Split a giant XML file in few simple steps

Learn how to examine big XML files a fragment at time and how to split them in a smaller pieces.



Sometimes, working with XML data, there is the need to inspect a giant file that does not fit in memory of the computer or, if it does it, slows down the system. In fact, being the XML a sort of data file with meta informations, it is well suited to be shredded in pieces, taking on only the fragment of information we are looking for.



2 Decide if you want to extract information or browse it

If you want only navigate the data without creating any file, uncheck 'Extract Documents'. Even if you don't extract the XML fragments into smaller file, QXmlEdit scans the file and records the information for an interactive review.

Parameters		
Source file path:	/home/lbell/Desktop/test xml/standard.xml	
Split documents using this path	/site/regions/africa/item/n	e.g.: /element/child

Extract all the documents	
Extract a range from: 1	🗘 to: 50000 🗘
xtract a range from: 1	🕻 to: 50000 🗘

4 How many fragments?

Selecting a range permits to extract only a small subset of the original file. No wasted space on your disk.

	Welcome to QXmIEdi	it
ou can choose a task clicking the icon:		
Examine a XML Schema (XSD)	Split a XML file	Choose a recent file
		eftail verifice*1.0*
Examine a XML Schema file showing it in a graphical editor. You can print the schema or export it as an image. ecent files list:	Split a XML file in smaller fragments or simply browse it, a fragment at time.	Select one of the recent files shown in the list.
/qxmledit/xsd/test/data/test_split_bad.xml /gxmledit/Desktop/pom.xml		
/qxmledit/Desktop/pom.xml /qxmledit/xsd/data/test/pom.xml		
/qxmledit/Desktop/pom.xml		
/qxmledit/Desktop/pom.xml /qxmledit/xsd/data/test/pom.xml		
/qomledidDe sidopipom.xmi gimmelditissdidate sigom.xmi /qomleditissdidata/test/maven-4.0.0.xsd	ctivate this option using the preferences dialog)

1 Start QxmlEdit

When starting the program will open a welcome dialog leading to the most common operations. Choose "Split a file". If you choose to hide the dialog next time the program restart and change your mind, you can reactivate it in the configuration panel.

5 Decide the output folder and the naming

Browse to the file to examine, the insert the XPath that identifies a fragment. Insert a string

of the form

/ROOT/DATA and so on.

Enter the location where write the extracted fragments and decide if you want to create a folder each N files. Decide also how to name the files and the folders. The combo boxes have some predefined values to ease the task, like a timestamp, or a progressive counter.

14	encoding:UTF8
	Add Child
	<u>D</u> elete
	A Move Up
	Hove Down
	Current fragment
1	G0 _ ↔ 550
¢	

Extract in the following di	rectory: //tmp/y									
✓ Make subfolders									View in brow	ser
each: 50	files									
Subfolder name pattern	%date%	~	_	~	%time%	×	_	~	%number%	~
File name pattern	%date%	~	%time%	~	_	~	%number%	×		~

6 Go and examine data

After the start of the operation, in the the main window a navigation box appears. The fragments found in the input file are accessible directly with a random access using the informations collected in the previous phases. If you choose to split the file in fragments, you can examine them in the

Source file path:			nl/stand:	ard.xml					
plit documents using this path	/site/regions	africalitem						e.g.: /element	child
Extract all the documents									
Extract a range from: 1	0 10	50000	0						
Extraction									
Extract in the following directory:	/tmp/y								
								View in brow	ser
 Make subfolders 									
each: 50 O	ties								
eacir. 00 V	neo								
Subfolder name pattern %date	% v		~	%tme%	~		~	%number%	~
	16 V	%time%	~		~	%number%	~		~
File name pattern %date									

Extraction