ICDAR2021 Competition on Mathematical Formula Detection

Dan Anitei[0000−0001−8288−6009], Joan Andreu Sánchez[0000−0003−0423−2020], José Manuel Fuentes[0000−0002−5827−7489], Roberto Paredes[0000−0002−5192−0021], and José Miguel Benedí[0000−0001−6516−2746]

Pattern Recognition and Human Language Technologies Research Center
Universitat Politècnica València
Valencia 46022, Spain
{danitei,jofuelo1,jandreu,rparedes,jmbenedi}@prhlt.upv.es

Abstract. This paper introduces the Competition on Mathematical Formula Detection that was organized for the ICDAR 2021. The main goal of this competition was to provide the researchers and practitioners a common framework to research on this topic. A large dataset was prepared for this contest where the GT was automatically generated and manually reviewed. Fourteen participants submitted their results for this competition and these results show that there is still room for improvement especially for the detection of embedded mathematical expressions.

1 Introduction

Currently, huge amounts of documents related to science, technology, engineering, and mathematics (STEM) are being published by online digital libraries worldwide. Searching in STEM documents is one of the most usual activities for researchers, scholars, and scientists worldwide. Searching for plain text in large electronic STEM collections is considered a solved problem when queries are simple strings or regular expressions. However, searching for complex structures like chemical formulas, plots, draws, maps, tables, and mathematical expressions, among many others, remains scarcely explored [11, 6]. This paper describes the competition that was performed in the context of the International Conference on Document Analysis and Recognition 2021 for searching mathematical expressions (MEs) in large collections of STEM documents. This competition was stated as a visual detection problem and therefore the dataset was just provided with visual information.

The searching can be based just on visual features or it can also use textual information around the ME. This paper focuses on the former approach, while the latter approach is left for future research. It is worth mentioning that searching in large datasets may require performing some pre-processing since recognition and search in query time can be prohibitive.

Searching for MEs requires as a first step locating them in the document. MEs can be either embedded along the lines of the text or displayed. We refer to

the first ones as *embedded* MEs and the second ones as *displayed* MEs. Locating displayed MEs can be easily performed with profile projection methods since these expressions are separated from the text, although they can be confused with other graphic elements (e.g. tables, figures, plots, etc.). However, it is more difficult to locate embedded MEs since they can be easily confused with running text. Given the relevance of this problem for the STEM research community we considered it very challenging and proposed this first competition to know how far the current technology is from providing good results for this problem.

Current technology for ME recognition in documents is based on Machine Intelligent methods that need a large amount of data [1, 6] with the necessary ground truth, both for training and testing. However, this ground truth has to be prepared manually. In fact, this is one of the bottlenecks for researching automatic methods. Developing techniques for preparing large datasets with ground truth (GT) is a real need. This paper introduces the IBEM dataset[1] for this competition that is currently made up of 600 research papers, more than 8 000-page images, and more than 160 000 MEs. The dataset has been automatically generated from the LaTeX version of the documents and consequently can be enlarged easily. The ground truth includes the position at the page level for each ME both embedded in the text and displayed.

This paper is organized as follows: first, the literature about the problem of searching MEs is reviewed. Then, we describe the dataset used in this competition and how it is structured. The competition is described in Section 4 and then we present the results and the conclusions.

2 Related Work

Searching MEs in documents is not a very researched problem [11] although STEM researchers devote a lot of time to look for information for their daily research. Looking for MEs in the printed text has recently received attention as some competitions have shown [12].[2] One problem that the authors of this paper have identified related to the searching of MEs is that no paper exists about searching in large datasets of STEM documents and this is one of the important targets in this competition.

The prevalent technology for pattern recognition is based on machine learning techniques and these techniques require large amounts of training sets. There are several datasets of typeset MEs that have been used in the past for various research projects. The UW-III dataset [7] is well-known, but the amount of data it contains is limited. It provides GT for symbol classification and the LaTeXversion is also available. A second well-known dataset is the InftyCDB-1 dataset [10] that contains 21 056 MEs. This dataset does not include matrices, tables, and figures. The relationship among symbols in a ME was defined manually, and the markup language is not included in the GT. Another important dataset is the

[1] http://ibem.prhlt.upv.es/en/
[2] https://ntcir-math.nii.ac.jp/introduction/

IM2LATEX-100K [1], that includes 103 556 different LaTeX MEs along with rendered pictures. The MEs were extracted by parsing the LaTeX sources of papers from tasks I and II of the 2003 KDD cup [2]. The problem that we identified in this dataset is that it is useful for researching the recognition of MEs, but not for the detection of MEs in the context of the article they come from. This issue is pointed in [6], where the proposed solution was to build a new dataset that contains 47 articles with 887 pages, but the total number of MEs is not provided. Finally, it is worth mentioning that this last dataset has been used in a competition on MEs detection [5].

The limitations pointed above, lead to the conclusion that a dataset that includes large numbers of images of pages from scientific documents, where MEs are located and annotated, and where the markup language is available, needed to be built.

3 Dataset description

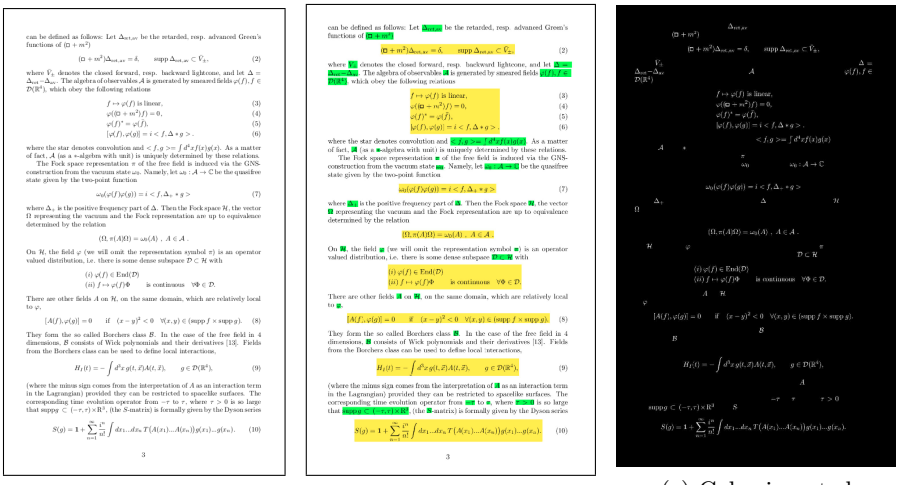
We chose the KDD Cup dataset [2], which has been used for knowledge extraction and data-mining purposes. More importantly, this collection of documents is publicly available and it allowed us to overcome copyright issues.

The KDD Cup dataset is a large collection of research papers ranging from 1992 to 2003 inclusive, with approximately 29 000 documents in total. The LaTeX sources of all papers are available for download. We chose 600 documents to create the dataset presented in this paper.

The GT prepared for the documents includes the location and dimension of MEs (upper left corner coordinates, width, and height), and their type (embedded or displayed). The bounding boxes of the MEs were highlighted, which allowed to filter out the running text, resulting in images containing only ME definitions. Highlighting the MEs would also provide a way to visually verify and validate the GT. An example of the output described before can be seen in Figure 1.

To obtain the GT and the output shown in Figure 1, we divided the extraction process into two parts. The first part consisted of designing LaTeX macros for highlighting and extracting the MEs. The second part consisted of building regular expressions that would detect the delimiters used to define LaTeX mathematical environments and automate the insertion of the LaTeX macros created in the first part. For this reason, we decided to only focus on documents that did not rename these delimiters.

For each type of MEs, embedded or displayed, we used different approaches to compute the location of the bounding boxes. We made use of the `savepos` module provided by the `zref` [3] package to obtain the absolute coordinates of the starting and ending points of the expressions as rendered on the page image. Once these coordinates were calculated, the dimensions of the bounding box were computed.



(a) Original page. (b) Highlighted bounding boxes. (c) Color inverted.

Fig. 1. An example of the output obtained from processing a page from the collection.

In the case of displayed MEs, the first and last symbols of the expression would not always be rendered in the upper left and lower right corner of the bounding box. Figure 2 shows an example of this situation.

$$\sum_{Y_1 \sqcup Y_2 = Y,\, X_1 \sqcup X_2 = X} (-1)^{(|Y_1 \cap Y_4| + |X_1 \cap X_4|)} [R^*(Y_1 \cap Y_4, X_1 \cap X_4) \times_\hbar A \times_\hbar R(Y_2 \cap Y_4, X_2 \cap X_4)] \cdot (-1)^{(|Y_1 \cap Y_3| + |X_1 \cap X_3|)} [R^*(Y_1 \cap Y_3, X_1 \cap X_3) \times_\hbar R(Y_2 \cap Y_3, X_2 \cap X_3)] = 0. \quad \square$$

Fig. 2. Example of a displayed ME in which the coordinates of the first symbol don't correspond to the upper leftmost pixel of the ME bounding box.

To take into account situations like the one in Figure 2, macros were inserted to take coordinate measurements before and after each newline symbol and also between the symbols of some mathematical elements such as fractions, sums, products, integrals, etc., which could have superscript or subscript elements making them be rendered in an upper or lower position than the first or last symbol, respectively. As in the case of embedded MEs, we had to take into account that the ME could be split over two pages or columns.

Once the LaTeX macros for highlighting and extracting MEs were created, we proceeded with the second phase, which implied designing regular expressions for automating the insertion of these macros. Since the number of regular ex-

pressions would be quite large given the many alternatives of environments that LaTeX provides for defining MEs, in this project we used the `sed` tool [8] as the preferred text stream editor.

Once the MEs were highlighted (green for embedded, yellow for displayed), we compiled the resulting LaTeX file to obtain a pdf format, which we broke down into images of the corresponding pages. Each such image was later processed[3] and passed through a yellow and green color filter to obtain a negative image similar to the one that is shown in Figure 1.

Bearing in mind that the GT of the dataset presented in this paper was generated automatically from a collection of scientific papers, there was no previous information about the position of MEs in these documents. In light of this fact, data validation had to be done by visually checking that the bounding boxes of the MEs present were in the correct position and of the right dimension.

We manually chose 600 documents that were visually validated and we proceeded with extracting the GT. Table 1 highlights the characteristics of the resulting dataset[4]. It is important to remark that the GT includes some ME that may seem erroneous. For example, sometimes multi-line ME are marked with just one bounding box while on other occasions, there is a separated bounding box for each line. We do not consider these two situations as an error because in the end the GT just represents how the LaTeX users are using this editor for writing MEs. This fact was relevant enough in the competition since distinguishing these situations affected the results obtained in the competition as we illustrate in the Results section.

Table 1. Statistics about the dataset.

Total no. of documents	600
Total no. of pages	8 272
No. of displayed MEs	29 593
No. of embedded MEs	136 635
Average no. of pages per document	13.79
Average no. of displayed MEs per document	49.32
Average no. of embedded MEs per document	227.73

4 Competition description

4.1 Protocol

The protocol for the competition followed these steps:

[3] https://opencv.org/

[4] The dataset is free available at http://doi.org/10.5281/zenodo.4757865

1. A web page was prepared for the competition. This web was used to provide information, to register the participants, to deliver the data, and for results submission.
2. The training set was provided to the participants at the beginning of the competition, approximately two months in advance of the deadline of results submission.
3. The evaluation tool was available at the beginning of the competition.
4. One week before the deadline of the competition, the test set was made available to the entrants. This test set did not have the associated ground truth available.

The test set provided to the participants was merged with several thousands of page images for which there was no ground truth. The participants were not able to distinguish the actual test set from the other page images. This was done for two reasons, namely:

- To prevent participants to overfit their system on the actual test set.
- To disseminate the idea that this type of task has to be defined for large datasets.

The participants had to provide the results on the merged data set, but they were ranked according to the actual test set.

The participants had to use only the provided data for training their systems. External data was not allowed for training the systems. This was done to make it easier to discriminate among systems, and therefore the results should not depend on the amount of training data.

The participants did not receive any feedback about their results on the test set. Providing evaluation results while the competition was open may help the participants to fit their systems. Besides, several submissions per participant were allowed, but just the best one was used to rank the participants.

4.2 Evaluation of systems

The evaluation was performed using Intersection-over-Union (IoU), and systems were ranked based on their F-measure after matching output formula boxes to ground truth formula regions. IoU overlap threshold was 0.7. Any predicted bounding box that surpassed this threshold was considered as a true positive (TP), while those that did not were considered as false positives (FP). For a detection box to be considered as TP, it also had to be of the correct class. The GT bounding boxes that were not detected were considered false negatives (FN). Thus, the precision of the system was calculated by the formula: TP / (TP + FP), while the recall of the system was calculated by the formula: TP / (TP + FN).

Considering that detection models usually output various candidates for a given region, before matching the output formula boxes with the GT boxes, the

entrants were recommended to apply a bounding box suppression algorithm. In case that two or more bounding boxes overlapped in the submitted solution, a non-maximum suppression strategy was applied as a sanity check to avoid multiple true positives corresponding to the same GT box. This algorithm only allowed us to keep the bounding boxes with the highest confidence score for each of the classes while removing bounding boxes that had an overlap of more than 0.25 score of IoU with lower confidence. Care had to be given to not providing overlapped bounding boxes if the entrants wanted to avoid a non-maximum suppression strategy.

4.3 Dataset partition

The IBEM dataset was divided into two sets (training and test sets) and delivered in two phases (training and test phases) to allow for performing different types of experiments. A first training set was provided in the training phase, and a second small training set was provided in the evaluation phase. The test set was released in the test phase.

First, the 600 documents contained in the dataset were shuffled at the document level. The set of documents prepared for the training phase was created by choosing the first 500 documents. The set of 100 documents prepared for the evaluation phase were distributed as follows:

1. the first 50 documents were used for testing (Ts10).

Then, the remaining 50 documents were shuffled at the page level.

2. 50% (329 pages) of these images could be used for training (Tr10);
3. 50% (329 pages) of these images were used for testing (Ts11).

Note that 1 was used for performing a task-independent evaluation and 3 was used for performing a task-dependent evaluation (about 25 documents).

The available data for the training datasets consisted of:

1. The original images of all the training pages.
2. A text file per training page, containing the corresponding ground truth.

The goal of the competition was to obtain the best mathematical expression detection rate on the Ts10 and Ts11 datasets.

5 Participant systems

More than 47 participants registered in the competition but in the end, only 14 participants submitted some results. The list of participants that submitted some result are listed in chronological order as they registered to the competition:

1. Lenovo Ocean from Lenovo Research, China (Lenovo).
2. HW-L from Huazhong University of Science and Technology, China (HW-L).

3. PKU Founder Mathematical Formula Detection from State Key Laboratory of Digital Publishing Technology, China (PKUF-MFD).
4. PKU Study Group from Peking University, China (PKUSG).
5. Artificial Intelligence Center of Institute of New Technology from Wuhan Tianyu Information Industry Co., Ltd., China (TYAI).
6. TAL Education Group, China(TAL).
7. SCUT-Deep Learning and Vision Computing Lab from Netease Corporation, China (DLVCLab).
8. Visual Computing Group from Ping An Property & Casualty Insurance Company of China Ltd., China (PAPCIC).
9. Autohome Intelligence Group from Autohome Inc., China (AIG).
10. Vast Horizon from Institute of Automation, Chinese Academy of Sciences, China (VH).
11. Shanghai Pudong Development Bank, China (SPDBLab).
12. University of Gunma, Japan (Komachi).
13. Augmented Vision from German Research Center for Artificial Intelligence, Germany (AV-DFKI).
14. University of Information Technology, China (UIT).

All participants used CNN for their solutions by adapting pre-trained models and/or architectures in most cases. The winner system was based on the Generalized Focal Loss (GLF) [4] since the scale variation is huge in this task between the embedded formula and the displayed formula. GFL can well eliminate the imbalance issue of positive/negative sampling on large or small objects. They adopted Ranger as an optimizer. Several GFL models were assembled via Weighted Box Fusion (WBF) [9].

6 Results

Results obtained in the competition are shown in Table 2 sorted according to the F1 score. Note that the best results were obtained by the PAPCIC group that was able to keep good results for embedded formulas. This was a key difference since detecting displayed MEs was easily solved for most of the participants. Note also that the PAPCIC system was able to get better results in the task-dependent scenario for embedded MEs than in the task-independent scenario. It is important to remark that the context can be relevant to detect embedded MEs. Consequently, PAPCIC was nominated as the winner of the competition.

Table 2. Results obtained in the competition. The *E, I, S* characters of the *Type* column stand for *Embedded, Isolated*, and whole *System*, respectively. For each type of evaluation, the table shows the F1 score with the corresponding precision and recall scores enclosed in parentheses.

Group ID	Type	F1 score (Ts10 + Ts11)	F1 score Task dependent (Ts11)	F1 score Task independent (Ts10)
PAPCIC	E	94.79 (94.89 , 94.69)	95.11 (95.11 , 95.11)	94.64 (94.79 , 94.50)
	I	98.76 (98.25 , 99.28)	98.70 (98.34 , 99.07)	98.79 (98.21 , 99.37)
	S	**95.47** (95.47 , 95.47)	**95.68** (95.62 , 95.73)	**95.37** (95.40 , 95.35)
Lenovo	E	94.29 (95.36 , 93.25)	93.98 (95.41 , 92.60)	94.44 (95.34 , 93.56)
	I	98.19 (98.26 , 98.12)	97.85 (98.04 , 97.67)	98.33 (98.35 , 98.31)
	S	**94.96** (95.86 , 94.08)	**94.60** (95.84 , 93.40)	**95.13** (95.87 , 94.39)
DLVCLab	E	93.79 (94.54 , 93.05)	93.88 (94.70 , 93.07)	93.75 (94.46 , 93.04)
	I	98.54 (98.19 , 98.89)	98.61 (98.33 , 98.88)	98.51 (98.13 , 98.90)
	S	**94.60** (95.17 , 94.04)	**94.64** (95.29 , 93.99)	**94.59** (95.12 , 94.07)
TYAI	E	93.39 (94.43 , 92.38)	93.94 (95.05 , 92.86)	93.13 (94.13 , 92.15)
	I	98.55 (98.19 , 98.92)	98.42 (98.15 , 98.70)	98.61 (98.21 , 99.02)
	S	**94.28** (95.08 , 93.49)	**94.66** (95.55 , 93.79)	**94.1** (94.86 , 93.35)
SPDBLab	E	92.80 (93.25 , 92.36)	92.14 (92.83 , 91.46)	93.12 (93.45 , 92.79)
	I	98.06 (98.06 , 98.06)	97.76 (97.85 , 97.67)	98.19 (98.15 , 98.23)
	S	**93.70** (94.08 , 93.33)	**93.03** (93.63 , 92.44)	**94.01** (94.28 , 93.74)
YoudaoAI	E	92.73 (93.57 , 91.91)	92.71 (93.95 , 91.51)	92.74 (93.39 , 92.10)
	I	98.34 (97.66 , 99.03)	98.38 (97.97 , 98.79)	98.32 (97.52 , 99.13)
	S	**93.70** (94.28 , 93.12)	**93.63** (94.61 , 92.67)	**93.74** (94.14 , 93.34)
PKUF-MFD	E	91.94 (92.18 , 91.70)	92.32 (92.93 , 91.72)	91.76 (91.82 , 91.69)
	I	96.56 (96.88 , 96.24)	96.87 (97.28 , 96.46)	96.43 (96.72 , 96.15)
	S	**92.72** (92.98 , 92.47)	**93.04** (93.62 , 92.47)	**92.57** (92.68 , 92.47)
HW-L	E	90.53 (91.55 , 89.53)	90.57 (91.82 , 89.35)	90.51 (91.42 , 89.61)
	I	98.94 (98.81 , 99.06)	98.61 (98.33 , 98.88)	99.08 (99.02 , 99.13)
	S	**91.97** (92.81 , 91.15)	**91.86** (92.88 , 90.86)	**92.02** (92.77 , 91.28)
Komachi	E	90.39 (90.92 , 89.86)	89.69 (90.43 , 88.97)	90.72 (91.16 , 90.28)
	I	98.57 (98.27 , 98.87)	98.6 (98.69 , 98.51)	98.55 (98.09 , 99.02)
	S	**91.79** (92.19 , 91.39)	**91.11** (91.75 , 90.48)	**92.10** (92.39 , 91.81)
AIG	E	89.71 (90.18 , 89.25)	89.19 (89.96 , 88.44)	89.95 (90.28 , 89.63)
	I	95.95 (99.00 , 93.09)	96.07 (99.01 , 93.30)	95.90 (99.00 , 93.00)
	S	**90.75** (91.61 , 89.90)	**90.26** (91.34 , 89.21)	**90.97** (91.74 , 90.22)
PKUSG	E	89.10 (90.26 , 87.97)	88.59 (90.23 , 87.00)	89.34 (90.27 , 88.42)
	I	97.96 (97.85 , 98.06)	97.94 (98.31 , 97.58)	97.96 (97.66 , 98.27)
	S	**90.62** (91.58 , 89.68)	**90.09** (91.54 , 88.68)	**90.87** (91.60 , 90.15)
TAL	E	87.87 (88.56 , 87.20)	88.51 (89.12 , 87.92)	87.57 (88.29 , 86.85)
	I	96.85 (96.31 , 97.40)	97.09 (96.33 , 97.86)	96.75 (96.30 , 97.21)
	S	**89.42** (89.91 , 88.93)	**89.89** (90.29 , 89.49)	**89.19** (89.73 , 88.67)
UIT	E	86.04 (85.60 , 86.49)	85.64 (85.62 , 85.65)	86.23 (85.58 , 86.89)
	I	97.05 (95.12 , 99.06)	98.11 (97.08 , 99.16)	96.60 (94.31 , 99.02)
	S	**87.94** (87.26 , 88.63)	**87.63** (87.47 , 87.80)	**88.08** (87.16 , 89.02)
AV-DFKI	E	85.35 (87.37 , 83.41)	84.75 (86.96 , 82.66)	85.63 (87.57 , 83.77)
	I	97.48 (97.12 , 97.84)	97.40 (97.30 , 97.49)	97.52 (97.04 , 97.99)
	S	**87.45** (89.10 , 85.87)	**86.80** (88.67 , 85.01)	**87.76** (89.30 , 86.27)
VH	E	84.25 (83.49 , 85.02)	84.39 (83.76 , 85.02)	84.18 (83.37 , 85.01)
	I	98.59 (98.51 , 98.67)	98.51 (98.42 , 98.60)	98.62 (98.55 , 98.70)
	S	**86.67** (86.01 , 87.34)	**86.61** (86.06 , 87.18)	**86.70** (85.99 , 87.41)

Figure 3 shows the differences obtained by the winner of the competition and the second competitor. The differences are remarked with red circles. We can observe that the significant differences were due to the detection of embedded MEs. Note that this type of expression could be better detected if the context was taken into account.

It is also interesting to remark a key difference between the systems that were ranked in first and second positions. The PAPCIC system was able to better distinguish multi-row expressions than the Lenovo system, as Figure 4 shows. As we mentioned in Section 3, we did not include any change in the GT when dealing with multi-row ME, that is, if the authors decided to write the ME in several consecutive math environments rather than in just one math environment in the LaTeX source, we left them in that way. In the example that is shown in Figure 4, the PAPCIC system got 100% F1 score while the Lenovo system got 83.72%.

Figure 5 shows an example in the PAPCIC system that did not get good results. We can observe that most of the errors were produced in the embedded MEs for which the layout was complicated and consequently the context was not helpful.

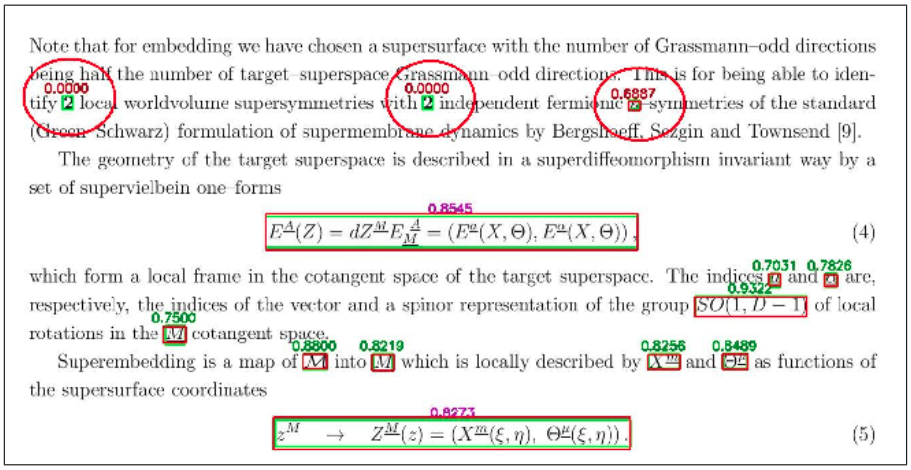
7 Conclusions

This paper has introduced a competition for MEs detection in STEM documents. The competition was raised as an image-based information retrieval problem. Several participants sent their results that were tested on a dataset that was not seen in the training phase. We observed that excellent results can be achieved in terms of F1 score, but there is still room for improvement especially in embedded formulas.

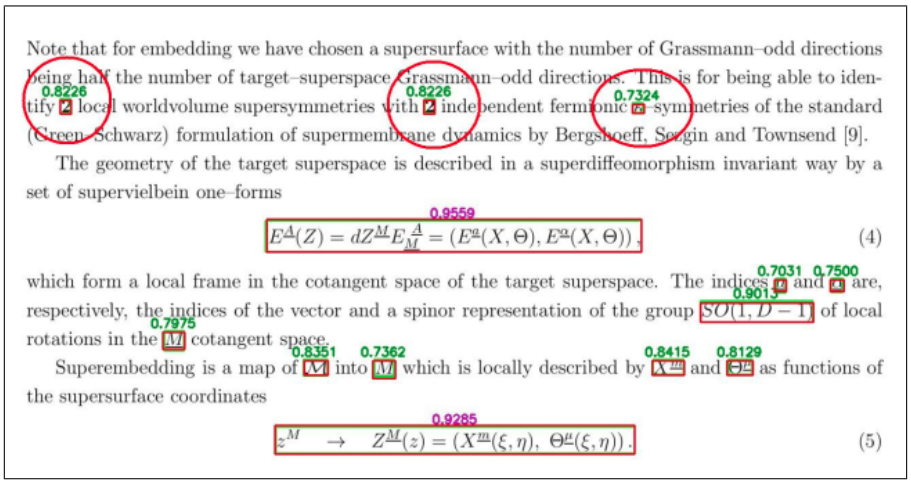
For future work we plan to extend this competition in several dimensions: i) enlarging the dataset with more STEM documents; ii) making possible the searching by combining visual information and text information around the MEs; and iii) making possible the searching by using semantic information. This last feature means that when looking for MEs, dummy variables should be ignored.

Acknowledgements

This work has been partially supported by the Ministerio de Ciencia y Tecnología under the grant TIN2017-91452-EXP (IBEM) and by the Generalitat Valenciana under the grant PROMETEO/2019/121 (DeepPattern).

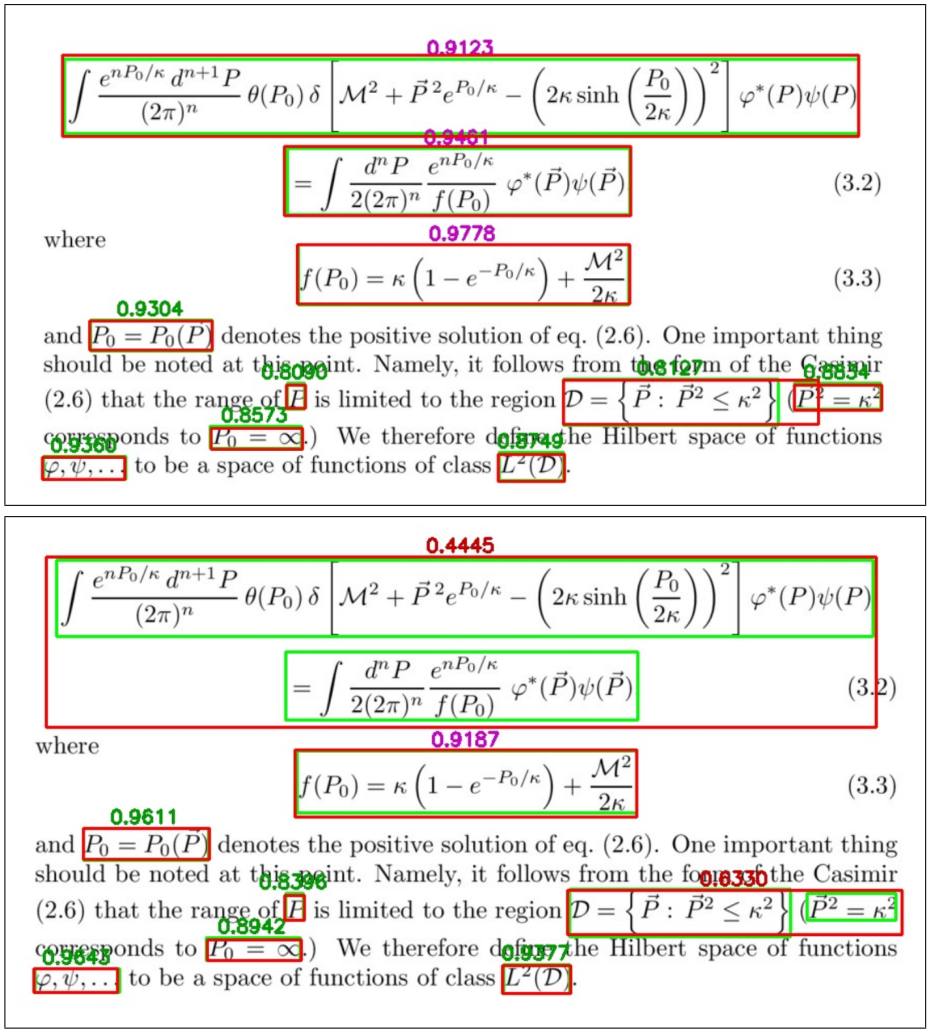


(a) Projected detections of Lenovo system.



(b) Projected detections of PAPCIC system.

Fig. 3. Part of a page from the Ts11 set, shown as a comparison between the two best systems that have been submitted, in which Lenovo system did not detect the first two embedded MEs and partially detected (IoU < 0.7) the third embedded ME encircled in red. For this example the PAPCIC system outperformed the other systems, having reached a perfect F1 score.



$$\int \frac{e^{nP_0/\kappa}\, d^{n+1}P}{(2\pi)^n}\, \theta(P_0)\, \delta\left[\mathcal{M}^2 + \vec{P}^{\,2} e^{P_0/\kappa} - \left(2\kappa \sinh\left(\frac{P_0}{2\kappa}\right)\right)^2\right] \varphi^*(P)\psi(P)$$

$$= \int \frac{d^n P}{2(2\pi)^n} \frac{e^{nP_0/\kappa}}{f(P_0)}\, \varphi^*(\vec{P})\psi(\vec{P}) \tag{3.2}$$

where

$$f(P_0) = \kappa\left(1 - e^{-P_0/\kappa}\right) + \frac{\mathcal{M}^2}{2\kappa} \tag{3.3}$$

and $P_0 = P_0(P)$ denotes the positive solution of eq. (2.6). One important thing should be noted at this point. Namely, it follows from the form of the Casimir (2.6) that the range of P is limited to the region $\mathcal{D} = \left\{\vec{P} : \vec{P}^2 \leq \kappa^2\right\}$ ($P^2 = \kappa^2$ corresponds to $P_0 = \infty$.) We therefore define the Hilbert space of functions $\varphi, \psi, \ldots$ to be a space of functions of class $L^2(\mathcal{D})$.

$$\int \frac{e^{nP_0/\kappa}\, d^{n+1}P}{(2\pi)^n}\, \theta(P_0)\, \delta\left[\mathcal{M}^2 + \vec{P}^{\,2} e^{P_0/\kappa} - \left(2\kappa \sinh\left(\frac{P_0}{2\kappa}\right)\right)^2\right] \varphi^*(P)\psi(P)$$

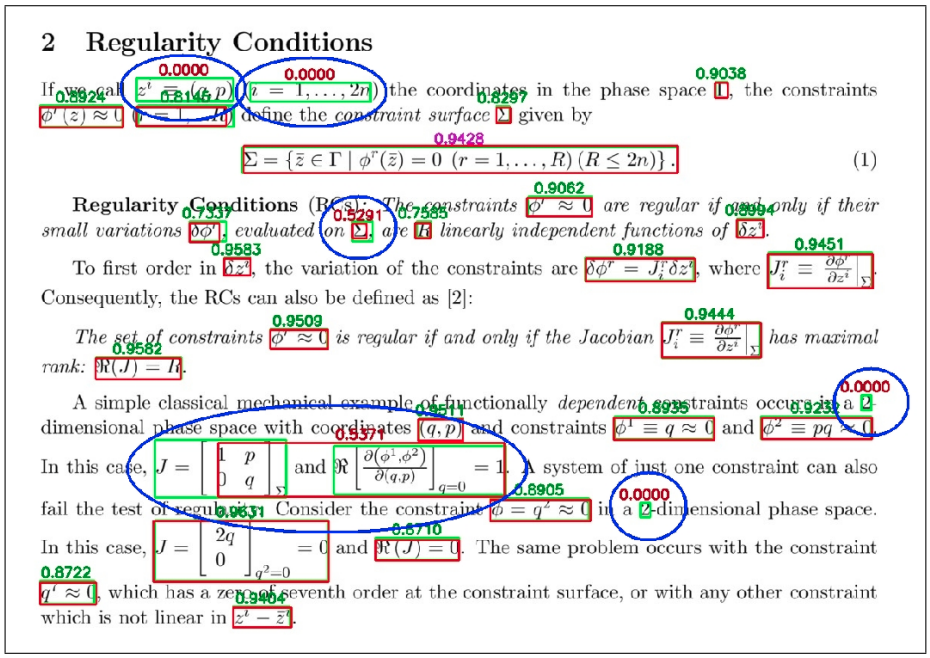
$$= \int \frac{d^n P}{2(2\pi)^n} \frac{e^{nP_0/\kappa}}{f(P_0)}\, \varphi^*(\vec{P})\psi(\vec{P}) \tag{3.2}$$

where

$$f(P_0) = \kappa\left(1 - e^{-P_0/\kappa}\right) + \frac{\mathcal{M}^2}{2\kappa} \tag{3.3}$$

and $P_0 = P_0(P)$ denotes the positive solution of eq. (2.6). One important thing should be noted at this point. Namely, it follows from the form of the Casimir (2.6) that the range of P is limited to the region $\mathcal{D} = \left\{\vec{P} : \vec{P}^2 \leq \kappa^2\right\}$ ($P^2 = \kappa^2$ corresponds to $P_0 = \infty$.) We therefore define the Hilbert space of functions $\varphi, \psi, \ldots$ to be a space of functions of class $L^2(\mathcal{D})$.

Fig. 4. Differences between multi-row ME located by the PAPCIC system (top) and the Lenovo system (bottom).



2 Regularity Conditions

If we call $z^t \equiv (q,p)$ $(t = 1,\dots,2n)$ the coordinates in the phase space Γ, the constraints $\phi^r(z) \approx 0$ $(r = 1,\dots,R)$ define the *constraint surface* Σ given by

$$\Sigma = \{\bar{z} \in \Gamma \mid \phi^r(\bar{z}) = 0 \ (r = 1,\dots,R)\,(R \le 2n)\}\,. \tag{1}$$

Regularity Conditions (RCs): *The constraints $\phi^r \approx 0$ are regular if and only if their small variations $\delta\phi^r$, evaluated on Σ, are R linearly independent functions of δz^t.*

To first order in δz^t, the variation of the constraints are $\delta\phi^r = J_i^r \delta z^i$, where $J_i^r \equiv \left.\frac{\partial \phi^r}{\partial z^t}\right|_\Sigma$. Consequently, the RCs can also be defined as [2]:

The set of constraints $\phi^r \approx 0$ is regular if and only if the Jacobian $J_i^r \equiv \left.\frac{\partial \phi^r}{\partial z^t}\right|_\Sigma$ has maximal rank: $\Re(J) = R$.

A simple classical mechanical example of functionally *dependent* constraints occurs in a 2-dimensional phase space with coordinates (q,p) and constraints $\phi^1 \equiv q \approx 0$ and $\phi^2 \equiv pq \approx 0$. In this case, $J = \left[\begin{array}{cc} 1 & p \\ 0 & q \end{array}\right]_\Sigma$ and $\Re\left[\frac{\partial(\phi^1,\phi^2)}{\partial(q,p)}\right]_{q=0} = 1$. A system of just one constraint can also fail the test of regularity. Consider the constraint $\phi = q^2 \approx 0$ in a 2-dimensional phase space. In this case, $J = \left[\begin{array}{c} 2q \\ 0 \end{array}\right]_{q^2=0} = 0$ and $\Re(J) = 0$. The same problem occurs with the constraint $q^t \approx 0$, which has a zero of seventh order at the constraint surface, or with any other constraint which is not linear in $z^t - \bar{z}^t$.

Fig. 5. Part of a page from the Ts10 set, in which the PAPCIC system did not detect four embedded MEs, partially detects (IoU < 0.7) one embedded ME and incorrectly joins two adjacent embedded MEs, encircled in blue. For this example, the system ranks as second to last with an F1 score of 87.67, while the best F1 score obtained is 94.74.

References

1. Deng, Y., Kanervisto, A., Rush, A.M.: What you get is what you see: A visual markup decompiler. ArXiv **abs/1609.04938** (2016)
2. Gehrke, J., Ginsparg, P., Kleinberg, J.: Overview of the 2003 kdd cup. SIGKDD Explor. Newsl. (2), 149–151 (Dec 2003)
3. Heiko Oberdiek: The zref package, https://osl.ugr.es/CTAN/macros/latex/contrib/zref/zref.pdf
4. Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J., Yang, J.: Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection (2020)
5. Mahdavi, M., Zanibbi, R., MouchÃšre, H., Viard-Gaudin, C., Garain, U.: ICDAR 2019 CROHME + TFD: Competition on recognition of handwritten mathematical expressions and typeset formula detection. In: International Conference on Document Analysis and Recognition (2019)
6. Ohyama, W., Suzuki, M., Uchida, S.: Detecting mathematical expressions in scientific document images using a u-net trained on a diverse dataset. IEEE Access **7**, 144030–144042 (2019)
7. Phillips, I.: Methodologies for using UW databases for OCR and image understanding systems. In: Proc. SPIE, Document Recognition V. vol. 3305, pp. 112–127 (1998)
8. Pizzini, K., Bonzini, P., Meyering, J., Gordon, A.: GNUsed, a stream editor, https://www.gnu.org/software/sed/manual/sed.pdf
9. Solovyev, R., Wang, W., Gabruseva, T.: Weighted boxes fusion: Ensembling boxes from different object detection models. Image and Vision Computing **107**, 104117 (2021). https://doi.org/https://doi.org/10.1016/j.imavis.2021.104117, https://www.sciencedirect.com/science/article/pii/S0262885621000226
10. Suzuki, M., Uchida, S., Nomura, A.: A ground-truthed mathematical character and symbol image database. In: Proc. 8th International Conference on Document Analysis and Recognition (ICDAR'05). pp. 675–679 (2005)
11. Zanibbi, R., Blostein, D.: Recognition and retrieval of mathematical expressions. International Journal on Document Analysis and Recognition **14**, 331–357 (2011)
12. Zanibbi, R., Oard, D.W., Agarwal, A., Mansouri, B.: Overview of arqmath 2020: Clef lab on answer retrieval for questions on math. In: Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Joho, H., Lioma, C., Eickhoff, C., Névéol, A., Cappellato, L., Ferro, N. (eds.) Experimental IR Meets Multilinguality, Multimodality, and Interaction. pp. 169–193. Springer International Publishing, Cham (2020)