

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Отчет по лабораторной работе №3
по курсу «Распределенные системы»

Выполнил: студент 1-го курса магистратуры, гр. 21224

Гафиятуллин А.Р

Новосибирск, 2022

1. ОПРЕДЕЛЕНИЯ

Профиль базы данных — формальное определение базы данных, которое включает в себя описание абстрактной структуры записи, поисковых атрибутов и процессов обработки записи, таких как ввод, индексирование, хранение и извлечение.

2. СОКРАЩЕНИЯ

DCMES — Dublin Core Metatdata Element Set

XML — eXtensible Markup Language

XSLT — eXtensible Stylesheet Language Transformations

XPath — XML Path Language

3. ОБЪЕКТ ИССЛЕДОВАНИЯ

Профиль базы данных на основе набора элементов "Дублинское ядро".

4. ЦЕЛИ РАБОТЫ

Создать профиль базы данных для сервера Zebra используя набор элементов метаданных "Дублинское ядро".

5. МЕТОДОЛОГИЯ/ПОРЯДОК ПРОВЕДЕНИЯ РАБОТ

5.1. Развернута виртуальная машина.

5.2. Тестовые данные:

```
libuser@ds102:~/wdir % cat/repos/ds/datasets/fit.nsu.ru/collection.xml
<?xml version="1.0" encoding="UTF-8"?>
<dc:collection xmlns="http://fit.nsu.ru/ds" xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:metadata>
    <dc:identifier>https://www.nsu.ru/</dc:identifier>
    <dc:title>Новосибирский государственный университет</dc:title>
    <dc:title>НГУ</dc:title>
  </dc:metadata>
  <dc:metadata>
    <dc:identifier>https://www.nsu.ru/n/information-technologies-department/</dc:identifier>
    <dc:title>
      Факультет информационных технологий Новосибирского государственного университета
    </dc:title>
    <dc:title>ФИТ НГУ</dc:title>
  </dc:metadata>
</dc:collection>
libuser@ds102:~/wdir %
```

5.3. Установлен программный пакет **zoopark-ds-1.0.2**:

```

libuser@ds102:~/tmp/zoopark-ds-1.0.2 % make install
Making install in src
../config/install-sh -c -d '/home/libuser/local/bin'
/bin/sh ../libtool --mode=install /usr/bin/install -c zebra-client zebra-worker zebrash '/home/libuser/local/bin'
libtool: install: /usr/bin/install -c zebra-client /home/libuser/local/bin/zebra-client
libtool: install: /usr/bin/install -c zebra-worker /home/libuser/local/bin/zebra-worker
libtool: install: /usr/bin/install -c zebrash /home/libuser/local/bin/zebrash
Making install in include
Making install in cuba
../config/install-sh -c -d '/home/libuser/local/include/zoopark-ds'
/usr/bin/install -c -m 644 ctrl.h funupdate.h param.h version.h worker.h '/home/libuser/local/include/zoopark-ds'
Making install in etc
../config/install-sh -c -d '/home/libuser/local/etc/zoopark-ds'
/usr/bin/install -c -m 644 rc.conf.dist zebra-worker-start.sh zebra-worker-stop.sh yazgfs.xml pqf.properties '/home/libuser/local/etc/zoopark-ds'
Making install in profiles
Making install in dom
Making install in tab
../../../../config/install-sh -c -d '/home/libuser/local/share/zoopark-ds/profiles/dom/tab'
/usr/bin/install -c -m 644 string.chr urx.chr dom.xml extract.xml input.xml store2dc.xml '/home/libuser/local/share/zoopark-ds/profiles/dom/tab'
../../../../config/install-sh -c -d '/home/libuser/local/share/zoopark-ds/profiles/dom'
/usr/bin/install -c -m 644 zebra.cfg zebra.json '/home/libuser/local/share/zoopark-ds/profiles/dom'
Making install in tests
Making install in dom
libuser@ds102:~/tmp/zoopark-ds-1.0.2 % ls -l
distfiles      local          tmp
libuser@ds102:~/tmp/zoopark-ds-1.0.2 %

```

5.4. Созданы репозиторий и база данных на основе профиля:

```

libuser@ds102:~/wdir % $HOME/local/bin/zebra-client -w . -p $HOME/repos/ds/profiles create -r r1 -t fit.nsu.ru
{"success": true}

```

5.5. Выполнено индексирование записей из тестового набора данных:

```

libuser@ds102:~/wdir % $HOME/local/bin/zebra-client -w . update-file -r r1 -d db1 -f $HOME/repos/ds/datasets/
23:21:00-04/04 [log] Command UPDATE FILE:
23:21:00-04/04 [log]     repo='r1'
23:21:00-04/04 [log]     databasename='db1'
23:21:00-04/04 [log]     file='/home/libuser/repos/ds/datasets/'
23:21:00-04/04 [log]     wdir='.'
23:21:00-04/04 [log]     pdir='(null)'
23:21:00-04/04 [log]     cql2pqf='(null)'
23:21:00-04/04 [log] zebra_start 2.0.55 bd2bc9360225e695bbaba2c2d1cd6925c4eb23a5
23:21:00-04/04 [log] config ./r1/zebra.cfg
23:21:00-04/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-alvis.so
23:21:00-04/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-dom.so
23:21:00-04/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-grs-marc.so
23:21:00-04/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-grs-regx.so
23:21:00-04/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-grs-xml.so
23:21:00-04/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-safari.so
23:21:00-04/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-text.so
23:21:00-04/04 [log] enabling shadow spec=./r1/shadow:10G
23:21:00-04/04 [log] cache_fname = ./r1/shadow/cache
23:21:00-04/04 [log] dir /home/libuser/repos/ds/datasets/
23:21:00-04/04 [log] dir /home/libuser/repos/ds/datasets/dom/
23:21:00-04/04 [log] dom.xml dom filter: loading config file ./r1/tab/dom.xml
23:21:00-04/04 [log] XML: Entity: line 1:
23:21:00-04/04 [log] XML: parser
23:21:00-04/04 [log] XML: error :
23:21:00-04/04 [log] XML: Document is empty
23:21:00-04/04 [log] XML: 02037nam0 2200361 i 450 00100170000000500170001702100230003410000410005710100080
23:21:00-04/04 [log] XML: ^
23:21:00-04/04 [log] dir /home/libuser/repos/ds/datasets/fit.nsu.ru/
23:21:00-04/04 [log] add dom.dom.xml /home/libuser/repos/ds/datasets/fit.nsu.ru/collection.xml 0
23:21:00-04/04 [log] add dom.dom.xml /home/libuser/repos/ds/datasets/fit.nsu.ru/collection.xml 770
23:21:00-04/04 [log] Iterations: isam/dict 171/78
23:21:00-04/04 [log] Dict: inserts/updates/deletions: 78/0/0
23:21:00-04/04 [log] Records: 2 i/u/d 2/0/0
23:21:00-04/04 [log] enabling shadow spec=./r1/shadow:10G
23:21:00-04/04 [log] cache_fname = ./r1/shadow/cache
23:21:00-04/04 [log] zebra_stop: 0.08 0.00 0.04
{"success": true}
libuser@ds102:~/wdir %

```

5.6. Задача **search** с использованием языка **PQF**:

```
21:30:20-12/04 [log] zebra_stop: 0.00 0.00 0.00
{"success": true, "data": {"numberOfRecords": 2, "records": [{"recordIdentifier": null, "recordPosition": 1, "recordPacking": "string", "recordSchema": "dc", "recordData": "<?xml version='1.0' encoding='UTF-8'>\n<dc:metadata xmlns:dc='http://purl.org/dc/elements/1.1/'>\n    <dc:identifier>https://www.nsu.ru</dc:identifier>\n    <dc:title>Новосибирский государственный университет</dc:title>\n    <dc:title>НГУ</dc:title>\n    </dc:metadata>\n</dc:record>"}, {"recordIdentifier": null, "recordPosition": 2, "recordPacking": "string", "recordSchema": "dc", "recordData": "<?xml version='1.0' encoding='UTF-8'>\n<dc:metadata xmlns:dc='http://purl.org/dc/elements/1.1/'>\n    <dc:identifier>https://www.nsu.ru/n/information-technologies-department</dc:identifier>\n    <dc:title>\n        Факультет информационных технологий Новосиби́рского государственного университета\n    </dc:title>\n    <dc:title>ФИТ НГУ</dc:title>\n    </dc:metadata>\n</dc:record>"}]}}
```

5.7. Задача **search** с использованием языка **CQL**:

```
{"success": true, "data": {"numberOfRecords": 2, "records": [{"recordIdentifier": null, "recordPosition": 1, "recordPacking": "string", "recordSchema": "dc", "recordData": "<?xml version='1.0' encoding='UTF-8'>\n<dc:metadata xmlns:dc='http://purl.org/dc/elements/1.1/'>\n    <dc:identifier>https://www.nsu.ru</dc:identifier>\n    <dc:title>Новосибирский государственный университет</dc:title>\n    <dc:title>НГУ</dc:title>\n    </dc:metadata>\n</dc:record>"}, {"recordIdentifier": null, "recordPosition": 2, "recordPacking": "string", "recordSchema": "dc", "recordData": "<?xml version='1.0' encoding='UTF-8'>\n<dc:metadata xmlns:dc='http://purl.org/dc/elements/1.1/'>\n    <dc:identifier>https://www.nsu.ru/n/information-technologies-department</dc:identifier>\n    <dc:title>\n        Факультет информационных технологий Новосиби́рского государственного университета\n    </dc:title>\n    <dc:title>ФИТ НГУ</dc:title>\n    </dc:metadata>\n</dc:record>"}]}}
```

5.8. Задача **scan** с использованием языка **PQF**:

```
libuser@ds102:~/wdir % $HOME/local/bin/zebra-client -w . scan -r r1 -d db1 -e dc -t pqf -q '@1=4 нгу'
21:36:44-12/04 [log] Command SCAN:
21:36:44-12/04 [log]   repo='r1'
21:36:44-12/04 [log]   dbname='db1'
21:36:44-12/04 [log]   type='pqf'
21:36:44-12/04 [log]   term='@1=4 нгу'
21:36:44-12/04 [log]   position=-1
21:36:44-12/04 [log]   number=-1
21:36:44-12/04 [log]   wdir='.'
21:36:44-12/04 [log]   pdir='(null)'
21:36:44-12/04 [log]   cql2pqf='(null)'
21:36:44-12/04 [log] zebra_start 2.0.55 bd2bc9360225e695bbaba2c2d1cd6925c4eb23a5
21:36:44-12/04 [log] config ./r1/zebra.cfg
21:36:44-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-alvis.so
21:36:44-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-dom.so
21:36:44-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-grs-marc.so
21:36:44-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-grs-regx.so
21:36:44-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-grs-xml.so
21:36:44-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-safari.so
21:36:44-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-text.so
21:36:44-12/04 [log] dom.xml dom filter: loading config file ./r1/tab/dom.xml
21:36:44-12/04 [log] zebra_stop: 0.00 0.00 0.01
{"success": true, "data": {"terms": [{"numberOfRecords": 2, "value": "нгу", "displayTerm": "НГУ"}, {"numberOfRecords": 1, "value": "новосибирский", "displayTerm": "Новосибирский"}, {"numberOfRecords": 1, "value": "новосибирского", "displayTerm": "Новосибирского"}, {"numberOfRecords": 1, "value": "технологий", "displayTerm": "технологий"}, {"numberOfRecords": 1, "value": "университет", "displayTerm": "университет"}, {"numberOfRecords": 1, "value": "университета", "displayTerm": "университета"}, {"numberOfRecords": 1, "value": "факультет", "displayTerm": "Факультет"}, {"numberOfRecords": 1, "value": "фит", "displayTerm": "ФИТ"}]}}
```

5.9. Задача **scan** с использованием языка **CQL**:

```
libuser@ds102:~/wdir % $HOME/local/bin/zebra-client -w . -c ./pqf.properties scan -r r1 -d db1-e dc --
tcql -q "dc.title=нры"
21:38:27-12/04 [log] Command SCAN:
21:38:27-12/04 [log]   repo='r1'
21:38:27-12/04 [log]   databasename='db1'
21:38:27-12/04 [log]   type='cql'
21:38:27-12/04 [log]   term='dc.title=нры'
21:38:27-12/04 [log]   position=-1
21:38:27-12/04 [log]   number=-1
21:38:27-12/04 [log]   wdir='.'
21:38:27-12/04 [log]   pdir='(null)'
21:38:27-12/04 [log]   cql2pqf='./pqf.properties'
21:38:27-12/04 [log] CQL transform result: '@attr 6=1 @attr 2=3 @attr 4=1 @attr 3=3 @attr 6=1 @attr 5=
100 @attr 1=4 "нры" '
21:38:27-12/04 [log] zebra_start 2.0.55 bd2bc9360225e695bbaba2c2d1cd6925c4eb23a5
21:38:27-12/04 [log] config ./r1/zebra.cfg
21:38:27-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-alvis.so
21:38:27-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-dom.so
21:38:27-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-grs-marc.so
21:38:27-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-grs-regx.so
21:38:27-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-grs-xml.so
21:38:27-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-safari.so
21:38:27-12/04 [log] Loaded filter module /usr/local/lib/idzebra-2.0/modules/mod-text.so
21:38:27-12/04 [log] dom.xml dom filter: loading config file ./r1/tab/dom.xml
21:38:27-12/04 [log] zebra_stop: 0.00 0.01 0.00
{"success": true, "data": {"terms": [{"numberOfRecords": 2, "value": "нры", "displayTerm": "НРЫ"}, {"numberOfRe
cords": 1, "value": "новосибирский", "displayTerm": "Новосибирский"}, {"numberOfRecords": 1, "value": "новосиби
рского", "displayTerm": "Новосибирского"}, {"numberOfRecords": 1, "value": "технологий", "displayTerm": "техно
логий"}, {"numberOfRecords": 1, "value": "университет", "displayTerm": "университет"}, {"numberOfRecords": 1, "
value": "университета", "displayTerm": "университета"}, {"numberOfRecords": 1, "value": "факультет", "displayT
erm": "Факультет"}, {"numberOfRecords": 1, "value": "фит", "displayTerm": "ФИТ"}]}}
```

5.10. Удаления репозитория:

```
libuser@ds102:~/wdir % $HOME/local/bin/zebra-client -w . destroy -r r1
21:38:50-12/04 [log] Command DESTROY:
21:38:50-12/04 [log]   repo='r1'
21:38:50-12/04 [log]   wdir='.'
21:38:50-12/04 [log]   pdir='(null)'
21:38:50-12/04 [log]   cql2pqf='(null)'
{"success": true}
```

5.11. Описание профиля базы данных, файл extract.xml:

```

libuser@dsl02:~/wdir % cat/repos/ds/profiles/fit.nsu.ru/tab/extract.xsl
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:z="http://indexdata.com/zebra-2.0"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  exclude-result-prefixes="dc"
>

  <!--
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  -->

  <xsl:output method="xml" encoding="UTF-8"/>

  <xsl:template match="text()"/>

  <xsl:template match="/">
    <xsl:apply-templates />
  </xsl:template>

  <xsl:template match="dc:collection">
    <xsl:apply-templates />
  </xsl:template>

  <xsl:template match="dc:metadata">
    <z:record z:id="{normalize-space(dc:identifier)}" type="update">
      <xsl:apply-templates />
    </z:record>
  </xsl:template>

  <!-- Bib-1 12 (Local-number) -->
  <xsl:template match="dc:identifier">
    <z:index name="Local-number:w Local-number:p Any:w">
      <xsl:value-of select="."/>
    </z:index>
  </xsl:template>

  <!-- Bib-1 4 (title) -->
  <xsl:template match="dc:title">
    <z:index name="Title:w Title:p Any:w">
      <xsl:value-of select="."/>
    </z:index>
  </xsl:template>

  <!-- Bib-1 1003 (author) -->
  <xsl:template match="dc:creator">
    <z:index name="Author:w Author:p Any:w Any:p">
      <xsl:value-of select="text()"/>
    </z:index>
  </xsl:template>

</xsl:stylesheet>

```

5.12. Таблица соответствия элементов DCMES и атрибутов Bib-1:

Элемент DCMES	Атрибуты Bib-1
title	Title (4), Any(1016)
identifier	Local number (12), Any(1016)
creator	Author (1003), Any(1016)
subject	Subject heading (21), Any(1016)
description	Abstract (62), Note (63), Any(1016)
publisher	Publisher (1018), Any(1016)
contributor	Editor (1020), Any(1016)
date	Date (30), Any(1016)
type	Content-type (1034), Any(1016)
format	Authority/format id (1013), Any(1016)
source	Note (63), ISBN (7), ISSN (8), Identifier—standard (1007), Doc-id(1032), Any(1016)
language	Code--language (54), Any(1016)
relation	Note (63), ISBN (7), ISSN (8), Identifier—standard (1007), Doc-id(1032), Any(1016)
coverage	Geographic-class (1022), Name geographic (58), Subject subdivision (47), Any(1016)
rights	No. legal deposit (49), Any(1016)

6. ОПИСАНИЕ РЕЗУЛЬТАТА

Изучен набор элементов метаданных "Дублинское ядро", приведена таблица соответствия элементов DCMES и атрибутов Bib-1 и проведено тестирование приведенного профиля базы данных задачами scan и search на языках PQF и CQL.

7. ССЫЛКИ

1. ANSI/NISO Z39.85-2012 — URL:
<https://groups.niso.org/higherlogic/ws/public/download/10258>
2. Zebra - User's Guide and Reference — URL:
<https://software.indexdata.com/zebra/doc/index.html>
3. Профиль fit.nsu.ru — URL:
https://bitbucket.org/oleg_kolobov/ds/src/master/
4. XSL Transformations (XSLT) Version 1.0 — URL:
<https://www.w3.org/TR/1999/REC-xslt-19991116>
5. XML Path Language (XPath) Version 1.0 — URL:
<https://www.w3.org/TR/1999/REC-xpath-19991116/>
6. Prefix Query Format (PQF) — URL:
<https://software.indexdata.com/yaz/doc/tools.html#PQF>
7. CQL — URL: <https://www.loc.gov/standards/sru/cql/index.html>
8. Bib-1 Attribute Set — URL:
<https://www.loc.gov/z3950/agency/defns/bib1.html>