Freie Universität Berlin

XML Übung

Willkommen

Welcome

Bienvenido

Internationalisierung...

Freie Universität Berlin

- heisst: als Informatiker brauchen Sie Englisch, so warum nicht es in der Universität auch nutzen?

- Fragen kann man auf Deutsch oder Englisch
- Antworten kann ich auch auf Deutsch oder Englisch
- Blätter sind auf Deutsch
- Folien sind teilweise Deutsch, teilweise Englisch

- Ich werde auf Englisch reden!

Struktur

Freie Universität Berlin

- Fragen zur vorherigen Vorlesung(en)
- “In Depth”: ein Thema aus der Vorlesung ins Tiefe
- Musterlösungen (wenn Übungsblatt fertig)
- Musterfragen
- XML Tools: einige Screenshots aus Eclipse SDK
- “XML Extra”: praktische Beispiele von XML Verwendung

Übungsplan

Freie Universität Berlin

Datum	Übungsblatt	Übung	Musterlösung
23.4.	1	1	
30.4.	2		
7.5.	3	2	1 & 2
14.5.		3	3
21.5.	4	4	
28.5.	5	5	4
18.6.	6	6	5
25.6.	7		
2.7.		7	6 & 7

Übung 1

Struktur

Freie Universität Berlin

- Fragen zur vorherigen Vorlesung
- In Depth: XML Syntax, Namensräume
- Musterfragen
- XML Extra: Meine erste XHTML Webseite

In Depth:

XML Syntax, Namensräume

XML-Deklaration

Freie Universität Berlin

<?xml version="1.0" encoding="UTF-8"?>

Attribut version

- verwendete XML-Version: "1.0" oder "1.1"
- obligatorisch

Attribut encoding

- Kodierung der XML-Datei
- optional (default: UTF-8)

Attribut standalone

- Gibt an, ob es eine zugehörige DTD oder ein XML-Schema gibt ("no") oder nicht ("yes").
- optional

Beachte: immer in dieser Reihenfolge!

Wie kann die Bedeutung von XML-Elementen festgelegt werden?

Freie Universität Berlin

- durch Zuordnung des Element-Namens zu einem Namensraum
- Namensraum wird mit einer URI identifiziert:

  z.B. http://www.w3.org/1999/xhtml
- zwei Möglichkeiten:

1. expliziter Namensraum-Präfix
   - zuerst: xmlns:myns="http://www.w3.org/1999/xhtml"
   - dann: z.B. <myns:p>
   - Wahl des Präfixes (ziemlich) egal!
2. Standard-Namensraum
   - zuerst: xmlns="http://www.w3.org/1999/xhtml"
   - dann: z.B. <p>

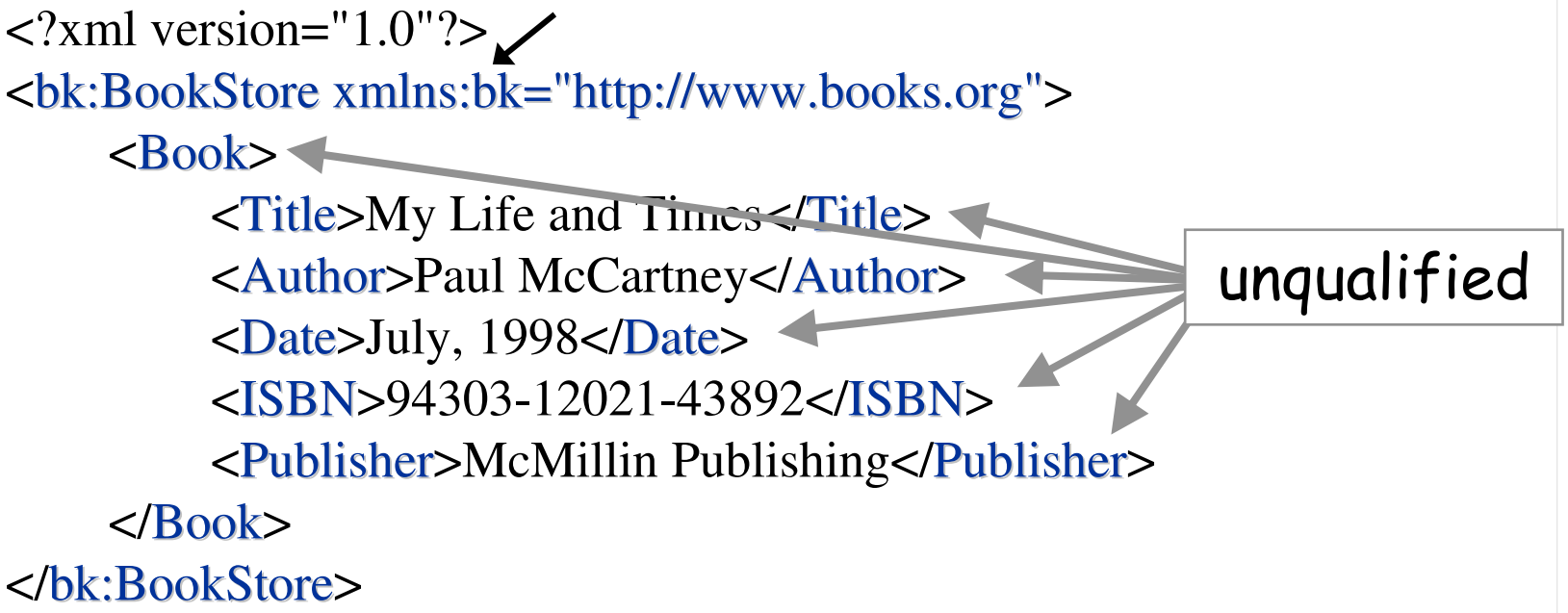
Welche Element-Namen sind qualified?

Freie Universität Berlin

```
<?xml version="1.0"?>
<BookStore xmlns="http://www.books.org">
    <Book>
        <Title>My Life and Times</Title>
        <Author>Paul McCartney</Author>
        <Date>July, 1998</Date>
        <ISBN>94303-12021-43892</ISBN>
        <Publisher>McMillin Publishing</Publisher>
    </Book>
</BookStore>
```

- alle Element-Namen (einschl. BookStore!) dem Standard-Namensraum zugeordnet
- alle Element-Namen daher namensraumeingeschränkt (qualified)

Welche Element-Namen sind qualified?



- hier kein Standard-Namensraum festgelegt
- bk:Bookstore: namensraumeingeschränkt (qualified)
- alle anderen Element-Namen: nicht namensraumeingeschränkt (unqualified)

Welchem Namensraum sind die markierten Element-Namen zugeordnet?

Freie Universität Berlin

```
<course xmlns="http://www.w3.org/2004/course">
  <title>Semantic Web</title>
  <lecturers>
    <name xmlns:pers="http://www.w3.org/2004/pers">
      <title>Priv.-Doz. Dr. M.S.E</title>
      <first>Steffen</first>
      <last>Staab</last>
    </name>
  </lecturers>
  <date>12/11/2004</date>
  <abstract>...</abstract>
</course>
```

- kein Namensraum-Präfix
- daher entweder *unqualified* oder einem Standard-Namensraum zugeordnet
- Standard-Namensraum ist hier "http://…/course"

Zu welchem Namensraum gehören die id-Attribute?

Freie Universität Berlin

```
<course xmlns="http://www.w3.org/2004/course">
  <title id="123">Semantic Web</title>
  <lecturers>
    <name id="999" xmlns="http://www.w3.org/2004/pers">
        <title>Priv.-Doz. Dr. M.S.E</title>
        <first>Steffen</first>
        <last>Staab</last>
    </name>
  </lecturers>
  <date>12/11/2004</date>
  <abstract>...</abstract>
</course>
```

- Attribute gehören *nicht* zum Standard-Namensraum.
- deshalb sind beide id-Attribute *keinem* Namensraum zugeordnet!

Wie kann ein Attribut dem Standard-Namensraum zuordnet werden?

Freie Universität Berlin

```
<course xmlns="http://www.w3.org/2004/course"
        xmlns:course="http://www.w3.org/2004/course">
  <title course:id="123">Semantic Web</title>
  <lecturers>
    <name pers:id="999" xmlns="http://www.w3.org/2004/pers"
                  xmlns:pers="http://www.w3.org/2004/pers">
      <title>Priv.-Doz. Dr. M.S.E</title>
      <first>Steffen</first>
      <last>Staab</last>
    </name>
  </lecturers>
  <date>12/11/2004</date>
  <abstract>...</abstract>
</course>
```

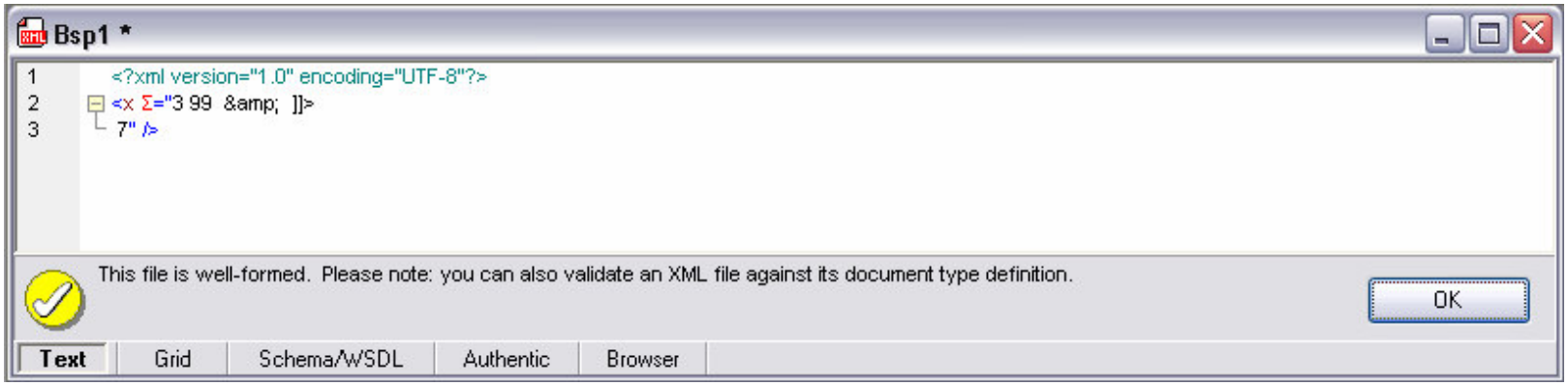
Weitere Musterfragen

Quelle: Altova

Wohlgeformt?

Freie Universität Berlin

```
<?xml version="1.0" encoding="UTF-8"?>
<x Σ="3 99  &amp;  ]]>
7" />
```



Was ist richtig? (Choose all that apply.)

Freie Universität Berlin

A. The prefix xml is by definition bound to the namespace name http://www.w3.org/XML/1998/namespace.

B. The prefix xmlns is bound to the namespace name http://www.w3.org/XML/1998/namespace.

C. The prefix xmlns is used only for namespace bindings.

D. The prefix xmlns may be arbitrarily reused for any purpose by the XML author.

Was ist richtig?

Freie Universität Berlin

A. An XML document may contain several root elements.

B. XML element names can contain spaces.

C. XML documents cannot contain any attribute which is not specified in the content model.

D. XML element and attribute names must be lower-case.

E. XML element names may begin with the underscore ('_') character.

Was ist wohlgeformt?

Freie Universität Berlin

A.
```
<x xmlns:foo="http://www.example.com"
    xmlns:bar="http://www.example.com" >
        <anchor x="1" x="2" />
</x>
```

B.
```
<x xmlns:foo="http://www.example.com"
    xmlns="http://www.example.com" >
        <anchor x="1" y="1" />
</x>
```

Was ist wohlgeformt?

Freie Universität Berlin

C.
```
<x xmlns:foo="http://www.example.com"
    xmlns:bar="http://www.example.com" >
        <anchor foo:x="1" bar:x="2" />
</x>
```

D.
```
<x xmlns:foo="http://www.example.com"
    xmlns="http://www.example.com" >
        <anchor x="1" foo:x="2" />
</x>
```

Korrekte Syntax?

Freie Universität Berlin

A. <!--Extractor Module: Utility Templates -->

B. <!--Extractor Module -- Utility Templates -->

C. <!-- Extractor Module Utility Templates -->

D. <!--Extractor Module Utility Templates --->

E. <!-- Extractor - Module - Utility - Templates-->

Was ist richtig?

Freie Universität Berlin

What is the construct for specifying to an XML processor to render text in an XML document without parsing it?

1. <PCDATA>How now, brown cow?</PCDATA>
2. <CDATA>How now, brown cow?</CDATA>
3. <![CDATA[How now, brown cow?]]>
4. <!PCDATA[How now, brown cow?]]>
5. <![#PCDATA[How now, brown cow?]]>

Was ist richtig?

Freie Universität Berlin

The pseudo-attribute in the XML prolog (<?xml ... ?>) which specifies the character encoding must be:

1. Lower case, but the value is case-sensitive.
2. Upper case, but the value is case-insensitive.
3. Lower case, but the value is case-insensitive.
4. Lower case, but the value must be in upper case.
5. Upper case, and the value must also be in upper case.

Was ist richtig?

Freie Universität Berlin

By setting the default namespace to an empty string, what would happen within the scope of the declaration?

1. The empty string becomes the namespace.
2. The empty string becomes the default namespace.
3. It has the same effect of there being no default namespace.
4. The default namespace does not change.
5. None of the above.

Using

XML

HTML: Building block of the Web

Freie Universität Berlin

- HTML is part of the success of the World Wide Web
  - Simple enough that people could write their own pages
  - Forgiving enough that people could make mistakes and the page would still be displayable
  - Expressive enough for simple layout of textual and graphical content
  - Hyperlinks as key aspect of the Web (as *hypertext system*)

Problems with HTML

Freie Universität Berlin

- As the Web evolved, so did the requirements on Web documents
  - Increasingly machine *generated* content
    - Data dynamically extracted from databases
    - Response based on individual queries or other variables
  - Increasingly machine *processed* content
    - Web content not longer only for human consumption
    - Ubiquity of Internet meant that content needs to be adapted to different devices, networks and even users

- Emergence of "Web 2.0"
  - Focus on reusability of Web content (e.g. mashups)
  - ***Screen scraping*** as less than ideal means to acquire data from HTML web pages

The solution: XHTML

Freie Universität Berlin

- The problems with HTML relate to its *unstructuredness* and its *non-integratability*
- Solution: define a XML conformant version of HTML
  - **XHTML** (Extensible HTML)
  - Promoted strongly by the World Wide Web Consortium
  - Version 1.0 available since January 2000

- Defines how HTML documents can be well formed
- Provides various DTDs for different “flavours” of HTML
  - Enables validity checking of XHTML Web pages
- Allows for mixing of vocabularies through namespaces

Example: from HTML to XHTML

Freie Universität Berlin

a. <HTML>
b. <head>
c. <title> An example </title>
d. </head>
e. <body bgcolor="#ffffff">
f. <img src="/A&P/example.gif" BORDER=0 >
g. <h2> Info </h2>
h. <P> This is some text. Isn't this cool!
i. 5 is less than 3, i.e. 5 < 3
j. </body>
k. </HTML>

Example: (1) add missing tags

Freie Universität Berlin

a. <HTML>
b. <head>
c. <title> An example </title>
d. </head>
e. <body bgcolor="#ffffff">
f. <img src="/A&P/example.gif" BORDER=0 >
g. <h2> Info </h2>
h. <P> This is some text. Isn't this cool! </P>
i. 5 is less than 3, i.e. 5 < 3
j. </body>
k. </HTML>

Example: (2) close open tags

Freie Universität Berlin

a. <HTML>
b. <head>
c. <title> An example </title>
d. </head>
e. <body bgcolor="#ffffff">
f.
g. <h2> Info </h2>
h. <P> This is some text. Isn't this cool! </P>
i. 5 is less than 3, i.e. 5 < 3
j. </body>
k. </HTML>

Example: (3) nest tags, not overlap

Freie Universität Berlin

a. <HTML>
b. <head>
c. <title> An example </title>
d. </head>
e. <body bgcolor="#ffffff">
f.
g. <h2> Info </h2>
h. <P> This is some text. Isn't this cool! </P>
i. 5 is less than 3, i.e. 5 < 3
j. </body>
k. </HTML>

Example: (4) add quotes to attribute values

Freie Universität Berlin

a. <HTML>
b. <head>
c. <title> An example </title>
d. </head>
e. <body bgcolor="#ffffff">
f.
g. <h2> Info </h2>
h. <P> This is some text. Isn't this cool! </P>
i. 5 is less than 3, i.e. 5 < 3
j. </body>
k. </HTML>

Example: (5) use escape characters

Freie Universität Berlin

a. <HTML>
b. <head>
c. <title> An example </title>
d. </head>
e. <body bgcolor="#ffffff">
f.
g. <h2> Info </h2>
h. <P> This is some text. Isn't this cool! </P>
i. 5 is less than 3, i.e. 5 < 3
j. </body>
k. </HTML>

Example: (6) all lower case

Freie Universität Berlin

a. <html>
b. <head>
c. <title> An example </title>
d. </head>
e. <body bgcolor="#ffffff">
f.
g. <h2> Info </h2>
h. This is some text. Isn't this cool!
i. 5 is less than 3, i.e. 5 < 3
j. </body>
k. </html>

Example: (7) add the XHTML namespace

Freie Universität Berlin

a. <html xmlns="http://www.w3.org/1999/xhtml">
b. <head>
c. <title> An example </title>
d. </head>
e. <body bgcolor="#ffffff">
f.
g. <h2> Info </h2>
h. This is some text. Isn't this cool!
i. 5 is less than 3, i.e. 5 < 3
j. </body>
k. </html>

Example: (8) make the document DTD valid

a. <html xmlns="http://www.w3.org/1999/xhtml">
b. <head>
c. <title> An example </title>
d. </head>
e. <body style="backgroundColor: #FFFFFF">
f.
g. <h2> Info </h2>
h. This is some text. Isn't this cool!
i. 5 is less than 3, i.e. 5 < 3
j. </body>
k. </html>

Example: (9) add the DTD declaration

Freie Universität Berlin

<!DOCTYPE html PUBLIC "..." http://...>

a. <html xmlns="http://www.w3.org/1999/xhtml">

b. <head>

c. <title> An example </title>

d. </head>

e. <body style="backgroundColor: #FFFFFF">

f.

g. <h2> Info </h2>

h. This is some text. Isn't this cool!

i. 5 is less than 3, i.e. 5 < 3

j. </body>

k. </html>

First steps to your own XML

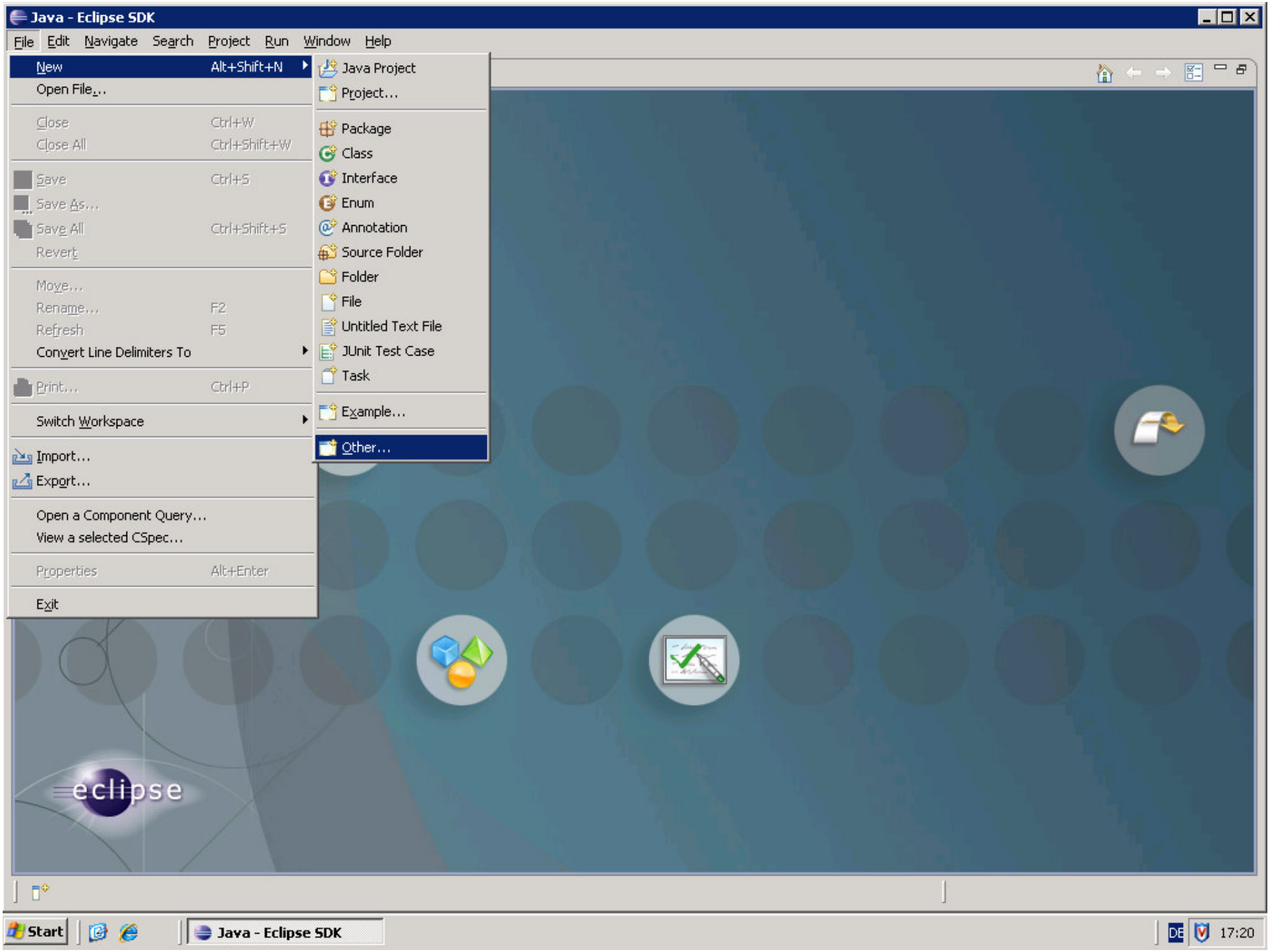
Freie Universität Berlin

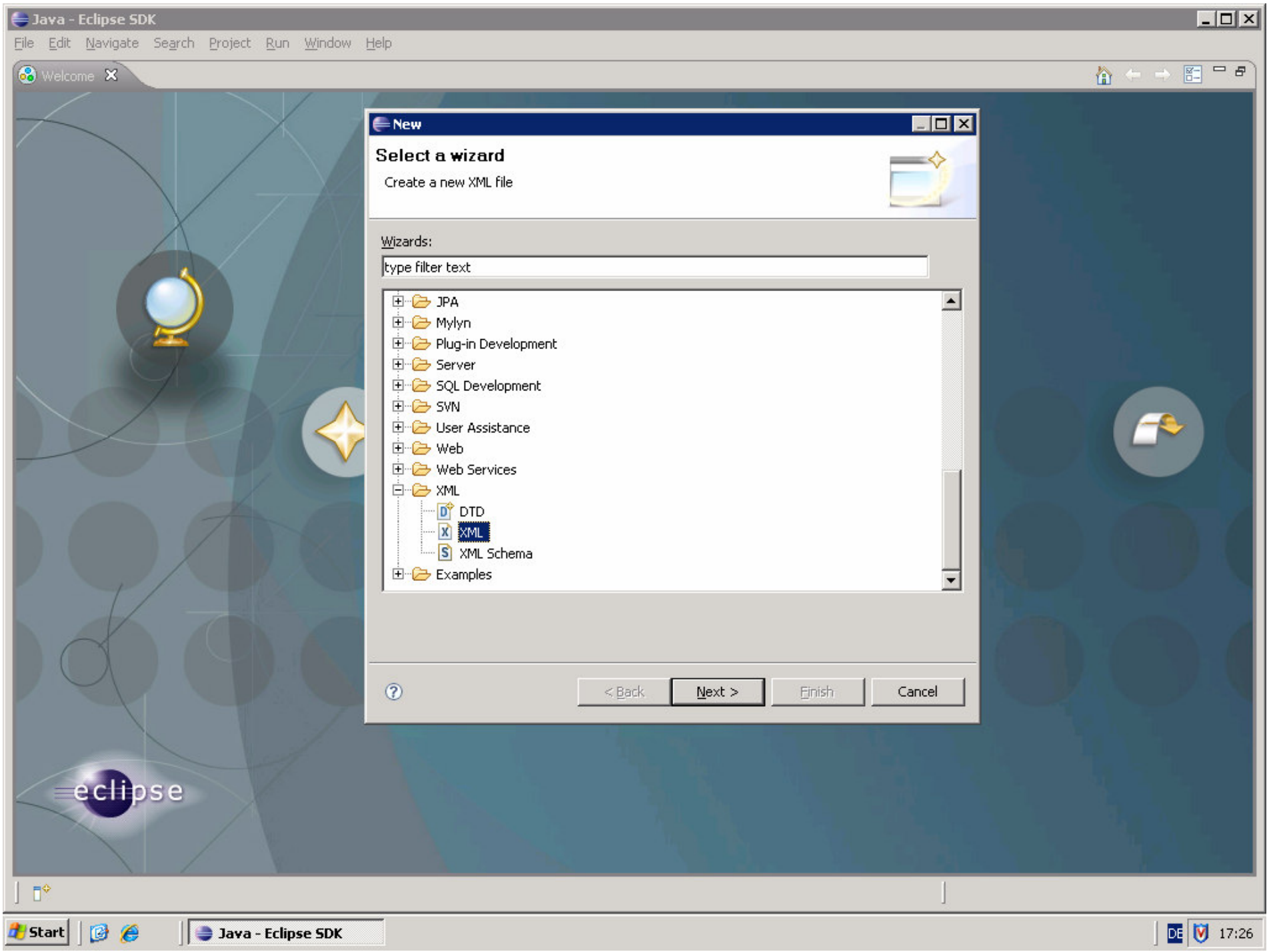
- If you have a HTML based Web page…
  - Please no MySpace etc.
  - You're Computer Scientists – work at the tag level!

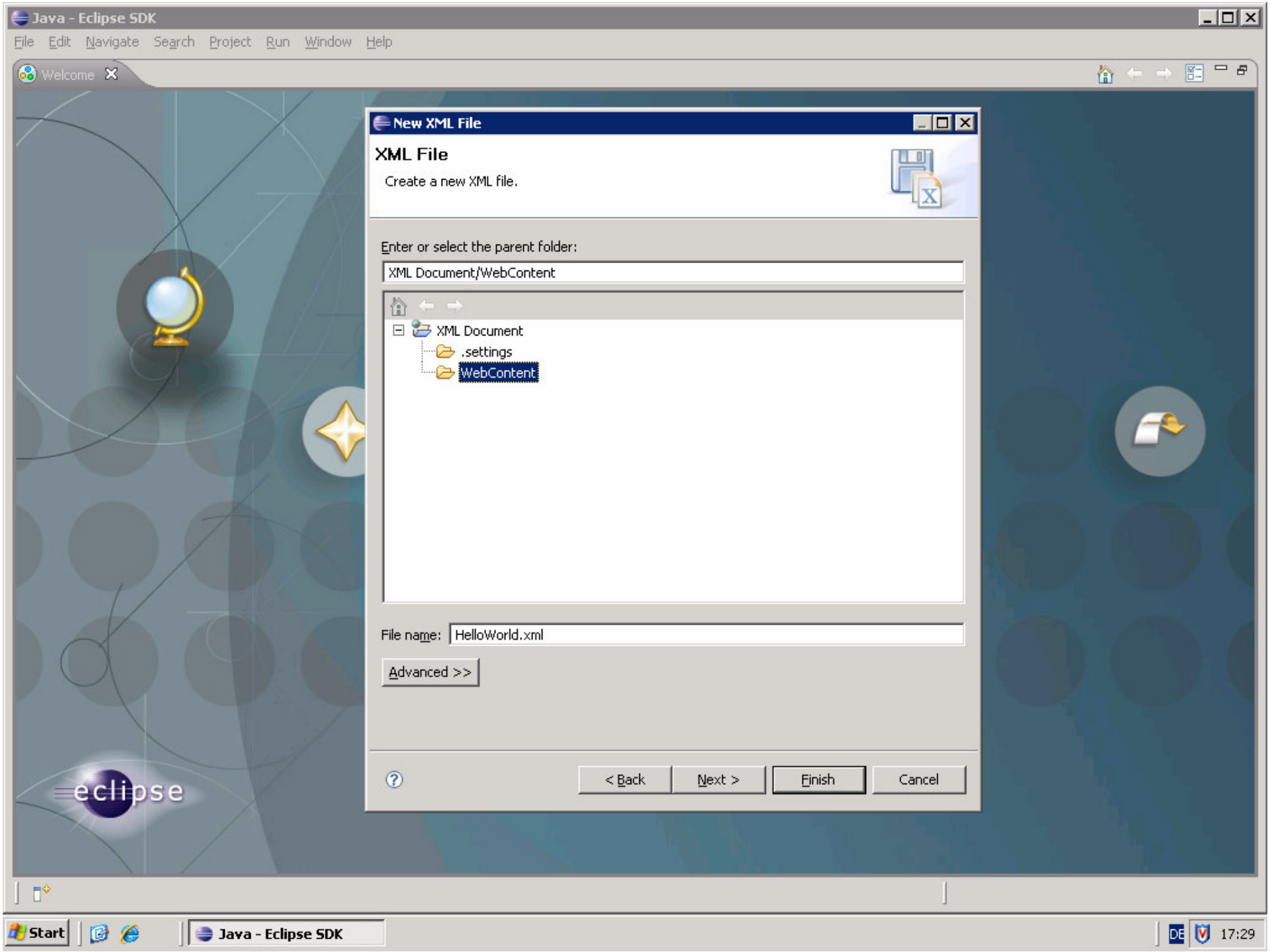
- Yes, there are tools to generate XHTML from HTML
  - But you'll want to edit it manually, right?
  - You can check the result with HTML Tidy (http://tidy.sourceforge.net/)

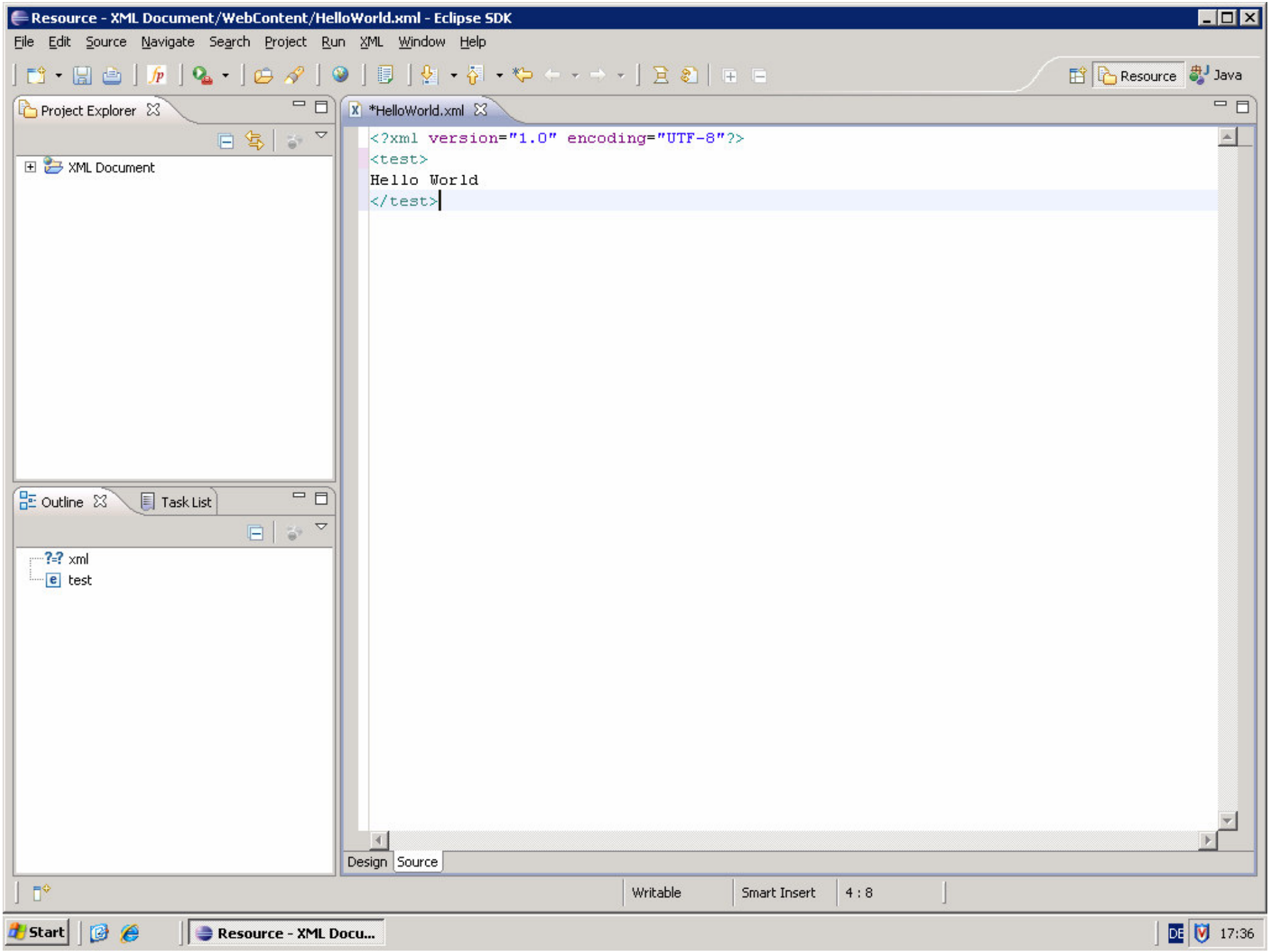
- Congratulations, you now have a XML document!

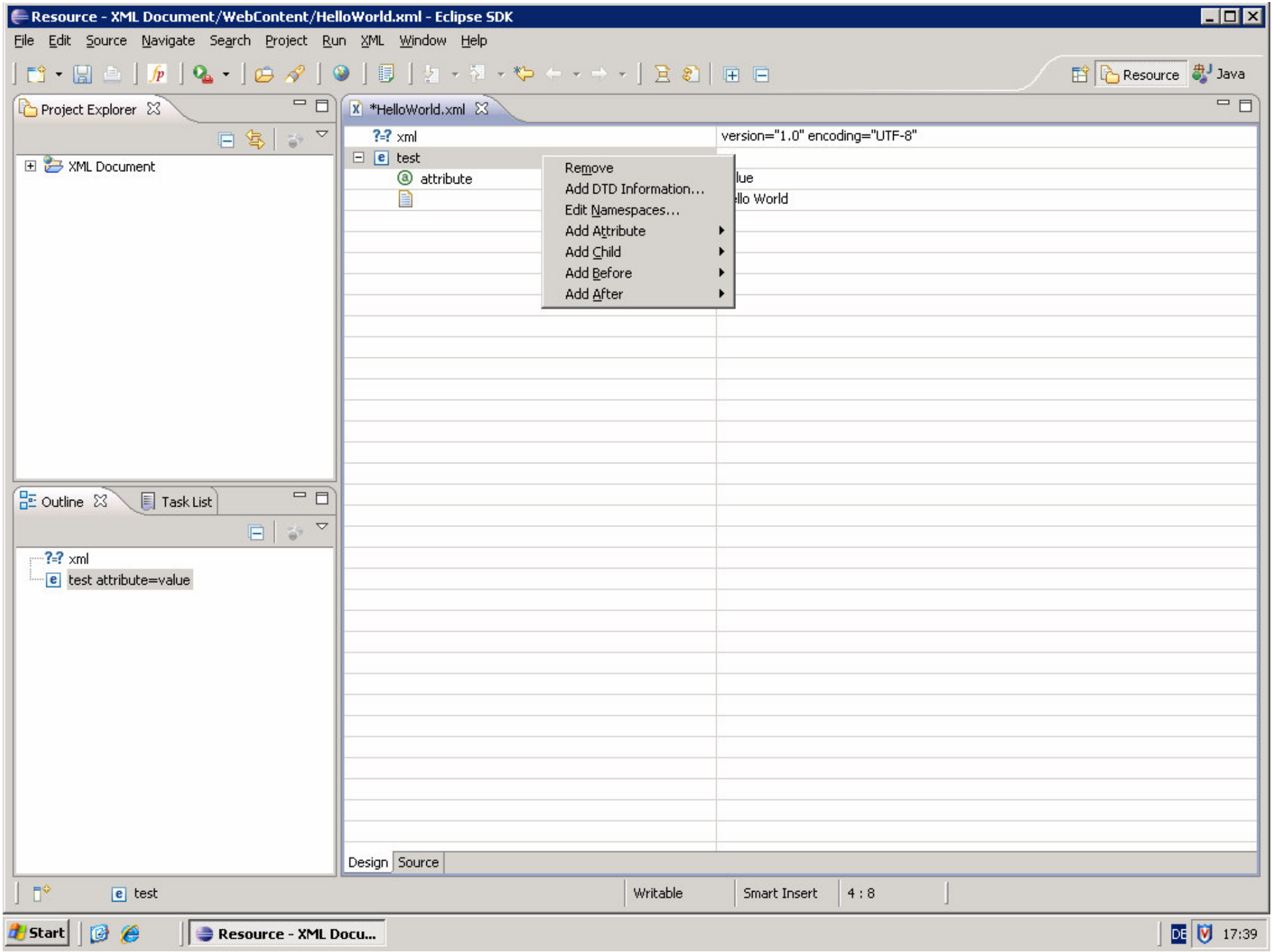
XML Tools

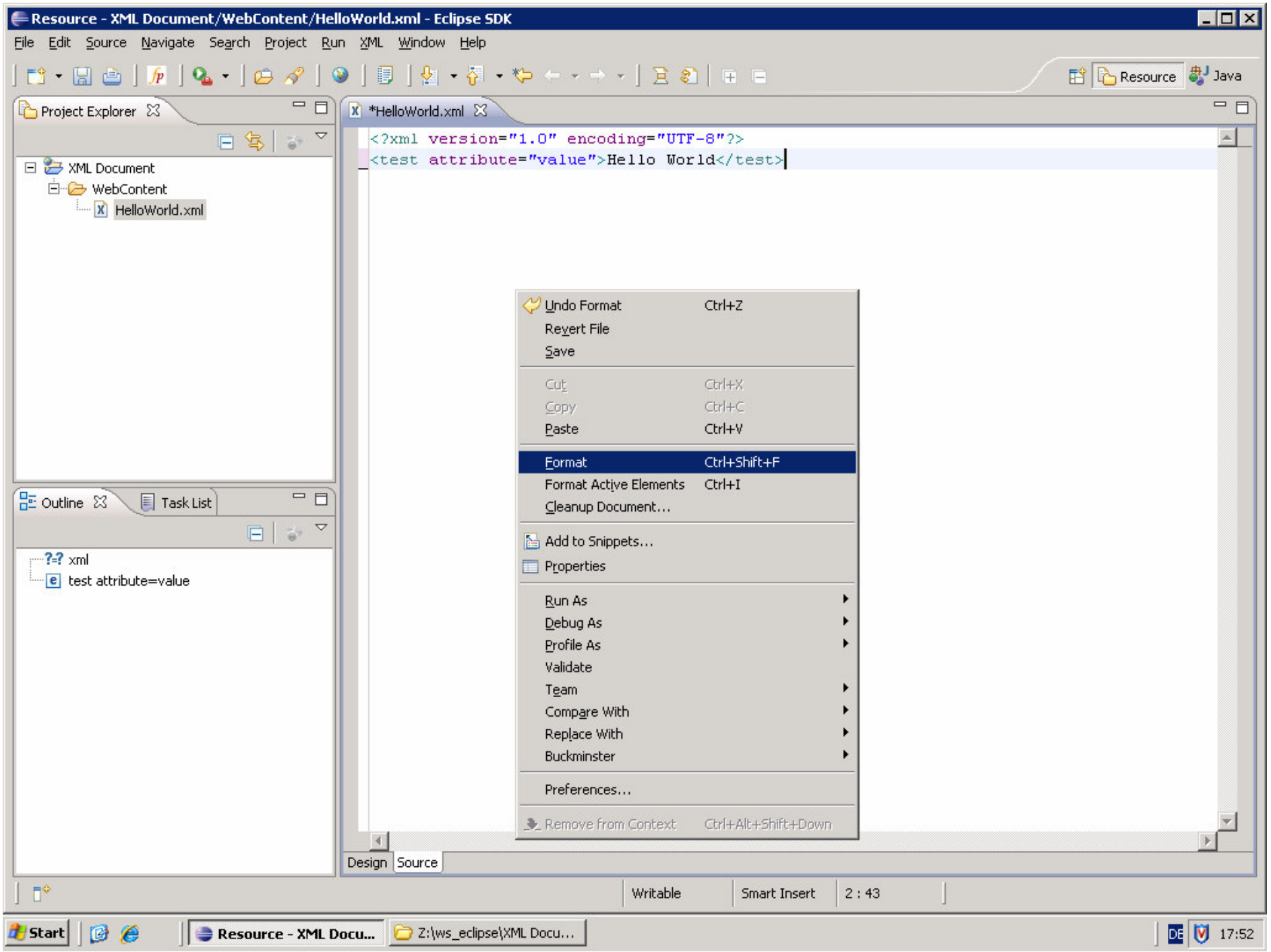

Java - Eclipse SDK
File Edit Navigate Search Project Run Window Help
New Alt+Shift+N
Open File...
Close Ctrl+W
Close All Ctrl+Shift+W
Save Ctrl+S
Save As...
Save All Ctrl+Shift+S
Revert
Move...
Rename... F2
Refresh F5
Convert Line Delimiters To
Print... Ctrl+P
Switch Workspace
Import...
Export...
Open a Component Query...
View a selected CSpec...
Properties Alt+Enter
Exit
Java Project
Project...
Package
Class
Interface
Enum
Annotation
Source Folder
Folder
File
Untitled Text File
JUnit Test Case
Task
Example...
Other...
eclipse
Start
Java - Eclipse SDK
17:20

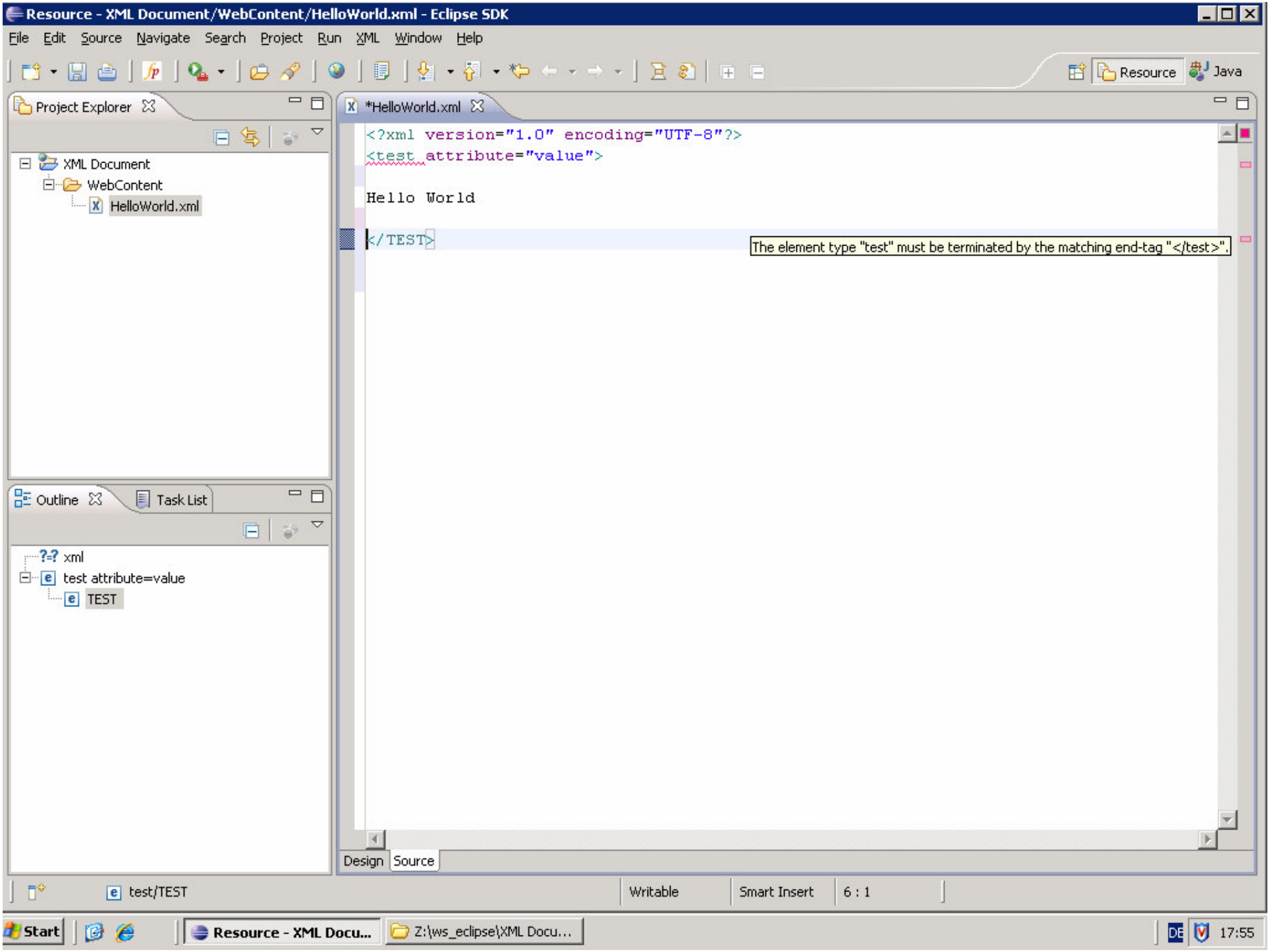

Java - Eclipse SDK
File Edit Navigate Search Project Run Window Help
Welcome
New
Select a wizard
Create a new XML file
Wizards:
type filter text
JPA
Mylyn
Plug-in Development
Server
SQL Development
SVN
User Assistance
Web
Web Services
XML
DTD
XML
XML Schema
Examples
< Back
Next >
Finish
Cancel
eclipse
Start
Java - Eclipse SDK
17:26

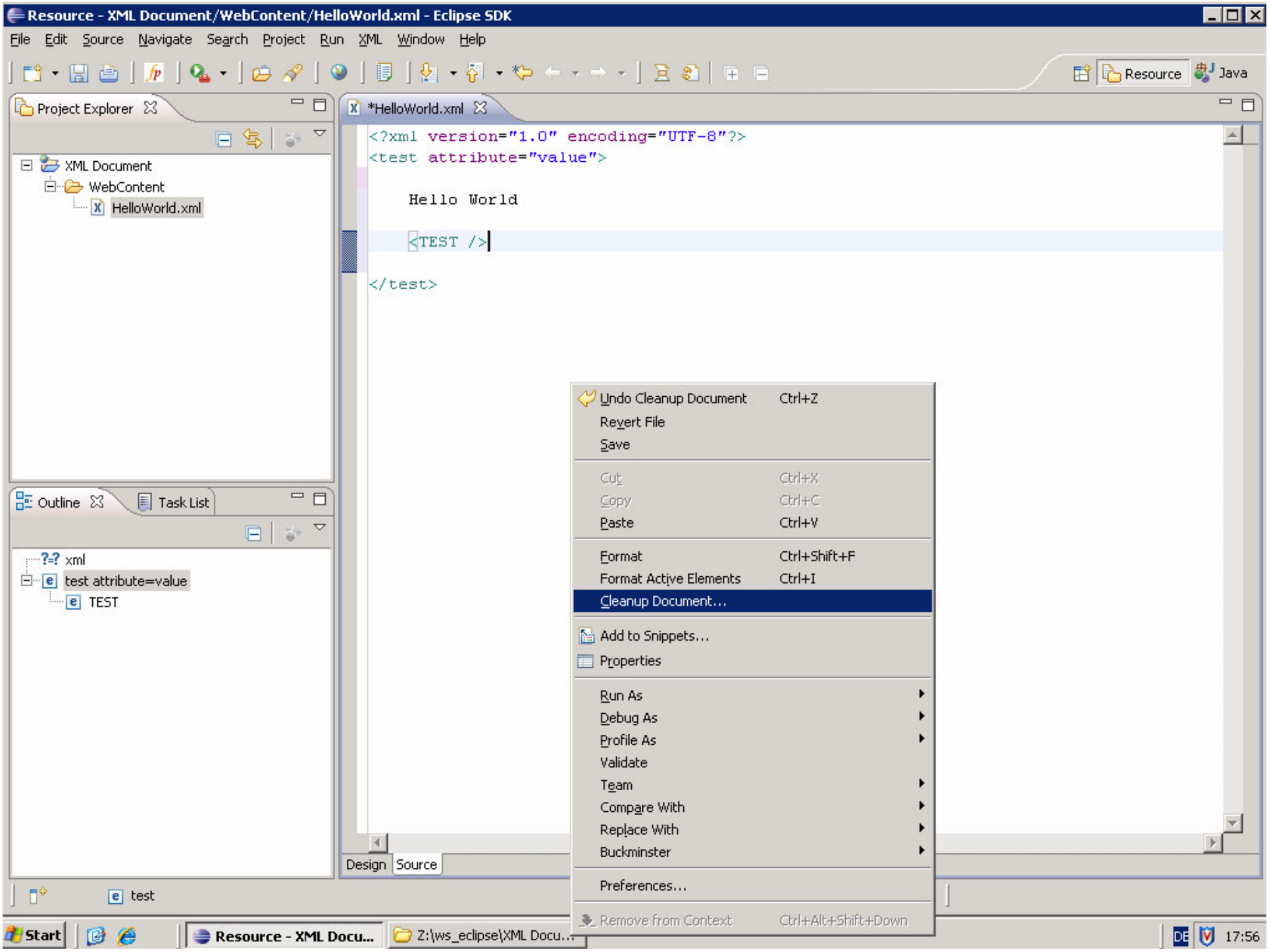

Java - Eclipse SDK
File Edit Navigate Search Project Run Window Help
Welcome
New XML File
XML File
Create a new XML file.
Enter or select the parent folder:
XML Document/WebContent
XML Document
.settings
WebContent
File name: HelloWorld.xml
Advanced >>
< Back
Next >
Finish
Cancel
eclipse
Start
Java - Eclipse SDK
17:29


Resource - XML Document/WebContent/HelloWorld.xml - Eclipse SDK
File Edit Source Navigate Search Project Run XML Window Help
Resource
Java
Project Explorer
XML Document
*HelloWorld.xml
<?xml version="1.0" encoding="UTF-8"?>
<test>
Hello World
</test>
Outline
Task List
xml
test
Design
Source
Writable
Smart Insert
4 : 8
Start
Resource - XML Docu...
17:36


Resource - XML Document/WebContent/HelloWorld.xml - Eclipse SDK
File Edit Source Navigate Search Project Run XML Window Help
Resource
Java
Project Explorer
XML Document
*HelloWorld.xml
xml
version="1.0" encoding="UTF-8"
test
attribute
Remove
Add DTD Information...
Edit Namespaces...
Add Attribute
Add Child
Add Before
Add After
Outline
Task List
xml
test attribute=value
Design
Source
test
Writable
Smart Insert
4 : 8
Start
Resource - XML Docu...
17:39


Resource - XML Document/WebContent/HelloWorld.xml - Eclipse SDK
File Edit Source Navigate Search Project Run XML Window Help
Resource
Java
Project Explorer
XML Document
WebContent
HelloWorld.xml
*HelloWorld.xml
<?xml version="1.0" encoding="UTF-8"?>
<test attribute="value">Hello World</test>
Undo Format Ctrl+Z
Revert File
Save
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Format Ctrl+Shift+F
Format Active Elements Ctrl+I
Cleanup Document...
Add to Snippets...
Properties
Run As
Debug As
Profile As
Validate
Team
Compare With
Replace With
Buckminster
Preferences...
Remove from Context Ctrl+Alt+Shift+Down
Outline
Task List
xml
test attribute=value
Design
Source
Writable
Smart Insert
2 : 43
Start
Resource - XML Docu...
Z:\ws_eclipse\XML Docu...
17:52


Resource - XML Document/WebContent/HelloWorld.xml - Eclipse SDK
File Edit Source Navigate Search Project Run XML Window Help
Resource
Java
Project Explorer
XML Document
WebContent
HelloWorld.xml
*HelloWorld.xml
<?xml version="1.0" encoding="UTF-8"?>
<test attribute="value">
Hello World
</TEST>
The element type "test" must be terminated by the matching end-tag "</test>".
Outline
Task List
xml
test attribute=value
TEST
Design
Source
test/TEST
Writable
Smart Insert
6 : 1
Start
Resource - XML Docu...
Z:\ws_eclipse\XML Docu...
17:55


Resource - XML Document/WebContent/HelloWorld.xml - Eclipse SDK
File Edit Source Navigate Search Project Run XML Window Help
Resource
Java
Project Explorer
XML Document
WebContent
HelloWorld.xml
*HelloWorld.xml
<?xml version="1.0" encoding="UTF-8"?>
<test attribute="value">
Hello World
<TEST />
</test>
Undo Cleanup Document Ctrl+Z
Revert File
Save
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Format Ctrl+Shift+F
Format Active Elements Ctrl+I
Cleanup Document...
Add to Snippets...
Properties
Run As
Debug As
Profile As
Validate
Team
Compare With
Replace With
Buckminster
Preferences...
Remove from Context Ctrl+Alt+Shift+Down
Outline
Task List
xml
test attribute=value
TEST
Design
Source
test
Start
Resource - XML Docu...
Z:\ws_eclipse\XML Docu...
17:56

The End

Freie Universität Berlin

- Übungsblatt 2 ist online 30.4.
- Übung 2 ist am 7.5.
  - Mit Musterlösungen zu Übungsblätter 1 und 2