Intermediate Excel Tips and Tricks for Manipulating Text

Introduction
This is a brief introduction to intermediate Excel tips and tricks for the business analysis environment.

I am a wayward industrial engineer who has been making a living at business analysis for many years. I have used spreadsheets professionally because often it’s the only horse I am given to ride.

Caveats
I do not endorse spreadsheets for every problem, especially data storage. I do not endorse Microsoft, though I have floated much of my career on their tools. I’ll be working in Excel 2007.

I’m going to assume that you know spreadsheet basics.

I’m going to leave out a bunch of stuff, I only have 20 minutes. I won’t be going in depth on the tips and tricks presented.

I will not be using VBA (MS Office’s Programming Language). I will present some add-ins though.

Structure
I’ll cover tips and tricks organized into general text manipulation activities:

Delimitating, Concatenating, and Cleaning Text
Changing Cases
Finding and Replacing Text
Text String Parts
Regular Expressions (RegEx)
Intermediate Excel Tips and Tricks for Manipulating Text

Delimitating, Concatenating, and Cleaning Text

Text to Column
Data>Data Tools>Text to Column splits the text contents of a cell into columns based on a delimiter or fixed width. This is helpful when multiple values are stored in a single cell but really need to be in separate columns. Caution! Once you use this tool, it changes the original cell contents and will over write any data in columns to the left of the source cell. If keeping the original cell contents is important, you will want to use formulas to split out the text instead or make a copy of the original cells.

Concatenating Text with “&” or CONCATENATE(text1,text2,...)
Concatenating is when you want to join together two or more text strings. There are two ways to do this. The fist way is to use and ampersand (&) in between each text string you are concatenating. The ampersands can make reading a long concatenation difficult. It’s often better to use the CONCATENATE function to compose the concatenation. For either method of concatenation, you can use cell references, table references, literal text, and nest other functions.

Removing Whitespaces and Non-printing Characters
TRIM(text) will remove leading and trailing whitespace. CLEAN(text) will remove non-printing characters like a line break.
Intermediate Excel Tips and Tricks for Manipulating Text

Changing Cases

Text in a different case than what you need can be a very annoying problem. Often this is a result of how the data is stored and the practice of using ALL CAPS FOR VALUES; however, a mixture of upper and lower case is more readable for humans.

**LOWER(text)** will convert THIS TEXT and This Text to this text

**UPPER(text)** will convert This Text and this text to THIS TEXT

**PROPER(text)** will convert THIS TEXT and this text to This Text

Sometimes, we only want some of the words in a string to be lower, upper, or proper case. In these cases, you can use other text functions to be more specific about what part of a string you want to change.
Intermediate Excel Tips and Tricks for Manipulating Text

Finding and Replacing Text

A very common text task is finding some text within other text and then replacing it. There are two ways of doing this.

Find & Replace Tool

The first is to use the **Home>Editing>Find & Replace Tool (shortcut is Ctrl + F)**. This allows for both a search and replace. It allows you to see and navigate to all the places where your string was found and it supports basic “wildcarding”. The down sides are that it changes the original data and does not support nesting formulas or cell, table, or range references. Similar functionality is possible in columns where filtering is turned on.

FIND() and SEARCH()

Both of these functions find the starting position of one text string in another.

- **FIND(find this text, within this text, [start looking here]):** Case Sensitive, Wildcard not allowed.
- **SEARCH(find this text, within this text, [start looking here]):** Not Case Sensitive, Wildcard allowed.

Find and search are not just useful for finding the position of text but for determining if a term or phrase is present in a text string like a comment field.

SUBSTITUTE() and REPLACE()

Both of these functions replace text in a text string with alternate text.

- **SUBSTITUTE(in this text, replace this text, with this text)** replaces specific text in a text string.
- **REPLACE(in this text, start here, then replace the next so many characters, with this text)** replaces text at a specific position in a text string.
Text String Parts

Once you can determine the position of specific text within other text, you can use a set of functions to return parts of a string.

LEN(text) returns the length of a text string.

LEFT(text) returns text starting from the beginning of a string to some end position.

MID(text, starting position, number of characters) returns text from a position to some number of characters within a text string.

RIGHT(text) returns text starting from the end of a text string.

All Together Now! By combining the locating functions with the string part functions, you can cut up and return specific parts of a text string.
Intermediate Excel Tips and Tricks for Manipulating Text

Regular Expressions (RegEx)

The regular expression "((R|r)egu\la(E|e)xpressions?)|((R|r)eg(E|e)x)" will match the text strings: Regular Expressions, Regular Expression, regular expressions, regular expression, RegEx, or regex.

Regular Expressions are like super wildcards. They are a feature of programming languages that allow for powerful text pattern matching; however, the syntax is very cryptic and requires a learning curve. Regular expressions are not a default feature of Excel, but they are available through VBScripts that are part of Excel. You can turn them on manually and write user defined expressions to expose them as detailed here, [http://lispy.wordpress.com/2008/10/17/using-regex-functions-in-excel/](http://lispy.wordpress.com/2008/10/17/using-regex-functions-in-excel/), or you can use and add-in like the one here, [http://www.codedawn.com/excel-add-ins.php](http://www.codedawn.com/excel-add-ins.php). The add-in has a new find and replace wizard. Both expose four new regular expression formulas:

- **RegExFind** returns text in a string that matches a pattern.
- **RegExReplace** replaces text in a string that matches a pattern with alternate text.
- **RegExTest** returns True if a pattern is matched in a text string.
- **RegExMatchCount** returns the number of times a pattern is matched in a text string.