

# APE 2008 Academic Publishing in Europe, Quality and Publishing

*January 22–23, 2008, Berlin*

Ronald Schild

*MVB, German Publishers and Booksellers Association, Frankfurt am Main, Germany*

Those of you who were here for last year's conference might remember that I was here too and chaired one of the sessions. Last year, I mentioned that we had just started a new project called "Volltextsuche-Online" which stands for "full-text search online". Today, I am here to present this project again. I am quite proud to be able to tell you that it is now up and running, following the launch of the website at the last Frankfurt Book Fair. Furthermore, we have changed the name of the project, because full-text search, or searches in the full text, is just one small functionality in the whole set of features we want to offer. The project's new name is libreaka!. We also wanted to make the point that we are not a "YACS" – "Yet Another Californian Start-up". On the contrary, MVB is based at the heart of the German publishing industry and is wholly owned by the Börsenverein, the German Publishers and Booksellers Association.

Now, what do we want to achieve with libreaka!?! To put it very simply: we want to put every German-language book in print online. We want to be the leading internet platform for the German book trade or, figuratively speaking, we want to be the Swiss army knife of *e*-content. First of all, we have a marketing objective. We want to increase the reach of publishers and booksellers. We want to open up new business models and sales channels for booksellers and publishers and want to protect intellectual property. We want to give publishers control over how much and what they publish on the internet.

Now, what do I mean when I say: "We want to increase the reach of the publisher or bookseller"? First of all, by putting every book in print online with libreaka, we want to give end-users access via the internet to culture, to knowledge, to education. At the same time, we offer every bookseller a service similar to the "Search Inside" offered by a large online retailer. Furthermore, we want to integrate the content on libreaka! in the main search engines Google, Yahoo and Microsoft. If you look for a book that is on libreaka! today, you will probably find it on Google as well. But what we want is to drive initiatives like "Google Book Search" or "Microsoft Life Search" by adding our content and therefore making it more easily available. At the same time, libreaka! gives booksellers and publishers the opportunity to offer their customers a search inside the book-function and a chance to read examples of the books on libreaka!.

The integration in the search engines basically works in two ways. One way is very easy: the search engine indexes our html-pages. This is what they are already doing today. The second model is more complicated: the data we get from publishers is in pdf-format, though it is not a perfect format to use. We want to move to xml-format, but this data is rather difficult to get. That is why we are stuck with pdf for the moment. What we do is to extract the text from the pdf and build our own text index – this is purely text, no images and no meta-information about the book. Then we give access to that text

index to the search engines that enrich their own search index with it. They generate search results and display a search list where you see basic information on the books, such as the title. But in order to view and to access the book, you have to go back to libreaka!, you read and view the book on our platform. This gives the publisher the biggest possible control over the content. But we still have to suit different publisher's needs. There might be publishers who prefer open access models and want to give the content away freely. Other publishers favour more restrictive models. We are not in the position of telling the publishers what model to use. But we want to be open to any business model that a publisher might want to use. At the moment, we are negotiating with the big search engines and we have arrived at a certain understanding that they will at some time integrate our data.

As I have already mentioned, to offer full-text search is not our main objective. What is more important is the relevance of the search results. Already now, with over 42,000 books online, we get at least several hundred search results for more general search terms. That is obviously too much. We know from usability studies that people who use a search engine usually stop looking after the first three listings or at least after the first or second page of the result list. Search results on page 10 are not really important. With this in mind, the real challenge is to show the most relevant search results on the first page of our result list. We do this by combining the search of the full texts of the books on libreaka! with their metadata. We have access to the metadata of the German-language books in print that are available via the "Verzeichnis Lieferbarer Bücher (VLB)", which is published by MVB. It lists more than 1.2 million books. This gives us the opportunity to add to the search results the author's name, the publisher, the publishing year and more information on the title. Furthermore, we have a research project running at the moment with the Ludwig-Maximilians-Universität (LMU), a Munich university, to use linguistic methods to increase the relevance of search results.

What I have wanted to point out so far: libreaka! means more marketing for publishers and booksellers. They can increase the reach of their target audience and give more information to their customers. But, of course, they want to do business as well. What we are planning to introduce this year are various *e-commerce* models. The easiest model is direct consumer sale of *e-books*. You can sell *e-books* that are downloadable to PCs or reading devices. This model is the most discussed at the moment. Personally, I think that online access to information will become more important and might be one main driving force when talking of *e-content*. And that applies as much to individuals as it does to institutions – libraries, research centres and so on. We can think about different subscription models and pay per use models. Here, again, the publisher decides, what business model he wants to implement for his *e-content*.

And, of course, we are open to booksellers too. Our objective is very clearly to offer every bookseller, regardless of size or online affinity, the ability to sell *e-content*. We don't want a monopoly or very few dominant players to control the *e-content-market*. That will be neither good for publishers, nor for cultural diversity nor for access to information.

And, last point, what we want is to protect intellectual property. Our mission is to provide a distribution platform; we are not a proprietary system like many other book search engines. libreaka! is a platform where the publisher can list his products, put up his content and we give it away in accordance with the publisher's needs and wishes. They keep full control over their content – on a per page basis: the publisher decides how many pages of each of his books in libreaka! can be viewed: for one book he may grant access to the general public only for pages 1–10. For the next book, everyone is allowed to see everything. And the books can be distributed across all sales channels that cooperate with the publisher. Of course, we offer DRM systems or DRM functionalities, but I am not going into detail about that here.

I would now like to show very briefly how libreaka! works. Search for any given term and you get a list with your search results. For every book in this list, you see the cover, the title and a short description.

When you click on one of the books you get access to it. The virtual book is displayed the same way as the printed book. If one of the pages is blank in the printed book, it's displayed the same way in libreka!. More importantly, in a second step, you can look for your search term "inside the book". libreka! then gives you the pages of the book, where your search term appears. The pages the publishers wants to exclude from being seen by the general public are not displayed in libreka!. I can do many more things, like browse various book categories, such as fiction, non-fiction or children's books, but I am not going to show you that now, it would take the whole afternoon.

As I have already said, we are on the way to implementing *e-commerce* functionalities, they are going to follow pretty soon – as well as a privileged access for booksellers. The objective is to rationalise the interaction between publisher and bookseller. At the moment, publishers spend lots of money on transporting sample books to the booksellers, like reading samples. With libreka! they do not need this anymore.

At the moment more than 850 German-language publishers are partners of libreka! We ran a special promotion, advertising free scans for books where no data is available. This brought us 60,000 books to scan – which we are doing now. Furthermore, we were promised 30,000 books in digital format. At the end of this presentation the last slide shows some of the participating publishers. Some of the biggest publishers in Germany are on board with libreka!.

Thank you very much for listening.