



IFLA
2005
OSLO

World Library and Information Congress: 71th IFLA General Conference and Council

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

Conference Programme:

<http://www.ifla.org/IV/ifla71/Programme.htm>

Июль 15, 2005

Code Number:

151-R

Meeting:

119 Acquisitions and Collection Development

Различные подходы к пользовательским коллекциям.

Наше цифровое наследие как источник информации для пользователей: сбор сетевых публикаций и организация доступа к ним в Национальной библиотеке Норвегии

Керсти Рустад,
Национальная библиотека Норвегии

Translated by: Natalia Litvinova
Nlit@rsl.ru

Аннотация

Национальная библиотека Норвегии – центр мультимедийной информации, который предлагает своим пользователям документы на любых носителях информации. В этом докладе рассказывается, как Национальная библиотека собирает сетевые публикации в соответствии с Законом об обязательном экземпляре и каким образом эти документы можно сделать доступными для пользователей в качестве источников информации.

Организация доступа к Веб-архиву, содержащему миллионы страниц, собранных за несколько лет, – сложная проблема. Каким образом конечный пользователь, исследователь или документовед, получает доступ к Интернет-архиву? В докладе рассматриваются возможности доступа и ограничения, налагаемые действующим законодательством.

1. Введение

Недавно в Национальную библиотеку Норвегии обратился исследователь, занимающийся историей норвежской прессы. Он проводил сравнительное

исследование подачи новостей в печатных газетах, по радио и телевидению и в Интернет-газетах, поэтому ему нужен был доступ к источникам. Пожалуй, этот эпизод иллюстрирует типичный пользовательский запрос. Вопрос в том, может ли национальная библиотека предоставить такую информацию?

Национальная библиотека Норвегии – хранилище памяти нации; в миссии библиотеки провозглашается, что она является мультимедийным центром информации. Национальная библиотека хранит норвежское культурное наследие и обеспечивает доступ к документированным источникам о норвежской культурной и общественной жизни для нынешних и будущих пользователей.

Поэтому ответ на поставленный вопрос – утвердительный. Предполагается, что Национальная библиотека обеспечивает пользователей источниками на любых носителях информации. Основное средство достижения этой цели – наш Закон об обязательном экземпляре.

Когда в 1990 году была введена в действие последняя редакция закона об обязательном экземпляре, Норвегия одной из первых стран в мире включила в законодательный акт обязательный экземпляр электронных документов. Закон основан на принципе, согласно которому вся общедоступная информация, независимо от формы и носителя, должна сохраняться и предоставляться в доступ в целях научных исследований и документооборота. И хотя Интернет в его нынешнем виде не существовал в 1990 году, действующий закон об обязательном экземпляре распространяется на общедоступные документы Интернета. Это является законодательной базой деятельности Национальной библиотеки Норвегии по сбору сетевых публикаций.

2. Сбор сетевых публикаций.

Сбор сетевых публикаций в тестовом режиме выполнялся с середины 90-х годов, а первый тотальный отбор всего норвежского сектора веба был выполнен в декабре 2002 года. Однако только в 2005 году Норвежская национальная библиотека начала регулярный сбор норвежских сетевых публикаций.

Наша стратегия отбора основана на собирании всех общедоступных документов норвежского веб-пространства, то есть документов домена “no”. Поскольку Закон об обязательном экземпляре распространяется на все общедоступные документы, как опубликованные норвежскими издательствами, так и специально предназначенные для норвежских пользователей, документы других доменов, таких как org и com, также можно собирать в рамках этого закона. Это означает, что в будущем мы планируем подключить другие домены. Кроме того, мы не исключаем сайты с robots.txt, то есть наша программа-сборщик игнорирует протокол robots.txt. К этому решению мы пришли в результате долгих дискуссий.

Как вам, наверное, известно, robots.txt – стандартный протокол, используемый для того, чтобы поисковые машины не выгружали сайты или их части, которые веб-мастера сочли не подлежащими индексированию. Общепринятое правило Интернета – уважать протокол robots.txt. Но если речь идет о Национальной библиотеке и ее ответственности за сохранение всех общедоступных документов Интернета, нам приходится игнорировать robots.txt, чтобы важные документы были сохранены. Практика нашей работы показала, что следствием соблюдения robots.txt бывает потеря целых сайтов. Это не соответствовало бы Закону об обязательном экземпляре. Национальной библиотеке пришлось вступить в дискуссию с издателями сетевых

публикаций и доказывать, что игнорирование robots.txt в практике ее сбора законно и этим отличается от аналогичных действий других поисковых машин.

Мы выполняем сбор веб-страниц с помощью Heritrix – программы-сборщика, разработанной в рамках проекта Internet Archive, в котором участвовала и Национальная библиотека Норвегии. Первый цикл сбора сетевых публикаций мы начали в феврале текущего года. Мы рассчитывали делать полный сбор страниц домена “no” 4 раза в год. Этот расчет основывался на предварительной оценке, согласно которой каждый цикл сбора займет примерно три недели и будет собрано около 10 миллионов документов, то есть веб-страниц со своими URL. Вскоре мы убедились, что предварительная оценка абсолютно не соответствовала действительности. Когда закончились запланированные три недели сбора, программа-сборщик отобрала более 30 миллионов веб-страниц, и с каждым часом это количество возрастало.

Сбор веб-страниц всего норвежского домена обеспечивает сохранение «снимков», фиксирующих состояние норвежского Интернета в определенный момент времени. Для пользователей эти источники информации очень важны. Но важно также обеспечить полное сохранение отдельных документов. Например, все выпуски журналов, все отчеты одной серии, ежедневные номера газет. Поэтому важным дополнением тотальному сбору норвежских сайтов является селективный подход. Он будет реализован на втором этапе и явится второй целью сбора сетевых публикаций. Уже в 2005 году будет начато ежедневное сохранение выпусков газет, размещенных в Интернете. Мы также намереваемся начать сохранение всех сетевых периодических изданий Норвегии в 2005 году.

3. Международное сотрудничество

Национальная библиотека Норвегии принимает участие в различных международных проектах архивирования ресурсов Интернета. Библиотека – член International Preservation Consortium (Международный консорциум по проблемам сохранности), в котором мы сотрудничаем с другими национальными библиотеками, и американским Internet Archive (Интернет-архивом).

Кроме того, библиотека участвует в проекте Nordic Web Archive, форуме национальных библиотек скандинавских стран, посвященном координации деятельности по сбору и архивированию документов Интернета. Хотя название форума наводит на мысль, что у нас общий веб-архив, это не соответствует действительности. Скандинавские национальные библиотеки формируют отдельные веб-архивы. Но в рамках этого проекта разработаны программные средства распределенного доступа к веб-архивам: NWA Toolset.

4. Сохранение сетевых публикаций

Национальная библиотека Норвегии ориентируется на тысячелетний срок хранения. Это означает, что наша цель – обеспечить доступность информации о современных документах в любых форматах через тысячу лет. Реалистична ли эта цель, узнают только наши будущие пользователи, но, тем не менее, это наша цель, и мы планируем свою деятельность по сохранению в соответствии с ней. Сетевые публикации, собираемые Heritrix, хранятся в Электронном репозитории долгосрочного хранения Национальной библиотеки.

5. Доступ к сетевым публикациям

Каким образом обеспечивается доступ к сетевым публикациям, помещенным в Интернет-архив, для пользователей, занимающихся исследованиями и документоведением? Есть по крайней мере два аспекта, которые следует учесть при решении этой проблемы: метаданные и правовые вопросы.

5.1. Метаданные.

В текущем году Национальная библиотека примет решение, какие документы, помещаемые в сетевой архив, следует каталогизировать и включать в национальную библиографию. При этом будут учитываться следующие факторы: необходимость библиографического контроля электронных документов, вариант предоставления их в доступ и количество документов в архиве. Как бы то ни было, каталогизация сетевого архива, вероятно, будет выполнена лишь для небольшой части документов.

Осознавая, что, скорее всего, менее 1% документов, собранных в норвежском секторе Интернета, будут снабжены библиографическими записями, мы должны разработать другие средства доступа к веб-архиву. Решение таково: все документы архива будут индексироваться программой FAST. Это позволит пользователям вести поиск по полным текстам документов. Заиндексированные документы будут доступны через NWA Toolset. Эти программные средства обеспечат доступ к более чем 99% материалов архива.

Ниже приведены слайды, которые показывают, как выглядит сетевой архив для пользователя: показан интерфейс NWA Toolset, основным программным средством доступа к норвежскому Интернет-архиву.

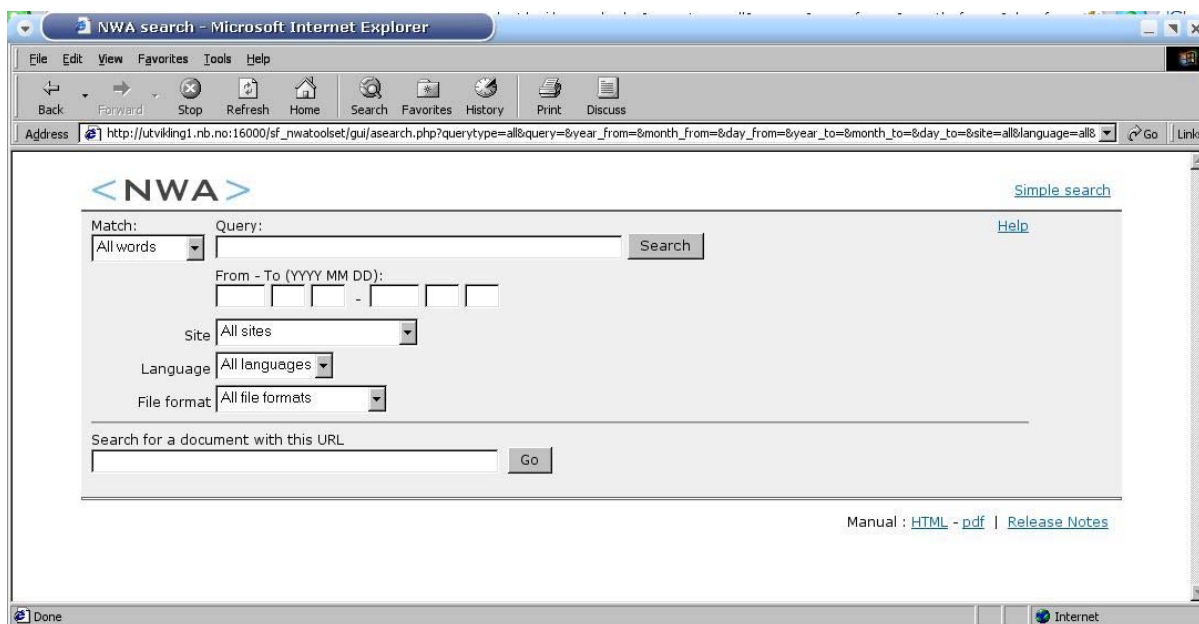


Рисунок 1. Поисковый интерфейс расширенного поиска NWA Toolset. В числе поисковых возможностей – поиск по полному тексту с использованием булевых операторов и поиск по определенному URL. Поиск можно ограничить датами, языками документа, форматами файлов и сайтами.

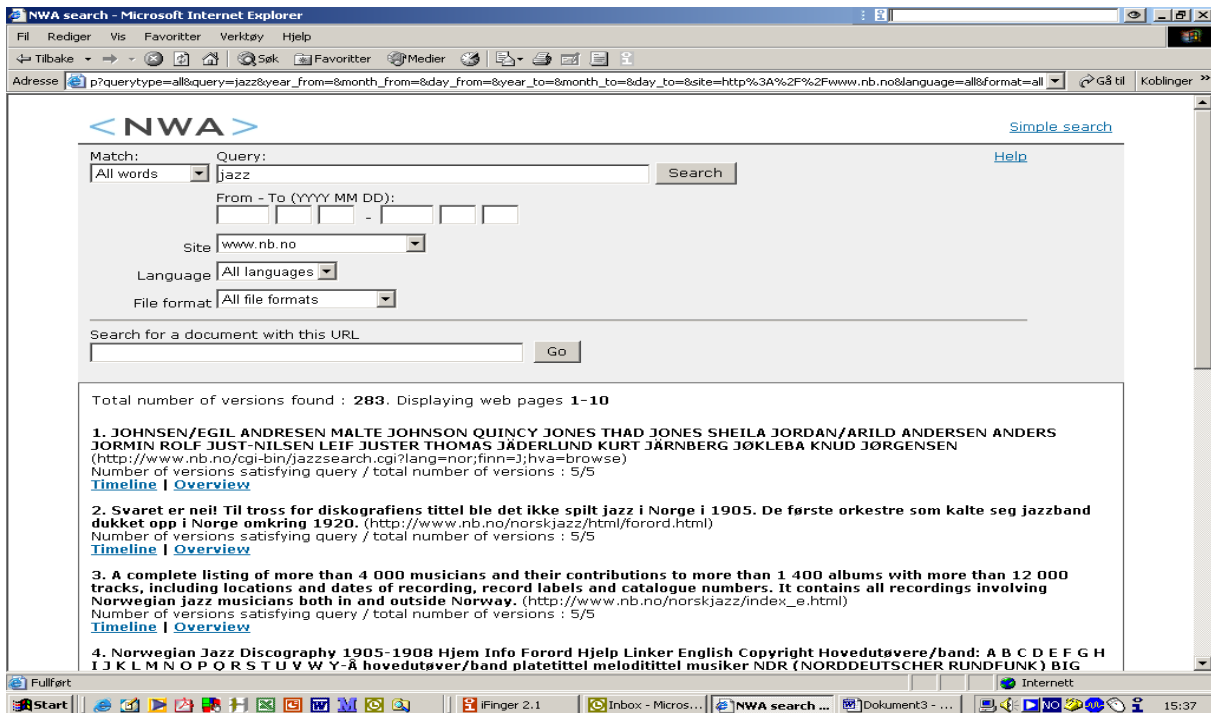


Рисунок 2 показывает пример поиска по ключевому слову «джаз», ограниченного сайтом Национальной библиотеки Норвегии. Результаты поиска можно выводить в двух вариантах: по истории сохранений на оси хронологии или в режиме обзора.

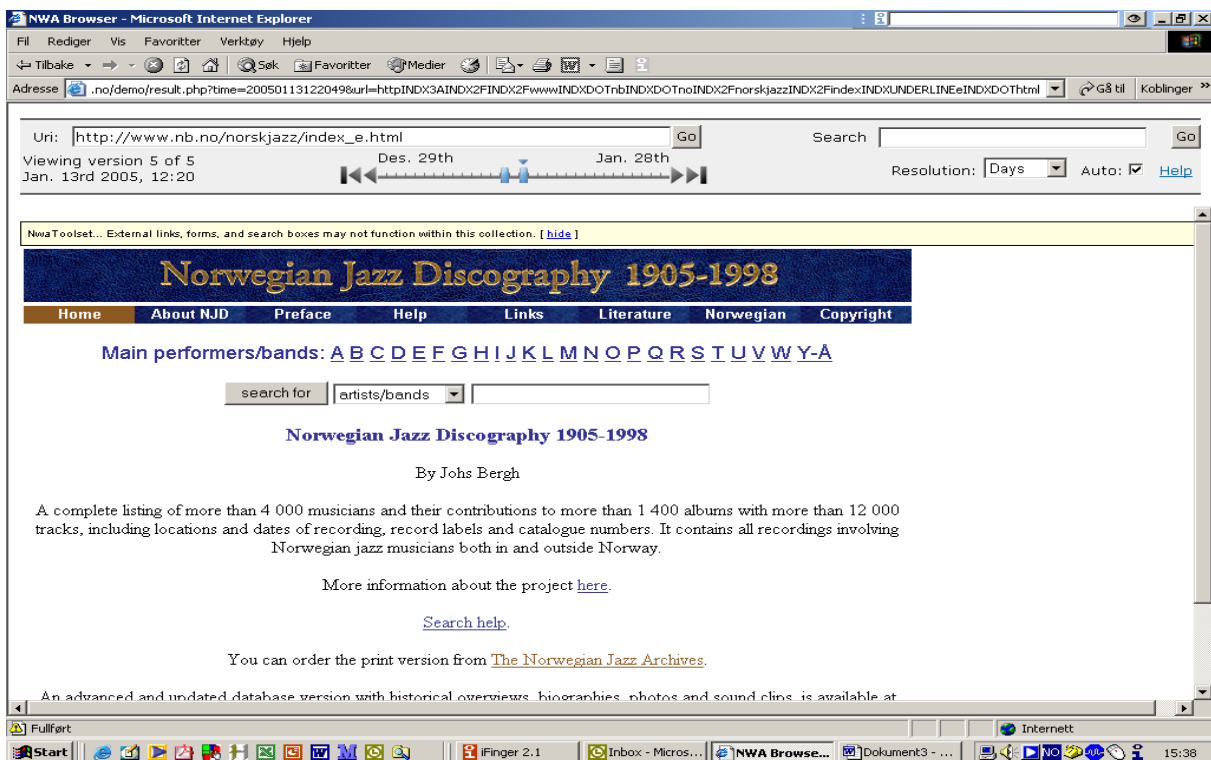


Рисунок 3 демонстрирует показ истории сохранения документа на оси хронологии. Вы видите, что предъявлена 5-я версия из 5 сохранений этого документа, выполненная 13 января 2005 года, в 12.20. Пользователь может проследить историю документа, передвигаясь по оси хронологии. Ось можно представить по годам, месяцам, дням, часам и минутам. Это дает пользователю возможность путешествовать во времени!

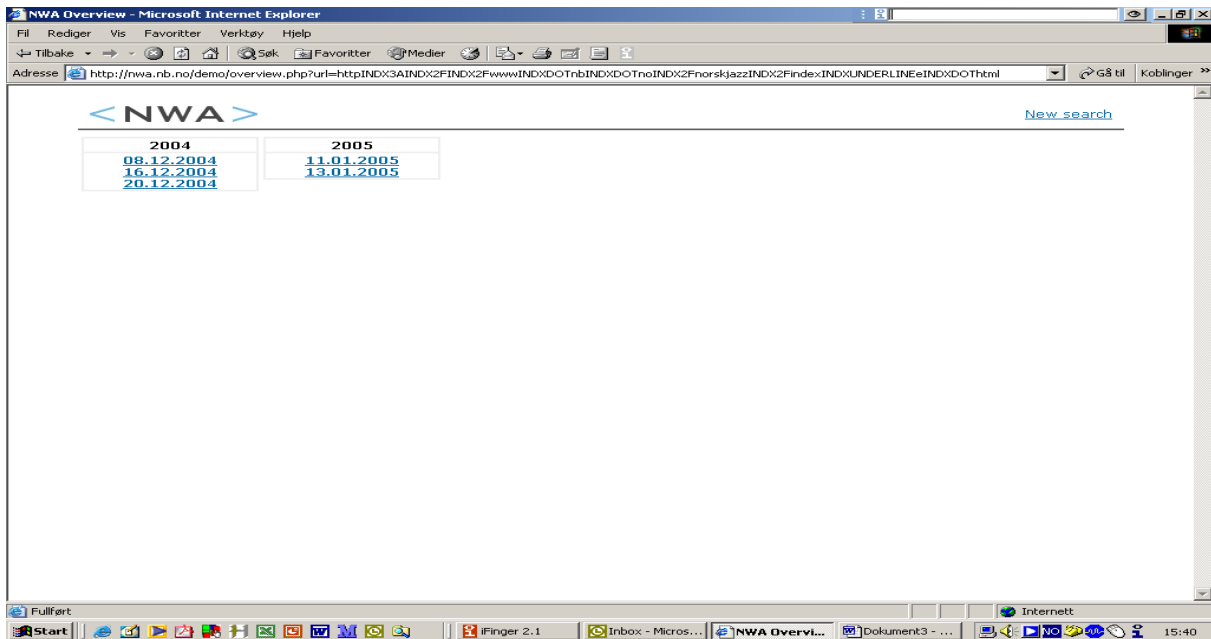


Рисунок 4 показывает представление истории документа через обзор количества его выгрузок (сохранений).

5.2. Правовые вопросы.

Сбор сайтов норвежского сектора Интернета позволяет Национальной библиотеке сохранить национальное цифровое наследие. Но доступ к нему пользователей ограничивается норвежским законодательством.

При решении вопроса о предоставлении доступа к материалам архива Национальная библиотека должна учитывать нормы законов об обязательном экземпляре, о персональных данных и об охране авторского права.

Закон об обязательном экземпляре устанавливает, что документы, собранные в соответствии с ним, могут предоставляться в пользование в целях исследовательской деятельности и документоведения. Это означает, что закон ограничивает доступ к документам для всех пользователей. Предоставление доступа к документу определяется целями, в которых он будет использован. Например, к архиву сетевых документов могут получить доступ исследователи, преподаватели, студенты и некоторые другие группы пользователей со специфическими интересами или хобби, такими как генеалогия и т.п.

Сетевые публикации, которые становятся доступны пользователям, могут содержать персональные данные и секретные персональные данные. Их использование регламентируется Законом о персональных данных. Этот закон помогает обрабатывать персональные данные в соответствии с фундаментальными принципами уважения к праву на частную жизнь, исключая возможность вторжения в нее. Возможность индексирования персональных данных для поиска в Интернет-архиве делает их гораздо более доступными, чем при публикации в традиционных документах. Поэтому Национальная библиотека Норвегии должна получить лицензию Норвежской инспекции по данным, чтобы осуществлять сбор сетевых публикаций. Кроме того, прежде чем давать доступ к архиву пользователям, библиотека должна получить еще одну лицензию на этот вид деятельности.

Согласно Закону об авторском праве Национальная библиотека не имеет права предоставлять доступ к архивным публикациям, если этот доступ ущемляет права авторов публикаций, имущественные и иные. Согласно действующему законодательству, доступ может предоставляться только с компьютеров, находящихся в библиотеке.

Обобщая вышесказанное, перечислим основные требования, которые должна выполнить библиотека согласно действующему законодательству, прежде чем предоставлять доступ пользователям к сетевым публикациям Интернет-архива:

- документы должны использоваться только в целях исследовательской деятельности и документоведения;
- должна быть получена генеральная лицензия Норвежской инспекции по данным, оформленная в соответствии с Законом о персональных данных;
- пользователи должны вести поиск в архиве только с компьютеров библиотеки.

Мы видим, что действующее законодательство дает конечным пользователям ограниченные возможности доступа к архиву электронных документов по сравнению с возможностями пользования печатными документами, поступившими в библиотеку по Закону об обязательном экземпляре.

Изменения, которые сейчас предлагается внести в норвежский Закон об авторском праве, включают разрешение предоставления электронных документов по запросам через межбиблиотечный абонемент. Это означает, что Национальная библиотека Норвегии сможет посылать копии электронных документов по запросам пользователей, обслуживаемых в других библиотеках. Эти библиотеки должны будут уничтожать электронный документ после получения.

Учитывая, что Норвегия – страна с большой протяженностью и неравномерной плотностью населения, мы считаем, что организация доступа к Интернет-архиву только в пределах Национальной библиотеки в Осло и Ране не является оптимальным решением.

Можно выделить четыре уровня доступа для пользователей:

- Доступ с компьютеров Национальной библиотеки. Согласно изменениям, вносимым в норвежский Закон об авторских правах, это будет разрешено.
- Доступ с компьютеров университетов, колледжей, публичных библиотек. Этот вариант противоречит действующему законодательству, но во многих отношениях был бы прекрасным решением проблемы доступа.
- Доступ с персональных компьютеров пользователей.
- Электронные документы можно выставлять на сайт Национальной библиотеки, например, как экспонаты выставки. На это понадобится оформить лицензию.

Доступ к Интернет-архиву, как ясно из вышесказанного, предоставляется только в целях исследовательской деятельности и документоведения. Это означает, что согласно нынешнему законодательству нельзя открыть доступ для любых пользователей. Однако есть возможность организовать свободный доступ к отдельным частям Интернет-архива. Национальная библиотека может оформить договоры или лицензии с держателями копирайта, которые позволят библиотеке открыть свободный доступ к сетевым публикациям, например, через организацию выставки на сайте библиотеки.

Изменения, предлагаемые для внесения в норвежский закон об авторском праве, предусматривают возможность подписания электронных лицензий между библиотеками и держателями копирайта. Это может помочь Национальной библиотеке сделать электронные документы доступными конечным пользователям, но это будет зависеть от условий оплаты таких лицензий.

Ближайшей целью Национальной библиотеки является организация доступа к архиву с компьютеров библиотеки. Это неизбежно ограничит доступ географическими условиями. Национальная библиотека рассматривает сейчас возможности расширения доступа в будущем.

5.3. К каким документам необходимо обеспечить доступ пользователей?

В принципе, все сетевые публикации, помещенные в архив на законных основаниях, следует предоставлять в доступ конечным пользователям в целях исследовательской деятельности и документоведения. В случае, если документ защищен копирайтом, доступ конкретного пользователя будет ограничен частью архива, релевантной его потребностям. Для документов, не подпадающих под действие Закона об авторском праве, таких как законодательные акты, административные документы и другие документы властных структур, никаких ограничений доступа не предусмотрено.

6. Заключение

Со временем Норвежский Интернет-архив будет содержать миллиарды веб-страниц. Национальная библиотека должна обеспечить пользователю навигацию по ним. Проблемы сбора, сохранения и доступа к веб-документам очень сложны, но Национальная библиотека верит, что она обладает соответствующей инфраструктурой, службами и организационными ресурсами, необходимыми для их успешного решения.

С точки зрения технических аспектов, сетевые публикации легко сделать доступными для пользователей. Однако доступ ограничивается действующим законодательством. Для Национальной библиотеки очень важно, чтобы использование Интернет-архива проходило в соответствии с законами об обязательном экземпляре, о персональных данных и об авторском праве.

Но эти законодательные ограничения не должны препятствовать доступу пользователей к сетевым публикациям, являющимися свидетельствами норвежской культурной и общественной жизни, если они используются как источники информации в целях исследовательской деятельности и документоведения.