Методы автоматизированного порождения поисковых эвристик по предметной области «информационная безопасность»

Тян Алексей

20 июня 2013 г.

Цель

На базе метапиосковой системы «виртуальный каталог» реализовать поиск Интернет ресурсов определённого типа для любой выбранной рубрики.

Задачи

1. Изучение основных принципов работы виртуального каталога.
2. Исследование методов автоматического порождения эвристик для виртуального каталога.
3. Реализация возможности порождения эвристик для пар рубрика-тип ресурса.
4. Порождение эвристик.
5. Создание рубрикатора для предметной области «информационная безопасность».

Определения

Релевантность - это соответствие полученной в результате поиска информации информационному запросу.

Пертинентность - это соответствие полученной в результате поиска информации информационной потребности.

Точность - это отношение числа найденных релевантных ресурсов к общему количеству найденных ресурсов.

Полнота - это отношение числа найденных релевантных ресурсов к общему количеству релевантных ресурсов в сети.

Поиск (формализация)

1. Потребность пользователя в информации.
2. Формализованный запрос. Представленный в виде с которым работает поисковая система, и отображающий потребность пользователя в информации.
3. Ответ на запрос.

Поисковые системы

Плюсы

1. Высокая релевантность.
2. Полнота.
3. Количество ресурсов, по которым ведётся поиск.
4. Точность.

Минусы

Необходимо выразить информационную потребность в виде 2-3 ключевых слов.

Интернет-каталоги

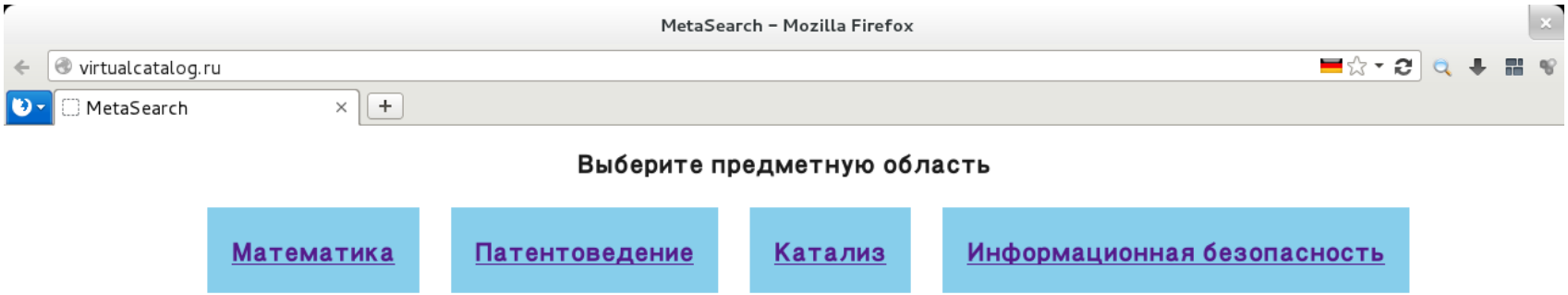
Плюсы

Интерфейс позволяет точно выразить информационную потребность.

Минусы

1. Низкая полнота.
2. В сравнении с поисковыми системами, мало ресурсов по которым ведётся поиск.
3. Отсутствие свежей информации.

Виртуальный каталог



Виртуальный каталог - синтез Интернет-каталога и поисковой системы.
Интерфейс - аналогичен интерфейсу Интернет-каталога.
Виртуальный каталог в отличие от Интернет-каталога, не хранит ссылки на ресурсы, а составляется запрос к поисковой системе.

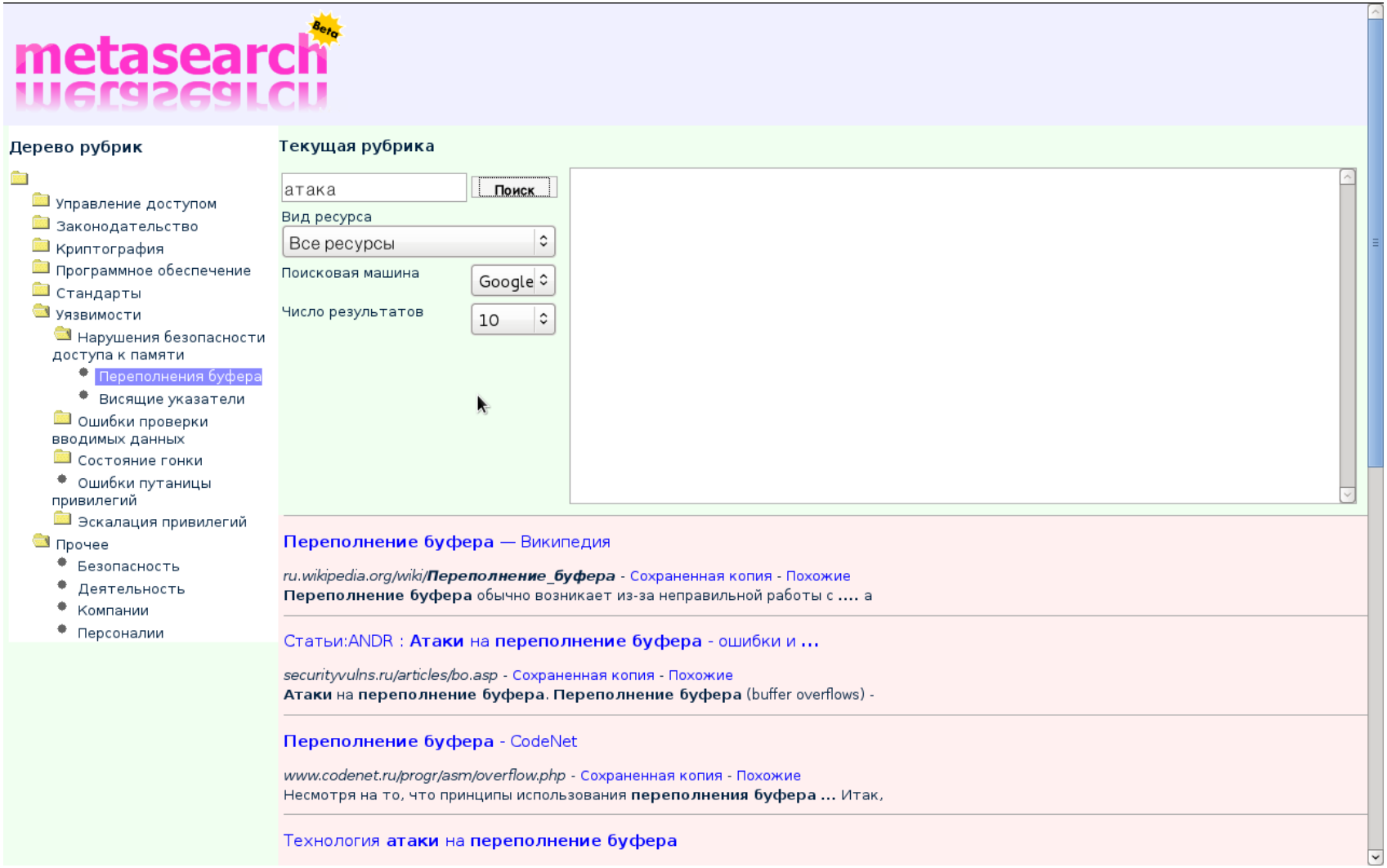
Виртуальный каталог: информационная безопасность

1. Управление доступом.
2. Законодательство.
3. Криптография.
4. ПО.
5. Стандарты.
6. Уязвимости.
7. Атаки.
8. Прочее.

Виртуальный каталог: Атаки

- Вредоносные программы.
  - Эксплоиты.
  - Вирусы.
    - Стелс-вирус.
    - Полиморфные вирусы.
  - Троянская программа.
  - Руткиты.
- Инъекции.
  - SQL-инъекция.
  - PHP-инъекция.
  - XSS.
- Отказ в обслуживании.
  - Переполнение буфера.
  - DoS-атака.
    - DDoS-атака.
    - Флуд.
- MITM-атака.
- IP-спуфинг.

Виртуальный каталог: информационная безопасность



Автоматизация подбора эвристик

1. Обучение.
   Выбрать пертинентные и не пертинентные тексты.

2. Генерация эвристик.
   - Получить множество $M1$ (лексем из пертинентных текстов).
   - Получить множество $M2$ (лексем из не пертинентных текстов).
   - Составить всевозможные конъюнкции из элементов $M1$.
   - Удаление конъюнкций состоящих из элементов $M2$.

3. Проверка результатов.

Сравнение

	google	**yandex**	**виртуальный каталог**
инъекции	50%	60%	80%
вирус	20%	40%	60%
аутентификация	70%	60%	90%
законодательство	0	0	40%
троянский конь	30%	40%	50%

В таблице сравниваются результаты поиска статей.

Тут пертинентность считается как отношение найденных ресурсов соответствующей тематики ко всем найденным ресурсам.

Результаты

1. Создана рубрика в виртуальном каталоге по предметной области «информационная безопасность».
2. Реализован выбор типа ресурса при поиске в виртуальном каталоге.
3. Автоматизировано порождение эвристик для пар рубрика-тип ресурса.

Публикации

- Тян А.Ю. Виртуальный каталог по предметной области информационная безопасность // Материалы 51 Международной Научной Студенческой Конференции «Студент и научно-технический прогресс»: Информационные технологии / Новосиб. гос. ун-т, Новосибирск, 2013, С. 125.

- Тян А.Ю. Виртуальный каталог по предметной области «информационная безопасность» // Современные инновации в науке и технике. Материалы III Международной научно-практической конференции / Юго-Западный гос. ун-т, Курск, 2013, С. 185-192.

Спасибо за внимание.