

Full-Documentation of BaseX

BaseX-Team

Full-Documentation of BaseX

BaseX-Team

Publication date June 2009

Copyright © 2009 DBIS, University of Konstanz

Table of Contents

1. Introduction	1
1. Prerequisites	1
1.1. Operating System	1
1.2. Software	1
1.3. Hardware	1
2. Installation	1
3. Quick Start	2
4. Starting with BaseX GUI	2
5. Starting with BaseX Console	2
6. Using BaseX as Server-Client	3
7. Starting BaseX as a service	3
2. Working with the BaseX Console	4
1. Launching BaseX in console mode	4
2. What can i do in the console mode?	4
3. Description of the Console Features	4
3.1. Database Commands	5
3.2. Query Commands	5
3.3. Update Commands	5
3.4. General Commands	5
3. Working with the BaseX GUI	6
1. Description of all the Views in the GUI	6
2. Description of the GUI Features	6
2.1. What does the Filter button exactly do?	6
2.2. What is Realtime Filtering?	6
2.3. How can I search without XQuery?	6
2.4. The search field in the query panel and the search field right beyond the buttons, do they differ?	6
2.5. How can I configure the look and feel of BaseX?	7
2.6. What is the Main-Memory Mode?	7
2.7. How can I import my file system? What does that mean?	7
2.8. How can I import my data aka how do i create a new database instance aka how to shred an XML file?	7
4. General Information	8
1. Location of BaseX Files	8
2. How do you measure the performance of the queries?	8
3. What indexing techniques are available and what do they do?	8
4. How can i create a Database for a couple of XML files?	8
5. Is there a way to save large results of a query in a file?	9
5. Description of BaseX-Features	10
1. XQuery Implementation	10
2. XQuery Full-Text Implementation	10
3. Description of the Indexes	10
4. What about the different indexes BaseX offers: which data structures are applied?	11
5. Do you support XML Updates?	11
6. What can i do if i get Exceptions on huge XML files as input?	11
7. Can i protect a DB in BaseX with a password?	12
6. Developer's Guide	13

1. Programming with BaseX API	13
2. Programming with XML:DB API	13
3. Programming with XQuery API	13

Chapter 1. Introduction

BaseX is a native, open-source XML database and efficient XQuery/XQuery Full Text processor. It supports very large XML instances and offers a visual, interactive frontend. BaseX is written in Java and freely available for download. It is developed by the Database and Information Systems Group at the University of Konstanz.

1. Prerequisites

1.1. Operating System

BaseX was tested on the following platforms:

- Windows 2000, XP, Vista, 7
- Linux: Debian/Ubuntu, SuSE, Redhat
- Max OS X (10.4+)
- OpenBSD (4.3)

Every other platform with a Java VM should be able to run BaseX.

1.2. Software

To run any version of BaseX, you need version 1.5 or later of the Java runtime environment (JRE) or development kit (JDK).

The latest version of Java can be downloaded at <http://www.java.com> or <http://java.sun.com>.

1.3. Hardware

BaseX works with very limited resources (400 MHz, 64 MB RAM). To work with larger databases, however, you might benefit from more RAM. If you encounter main memory limits, there are several things you can try:

- Increase Java's virtual memory
- Switch off database indexes
- Use BaseX's internal XML parser

2. Installation

BaseX works without any installation. The following files can be downloaded from the homepage:

- *BaseX.jar*: simple JAR file
- *BaseX.exe*: a JAR file, which is wrapped into a Windows executable
- *BaseX-XQJ.jar*: XQuery for Java Database API

- *BaseX-XMLDB.jar*: XML:DB Database API
- *BaseX-Complete.zip*: JAR files, sources, documentation and files to build the project

The JAR file and the Windows executable can be launched with a simple double-click. The "Complete" package contains the following files and directories (selection):

Project Directory

doc	Project documentation
etc	Various files for project handling
src	Project sources
.project	Eclipse project file
makefile	Linux Makefile
BaseX.jar	Public JAR file
license.txt	Licensing Information
readme.txt	Project Information
build.xml	Ant Build file
input.xml	XML sample document

In the *etc* folder, you find Linux scripts and Windows batch files which can be moved into your PATH environment to speedup execution.

3. Quick Start

By double-clicking the JAR file, the graphical interface is launched. If you start BaseX on the command line, or use the scripts mentioned above, you have some more options:

Console version

```
java -cp BaseX.jar org.baseX.BaseX
```

Running the server and a client instance

```
java -cp BaseX.jar org.baseX.BaseXServer
```

```
java -cp BaseX.jar org.baseX.BaseXClient
```

GUI version (granting more memory)

```
java -Xmx512m -jar BaseX.jar
```

4. Starting with BaseX GUI

Just download the BaseX.jar or BaseX.exe to start the GUI Version of BaseX. If you download the source, you have to start the BaseXWin.java class.

5. Starting with BaseX Console

Just enter `java -cp BaseX.jar org.baseX.BaseX` on the command line. The Java option `-Xmx...` reserve more memory, and the `-h` flag of BaseX lists all available flags.

6. Using BaseX as Server-Client

Start the BaseXServer.java class and the Server is running. Now you can connect with BaseXClient.java.

- To start BaseX in server mode, type in: `java -cp BaseX.jar org.basex.BaseXServer`
- To run a BaseX client, please type in: `java -cp BaseX.jar org.basex.BaseXClient -s server.org`

7. Starting BaseX as a service

Chapter 2. Working with the BaseX Console

1. Launching BaseX in console mode

Enter `java -Xmx512m -cp BaseX.jar org.basex.BaseX -h` to get all available flags.

```
BaseX 5.5; DBIS, University of Konstanz
Usage: BaseX [options] [query]
[query] specify query file
-c chop whitespaces
-d debug mode
-e skip entity parsing
-o [file] specify output file
-q "cmd" send BaseX commands
-v/V show (all) process info
-x print result as xml
-z skip query output
```

2. What can i do in the console mode?

Type in "help" to get a list of all BaseX commands. Several commands can be separated by semicolons. To evaluate commands without entering the console mode, you can use the -q option on the command line. See the following example.

```
java -Xmx512m -cp BaseX.jar org.basex.BaseX -vq "create xml input.xml;

Database 'input' created in 155.17 ms.

Parsing      : 110.35 ms
Compiling    : 0.81 ms
Evaluating   : 0.05 ms
Printing     : 2.06 ms
Total Time: 113.35 ms
Results      : 1 Item
Printed      : 360 Bytes
```

3. Description of the Console Features

You can type in help in the console to get a list of all BaseX commands. Several commands can be separated by semicolons. To evaluate commands without entering the console mode, you can use the -q option on the command line

3.1. Database Commands

- Create (create [DB|FS|INDEX] [...]): Creates database from XML or filesystem, or creates index
- Open (open [database]): Opens the specified [database].
- Info (info [DB|INDEX|TABLE]?): Shows information on the currently opened database.
- Close (close): Closes the current database.
- List (list): Lists all available databases.
- Drop (drop [DB|INDEX] [...]): Drops a database or an index.
- Export (export [file]): Exports the current context set to an XML [file].
- Optimize (optimize): Optimizes the current database structures.

3.2. Query Commands

- XQuery (xquery [query]): Evaluates an XQuery and prints the result.
- Find (find [query]): Evaluate a simple keyword [query] and print its results. This command is used in the simple search mode in the GUI.
- CS (cs [query]): Evaluates the specified XPath [query] and set the result as new context set.

3.3. Update Commands

- Copy (copy [pos] [source] [target]): Copy database nodes. Evaluates the [source] query and copies the resulting nodes as child nodes into the [target] query. [pos] specifies the child position; if 0 is specified, the nodes are inserted as last child. The queries should be enclosed by brackets.
- Delete (delete ["target"]): Delete database nodes resulting from the specified [target] query. The query should be enclosed by brackets.
- Insert (insert [fragment|element|attribute|text|comment|pi] [...]): Insert database nodes. Insert a fragment or a specific node at the specified child [pos] of the specified [target] query.
- Update (update [element|attribute|text|comment|pi] [...]): Update database nodes satisfying the specified [target] query.

3.4. General Commands

- Help (help [command]): Get help on BaseX commands. If [command] is specified, information on the specific command is printed; otherwise, all commands are listed. If 'all' is specified, hidden commands are included.
- Set (set [option] [value?]): Sets global options. The currently set values can be shown with the info command.
- Exit (exit/quit): Leave the console mode of BaseX.

Chapter 3. Working with the BaseX GUI

1. Description of all the Views in the GUI

- Text View: Displays query results and other textual output.
- Map View: This visualization represents all data in a TreeMap. All nodes of the XML document are represented as rectangles, filling the complete area. You can choose in the Menu Options/Map Layout the different Algorithms: Split, Strip, Squarified and Slice and Dice Layout.
- Folder View: This visualization displays all XML nodes in a usual tree view.
- Table View: This visualization displays all XML nodes in a table with rows and columns.
- Scatterplot View: This visualizations displays all XML nodes in a scatterplot.

2. Description of the GUI Features

2.1. What does the Filter button exactly do?

After pressing this button, the visualizations display the previously highlighted XML nodes, omitting all the other nodes of the document. If realtime filtering is enabled, the Filter button will be disabled, and all results are automatically filtered.

2.2. What is Realtime Filtering?

If realtime filtering is enabled, all visualizations directly show the query results while entering the query. If this feature is disabled, the query results are highlighted in the visualizations and can be explicitly filtered using the 'Filter' button.

2.3. How can I search without XQuery?

The Search field triggers a simple search query in the XML document. The following syntax is supported:

Query	Description
foo	Find tags and texts containing foo
=foo	Find exact matching text nodes
@foo	Find attributes and attribute values
@=foo	Find exact attribute values

2.4. The search field in the query panel and the search field right beyond the buttons, do they differ?

Not really, both offer the same functionality, but in the query panel there is a dropdown menu where it's possible to choose tags or attributes of the XML document.

2.5. How can I configure the look and feel of BaseX?

All panels are freely adjust- and draggable. Changing of colors, fonts and the layout is possible via the 'Layout' entry in the 'Options' menu.

2.6. What is the Main-Memory Mode?

The Main-Memory Mode speeds up querying but disables updates. The table data is kept in memory and the text of a document is still accessed from disk.

2.7. How can I import my file system? What does that mean?

To import a file system, go to 'File/Import Filesystem...'. You will be prompted for the next steps. A file hierarchy traversal is performed and the directory structure of the filesystem is mapped into an XML representation. Additionally some metadata is extracted from some known file types. What you get is an 'XML view' of your current file system. You can query the filesystem just as any other database instance. Manipulation (update, deletion) of the XML database instance representing a file hierarchy has no effect on the real filesystem.

2.8. How can I import my data aka how do i create a new database instance aka how to shred an XML file?

There are 2 ways:

- If your input data is an XML file, you can use the 'New' command. BaseX will create a new database with the name of your XML file. (GUI only)
- You can use the create command. (GUI and console)

Chapter 4. General Information

1. Location of BaseX Files

The databases are stored in a BaseXData directory in your home directory. The path can be changed in the GUI via the Options/Preferences menu. The two configuration files .basex/.basewin are stored in the same directory. This path can't be changed by the user.

2. How do you measure the performance of the queries?

The measurements include parsing, compilation, evaluation and printing time of a query. There are different ways to retrieve the performance info:

- Console: use "-v" flag as command line argument
- Console Mode: enter set info or set info all in the console mode
- GUI Mode: display performance results in the QueryInfo view

3. What indexing techniques are available and what do they do?

Indexes can speedup queries by magnitudes. Currently, three indexes exist:

- Text Index: All text nodes are indexed to speedup XPath predicates.
- Attribute Value Index: All attributes are indexed to speedup XPath predicates.
- Full-Text Index: A full-text index is created to speedup content based queries in XPath.

Note that the indexes only speedup XPath and simple user queries. The query processor will optimize queries completely automatic whenever possible. Expect XQuery to make use of the indexes in the next official release.

4. How can i create a Database for a couple of XML files?

There are 2 ways to do this:

- GUI Mode: Open the GUI, choose File -> New -> Browse -> (choose directory)
- Console Mode: `java -cp BaseX.jar org.basex.BaseX -q "set chop; create db /path/to/collection dbname"`

Next, you can open a collection and run an XQuery as follows: for \$doc in collection("dbname") return

5. Is there a way to save large results of a query in a file?

The best way to save large results is to work on the command line and use the "-o" flag, as shown below:

```
sample.xq: for $node in doc('dbname')//etc return $node/something
BaseX-Command:
java -cp BaseX.jar org.basex.BaseX -o result.xml sample.xq
```

Chapter 5. Description of BaseX-Features

1. XQuery Implementation

All XPath optimizations are integrated in XQuery and the XPath implementation was obsolete and so it was replaced by XQuery. As XPath is a subset of XQuery, you won't actually notice many differences, but XQuery directly allows to work with several databases/documents, so you won't have to manually change the Context references. A query like...

```
for $n in 1 to 10
for $doc in doc( concat('document', $n) )
return $doc//any/path/etc
```

could then be used to access several databases in one query. You can just type all your XPath queries in the XQuery textfield.

2. XQuery Full-Text Implementation

The full-text features can be used in XPath as well as in XQuery. Again, queries will be evaluated much faster in XPath, and XQuery covers more features of the language specification. The following example queries are based on the W3C Use Cases:

```
Return all title tags which contain the word 'Usability'
//title[text() ftcontains 'Usability']
```

```
Return all authors, containing 'Marigold'
for $i score $s in //author
where $i/text() ftcontains 'Marigold'
return score="{ $s }">{ $i/text() }
```

3. Description of the Indexes

Text indexes allow a speedup of order of magnitudes for text-based queries. Here are some examples for queries which are rewritten for index access:

Text-Based Queries:

- `//node()[text() = 'Usability']`
- `//div[p = 'Usability' or p = 'Testing']`
- `path/to/relevant[text() = 'Usability Testing']/and/so/on`

Attribute Index:

- `//node()[@align = 'right']`
- `descendant::elem[@id = '1']`
- `range/query[@id >= 5]`

Full-Text Index:

- `//node[text() ftcontains 'Usability']`
- `//node[text() ftcontains 'Usebiliti' with fuzzy]`
- `//book[chapter ftcontains ('web' ffor 'WWW' without stemming) ftand 'diversity' case sensitive with stemming distance at most 5 words]`

The full-text index is optimized to support all full-text features of the XQuery Full-Text recommendation. BaseX extends the specification by offering a fuzzy match option. Fuzzy search is based on the Levenshtein algorithm; the longer query terms are, the more errors will be tolerated. Default "Case Sensitivity", "Stemming" and "Diacritics" options will be considered in the index creation. Consequently, all queries will be sped up which use the default index options.

4. What about the different indexes BaseX offers: which data structures are applied?

- Text Index: B-Tree: Both the text and attribute index are based on a B-Tree and support fast exact and range queries.
- Full-Text Index (Fuzzy Version): The standard full-text Index is implemented as sorted array structure. It is optimized for simple and fuzzy searches.
- Full-Text Index (Trie Version): A second full-text Index is implemented as a compressed trie. Its needs slightly more memory than the standard full-text index, but it supports more features, such as full wildcard search.

5. Do you support XML Updates?

BaseX provides internal commands to perform updates on XML (see also BaseX Commands). Moreover, the GUI provides a convenient way to perform updates on the data.

6. What can i do if i get Exceptions on huge XML files as input?

You can try to use the internal XML parser of BaseX instead of Java's own XML parser. To do this there are 2 ways: Moreover, the GUI provides a convenient way to perform updates on the data.

- Console Mode: `basex -q "set intparse; ... create db ..."`
- GUI Mode: File -> New -> Parsing -> Use internal XML parser

7. Can i protect a DB in BaseX with a password?

No, there is no password protection inside the BaseX architecture - but (at least in Linux) you can restrict the BaseX database directory to a specific user or group. Moreover, the GUI provides a convenient way to perform updates on the data.

Chapter 6. Developer's Guide

1. Programming with BaseX API

Your Text

2. Programming with XML:DB API

Your Text

3. Programming with XQuery API

Your Text