

Re: text to bibliography?

Source:

<http://www.tech-archive.net/Archive/Word/microsoft.public.word.docmanagement/2008-08/msg02304.html>

- *From:* p0 <yves.dhondt@xxxxxxxxxx>
 - *Date:* Mon, 18 Aug 2008 02:23:29 -0700 (PDT)
-

convincing any user of reading even 25 pages before he can start using a program, let alone 500. People just aren't patient enough anymore for looking up things in a help file (see also my comment below on creating new formatting styles).

No, tht's not it at all. You cannot use a "Help" file unless you happen to know the exact name that the writers of the "Help" have assigned to a feature/bug. Look how many times a day it is asked here how to get rid of the dots between words, or why there's suddenly no vertical space between the pages.

Help files are perfectly searchable nowadays. You write about the 'dots between words' question. It is indeed a frequently asked question.

So I just started Word pressed F1 and entered the following in the help box "dot words" (without the quotes). And guess what, the 8th entry in the list of results is titled: "I see dots and arrows in my document". I click the link, and yes, it tells me exactly how to get rid of those dots. It is there in the help, seconds away for people to find it. Yet they still go to newsgroups to get an answer to their question which they could easily find themselves. And they do get the answer in here. What is more, the question and the answer in the newsgroup are both indexed by Google. A 2 second Google search would prevent the next person from asking the exact same question.

Note that this is not a complaint about the (quality of) posts in a newsgroup. I merely wish to point out that people don't read help pages or look for answers already out there. And remarks like 'but it is easier to just ask the question' don't count. Newsgroups are slow. They sometimes have to wait 24 hours for an answer they could have found in 2 seconds.

Re: text to bibliography?

And actually, your "I do?" remarks are plain examples of you not willing to look for help on the subject. I don't blame you for not wanting to look things up, but at least don't say the answers aren't available.

That aside, the bibliographic tools of Word 2007 lack almost all documentation; the promised SDK is almost a year overdue now (I doubt it will ever be released); and non of the people originally working on the academic features seem to be still doing that job nowadays.

What's an SDK? You use their jargon, you must be one of them!

An SDK is a Software Development Kit, and it is a term widely used in the software business. It is most certainly not coined by Microsoft.

It doesn't seem too much to ask that "Text to Table" could come up with a tabular presentation, which some other module could then convert to the "format" used by the bibliographic database: if it knows that col. 1 is the author, col. 2 is the date, col. 3 is the title, col. 4 is the place, and col. 5 is the publisher (that's a basic Book entry), why can't it simply do that?

Now you are no longer talking static text, you are talking (poorly)

"Poorly"? CMS has been around since 1906 and is by far the leading style guide in the US.

Re: text to bibliography?

If I would ask the medical doctors what the leading US style guide would be, they would say AMA.

If I would ask the psychologists what the leading US style would be, they would say APA.

If I would ask the legal people what the leading US style would be, they would say Bluebook.

If I would ask the average school kid writing a little science paper what the leading US style would be, they would say Turabian.

Gotcha. Turabian is based on Chicago (which she was the editor of for 50 years or so).

Gotcha? Does it matter which one is derived from which one? They are different. Hence, any static text parser would have to work differently on both styles.

The other three probably do not predate 1906.

And if you would ask the rest of the world what the most commonly used style would be, they would probably say Harvard (which is also the oldest one if I'm not mistaken).

I'm not aware that a style called "Harvard" is used in the US. In what publication is it codified?

I personally don't use it. But it is used in almost all science fields in Western Europe and the British Commonwealth (so that's including Australia and New Zealand). So I think it is probably the most widely used system.

I am not trying to say that CMS isn't important or widely used, it is just that everybody feels that his or her style is the most important and commonly used one while it isn't.

On a side note, of the above list, only APA and Turabian are supported in Word 2007.

I've noticed that. Kinda leaves the humanists, who tend to use MLA, up the creek.

Re: text to bibliography?

Not really, Word 2007 supports MLA out of the box.

formatted text. And you would have to have a tool to map columns to fields, since in my case, year should be the last entry (except maybe for pages) in my bibliography and most certainly not the second.

So you don't do author-date references in the text? Fine. That would be Chicago's "Humanities" style.

The style I use is not supported by Word 2007 at all. I did write the transformation stylesheet for it from scratch.

And in your case, how is your book displayed if it is an anonymous work? I would guess col. 1 is the title, col. 2 is the date, col. 3 is the place, and col. 4 is the publisher. So even between 2 entries of the same type, the ordering of data would be different.

No, col. 1 would be empty. (Though there are circumstances in which the author is entered as "Anonymous"; see CMS.)

And how do you expect your static text parser to guess that column one is empty? Once you start adding delimiters, you can just as well use the delimiters Microsoft defined. Those delimiters being xml tags. They might not be what you prefer as delimiters, but they are delimiters.

I don't know what a "static text parser" is. Did you again forget that I've put tabs between the fields, in order to do Text to Table? (The

Re: text to bibliography?

punctuation between each pair of fields differs through each paragraph, so it can't go by comma or period or colon.)

Static text is text without any kind of markup or delimiters indicating clearly where fields start and/or end. It is what you would call text without any codes.

Once again, you want to use your tabs, Microsoft wants you to use their tabs, their tabs being the xml tags I showed earlier. It is not because their tabs are longer than yours (and a lot more descriptive), that they are worse. It is all a matter of taste.

Maybe you don't have anonymous works, but it doesn't matter. What you require is so specific that you will probably be the only one using the 'import filter' anyway. The point is, Microsoft provides a set of generic tools which works for 80% of their customers. There is no

You are again losing sight of the point. They provide no tool for going from an existing bibliography to the bibliography database.

The tools are there, they are just not obvious in use for the average Word user.

You just recently told me that it's not possible.

It is possible, I showed you which tags to use in the simple example I gave in one of the previous posts.

However, to automate the entire process, the input format has to be perfectly known. That is, no single exception can be left aside (anonymous works, corporate authors, ...). Once you have fully defined your format, all you have to do is provide a mapping between your fields and the fields defined by Microsoft.

So is it possible to do it in an automated way? Yes. Is it doable? No.

Re: text to bibliography?

Re: text to bibliography?

There are so many versions of every style format that your 'translator' would be either just working in your specific case, or be a huge monster which takes years to make and would even then not cover some exceptions. Microsoft decided not to create the monster (and I can't blame them). Instead, they decided to give you the tools to create your translator for your specific case. But for someone without any programming skills, those tools are too hard to use.

The format of a b:Source element is entirely defined by an xml schema. All you have to do is write a (simple) XSLT which transforms your format into the format described by that schema. Of course, if your format happens to be an incomprehensible static text, your XSLT will be very complicated. But you can not blame Microsoft for that.

I have no idea what a "b:Source element," an "xml schema," an "XSLT schema," however simple or complex, or an "incomprehensible static text" may be.

And that is the main problem. I do not blame you for not knowing them. But they are available to you. If you want to learn how to use them, you can. It is all about reading the documentation on those technologies.

I never used XSLT before I started using Word 2007. It took me a couple of hours to figure out how it worked and I could start creating my own stuff. I agree that coming from a computer science background gave me an advantage, but still ... I had to start from zero.

point for them in developing a tool which will work in a very specific case (yours) and therefore will target 1% or less of their customer base. If you want one, you will have to write it yourself. They provide a specification of the bibliography format and even provide a programming interface (I have no experience with it). They try to help you a long way, but the last few steps you will have to take yourself.

On the occasions when programming new reference styles has been mentioned here, the MVPs have stated it appears to be

Re: text to bibliography?

impossibly
complicated to do so.

It is not. It is pretty basic XSLT, nothing fancy at it.

Like you said above, all people have to do is read the available
help:

* you have the open xml specification;

I do?

Yes. It is an ECMA standard (and now even an ISO standard). The
specification is open and freely available.

ECMA: <http://www.ecma-international.org/publications/standards/Ecma-376.htm>

Microsoft: <http://msdn.microsoft.com/en-us/office/aa905545.aspx>

ISO: they are still finalizing the text

* you have blog articles by Microsoft people;

I do?

Yes. Probably the best example out there to get you started on
creating your own bibliographic style is

http://blogs.msdn.com/microsoft_office_word/archive/2007/12/14/bibliography-citations-1011.aspx

but there are others.

* you have MSDN articles describing the format (not extensively
though);

I do?

Yes. For example <http://msdn.microsoft.com/en-us/library/bb258052.aspx>

Re: text to bibliography?

Re: text to bibliography?

* you have 10 predefined styles, each consisting out of a couple of 1000 lines of XSLT code (that is over 10000 lines of example code)

That's an awful lot of code.

Yes, that is an awful lot of EXAMPLES. Since when is having too much examples a bad thing? Besides, if you want an example of it being broken down to its bare minimum, you can check the blog article above.

So there is plenty of information around. Maybe it is not perfectly organized, but it is there if you want to learn how to use it.

But the MVPs are correct, as long as there is no point and click solution, it is too complicated for the average Word user. And no matter how much help files you are going to add, it will remain too complicated.

Yet somehow I didn't find Papyrus the least bit complicated --- though apparently it's too much for you??

I haven't tried it, I just pointed out that it comes with a lot of documentation. Word isn't complicated to use either if you stick to the basic tasks. It still comes with a huge documentation though.

I will soon find out how greatly it respects CMS style, especially for complicated entries.

Well I never use the style, but since Word only defines one version, and the Chicago style is different for different research fields, I would not get my hopes up if I were you.

Since the 14th ed., CMS has had two different and parallel schemata, the old humanities style, and the author-date style favored in the social sciences. The U of C Press does little or nothing in hard sciences that might need other provisions. (And the 15th has grown intolerably permissive, perhaps as those who knew Mrs. Turabian --- unfortunately I never met her --- themselves retire.)

Re: text to bibliography?

This is the last post I make to this thread, because I feel we are starting to just argue for the sake of arguing.

To come back to your original question: "text to bibliography?" Yes it is possible to automate that process but highly complex and therefore 99% of the people out there will not be able to do it and the practical answer is: No.

If you have specific questions about changing existing styles or need help on creating your own style, just post a message to the newsgroup and if I am around, I will try to help you.

Yves

.