What will it take to get (end user) XML editors that people will use?

Norman Walsh
MarkLogic Corporation
2 August 2011

What will it take to get (end user) XML editors that *don't suck?*

Norman Walsh
MarkLogic Corporation
2 August 2011

Any questions?

A miracle

MarkLogic

But, seriously...

Let's take a survey.

Survey question: 1 of 9

Have you ever purchased something over the web?

- (•) All
- ( ) Almost all
- ( ) Most
- ( ) Some
- ( ) A few
- ( ) None

MarkLogic

Survey question: 2 of 9

Do you regularly drive an automobile, motorcycle, etc.?

- ○ All
- ◉ Almost all
- ○ Most
- ○ Some
- ○ A few
- ○ None

MarkLogic

Survey question: 3 of 9

Have you used Emacs?

- ○ All
- ○ Almost all
- ○ Most
- ◉ Some
- ○ A few
- ○ None

Survey question: 4 of 9

Have you used that other editor?

- ○ All
- ○ Almost all
- ○ Most
- ◉ Some
- ○ A few
- ○ None

MarkLogic

Survey question: 5 of 9

Have you used oXygen?

- All
- Almost all
- Most
- Some
- A few
- None

MarkLogic

Survey question: 6 of 9

Have you used Notepad?

- ○ All
- ○ Almost all
- ○ Most
- ○ Some
- ◉ A few
- ○ None

MarkLogic

Survey question: 7 of 9

Have you used Arbortext Editor?

- ○ All
- ○ Almost all
- ○ Most
- ◉ Some
- ○ A few
- ○ None

MarkLogic

Survey question: 8 of 9

Have you used a recent version of Microsoft Word?

- ○ All
- ◉ Almost all
- ○ Most
- ○ Some
- ○ A few
- ○ None

MarkLogic

Survey question: 9 of 9

Have you commented on a weblog?

- ○ All
- ◉ Almost all
- ○ Most
- ○ Some
- ○ A few
- ○ None

What do we really mean?

What will it take to get (end user) XML editors that people will use?

- What do we mean by XML?
- What do we mean by editor?
- What do we mean by people?

MarkLogic

What do we mean by XML?

```
<purchase/>
```


MarkLogic

What do we mean by XML?

```
<purchase/>
<drive/>
<emacs/>
```

Sidebar

- Edit only well-formed XML documents?
- Edit any XML fragment?
- Edit XML-like documents that aren't well-formed or are momentarily not well-formed?
  - Is `~/.bashrc` or `c:\autoexec.bat` a not-well formed XML file?

What do we mean by XML?

```
<purchase>all</purchase>
<drive>almost</drive>
<emacs>some</emacs>
```


What do we mean by XML?

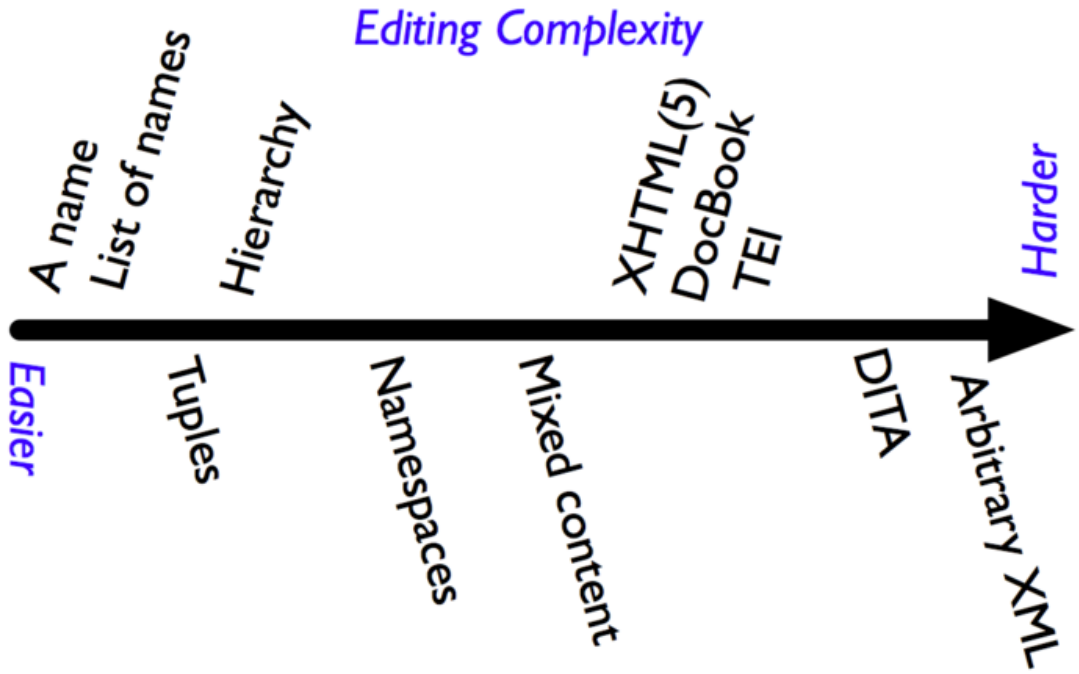
```
<survey>
  <purchase>all</purchase>
  <drive>almost</drive>
  <emacs>some</emacs>
  …
```


What do we mean by XML?

- Namespaces …
- Mixed content …
- HTML, DocBook, TEI, DITA, …
- …
- Arbitrary XML

MarkLogic

What do we mean by XML?



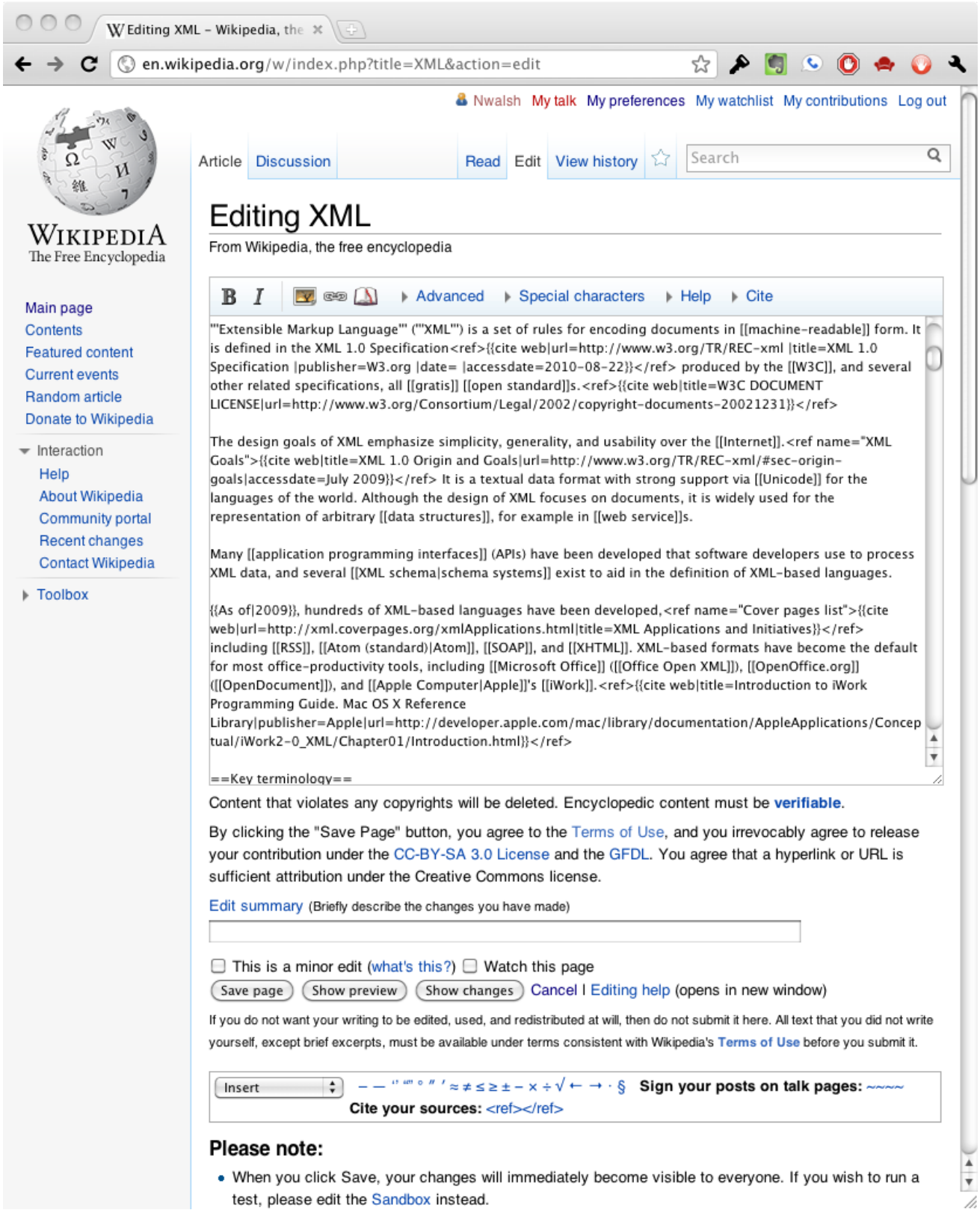
MarkLogic

What do we mean by editor?

Survey says!

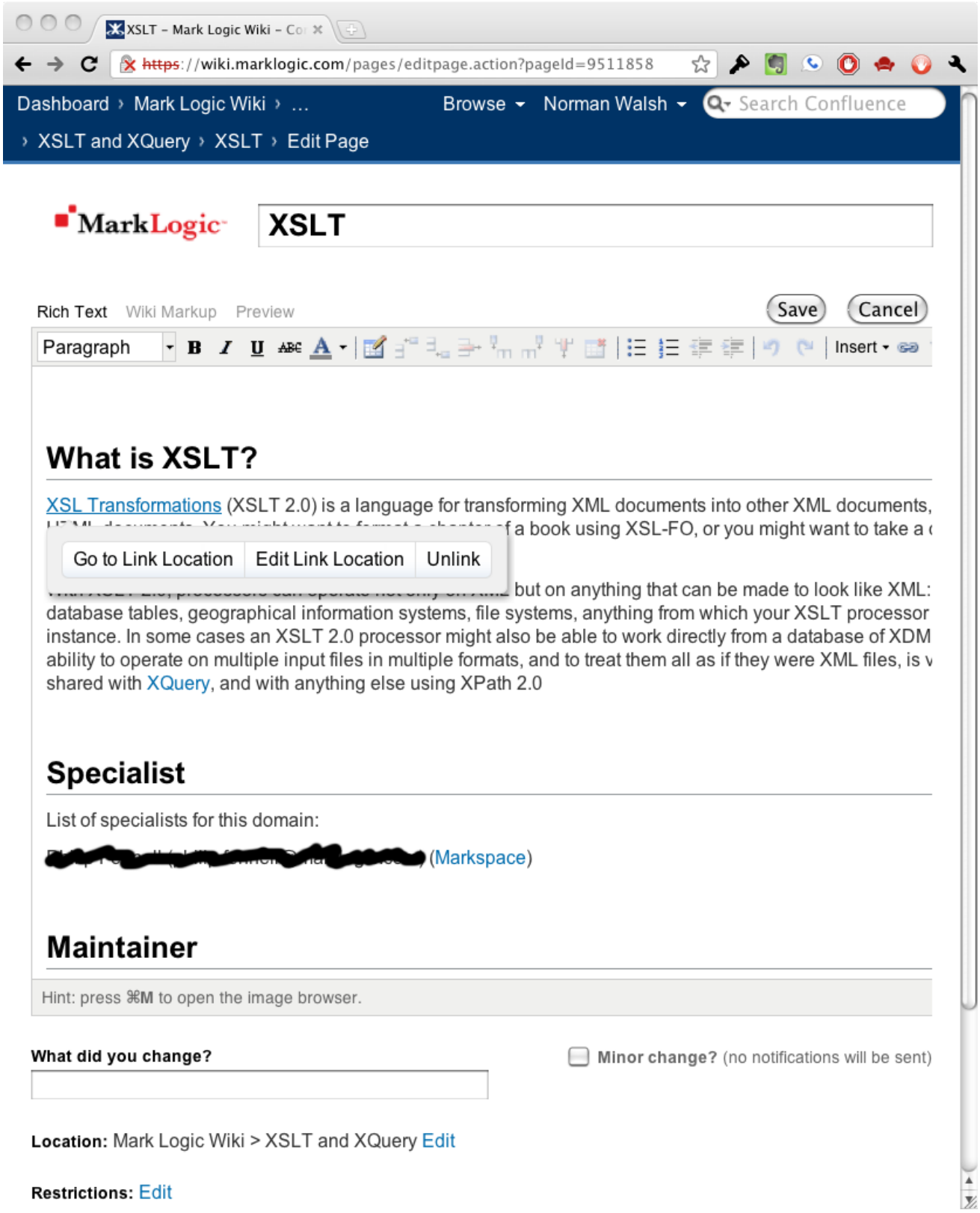
```
<survey>
  <purchase>all</purchase>
  <drive>almost</drive>
  <emacs>some</emacs>
  <other>some</other>
  <oxygen>most</oxygen>
  <notepad>few</notepad>
  <epic>some</epic>
  <msword>almost</msword>
  <comment>almost</comment>
</survey>
```


What do we mean by editor?



MarkLogic

What do we mean by editor?



MarkLogic

What do we mean by editor?

ch01.xml

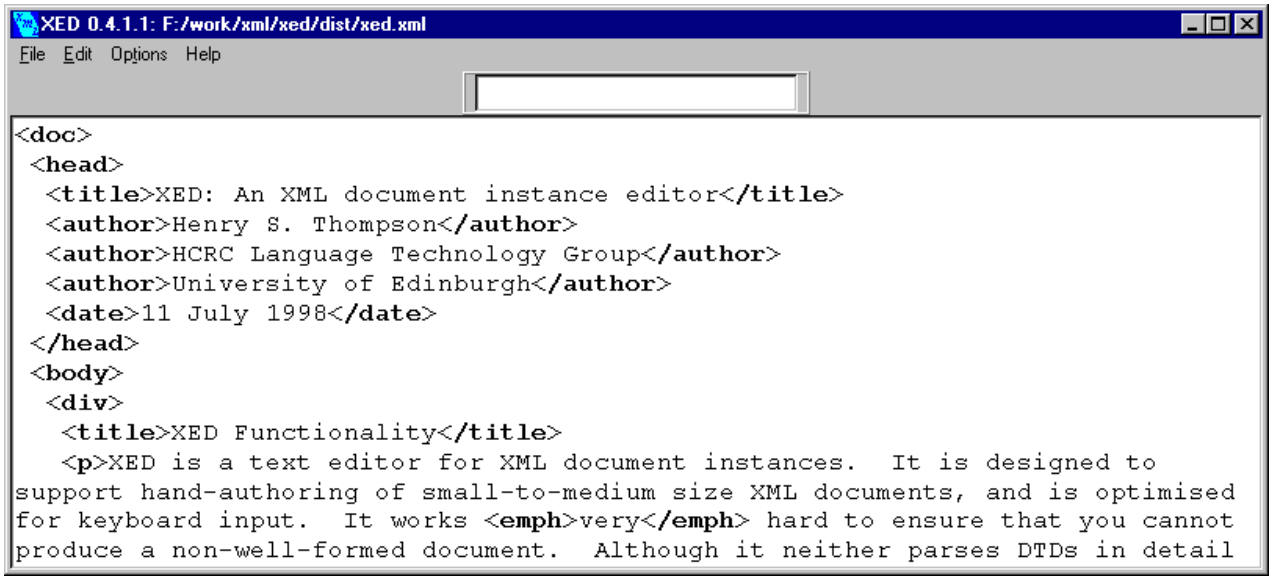
```
<chapter xmlns="http://docbook.org/ns/docbook"
         xmlns:cx="http://xmlcalabash.com/ns/extensions"
         xmlns:xi="http://www.w3.org/2001/XInclude"
         xmlns:xlink="http://www.w3.org/1999/xlink"
         xmlns:p="http://www.w3.org/ns/xproc"
         version="5.0-xproc" xml:id="introduction">
<info>
<title>Introduction</title>
</info>

<para>In this chapter, we'll fill in some background about pipelines
and get you up to speed on the concepts before we dive into the actual
features of XProc. We're going to assume that you're familiar with
XML: you're comfortable the vocabulary of tags and attributes, parsing
and validation, and that you've used an XML application or two: XSLT,
for example.</para>

<section xml:id="whatis">
<title>What's a pipeline?</title>

<para>At a high level of abstraction, a pipeline is what you get
whenever the output of one process feeds into the input of another.
If you work with XML, you've probably already used pipelines
even if you didn't think about them in those terms. If you validate
a document and then transform it, or if you apply XInclude and then
validate, or if you use XSLT to run two differen transformations,
those are all simple "pipeline" operations.</para>

<para>In another context, pipelines are a common feature of Unix
command lines, for example:</para>

<screen>cat ch01.xml | grep "&lt;para" | wc -l</screen>

<para>What that says is:</para>

<orderedlist>
<listitem>
<para>Run the <command>cat</command> command over
<filename>ch01.xml</filename>. That will display the contents of the file.
</para>
</listitem>
<listitem>
<para>The "|" symbol means that instead of displaying those lines in
the shell window, that output will become the input to the
<command>grep</command> command. What <command>grep</command> will do
is display all the lines that contain "<literal>&lt;para</literal>"
and discard all the rest.</para>
</listitem>
<listitem>
<para>Finally, with another "|", the output of <command>grep</command> will
become the input to <command>wc</command>. The <command>wc</command> command
counts the words in a file, but we gave it the <option>-l</option> option, so
it will count the lines instead.</para>
</listitem>
</orderedlist>

<para>What this pipeline does is print the number of paragraphs in the
document:</para>

<screen>??FIXME:??</screen>

<para>By using a pipeline we've taken three useful utilities and
composed them together to create a pipeline that will count the
approximate number of paragraphs we've used in all the XML documents
in the current directory. I say approximately because those are all line-based
```

-&U:@--- ch01.xml Top L1 SVN-17 [(nXML Valid yas)]----------

Mark set

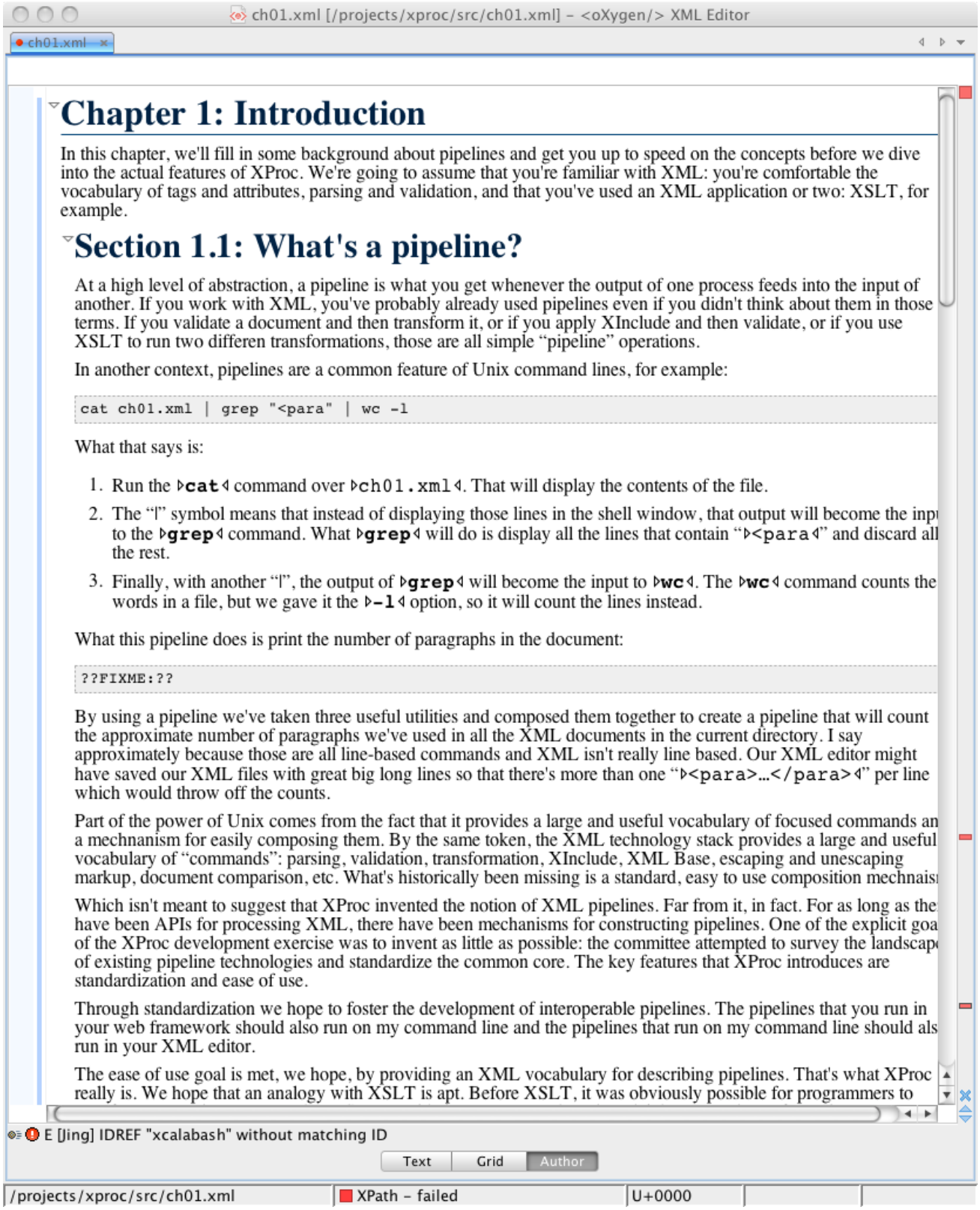
MarkLogic

What do we mean by editor?



MarkLogic

What do we mean by editor?

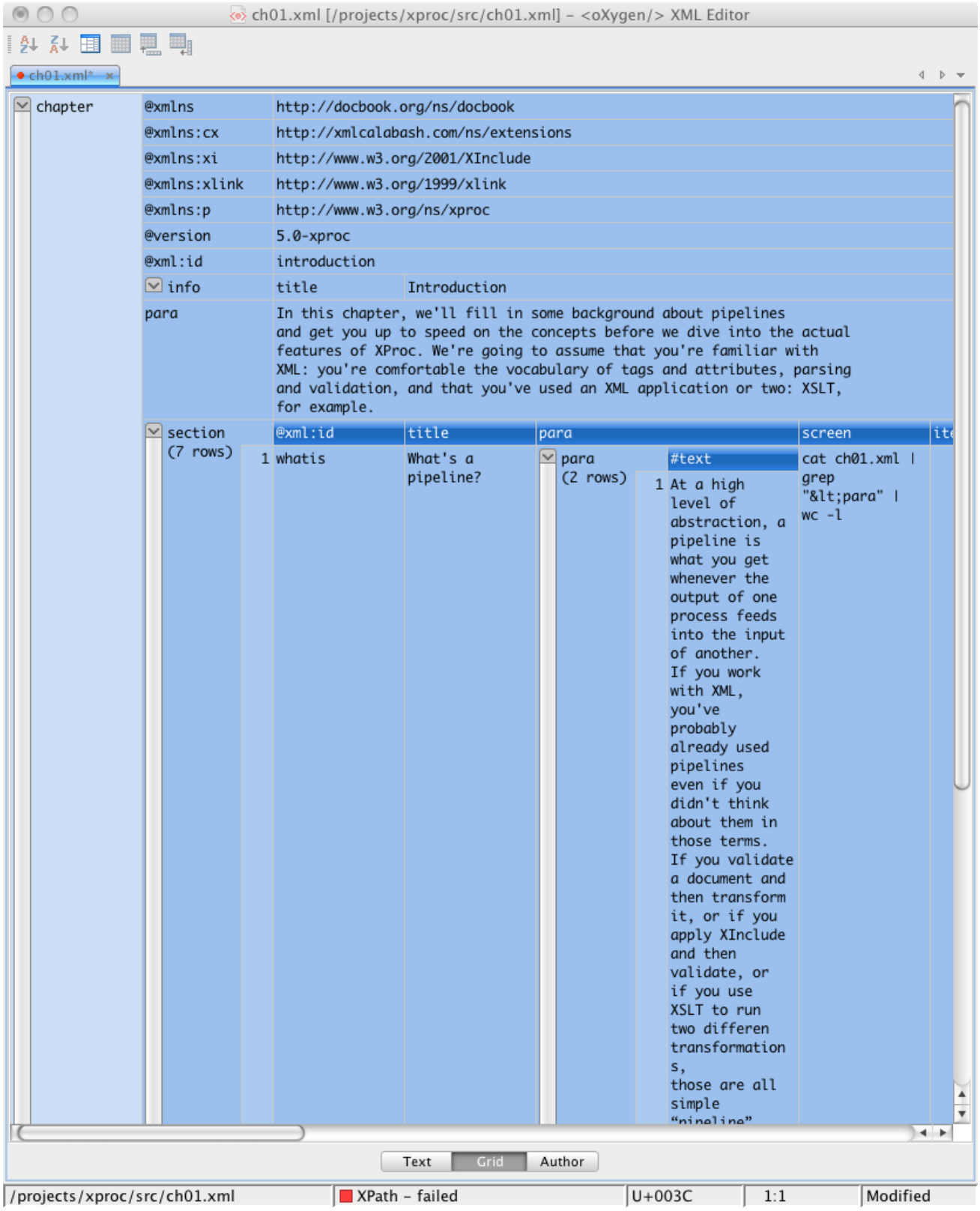


MarkLogic

What do we mean by editor?

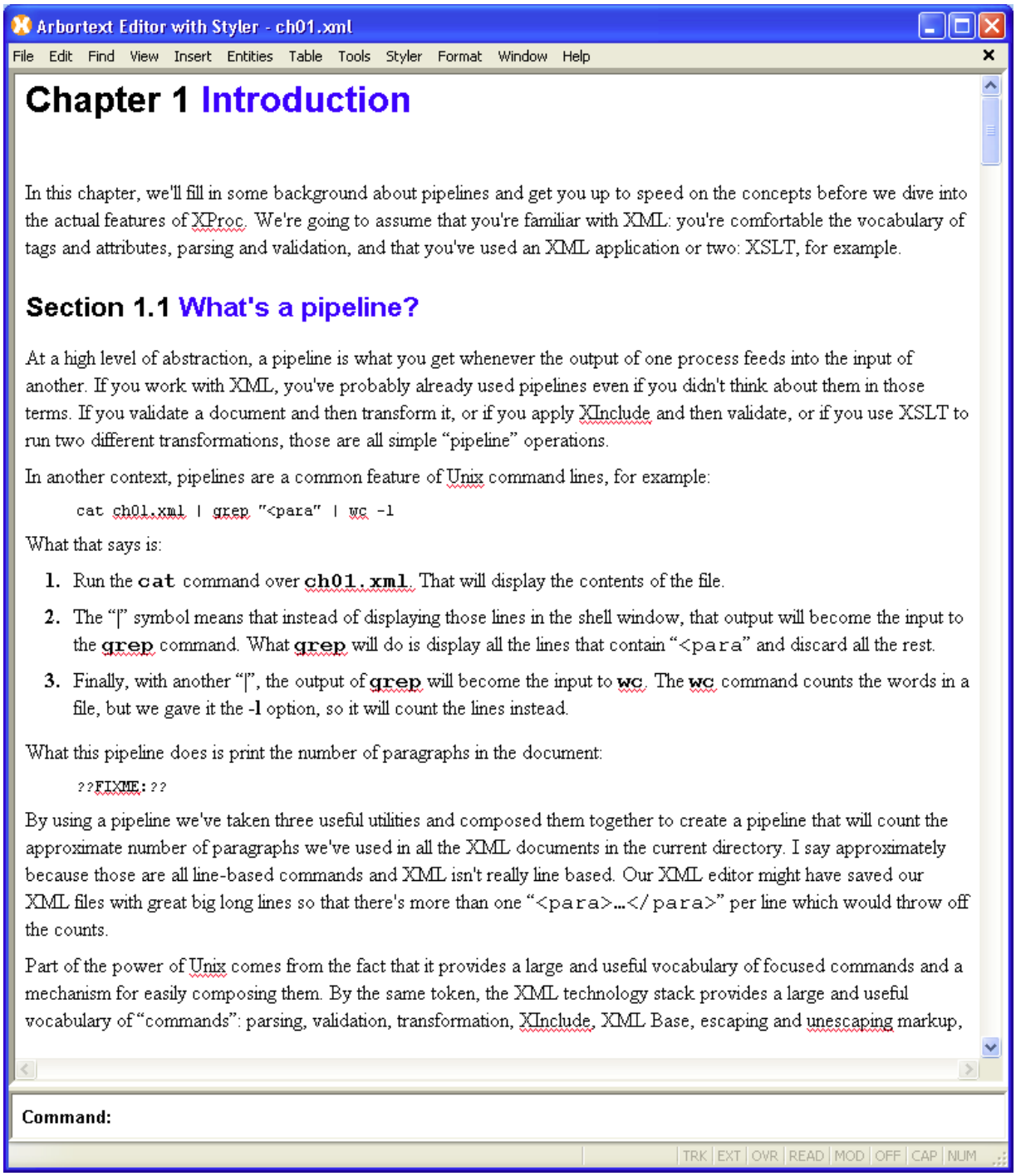
MarkLogic

What do we mean by editor?



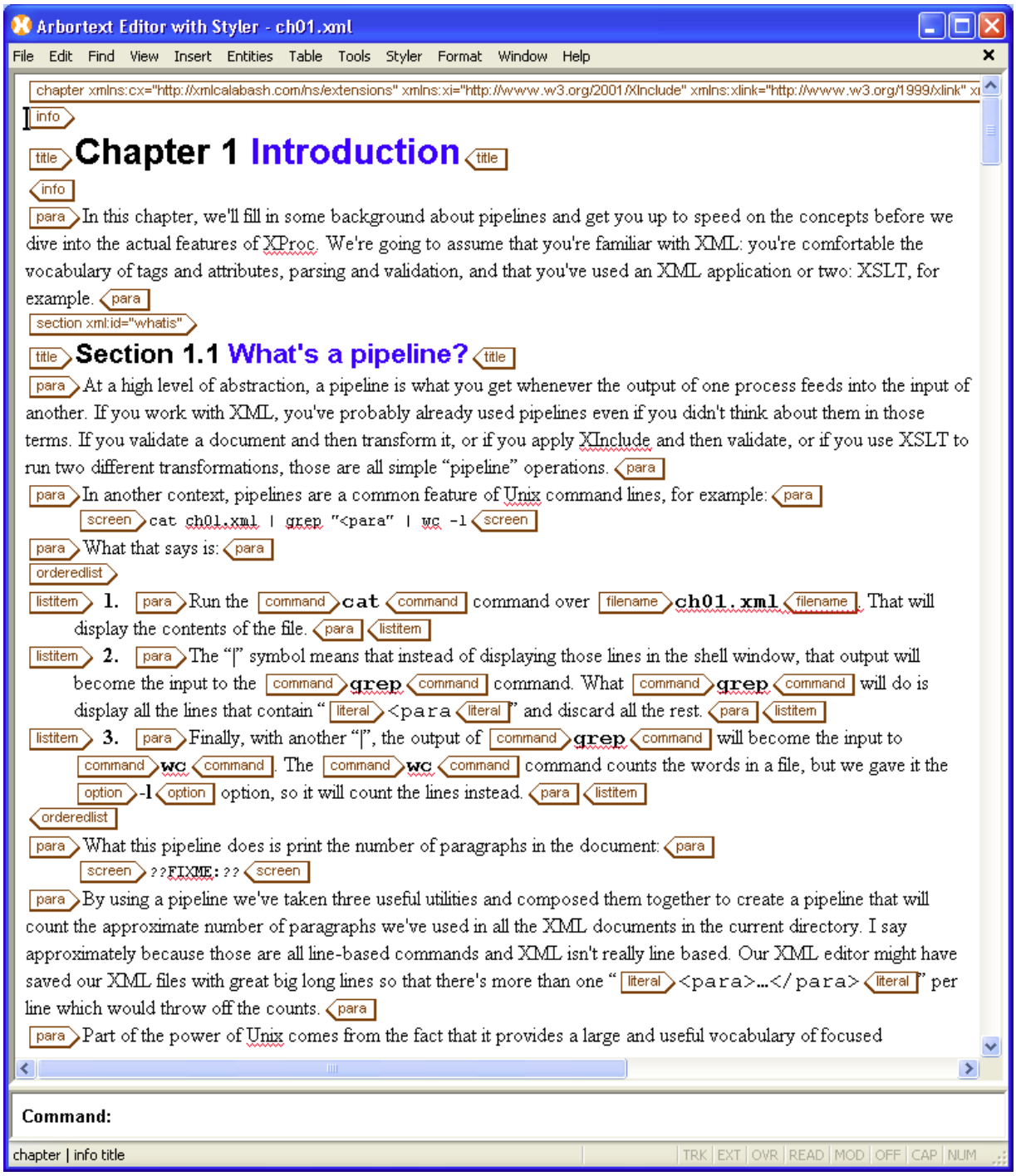
MarkLogic

What do we mean by editor?



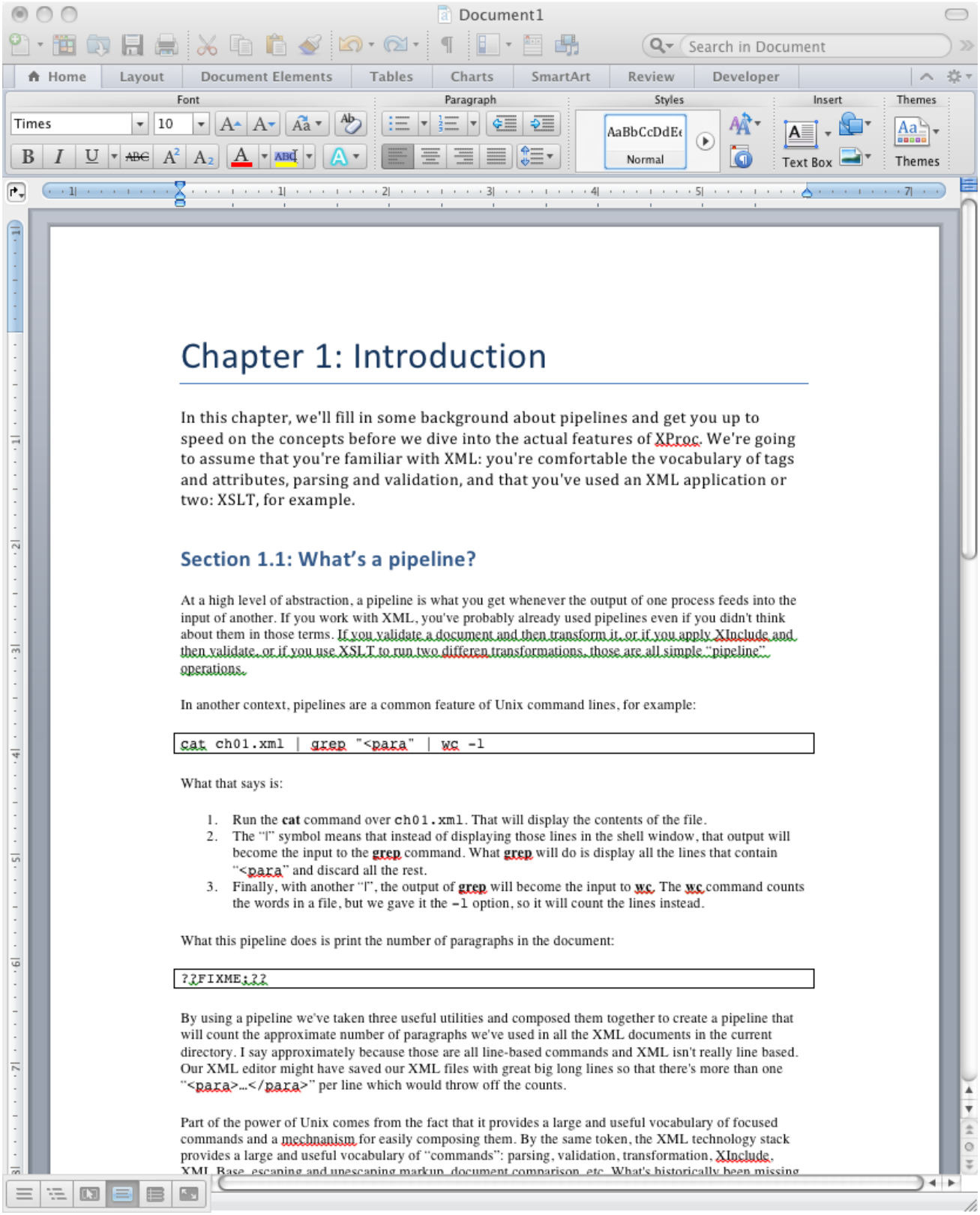
MarkLogic

What do we mean by editor?



MarkLogic

What do we mean by editor?



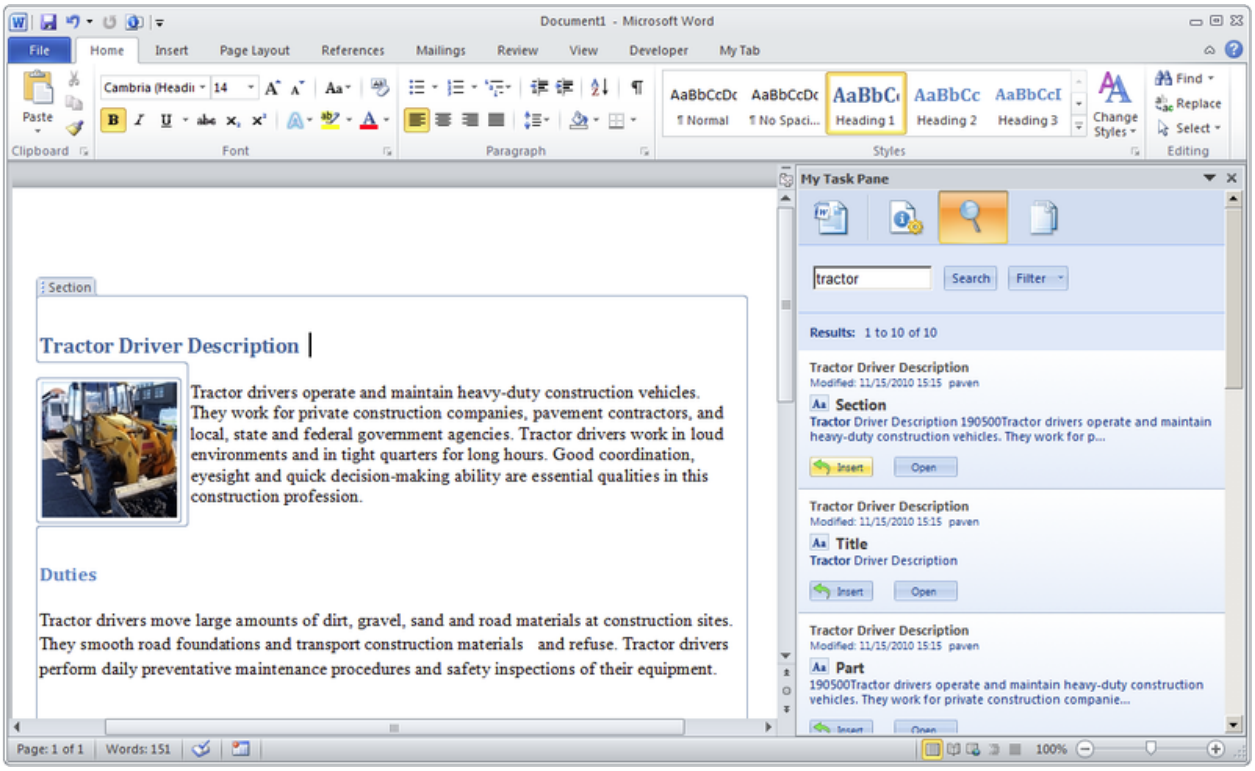
MarkLogic

Aside: WordML

Or, “is this what we mean by XML!?”

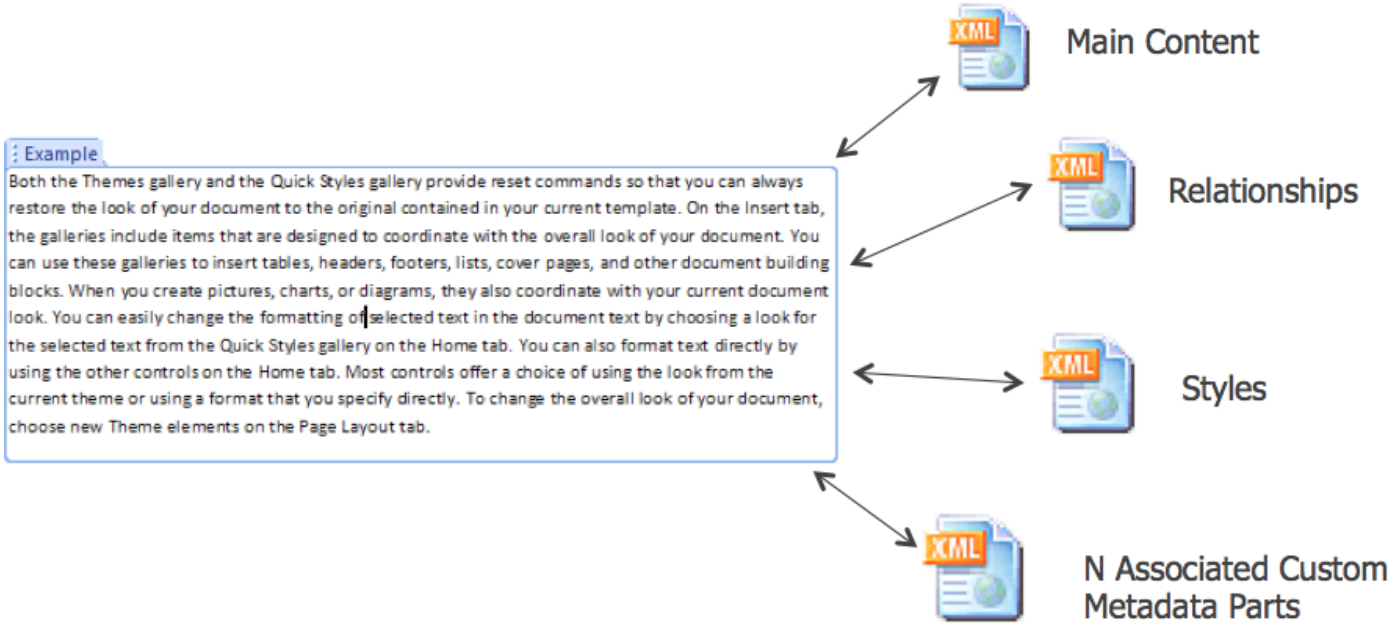
MarkLogic

MarkLogic Toolkit for Word



MarkLogic

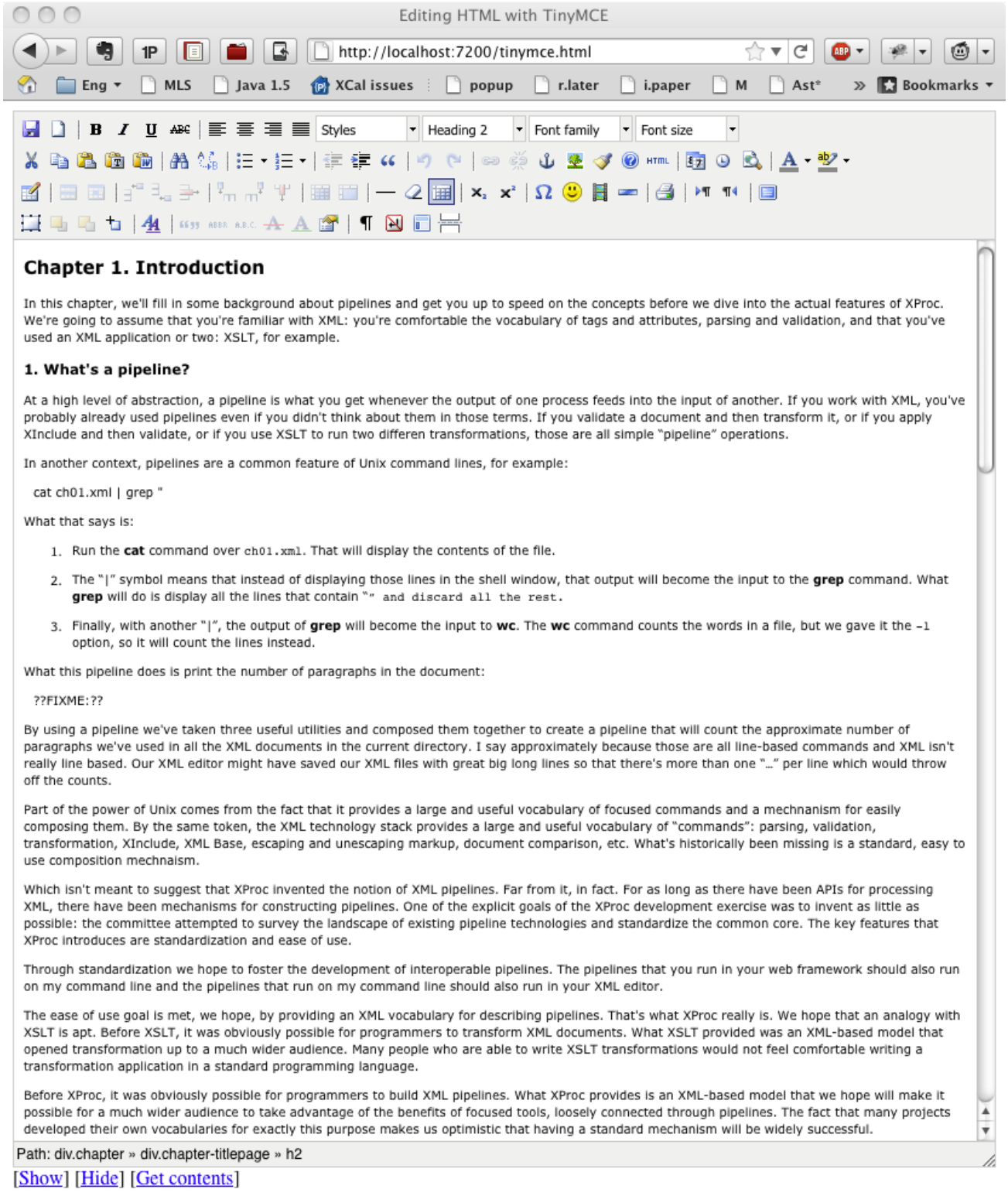
Word Components



Try it yourself

MarkLogic

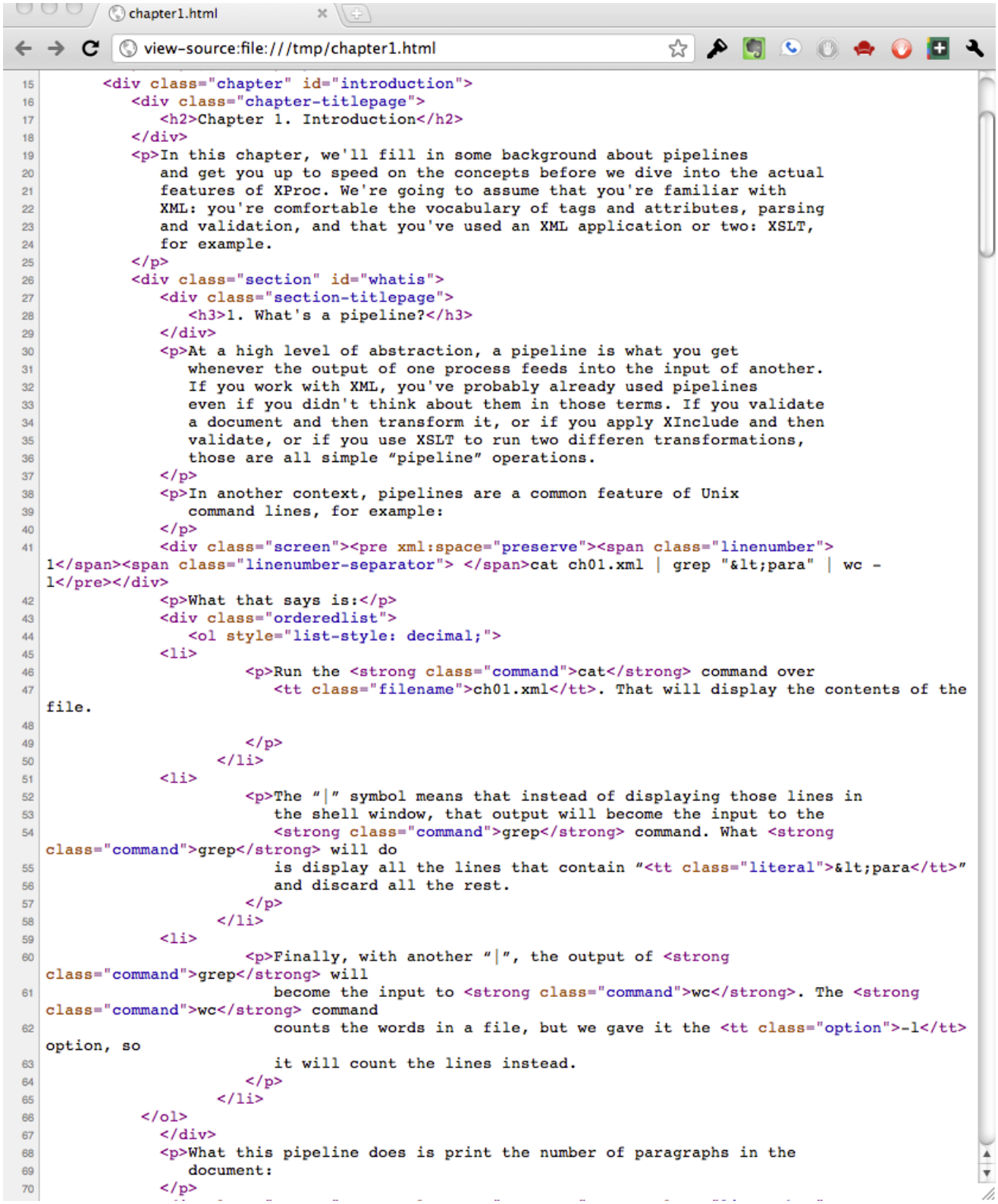
What do we mean by editor?



MarkLogic

Aside: (X)HTML(5)

Or, “is this what we mean by XML!?”




```
      <div class="chapter" id="introduction">
         <div class="chapter-titlepage">
            <h2>Chapter 1. Introduction</h2>
         </div>
         <p>In this chapter, we'll fill in some background about pipelines
            and get you up to speed on the concepts before we dive into the actual
            features of XProc. We're going to assume that you're familiar with
            XML: you're comfortable the vocabulary of tags and attributes, parsing
            and validation, and that you've used an XML application or two: XSLT,
            for example.
         </p>
         <div class="section" id="whatis">
            <div class="section-titlepage">
               <h3>1. What's a pipeline?</h3>
            </div>
            <p>At a high level of abstraction, a pipeline is what you get
               whenever the output of one process feeds into the input of another.
               If you work with XML, you've probably already used pipelines
               even if you didn't think about them in those terms. If you validate
               a document and then transform it, or if you apply XInclude and then
               validate, or if you use XSLT to run two differen transformations,
               those are all simple “pipeline” operations.
            </p>
            <p>In another context, pipelines are a common feature of Unix
               command lines, for example:
            </p>
            <div class="screen"><pre xml:space="preserve"><span class="linenumber">
1</span><span class="linenumber-separator"> </span>cat ch01.xml | grep "&lt;para" | wc -
l</pre></div>
            <p>What that says is:</p>
            <div class="orderedlist">
               <ol style="list-style: decimal;">
            <li>
                     <p>Run the <strong class="command">cat</strong> command over
                        <tt class="filename">ch01.xml</tt>. That will display the contents of the
file.

                     </p>
                  </li>
            <li>
                     <p>The “|” symbol means that instead of displaying those lines in
                        the shell window, that output will become the input to the
                        <strong class="command">grep</strong> command. What <strong
class="command">grep</strong> will do
                        is display all the lines that contain “<tt class="literal">&lt;para</tt>”
                        and discard all the rest.
                     </p>
                  </li>
            <li>
                     <p>Finally, with another “|”, the output of <strong
class="command">grep</strong> will
                        become the input to <strong class="command">wc</strong>. The <strong
class="command">wc</strong> command
                        counts the words in a file, but we gave it the <tt class="option">-l</tt>
option, so
                        it will count the lines instead.
                     </p>
                  </li>
          </ol>
            </div>
            <p>What this pipeline does is print the number of paragraphs in the
               document:
            </p>
```


MarkLogic

What do we mean by editor?

Or, “what was that driving question all about?”

Photo credit: Doug Wertzman

MarkLogic

What do we mean by people?

- You and me?
- Professional writers?
- Engineers?
- Interns?
- Managers?

What do we mean by people?

- XML experts
  - I'll see your CDATA section and raise you an external parsed entity
- XML users
  - I'll see your empty element and raise you a processing instruction
- Users being presented with XML for the first time
  - What's an XML?

XML experts

- Understand the separation of content and presentation
- Believe in the value of structured markup
- Can leverage XML to deliver value

What editing problem?

XML users

- Understand the separation of content and presentation
- Have been told about the value of structured markup
- May be seeing some value from structured markup
- May feel threatened by the loss of control

Possibly comfortable with oXygen/Arbortext Editor/etc.

First time XMLers

- May have been told about the separation of content and presentation
- May not understand the value of structured markup
- May be threatened or frightend by the prospect of changing editing tools
- Probably think in terms of document presentation

Think everyone uses Microsoft Word.

Aside: presentation

- "Probably think in terms of document presentation"

The presentation problem

Fact:

- Many users consider their branded output as primary, XML as secondary

Reality:

- Rendering arbitrary XML to high-quality print is hard
- "Lights out" formatting is fundamentally different than "by hand" layout
- Automatically rendering arbitrary XML to an arbitrary design is at least hard, if not impossible

The other presentation problem

- The rise of desktop publishing, and the WYSIWYG word processors that followed, have trained a generation (or more) of authors to believe that the important thing about a document is how it looks
- Authors believe that they add value by controlling the appearance of the document

A tale of two engagements

In which the presenter rambles on a bit

What will it take...

Education

MarkLogic

Editing XML

- Isn't harder just because the tools suck.
- It requires authors
  - to think diffrently about content
  - to know more about the content
  - to *unlearn* a set of skills of which they may be rightfully proud and for which they have historically been rewarded
  - to accept that they have less control over the presentation
- It requires managers
  - to change the structure of rewards for authors
  - to accept that the problems are harder
  - to understand that great rewards are possible, but they're not free

Or did you mean…

What will it take to get

- An XML editor that provides a tailored user interface for any vocabulary
- With a high-quality preview and "tags off" editing
- That supports semantic markup without any additional effort
- And can be used out-of-the-box without any customizaiton

That will take

A miracle

Conclusions

- For a lot of use cases and a lot of users, we have editors that don't suck. Or at least, don't suck as much as you might at first glance think they do.
- Tools like oXygen, Arbortext Editor (Serna, Altova, etc.) work reasonably well for many kinds of users.
- If you're willing to squint at the XML a bit, tools like Word and OpenOffice work today
- If you want to do better, it's as much about education as technology

Q&A

Norman Walsh
mailto:norman.walsh@marklogic.com
tel:+1-413-624-6676
http://www.marklogic.com/

Table of contents

- Aside: (X)HTML(5)
- What do we mean by editor?
- What do we mean by people?
- What do we mean by people?
- XML experts
- XML users
- First time XMLers
- Aside: presentation
- The presentation problem
- The other presentation problem
- A tale of two engagements
- What will it take…
- Editing XML
- Or did you mean…
- That will take
- Conclusions
- Q&A

