвенные свидетельства, о которых здесь шла речь, не являются неуязвимыми. Чтобы избавиться от невидимого шлейфа информации в бинарном формате .doc, достаточно сохранять документы в .rtf (Rich Text Format). Некоторые программы (и поздние версии Windows) позволяют чистить метаданные, однако вероятность сохранения артефактов все равно остается. Информацию, которую позволяют считать идентификаторы изменений в .docx, легко «сбросить», скопировав готовый текст в новый файл (впрочем, на Word 12.0 даже в этом случае RSID переходят в новый документ, см.: [Fu et al. 2011]). Почти неизвестной областью остаются «облачные» инструменты письма, над которыми у пользователя почти нет никакого контроля [Roussev et al.]. Google Docs, например,

сохраняют полную историю действий в каждом документе с точностью до символа и микросекунды [Somers][7]. Полная информация доступна всем, у кого есть права редактуры, т. е. не только создателю документа. Можно ожидать, что это окажется беспрецедентным источником по изучению письма. Однако возможность писать, не сохраняя исправлений, должна остаться.

Литература

Шварц: *Шварц Е. А.* Труды и дни Лавинии, монахини из ордена Обрезания Сердца. Анн-Арбор, 1987.

Al-Sharif et al.: *Al-Sharif Z. A., Bagci H., Zaitoun T. A., Asad A.* Towards the Memory Forensics of MS Word Documents // Information Technology — New Generations. Advances in Intelligent Systems and Computing. 2018. Vol. 558. P. 179–185.

Blanchette: *Blanchette J.-F.* A material history of bits // Journal of the Association for Information Science and Technology. 2011. Vol. 62(8). P. 1042–1057.

Castiglione et al.: *Castiglione A., De Santis A., Soriente C.* Taking advantages of a disadvantage: Digital forensics and steganography using document metadata // *Journal of Systems and Software.* 2007. Vol. 80(5). P. 750–764.

Fu et al.: *Fu Z., Sun X., Liu Y., Li B.* Forensic investigation of OOXML format documents // Digital Investigation. 2011. Vol. 8(1). P. 48–55.

Garfinkel & Migletz: *Garfinkel S. L., Migletz J. M.* New XML-Based Files Implications for Forensics // IEEE Security and Privacy Magazine. 2009. Vol. 7(2). P. 38–44.

Kirschenbaum 2008: *Kirschenbaum M.* Mechanisms: New Media and the Forensic Imagination. Cambridge, 2008.

[7] Всю историю изменений можно получить и визуализировать при помощи расширения Draftback для Chrome.

Kirschenbaum 2013: *Kirschenbaum M.* The .txtual Condition: Digital Humanities, Born-Digital Archives, and the Future Literary // Digital Humanities Quarterly. 2013. Vol. 7(1).

Owens: *Owens T.* The Theory and Craft of Digital Preservation. LIS Scholarship Archive, 2017. 10.31229/osf.io/5cpjt (доступно на 18.12.2018).

Ries: *Ries T.* The rationale of the born-digital dossier génétique: Digital forensics and the writing process: With examples from the Thomas Kling Archive // Digital Scholarship in the Humanities. 2018. Vol. 33 (2–1). P. 391–424.

Roussev et al.: *Roussev V., Ahmed I., Barreto A., McCulley S., Shanmughan V.* Cloud forensics — Tool development studies & future outlook // Digital Investigation. 2016. Vol. 18. P. 79–95.

Somers: *Somers J.* How I Reverse Engineered Googe Docs to Play Back Any Document's Keystroke. Публикация 05.11.2014. http://features.jsomers.net/how-i-reverse-engineered-google-docs/ (доступно на 18.12.2018).