

The Editors' WebWatch

The Editors' WebWatch is a membership-driven resource guiding editors and writers in the sciences to websites and services of interest. Suggestions for the February issue should be sent to ese.webwatch@gmail.com. We are using the Editor's Bookshelf blog at <http://ese-bookshelf.blogspot.com> to collect entries; contributions are welcome.

Regular expressions

<http://office.microsoft.com/en-us/help/HA010873051033.aspx>
<http://office.microsoft.com/en-us/help/HA010873041033.aspx>
http://wiki.services.openoffice.org/wiki/Documentation/How_Tos/Regular_Expressions_in_Writer

One of the most powerful features of any text-editing package is its support for regular expressions. Using regular expressions as part of a find-and-replace operation enables you to make systematic edits to a document with remarkable speed.

Querying a search engine for something like "regular expressions" produces lots of helpful results, but in Microsoft Word and OpenOffice the syntax used for regular expressions, or regexes for short, is rather different.

Say one wanted to change "vapor" into "vapour", but only in those cases where it was a word on its own, not, for example, in "vaporous" or "evaporate". `<(vapor)>` (Word) and `\<vapor\>` (OOo) catch these cases, where the `<` and `\<` stand for the beginning of a word. Likewise, word-final "-ize" can be identified by `(ize)>` (Word) and `ize\>` (OOo). This, however, would be an unwise thing to do as it would also catch "size", "prize", "seize" and so forth. One can restrict the search to words longer than, say, six letters by writing `[a-z]{3,}(ize)>` (Word) and `[a-z]{3,}ize\>` (OOo), where the square brackets introduce a set of characters to be matched (this can be as simple as a single character,

`[n]`, or a range, `[A-Z]`, or a mixture, `[aeiou0-9]`), and the braces indicate the number of characters to be matched, in this case at least three.

In general, the trickier the task being asked of the regular expression, the more similar the syntax in Word, OOo, and in programming languages such as Perl.

The above examples might seem somewhat contrived, but regular expressions come into their own for tasks such as rearranging formulaic subject matter such as addresses and bibliographies, and the Microsoft Word-specific pages are particularly focused on the tasks of rearranging names and dates. It's confusing that the full stop in Word regex syntax stands for a full stop, whereas in OOo it stands for any character, while in Word a question mark stands for any character and in OOo it means zero or one of the preceding character.

In short, regular expressions are so powerful that it is a very good idea to practise on test documents and compare the results carefully with the original before trying them out on a live piece. But they may well change your life.

As with many things, this has apparently changed in Word 2007 – but that is a matter for another column.

What is this file?

<http://extensions.pndesign.cz/>

RAR, SIT, CDR... A special case of three-letter abbreviations, and one that plagues technical editors, is the file extension. Authors are endlessly inventive at submitting files in peculiar formats, and the first question you need to answer is "What program did they use to generate that?" This page answers that quickly and neatly. The site is lightweight and simple to navigate.

I should note in passing that while Wikipedia is always something of a lottery, documenting the peculiar programs that can be used to open peculiar files is one of its strengths.

Abbreviations for editors

<http://www.nactem.ac.uk/software/acromine/>

Anyone with experience of lists of abbreviations and acronyms will have spotted that they're seldom up to date and often contain abbreviations and acronyms which, from a cursory internet search, seem to exist only in lists, rather than out in the wild. So an abbreviation list that is somehow automatically generated from current material would be extremely welcome.

AcroMine has been around for a few years but you may not have seen it before. The idea is to take all of PubMed and look for word sequences that regularly co-occur with expressions in brackets that match.

But how well does it work across disciplines? My first attempt was a term used in nuclear magnetic resonance spectroscopy: INEPT. AcroMine correctly identifies this as "insensitive nuclei enhanced by polarization transfer". AcroMine offers 22 hits for "MMR"; the most common is, surprisingly, not the vaccine against measles, mumps, and rubella but "mismatch repair". AcroMine even correctly offers "Large Hadron Collider" as an expansion for "LHC", demonstrating that these days PubMed is a resource for scientific, technical, and medical editors of all disciplines.

Online biochemical and chemical nomenclature

<http://www.chem.qmul.ac.uk/iupac/>

This website may look old-fashioned, but it will still be invaluable when the likes of Facebook are a distant memory. Here on a single page are many of the recommendations of the International Union of Biochemistry and Molecular Biology (IUBMB) and those recommendations of the International Union of Pure and Applied Chemistry (IUPAC) that are particularly relevant to editors working in the life sciences.

Blog update

<http://scienceblogs.com/notrocketscience/>
<http://www.badsience.net/>
<http://david-crystal.blogspot.com/>

Many journal publishers (BMJ, OUP, NPG, RSC, ACS) are using blogging to promote their own articles, but I thought it would be more interesting to look at some contrasting blogs from the very different worlds of science and medicine.

Ed Yong's *Not Exactly Rocket Science* concentrates on readable accounts for non-experts of peer-reviewed published material. Readers outside the UK might be unfamiliar with Ben Goldacre's *Bad Science*, which is both a column in *The Guardian* and a rather longer and more-often updated blog covering

how the media reports medicine. It is a sobering but immensely entertaining read.

Lastly and most relevantly for people whose job is working with language, David Crystal, scourge of prescriptivists, has a blog about how language *actually* works which is well worth following and learning from. An excellent comment on style guides is worth repeating in full:

Style guides should be explaining to people what English allows us to say and write, and pointing out the strengths and weaknesses of different usages in different contexts. Blanket bans are a nonsense.

Physics and physical chemistry

<http://old.iupac.org/publications/books/gbook/>

The International Union of Pure and Applied Physics (IUPAP) doesn't put its recommendations online, but IUPAC's Green Book, which is based in part on the recommendations of IUPAP and also on ISO 31 (now superseded by ISO 80000-3:2006, which costs only CHF 96), is online as a PDF of its second edition, and the third edition will appear online soon.

It's an indispensable *vade mecum* for editing equations and the kind of mathematical expressions that one sees in running scientific text and was immensely useful throughout my doctorate and my years as a technical editor.

Colin Batchelor
 batchelor@csc.rsc.org

With thanks to Richard Hurley.

Technical Tips: Synchronizing files on different computers

"My Briefcase" in Windows is useful for synchronizing files between one or more computers (eg at work and home) and an intermediary pen drive (or other electronic media). You can work on files in either location - the Briefcase or the pen drive - and then with a single click at the end of your session synchronize the copies (and then do the same on your other computer).

Start by having one Briefcase on computer A and another on computer B and then synchronizing each with the files on the pen drive - rather than having a single Briefcase on the pen drive and then synchronizing files in each computer with what's in the pen drive. This is because the file locations of the documents you are synchronizing would inevitably be different on the two computers.

So on each computer you need to:

1. Right click in the folder where you want to locate the Briefcase, eg desktop or "My Documents" then go down to "New" and then "Briefcase".

2. Rename the Briefcase to something memorable, eg "Current work" or "Portable files".

3. Select (highlight) one or more documents from the pen drive that you want to keep synchronized. You can select whole folders and subfolders if you want. You then drag the items into your Briefcase, but using the *right click* button, not the left. You are then offered the choice of "Make synch copy", "Move", "Create shortcut", or "Cancel". You need to select "Make synch copy".

4. Later do the same thing on computer B.

You can now work on documents in either location (the Briefcase or the pen drive). At the end of the day or whenever you want, you insert the pen drive, open the Briefcase (double click on it, or you can create a shortcut on your taskbar and click on that), and click the button with file icons and green arrows going in opposing directions ("Update all"). This synchronizes all documents — or you can update documents singly by using the button to the right.

Hitches etc:

- * You have to close any document that you want to synchronize before you "Make synch copy".

- * The non-Briefcase location obviously has to be on the pen drive. Clearly you can make "synch copies" in your Briefcase from documents in another folder on the computer, say, "My Documents", but then you won't have copies in your pen drive to synchronize with computer B. "My Briefcase" is useful only for things you'll be working on over time in both locations.

- * Very rarely (maybe twice in the several years I've been using it) the Briefcase seems to become corrupt, and you have to delete it and create a new one. Not too much of a problem, as you will still have copies of all the files in the other location, though they may not be up to date, depending on how recently you have clicked the "Update all" button.

Karl Sharrock
 Technical editor, BMJ
 ksharrock@bmj.com