

CS 1520 / CoE 1520: Programming Languages for Web Applications (Spring 2012)
Department of Computer Science, University of Pittsburgh

Assignment #1: Perl

Released: September 22nd, 2012

Due: 11:59pm, Wednesday, October 3rd, 2012

Goal

Gain familiarity with Perl.

Description

In this assignment, you will parse a set of similarly-formatted HTML pages to collect raw data, store the data in appropriate data structures, and answer user queries based on the data. The pages you are asked to parse are the NHL statistics pages for the different teams, starting from the 2000-2001 season. We will provide trimmed-down version of the HTML files, for you to parse, and also the correct output for two years, so that you can verify your first Perl program or start with the second program if you prefer.

You will need to write two different Perl programs:

1. `parse.pl` parses a specified HTML page stored in file (`input_file_name`) and converts the data into a tab-separated format, stored in file (`output_file_name`).

Usage: `parse.pl -input <input_file_name> -output <output_file_name>`
(for example: `parse.pl -input season10.html -output 10.txt`)

Note that the expectation is the program will be called multiple times (i.e., once per file, for all years), and that the output file will be overwritten every time the program is executed.

2. `answer.pl` reads in **all the .txt files** in the current directory, expecting them to be the format that is output by `parse.pl`. In doing so, the program is building and populating the appropriate data structures. Then the program will prompt the user which question he/she wants to be answered and provide the answer. There are five questions that should be supported:

- (a) Show the rank (i.e., order on the table) of a specific team (provided by the user) for each season. For example, if the team name provided by the user is Pittsburgh, then the program should produce:

```
Statistics for PITTSBURGH:
2011-2012: 3
2010-2011: 2
2009-2010: 7
etc
```

An error message should be printed if the provided team name does not exist.

- (b) Show the final rankings of all teams for a specific season (provided by the user). For example, if the year provided is 2012, then the program should produce:

```
Statistics for the 2011-2012 season:
1 VANCOUVER
2 NY RANGERS
3 PITTSBURGH
etc
```

An error message should be printed if the provided year is outside the data provided.

- (c) Compute the average winning percentage (i.e., the percentage of games won as compared to all games played), over all seasons for each team. Produce a listing of the teams along with their winning percentages, in descending order (one line per team).
- (d) Compute the total number of goals over all seasons for each team. Produce a listing of the teams along with their total number of goals, in descending order (one line per team).

Command Line Arguments

To simplify things, you should utilize the `Getopt` Perl module to process the command-line arguments. You can get more information about it at the following URL:
<http://perldoc.perl.org/Getopt/Long.html>

Input Data

A zip file containing all files is provided as part of this assignment, at the following URL:

<http://db.cs.pitt.edu/courses/cs1520/fall2012/assign/data/assign1.zip>

Since the file contains copyrighted material that is not allowed to be publicly reposted, but it is allowed to be used for academic purposes, the zip file is only accessible from computers within the `pitt.edu` domain (i.e., on campus). If you are outside campus and want to access this, please use the VPN service provided by the University, which is available at <http://sremote.pitt.edu>.

The zip file contains 12 HTML files, named `season1.html`, ..., `season12.html` and two `.txt` files, named `11.txt` and `12.txt`. The two files are the proper output of the `parse.pl` script. Each season's HTML file is marked with the year it was terminated, so `season12.html` corresponds to the 2011-2012 season. Ditto for `12.txt`.

Note that although all the statistics are provided for all teams, the only statistics relevant to this assignment are the following:

- rank (i.e., placement in the table)
- wins "W",
- losses "L",
- points "P",
- power play (pp%)

What to submit

Two Perl programs that perform the tasks listed above, along with any additional libraries that you have developed (and are shared by the programs). Name your programs `index.pl` and `search.pl`. There is no standard naming scheme for the libraries.

Academic Honesty

The work in this assignment is to be done *independently*, by you and only you. Discussions with other students on the assignment should be limited to understanding the statement of the problem. **Cheating in any way, including giving your work to someone else, will result in an F for the course and a report to the appropriate University authority for further disciplinary action.**

How to submit your assignment

We will use a Web-based assignment submission interface. To submit your assignment:

- If you have more than one file to submit, prepare your assignment for uploading, by generating a single zip file with all the files.
- Go to the class web page <http://db.cs.pitt.edu/courses/cs1520/spring2012> and click the Submit button.
- Use your pittID as the username and the password you specified at the contact information form for authentication. There is a reminder service via email if you forgot your password. You must have already submitted your contact information, if you have not yet you need to do so now.
- Upload your assignment file to the appropriate assignment (from the drop-down list).
- Check (through the web interface) to verify what is the file size that has been uploaded and make sure it has been submitted in full. **It is your responsibility to make sure the assignment was properly submitted.**

You must submit your assignment before the due date (11:59pm, Wednesday, October 3rd, 2012) to avoid getting any late penalty. The timestamp of the electronic submission will determine if you have met the deadline. There will be no late submissions allowed after 11:59pm, Friday, October 5th, 2012.

[Last updated on September 22, 2012 at 12:59am EST]