

The *EJIW* index: were new publishing technologies a burden or a bonus?

Caroline Diepeveen

Caroline Diepeveen reports on the making of the Wheatley Medal-winning index to the Encyclopedia of Jews in the Islamic world (*EJIW*), the problems that had to be resolved along the way, and the pleasures and the promise for the future of indexer teamworking. She takes a look at the online version of the index, and finally offers thoughts on why indexers must embrace the new technology publishers are introducing to ease their production process.

Introduction

The index to the *Encyclopedia of Jews in the Islamic world* (*EJIW*), published by Brill (Leiden), won the 2011 Wheatley Medal for excellence in indexing. In the early stages, this seemed a highly unlikely outcome. In this article I describe the unusual *EJIW* production process, and what it meant for the indexing process as an illustration of the technologies that publishers are currently developing to make the publishing process more efficient and to allow them to publish simultaneously on multiple platforms.

Brill

Brill is a comparatively small Dutch publisher, founded over three centuries ago when Holland (a province of the Netherlands) was flourishing and many people from the southern Netherlands and beyond came there because of its freedom of the press, allowing them to publish freely without fear of prosecution. The Dutch publishing industry flourished as a result. Nowadays, Brill has a very sound worldwide reputation as a scholarly press. It does not need convincing of the value of indexes. All Brill books have indexes, mostly of good quality. It publishes many reference works that are usually also available online. Quality has a price, though. The books are certainly not cheap, and Brill's online subscriptions are very expensive.

In preparation for my presentation at the SI Brighton conference 2012 and this article, Brill allowed me temporary access to the *EJIW* online publication. A complementary copy of the whole print *EJIW* version was already in my possession. This allowed me to compare the hard copy index with the online index, which proved to be an interesting exercise. I discuss this further below.

Team indexing

The *EJIW* index was compiled by a team of three indexers. All three of us look back on this as a happy experience. Many fellow indexers expressed their surprise that team indexing could ever be a possibility, let alone be a happy experience. To be honest, when we were doing it, we never thought twice about it. As I explain, it came about under pressure, but I am

now convinced that new publishing technologies open the way for team indexing on large projects.

The content management system (CMS)

The *EJIW* publication consists of five volumes, four volumes of alphabetically arranged encyclopedia articles and one volume (vol. 5) of resources. The index forms the main part of Volume 5, consisting of 350 of the 500 pages of this volume. To organize work on this large publication, Brill used a content management system (CMS), development of which was outsourced to a software development firm. The CMS included an indexing module.

When I tested the indexing module it was immediately apparent that it was grossly inadequate, and had been developed by people who had no idea about indexing. My Brill editors made it clear that indexing outside the CMS was not an option, but they were willing to look at improving the module. In the first version of the module you could select a term in the text (or several terms, but not ranges) and you could then replace this term or terms with your own preferred index terms. No locator reference or ID was assigned to the term, up to one level of subheading was allowed, and I was told the index output was in some sort of Excel.

Unique locators/IDs

First of all, I felt that a unique locator/ID number was needed for each anchor/tag placed in the text, preferably one that would also tell me where in the text the term was located. The encyclopedia articles already had a number/letter combination assigned to them in the CMS. The letters referred to the editor who was responsible for the article in question. These could be translated into numbers unique to each article. Within the article each index term would be given a sequential number, for instance 001 for the first index term in that article. The index ID number would then consist of the article number plus the index term number. Most articles were short (usually 300–600 words), so it would be easy to locate the index term number when editing the index. The ID numbers would be allocated automatically by the CMS.

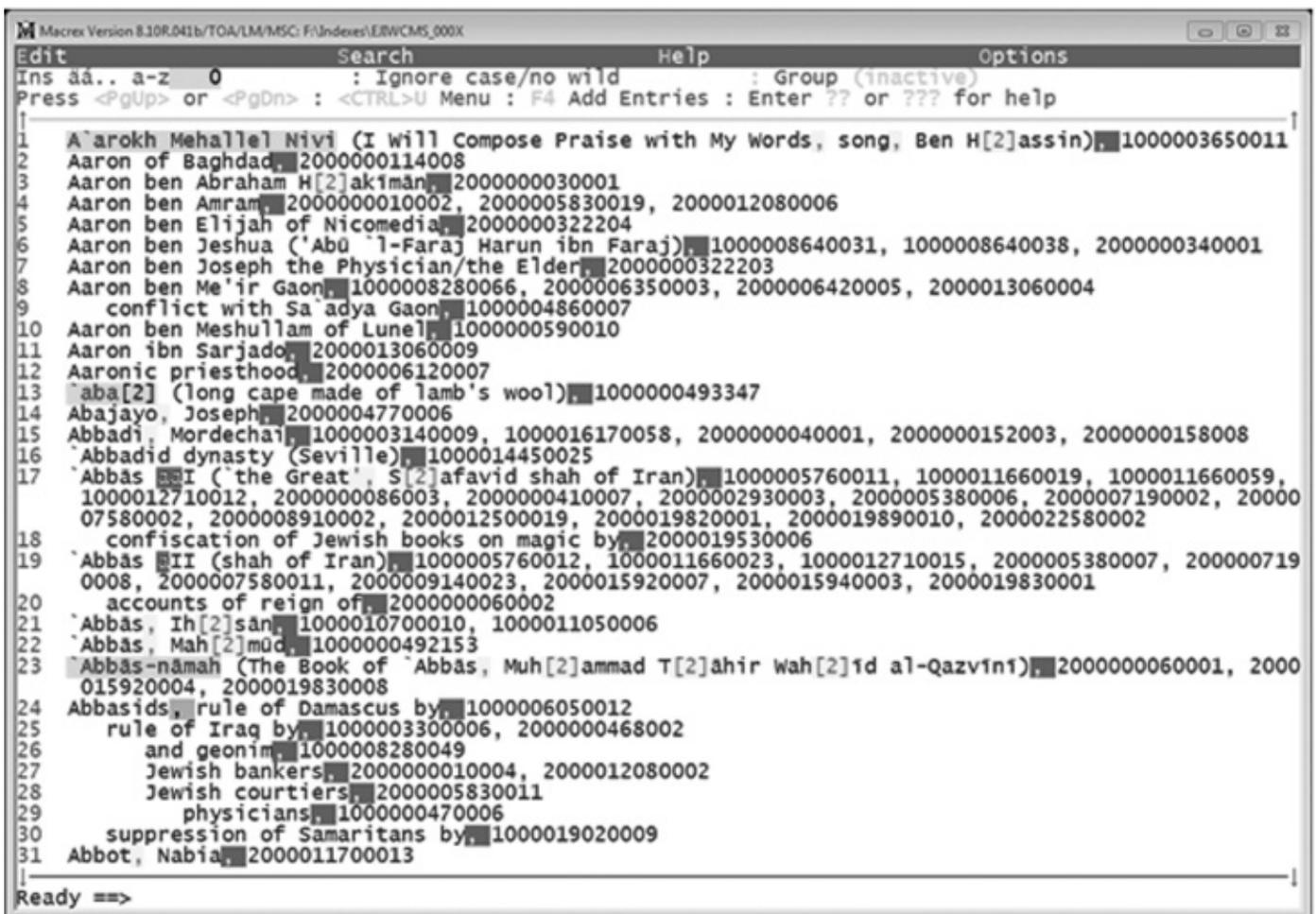


Figure 1 Macrex screenshot of the EJIW index with CMS locator numbers

Ranges

We decided to dispense with ranges. It would make things technically too complicated and as most of the encyclopedia articles were short, ranges were not really needed. In the longer articles, I repeatedly selected the same term in the same article in the expectation that this would result in page references covering the whole article.

Unfriendly software interfaces

Our original intention was to do the term selection in the CMS, generate an Excel™ output, read that into Macrex, edit in Macrex, produce an Excel output from the Macrex file and feed that output back into the CMS. The Macrex developers, Drusilla and Hilary Calvert, helped with this process, and we were able to read the Excel CMS output into Macrex and produce Excel output from Macrex. Unfortunately, the software firm hired by Brill was unable to adapt the CMS to allow the Excel output to be fed back into the CMS.

When this became clear, we decided to abandon the idea of a fully embedded index (that is, one using terms embedded in the text to generate the index automatically as an integral part of the text). We would go for tagging instead. I would copy the index ID numbers into Macrex

and would produce a Macrex index in the usual way. The index ID numbers would be converted into page numbers in the Philippines (where most of Brill's labour-intensive work is done). Producing the online index from this would also be fairly straightforward. I had hoped that the ID numbers, which were quite long of course, could be copied to the Windows™ clipboard. Unfortunately this also proved to be beyond the capabilities of the CMS, so each index-ID number had to be manually copied and pasted into Macrex. (There is a key stroke to increment the locator by one in Macrex, but still you would have to be careful not to make mistakes with such complicated numbers.) The index was sent to the Philippines where the locators would be converted into page numbers. A sample of the hard copy version of the EJIW index is shown in Figure 2.

The work flow

The original idea was to provide me with a more or less steady flow of encyclopedia articles to index. The 'tagging' approach meant that the articles did not have to come to me in the correct order. Articles that were ready could be marked 'for indexing' and I would see them once I was logged into the CMS. I would also get an email message notifying me each time new 'for indexing' articles were placed in the CMS. The 'for indexing' mark meant that

A	
<i>À bout de souffle</i> (Breathless, film, Godard), music of, 4.397	religious intolerance of, 4.438
<i>A'arokh Mehallem Nivi</i> (I Will Compose Praise with My Words, poem, Ben Hassin), 1.389	persecution of Samaritans, 4.232
Aaron, spiritual relationship to Moses, 4.357	rule of
Aaron of Baghdad, 4.411	forced conversion of non-Muslims, 4.690
Aaron ben Aaron, 3.14	interfaith relations, 2.570
Aaron ben Amram, 1.1, 682; 3.30	Iraq, 1.313; 2.603; 3.55
Aaron ben Asher, 4.437	economic decline, 2.606
Aaron ben Elijah of Nicomedia, 1.460	and Jews, 2.603
Aaron ben Jeshua see 'Abû l-Faraj Hârûn ibn Faraj	courtiers, 1.682; 2.608
Aaron ben Joseph, 1.460; 3.95	bankers, 1.1; 3.30
Aaron ben Mashiah, 3.79	physicians, 1.32, 683
Aaron ben Me'ir (gaon of Palestinian yeshiva, 10th cent.), 1.2; 2.36, 41, 264; 3.174; 4.9, conflict with Sa'adya Gaon, 1.558	geonim, 2.263
family of, 3.32	peasants, 2.604
Aaron ben Meshullam of Lunel, 1.49	and Jews, astronomers, 4.270
Aaron the Elder, 3.178	Palestine, 4.5, 13
Aaron Hâkimân, 1.2	Syria, 4.433, 438
Aaron ibn Sarjado, 3.174	Damascus, 2.6
Aaronic priesthood, 2.18	Abbot, Nabia, 2.636
'abî (long cape made of lamb's wool), 1.638	Abbou, Isaac, 3.474
Abajayo, Joseph/Yosef, 1.524, 3.586	Abdoud (Abud), Hâyyim Saul, 3.505
Abalioğlu, Yunus Nadi see Nadi, see Yunus	'Abd Allâh (Boabdil, amir of Granada), 2.351
Abûlqâ (II-Khanid ruler, 1265–1282), 2.583	'Abd Allâh (Sultan of Granada), 1.183; 2.308, 488
abargualas (ululations), 4.492	'Abd Allâh al-Ghâlib (Mawlay) (Moroccan sultan, 1557–1574), 1.400; 3.344
Abbadi, Mordechai, 1.3, 340; 2.105, 348; 3.504	'Abd Allâh ibn Buluggin (Zirid king of Granada), 2.524; 3.292
'Abbadid kingdom (Seville), 3.323	'Abd Allâh ibn Saba', 1.7; 4.357
'Abbas I ('the Great', shah of Iran), 1.4, 100, 197, 610, 672; 2.42, 189, 330, 620, 623–624; 3.86, 123–214, 372; 4.665	'Abd Allâh ibn Salâm, 1.216, 338
anti-Jewish measures by, 2.611	'Abd Allâh ibn Tâhir (Abbasid governor of Alexandria), 1.133
confiscation of Jewish books on magic by, 4.150	'Abd Allâh, Imâm Hanaifi Sayyid, 4.239
in legends on Serah bat Asher, 4.315	'Abd Allâh, Ismâ'îl, 1.242
persecution of Jews by, 2.584; 3.167	'Abd Allâh, Joseph, 2.317
'Abbas II (shah of Iran), 1.4, 610, 672; 2.156, 189, 346, 436, 620; 3.123, 156, 490	'Abd Allâh Pasha (governor of Sidon), 2.193; 3.618
accounts of reign of, 1.6; 2.584	'Abd Allâh, Ya'qûb Faraj, 3.111
persecution of Jews by, 2.585; 3.167, 245, 496; 4.362	'Abd al-'Aziz (Amirid ruler of Valencia, 1021–1061), 3.288; 4.589
al-'Abbâs (uncle of Muḥammad), 1.682	'Abd al-'Aziz ibn Mûsâ (Umayyad ruler of al-Andalus, 714–716), 4.325
Abbas, Ferhat, 1.153	'Abd al-'Aziz (Mawlay) (Moroccan sultan, 1894–1908), 2.202
Abbas, Ihsân, 2.498, 537	'Abd al-Bâsit, 2.201; 4.503
Abbas, Maḥmûd, 1.235	'Abd al-Ḥaqq ibn Abi Sa'id (Marinid sultan of Morocco, 1420–1465), 1.603; 2.201, 467; 3.467
'Abbas-nâma (The Book of 'Abbas, Muḥammad Tâhir Wahîd al-Qazvini), 1.5, 6; 3.158, 489	'Abd al-Ḥaqq al-Islâmi, Abû Muḥammad, 1.6
Abbasids	'Abd al-Ilâh (regent of Iraq), 2.23, 592
decline of, 2.606	'Abd al-Majid I (Ottoman Sultan) see Abdûlmecid I

Brill, Leiden 2010
Also available online: brillonline.nl

EJIW

Figure 2 The EJIW paper index

the editors had to stay away from the article since indexing and editing simultaneously can, of course, cause problems. Sometimes articles were edited after indexing, and I would get a warning as the changes would not result in automatic changes to the index. The editor could tell me if a term with a tag attached to it had been deleted or changed. Sometimes the editor would forget to note the locator number, which made retracing that particular index entry somewhat of a challenge, but they soon came to realize the importance of the locator number.

Once an article had been indexed, it could be marked as 'indexed' within the CMS. However, if we did that, we would 'lose' the article, and could no longer retrieve it. This was no good, of course, if we needed to retrace entries at the editing stage. Again, this was something the software engineers had devised without realizing how an indexer works. I solved this by not marking any of the articles as 'indexed'. This meant that I had to keep a separate Excel worksheet alongside my work. This worksheet had the complete list of all the articles

that were to appear in the encyclopedia and their unique numbers. I could mark on the sheet which articles were indexed, giving me some idea of how much more work there was still to come. Of course, this bit could have been made much easier to manage had the software engineers known more about the indexing process.

Turning to a team

The steady flow of articles unfortunately did not materialize. In practice, most of the articles were late. We had also lost a lot of time making the indexing module workable. The upshot was that I started indexing much later than originally planned. When the bulk of the articles started to materialize somewhat belatedly, I was looking at more work than one indexer could handle. This left us with the choice of either postponing the publication date or doing the indexing with more than one person. Obviously, the latter option was the more attractive one for the publisher.

I had worked with Jacqueline Pitchford and Pierke Bosschieter (both Dutch indexers) before, but always on separate occasions and as pairs, never as a threesome. I knew their way of indexing and was confident we could work well together. Brill agreed that the remainder of the indexing work (the bulk, I have to emphasize) could be shared with other indexers, provided I retained overall responsibility for the work and did the final editing.

Pierke and Jacqueline agreed to step in and index as many articles as they could alongside their other indexing work. When I got a new list of articles from the Brill editor, I would indicate which were to go to Pierke, which to Jacqueline and which to remain with me. When Pierke or Jacqueline logged into the CMS, they would see the list of articles ready for them to

index. Once they had done a bunch of articles, say about 25, the Brill desk editor would allocate the articles back to me, and Pierke and Jacqueline would send me the index entries attached to these articles. I would do a little pre-editing and then feed the entries into my master index. We could specialize among ourselves. Pierke mostly got the articles dealing with art, music and literature. Jacqueline did the economics articles and town descriptions, and I concentrated on the religious and historical articles. Jacqueline and Pierke would let me know how many hours they could devote to the EJIW indexing for the coming weeks, and I could figure out approximately how many articles they could do in those hours. So, we could vary the workload, and if either Pierke or Jacqueline had too many articles they could return some, or they could ask for more if they happened to have more time on their hands.

All the articles were in the CMS, of course, so we did not need to send them to one another and it was easy to keep an overview. Jacqueline, like me, works with Macrex

so would send me Macrex files. Pierke works with Sky. Sky allows you to make a Macrex backup file as output file. This worked fine. This is an index with many, many diacritics. Provided Pierke used the Macrex coding for them (you can make macros for that), there would be no problem with the diacritics.

Achieving seamlessness

By the time Jacqueline and Pierke joined me, I already had a basic framework for the index which I had to ask them to conform to. I would send Jacqueline a table of authorities file (TOA) and Pierke an RTF version. This worked surprisingly well. But double entries were too difficult to manage in such a large index with three different indexers, so these had to be changed into cross-references. The team indexing worked like a peer review on the go. We kept each other sharp on indexing decisions. Queries from Jacqueline and Pierke sometimes meant I had to rethink my earlier decisions. I could see at a glance what decisions they had taken because ‘tagged’ terms are automatically highlighted in green by the CMS. I found that Jacqueline (for the purposes of this index) tended to overindex at first, a common tendency when you start to do ‘tagged’ indexes. Pierke, by contrast, tended to under-index. After only a couple of articles we got things into balance, and we thereafter managed to index to about the same level of detail. With traditional indexing, comparing each other’s indexing styles would not be so easy.

A happy team

I have asked Jacqueline and Pierke what they thought of the cooperation. Pierke said she found it very instructive to see how the index framework I had made was structured. It was very rigid, she said. It was indeed more rigid than usual, but this was born out of the need to achieve consistency of approach between the three indexers. Working on this project has also opened Pierke’s eyes to the flexibility of Macrex. This is perhaps something for the software developers to ponder on, but we certainly have no complaints about the ease of communication between the two indexing programs. Both Pierke’s and Jacqueline’s final conclusion about the team indexing was: we should do this more often! And I can wholeheartedly subscribe to that conclusion.

The hard-copy index

The hard-copy version is a pretty straightforward although somewhat complicated index. This is the index that won us the Wheatley Medal, of course (see Figure 2). We kept a fairly rigid structure. This was necessary because of the teamworking. However, because many

articles were structured in similar ways it was also easy to carry out.

For example, the articles dealing with towns all basically dealt with the same subjects. There was, of course, always a description of the Jewish community. Within that community, sub-subjects such as ‘size’, ‘migration to Israel’ and ‘leadership’ nearly always came up, so I tried to use the same subheadings and sub-subheadings for each community.

Names were, of course, a major challenge. There were modern, medieval and pre-medieval names, and there were Muslim and Jewish names, each with their diacritics. The Muslim names could be further divided into Arabic, Turkish and Persian names, each with their own titles, honorifics and so on. The editing of the whole publication was well controlled, but it was sometimes impossible to stick to one transcription/spelling of a name, for example because of the differences in spelling in the Turkish and Arabic languages. This was explained in the introduction in Volume 1. My introductory note to the index was deleted by the editors (Stillman, 2010: xi), so users of the index, in Volume 5, need to go back to Volume 1 for explanations about transcriptions, treatment of modern/pre-modern names and similar issues. Of course, the explanation in the Introduction is more comprehensive than my introductory note, but I still feel some explanation of the treatment of names at the start of the index would not have been superfluous.

The online index

The online version of the index (which was prepared by Brill from our RTF file), was not published until a year after the hardcopy version. The original idea had been to publish the two indexes simultaneously. I queried Brill about this. Their response was that they simply did not have the labour power

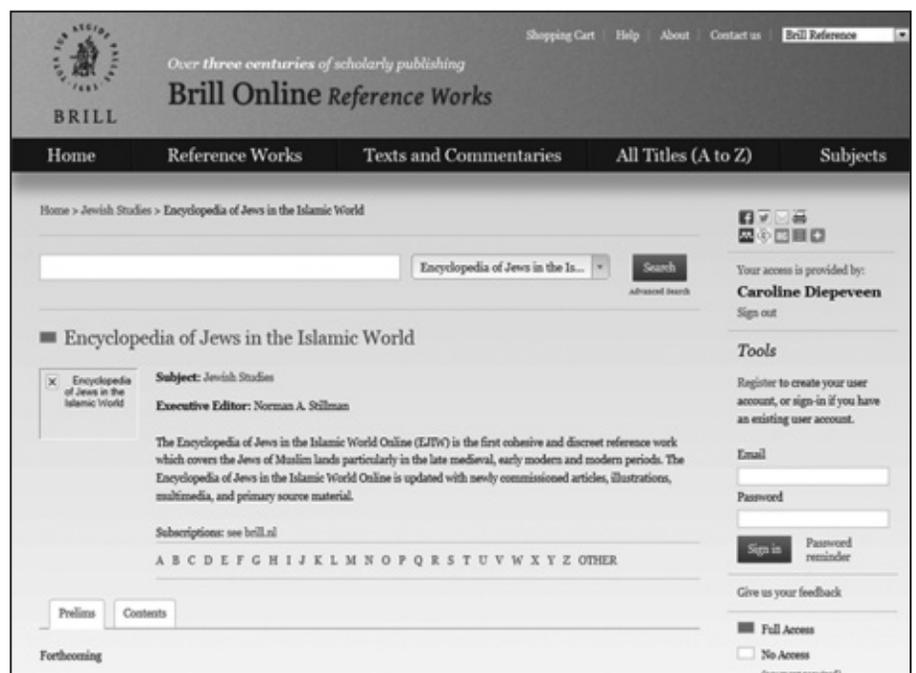


Figure 3 EJIW online startup screen

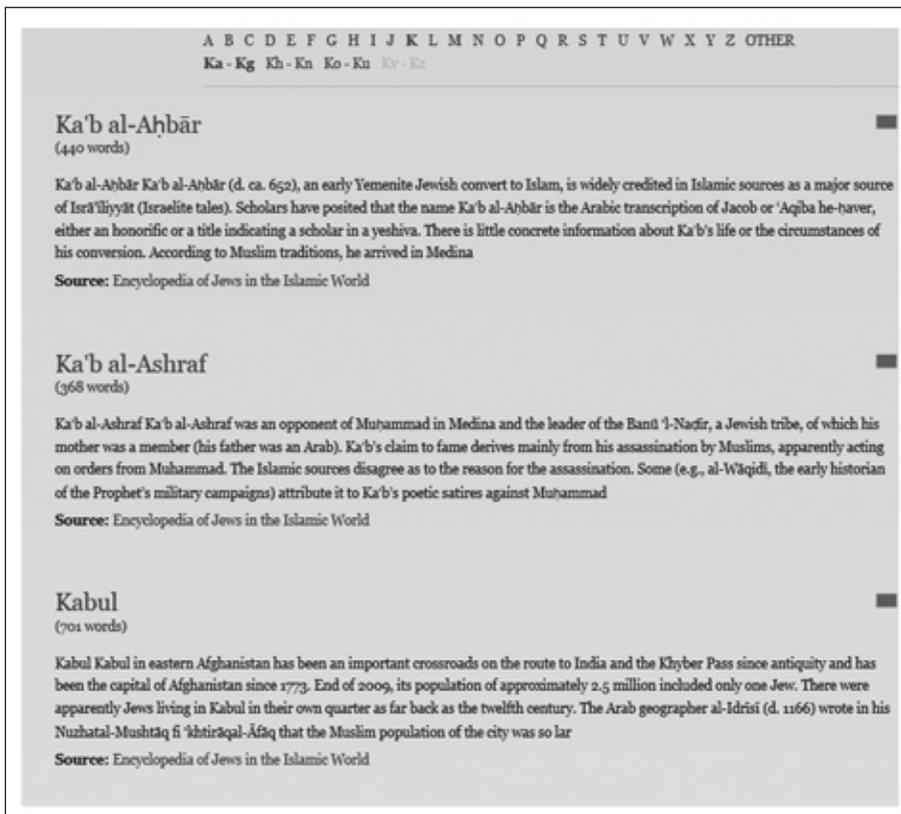


Figure 4 EJIW online screen after alphabet search

available to make everything ready in time, and therefore a decision was made to postpone online publication.

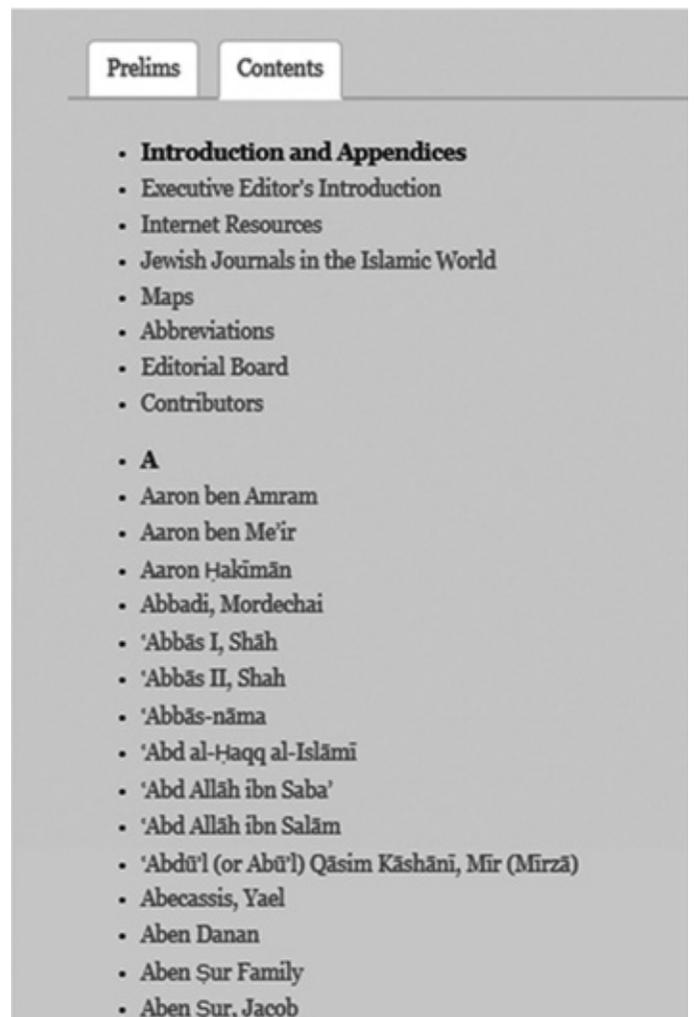
In the introduction to this article, I indicated that Brill allowed me three months free use of the online *EJIW* publication. This allowed me to explore what became of our index in the online version. Here I will describe my online experience with *EJIW*. When you log into the online publication of *EJIW*, the first screen you see is the one shown in Figure 3.

Unfortunately, the user is first presented with a full-text search tool. It has been demonstrated elsewhere how much inferior full-text search is to an index. For example, a usability study by Bloomberg BNA, a leading firm providing legal, regulatory and business information for professionals, found that text searchers had only a 23 per cent success rate, as opposed to an 86 per cent success rate for index users (<http://tinyurl.com/BNAindexstudy>).

Below the full text search box there is a clickable alphabet. You might expect this to lead to the index, but unfortunately it does not. In fact, it brings you to the alphabetical list of encyclopedia articles. Figure 4 shows the screen for 'k'.

I still had not found the index so I decided to click on the contents tab below the clickable alphabet (see Figure 3). This produced the screen shown in Figure 5. The first bit is largely similar to the table of contents of Volume 5, which also contains the index. The index, however, is missing from this list. Below the first bit, under the letter A, is

Figure 5 EJIW online Contents tab (note: the right-hand side of the screen is blank and not shown here)



again an alphabetical list of the encyclopedia articles. Still, no sight of the index. Wherever could it be?

The answer was that you had to scroll all the way down the list of encyclopedia articles until, at the very end, you got to the sequence shown in Figure 6. The clickable entries below the letter A and B allow you to access bits of the index. Clicking on the first entry below A finally brings you to the online index (Figure 7). We might wonder how many users will manage to get there.

Figure 7 might look a bit bewildering at first. I have chosen a screenshot of the entries for the Persian Shahs Abbas I and Abbas II. What is very interesting is that Brill has opted to use the titles of the encyclopedia articles as locators. The locators are clickable, of course, and directly lead you to the paragraph in the text that deals with the index term. The first three locators after Abbas I are conversion, Isfahan and Isfahan. This means that there are bits of text on Abbas I in the article on conversion and that Abbas I makes two significant



Figure 6 EJIW online index access screen

appearances in the article on Isfahan. If we look a bit further down the list of locators, we see that Abbas I also has an entire article devoted to himself. So in fact, the locators offer quite a bit of extra information compared with mere page numbers. They are lengthier than page numbers, of course. Perhaps an alphabetical ordering of the locators would allow for easier grasp of them. But on the whole I am happy with the way the locators are presented.

When I finally set eyes on the online version of our index it became clear to me how appropriate the following quote from Jan Wright is:

We can let go of the page, and think of locators as just that: locators. We do not care if they are anchor points in HTML, unique ID strings in an InDesign file, or time codes in a video. The human analysis of the aboutness at that location is our unique contribution. The technology for presenting that location and accessing that location can change, but we are still the best resource for letting readers know whether a location is worth a visit, and where else they can look.

(Wright, 2012: 46)

I would like to add that some locators are more informative than others, though.

Unfortunately, navigation around the online index, when you get to it, is cumbersome. To do a new search of the index, you have to return to the 'contents' tab and scroll all the way down again, or do a couple of back screens to retrace your steps. In conclusion, the online version of the index was done brilliantly, but I do wonder whether the poor user will ever be able to find it, or having found it, have the patience to use it.

Brill feedback and update

I have informed Brill about my experiences with the online index and invited them to comment on it. To my main point about the difficulty of finding the online index, Brill's staff responded that they are aware of the problem. They explained that they have changed to a new platform, and are currently working on a better solution for the index which they want to implement for all their reference works. So with luck this will be solved in the near future.

Since I wrote this article, as is so often the case, things have moved on. A taster edition of the EJIW online index is now available online at: <http://referenceworks.brillonline.com/browse/encyclopedia-of-jews-in-the-islamic-world>, to entice you to purchase the online subscription. Only the main

headings and the first subheading are given, but this does give an idea of how the locators work. The index is now placed under the 'prelims' tab, making it accessible from the home page. This may not be the most logical place for the index, but it is certainly a big improvement from where it was originally.

Brill staff said they have received equally positive feedback from other users about the locators in the online index. This is excellent news in that it means that there are users able to find the online index!

They said they were not aware that navigation of the online index was cumbersome. I hope this is something that will be ameliorated in the future.

New technologies: threat or opportunity

As yet this is my only indexing project using the Brill CMS, which Brill has said it will reserve for very large multi-volume projects, the next of which is not scheduled until the end of 2013. However, it is clear that many other publishers are currently developing new technologies to improve their production process and to allow them to publish hard copies and online versions simultaneously.

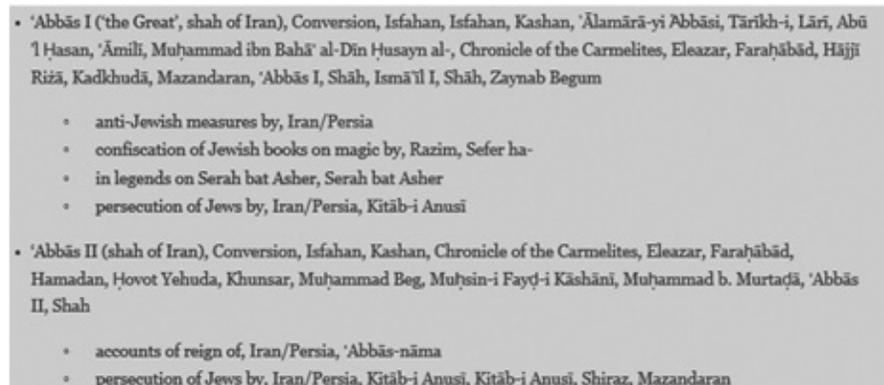


Figure 7 The EJIW online index

Indexers are unfortunately rarely involved when the software is being developed, and the software developers seldom know much about indexing. The SI Publishing Technology Group and the ASI Digital Trends Task Force do important work for the indexing profession at a general level. It is up to individual indexers to liaise as much as possible with publishers developing new technologies. To do this they have to be aware of new technological developments, to have a reasonable understanding of what these actually involve, and to be flexible. As Jan Wright wrote in the March 2012 *Indexer*:

We need to understand the level of technical expertise our clients have. They are having to learn new techniques at the same time we are, and are often not sure how to go forward. Discussing the hows and whys, and the eventual outputs they want to create, becomes crucial for indexers now. We all need to learn more about technologies so that we can discuss project processes, know the implications of decisions, and understand the final displays of our indexes.

(Wright, 2012: 1)

A threat? At first, the indexing module developed in the Brill CMS seemed hopeless. The software developers clearly had no idea what indexing was about. Persuading the Brill IT people and the software firm of the changes needed in the CMS was not easy, and I achieved only limited success. For example, it took me a very long time to convince them that the relationship between the text anchor/tag and locator is not a one-to-one relationship, but a one-to-many relationship. One anchor/tag may relate to several index terms.

An opportunity? The team option was not something that we had envisaged from the start. It was born out of necessity, but the CMS turned out to be very well suited for team work. We tend to think of indexing as a lonely profession, but it does not have to be so. New technologies will make it easier for us to work in teams, which means we can handle large indexing projects in shorter time spans. Surely there is a big advantage here for publishers. The software allows for

easy management of the workflows between team members, and the CMS 'tagging' technology makes every indexing decision clearly visible, thus making for ease in editing and for control of the density of indexing between team members.

The future? The main thing I learned from the *EJIW* experience is to keep an open mind about every new development. Even word-spotting software may have useful features, and a lousy indexing module in a CMS can still allow someone to produce an award-winning index. Walking away from new developments and saying that you only want to index in the traditional way is not an option. Dismissing a new technology without knowing what it is about is a luxury we certainly cannot afford.

Acknowledgements

I would like to thank Pierke Bosschieter and Jacqueline Pitchford for commenting on the first drafts of this article, and Joed Elich, publishing manager at Brill, for arranging three months free access to the *EJIW* online publication.

References

- Stillman, N.A. (ed.) (2010) *Encyclopedia of Jews in the Islamic World*. Leiden, the Netherlands: Brill. Available at: <http://referenceworks.brillonline.com/browse/encyclopedia-of-jews-in-the-islamic-world> (subscription only)
- Wright, J. (2012) 'Editorial: digital trends.' *The Indexer* 30(1), 1, 46.

Websites

- ASI Digital Trends Task Force: <http://www.asindexing.org/i4a/pages/index.cfm?pageid=3647>
- BNA Bloomberg study on using online indexes: <http://tinyurl.com/BNAindexstudy>
- SI Publishing Technology Group: <http://www.ptg-indexers.org.uk> (member login required).

Caroline Diepeveen is a freelance indexer, based in Middelburg (the Netherlands). She has been indexing for well over ten years, first when living in Scotland and from 2000 onwards in the Netherlands. She is one of the founding members of NIN (the Netherlands Indexing Network) and co-editor of the NIN newsletter (NINieuws). In 2011 she won the Wheatley Medal, together with fellow NIN members Pierke Bosschieter and Jacqueline Belder, for their index to the Encyclopedia of Jews in the Islamic World published by Brill.
Email: cdiep@zeelandnet.nl



Caroline Diepeveen speaks at the SI conference, July 2012