Arch. Math. 77 (2001) 116–128
0003-889X/01/010116-13 $ 4.10/0
© Birkhäuser Verlag, Basel, 2001

Archiv der Mathematik

Report on the retrodigitization project "Archiv der Mathematik"

By

GERHARD O. MICHLER

To Professor Erich Lamprecht on the occasion of his 75th birthday

1. Introduction. Nowadays many mathematical journals are published electronically, in general in portable document format (PDF). For example, the servers of the American Mathematical Society (AMS) contain at present more than 4700 articles published electronically from 1996 to date in the journals of the AMS that are accessible and searchable online to subscribers via the internet address: www.ams.org/journals. The PDF-format offers many advantages over the traditional paper format of a printed article. Authorized users can also link from journal articles references to the reviews in the Mathematical Reviews of the AMS, and in due course from there to the original quoted articles, if available in PDF. The online access also allows to make printouts of wanted articles at the user's printer and to view abstracts of recently posted articles to be published shortly.

In view of the vaste literature in mathematics the search functions provided by the software of a distributed digital library for specific articles and/or specific results or concepts explained in such an article are very useful for researchers and graduate students. They also would like to have such a comfortable access to the back issues of the mathematical journals published only in paper format. The proofs of many important old theorems have never been incorporated into standard textbooks because they are still too complicated to be derived from generally known theoretical results. Therefore modern research articles have to quote arguments, methods, partial and final results of old papers.

Thus it is desirable to retrodigitize the back issues of a mathematical journal and combine them with the electronically published recent issues in such a way that all the volumes of the resulting digital journal are searchable and linked to "Mathematical Reviews" or "Zentralblatt Mathematik". Through these review journals the retrodigitized volumes could then be linked to all other mathematical journals belonging to a distributed digital mathematical library. In particular, an authorized user of such a library would be able to read on screen an original article and a cited article or a review at the same time.

This retrodigitization task is very demanding and requires a deep understanding of mathematics, computer science and the support of the publisher. It is the purpose of this survey article to describe a solution found by the author's study group in the pilot project "Retrodigitization of the journal Archiv der Mathematik" financially supported by the Deutsche Forschungsgemeinschaft from 1 April, 1997 until 30 September, 2000.

Mathematics Subject Classification (2000): Primary 94A08; Secondary 68U15, 68U35.

Using the IBM digital library database [4] and the MILESS software [7] developed by the Computer Center of Essen University we have been able to construct a prototype of a searchable and retrievable digital library "Archiv der Mathematik". It contains

1) the retrodigitized volumes **60** (1993) – **67** (1996) in MVD format,
2) the electronically published volumes **68** (1997) – **73** (1999) in PDF format.

The cited journal articles of the references of all 13 volumes are recognized automatically as well as the layout of the first page of each article. Thus it is possible to produce the bibliographic data including the ISSN of all journal articles of the 13 volumes and their references in XML format. In particular, all these articles and the cited journal articles of their references can be linked automatically to MathSciNet or Zentralblatt MATH.

The methods developed in this pilot project can be adjusted for the retrodigitization of other mathematical journals and older back issues of "Archiv der Mathematik".

2. Retrodigitizing the ordinary text. The best solution of the retrodigitization task would be the *automatic production* of the Latex file of a printed mathematical article, even if the original manuscript was not submitted for publication in Tex or Latex format. Since no computer can understand the mathematical contents of a printed article, only probabilistic and heuristic methods can be applied to recover the mathematical contents of a printed page of the given article from its digital image obtained from a scanner.

Each page of a mathematical article consists of ordinary text, geometric pictures, diagrams, tables and mathematical formulas. It is scanned with 600 dots per inch (dpi). Therefore a very good picture of each scanned page can be viewed on screen. Furthermore, the TIFF file of a scanned page allows complete printouts of it at the user's printer. But it does not provide tools for searching for mathematical expressions or formulas within a page. As one does not search for geometric pictures, diagrams or tables, and since there are excellent pictures of them in the TIFF file, they are neglected in the further steps of the retrodigitization process.

Since the complete Latex file of the text of a printed page occupies much more space than the image of the scanned page it is clear that there is only a 3-dimensional solution of the retrodigitization problem. For its realization T. Phelps' (UC California, Berkeley) multivalent document system [10] is used. It is a general paradigm that regards complex documents as multivalent documents comprising multiple layers of distinct but intimately related content, see [11].

In our application it consists of two multiple active semantic layers of the contents of a scanned page. In the first layer its GIF file is stored which is automatically produced from its TIFF file. The second layer contains the Xdoc file of that page. It consists of the ASCII file of its ordinary text and coordinates of each word shown on the first layer. Mathematically, the two layers are linked by the same coordinates of a given word or formula. T. Phelps' software [10] provides the tools to view the results of the computer search in the ASCII file on screen in the TIFF file by highlightening the mathematical expression or ordinary word in a special colour like yellow or by means of a red frame.

For the production of the Xdoc file from the TIFF file a commercial optical character recognition software (OCR) like FineReader [3] is used. It is able to recognize the ordinary words in the TIFF file of a mathematical document and give the coordinates of each recognized word or formula. Using these data a special software has been written by the author's colloborators to obtain the Xdoc format of a scanned page in such a way that it can be read by T. Phelps' system [10].

In order to improve the recognition rate of FineReader several other programs have been written to include a database of titles and abbreviations of mathematical journals used by Mathematical Reviews of the AMS, a mathematics dictionary containing classical abbreviations (like *sup* or *inf*) and special expressions (*normal subgroup*) in English. There is also a similar dictionary in German.

As an example we now show a copy of the TIFF file of the upper 14 lines of p. 183 of volume 65 (1995) of Archiv der Mathematik.

Let f_8 be the interpretation of $\bar{f}_8$ as a function $\mathbb{R}^2 \to \mathbb{R}^2$. Then we have

$$f_8'': \mathbb{R}^2 \to \mathbb{R}: (x, y) \mapsto \tfrac{1}{4}(b^6 + 2(a^2+4)b^4 + (a^4 + 16a^3 - 40a^2 + 16a + 16)b^2 + 16a(a-1)^3(a-4))(x^2+y^2)^{2a-4}.$$

So, $f_8'' \leqq 0$ corresponds to the last condition above.

These last examples also show that if a family of partition functions depends on some parameters, then $f'' \leqq 0$ can give constraints on these parameters, which often coincide with the actual necessary and sufficient conditions for these functions to be planar. This suggests to use the inequality $f'' \leqq 0$ in attempts to construct new planar functions from existing ones by introducing (small) deformations of the original functions and checking whether the inequality is still fulfilled for the modified functions. Successful attempts of this kind of construction have been made by the author by replacing the piecewise constant functions u and v in Examples 5–8 certain twice continuously differentiable functions $\mathbb{R} \to \mathbb{R}$.

The Xdoc format of the ordinary text of the first 8 lines of this section of p. 183 is automatically produced and printed out as follows.

```
[a;"XDOC.10.0";E;"FrEngine40-CB1"]
[d;"0000652183.xdc"]
[p;1;P;83;S;0;1666;0;0;3161;4665]
[t;1;1;0;0;A;"";"";"";0;0;0;0;1]
[f;0;"<DEFAULT>";R;q;10;V;66;16;0;33;100]
[f;1;"Courier";R;q;12;F;100;25;10;50;100]
[f;2;"F";R;q;49;V;52;13;0;26;100]
[f;3;"F";R;q;49;V;60;15;0;30;100]
[s;1;87;0;0069;p;2]Vol.[h;204;24]65,[h;312;33]1995[h;476;698]
Planar[h;1372;24]partition[h;1658;27]functions[h;1962;980]183
[y;3052;0;71;0;H]
[s;1;88;0;0265;p;3]Let[h;198;24]/8[h;287;32]be[h;394;32]the[h;
526;29]interpretation[h;1020;31][c;4]$\sf{o} f f_{ 8 } x a$[h;
1369;0][c;3][h;1369;27]function[h;1673;30]R2[h;1802;27][c;5]
$\rightarrow$[h;1892;0]
[c;3][h;1892;20]R2.[h;2038;32]Then[h;2242;30]we[h;2361;30]
```

```
have[y;2546;0;266;0;H]
[s;1;414;0;394;p;6][c;6]$f _{8 }^{' ' } :
 \mathbb{R} _{}^{2 } { \rightarrow _{  }^{  } }\mathbb{R} :
\left( x , y \right) \mapsto \frac{\displaystyle {1 }}{
\displaystyle{4 }}\left( b _{}^{6 } + 2 \left( a _{}^{2 } + 4
\right) b _{}^{4 } + \left( a _{}^{4 } + 1 6 a _{}^{3 } -
\infty a _{}^{2 } + 1 6 a + 1 6 \right) b _{}^{2 }i \right.
$[y;3045;0;394;0;H]
[s;1;1303;0;0527;p;3][c;7]$\left. + 1 6 a
\left ( a - \tt{1}\right )^{ 3 } \left ( a - 4 \right )
\right ) \left ( x^{ 2 } + y^{ 2 } \right )^{ 2 a - 4 }$
[h;2553;0][c;3][h;2553;11].
[y;2575;0;536;0;H]
[s;1;87;0;0673;p;3]So,[h;186;24][c;8]$f^{ ' ' }_{ 8 }\leqq 0$
[h;441;0][c;3][h;441;27]corresponds[h;876;30]to[h;972;32]the
[h;1105;29]last
[h;1250;28]condition[h;1598;33]above.[y;1844;0;685;0;H]
[s;1;163;0;0771;p;3]These[h;356;26]last[h;497;24]
examples[h;828;26]also[h;985;27]show[h;1184;25]that[h;1342;25]
if[h;1416;17]a[h;1470;24]family[h;1702;26]of[h;1797;17]
partition[h;2109;24]functions[h;2441;26]depends[h;2740;28]
on[h;2852;26]some[y;3050;0;772;0;H]
[s;1;83;0;0870;p;3]parameters,[h;473;28]then[h;646;22][c;9]$
f^{ ' ' } \leqq 0$[h;899;0][c;3][h;899;25]can[h;1039;28]give
[h;1200;26]constraints[h;1596;30]on[h;1708;28]these[h;1903;28]
parameters,[h;2323;28]which[h;2549;28]often[h;2744;27]
coincide[y;3048;0;885;0;H][s;1;83;0;0966;p;3]with[h;228;26]
the[h;355;26]actual[h;583;26]necessary[h;925;26]and[h;1075;26]
sufficient[h;1397;26]conditions[h;1775;24]for[h;1894;26]these
[h;2088;25]functions[h;2420;27]to[h;2513;27]be[h;2615;26]
planar.[h;2874;28]This[y;3048;0;967;0;H]
```

Since the Xdoc format uses very much space the remaining lines of the above text are not restated here. Furthermore, the Xdoc format contains information about recognition of the special fonts used by the printers of Archiv der Mathematik. For the above example this information is given by

```
[f;4;"TeX";R;s;10;V;55;22;0;5;100]
[f;5;"TeX";R;s;10;V;27;0;0;5;100]
[f;6;"TeX";R;s;10;V;48;48;0;5;100]
[f;7;"TeX";R;s;10;V;65;22;0;5;100]
[f;8;"TeX";R;s;10;V;60;16;0;5;100]
[f;9;"TeX";R;s;10;V;62;13;0;5;100]
[f;10;"TeX";R;s;10;V;61;14;0;5;100]
[f;11;"TeX";R;s;10;V;36;2;0;5;100]
[f;12;"TeX";R;s;10;V;35;3;0;5;100]
[f;13;"TeX";R;s;10;V;57;0;0;5;100]
```

3. Recognizing the mathematical formulas. All parts of the scanned page which are not recognized as ordinary text are stored as empty space in the Xdoc file at this stage of the retrodigitization process.

In order to recognize the mathematical formulas another program has been written by the author's collaborators which separates the formula text from the TIFF file and preserves the coordinates of the areas of the formulas. Thus from each scanned page two files are produced at the same time, the Xdoc file explained in the previous section and a new file, the Segmentation file. This Segmentation file and the TIFF file are then taken as the input of Professor Okamoto's (Shinshu University, Nagano) special OCR program EXP [8] which in many cases is able to recognize the mathematical formula in a given area and to produce a Tex output of it together with the original coordinates of its corresponding image in the TIFF file.

In order to improve the recognition rate of Okamoto's program EXP the author's collaborators have extended its training system by providing copies of the special mathematical fonts used by the printers of Archiv der Mathematik. Furthermore, they have written additional programs to improve the parsing and the Tex production of the recognized formulas.

As an example of a result achieved by Okamoto's program EXP we now give its Tex output of the mathematical formula originally printed on lines 2 and 3 of page 183 of volume 65 shown above. It is also contained in the Xdoc file restated in Sect. 2.

```
$f _{8 }^{' ' } : \mathbb{R} _{}^{2 } { \rightarrow _{  }^{ }
} \mathbb{R} : \left( x , y \right) \mapsto \frac{
\displaystyle {1 }} {\displaystyle {4 }}\left( b _{}^{6 } + 2
\left( a _{}^{2 } + 4 \right) b _{}^{4 } + \left( a _{}^{4 }
+ 1 6 a _{}^{3 } - \infty  a _{}^{2 } + 1 6 a + 1 6 \right)
b _{}^{2 }i \right. $ $\left. + 16 a \left ( a - \tt{1}
\right )^{ 3 } \left ( a - 4 \right ) \right ) \left ( x^{ 2 }
+ y^{ 2 } \right )^{ 2 a - 4 }$
```

The printout of this Tex file looks as follows.

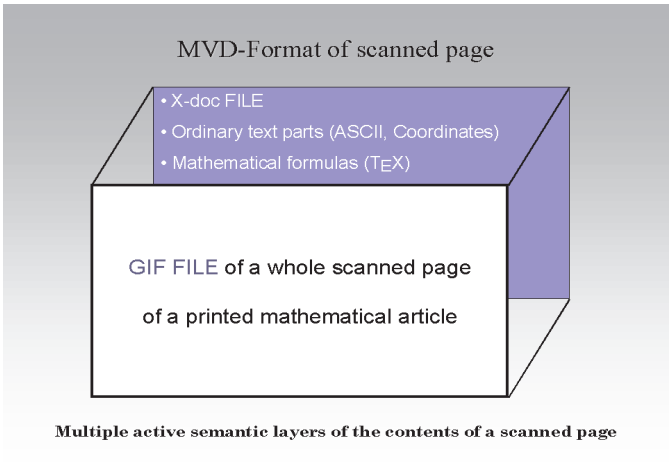
$$f_8'' : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto \frac{1}{4}\left(b^6 + 2\left(a^2 + 4\right)b^4 + \left(a^4 + 16a^3 - \infty a^2 + 16a + 16\right)b^2 + 16a\left(a - 1\right)^3\left(a - 4\right)\right)\left(x^2 + y^2\right)^{2a-4}$$

Looking at this printout one observes that the coefficient 40 of the term $-40a^2$ has been recognized as ∞ by EXP. This is caused by the touching of the digits 4 and 0 in the original printed page. So far all OCR systems cannot deal with touching characters efficiently without hand corrections.

In order to combine the recognized ordinary text contained in the Xdoc file produced so far with the recognized mathematical formulas another program was written by the author's collaborator Dr. Waki which places the Tex format of a formula at the right place in the above mentioned empty spaces of the Xdoc file of the scanned page. For that achievement the coordinates have been essential. Since the Tex code of a formula occupies much more space than its image on the TIFF file this text is included in an appropriate small scale at its correct space. Thus it is still possible to view it on screen, to export it for printouts or incorporate it into new mathematical manuscripts. Furthermore, it is now possible to search also for recognized mathematical formulas by using their LaTex code as input.

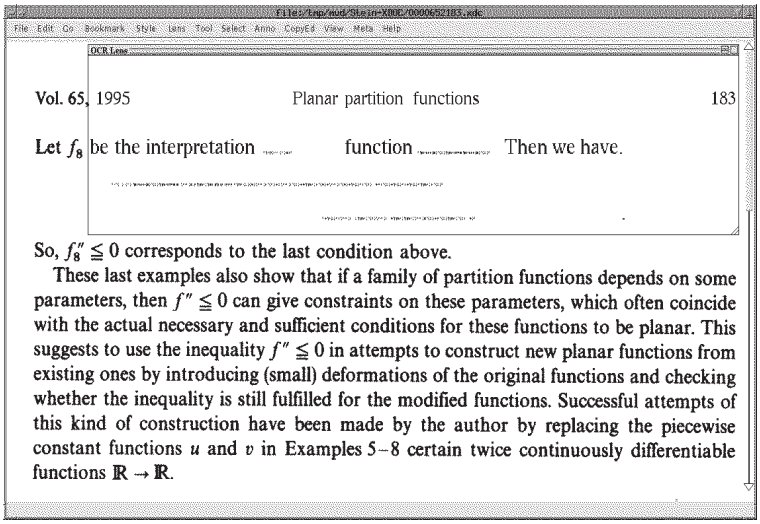
Thus the multivalent document format (MVD) of a scanned page consists of two layers, where the first layer contains the GIF file of the scanned page, and the second layer contains

the corrresponding Xdoc-file with its ordinary text and the recognized formulas together with their correct coordinates.



Multiple active semantic layers of the contents of a scanned page

For our given example (vol. 65, p. 183, lines 1-4) the GIF file looks the same as the printout of the TIFF file given before. Since already the Xdoc file containing only the ordinary text is almost unreadable for a mathematician we now state a printout of recognized ordinary text with the squeezed in Tex commands for printing the recognized mathematical formula. For a reader these Tex parts of the text are unreadable. However, as mentioned before, the Tex code of the recognized mathematical formula is stored in the Xdoc file, and the complete mathematical formula can be reproduced automatically.



Vol. 65, 1995 Planar partition functions 183

Let f_8 be the interpretation function Then we have.

So, $f_8'' \leqq 0$ corresponds to the last condition above.

These last examples also show that if a family of partition functions depends on some parameters, then $f'' \leqq 0$ can give constraints on these parameters, which often coincide with the actual necessary and sufficient conditions for these functions to be planar. This suggests to use the inequality $f'' \leqq 0$ in attempts to construct new planar functions from existing ones by introducing (small) deformations of the original functions and checking whether the inequality is still fulfilled for the modified functions. Successful attempts of this kind of construction have been made by the author by replacing the piecewise constant functions u and v in Examples 5–8 certain twice continuously differentiable functions $\mathbb{R} \to \mathbb{R}$.

4. Recognizing the layout of the first page. Since the number of printed old journal articles is still much larger than the number of electronically published recent articles, any extensive retrodigitization project cannot afford to produce the bibliographic data of the articles of a traditional journal by hand typing. Therefore the author's collaborators Dr. G. Hennecke and Dr. R. Staszewski have written a program which automatically recognizes the layout of the first page of each article which one wants to retrodigitize. It reads any scanned page. By analyzing the first recognized letters of it, it is able to decide whether this is the first page of the article.

As an example we include here the postscript copy of the TIFF file of the first 8 lines of the first page of the following article.

Arch. Math., Vol. 65, 176–184 (1995) 0003-889X/95/6502-0176 $ 3.30
© 1995 Birkhäuser Verlag, Basel

A sufficient condition for planar partition functions

By

BURKARD POLSTER *)

1. Introduction. A twice differentiable function $\mathbb{R} \to \mathbb{R}$ that has a positive (negative) second derivative is convex (concave). If it satisfies one more condition, it is a planar partition function.

The output of the program consists of the following bibliographic data of this article:

JOURNAL: Archiv der Mathematik
ISSN: 0003-889X/95/6502-0176
COPYRIGHT: ©1995 Birkhäuser Verlag, Basel
VOLUME: 65
FIRST PAGE: 176
LAST PAGE: 184
YEAR: 1995
AUTHOR: Burkard Polster
TITLE: A sufficient condition for planar partition functions
LANGUAGE: English

The bibliographic data are produced in XML format as required by the librarians and the publisher. They are essential for the integration of the MVD documents of all the retrodigitized pages of a journal article into a digital library like the IBM DB2 Digital Library product [4]. This will be explained in a later section. Because of the limited space we only restate the XML format providing the information on the title of Polster's article here.

```
<titles>
<title lang="eng">
A sufficient condition for planar partition functions
</title>
</titles>
<types>
<category
  id="2"
  label="Artikel"
  classification="TYPE"
>
```

In the meantime "Archiv der Mathematik" has changed the layout of the first page of an article. Therefore the recognition program has been extended in such a way that it is very easy to adjust it for the specific task.

5. Recognizing the references of an article. The author's collaborators Dr. G. Hennecke, Dr. H. Gollan and Dr. R. Staszewski have written another special program for the recognition of the references. It reads any scanned page of the given article and decides automatically, whether this page contains the beginning or the remaining part of the references of the article.

From the TIFF file the program recognizes the name's of the author(s) the initials of their first names and the ISSN numbers of the originally abbreviated journals names, volumes, years and page numbers. Thus it is possible to retrieve the cited article from the digital library, if it is in PDF format or MVD format. As an example we now show a copy of the references of Polster's article "A sufficient condition for planar partition functions", vol. **65** (1995), p. 183.

References

[1] D. Betten, Komplexe Schiefparabel-Ebenen. Abh. Math. Sem. Univ. Hamburg **48**, 76–88 (1979).

[2] H. Groh, Point homogeneous flat affine planes. J. Geom. **8**, 145–162 (1976).

[3] D. Betten and N. Knarr, Rotationsflächen-Ebenen. Abh. Math. Sem. Univ. Hamburg **57**, 227–234 (1988).

[4] D. Betten and B. Polster, Topological ovals in 4-dimensional translation planes. Arch. Math. **63**, 284–288 (1994).

[5] D. Braess, Nonlinear approximation theory. Berlin-Heidelberg-New York 1986.

[6] A. W. Goodman, Univalent functions, vol. I. Tampa 1983.

[7] E. Kirjackis, On functions with non-zero *n*-th divided differences (Russian). Litovsk. Math. Sb. **1**, 109–115 (1961).

The output of the recognition program looks as follows

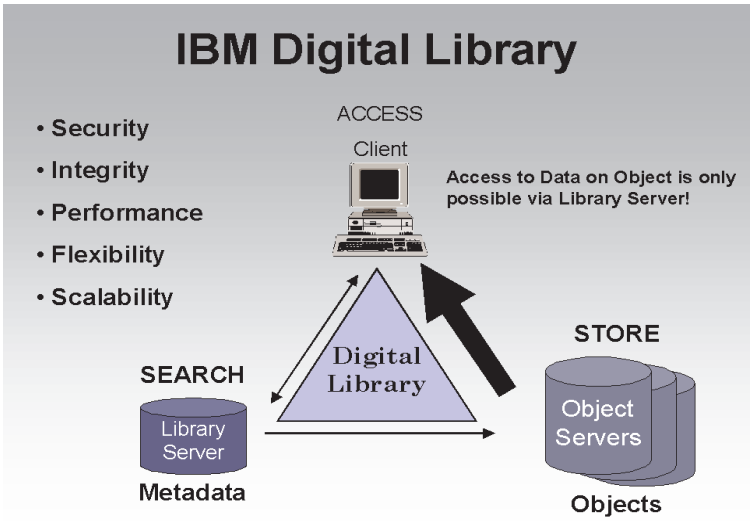
```
D. BETTEN; Komplexe Schiefparabel-Ebenen;Abhandlungen aus dem
Mathematischen Seminar der Universitaet Hamburg ISSN
0025-5858. 48, 76-88(1979).
H. GROH; Point homogeneous flat affine planes;Journal of
Geometry, ISSN 0047-2468.8, 145-162 (1976).
```

D. BETTEN, N. KNARR; Rotationsflaechen-Ebenen;Abhandlungen aus dem Mathematischen Seminar der Universitaet Hamburg, ISSN 0025-5858.57,227-234 (1988).
D. BETTEN, B. POLSTER; Topological ovals in 4-dimensional translation planes;Archiv der Mathematik. Archives of Mathematics. Archives Mathematiques, ISSN 0003-889X. 63, 284-288 (1994).
D. BRAESS; Nonlinear approximation theory;Berlin-Heidelberg-New York 1986.
A.W. GOODMAN; Univalent functions, vol.I.; Tampa 1983.
E. KIRJACKIS; On functions with non-zero n-th divided differences (Russian);Litovsk. Math. Sb. 1, 109-115(1961).

We remark that the journal "Litovsk Math. Sb." has not been recognized, because the chosen abbreviation of its title has not been added to our database. This can easily be mended.

6. Digital library Miless. Using the automatically produced bibliographic XML-data of each retrodigitized article of the volumes 60 to 67 of "Archiv der Mathematik" all these articles have been incorporated into the IBM digital library system MILESS of the "Hochschulrechenzentrum der Universität Essen". It is based on an IBM hardware for which there exists a software such that the library server containing the metadata of the journal articles and the object server storing the retrodigitized articles in MVD format provide a library environment which warrants:

1) security, which means only authorized students or scientists have access to this digital library,
2) integrity of the articles contained in the library,
3) reliable performance of the system,
4) scalability of the digital library.



By means of the MILESS software and our programs for the recognition of the articles mentioned in the references of a given research paper this digital library system provides also correct links between the articles it contains. This function is already described in [5] by H. Gollan, F. Lützenkirchen and D. Nastoll, see also [6].

At present MILESS contains only retrodigitized and recent articles of the journal "Archiv der Mathematik". Therefore only cited articles of that journal can be linked to a given article contained in MILESS in MVD or PDF format. This is explained by the following example.

Reading on screen the article "A sufficient condition for planar partition functions" by B. Polster published in Arch. Math. **65** (1995), the automatic program recognizing the cited references can be started. Among the cited articles stated in Section 5 it finds the XML-data and the ISSN number of the article

[3] D. Betten and B. Polster, Topological ovals in 4-dimensional translation planes, Arch. Math. **63**, 284–288 (1994).

Now another software produces automatically the link from the first to the second article as

http://miless.uni-essen.de/iem/ArchivderMathematik/63,284

where the volume number and the first page are used to identify the cited article. Upon this request the MILESS system starts an internal search to retrieve the reference article.

As is mentioned in [5] this linkage function is not necessarily restricted to the articles of one journal and can be applied in a distributed digital library containing both recently electronically published articles in PDF format and retrodigitized back issues of mathematical journals in MVD format.

The document management system MILESS is flexible and scalable. The contents of its object servers can easily be extended. Therefore it can hold and manage the data of a large digital library.

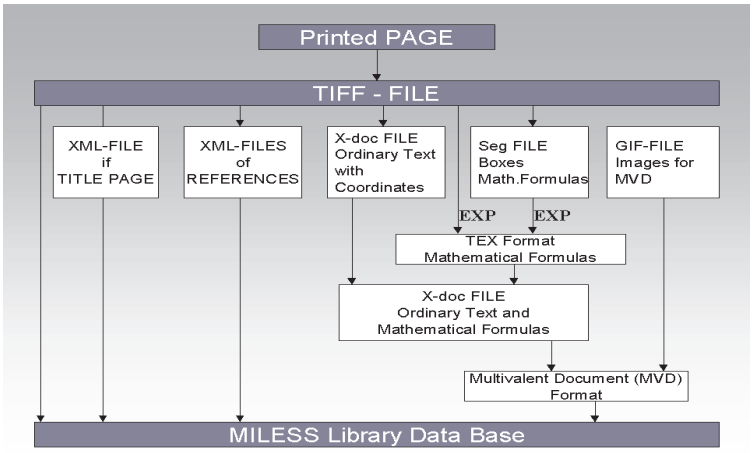
7. Combining retrodigitized back issues with recent electronically published articles. Since 1997 all articles of "Archiv der Mathematik" have been published electronically in PDF format. Using the public domain program *Image Magick 4.2.7* to the PDF files of the digital pages of a recent article we have produced a complete TIFF file of each page of the volumes **68** (1997) – **73** (1999). Applying then our programs for the recognition of the layout of the first page of an article we have obtained automatically all the bibliographical data of these articles in XML-format. Thus it is possible to include all the data of the 6 recent volumes of Archiv der Mathematik into the digital library MILESS. Furthermore, applying our programs for the recognition of the references to the TIFF files of these articles it is possible to link the recent with the back issues of the journal contained in MILESS.

8. Conclusion and outlook. In this project computer programs have been developed which enabled the author's collaborators to build a prototype of a searchable and retrievable digital library containing recently electronically published and retrodigitized back issues of mathematical journals. All digital documents of the articles of volumes **60** (1993) – **73** (1999) of Archiv der Mathematik are stored in the digital library database MILESS of the Computer Center (HRZ) of Essen University, at present. Its software is another essential tool besides the new programs developed by the author's collaborators.

It is intended to transfer all digital documents of the 8 retrodigitized volumes of Archiv der Mathematik from Essen to the "Niedersächsische Staats- und Universitätsbibliothek Göt-

tingen", which also has access to the IBM digital library database of the Computer Center of Göttingen University. Using then the MILESS software of the Computer Center of Essen University and the programs developed in this project it is possible to link these retrodigitized volumes of Archiv der Mathematik with the recently electronically published issues. Furthermore, all these articles can be linked to Mathematical Reviews and Zentralblatt at the Library of Göttingen University. Thus the authorized mathematicians can use these data via the internet.

The following picture summarizes all the necessary steps to produce the MVD format of a retrodigitized journal page stored in MILESS from the original printed page.



Our software producing the bibliographical data of an article in XML format can easily be extended in such a way that these data can be transferred to the CrossRef meta database [2]. The CrossRef project is a new reference-linking service of about 40 leading scientific and scholarly publishers. It enables linking of reference citations to the online content that those references cite, typically located on a different server and published by a different publisher. Sending an XML-query to CrossRef this metadata database returns the DOI to the article. A DOI consists of a prefix which is assigned to each publisher by the DOI agency, and a suffix which has been assigned by the publisher to the article, see [9]. The DOI can then be used to link to the URL of the article.

Also retrodigitized articles can be linked to this globally distributed digital library system. All what is necessary by the publishers is to order a DOI with the DOI agency for these cited articles for which an ISSN number of the journal, the volume and the first page have been assigned by our software. "Archiv der Mathematik" is part of the reference-linking system CrossRef through the Springer Link, see link.springer.de.

Finally it should be mentioned that in another joint project with the Japanese professors Okamoto (Shinshu University, Nagano) and Suzuki (Kyushu University, Fukuoka) new programs will be developed for improving the recognition rate of mathematical formulas. Since the reader of a retrodigitized article sees the complete picture of a mathematical formula in

the TIFF file of a scanned page, mistakes in its MVD code are not disturbing as long as one does not want to search for the complete formula.

The results of this project have been explained and demonstrated by the author and his collaborators Dr. Nörenberg, Dr. Staszewski and Dr. Waki at the international conference "Retrodigitization of mathematical journals and their integration into searchable digital libraries" organized by Prof. R. Fateman (UC Berkeley), Professor E. Mittler (University of Göttingen) and the author. It was held at the Institute for Experimental Mathematics of Essen University from August 1 to August 3, 2000 and financially supported by the Deutsche Forschungsgemeinschaft.

Acknowledgements. The author kindly acknowledges financial support by DFG grant III N 2 - 542 81(1) Essen BIB45 ENug 01-02, and technical support by the director Dr. B. Lix of the Computer Center of Essen University by providing access to MILESS.

He is very grateful to the Birkhäuser Verlag for its generous technical and legal support, in particular to Mr. B. Luchner. The author also thanks Professor R. Fateman, T. A. Phelps Ph.D and Professor D. Wilensky of the University of California (Berkeley) and Professor Okamoto for their advice and their programs [1], [10] and [8], respectively. Finally, he owes thanks to his previous or present collaborators C. Begall, Dr. H. Gollan, Dr. G. Hennecke, Mrs. B. Kratzer, Mrs. D. Mirrouch, Dr. R. Nörenberg, Dr. J. Rosenboom, Dr. R. Staszewski and Dr. Waki who have done or do all the programming and hard work during their time at the Institute for Experimental Mathematics.

This article is dedicated to Professor Dr. Erich Lamprecht, the editor in chief of Archiv der Mathematik, because he has not only ensured publishing highly interesting articles in our journal over many years, but he also has taken great care to ask the publisher and printers to produce high quality prints of the articles in uniform style. This is not easy to achieve. The retrodigitization project has benefitted very much from the high quality printing of this journal.

References

[1] B. P. BERMAN, R. J. FATEMAN, N. MITCHELL and T. TOKUYASU, Optical character recognition and parsing of typeset mathematics. Journal of Visual Communication and Image Representation, **7**, 2–15 (1996).

[2] Crossref, The central source for reference linking, http://www.crossref.org

[3] FineReader OCR Engine, Developer's Guide, ABBYY Software House (BIT Software), Moscow, 1993–1997.

[4] IBM DB2 Digital Library, http://www.software.ibm.com./is/dig-lib/

[5] H. GOLLAN, F. LÜTZENKIRCHEN and D. NASTOLL, MILESS – a learning and teaching server for multi-media documents. In G. COOPERMAN, E. JESSEN, G. O. MICHLER eds., Workshop on wide area networks and high performance computing. Lecture Notes in Control and Information Sciences **249**, 143–150, Heidelberg 1999.

[6] G. O. MICHLER, A prototype of a combined digital and retrodigitized searchable mathematical journal. Lecture Notes in Control and Information Sciences **249**, 219–234 (1999).

[7] MILESS – Multimedialer Lehr- und Lernserver Essen, http://miless.uni-essen.de

[8] M. OKAMOTO and H. MSAFIRI, Mathematical expression recognition by projection profile characteristics. Trans. IEICE Japan, J78-D-11, No. 2, 366–370 (1995).

[9] N. PASKIN, Digital object identifier DOI: Current status and outlook. DLib Magazine, http://dx.doi.org./10.1045/may99-paskin

[10] T. A. PHELPS, Multivalent Documents: Anytime, Anywhere, Any Type, Every Way User-Improvable Digital Documents and Systems. Dissertation, UC Berkeley, 1998.
[11] T. A. PHELPS and R. WILENSKY, Multivalent Documents: Inducing Structure and Behaviors in Online Digital Documents. In Proceedings of the 29th Hawaii International Conference on System Sciences Maui, Hawaii, January 3–6, 1996.

Eingegangen am 18. 11. 2000

Anschrift des Autors:

G. O. Michler
Institute for Experimental Mathematics
Essen University
45326 Essen, Germany