

Dover Society Newsletters

Paul Skelton

Being a bona fide 'Dover Shark' (one who was born and bred in Dover) a couple of years ago I decided that it was about time I joined the Dover Society and did so in 2017. I was very glad to receive my first edition of the newsletter, which I believe was number 90, November 2017. "What a wonderful read," I thought, and wished I had joined earlier and wondered just what I had missed in previous issues.

Two years later, at the start of 2019, I was contemplating asking the committee whether they had any back issues that I could 'borrow' in order to catch up on missed articles. Then saw in the latest newsletter that our Membership Secretary, Sheila Cope, had a surplus of old editions that she wanted to reduce and she said she would offer them around at subsequent meetings. What an opportunity for me to gain these wonderful archives. What was also made clear, though, was that if someone was looking for a particular article, at present it was impossible to find it as there was no full index, and she asked whether there were any volunteers who would like to take on the task of creating one?

I envisaged someone sitting down in front of a desk, wading through these newsletters page by page, and making notes regarding the contents on a rolodex or some sort of database on a computer, and then trying to put them into an order that could be searched.

That sounded like time consuming work to me and most certainly a labour of love, especially to get it right. But in a mad

moment I decided offer my services, much to the dismay of my family and friends, who told me that I didn't have the time, that it was too much work and that I'd never get it finished, along with all the other activities I was currently doing.

To give a few statistics of the size of this project there are 94 editions of the newsletter to date, most containing over 50 pages and, a quick calculation tells me, over two and a quarter million words in total.

They say there's a small line between insanity and genius. They told me I was barmy in taking on this task, being in full time work as well. Surely it's something for someone who is retired? Then again you have only got to see me in the street to know that I am a little bit unorthodox and find it difficult to hide in a crowd. Eccentric, some call it, not quite the norm others. I think I know which side of the line I stand.

The task that I believe they requested, wasn't quite what I had in mind for an end result. I wanted better. I wanted to reference not just every article, but every word, 2.5 million of them. What the devil am I taking on here? Madness has finally caught me, surely. I always seem to take on huge and seemingly impossible tasks.

As soon as I picked the editions up, I knew this would be a bit of a challenge, but it was Easter, and I had two weeks off work as the schools were closed, so I knew I had a bit of free time. I had another thing in my favour as well. A computer, scanner,

and programs that would help me with this task and, better still, I had done similar work before.

But just how much work had I actually taken on? I'll let you decide with the following account of what was necessary. I will keep this as technically simple as I can, by the way.

My first task was literally to scan every page to capture a digital image. The last few copies, I was glad to find, were already available as PDF files (Printable Document Format), so they were ready for computer manipulation. The paper newsletters already being bound, I scanned two pages at a time, side by side. That totalled no less than 2,279 scans. However, I really wanted every page to be able to be viewed separately and in the correct order, not two at a time. I could have scanned each page separately, but that would have doubled the scanning time. I needed every page on its own, and started to do this with an image manipulation program Paint Shop Pro (very similar to Photoshop). However, the next program I used would do this automatically for me I discovered, thus saving more time in splitting over 2,000 images I am glad to say.

This next program was an OCR (Optical Character Recognition) program, which literally identifies the scans I had produced and distinguishes between pictures and text. This is very important and makes the contents of the files it produces searchable. The images were now not just bitmaps, (a picture made up like a huge sheet of graph paper with every square coloured in a different colour - you only have to enlarge one to see what I mean by this), but the part of the scan that looks like a letter is now actually recognised as a letter of the alphabet and

not a series of dots. This program not only does this, but also splits the images where there are two pages on a scanned image, giving separate pages, and also allows me to change the order if required and rotate them too. Of course, I had to move the scan of the front and back cover into the correct order. There was also one edition that, for some unknown reason, had its pages printed in the wrong order. That has now been corrected. The program referred to, but the way is ABBYY Fine Reader, but there are others on the market.

Having now processed all 94 newsletters, put them in the correct order, page by page, and converted them into a computer readable format, I saved each complete newsletter as an individual PDF file, each individually searchable on the computer using the "Find command (Ctrl F)" from within the program that will open the file.

However, this still wasn't the end result I wanted and I still didn't have the facilities to search all newsletters 94 at once for a required article or, indeed, author or word.

I also wanted this to be user friendly, otherwise who would know what to do with a disk containing nothing more than files with titles such as 001.pdf, 002.pdf and so on.

Next was to create an index page that would allow you to open each file on any web viewer such as Internet Explorer or Google Chrome, (or Safari etc. for those who prefer an Apple machine), and to do this I needed to go into web design mode and write some HTML (Hypertext Markup Language) files. This actually isn't as difficult as it sounds, I use a program called Microsoft Frontpage, although if you know your coding you could do this in a program as basic as Notepad.

I then had another file burnt onto the disk that contains a menu system viewable just like a website, but which will run from the disk, so is transportable to any computer.

Now, I have already said that each file is individually searchable using the "Find" command, and we all know that you can also search for things on the web using a program like Google, but Google doesn't know this disk exists, and I wanted all files to be searched at once, directly from the disk and for any word, just like Google.

I have another program called Zoom Search Engine Indexer, which will look at all the PDF and web files on the disk, and create its own Search Engine from the data now contained within those files. Those files are also saved to the disk and will run it just like the Google search engine so I wrote a little more HTML into the menu to point to that search file.

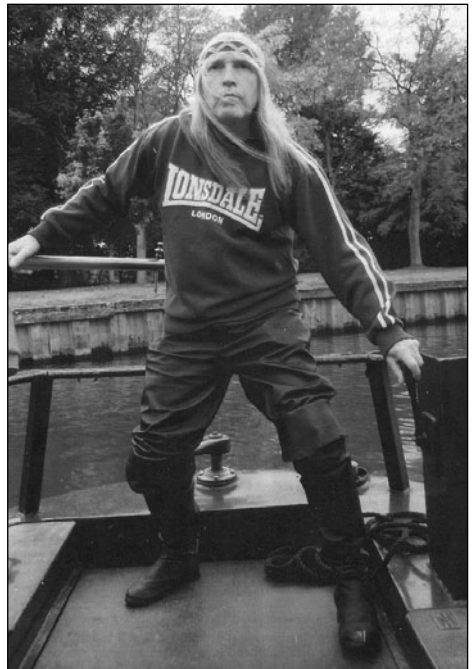
Here I need to explain in more detail what this search engine allows you to do. Should you, for instance, wish to find a specific word or words, this search engine will tell you every newsletter in which they exist and allow you to open that newsletter. Although quick, that's not as good as I really want, as you would then have to do a secondary search using the "Find" command to find where in that file your chosen word occurs. But, again I have the solution. And another mad moment, probably, as surely I have completed the original task. Yes, but not to the standard I had planned.

The next stage may seem to some like as if I was going backwards, but this was necessary, unfortunately, to enable me to make the search results user friendly. I used a program to split each newsletter into individual pages again and save them as individual files, this time not scanned

JPG picture files but as searchable PDF files.

Then came the manual work I was trying to avoid at the beginning, I had to identify each article but, luckily for me, each edition of the newsletter has its own "contents" page. My original menu identified only each edition by date and opened that newsletter with a click of the mouse, I changed those links to open a sub-menu containing the contents list of all the articles that appeared in each edition.

From my now split files just saved, I collected only those pages for each article, Again, is written on and kind of stitch them back together again and save that as another file. Thus producing an individual file for each article. AMMY does this very well. This secondary menu points to each



Paul Skelton

article, allowing you to access just that and not the whole newsletter. Of course, the search engine will now find not only the edition in which the chosen article was written, as I still have a file containing the complete newsletter as printed, but also individual files containing only that article. This method also allows me to remove all the advertisements, if they split an article in half.

Some may think that work is now complete, but I can guarantee that although it has been done with utmost care, there will always be a few errors, so the next step is to test that all the links work. I am glad to say that, although this could be done manually, with the aid of a program called Xenu Link Sleuth I am able to check automatically that every link points at least to an article. What can't be checked, though, is that they all point to and open the correct file. Although the

menu from the search engine is accurate, links from my web based file menu will have to be checked manually, but I'll leave that to people using the disks and ask them to report back to me any anomalies found.

So, that's all it takes really. Oh, and, if I have created the disk correctly, adding one last file, it should open automatically when you put the disk into your computer, presenting you with the main menu in your chosen internet browser.

My moment of madness is now complete I reckon. It's now over to the committee to decide whether they want the entire disk's contents uploaded to their web site to be available to the world, or whether they wish to produce disks to be used without internet access for anyone who wishes to acquire a copy. This should work either way.

Dover Society Badges



Available at £2 each

Contact the editor at:

**8 Cherry Tree Avenue
Dover, CT16 2NL
Tel: 01304 213668**

Email: Alan.Lee1947@ntlworld.com

WEARING A BADGE ASSISTS IN OUR DRIVE FOR
INCREASED MEMBERSHIP

FABRIC, YARN & HABERDASHERY

**Dover
Knit and Stitch**

**Unit 6, Charlton Centre, High Street
Dover CT16 1TT
Tel: 01304 208491**

E-mail: doverknitandstitch@outlook.com

Facebook: Dover Knit and Stitch