Lilacs documents & DB XML

un prototipo

utilización de la base de datos Lilacs
con Oracle Berkeley DB XML

Propósito del prototipo –

Evaluación de carga y operación de la base de datos LILACS por medio del sistema de gestión de base de datos Oracle Berkeley DB XML

Exploración y evaluación de la viabilidad técnica de utilizar el sistema de gestión de bases de datos Oracle Berkeley DB XML en la explotación de la base de datos LILACS publicada acumulada (base de datos retrospectiva) en formato XML nativo

El prototipo pretende ser un instrumento de pruebas y evaluación de las condiciones que producirán determinados niveles de desempeño en el uso de la base de datos LILACS bajo DB XML

Funciones del prototipo –

Permitir la experimentación, bajo diversas condiciones, respecto al desempeño de la combinación LILACS - DB XML, tratando de aprovechar las características y potencialidades del DB XML en la gestión de esa base de datos en formato XML nativo

- evaluación de carga de documentos
- evaluación de la operación de la base de datos

Características del prototipo –

Operación de la base de datos

- base de datos LILACS centralmente producida, publidada y distribuída
- base de datos cargada en DB XML centralmente en tiempo real
- actualizaciones semanales de la base de datos
- almacenamiento en XML nativo
- replicación (no aún configurada)

Evaluación

- almacenamiento requerido
- tiempo de carga
- tiempo de respuesta
- interface usuario final

Implementación actual del prototipo –

Características

- tipo de base de datos DB XML: B-tree
- tipo de almacenamiento: node storage
- índices: Abstract, Title, TitleInEnglish, LilacsID, dbxml:metadata[‘name’]
- granularidad de la indización: node
- transacciones: funcionando
- volúmen previsto: 1 000 | 10 000 | 100 000 | 500 000+ documentos
- replicación: no aún configurada

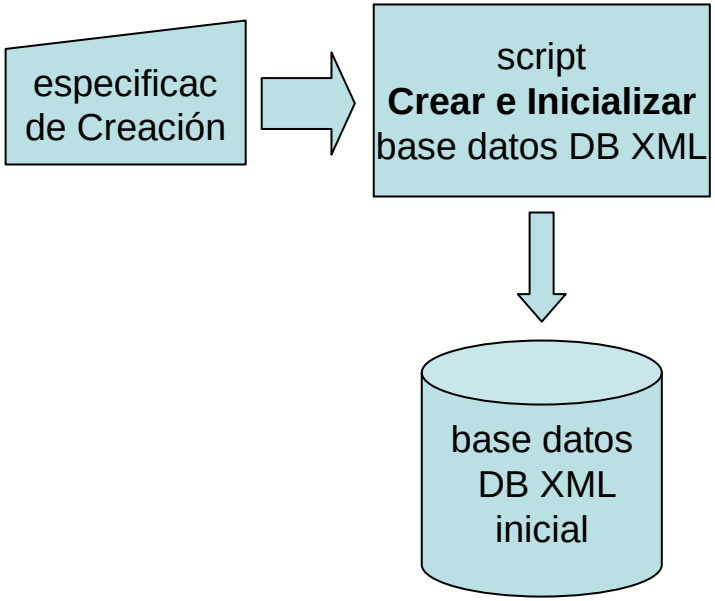
Desarrollo con scripts dentro de familias de componentes

- familia de creación, inicialización y modificación de característisticas
- familia de almacenamiento y mantenimiento de documentos
- familia de recuperación de documentos

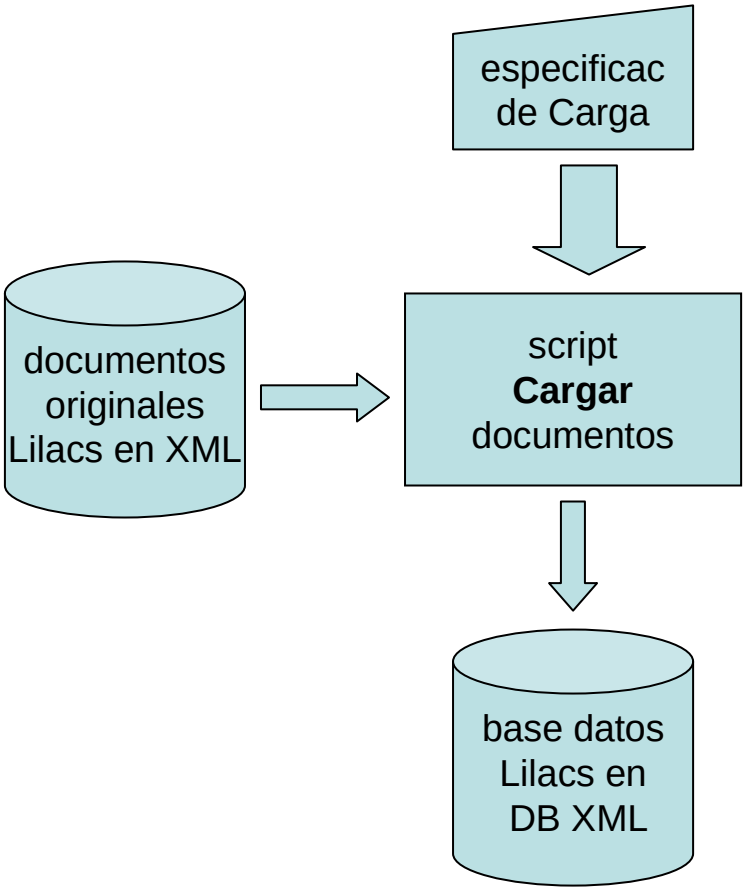
Componentes del prototipo –

1. creación de base de datos inicial - batch, offline
2. carga de docs - batch, incremental, transactional, non-exclusive real-time online

Crear e Inicializar base datos



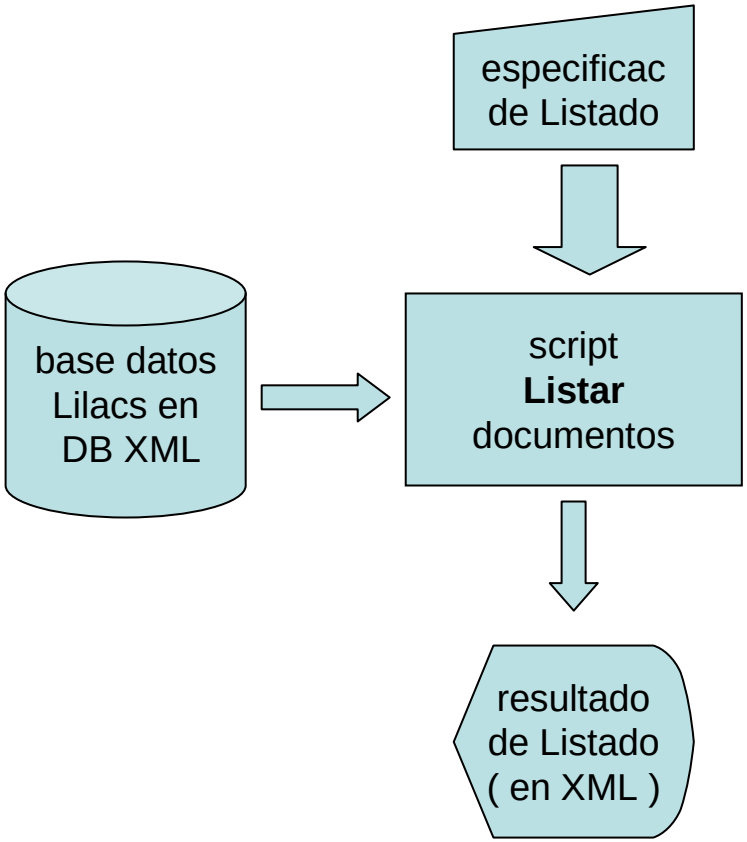
Cargar documentos Lilacs



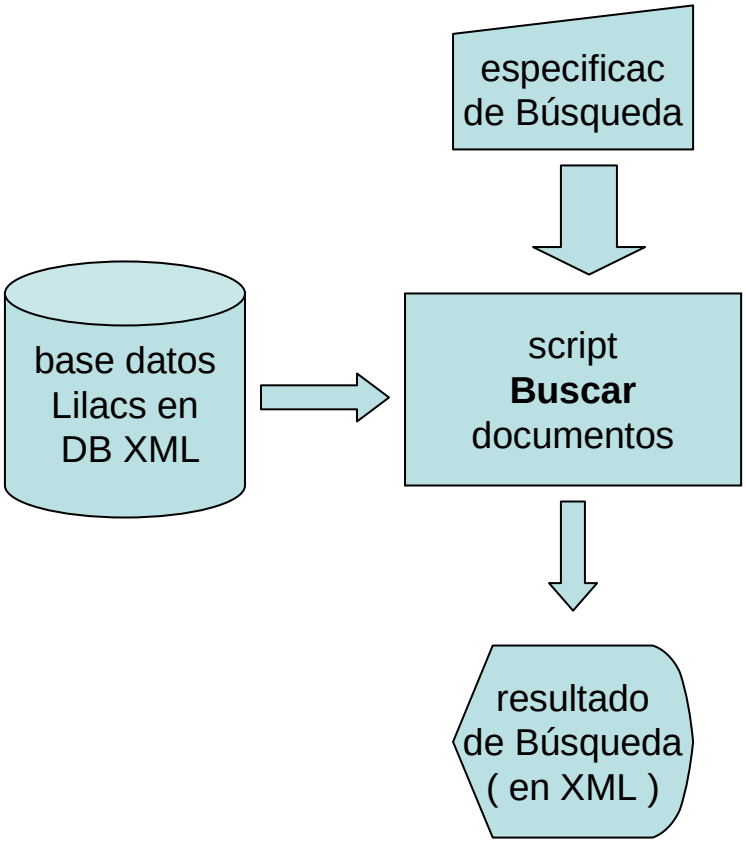
Componentes del prototipo –

3. listado de documentos - non-exclusive real-time online
4. búsqueda de documentos - non-exclusive real-time online

Listar documentos Lilacs

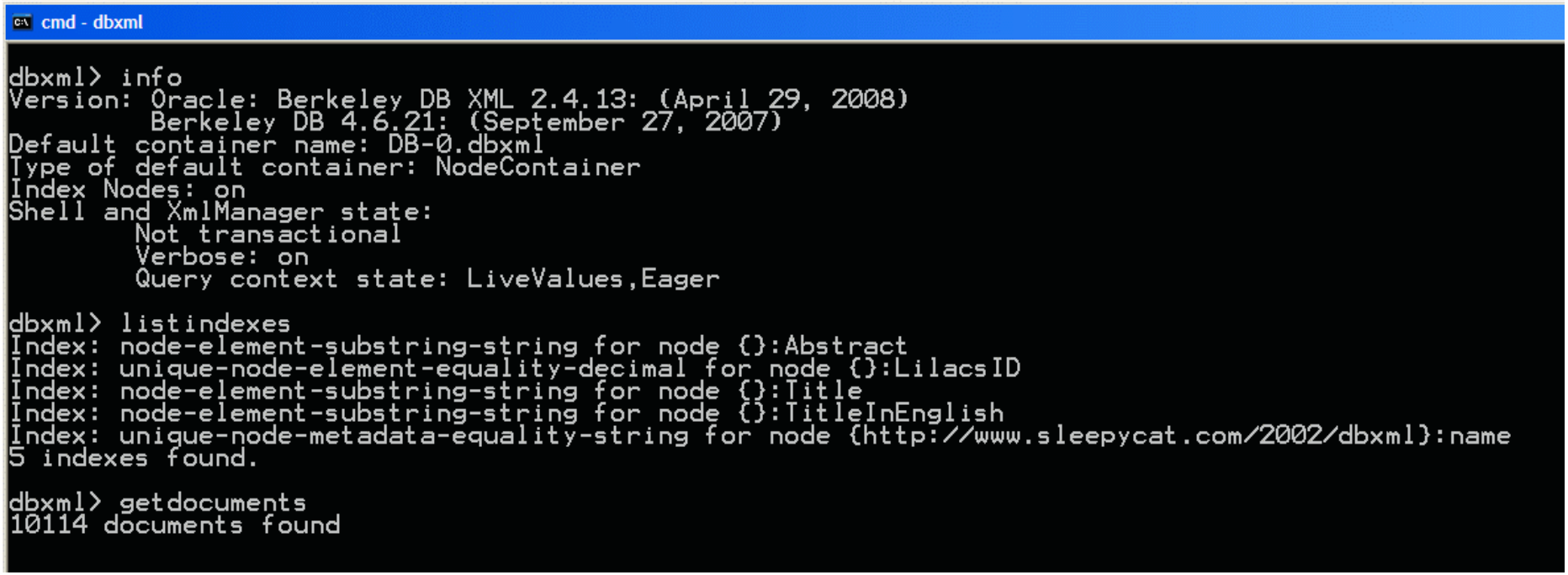


Buscar documentos Lilacs



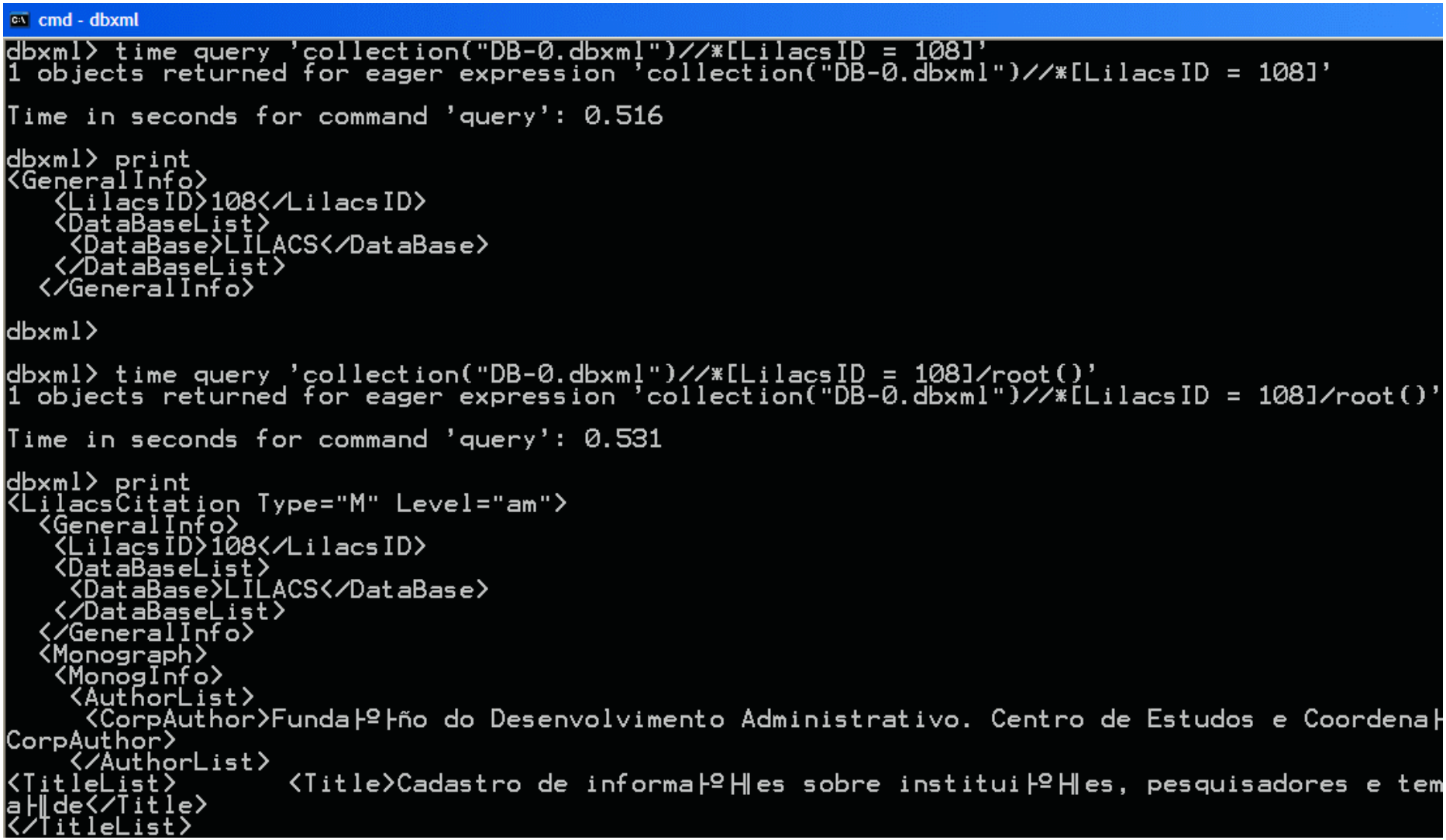
Acceso a la base de datos con el shell de DB XML

Características - info, listIndexes, getDocuments



Listar un documento con el shell de DB XML

dbxml> time query 'collection()//*[LilacsID = 108]‘
dbxml> time query 'collection()//*[LilacsID = 108]/root()'

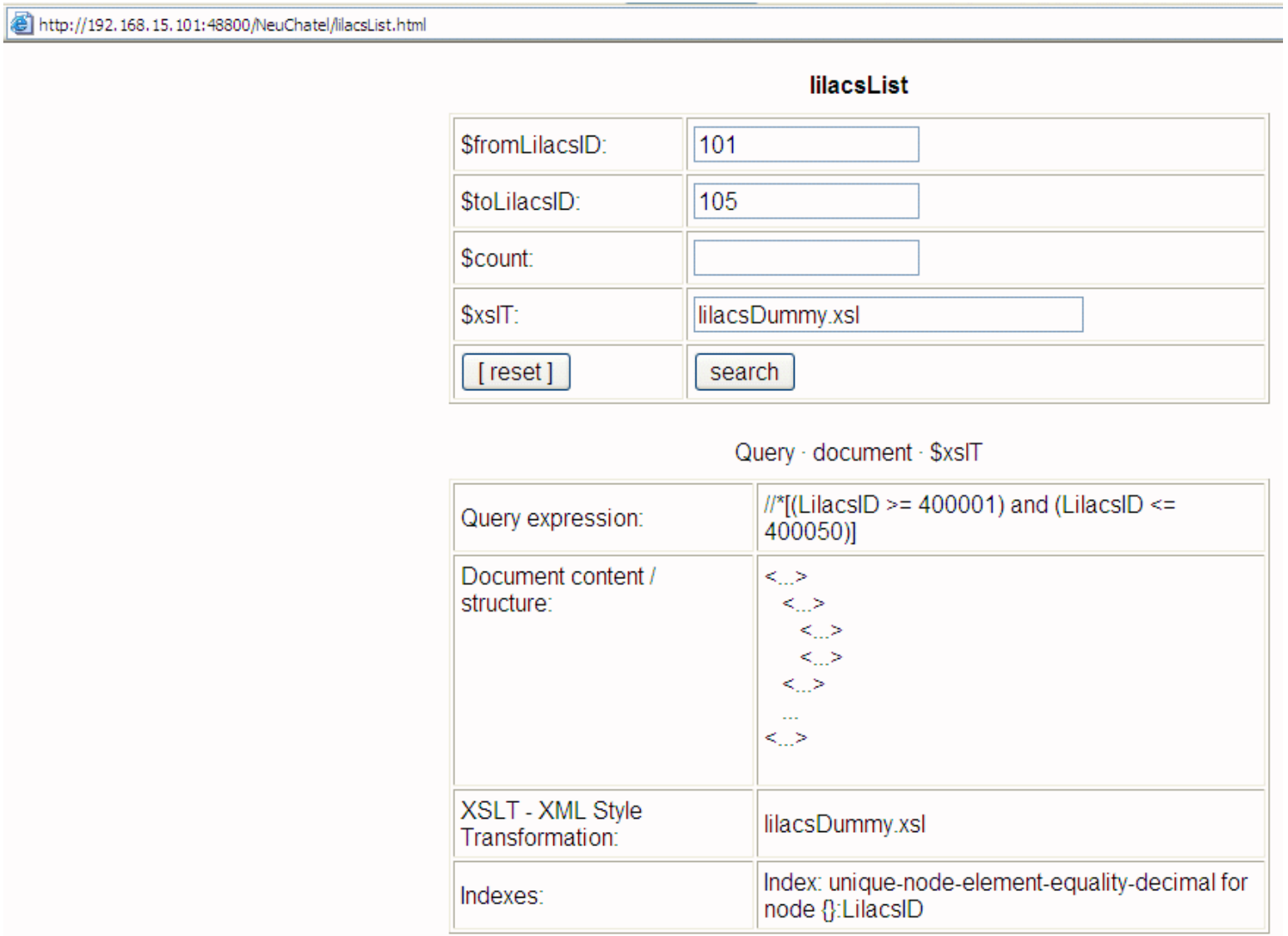


Buscar un documento con el shell de DB XML

dbxml> time query 'collection()//*[dbxml:contains(Abstract, "crianca") and
dbxml:contains(Abstract, "gasto federal")]'

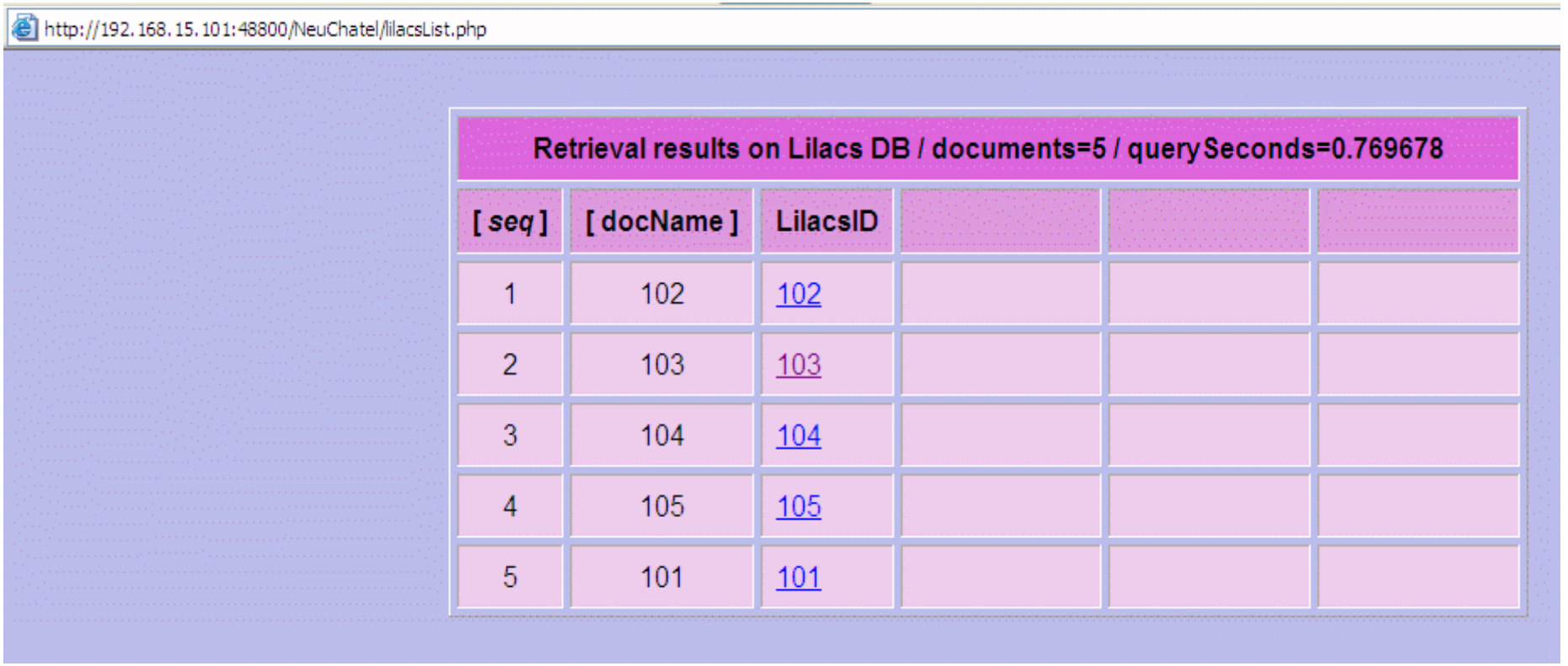
Acceso por script Listar del prototipo Lilacs & DB XML

una interface de pedido



Acceso por script Listar del prototipo Lilacs & DB XML

resultado transformado por XSLT



Acceso por script Listar del prototipo Lilacs & DB XML

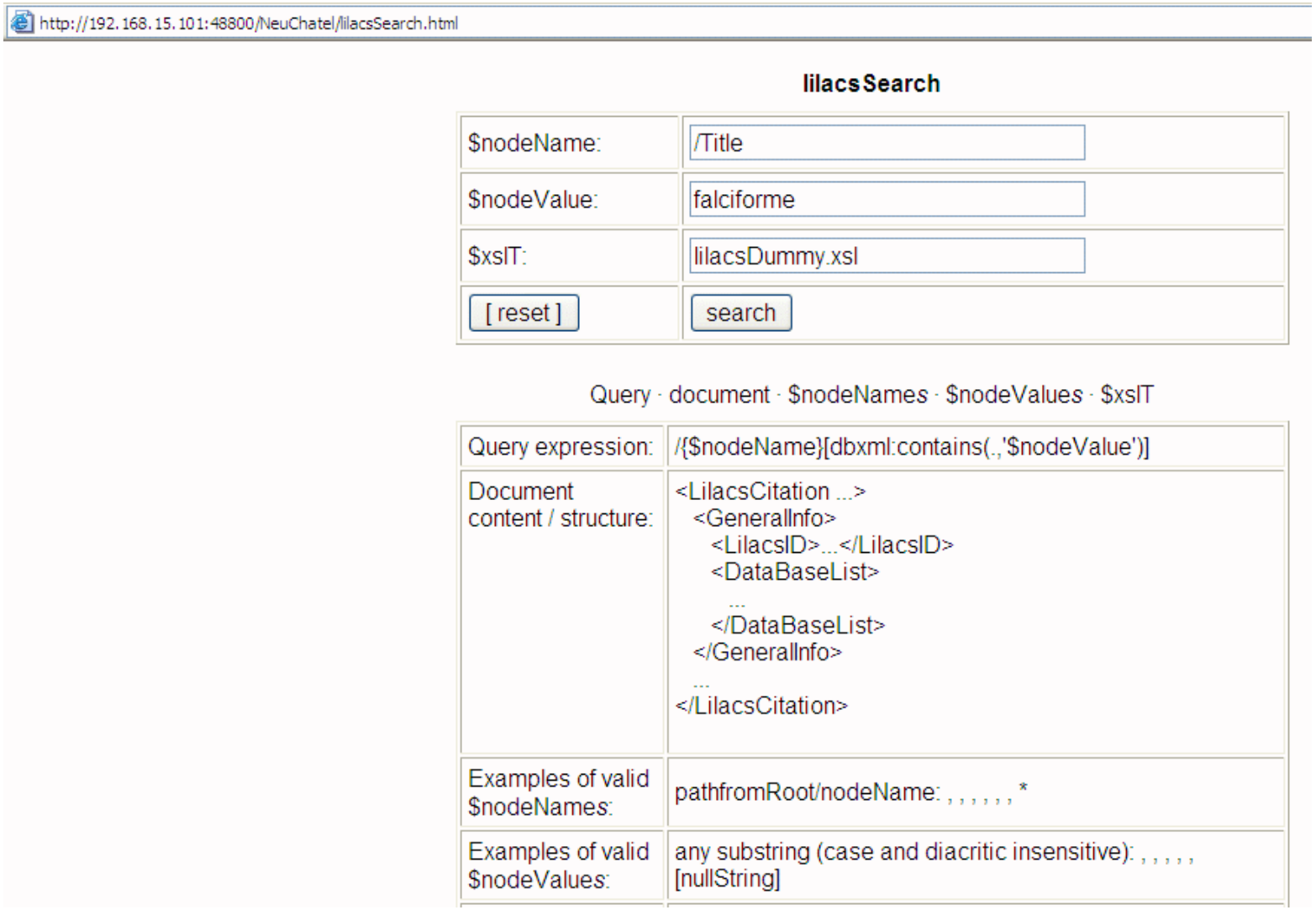
documento “103” del resultado, listado individualmente y sin transformar por XSLT

Address http://192.168.15.101:48800/NeuChatel/lilacsList.php?fromLilacsID=103&toLilacsID=103&xslT=


```
<?xml version="1.0" encoding="UTF-8" ?>
- <neuChatel version="20-Aug-2008" release="20-Aug-2008-1700CST" function="neuChatel_list" date="20080913" time="104931 UTC">
  - <parameters>
      <!--  scriptName="lilacsList.php"  -->
      <!--  ...listing docNames fromLilacsID="103" toLilacsID="103" count="-1"  -->
      <!--  ...querying with (XPath) expression "//*[(LilacsID &gt;= 103) and (LilacsID &lt;= 103)]"  -->
      <!--  ...output document transformation by xslT=""  -->
    </parameters>
  - <result status="success" retrieved="1" querySeconds="0.747438" fromLilacsID="103" toLilacsID="103" count="-1">
    - <document seq="1" docName="103">
      - <LilacsCitation Type="M" Level="am">
        - <GeneralInfo>
            <LilacsID>103</LilacsID>
          - <DataBaseList>
              <DataBase>LILACS</DataBase>
            </DataBaseList>
          </GeneralInfo>
        - <Monograph>
          - <MonogInfo>
            - <AuthorList>
                <CorpAuthor>Fundação Instituto Brasileiro de Geografia e Estatística, ed</CorpAuthor>
              </AuthorList>
            - <TitleList>
                <Title>Perfil estatístico de crianças e mães no Brasil: situação de saúde 1981</Title>
              </TitleList>
              <NumberOfPages>264</NumberOfPages>
            </MonogInfo>
          - <AnalyticalInfo>
            - <AuthorList>
              - <Author>
                  <Name>Oliveira, Luiz Antonio Pinto de</Name>
                </Author>
              - <Author>
                  <Name>Simões, Celso Cardoso da Silva</Name>
                </Author>
```

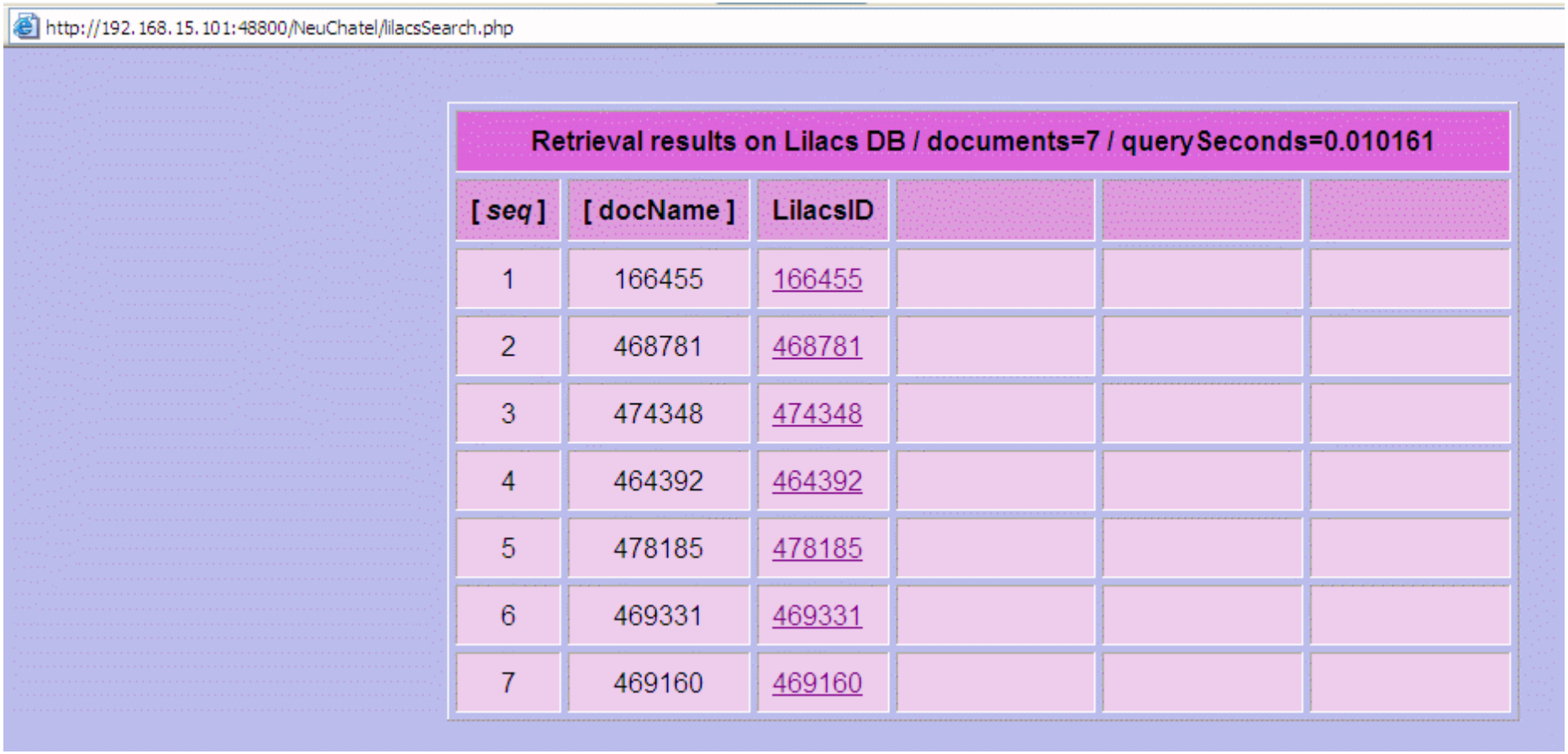
Acceso por script Buscar del prototipo Lilacs & DB XML

una interface de pedido



Acceso por script Buscar del prototipo Lilacs & DB XML

resultado transformado por XSLT



Acceso por script Buscar del prototipo Lilacs & DB XML

documento “166455” del resultado, listado individualmente y sin transformar por XSLT

Script “Crear base de datos” del prototipo Lilacs & DB XML

un proceso de creación - **batch, offline**

```
D:\~all Lilacs DB XML>php lilacsCreate.php

lilacsCreate.php version C3

Date Thursday 11 September 2008, time 23:41:55 UTC

Create and initialize a new database -



ContainerName - DB-0.dbxml

Indexes -
Index: node-element-substring-string for node {}:Abstract
Index: unique-node-element-equality-decimal for node {}:LilacsID
Index: node-element-substring-string for node {}:Title
Index: node-element-substring-string for node {}:TitleInEnglish
Index: unique-node-metadata-equality-string for node
          {http://www.sleepycat.com/2002/dbxml}:name
5 indexes found.



Created containerName - DB-0.dbxml.

End of program.



D:\~all Lilacs DB XML>
```

Script “Cargar documentos” del prototipo Lilacs & DB XML

final de un proceso de carga de documentos - **batch, incremental, transactional, non-exclusive real-time online**

```
.
..
...

begin transaction 1344
opNumber: 5373, addDoc: 366643.
opNumber: 5374, addDoc: 468884.
opNumber: 5375, addDoc: 368971.
opNumber: 5376, addDoc: 268695.
end transaction   1344

begin transaction 1345
opNumber: 5377, addDoc: 373601.
opNumber: 5378, addDoc: 469344.
opNumber: 5379, addDoc: 288551.
opNumber: 5380, addDoc: 373649.
end transaction   1345

begin transaction 1346
opNumber: 5381, addDoc: 373143.
opNumber: 5382, addDoc: 373144.
end transaction   1346

...operations ended.

New current number of documents in container: 10114.

Database document loader says: good-bye.

Total time (program): 10984.834393 seconds.
Total net time (loading): 10683.497536 seconds.

End of program.
```

Prototipo Lilacs & DB XML

reporte de dos procesos de carga de documentos

```
Carga de documentos Lilacs en DB XML
------------------------------------


batch           batch 1          batch 2

intervalo ID    0-250000         250001-0

inicio          17:45 CST        19:45 CST
final           19:25 CST        22:48 CST
tiempo          100 minutos      183 minutos

doc/transacc    4                4
transacc        1183             1346

documentos      4732             5382

rapidez prom    0,79 docs/seg    0,49 docs/seg



status:

total docs      4732             10114
almacenam       71.1 Mbytes      139 Mbytes



procesador intel 1.46 GHz
memoria principal real 512 Mbytes
MS Windows XP Home Edition 2002 Service Pack 3
PHP 5.2.2 - Oracle Berkeley DB XML 2.4.13
```

Comentarios / discusión –

Asuntos

- resultados de la evaluación hasta el momento
- aplicabilidad a IAH
- comparación con XML ISIS Script
- comparación con XIsis
- relación y aplicabilidad a ISIS NBP
- operación y características técnicas
- demostración de operación de componentes
- utilización de DB XML en producción de información