Northeastern University
College *of* Computer and Information Science

# CS1100: Excel Lab 3

To complete this assignment you must submit an electronic copy to Blackboard by the due date. Use the starter file *cs1100.e3.xls*. In this lab you are asked to process text and calculate relevant statistics.

## Knowledge Needed

This lab requires the following Excel functions and techniques:

- Cell ranges, borders, shading, cell formatting, number formatting
- Text processing functions: **MID**, **LEFT**, **RIGHT**, **FIND**, **LEN, TRIM**
- **IFERROR** and **IF** functions to build robust models
- **COUNTA** to count cells that are not blank and **COUNTBLANK** to count cells that are blank
- Copying of formulas
- Excel help and online documentation

## Problem 1 (25 points)

The sales department at *Symbian* has a list of popular apps, shown in the sheet labeled "Problem 1". The data was downloaded in one field instead of using separate fields for name, price, seller, size and category. You may assume that name, price and seller are always separated by a comma and a space as shown in the spreadsheet. The size is preceded by a dash and a space and the category is surrounded by angled brackets.

Tasks:

- Add your first and last name to cell A22.
- Use text processing functions to extract the fields of name, price, seller, size and category and put them into the appropriate columns
- Formulas must be copyable down

Your solution should look like this:

| Data | Name | Price | Seller | Size | Category |
|---|---|---|---|---|---|
| Dog Boogie, 2.99, Dog Town Media - 17 MB <Game> | Dog Boogie | $2.99 | Dog Town Media | 17 MB | Game |
| Skype, 0.00, Skype Communications - 29.6 MB <Social Networking> | Skype | $0.00 | Skype Communications | 29.6 MB | Social Networking |
| Angry Birds, 0.99, Rovio Entertainment - 39.9 MB <Game> | Angry Birds | $0.99 | Rovio Entertainment | 39.9 MB | Game |
| Instagram, 0.00, Burbn Inc. - 14.5 MB <Photo & Video> | Instagram | $0.00 | Burbn Inc. | 14.5 MB | Photo & Video |
| SimplePhysics, 1.99, Jundroo - 7.5 MB <Education> | SimplePhysics | $1.99 | Jundroo | 7.5 MB | Education |
| Facebook, 0.00, Facebook Inc - 33.2 MB <Social Networking> | Facebook | $0.00 | Facebook Inc | 33.2 MB | Social Networking |
| Cut the Rope, 0.99, Chillingo Ltd - 25.1 MB <Game> | Cut the Rope | $0.99 | Chillingo Ltd | 25.1 MB | Game |
| Fruit Ninja, 0.99, Halfbrick - 43.9 MB <Game> | Fruit Ninja | $0.99 | Halfbrick | 43.9 MB | Game |
| YouTube, 0.00, Google Inc. - 11.2 MB <Photo & Video> | YouTube | $0.00 | Google Inc. | 11.2 MB | Photo & Video |
| Mathway, 1.99, Bagatrix Solutions - 28.5 MB <Education> | Mathway | $1.99 | Bagatrix Solutions | 28.5 MB | Education |

## Problem 2 (30 points)

The marketing department at *Symbian* is interested in how many of their online users are supplying an e-mail address when registering for access to their web portal. They have asked the database administrator to download the user registration information to an Excel file shown in the worksheet labeled "Problem 2". Unfortunately, the export function provided by the database

combines first name, last name, and e-mail address into a single field. Your task is to create an Excel worksheet to separate the components and then determine what percentage of users does not have an e-mail address.

Tasks:

- Extract the first name and put it into the First Name column
- Extract the last name and put it into the Last Name column
- Extract the e-mail address and put it into the E-Mail Address column keeping in mind that some users do not have an e-mail address
- Calculate the percentage of users that do not have an e-mail address. (Use COUNTA and COUNTBLANK to compute the percentage)
- Formulas must be copyable down
- Use IFERROR to handle fields where data is missing

Here's what the output should look like. Format the output as shown:

| **Symbian User List** | | | |
|---|---|---|---|
| | | | |
| **User Information** | **First Name** | **Last Name** | **E-Mail Address** |
| Sheffield, John [john.s@yahoo.com] | John | Sheffield | john.s@yahoo.com |
| Mazzoli, Sean [s3662@aol.com] | Sean | Mazzoli | s3662@aol.com |
| White, Jennifer [j.white@iti.org] | Jennifer | White | j.white@iti.org |
| Ramamoorthy, Krisnan | Krisnan | Ramamoorthy | |
| Schneider, Markus [sm@gmail.com] | Markus | Schneider | sm@gmail.com |
| Heider, Patricia [patricia.heider@neu.edu] | Patricia | Heider | patricia.heider@neu.edu |
| O'Donnel, Tim [timod@tyne.net] | Tim | O'Donnel | timod@tyne.net |
| Weinstein, Liz | Liz | Weinstein | |
| Reiker, Pat [reiker@me.com] | Pat | Reiker | reiker@me.com |
| *% of users without e-mail address =* | *22.2%* | | |

## Problem 3 (45 points)

A bookstore would like to determine what kinds of books were sold on a particular day. Data has been downloaded into a spreadsheet shown in the worksheet labeled "Problem 3". The downloaded data may contain leading or trailing spaces due to a defect in the download program.

Tasks:

1. In column B, extract the title of each book.
2. In column C, extract the author of each book
3. In column D, extract the format code of each book. Be sure to handle codes with trailing spaces.
4. In the sheet named "Lookup Table", define a lookup table as a named range with the name "FormatTable". The codes for Format are as follows:
   HC = Hardcover
   PB= Paperback
   EB = Ebook
   AB = Audiobook
5. In column E, use a VLOOKUP function to lookup the Format from the lookup table
6. Filter the data for each book format.
7. Find the percentage of each format sold
8. All formulas should be copyable down. Filter formula should be copyable down and across
9. Format your solution as shown below using bold, shading, borders, italics and percentages

Here is what your output should look like:

| Data | Title | Author | Format Code | Format | HC | PB | EB | AB |
|---|---|---|---|---|---|---|---|---|
| Gone Girl by Gillian Flynn, HC | Gone Girl | Gillian Flynn | HC | Hardcover | 1 | | | |
| Life of Pi by Yann Martel, PB | Life of Pi | Yann Martel | PB | Paperback | | 1 | | |
| The Immortal Life of Henrietta Lacks by Rebecca Skloot, PB | The Immortal Life of He | Rebecca Skloot | PB | Paperback | | 1 | | |
| Team of Rivals by Doris Kearns Goodwin, EB | Team of Rivals | Doris Kearns Goodwi | EB | Ebook | | | 1 | |
| Outliers by Malcolm Gladwell, AB | Outliers | Malcolm Gladwell | AB | Audio Book | | | | 1 |
| Duck & Goose by Tad Hills, PB | Duck & Goose | Tad Hills | PB | Paperback | | 1 | | |
| The New Jim Crow by Michelle Alexander, EB | The New Jim Crow | Michelle Alexander | EB | Ebook | | | 1 | |
| The Zombie Survival Guide by Max Brooks, PB | The Zombie Survival G | Max Brooks | PB | Paperback | | 1 | | |
| Madline by Ludwig Bemelmans, EB | Madline | Ludwig Bemelmans | EB | Ebook | | | 1 | |
| Griftopia by Matt Taibbi, EB | Griftopia | Matt Taibbi | EB | Ebook | | | 1 | |
| On the Map by Simon Garfield, HC | On the Map | Simon Garfield | HC | Hardcover | 1 | | | |
| The Moonstone by Wilkie Collins, AB | The Moonstone | Wilkie Collins | AB | Audio Book | | | | 1 |
| | | | | *Percentage* | *17%* | *33%* | *33%* | *17%* |

**Hints**:

- Create additional (hidden) columns to store temporary data
- Determine what the right delimiters are for the different components
- Use **IFERROR** to deal with error conditions when substrings cannot be found
- For Problem 2, you may assume that all users have a first name and a last name, but that the e-mail address is optional
- Do not change any data that is given (e.g. do not remove or add spaces in Problem 3)

## GRADING RUBRIC

This rubric is intended to guide graders in their evaluation of the students' submissions.

### Problem 1 (25 points)

| Criterion | Grading |
|---|---|
| *Student's name is added to data* | -5 if name not added |
| *Correct name* | -5 if app name is not extracted |
| *Correct price* | -5 if price not extracted |
| *Correct seller* | -5 if seller not extracted |
| *Correct size* | -5 if size not extracted |
| *Correct category (no brackets)* | -5 if category not extracted |

### Problem 2 (30 points)

| Criterion | Grading |
|---|---|
| *Correct first name* | 5 points |
| *Correct last name* | 5 points |
| *Correct email address* | 5 points |
| *Handling of missing email* | 5 points |
| *Correct percentage of users without emails* | 5 points |
| *Formatting* | 5 points (1 point shading, 1 point bottom border, 1 point upper border, 2 points for percentage with 1 digit of accuracy) |

**Problem 3 (45 points)**

| Criterion | Grading |
|---|---|
| *Extraction of title, author and format code* | 5 points each (15 points total) |
| *Correct lookup table and named range* | 5 points |
| *Correct lookup formula* | 5 points |
| *Handling of trailing spaces* | 5 points |
| *Filtering of formats* | 10 points |
| *Percentages* | 5 points |