Regex Exercise: Convert the text of a voyage narrative into XML

Consult the following resources as you work with Regular Expressions:

- Our newtFire tutorial on Autotagging with Regular Expressions (Regex)
- Regular-Expressions.info Tutorial: a mine of helpful detail on regular expression matching,

Your challenge is to up-convert to XML the complete plain text file of *A VOYAGE round the WORLD* by Georg Forster, using the Find and Replace window . This file is quite large, so autotagging using Regular Expressions (regex) is really the only option we have to make this text into an XML document. Begin by downlading the text file and opening it in <oXygen/>. Use the Find and Replace window in <oXygen/> to autotag the document, and consult our Guide to Autotagging with Regular Expressions and the Regular Expressions Quick Start tutorial as you work.

Record each step of your process carefully, in a separate plain-text file. This plain text file is what you will submit for your homework. Record step-by-step your global Find-and-Replace operations with Regular Expressions in oXygen. Your goal is to produce an XML document like our model XML file but even if you have have trouble, what is most important is that you document the steps you took.

Your XML markup should accomplish the following:

1. Indicate the structure of the file by marking book divisions, chapter divisions, and paragraphs. (You do not necessarily want to do this in that order! You might want to start "from the inside out", with the paragraphs first, and then work your way up.) Think about a strategy that makes sense to you to help you match the distinctive patterns that designate the structure of this document.
2. Tag the dates, at least the dates that are sitting in square brackets. Ideally, you should remove any pseudo-markup around them.

Your complete text should look like our model, only you could go one better by removing the pseudo-markup (the brackets) around the dates. Can you locate and tag more dates than those in the square brackets?

Upload two files on Courseweb for this exercise:

1. a plain-text file in which you recorded your steps, and
2. your end result: the XML file you have created.