GATE and UIMA in Language Technology Teaching

Graham Wilcock
University of Helsinki
graham.wilcock@helsinki.fi
Outline

• Current course materials
  • Shakespeare’s Sonnets
  • GATE & ANNIE
• New course materials
  • UI MA & OpenNLP
  • UI MA & Stanford NLP
• Appendix
  • Eclipse & Stanford Eclipse
IBM: Tidwell XSLT tutorials

Part 1. Transforming XML into HTML

Doug Tidwell
Cobar Evangelist, development tools.
Published November 1999
Updated November 2000

Part 1 of this Transforming XML documents tutorial shows you how to transform XML documents into HTML.

Let's meet our content:

For our transformations, we'll use six source documents: a Shakespearean sonnet, a business letter, definitions for several technical terms, some spreadsheet data, a section from a technical manual, and a short section from Henry Fielding's 18th-century British novel Tom Jones. This will give us a wide range of document types to transform.

As we discuss our different transformations, some of these documents will be more relevant than others. For example, converting spreadsheet data into a pie chart would be useful to many people, converting a sonnet into a pie chart, while an intriguing exercise, probably has a more narrow appeal. We'll present a simple rendering of all our documents here. If you'd like to see the XML source for the documents and their DTDs, check out Appendix A: Sample XML Documents.

A Shakespearean sonnet:
This document is rendered line-by-line, with an indication of the rhyme scheme at the start of each line.

Sonnet 130
Author: William Shakespeare (British, 1564-1616)

A  My mistress' eyes are nothing like the sun,
B  Coral is far more red than her lips red
A  If snow be white, why then her breasts are dun,
B  If hair be wires, black wires grow on her head.
C  I have seen roses damask'd, red and white,
D  But no such rose see I in her cheeks;
C  And in some perfumes is there more delight

[Diagram of IBM developerWorks interface]
Sonnet 130

by William Shakespeare (1564-1616)

My mistress' eyes are nothing like the sun,
Coral is far more red than her lips red.|
If snow be white, why then her breasts are dun,
If hairs be wires, black wires grow on her head.

I have seen roses damasked, red and white,
But no such roses see I in her cheeks.
And in some perfumes is there more delight
Than in the breath that from my mistress reeks.

I love to hear her speak, yet well I know
That music hath a far more pleasing sound.
I grant I never saw a goddess go,
My mistress when she walks, treads on the ground.

And yet, by Heaven, I think my love as rare
As any she belied with false compare.
Gutenberg: Sonnets corpus
Outline

• Current course materials
  • Shakespeare’s Sonnets
    • GATE & ANNIE

• New course materials
  • UI MA & OpenNLP
  • UI MA & Stanford NLP

• Appendix
  • Eclipse & Stanford Eclipse
GATE & ANNIE

- Start with ANNIE
  - Ready-to-run NLP tools
  - ANNIE NE Recognizer
    - ANNIE POS Tagger
- Add JAPE annotations
  - Students write NP, PP rules
Sonnet 130
by William Shakespeare (1564-1616)

My mistress' eyes are nothing like the sun,
Coral is far more red than her lips red.
If snow be white, why then her breasts are dun,
If hairs be wires, black wires grow on her head.

I have been roses damasked, red and white,
But no such roses see I in her cheeks.
And in some perfumes is there more delight
Than in the breath that from my mistress reeks.

I love to hear her speak, yet well I know
That music hath a far more pleasing sound.
I grant I never saw a goddess go,
My mistress when she walks, treads on the ground.

And yet, by Heaven, I think my love as rare
As any she belied with false compare.
GATE & ANNIE

• Start with ANNIE
  • Ready-to-run NLP tools
  • ANNIE NE Recognizer
    • ANNIE POS Tagger
• Add JAPE annotations
  • Students write NP, PP rules
Sonnet 130

By William Shakespeare (1564-1616)

My mistress' eyes are nothing like the sun,
Coral is far more red than her lips red,
If snow be white, why then her breasts are snow,
If hairs be wires, black wires grow on her head.

I have seen roses damasked, red and white,
But no such roses see I in her cheeks.
And in some pictures these more delight
Than in the breath that from my mistress reeks.

I love to hear her speak; yet well I know
That music hath a far more pleasing sound.
I grant I never saw a goddess go,
My mistress when she walks, treads on the ground.

And yet, by Heaven, I think my love as rare
As any she; think'st thou not mine good fair?
GATE & ANNIE

- Start with ANNIE
  - Ready-to-run NLP tools
  - ANNIE NE Recognizer
  - ANNIE POS Tagger
- Add JAPE annotations
  - Students write NP, PP rules
Sonnet 130

By William Shakespeare (1564-1616)

My mistress' eyes are nothing like the sun,
Lanterns are more red than her lips red.
If snow be white, why then her breasts are dun.
If hairs be wires, black wires grow on her head.

I have seen roses damasked, red and white,
But no such roses see I in her cheeks.
And in some pictures is there more delight
Than in the breath that from my mistress' necks.

I love to hear her speak, yet well I know
That music hath a far more pleasing sound.
I grant I never saw a goddess go,
My mistress when she walks, treads on the ground.

And yet, by Heaven, I think my love as rare
As any she belied with false compare.
Outline

• Current course materials
  • Shakespeare’s Sonnets
  • GATE & ANNIE

• New course materials
  • UI MA & OpenNLP
  • UI MA & Stanford NLP

• Appendix
  • Eclipse & Stanford Eclipse
UI MA & OpenNLP

- OpenNLP tools
  - First, run from command line
  - Install in UI MA (assignment)

- UI MA & OpenNLP
  - OpenNLP POS Tagger
  - OpenNLP NE Recognizer
UI MA & OpenNLP

- OpenNLP tools
  - First, run from command line
  - Install in UIMA (assignment)

- UI MA & OpenNLP
  - OpenNLP POS Tagger
  - OpenNLP NE Recognizer
My mistress' eyes are nothing like the sun,
Corinna's hair, the sun's disc more red;
If snow be white, why then her breasts are dun;
If hairs be wires, black wires grow on her head.

I have seen roses dearest, red and white,
But not such roses see I in her cheeks;
And in some perfumes is there more delight
Than in the breath that from my mistress reeks.

I love to hear her speak; yet well I know
That music hath a far more pleasing sound,
I grant I never saw a goddess go,
My mistress when she walks, treads on the ground.

And yet, by Heaven, I think my love as rare
As any she belied with false compare.
UIMA & OpenNLP

- OpenNLP tools
  - First, run from command line
  - Install in UIMA (assignment)

- UIMA & OpenNLP
  - OpenNLP POS Tagger
  - OpenNLP NE Recognizer
Sonnet 130

by William Shakespeare (1564-1616)

My mistress' eyes are nothing like the sun, 
Corin is far more red than her lips red. 
If snow be white, why then her breasts are dun, 
If hairs be wires, black wires grow on her head. 
I have seen roses damasked, red and white, 
But no such roses see I in her cheeks. 
And in some perfumes is there more delight 
Than in the breath that from my mistress reeks. 
I love to hear her speak; yet well I know 
That music hath a far more pleasing sound. 
I grant I never saw a goddess go. 
My mistress when she walks, treads on the ground. 
And yet, by Heaven, I think my love as rare 
As any she belied with false compare.
UIMA & OpenNLP

- OpenNLP tools
  - First, run from command line
  - Install in UIMA (assignment)
- UIMA & OpenNLP
  - OpenNLP POS Tagger
  - OpenNLP NE Recognizer
- Add Java annotators
  - Students write NP, PP annotators?
Outline

• Current course materials
  • Shakespeare’s Sonnets
  • GATE & ANNIE

• New course materials
  • UIMA & OpenNLP
  • UIMA & Stanford NLP

• Appendix
  • Eclipse & Stanford Eclipse
Stanford NLP Group

- Stanford NLP tools
  - Stanford POS Tagger
  - Stanford NE Recognizer
  - Stanford Parser
UI MA & Stanford NLP

• Stanford NE Recognizer
  • Use Stanford NER-GUI

• UI MA & Stanford NER
  • Install in UI MA (assignment)
  • UI MA wrapper by F. Laws
Sonnet 130

by William Shakespeare (1564–1616)

My mistress' eyes are nothing like the sun, 
Corinna is far more red than her lips red. 
If snow be white, why then her breasts are dun, 
If hairs be wires, black wires grow on her head.

I have seen roses damasked, red and white, 
But no such roses see I in her cheeks. 
And in some perfumes is there more delight 
Than in the breath that from my mistress reeks.

I love to hear her speak, yet well I know 
That music hath a far more pleasing sound. 
I grant I never saw a goddess go, 
My mistress when she walks, treads on the ground.

And yet, by Heaven, I think my love as rare 
As any she belied with false compare.
UI MA & Stanford NLP

- Stanford NE Recognizer
  - Use Stanford NER-GUI

- UI MA & Stanford NER
  - Install in UIMA (assignment)
  - UI MA wrapper by F. Laws
Unstructured Information Management Architecture

Input Directory: C:\Tools\IBMworkspace\uma_examples\data\sonnet
Output Directory: C:\Tools\IBMworkspace\uma_examples\data\processed
Location of Analysis Engine XML Description: C:\Tools\IBMworkspace\StanfordNLP\dof\lorianfiles\uma\StanfordNERAnn
XML Tag containing Text (optional): 
Language: en
Character Encoding: UTF-8

Sonnet 130
by William Shakespeare (1564-1616)

My mistress' eyes are nothing like the sun,
Corinna is far more red than her lips red.
If snow be white, why then her breasts are dun,
If hairs be wires, black wires grow on her head.
I have seen roses damasked, red and white,
But no such roses see I in her cheeks.
And yet, by Heaven, I think my love as rare
Than any she belied with false compare.

Annotation Results for sonnet130.txt.xml in C:ToolsIBMworkspaceuma_examples\data\processed

Click In Text to See Annotation Detail

Legend
- Document
- Location
- Organization

Problems: Javadoc
UIMA Document A
Loading data
getting doc.
analyzing.
Outline

• **Current course materials**
  • Shakespeare’s Sonnets
  • GATE & ANNIE

• **New course materials**
  • UIMA & OpenNLP
  • UIMA & Stanford NLP

• **Appendix**
  • Eclipse & Stanford Eclipse
Eclipse & Stanford Eclipse

- jEdit vs. Eclipse
  - Students currently use jEdit
  - Eclipse learning curve is harder
- Stanford Eclipse
  - Stanford CS: simplified Eclipse
  - *Karel the Robot Learns Java*
  - Starter projects for assignments
At present, the CollectNewspaperKarel subclass does nothing. Your job in the assignment is to add the necessary code to instruct Karel to walk to the door of its house, pick up the newspaper (represented by a beeper, of course), and then return to its initial position in the upper left corner of the house.

```java
import stanford.karel.*;

public class CollectNewspaperKarel extends SuperKarel {

    public void run() {
        move();
        turnRight();
        move();
        turnLeft();
        move();
        pickBeeper();
    }
}
```
Starter projects

The layout of the screen—which the Eclipse designers decided to call a perspective—corresponds to the Stanford Editor perspective, which is where you will want to start editing your programs. Because Eclipse is a professional tool, it contains many advanced features that only get in the way in an introductory course. The purpose of the Stanford Editor perspective is to hide all the scary-looking features that you don’t need. If you ever find yourself in a perspective that seems confusing, click on Stanford in the menu bar and then select Switch to Editor. You can also click on the Switch to Editor button from the main toolbar, which looks like this:

Once you have Eclipse loaded, your next step will usually be to create a new project in your workspace by importing a skeletal framework that we provide called a starter project. Using starter projects make your life much easier by allowing you to ignore the many details involved in creating a project from scratch. Every assignment will include a starter project for each problem, and your first task will be to download the starter project from the class website and then importing it into your workspace. The details for doing so are described in Handout #5 (Using Karel with Eclipse).
Sonnet 130

by William Shakespeare (1564 - 1616)

My mistress' eyes are nothing like the sun,
Corin is far more red than her lips red.

If snow be white, why then her breasts are dun.
If hair be wire, black wires grow on her head.

I have seen roses damasked, red and white,
But no such roses see I in her cheeks.

And in some perfumes is there more delight
Then in the breath that from my mistress reeks.

I love to hear her speak, yet well I know
That music hath a far more pleasing sound.

And yet, by heaven, I think my love as rare
As any she belied with false compare.
Sonnet 130

by William Shakespeare (1564-1616)

My mistress' eyes are nothing like the sun,
Corin is far more red than her lips red;
If snow be white, why then her breasts are dun;
If hairs be black, black is her velvet gown.

My mistress' eyes are Nothing like the sun,
Corin is far more red than her lips red;
If snow be white, why then her breasts are dun;
If hairs be black, black is her velvet gown.

I have seen roses damasked, red and white,
But no such rose as in her cheeks and in some perfumes there is more delight
Than in the breath that from her mistress walks.

I love to hear her speak, yet well I know
That music hath a far more pleasing sound.
I grant I never saw a goddess go,
My mistress when she walks, treads on the ground.

And yet, by Heaven, I think my love as rare
As any she bethed with fairest base.
Sonnet 130
by William Shakespeare (1564-1616)

My mistress' eyes are nothing like the sun,
Corin is far more red than her lips red.
If snow be white, why then her breasts are dun,
If hairs be wires, black wires grow on her head.

I have seen roses dyed with sunset's beam,
But such roses never saw the sun's brave stream.
And in some corner is there more delight
Than in the breath that from my mistress reeks.

I love to hear her speak, yet well I know
That music hath a far more pleasing sound.
I grant I never saw a goddess go,
My mistress when she walks,breathes, moves on the ground.

And yet, by Heaven, I think my love as rare
As any she belied with false compare.
Summary

• Current course materials
  • Shakespeare’s Sonnets
  • GATE & ANNIE

• New course materials
  • UI MA & OpenNLP
  • UI MA & Stanford NLP

• Appendix
  • Eclipse & Stanford Eclipse