Система обнаружения несанкционированного полного копирования документов электронных библиотек

Евгений Ивашко, Наталия Никитина

Институт прикладных математических исследований
КарНЦ РАН

RCDL–2011
19–22 октября 2011 г.

Определения

Определение

Электронная библиотека

Определения

Определение

Электронная библиотека — это упорядоченная коллекция разнородных электронных документов (в том числе книг, статей, изображений, аудио/видео файлов), снабженных средствами навигации и поиска.

Определения

Определение

Электронная библиотека — это упорядоченная коллекция разнородных электронных документов (в том числе книг, статей, изображений, аудио/видео файлов), снабженных средствами навигации и поиска.

Примеры электронных библиотек

- Библиотека Максима Мошкова

Определения

Определение

Электронная библиотека — это упорядоченная коллекция разнородных электронных документов (в том числе книг, статей, изображений, аудио/видео файлов), снабженных средствами навигации и поиска.

Примеры электронных библиотек

- Библиотека Максима Мошкова
- Либрусек / Флибуста

Определения

Определение

Электронная библиотека — это упорядоченная коллекция разнородных электронных документов (в том числе книг, статей, изображений, аудио/видео файлов), снабженных средствами навигации и поиска.

Примеры электронных библиотек

- Библиотека Максима Мошкова
- Либрусек / Флибуста
- eLIBRARY.RU

Определения

Определение

Электронная библиотека — это упорядоченная коллекция разнородных электронных документов (в том числе книг, статей, изображений, аудио/видео файлов), снабженных средствами навигации и поиска.

Примеры электронных библиотек

- Библиотека Максима Мошкова
- Либрусек / Флибуста
- eLIBRARY.RU
- Elsevier, JSTOR, ...

Определения

Определение

Электронная библиотека — это упорядоченная коллекция разнородных электронных документов (в том числе книг, статей, изображений, аудио/видео файлов), снабженных средствами навигации и поиска.

Примеры электронных библиотек

- Библиотека Максима Мошкова
- Либрусек / Флибуста
- eLIBRARY.RU
- Elsevier, JSTOR, ...
- Youtube.com, ...

Определение НПК

Информационная безопасность — это обеспечение

Определение НПК

Информационная безопасность — это обеспечение

- целостности

Определение НПК

Информационная безопасность — это обеспечение

- целостности
- доступности

Определение НПК

Информационная безопасность — это обеспечение

- целостности
- доступности
- конфиденциальности

Определение НПК

Информационная безопасность — это обеспечение

- целостности
- доступности
- конфиденциальности
- защиты от **несанкционированного полного копирования**

Определение НПК

Информационная безопасность — это обеспечение

- целостности
- доступности
- конфиденциальности
- защиты от **несанкционированного полного копирования**

Определение

Несанкционированное полное копирование (НПК) — получение копий большой части документов библиотеки без согласия владельцев

Угрозы НПК

Угрозы несанкционированного полного копирования:

Угрозы НПК

Угрозы несанкционированного полного копирования:

- нарушение авторских прав

Угрозы НПК

Угрозы несанкционированного полного копирования:

- нарушение авторских прав
- утрата уникальности ресурса

Угрозы НПК

Угрозы несанкционированного полного копирования:

- нарушение авторских прав
- утрата уникальности ресурса
- фишинг (подложный web-сайт)

Угрозы НПК

Угрозы несанкционированного полного копирования:

- нарушение авторских прав
- утрата уникальности ресурса
- фишинг (подложный web-сайт)
- распространение вредоносного ПО

Пример 1

"Условия пользования научной библиотекой РФФИ"

"Копирование целых томов или выпусков журналов, а также использование для этих целей автоматических поисковых систем (роботов) категорически запрещено. Организации, нарушившие это правило, лишаются доступа в библиотеку на год, а при повторном нарушении — навсегда."

Пример 2

ЛИЦЕНЗИОННОЕ СОГЛАШЕНИЕ ДЛЯ ОРГАНИЗАЦИИ
Научная электронная библиотека eLIBRARY.RU

Лицензионные материалы, представленные на данном Сайте, не могут прямо или косвенно использоваться для: **значительного по масштабам или систематического копирования или воспроизведения** (в частности, запрещается копирование целиком выпусков журналов); ...
В случае нарушения условий настоящего Лицензионного соглашения со стороны Организации или ее Пользователей Оператор имеет право временно блокировать доступ всех или части Пользователей данной Организации к Лицензионным материалам. При повторных нарушениях Оператор вправе в одностороннем порядке прекратить действие настоящего Лицензионного соглашения.

Законодательство

Законодательство

- **Россия**: база данных электронной библиотеки охраняется *авторским правом* как результат творческой деятельности по подбору или расположению включенных в нее материалов.

Законодательство

- **Россия**: база данных электронной библиотеки охраняется *авторским правом* как результат творческой деятельности по подбору или расположению включенных в нее материалов. База данных может охраняться и *смежным правом*, которое признается за ее изготовителем независимо от наличия авторских прав.

Законодательство

- **Россия**: база данных электронной библиотеки охраняется *авторским правом* как результат творческой деятельности по подбору или расположению включенных в нее материалов. База данных может охраняться и *смежным правом*, которое признается за ее изготовителем независимо от наличия авторских прав. *Исключительное право* распространяется на базы данных, создание которых требует существенных финансовых, материальных, организационных или иных затрат.

Законодательство

- **Россия**: база данных электронной библиотеки охраняется *авторским правом* как результат творческой деятельности по подбору или расположению включенных в нее материалов. База данных может охраняться и *смежным правом*, которое признается за ее изготовителем независимо от наличия авторских прав. *Исключительное право* распространяется на базы данных, создание которых требует существенных финансовых, материальных, организационных или иных затрат.
- В **Европейском союзе** существует специальный правовой режим *sui generis*: производитель базы данных вправе запрещать извлечение и/или повторное использование существенной части содержимого базы данных.

Законодательство

- **Россия**: база данных электронной библиотеки охраняется *авторским правом* как результат творческой деятельности по подбору или расположению включенных в нее материалов. База данных может охраняться и *смежным правом*, которое признается за ее изготовителем независимо от наличия авторских прав. *Исключительное право* распространяется на базы данных, создание которых требует существенных финансовых, материальных, организационных или иных затрат.
- В **Европейском союзе** существует специальный правовой режим *sui generis*: производитель базы данных вправе запрещать извлечение и/или повторное использование существенной части содержимого базы данных.
- В **США** ограничения на копирование значимой части базы данных зачастую прописываются в *контракте* и защищаются в рамках *контрактного права*.

Судебные процессы

Судебные процессы

- три ведущих американских издательства подали в суд на Georgia State University за предоставление возможности студентам университета *"systematic, widespread, and unauthorized copying and distribution of a vast amount of copyrighted works"* ("Copyright Lawsuit against Georgia State University", 2008);

Судебные процессы

- три ведущих американских издательства подали в суд на Georgia State University за предоставление возможности студентам университета *"systematic, widespread, and unauthorized copying and distribution of a vast amount of copyrighted works"* ("Copyright Lawsuit against Georgia State University", 2008);
- июль 2011 г.: Aaron Swartz скачал 4.8 миллионов статей и других документов (почти полную библиотеку) с сервиса JSTOR, специализирующегося на распространении в электронном виде научных статей.

Методы защиты от НПК

Как защищаются от несанкционированного полного копирования?

Методы защиты от НПК

Как защищаются от несанкционированного полного копирования?

- ограничение круга пользователей электронной библиотеки

Методы защиты от НПК

Как защищаются от несанкционированного полного копирования?

- ограничение круга пользователей электронной библиотеки
- платный доступ

Методы защиты от НПК

Как защищаются от несанкционированного полного копирования?

- ограничение круга пользователей электронной библиотеки
- платный доступ
- юридические ограничения

Методы защиты от НПК

Как защищаются от несанкционированного полного копирования?

- ограничение круга пользователей электронной библиотеки
- платный доступ
- юридические ограничения
- технические ограничения (ограничение интенсивности и количества загружаемых документов, CAPTCHA)

Технические ограничения доступа к ЭБ

Ограничение интенсивности/количества обращений к электронным документам:

Технические ограничения доступа к ЭБ

Ограничение интенсивности/количества обращений к электронным документам:

- какой порог установить?

Технические ограничения доступа к ЭБ

Ограничение интенсивности/количества обращений к электронным документам:

- какой порог установить?
- будет ли эффективное ограничение зависеть

Технические ограничения доступа к ЭБ

Ограничение интенсивности/количества обращений к электронным документам:

- какой порог установить?
- будет ли эффективное ограничение зависеть
  - от электронной библиотеки?

Технические ограничения доступа к ЭБ

Ограничение интенсивности/количества обращений к электронным документам:

- какой порог установить?
- будет ли эффективное ограничение зависеть
  - от электронной библиотеки?
  - от тематики контента?

Технические ограничения доступа к ЭБ

Ограничение интенсивности/количества обращений к электронным документам:

- какой порог установить?
- будет ли эффективное ограничение зависеть
  - от электронной библиотеки?
  - от тематики контента?
  - от количества документов ЭБ?

Технические ограничения доступа к ЭБ

Ограничение интенсивности/количества обращений к электронным документам:

- какой порог установить?
- будет ли эффективное ограничение зависеть
  - от электронной библиотеки?
  - от тематики контента?
  - от количества документов ЭБ?

...

Технические ограничения доступа к ЭБ

Ограничение интенсивности/количества обращений к электронным документам:

- какой порог установить?
- будет ли эффективное ограничение зависеть
  - от электронной библиотеки?
  - от тематики контента?
  - от количества документов ЭБ?

...

Недостаток подхода заключается в чрезмерной простоте: зная допустимое число скачиваемых документов, можно точно указать какое время необходимо для полного копирования всех документов ЭБ.

Цель защиты от НПК

Цель работы

Цель защиты от несанкционированного полного копирования — сохранение открытости и доступности публичных электронных библиотек.

Цель работы

Цель защиты от несанкционированного полного копирования — сохранение открытости и доступности публичных электронных библиотек.

Цель работы — создание *интеллектуальной* системы защиты от несанкционированного полного копирования.

Цель работы

Цель защиты от несанкционированного полного копирования — сохранение открытости и доступности публичных электронных библиотек.

Цель работы — создание *интеллектуальной* системы защиты от несанкционированного полного копирования.

Метод — модифицированный аномальный статистический подход в обнаружении вторжений.

Аномальный подход в обнаружении вторжений

Аномальный подход хорошо зарекомендовал себя в системах обнаружения вторжений.

Аномальный подход в обнаружении вторжений

Аномальный подход хорошо зарекомендовал себя в системах обнаружения вторжений.
Основная идея основана на двух следующих гипотезах (получивших подтверждение на практике):

Аномальный подход в обнаружении вторжений

Аномальный подход хорошо зарекомендовал себя в системах обнаружения вторжений.
Основная идея основана на двух следующих гипотезах (получивших подтверждение на практике):

1. количество злоумышленников составляет не более долей процента от общего числа пользователей;

Аномальный подход в обнаружении вторжений

Аномальный подход хорошо зарекомендовал себя в системах обнаружения вторжений.
Основная идея основана на двух следующих гипотезах (получивших подтверждение на практике):

1. количество злоумышленников составляет не более долей процента от общего числа пользователей;
2. действия злоумышленника значимо отличаются от действий обычных пользователей.

Аномальный подход в обнаружении вторжений

Аномальный подход хорошо зарекомендовал себя в системах обнаружения вторжений.
Основная идея основана на двух следующих гипотезах (получивших подтверждение на практике):

1. количество злоумышленников составляет не более долей процента от общего числа пользователей;
2. действия злоумышленника значимо отличаются от действий обычных пользователей.

Следовательно, на основании действий обычных пользователей можно построить шаблон "нормального" поведения; тогда значимое отклонение от этого шаблона будет свидетельствовать об аномальном поведении — т.е. обнаружении злоумышленника.

Аномальный подход в обнаружении НПК

Гипотеза

Аномальный подход в обнаружении НПК

Гипотеза

- электронные документы, скачиваемые обычными пользователями, семантически (по смыслу) связаны между собой;

Аномальный подход в обнаружении НПК

Гипотеза

- электронные документы, скачиваемые обычными пользователями, семантически (по смыслу) связаны между собой;
- документы, скачиваемые при полном несанкционированном копировании, имеют слабую семантическую (смысловую) связь.

Аномальный подход в обнаружении НПК

Гипотеза

- электронные документы, скачиваемые обычными пользователями, семантически (по смыслу) связаны между собой;
- документы, скачиваемые при полном несанкционированном копировании, имеют слабую семантическую (смысловую) связь.

Обоснование

Обращаясь за информацией в ЭБ, пользователь хочет получить ответ на определенный вопрос и/или подобрать список документов по определенной тематике. Даже с учетом смежных тем и варьирования тематики поиска в зависимости от вновь получаемой информации, все документы, к которым обращается пользователь, как правило, будут взаимосвязаны.

Семантические связи между документами

Семантические связи между документами

Применимость аномального подхода

Имея возможность определять, связаны ли между собой документы из определенного набора, можно выявлять и попытки несанкционированного полного копирования.

Семантические связи между документами

Применимость аномального подхода

Имея возможность определять, связаны ли между собой документы из определенного набора, можно выявлять и попытки несанкционированного полного копирования.

Для выявления связей между документами используется "поведенческий" подход — семантические связи между документами определяются на основе анализа поведения обычных пользователей электронной библиотеки.

Этапы построения системы защиты

При разработке системы, реализующей аномальный подход в обнаружении вторжений, возникают следующие основные задачи:

Этапы построения системы защиты

При разработке системы, реализующей аномальный подход в обнаружении вторжений, возникают следующие основные задачи:

1. построение шаблона «нормального» поведения пользователя;

Этапы построения системы защиты

При разработке системы, реализующей аномальный подход в обнаружении вторжений, возникают следующие основные задачи:

1. построение шаблона «нормального» поведения пользователя;
   - на основе исходных "нормальных" данных

Этапы построения системы защиты

При разработке системы, реализующей аномальный подход в обнаружении вторжений, возникают следующие основные задачи:

1. построение шаблона «нормального» поведения пользователя;
   - на основе исходных "нормальных" данных
   - на основе онтологий

Этапы построения системы защиты

При разработке системы, реализующей аномальный подход в обнаружении вторжений, возникают следующие основные задачи:

1. построение шаблона «нормального» поведения пользователя;
   - на основе исходных "нормальных" данных
   - на основе онтологий
2. разработка классификатора, позволяющего отличить "нормальную" последовательность действий от аномальной;

Этапы построения системы защиты

При разработке системы, реализующей аномальный подход в обнаружении вторжений, возникают следующие основные задачи:

1. построение шаблона «нормального» поведения пользователя;
   - на основе исходных "нормальных" данных
   - на основе онтологий
2. разработка классификатора, позволяющего отличить "нормальную" последовательность действий от аномальной;
3. определение граничных значений характеристик классификатора

Этапы построения системы защиты

При разработке системы, реализующей аномальный подход в обнаружении вторжений, возникают следующие основные задачи:

1. построение шаблона «нормального» поведения пользователя;
   - на основе исходных "нормальных" данных
   - на основе онтологий
2. разработка классификатора, позволяющего отличить "нормальную" последовательность действий от аномальной;
3. определение граничных значений характеристик классификатора
   - снижение числа пропусков атак

Этапы построения системы защиты

При разработке системы, реализующей аномальный подход в обнаружении вторжений, возникают следующие основные задачи:

1. построение шаблона «нормального» поведения пользователя;
   - на основе исходных "нормальных" данных
   - на основе онтологий
2. разработка классификатора, позволяющего отличить "нормальную" последовательность действий от аномальной;
3. определение граничных значений характеристик классификатора
   - снижение числа пропусков атак
   - снижение числа ложных срабатываний

Этапы построения системы защиты

При разработке системы, реализующей аномальный подход в обнаружении вторжений, возникают следующие основные задачи:

1. построение шаблона «нормального» поведения пользователя;
   - на основе исходных "нормальных" данных
   - на основе онтологий
2. разработка классификатора, позволяющего отличить "нормальную" последовательность действий от аномальной;
3. определение граничных значений характеристик классификатора
   - снижение числа пропусков атак
   - снижение числа ложных срабатываний
   - уменьшение среднего времени до обнаружения атаки

Этапы построения системы защиты

При разработке системы, реализующей аномальный подход в обнаружении вторжений, возникают следующие основные задачи:

1. построение шаблона «нормального» поведения пользователя;
   - на основе исходных "нормальных" данных
   - на основе онтологий
2. разработка классификатора, позволяющего отличить "нормальную" последовательность действий от аномальной;
3. определение граничных значений характеристик классификатора
   - снижение числа пропусков атак
   - снижение числа ложных срабатываний
   - уменьшение среднего времени до обнаружения атаки
4. периодическое обновление шаблонов «нормального» поведения.

Исходные данные

Для построения системы обнаружения несанкционированного полного копирования необходимы следующие наборы исходных данных:

Исходные данные

Для построения системы обнаружения несанкционированного полного копирования необходимы следующие наборы исходных данных:

- тренировочный набор заведомо безопасных для ЭБ действий пользователя (на его основе строится шаблон "нормального" поведения);

Исходные данные

Для построения системы обнаружения несанкционированного полного копирования необходимы следующие наборы исходных данных:

- тренировочный набор заведомо безопасных для ЭБ действий пользователя (на его основе строится шаблон "нормального" поведения);
- тестовый набор заведомо безопасных действий (используется для оценки числа ошибок типа false positive);

Исходные данные

Для построения системы обнаружения несанкционированного полного копирования необходимы следующие наборы исходных данных:

- тренировочный набор заведомо безопасных для ЭБ действий пользователя (на его основе строится шаблон "нормального" поведения);
- тестовый набор заведомо безопасных действий (используется для оценки числа ошибок типа false positive);
- тестовый набор заведомо аномальных действий (используется для оценки числа ошибок типа false negative).

Исходные данные

Для построения системы обнаружения несанкционированного полного копирования необходимы следующие наборы исходных данных:

- тренировочный набор заведомо безопасных для ЭБ действий пользователя (на его основе строится шаблон "нормального" поведения);
- тестовый набор заведомо безопасных действий (используется для оценки числа ошибок типа false positive);
- тестовый набор заведомо аномальных действий (используется для оценки числа ошибок типа false negative).

Наборы исходных данных могут быть получены из зафиксированной за определенный срок истории посещений ЭБ пользователями. С помощью этих наборов данных подбираются параметры шаблона "нормального" поведения и классификатора, чтобы минимизировать число ошибок классификации.

Алгоритм

Алгоритм

Этапы построения системы обнаружения несанкционированного полного копирования:

Алгоритм

Этапы построения системы обнаружения несанкционированного полного копирования:

1. с помощью тренировочного набора данных строится Марковская цепь, представляющая собой шаблон "нормального" поведения пользователя;

Алгоритм

Этапы построения системы обнаружения несанкционированного полного копирования:

1. с помощью тренировочного набора данных строится Марковская цепь, представляющая собой шаблон "нормального" поведения пользователя;
2. на основе шаблона "нормального" поведения строится классификатор поведения пользователя;

Алгоритм

Этапы построения системы обнаружения несанкционированного полного копирования:

1. с помощью тренировочного набора данных строится Марковская цепь, представляющая собой шаблон "нормального" поведения пользователя;
2. на основе шаблона "нормального" поведения строится классификатор поведения пользователя;
3. действия пользователя электронной библиотеки сравниваются с шаблоном "нормального" поведения — за отклонение от шаблона начисляется "штраф" (согласно классификатору);

Алгоритм

Этапы построения системы обнаружения несанкционированного полного копирования:

1. с помощью тренировочного набора данных строится Марковская цепь, представляющая собой шаблон "нормального" поведения пользователя;
2. на основе шаблона "нормального" поведения строится классификатор поведения пользователя;
3. действия пользователя электронной библиотеки сравниваются с шаблоном "нормального" поведения — за отклонение от шаблона начисляется "штраф" (согласно классификатору);
4. при большом числе "штрафов" последовательность действий пользователя считается аномальной.

Алгоритм

Этапы построения системы обнаружения несанкционированного полного копирования:

1. с помощью тренировочного набора данных строится Марковская цепь, представляющая собой шаблон "нормального" поведения пользователя;
2. на основе шаблона "нормального" поведения строится классификатор поведения пользователя;
3. действия пользователя электронной библиотеки сравниваются с шаблоном "нормального" поведения — за отклонение от шаблона начисляется "штраф" (согласно классификатору);
4. при большом числе "штрафов" последовательность действий пользователя считается аномальной.

Замечание

Разработанный подход не привязан к типу содержимого электронной библиотеки и легко может быть модернизирован для защиты коллекций аудио- и видео-записей (например, для таких сайтов как RuTube.ru, YouTube.com и др.). При этом шаблоны нормального поведения могут быть либо едиными для всех типов контента, либо строиться независимо.

Описание экспериментов

Описание экспериментов

Для оценки применимости подхода проведены эксперименты.

Описание экспериментов

Для оценки применимости подхода проведены эксперименты.

Объект экспериментов

- объект: Электронная библиотека Республики Карелия

Описание экспериментов

Для оценки применимости подхода проведены эксперименты.

Объект экспериментов

- объект: Электронная библиотека Республики Карелия
- исходные данные за период июнь 2007 г. - февраль 2009 г.

Описание экспериментов

Для оценки применимости подхода проведены эксперименты.

Объект экспериментов

- объект: Электронная библиотека Республики Карелия
- исходные данные за период июнь 2007 г. - февраль 2009 г.
- объем данных: >1000 документов, обращения зафиксированы к >700 документам

Описание экспериментов

Для оценки применимости подхода проведены эксперименты.

Объект экспериментов

- объект: Электронная библиотека Республики Карелия
- исходные данные за период июнь 2007 г. - февраль 2009 г.
- объем данных: >1000 документов, обращения зафиксированы к >700 документам

Описание исходных данных

- лог-файл с IP-адресами пользователей и названиями документов

Описание экспериментов

Для оценки применимости подхода проведены эксперименты.

Объект экспериментов

- объект: Электронная библиотека Республики Карелия
- исходные данные за период июнь 2007 г. - февраль 2009 г.
- объем данных: >1000 документов, обращения зафиксированы к >700 документам

Описание исходных данных

- лог-файл с IP-адресами пользователей и названиями документов
- отброшены потенциальные proxy-сервера

Описание экспериментов

Для оценки применимости подхода проведены эксперименты.

Объект экспериментов

- объект: Электронная библиотека Республики Карелия
- исходные данные за период июнь 2007 г. - февраль 2009 г.
- объем данных: >1000 документов, обращения зафиксированы к >700 документам

Описание исходных данных

- лог-файл с IP-адресами пользователей и названиями документов
- отброшены потенциальные proxy-сервера
- отброшены неинформативные сессии (<10 действий)

Описание экспериментов

Для оценки применимости подхода проведены эксперименты.

Объект экспериментов

- объект: Электронная библиотека Республики Карелия
- исходные данные за период июнь 2007 г. - февраль 2009 г.
- объем данных: >1000 документов, обращения зафиксированы к >700 документам

Описание исходных данных

- лог-файл с IP-адресами пользователей и названиями документов
- отброшены потенциальные proxy-сервера
- отброшены неинформативные сессии (<10 действий)
- сессия — обращения к ЭБ с одного IP-адреса

Описание экспериментов

Для оценки применимости подхода проведены эксперименты.

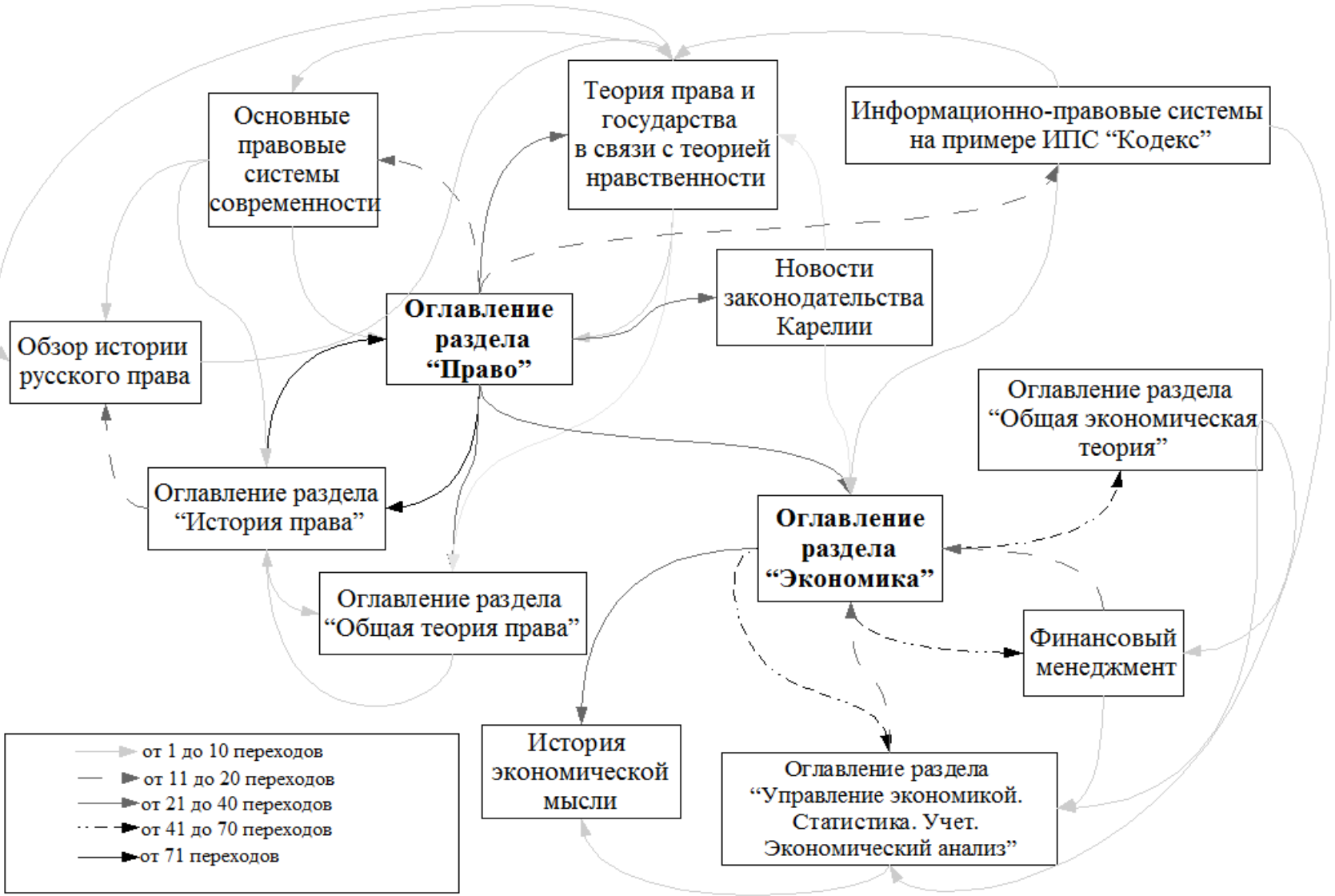
Объект экспериментов

- объект: Электронная библиотека Республики Карелия
- исходные данные за период июнь 2007 г. - февраль 2009 г.
- объем данных: >1000 документов, обращения зафиксированы к >700 документам

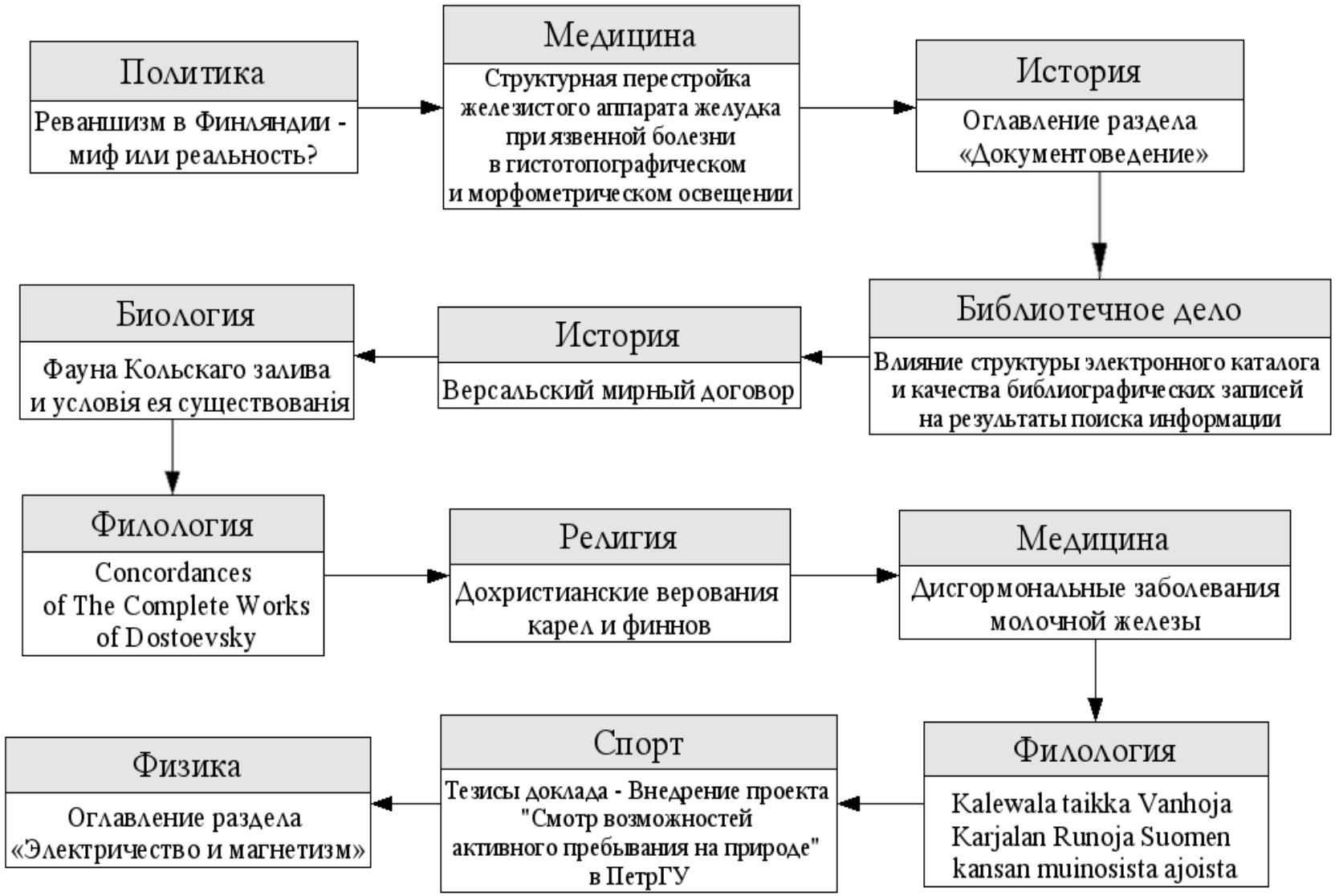
Описание исходных данных

- лог-файл с IP-адресами пользователей и названиями документов
- отброшены потенциальные proxy-сервера
- отброшены неинформативные сессии (<10 действий)
- сессия — обращения к ЭБ с одного IP-адреса
- лог-файл содержит 4718 сессий

Пример нормальной сессии



Пример аномальной сессии



Первые результаты

Первые результаты

1. разработана модель обнаружения полного несанкционированного копирования на основе модифицированного аномального статистического подхода к обнаружению вторжений;

Первые результаты

1. разработана модель обнаружения полного несанкционированного копирования на основе модифицированного аномального статистического подхода к обнаружению вторжений;
2. *разработан и отлажен прототип* системы защиты от несанкционированного полного копирования;

Первые результаты

1. разработана модель обнаружения полного несанкционированного копирования на основе модифицированного аномального статистического подхода к обнаружению вторжений;
2. *разработан и отлажен прототип* системы защиты от несанкционированного полного копирования;
3. проведены эксперименты, показавшие применимость подхода:

Первые результаты

1. разработана модель обнаружения полного несанкционированного копирования на основе модифицированного аномального статистического подхода к обнаружению вторжений;
2. *разработан и отлажен прототип* системы защиты от несанкционированного полного копирования;
3. проведены эксперименты, показавшие применимость подхода:
   - экспериментально установлено, что на основе анализа поведения пользователей *возможно автоматическое выявление семантических связей* между электронными документами;

Первые результаты

1. разработана модель обнаружения полного несанкционированного копирования на основе модифицированного аномального статистического подхода к обнаружению вторжений;
2. *разработан и отлажен прототип* системы защиты от несанкционированного полного копирования;
3. проведены эксперименты, показавшие применимость подхода:
   - экспериментально установлено, что на основе анализа поведения пользователей *возможно автоматическое выявление семантических связей* между электронными документами;
   - экспериментально установлено, что *возможно автоматическое выявление последовательностей обращений, противоречащих семантическим связям* между электронными документами.

Основные задачи

- Получение доступа к лог-файлам ЭБ, фиксирующей необходимый набор исходных данных;

Основные задачи

- Получение доступа к лог-файлам ЭБ, фиксирующей необходимый набор исходных данных;
  - увеличение объемов исходных данных;

Основные задачи

- Получение доступа к лог-файлам ЭБ, фиксирующей необходимый набор исходных данных;
  - увеличение объемов исходных данных;
  - устранение недостатков исходных данных первых экспериментов.

Основные задачи

- Получение доступа к лог-файлам ЭБ, фиксирующей необходимый набор исходных данных;
  - увеличение объемов исходных данных;
  - устранение недостатков исходных данных первых экспериментов.
- Определение параметров качественного классификатора:

Основные задачи

- Получение доступа к лог-файлам ЭБ, фиксирующей необходимый набор исходных данных;
  - увеличение объемов исходных данных;
  - устранение недостатков исходных данных первых экспериментов.
- Определение параметров качественного классификатора:
  - снижение числа ошибок типа false positive (”нормальные” действия классифицируются как аномальные);

Основные задачи

- Получение доступа к лог-файлам ЭБ, фиксирующей необходимый набор исходных данных;
  - увеличение объемов исходных данных;
  - устранение недостатков исходных данных первых экспериментов.
- Определение параметров качественного классификатора:
  - снижение числа ошибок типа false positive (”нормальные” действия классифицируются как аномальные);
  - снижение числа ошибок типа false negative (аномальные действия классифицируются как ”нормальные”);

Основные задачи

- Получение доступа к лог-файлам ЭБ, фиксирующей необходимый набор исходных данных;
  - увеличение объемов исходных данных;
  - устранение недостатков исходных данных первых экспериментов.
- Определение параметров качественного классификатора:
  - снижение числа ошибок типа false positive (”нормальные” действия классифицируются как аномальные);
  - снижение числа ошибок типа false negative (аномальные действия классифицируются как ”нормальные”);
  - уменьшение времени обнаружения.

Основные задачи

- Получение доступа к лог-файлам ЭБ, фиксирующей необходимый набор исходных данных;
  - увеличение объемов исходных данных;
  - устранение недостатков исходных данных первых экспериментов.
- Определение параметров качественного классификатора:
  - снижение числа ошибок типа false positive (”нормальные” действия классифицируются как аномальные);
  - снижение числа ошибок типа false negative (аномальные действия классифицируются как ”нормальные”);
  - уменьшение времени обнаружения.

Исходные данные

Исходные данные

Источник

Исходными данными для проведения экспериментов послужили лог-файлы доступа к документам *ЭБ Республики Карелия*, собранные за период с сентября 2004 г. по январь 2011 г.

Исходные данные

Источник

Исходными данными для проведения экспериментов послужили лог-файлы доступа к документам *ЭБ Республики Карелия*, собранные за период с сентября 2004 г. по январь 2011 г.

Для проведения экспериментов были отобраны сессии работы зарегистрированных пользователей, содержащие от 5 до 50 обращений к документам ЭБ. Сессией работы считалась последовательность обращений к цифровым документам, в которой время, прошедшее между двумя обращениями, не превышало 12 часов.

Исходные данные

Источник

Исходными данными для проведения экспериментов послужили лог-файлы доступа к документам *ЭБ Республики Карелия*, собранные за период с сентября 2004 г. по январь 2011 г.

Для проведения экспериментов были отобраны сессии работы зарегистрированных пользователей, содержащие от 5 до 50 обращений к документам ЭБ. Сессией работы считалась последовательность обращений к цифровым документам, в которой время, прошедшее между двумя обращениями, не превышало 12 часов.

Всего было получено 10393 сессии, принадлежащие 5561 пользователю и содержащие обращения к 2071 уникальному документу. Общее количество обращений к документам в выделенных сессиях составило 109847.

Нормальные и аномальные сессии

Нормальные и аномальные сессии

Для проведения экспериментов был создан набор заведомо аномальных псевдосессий, содержащих переходы между документами из различных тематических разделов библиотеки. Эти данные использовались для оценки эффективности системы обнаружения несанкционированного полного копирования документов.

Нормальные и аномальные сессии

Для проведения экспериментов был создан набор заведомо аномальных псевдосессий, содержащих переходы между документами из различных тематических разделов библиотеки. Эти данные использовались для оценки эффективности системы обнаружения несанкционированного полного копирования документов.

Все данные, полученные из Электронной библиотеки Республики Карелия, считались заведомо нормальными и использовались, во-первых, для создания шаблона «нормального» поведения пользователя, а во-вторых, для оценки числа ложных срабатываний.

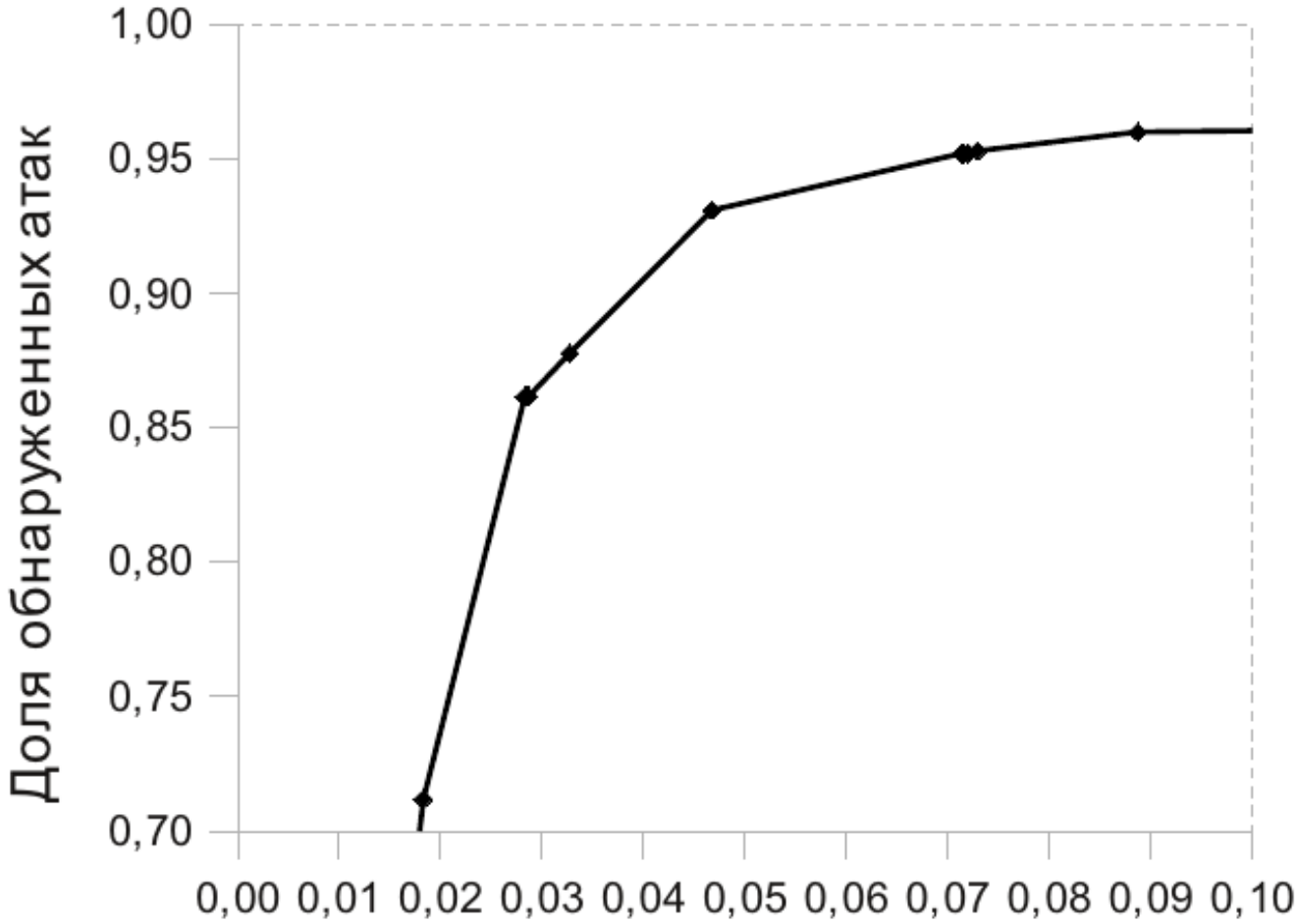


Рис.: Соотношение числа обнаруженных атак и ложных срабатываний

Что было сделано

Основные полученные результаты:

Что было сделано

Основные полученные результаты:

1. на основе лог-файлов доступа к документам Электронной Библиотеки Республики Карелия (www.elibrary.karelia.ru) за весь период существования библиотеки, проведена серия экспериментов, которая позволила:

Что было сделано

Основные полученные результаты:

1. на основе лог-файлов доступа к документам Электронной Библиотеки Республики Карелия (www.elibrary.karelia.ru) за весь период существования библиотеки, проведена серия экспериментов, которая позволила:
   - определить класс классификаторов поведения, позволяющих решить задачу обнаружения несанкционированного полного копирования документов электронной библиотеки;

Что было сделано

Основные полученные результаты:

1. на основе лог-файлов доступа к документам Электронной Библиотеки Республики Карелия (www.elibrary.karelia.ru) за весь период существования библиотеки, проведена серия экспериментов, которая позволила:
   - определить класс классификаторов поведения, позволяющих решить задачу обнаружения несанкционированного полного копирования документов электронной библиотеки;
   - оценить практическую и пиковую ресурсоемкость и времяемкость построения модели.

Что было сделано

Основные полученные результаты:

1. на основе лог-файлов доступа к документам Электронной Библиотеки Республики Карелия (www.elibrary.karelia.ru) за весь период существования библиотеки, проведена серия экспериментов, которая позволила:
   - определить класс классификаторов поведения, позволяющих решить задачу обнаружения несанкционированного полного копирования документов электронной библиотеки;
   - оценить практическую и пиковую ресурсоемкость и времяемкость построения модели.
2. достигнута предварительная договоренность о тестовой эксплуатации прототипа разрабатываемой системы в электронной библиотечной системе Электронной Библиотеки Республики Карелия.

Перспективы

Перспективы

На следующих этапах разработки запланированы:

Перспективы

На следующих этапах разработки запланированы:

- проведение экспериментов на данных ЭБ eLIBRARY.RU;

Перспективы

На следующих этапах разработки запланированы:

- проведение экспериментов на данных ЭБ eLIBRARY.RU;
- тестовая эксплуатация в ЭБ Республики Карелия;

Перспективы

На следующих этапах разработки запланированы:

- проведение экспериментов на данных ЭБ eLIBRARY.RU;
- тестовая эксплуатация в ЭБ Республики Карелия;
- разработка программного пакета, предназначенного для простого развертывания системы защиты в электронной библиотечной системе.

Спасибо за внимание!