A Ground-Truthed Mathematical Character and Symbol Image Database

Masakazu Suzuki^{*}, Seiichi Uchida^{**} and Akihiro Nomura^{***} * Faculty of Mathematics, ** Faculty of Information Science and Electrical Engineering, * * & Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi, 812-8581 Japan suzuki@math.kyushu-u.ac.jp

Abstract

This paper describes the specifications for our ground-truthed mathematical character and symbol image database, called InftyCDB-1. The ground-truth of each character is composed of type, font, quality (touched/broken) and link (relative position), etc. The database includes all the characters and symbols of 467 pages of 30 articles on mathematics, and is organized so that it can be used as word image database or as mathematical formula image database. InftyCDB-1 is a public database that is freely usable for research and development purposes.

1 Introduction

In this paper, we report the specifications for a groundtruthed mathematical character and symbol image database, called InftyCDB-1. The database is comprised of 467 pages of 30 articles on pure mathematics. InftyCDB-1 can be used, for example, in the following purposes:

- development and evaluation of character and scientific symbol recognition,
- development and evaluation of mathematical formula recognition, and
- analysis of words in mathematical documents.

InftyCDB-1 is freely usable for those research and development purposes.

All the character/symbol images in the 467 page images are included in the database with their *ground-truth*. Thus, users can get training data or test data for character/symbol recognition from the database. The ground-truth each character/symbol is composed of type, font (italic/upright, bold or not), quality (touched/broken) and link (relative position), and so on. Thus, for example, the database can be used for evaluations of font distinction ability in character recognition. Note that for all special (i.e., very rare) mathematical symbols, their own ground-truths have been carefully attached.

The image data are stored separated into word or math formula units. Thus, InftyCDB-1 can be used as a word image database or as a mathematical formula image database. The image data are arranged in alphabetic order independent of the content of papers. No whole page image is included in the database to avoid copyright problems.

Hereafter, the term *character* means not only ordinary characters (e.g., "A"), but also math symbols (e.g., "+"), unless otherwise noted. The term *category* means the finest level of character classification and the term *type* means a set of categories having a similar property. For example, "A", "B" and "C" are three categories belonging to the same type (Roman). In contrast, "A"(Roman), "A"(italic), "A"(calligraph), "A"(blackboard bold), "A" (German), and "A" (script) are six categories belonging to different types. Each character belongs to either the *text region* or the *math region*. The math region includes not only numbered equations but also in-line math formulae. Note that many in-line math formulae are composed of a single character, such as "x" in the sentence "The variable x denotes ...".

2 Outline of database

2.1 Data collection

The documents contained in the database are 30 English articles on pure mathematics (published $1970 \sim 2000$). The numbers of pages, characters, words and math expressions in the database are 467, 688,570, 108,914 and 21,056, respectively. For a quantitative analysis of the database, see $[1]^1$. This database is larger than past databases for research on math-OCR (e.g., about 15,000 characters in [2], about 10,000 characters in [3]). Note that matrices, tables, and figures are excluded from the database.

¹There are slight differences between the table1 below and the table in [1] because some errors were found after submission of the paper [1] and corrected

| | С <i>(</i> | i din | | | in the databe | 1001 | | | 1 |
|-----------------|-------------|--|-------------|-------|-----------------|---------|------------------|-------|-----------------|
| type | Iont | category | #predefined | | text region | | math region | | total |
| | | examples | categories | #cat. | #char (% |) #cat. | #char (% | #cat. | #char (% |
| accent | | ^ ~ = ·· \ • | 13 | 1 | 2 (<0.01 |) 7 | 2,700 (1.72) | 7 | 2,702 (0.39 |
| arrow | | ${\swarrow} {\to} {\to} {\leftarrow} {\checkmark}$ | 16 | 1 | 3 (<0.01 |) 7 | 1,103 (0.70) | 7 | 1,106 (0.16 |
| big symbol | | $\Sigma \int \Pi$ | 18 | 0 | 0 (0.00 |) 11 | 2,458 (1.57) | 11 | 2,458 (0.36 |
| blackboard bold | | ABCDEF | 26 | 0 | 0 (0.00 |) 9 | 427 (0.27) | 9 | 427 (0.06 |
| calligraphic | | ABCDEF | 26 | 0 | 0 (0.00 |) 19 | 592 (0.38) | 19 | 592 (0.09 |
| German | Upright | ABCabc | 52 | 0 | 0 (0.00 |) 25 | 1,041 (0.66) | 25 | 1,041 (0.15 |
| | Bold | ABCabc | 52 | 0 | 0 (0.00 |) 0 | 0 (0.00) | 0 | 0 (0.00) |
| Greek | Upright | $\Gamma \Delta \Theta$ | 11 | 0 | 0 (0.00 |) 10 | 2,148 (1.37) | 10 | 2,148 (0.31 |
| | Italic | $lphaeta\gamma$ | 29 | 5 | 19 (<0.01 |) 23 | 10,618 (6.76) | 23 | 10,637 (1.54 |
| | Bold | $\Gamma\Delta\Theta$ | 11 | 0 | 0 (0.00 |) 1 | 3 (<0.01) | 1 | 3 (< 0.01) |
| | Italic Bold | $lphaeta\gamma$ | 29 | 0 | 0 (0.00 |) 5 | 31 (0.02) | 5 | 31 (<0.01 |
| extended Latin | Upright | ÄÆè | 182 | 30 | 392 (0.07 |) 2 | 3 (<0.01) | 30 | 395 (0.06 |
| | Italic | Ä Æè | 182 | 9 | 55 (0.01 |) 2 | 10 (0.01) | 10 | 65 (0.01 |
| | Bold | ÄÆ è | 182 | 4 | 6 (<0.01 |) 0 | 0 (0.00) | 4 | 6 (< 0.01 |
| | Italic Bold | Ă Æ è | 182 | 0 | 0 (0.00 |) 0 | 0 (0.00) | 0 | 0 (0.00) |
| numeric | Upright | 012 | 10 | 10 | 12,018 (2.26 |) 10 | 15,294 (9.74) | 10 | 27,312 (3.97 |
| | Italic | 012 | 10 | 10 | 140 (0.03 |) 4 | 118 (0.08) | 10 | 258 (0.04 |
| | Bold | 012 | 10 | 10 | 923 (0.17 |) 4 | 26 (0.02) | 10 | 949 (0.14 |
| | Italic Bold | 012 | 10 | 0 | 0(0 |) 0 | 0 (0.00) | 0 | 0 (0.00 |
| operator | | $+ - \times / < \&$ | 92 | 6 | 154 (0.03 |) 49 | 20,359 (12.96) | 50 | 20,513 (2.98 |
| others | Upright | $\S @ \bigcirc \infty \forall \exists \partial$ | 42 | 10 | 2,903 (0.55 |) 15 | 1,797 (1.14) | 20 | 4,700 (0.68 |
| | Bold | §@© | 16 | 3 | 42 (0.01 |) 0 | 0 (0.00) | 3 | 42 (0.01) |
| parenthesis | Upright | () { } [] | 20 | 7 | 8,082 (1.52 |) 12 | 30,334 (19.31) | 12 | 38,416 (5.58 |
| | Bold | () { } [] | 20 | 2 | 112 (0.02 |) 0 | 0 (0.00) | 2 | 112 (0.02 |
| point | Upright | · · · · | 17 | 11 | 21,599 (4.06 |) 11 | 8,443 (5.41) | 14 | 30,042 (4.36 |
| | Bold | · · · · | 17 | 6 | 469 (0.09 |) 0 | 0 (0.00) | 6 | 469 (0.07 |
| Roman | Upright | A B C a b c | 61 | 57 | 414,825 (78.05 |) 55 | 8,259 (5.26) | 57 | 423,084 (61.44 |
| | Italic | ABCabc | 61 | 55 | 63,590 (11.96 |) 53 | 49,072 (31.24) | 56 | 112,662 (16.36 |
| | Bold | A B C a b c | 61 | 56 | 6,178 (1.16 |) 13 | 538 (0.34) | 56 | 6,716 (0.98 |
| | Italic Bold | A B C a b c | 61 | 0 | 0 (0.00 |) 19 | 1,508 (0.96) | 19 | 1,508 (0.22 |
| script | | ABC | 52 | 0 | 0 (0.00 |) 7 | 176 (0.11) | 7 | 176 (0.03 |
| total | | | 1,571 | 294 | 531,512 (100.00 | 373 | 157,058 (100.00) | 487 | 688,570 (100.00 |

Table 1. Characters in the database.

Notes: (1) Each "Roman" and "italic" type includes nine double letters (i.e., ligatures), such as "fi".

All pages were scanned in 600 dpi and binarized automatically by the same commercial scanner (RICOH Imagio Neo 450). The quality of the resulting page images varies with the quality of original print and/or copy. Several page images are noisy and include a lot of abnormal characters, such as touching or broken characters.

2.2 Ground-truth

The ground-truth for each character was attached *manually* by seven students in, or a graduate from, a university math department. The ground-truth of each character is composed of the following attributes:

- type, category and font
- text or math region
- normal or abnormal character
- size (height and width)
- link
- location in the word or formula image
- path (folder name + file name) to the image file.

The fifth attribute, link, represents the positional relationship to the preceding character, called *parent character*,



Figure 1. Link example of a math expression.

and was attached to describe the structure of a math formula (as a tree). There are six kinds of links: horizontal, right-superscript, right-subscript, left-superscript, left-subscript, upper, and lower. Figure 1 shows the link structure of a formula $x_t = x_{t-k^2} + p$. Note that the parent character of "+" is x and neither 2 nor k. The sixth attribute, location, is the rectangular coordinates (left,top,right,bottom) of the character image in the corresponding word/formula image. Each character data is connected to the word/formula image data by the seventh attribute, path, and the sixth attribute, location.



Figure 2. Word/formula segmentation.

2.3 Word segmentation

The segmentation of words is basically done by spacing. On the other hand, a set of consecutive math expressions in a line are unified into one formula regardless of the spacing, e.g., before and after a relative operators such as equal sign, etc. A word or a formula continued to next line is separated at the end of a line. Points (e.g. ",", ".", etc) are included in the word/formula just before the points. Opening parenthesis are included in the next word and closing parenthesis are included in the preceding words like points. Quotation marks are treated in a similar way.

In Figure 2, the arrows show the segmentation points and the sentence is separated into 11 words/formulae in this example.

3 Structure of database

The database InftyCDB-1 is composed of two parts: (i) text data and (ii) image data, related to each other. Text data is a Microsoft Access or CSV-format, the user's choice, while image data are systematically named PNG files.

3.1 Text data

For each character, the 29 attributes listed in the Table 2 are attached:

The attribute (5) is code defined in our laboratory to distinguish character/symbol categories in the math-OCR software called InftyReader[5]. The attribute (6) is a string to read the character: e.g., "int" for " \int ", "Omega" for " Ω ". The attribute (7) is "text" for text region character, and "math" for math region character. The attribute (8) is "True" (resp. "False") if the character is on the baseline (resp. in sub/super-script area). The attribute (9) (resp. (10) is "True" if the character is italic (resp. bold) font and "False" otherwise. The attribute (11) is "touched" for a touched character, "separate" for a broken character, "touch_and_sep" for touched and broken character, and "normal" otherwise. By using the attributes (14), (15), a user can reproduce the math tree structure for each formula.

As for the rule to define the path to the image file in the attribute (16), see 3.2 below.

Attributes (21) - (29) are the same for all the characters in a word/formula. Attribute (21) is the ID number attached to each word/formula. Attributes (22), (23) and (24) are string data that represent the corresponding word/formula

··· unbounded har-

monic function ...

Figure 3. SyllableAfter attribute for hyphenation.

in MathML format, LaTeX format, and IML format, respectively. IML format is the XML format used in the software developed in our laboratory, math-OCR "InftyReader" and an authoring tool "InftyEditor" for mathematical documents. According to those ground-truthed string data, users can evaluate the structural analysis procedures used in their own math-OCRs.

The attribute (29) is "True" for words at the beginning of line continued by hyphenation from the word at the end of previous line, and "False" for other cases. In Figure 3, the latter part "monic" of "harmonic" has the SyllableAfter attribute "True".

A sample of the text data for a formula " $\frac{d}{dt}h_{t\nu,z_0}|_{t=0}$ " and a word "and" is shown in Table 3. As noted in 2.2, the link represents the positional relationship to the parent character. Thus, for example, the "HORIZONTAL" link of "|" (vert, CharID=24) means that it and its parent character "*h*" (CharID=18) are on the same level. Then, using the "HORIZONTAL" links from "*h*" to the fractional line (CharID=14 and the first character), it is found that the "|" is on the baseline.

3.2 Image data

To reduce the number of image files, the images of a same word in a same article are stored in one image file. However, italic words and upright words are stored in different image files, and Roman capital/small letter are distinguished so that, e.g., "And" "and" are stored in different image files. Formula images in an article are also grouped in a same way, when the expressions are identical. Image file names are defined as follows:

- word · · · "string"(_FontFlag)_ "number".png
- formula ··· "number of characters in the formula" _"first three characters in the formula" (_ FontFlag)_ "number".png

For example, the file names of the word "(and)" and the formula " $\alpha \leq 1$ " are "LeftPar-and-RightPar 0.png" and "3_alpha-le-1_0.png", respectively. Figure 4 shows two image files in the database having the path names "Images¥ActaM_1970_37_63¥A¥approaches J_0.png" and "Images¥ActaM_1970_37_63¥MATH¥1-9¥7_D-Omega-BigLeftPar_0.png", respectively.

| Attribute | Explanation |
|---------------|--|
| CharID | ID number of each character |
| JornalID | ID number of the article |
| SheetID | Page number |
| Туре | Type name (see Table 1) |
| Code | Category code (OCR code) |
| Entity | Category name |
| Region | Distinction of text/math area |
| Baseline | Distinction of baseline/subscript |
| ItalicFlag | Italic flag |
| BoldFlag | Bold flag |
| Quality | Normal/touched/separate/touch_ and_ sep |
| Width | Width of the character |
| Height | Height of the character |
| ParentCharID | ID of the parent character |
| Link | Type of link, which represents the positional relation to the parent character |
| ImageName | Path (folder name + file name) to the image file including the character |
| Rect | Coordinates of the character in the image file (left,top,right,bottom) |
| WordID | ID number of the word/formula including the character |
| WordMathML | MathML string |
| WordTeX | LaTeX string |
| WordIML | IML string |
| WordRect | Coordinates of the word/formula in the image file (left,top,right,bottom) |
| SyllableAfter | Flag of word continued from the previous line by hyphenation |
| | Attribute CharID JornalID SheetID Type Code Entity Region Baseline ItalicFlag BoldFlag Quality Width Height ParentCharID Link ImageName Rect WordID WordMathML WordTeX WordIML WordRect SyllableAfter |

Table 2. List of attributes



Figure 4. Example of image files

4 Distribution

The database InftyCDB-1 is made and will be maintained in M.Suzuki laboratory, Faculty of Mathematics, Kyushu University. It is freely available for research and development purposes after user registration. (URL: http://www.inftyproject.org/).

The text data part is in CSV (16.8MB) or Microsoft Access format (26.7MB) as the user chooses, and the image data is in PNG format (202MB). The database is delivered in CD-ROM.

5 Conclusion

In this paper, we described our ground-truthed mathematical character and symbol image database, called InftyCDB-1. The database consists of two parts: text data and image data that are related to each other.

The ground-truth of each character is composed of type, font, quality (touched/broken) and link information to represent the tree structure of math formula, etc. The database includes all the 688,570 characters (and symbols) of 467 pages of 30 English articles on mathematics (published 1970 \sim 2000). Characters are grouped into words/formulae in the database. Total number of words and formulae in the database are 108,914 and 21,056, respectively. The database is freely usable for research, development and evaluation of math-OCRs.

References

- [1] S. Uchida, A. Nomura, M. Suzuki "Quantitative Analysis of Mathematical Documents," *Int. J. Doc. Anal. Recog.*, to appear.
- [2] H.-J. Lee and J.-S. Wang, "Design of a mathematical expression understanding system," *Pattern Recognition Letters*, 18(3):289–298, 1997.
- [3] M. Okamoto, H. Imai, and K. Takagi, "Performance evaluation of a robust method for mathematical expression recognition," *Proc. ICDAR*, 121-128, 2001.
- [4] A. Nomura, K. Michishita, S. Uchida, and M. Suzuki, "Detection and segmentation of touching characters in mathematical expressions," *Proc. ICDAR*, 1:126-130, 2003.

| | | | | | | | | | | | | | | uu | | | | | |
|---|---|--|---|---|--|--|---|--|---|--|--|--|--|--|---|---|---|--|--|
| | (1) | (2) | (3) | | (4) | (5) |) | (6) | | (7) | (8) | (8) | | (10) | () | 11) | (12) | (13) | |
| | 14 | 5 | 10 | Big | Symbol | 33D | 1 fra | fractionalLine | | math TRU | | E F | FALSE | FALSE | nor | mal | 82 | 12 | |
| | 15 | 5 | 10 | Roi | man | 016 | 4 d | | | math | TRU | TRUE F | | FALSE | normal | | 38 | 60 | |
| | 16 | 5 | 10 | Roi | man | 016 | 4 d | | | math | TRUE | | FALSE | FALSE | ALSE normal | | 38 | 59 | |
| | 17 | 5 | 10 | Roi | man | 0174 | 4 t | | | math | TRU | E F | FALSE | FALSE | nor | mal | 24 | 53 | |
| | 18 | 5 | 10 | Roi | man | 016 | 8 h | | | math | TRU | E F | FALSE | FALSE | nor | mal | 39 | 60 | |
| | 19 | 5 | 10 | Roi | man | 0174 | 4 t | | | math | FALS | SE F | FALSE | FALSE | nor | mal | 27 | 52 | |
| | 20 | 5 | 10 | Gre | eek | k 426D | | nu comma | | math | FALS | SE T | RUE | FALSE | FALSE nor | | 42 37 | | |
| | 21 | 5 | 10 | Poi | pint 142C | | C coi | | | math | FALS | SE F | FALSE | FALSE | nor | mal | 17 | 26 | |
| | 22 | 5 | 10 | Roi | man | 017. | A z | | | math | FALS | SE F | FALSE | FALSE | nor | mal | 38 | 39 | • • • |
| | 23 | 5 | 10 | Nu | meric | 013 |) zer | zero | | math | math FALS | | FALSE | SE FALSE | | normal | | 49 | |
| | 24 | 5 | 10 | Par | enthesis | 197 | C vei | vert | | math | th TRU | | FALSE | LSE FALSE | | mal | 11 | 187 | |
| | 25 | 5 | 10 | Roi | man | 0174 | 4 t | | | math | FALS | SE F | FALSE | FALSE | nor | mal | 26 | 52 | |
| | 26 | 5 | 10 | Ope | erator | 1D3 | D equ | ual | | math | FALS | SE F | FALSE | FALSE | tou | ched | 43 | 22 | |
| | 27 | 5 | 10 | Nu | meric | 013 |) zer | 0 | | math | FALS | SE F | FALSE | FALSE | nor | mal | 35 | 48 | |
| | : | : | : | : | | : | : | | | : | : | : | | : | : | | : | : | |
| - | | | | · Der | | | | | | TDUE | · | · | ALCE | EALCE | | | 27 | . 20 | |
| - | 804 | 2 | 13 | R01 | man | 010 | I a | | | TRUE | text | Г | ALSE | FALSE | 101 | aroto | 37 | 26 | |
| | 805 | 2 | 13 | Pot | man | 010 | | | | TRUE | text | L L | ALSE | FALSE | sep | mal | 42 | 56 | |
| | | | | | | | T U | | | | 11.701 | | | | | | | 50 | |
| | 14) | 2 | (15) | Roi | (16) | (17) | (19) | (10) | (20 | | 21) | (22) | (22) | (24) | (25) | (26) | (27) | (28) | (20) |
| (| 14) | тор | (15) | Ro | (16) | (17) | (18) | (19) | (20 |) (2 | 21) | (22) | (23) | (24) | (25) | (26) | (27) | (28) | (29) FALSE |
| (| 14) -1 14 | TOP | (15) ER | Rol | (16) *1 | (17) 0 12 | (18) 73 | (19) 82 50 | (20 8 |) (2 5 2800 1 2800 | 21) 05695 | (22) *2 *2 | (23) *3 | (24) *4 *4 | (25) 0 | (26) 0 | (27) 522 522 | (28) 205 205 | (29) FALSE FALSE |
| (| 14) -1 14 14 | TOP UND | (15) ER ER | Ro | (16) *1 *1 *1 | (17) 0 12 24 | (18) 73 101 | (19) 82 50 62 | (20 8 16 | $\begin{array}{c} 1102 \\ \hline 0 \\ 5 \\ 2800 \\ \hline 1 \\ 2800 \\ \hline 9 \\ 2800 \end{array}$ | 21) 05695 05695 | (22) *2 *2 *2 | (23) *3 *3 *3 | (24) *4 *4 | (25) 0 0 | (26) 0 0 | (27) 522 522 | (28) 205 205 205 | (29) FALSE FALSE FALSE |
| (| 14) -1 14 14 14 15 | TOP UND UPPE HOR | (15) ER ER IZONT | | (16) *1 *1 *1 *1 | (17) 0 12 24 54 | (18) 73 101 0 109 | (19) 82 50 62 78 | (20 8 16 5 | $\begin{array}{c} 1102\\ \hline \\ 15\\ 2800\\ 1\\ 2800\\ $ | 21) 05695 05695 05695 | (22) *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 | (25) 0 0 0 | (26) 0 0 0 0 | (27) 522 522 522 522 | (28) 205 205 205 205 | (29) FALSE FALSE FALSE FALSE |
| (| 14) -1 14 14 14 14 14 14 14 | TOP UND UPPE HOR HOR | (15) ER ER IZONT | | (16) *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 | $ \begin{array}{r} \hline 18) \\ 73 \\ 101 \\ 0 \\ $ | (19) 82 50 62 78 134 | (20 8 16 5 16 9 | $\begin{array}{c} 1102\\ \hline \\ 1 \\ 5 \\ 2800\\ \hline \\ 1 \\ 2800\\ \hline \\ 2 \\ 2800\\ \hline \\ 5 \\ 2800\\ \hline \\ 5 \\ 2800\\ \hline \end{array}$ | 21) 05695 05695 05695 05695 05695 | (22) *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 | (25) 0 0 0 0 0 0 | (26) (26) 0 0 0 0 0 0 | (27) 522 522 522 522 522 522 | (28) 205 205 205 205 205 205 | (29) FALSE FALSE FALSE FALSE FALSE |
| (| 14) -1 14 14 14 15 14 18 | TOP UND UPPE HOR HOR RSUE | (15) ER ER IZONT IZONT | TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 | $ \begin{array}{c} (18)\\ 73\\ 101\\ 0\\ 109\\ 35\\ 78\\ \end{array} $ | (19) 82 50 62 78 134 162 | (20) 8 16 5 16 9 13 | (1) (2) 5 2800 1 2800 9 2800 2 2800 5 2800 0 2800 | 21) 05695 05695 05695 05695 05695 05695 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 | (25) 0 0 0 0 0 0 0 | $ \begin{array}{c} (26) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0$ | (27) 522 522 522 522 522 522 522 522 | (28) 205 205 205 205 205 205 205 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE |
| (| 14) -1 14 14 14 15 14 15 14 19 | TOP UND UPPE HOR HOR RSUE | (15) ER ER IZONT IZONT 3 IZONT | TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 173 | $(18) \\ (18) \\ 73 \\ 101 \\ 0 \\ 109 \\ 35 \\ 78 \\ 96$ | (19) 82 50 62 78 134 162 215 | (20) 8 16 5 16 9 13 13 |) (1) 5 2800 1 2800 9 2800 2 2800 5 2800 5 2800 0 2800 3 2800 | 21) 05695 05695 05695 05695 05695 05695 05695 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | (25) 0 0 0 0 0 0 0 0 0 | $ \begin{array}{c} (26) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0$ | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 205 205 205 205 205 205 205 205 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | 14) -1 -1 14 14 15 14 15 14 19 20 20 | TOP UND UPPE HOR HOR HOR HOR | (15) ER ER IZONI IZONI IZONI IZONI | TAL TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 173 234 | $(18) \\ (18) \\ 73 \\ 101 \\ 0 \\ 109 \\ 35 \\ 78 \\ 96 \\ 119 \\ (18) \\ $ | (19) 82 50 62 78 134 162 215 251 | (20) 8 16 5 16 9 13 13 13 | 1 2800 5 2800 1 2800 9 2800 2 2800 5 2800 0 2800 3 2800 5 2800 | 21) 05695 05695 05695 05695 05695 05695 05695 05695 05695 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | (25) 0 0 0 0 0 0 0 0 0 0 0 0 | (26) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 205 205 205 205 205 205 205 205 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | 14) -1 14 14 14 15 14 15 14 19 20 21 | TOP UND UPPE HOR HOR HOR HOR HOR | (15) ER ER IZONT IZONT IZONT IZONT | TAL TAL TAL TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 173 234 262 | $(18) \\ (18) \\ 73 \\ 101 \\ 0 \\ 109 \\ 35 \\ 78 \\ 96 \\ 119 \\ 92$ | $(19) \\ 82 \\ 50 \\ 62 \\ 78 \\ 134 \\ 162 \\ 215 \\ 251 \\ 300 \\$ | (20) 8 16 5 16 9 9 13 13 13 14 14 | 1 2800 1 2800 9 2800 2 2800 2 2800 2 2800 3 2800 5 2800 3 2800 5 2800 1 2800 3 2800 5 2800 1 2800 | 21))5695)5695)5695)5695)5695)5695)5695)5695)5695)5695 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | (25) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | $\begin{array}{c} (26) \\ (26) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ $ | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 205 205 205 205 205 205 205 205 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | 14) -1 14 14 14 14 15 14 18 19 20 21 22 22 | TOP UND UPPE HOR HOR HOR HOR RSUI | (15) ER ER IZONT IZONT IZONT IZONT IZONT IZONT 3 | TAL TAL TAL TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 173 234 262 306 | $\begin{array}{c} (18) \\ \hline (18) \\ \hline 73 \\ 101 \\ \hline 0 \\ 109 \\ \hline 35 \\ \hline 78 \\ 96 \\ \hline 119 \\ 92 \\ \hline 116 \end{array}$ | (19) 82 50 62 78 134 162 215 251 300 340 | (20) 8 16 5 5 16 9 9 13 13 13 14 13 16 | (1) (2) 5 2800 1 2800 9 2800 2 2800 5 2800 0 2800 5 2800 3 2800 5 2800 1 2800 5 2800 5 2800 1 2800 5 2800 | 21))5695)5695)5695)5695)5695)5695)5695)5695)5695)5695)5695)5695)5695 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | (25) 0 0 0 0 0 0 0 0 0 0 0 0 0 | (26) (26) 00 00 00 00 00 00 00 00 00 0 | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 205 205 205 205 205 205 205 205 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | 14) -1 14 14 14 14 15 14 18 19 20 21 22 18 | TOP UND UPPE HOR HOR HOR HOR RSUI HOR | (15) ER ER IZONT IZONT IZONT IZONT IZONT 3 IZONT | TAL TAL TAL TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 173 234 262 306 363 | $\begin{array}{c} (18) \\ \hline (18) \\ \hline 73 \\ \hline 101 \\ \hline 0 \\ \hline 109 \\ \hline 35 \\ \hline 78 \\ \hline 96 \\ \hline 119 \\ \hline 92 \\ \hline 116 \\ \hline 18 \\ \end{array}$ | (19) 82 50 62 78 134 162 215 251 300 340 374 | (20) (20) 8 166 5 166 99 133 133 144 133 166 200 | (1) (2) (5) 2800 1 2800 9 2800 2 2800 2 2800 5 2800 0 2800 5 2800 0 2800 5 2800 5 2800 5 2800 5 2800 5 2800 | 21))5695)5695)5695)5695)5695)5695)5695)5695)5695)5695)5695)5695)5695)5695)5695 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | $\begin{array}{c} (25) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ $ | (26) (26) 0 0 0 0 0 0 0 0 0 0 0 0 0 | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 205 205 205 205 205 205 205 205 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | 14) -1 14 14 15 14 15 14 15 20 21 22 18 24 | TOP UND UPPE HOR HOR HOR HOR RSUI HOR RSUI | (15) ER ER IZONT IZONT IZONT IZONT 3 IZONT 3 IZONT 3 | TAL TAL TAL TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 173 234 262 306 363 386 | $\begin{array}{c} (18) \\ \hline (18) \\ \hline 73 \\ \hline 101 \\ \hline 0 \\ \hline 109 \\ \hline 35 \\ \hline 78 \\ \hline 96 \\ \hline 119 \\ \hline 92 \\ \hline 116 \\ \hline 18 \\ \hline 130 \\ \end{array}$ | (19) 82 50 62 78 134 162 215 251 300 340 374 412 | (20 8 16 5 16 9 13 13 13 14 13 16 20 18 | (1) (2) (5) (2800) (5) 2800 (1) 2800 (2) 2800 (2) 2800 (2) 2800 (3) 2800 (3) 2800 (3) 2800 (5) 2800 (1) 2800 (5) 2800 (1) 2800 (5) 2800 (5) 2800 (5) 2800 (2) 2800 (2) 2800 (3) 2800 (4) 2800 (5) 2800 (5) 2800 (2) 2800 | 21) 15695 1569 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | $\begin{array}{c} (25) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ $ | (26) 0 0 0 0 0 0 0 0 0 0 0 0 0 | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 205 205 205 205 205 205 205 205 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | 14) -1 14 14 15 14 15 14 15 14 15 14 15 14 15 14 15 14 15 14 15 14 15 14 15 14 15 14 15 14 15 14 15 14 15 14 15 20 21 22 18 24 25 | TOP UND UPPE HOR HOR HOR HOR RSUI HOR RSUI HOR | (15) ER ER IZONT IZONT IZONT IZONT B IZONT 3 IZONT 3 IZONT | TAL TAL TAL TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 173 234 262 306 363 386 423 | $\begin{array}{c} (18) \\ (73) \\ 101 \\ 0 \\ 109 \\ 35 \\ 78 \\ 96 \\ 119 \\ 92 \\ 116 \\ 18 \\ 130 \\ 154 \end{array}$ | (19) 82 50 62 78 134 162 215 251 300 340 374 412 466 | (20) | (1) (2) (5) 2800 (1) 2800 (2) 2800 (2) 2800 (2) 2800 (2) 2800 (2) 2800 (2) 2800 (3) 2800 (3) 2800 (3) 2800 (3) 2800 (3) 2800 (3) 2800 (5) 2800 (1) 2800 (5) 2800 (2) 2800 (3) 2800 (5) 2800 (2) 2800 (2) 2800 (2) 2800 (2) 2800 (2) 2800 (2) 2800 (2) 2800 (2) 2800 (3) 2800 (4) 2800 (5) 2800 (2) 2800 | 21))5695)569 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | $\begin{array}{c} (25) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0$ | (26) 0 0 0 0 0 0 0 0 0 0 0 0 0 | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 205 205 205 205 205 205 205 205 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | 14) -1 14 14 15 14 15 14 15 14 15 14 18 19 20 21 22 18 24 25 26 | TOP UND UPPE HOR HOR HOR HOR HOR RSUI HOR RSUI HOR HOR | (15) ER ER IZONT IZONT IZONT IZONT 3 IZONT 3 IZONT IZONT IZONT | TAL TAL TAL TAL TAL TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 173 234 262 306 363 386 423 487 | $\begin{array}{c} (18) \\ (73) \\ 101 \\ 0 \\ 109 \\ 35 \\ 78 \\ 96 \\ 119 \\ 92 \\ 116 \\ 18 \\ 130 \\ 154 \\ 135 \end{array}$ | (19) 82 50 62 78 134 162 215 251 300 340 374 412 466 522 | (20) 8 16 5 5 16 9 9 13 13 13 14 13 16 20 18 17 18 | (1) (2) (5) (2800) (1) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (3) (2800) (3) (2800) (5) (2800) (5) (2800) (5) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (2) (2800) (3) (2800) | 21))5695)569 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | $\begin{array}{c} (25) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ $ | (26) 0 0 0 0 0 0 0 0 0 0 0 0 0 | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 205 205 205 205 205 205 205 205 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | 14) -1 14 14 15 14 15 14 15 14 15 14 15 14 15 14 18 20 21 22 18 24 25 26 . | TOP UND UPPE HOR HOR HOR HOR RSUI HOR HOR HOR | (15) ER ER IZONI IZONI IZONI IZONI 3 IZONI 3 IZONI 3 IZONI 1 ZONI | TAL TAL TAL TAL TAL TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 173 234 262 306 363 386 423 487 | $\begin{array}{c} (18) \\ (73) \\ (10) \\ (1$ | (19) 82 50 62 78 134 162 215 251 300 340 374 412 466 522 | (20) (20) (20) (20) (20) (20) (20) (20) | (1) (2) (7) (7) (7) | 21) 05695 0569 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | (25) 0 0 0 0 0 0 0 0 0 0 0 0 0 | (26) 0 0 0 0 0 0 0 0 0 0 0 0 0 | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | $\begin{array}{c} 14) \\ \hline 14) \\ \hline -1 \\ 14 \\ \hline 14 \\ 15 \\ 14 \\ 18 \\ 19 \\ 20 \\ 21 \\ 22 \\ 18 \\ 24 \\ 25 \\ 26 \\ \hline \\ 26 \\ \hline \\ \\ \end{array}$ | TOP UND UPPE HOR HOR HOR HOR RSUI HOR RSUI HOR HOR | (15) ER ER IZONI IZONI IZONI IZONI 3 IZONI 3 IZONI 3 IZONI 1 ZONI | TAL TAL TAL TAL TAL TAL TAL | (16) *1 *1 *1 *1 *1 *1 *1 *1 *1 *1 | (17) 0 12 24 54 95 135 173 234 262 306 363 386 423 487 | $\begin{array}{c} (18) \\ (18) \\ 73 \\ 101 \\ 0 \\ 109 \\ 35 \\ 78 \\ 96 \\ 119 \\ 92 \\ 116 \\ 18 \\ 130 \\ 154 \\ 135 \\ \vdots \\ 225 \end{array}$ | (19) 82 50 62 78 134 162 215 251 300 340 374 412 466 522 | (20) 88 166 55 166 99 133 133 144 133 166 200 188 177 188 | (1) (2) (7) (7) (7) | 21) 05695 0569 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | $\begin{array}{c} (25) \\ (25) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ $ | | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | 14) -1 14) -1 14 14 15 14 18 19 20 21 22 18 24 25 26 : : -1 | TOP UND UPPE HOR HOR RSUI HOR HOR RSUI HOR RSUI HOR HOR HOR | (15) ER ER IZONT IZONT 3 IZONT IZONT 3 IZONT 3 IZONT 1 ZONT | TAL TAL TAL TAL TAL TAL TAL | $\begin{array}{c} (16) \\ *1 \\ *1 \\ *1 \\ *1 \\ *1 \\ *1 \\ *1 \\ *$ | $\begin{array}{c} (17) \\ 0 \\ 12 \\ 24 \\ 54 \\ 95 \\ 135 \\ 173 \\ 234 \\ 262 \\ 306 \\ 363 \\ 386 \\ 423 \\ 487 \\ \vdots \\ 0 \\ 0 \\ \end{array}$ | $\begin{array}{c} (18) \\ (73) \\ (10) \\ (1$ | $\begin{array}{c} (19)\\ 82\\ 50\\ 62\\ 78\\ 134\\ 162\\ 215\\ 251\\ 300\\ 340\\ 374\\ 412\\ 466\\ 522\\ \vdots\\ 37\\ 37\\ \end{array}$ | (20) (20) (20) (20) (20) (20) (20) (20) | (1) (2) (7) (7) (7) | 21))5695)569 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 *3 | $\begin{array}{c} (24) \\ *4 \\ *4 \\ *4 \\ *4 \\ *4 \\ *4 \\ *4 \\ $ | $\begin{array}{c} 1.0.1 \\ (25) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ $ | (26) 0 0 0 0 0 0 0 0 0 0 0 0 0 | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 5888 | (29) FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE |
| | 14) -1 14) -1 14 14 15 14 18 19 20 21 22 18 24 25 26 | TOP UND UPPE HOR HOR HOR HOR RSUI HOR RSUI HOR HOR TOP HOR | (15) ER ER IZONT IZONT IZONT IZONT IZONT IZONT IZONT | TAL TAL TAL TAL TAL TAL | $\begin{array}{c} (16) \\ *1 \\ *1 \\ *1 \\ *1 \\ *1 \\ *1 \\ *1 \\ *$ | $\begin{array}{c} (17) \\ 0 \\ 12 \\ 24 \\ 54 \\ 95 \\ 135 \\ 173 \\ 234 \\ 262 \\ 306 \\ 363 \\ 386 \\ 423 \\ 487 \\ \vdots \\ 0 \\ 40 \\ \end{array}$ | $\begin{array}{c} (18) \\ (73) \\ 101 \\ 0 \\ 109 \\ 35 \\ 78 \\ 96 \\ 119 \\ 92 \\ 116 \\ 18 \\ 130 \\ 154 \\ 135 \\ \vdots \\ 5850 \\ 5851 \end{array}$ | $\begin{array}{c} (19)\\ 82\\ 50\\ 62\\ 78\\ 134\\ 162\\ 215\\ 251\\ 300\\ 340\\ 374\\ 412\\ 466\\ 522\\ \vdots\\ 37\\ 82\\ \end{array}$ | (20) 88 166 55 166 166 99 91 33 133 144 133 166 200 188 177 188 5888 5588 |)) (7) (7) (7) (7) (7) | 21) 15695 1695 | (22) *2 *2 *2 *2 *2 *2 *2 *2 *2 *2 | (23) *3 *3 *3 *3 *3 *3 *3 *3 *3 *3 | (24) *4 *4 *4 *4 *4 *4 *4 *4 *4 *4 | $\begin{array}{c} (25) \\ (25) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ $ | $\begin{array}{c} (26) \\ (26) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ $ | (27) 522 522 522 522 522 522 522 522 522 52 | (28) 205 | (29)FALSE |

Table 3. Example of the text data in the database for a formula " $\frac{d}{dt}h_{t\nu,z_0}|_{t=0}$ " and a word "and".

(*1) AnnMS_1971_157_173¥MATH¥10-19¥14_fractionalLine-dd_0.png, (*5) ActaM_1970_37_63¥A¥and_1.png

(*2),(*3), (*4): Expression of the formula $\frac{d}{dt}h_{t\nu,z_0}|_{t=0}$ in MathML, in LaTeX and in IML respectively.

- [5] M. Suzuki, F. Tamari, R. Fukuda, S. Uchida, T. Kanahori "Infty- an integrated OCR system for mathematical documents," ACM Symposium on Document Engineering, 95-104, 2003
- [6] J. Ha, R. M. Haralick, and I. T. Phillips, "Understanding mathematical expressions from document images," *Proc. ICDAR*, 956-959, 1995.
- [7] Y. Eto and M. Suzuki, "Mathematical formula recognition using virtual link network," *Proc. ICDAR*, 762– 767, 2001.

List of articles in the database

Acta Math., 124(1-2), 37-63, 1970. *ibid.*, 181(2), 283-305, 1998.
Ann. Sci. Ecole Norm. Sup., 4d sér, t.3, 273-284, 1970.

ibid., t.30, 367-384, 1997. • Ann. Inst. Fourier, 20(1), 493-498, 1970. • *ibid.*, 49(2), 375-404, 1999. • Ann. Math., 91, 550-569, 1970. • Ann. Math. Studies, 66, 157–173, 1971. • Arkiv für Matematik, 9(1), 141-163 1971. • *ibid.*, 35(1), 185-199, 1997.
Bull. Amer. Math. Soc., 77(1), 157-159 1971. • *ibid.*, 77(1), 160-163 1971. • *ibid.*, 80(6), 1219-1222, 1974. • *ibid.*, 35(2), 123-143, 1998. • Bull. Soc. Math. France, 98, 165-192, 1970. *ibid.*, 126, 245-271, 1998. • Invent. Math., 9, 121-134, 1970. *ibid.*, 138, 163-181, 1999. • J. Math. Soc. Japan, 27(2), 281-288, 1975. • *ibid.*, 27(2), 289-293, 1975. • *ibid.*, 27(2), 497-506, 1975. • J. Math. Kyoto Univ., 11(1), 181-194, 1971. • *ibid.*, 11(1), 373-375, 1971. • *ibid.*, 11(2), 377-379, 1971. • *kyushu J.* Math., 53, 17-36, 1999. • Math. Ann., 225(3), 275-292, 1977. *ibid.*, 315, 175-196, 1999. • Tohoku Math. J., 25, 317-331, 1973. • *ibid.*, 25, 333-338, 1973. • *ibid.*, 42, 163-193, 1990.