

BACKGROUND

Hypertext Markup Language (HTML) is the standard markup language for documents designed to be displayed in a web browser. Web browsers receive HTML documents from a web server or local storage and render the documents into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document.

HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects such as interactive forms may be embedded into the rendered page. HTML provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. HTML elements are delineated by tags, written using angle brackets. Tags such as `` and `<input />` directly introduce content into the page. Other tags such as `<p>` surround and provide information about document text and may include other tags as sub-elements. Browsers do not display the HTML tags but use them to interpret the content of the page.

Tim Berners-Lee, best known as the inventor of the World Wide Web, proposed the language in a memo in 1989.

PROJECT DESCRIPTION

You need to code a program in Java that takes an HTML document file in the English language as input and produce a text document as output. The output document contains only the text information that a web browser display by reading the file, without the presentation details the markup language specified.

- Your program needs to Extract Titles, section headers, paragraphs, and tables.
- The output of the program would be recorded in a text file containing the text without any format.
- Tables rows would be rendered like a sequence of values separated by commas for each row in the table.

HTML in Browser

Matador Song

Fight, Matadors, for Tech!
Songs of love we'll sing to thee,
Bear our banners far and wide.
Ever to be our pride,
Fearless champions ever be.
Stand on heights of victory.
Strive for honor evermore.
Long live the Matadors!

Music by Harry Lemaire, words by R.C. Marshall



Text file contains

Matador Song

Fight, Matadors, for Tech!
Songs of love we'll sing to thee,
Bear our banners far and wide.
Ever to be our pride,
Fearless champions ever be.
Stand on heights of victory.
Strive for honor evermore.
Long live the Matadors!

Music by Harry Lemaire, words by R.C. Marshall

HTML in Browser

Text file contains

Our Customers

Company	Contact	Country
Alfreds Futterkiste	Maria Anders	Germany
Centro comercial Moctezuma	Francisco Chang	Mexico
Ernst Handel	Roland Mendel	Austria
Island Trading	Helen Bennett	UK
Laughing Bacchus Winecellars	Yoshi Tannamuri	Canada
Magazzini Alimentari Riuniti	Giovanni Rovelli	Italy

Our Customers

Company, Contact, Country
Alfreds Futterkiste, Maria Anders, Germany
Centro comercial Moctezuma, Francisco Chang, Mexico
Ernst Handel, Roland Mendel, Austria
Island Trading, Helen Bennett, UK
Laughing Bacchus Winecellars, Yoshi Tannamuri, Canada
Magazzini Alimentari Riuniti, Giovanni Rovelli, Italy

RESTRICTIONS

To solve this project you can use arrays, files, Strings and any numbers methods you need. You can use one data structure, like linked list. You cannot use prebuild or downloaded classes that remove markup language or use regular expressions. The purpose of this project is for you get familiar with coding and using Java. You are not required to use multiple classes; future projects will provide practice for that, once we cover more material and concepts.

SUBMISSION

Your solution should include the following documentation, missing elements would result in missing points.

- Each file needs to include a header containing the following.
 - Course, Section and Semester
 - Your Name
 - The title "Project 1"
- Methods / Functions should include a description of its purpose.
- Explaining input (parameters) and outputs (returns)

SUBMISSION REMINDERS

Once your program is complete with documentation...

- ***This Project is due on Monday, February 17th***
- ***Late submission on Tuesday February 18th will have a penalty of 20% on the grade.***
- ***Any submission after that will received a grade of zero.***
- Click on your project name and use your pop-up menu to clean it.
- Then Close NetBeans.
- Zip the whole Project folder, which is located inside your Documents folder under the NetBeans projects.
- Rename the Zip File using the following name format **your_eraider-ProjNo.x_Vy**
- Where **x** is the number of the assignment and **y** is the version.
- In case of multiple versions ONLY the latest version will be graded.
- Upload the Zip File containing the whole project folder to Blackboard, using Assignment Submission.