Semi-Automated Correction Tools for Mathematics-Based Exercises in MOOC Environments

Alberto Corbí and Daniel Burgos

Universidad Internacional de La Rioja

Abstract — **Massive Open Online Courses (MOOCs) allow the participation of hundreds of students who are interested in a wide range of areas. Given the huge attainable enrollment rate, it is almost impossible to suggest complex homework to students and have it carefully corrected and reviewed by a tutor or assistant professor. In this paper, we present a software framework that aims at assisting teachers in MOOCs during correction tasks related to exercises in mathematics and topics with some degree of mathematical content. In this spirit, our proposal might suit not only maths, but also physics and technical subjects. As a test experience, we apply it to 300+ physics homework bulletins from 80+ students. Results show our solution can prove very useful in guiding assistant teachers during correction shifts and is able to mitigate the time devoted to this type of activities.**

Keyword s— assignments, semiautomated correction, maths, physics, framework.

I. INTRODUCTION

MOOCs and online campuses nowadays represent an observable reality when it comes to self-education [5]. Together with *OpenCourseWare* platforms, they are definitively impacting our current TEL scene. Even in MOOC environments, students are usually required to carry out some homework. Nevertheless, these homework bulletins are hardly ever supervised by a tutor or a teacher. Quite the opposite, the students themselves are required to self-correct and self-assess their exercises based on correction grids, templates and answer keys. Peer reviewing also takes place, as we will discuss in section II. Fully automated quizzes are also commonly displayed and correction is normally done by the MOOC and/or e-learning platform.

Technical documents from the STEM fields (Science, Technology, Engineering and Mathematics) increase document richness with many sorts of structured objects: mathematical and chemical formulae, diagrams, tables and relations, etc. These additions usually carry essential information that complements the texts the student has to read. At first sight, homework assignments related to these disciplines are good candidates for automated correction processes. However, many teachers are interested not only in the accuracy of the result but also in the correctness of the resolution process, which might turn out to be as important as –or sometimes even more important than– the final outcome itself. Corrections performed by a human (a teacher/assistant) can also add value to the teacher's view on how his/her students learn and progress. The teacher's feedback on a correction sheet always entails a unique opportunity to improve the learner's knowledge and build a more robust awareness on the matter they are currently working on.

Exercises in physics deepen this reviewing philosophy and student-teacher interaction. Keeping an organized and coherent resolution flow is as relevant to the understanding of the underlying physical phenomena as the final output itself.

Besides, in physics, results can belong to a broad spectrum of mathematical natures and entities, ranging from simple and isolated numbers or scalars (e = 2.7182), vectors ($\vec{v} = 3\vec{i} + \vec{j}$), signed quantities (-k) and physical units (3.3 kΩ), to name a few that might appear on a basic physics course. In addition, slightly different numbers, notations and/or symbols can represent exactly the same correct result and account for the same reality. For instance, $\left|\vec{B}\right| = 5\,\text{N/Am}$ and $\|\mathbf{B}\| = 4.99\ \text{T}$ can both be labeled as correct and the student should receive a positive score/comment. If such minor discrepancies could be detected, an automated system might be able to send back an explicit recommendation as [26], for example, does. In the same sense, and as a last example, all of the following expressions have the exact same meaning: partial differentiation of function *f* with respect to an independent variable *x*:

$$f_x, f'_x, d_x f, \partial_x f, \nabla_x f, \partial f/\partial x, D_x f$$

Finally, students attending physics courses in online institutions and/or MOOCS come from very different backgrounds and behavior is easily altered over time, as described by [1]. The *human touch* in the reviewing process has always proven to be the key to success, independently of the academic environment: online, formal, higher education, etc.

All this being said, in MOOC environments, the amount of homework bulletins to be reviewed, and the substantial tutoring effort that takes place if every exercise from every

student is manually revised, can reach disproportionate levels. One of the goals of our project is concerned with assisting teachers during the correction phase. This target is achieved by pre-classifying student bulletins as ready to be *teacher-reviewed* or not. In the latter case, an automated message can be issued to the student, who can re-edit his/her own document before reissuing it to the teacher, for a second time. Of course, this assistant tool would heavily depend on the type of subject and content to be analyzed. In this paper, we focus on assisting teachers in online campuses and MOOCS when reviewing homework related to mathematical content.

II. Overview of the Current Status of MOOCs, Online Education and Student Assignment Management

MOOCs face nowadays a number of challenges: accreditation management, credit recognition, monetization implementation and content and methodology quality assurance. Among them, methodology quality becomes the foundation from which the other four are built. MOOCs are taking over the long-tradition role of Open Educational Resources. Some MOOCs also combine face-to-face strategies with online learning and even merge formal and informal settings. In addition, MOOCs highlight the current need for basic and specific competence acquisition, as a complement to the current courses, very much focused on personal interests and continuing education. They are also turning out the ultimate tool to fight against the lack of access to teaching resources (disadvantaged individuals, regions and countries).

MOOC platforms require support for teachers and tutors, based on their needs, skills, and teaching context. One of these has to do with grading essays and activities. Since MOOCs seeks the enrollment of hundreds or thousands of students, the evaluation becomes a real challenge. At present, some MOOCs rely on peer-assessment and counseling. Peer-to-peer seems significant and useful, so there is, at first sight, no need for a replacement. However, a complementary evaluation resource would be welcome by the educational community.

There are some approaches for automatic or semi-automatic assessment, like ontology networks [23], where the conceptualization of the domain model becomes the cornerstone to categorize and shapes the results properly. Another strategy involves the temporal hiring of additional teachers as *graders*, so they can act as complement to those professors officially assigned to the course. In addition, a detailed comment and assessment on the submitted final activity might not be compulsory, as long as the learner does not require a formal accreditation. This strategy scales down the number of assessments to those learners who actually send a formal/official request. At Universidad Internacional de La Rioja (UNIR), there is a prototype implemented, and under testing phase: A4Learning [30]. This tool is integrated into the Sakai LMS, and retrieves behavioral and academic information from users, so that they can be compared with previous records. Out of this comparison, the tool makes an estimate on every student on how his/her progress will be, based on similar profiles. In doing so, the professor gets a detailed analysis of every learner, 1 by 1, and clustered by similarity. With A4Learning, the teacher can analyze the student current status, anticipate potential academic future, and react in consequence. There is another early prototype, AppMOOC, which will retrieve basic requirements to grade activities, so that, when the professor gets an essay, a previous checking mechanism guarantees that the work fulfills these minimum information and/or structure. These two prototypes, A4Learning and AppMOOC, will be implemented along the next academic year at a larger scale, with the clear objective of supporting teachers on their functions as evaluators and feedback providers, big mid-size and large-size groups of learners, worldwide. The research work described in this paper is in intimate relation with the aforementioned projects.

III. Towards an Automated Helper System for Maths and Technical Student Homework Preliminary Screening

We have designed a special workflow and protocol that automatically analyses student assignments and checks whether they contain coherent mathematical information related to specific fields. This set of tools also takes into account equivalent expressions, exemplified in section I.

In order to check for this coherence, simple –but also highly configurable and easily editable– content-checking rules designed by the teacher are submitted to the correction engine. Then, for every exercise in the student digital notebook, mathematical expressions are semantically compared with the correction template submitted by the teacher. A more detailed review of the practical implementation is tackled below.

Of course, designing such a protocol is no easy task and has required working with state-of-the-art mathematical language-processing techniques and mathematics representation standards, also reviewed below.

A. State-of-the-Art Language Processing in Mathematics

Despite the fact that linguistic analysis of scientific documents is currently seen as an interesting line of research, the current work in the field is still limited. Mathematical literature represents a rather isolated linguistic niche embodying its own challenges. We can identify a significant contrast between this linguistic realm and, for instance, the domain of medical/healthcare research publications that have been studied by many scientific groups in recent years. Two of the current main issues that make mathematical texts challenging to work with are:

- Natural language –expressing complex symbolism– and mathematical representation are usually mixed and hosted in the same document.
- Almost a complete absence of accurately labeled linguistic compilations.

Indeed, state-of-the-art analyses largely try to bypass these problems by restricting their scope to *well-formed* sections of mathematical text and reports, as in the *controlled* approach reviewed below.

The first challenge of the recognition process is the recovery of the so-called *layout tree* [9] of the mathematical

expression. The next step involves creating operator trees. These trees are data structures that hold the logical relationships within an equation, as opposed to its horizontal and vertical links. The structure of the mathematical expression can then be made computationally transparent, which is necessary for any practical application involving a mathematics recognition process, like the one we are introducing in this paper. The layout tree also carries a burden of uncertainty in its correctness, which adds to the difficulty of establishing the expression's logical structure.

A holistic and detailed analysis of the processes of extracting and retrieving mathematical expressions and mathematics recognition has already been carried out by [28].

We will now review some lines of enquiry that have recently attracted interest in the research community around math semantics and language processing.

1) Controlled Natural Language

In this approach, a restrained natural language for mathematics is incrementally built [12]. With it, we are then capable of supporting a sufficient subset of natural language elements that would allow an author to write math expressions in a simple way but also be limited enough to allow unambiguous interpretation. Its primary goal is building formalized libraries of mathematical content, focusing on establishing pipelines over a narrow subset of language. Next, a systematic and careful widening takes place. Current projects implementing this view are:

- FMathL [21] described in `mat.univie.ac.at/~neum/FMathL.html`
- MathLang [13]
- MathNat [12]
- Naproche [3], [6], available at `naproche.net`

2) Natural Mathematical Discourse

The opposite 2of the controlled approach is to try to model the original language of real scientific documents [6, 29]. Consistent work in the area has been developed by [27] and [11], as well as by [4]. The corpus used for this work is based on the arXMLiv archiving project of scientific documents [24]. arXMLiv is hosted at the Cornell arXiv (`arxiv.org`) which contains one of the largest collections of scientific literature on the planet. Unfortunately, its texts are in the TeX/LaTeX format, which makes it rather useless for knowledge analysis engines, even though LaTeX can be considered a *de facto* global standard of typesetting. The goal of the project described in [10] is to translate all these documents to a common and agreed XML scheme, which can then serve as a basis for revealing math-related semantics.

B. Computer Representation of Mathematical Content with LaTeXML

LaTeXML [7] uses a context-free grammar to establish the logical structure of a document with mathematical content. It can then be exported to Content MathML and OpenMath [2]. Content MathML (also referred as MathML v3 from the W3C consortium and described in `w3c.org/TR/MathML3`) uses

just a few attributes and focuses on the meaning of the expression rather than its graphical layout. The `<apply>` element, for instance, represents the application of a function. Its first child element is the function itself and its operands and/or parameters are the remaining child elements.

Thanks to Content MathML and Open Math, digital libraries can be transformed into a more useful XML representation and be made more compliant with a mathematical knowledge-management approach. Two large-scale examples are arXMLiv and EuDML [22]. Only the first of those examples uses LaTeXML. The main challenges in this conversion step, in the case of arXMLiv, come from the fact that it is poorly knowledge-based, with minor exceptions in the form of clues provided via some infrequent and almost random in-line LaTeX annotations. It is then mandatory to infer additional semantics on all document levels. Fortunately, LaTeXML has proven to be extremely efficient at this task.

Consider the example in Fig. 1. There we have the standard mathematical notation –a simple equation of the form $f(x) = y$–, its Content MathML representation and, finally, the terms we extracted for indexing. Any mathematical construct can be represented in a similar way.
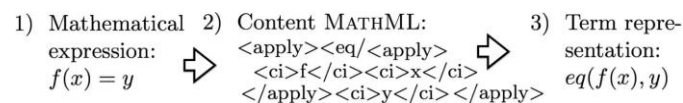

1) Mathematical expression: $f(x) = y$ ⇨ 2) Content MATHML: `<apply><eq/<apply> <ci>f</ci><ci>x</ci> </apply><ci>y</ci></apply>` ⇨ 3) Term representation: $eq(f(x), y)$

Fig. 1. From plain mathematical expressions to Content MathML.

LaTeXML also defines a conversion process and a set of tools that allow any plain LaTeX document to be translated [7]. LaTeXML can even work in daemon mode, which allows the deployment of server-centric conversion platforms [8] like the well-known *ltxMojo*, available at `latexml.mathweb.org`.

Once a mathematical text has been retyped as LaTeXML, search queries can take place. This topic is discussed in the following section.

C. The MathWebSearch Project

MathWebSearch [17], developed at the KWARC group (`kwarc.info`), processes XML-based content mathematics. Currently, the system supports MathML, OpenMath and LaTeXML (and any other document type that has been appropriately converted). It operates by computing an index term for each of the mathematical elements of a given XML document. Queries on this index are also expressed in a XML schema, reviewed below.

The MathWebSearch engine is used in our framework to analyse student-submitted mathematics assignments. On one hand, each student document is converted to Content MathML and indexed. On the other hand, a teacher's set of well-organized binary tests is coded as a variant of Content MathML –MathMLQ–. If all tests deliver a positive result, the assignment is flaged as *to be reviewed* by the teacher.

Finally, as MathWebSearch operates with terms, heuristics and semantics, it can understand a wider range of similar mathematical expressions. This ensures that the issues described in the introduction will hardly ever take place. Our engine is very tolerant to small variations of the same

mathematical expression. In other words, we are able to understand that $\sqrt{x}$ and $x^{1/2}$ have the same mathematical meaning and discern that 4.5 kJ is different from 4.5 Kj (the Joule energy unit in physics must always be capitalized, while the *kilo-* multiple should remain in lowercase). In this manner, the student is free to express him/herself with mathematical and syntactical independence. At the same time, the teacher is also able to demand exquisite precision, if so desired.

Algorithm 1 Example of a MATHMLQ query.

```
<mws:query
xmlns:mws="http://www.mathweb.org/mws/ns"
xmlns:m="http://www.w3.org/1998/Math/MathML"
limitmin="0" answsize="30" totalreq="yes"
output="json">
 <mws:expr>
  <m:apply>
   <m:csymbol cd="ambiguous">
    superscript
   </m:csymbol>
   <mws:qvar>q</mws:qvar>
   <m:cn type="integer">2</m:cn>
  </m:apply>
 </mws:expr>
</mws:query>
```

D. The MathWebSearch Query Language

MathWebSearch makes use of a content-oriented query language called MathMLQ. It is XML-based rather than being a genuine query language by itself. More detailed information on the syntax can be found in [18]. An example of application can be read in algorithm 1. The query described there is able to identify both the square of a function or a variable (x^2 or $[f(x)]^2$).

Apart from describing queries using the MathMLQ syntax just introduced, more simple instances can be expressed using the plain LaTeX math toolbox and syntax. This code can be then converted to MathMLQ. This conversion takes place with the tool `latexmlc`, presented in [16], which can also establish relations between LaTeX and a variety of office documents (WML from MS Word, ODT from Open Office, etc.) In this simplified LaTeX syntax, variables are labeled with the question mark symbol (`?`). For instance, the following expression:

```
latexmlc --address = latexml.mathweb.org/
convert --preload=mws.sty --whatsin=math -
whatsout = math --cmml 'literal:\sqrt{?c}^2'
```

Would produce the same XML output as the one displayed in algorithm 1.

E. Summary of Implementation

We now summarize the skeleton of our software implementation, which is graphically represented in Fig. 2. Students submit their homework in a variety of formats (Microsoft Office Word, OpenOffice, OpenDocument, Portable Document Format, LaTeX and LyX, etc.). Disciplines related to theoretical fields, such as mathematics, physics and computer science, almost exclusively use LaTeX. On the other hand, more applied fields of research, like life sciences, chemistry and engineering, usually typeset on the so-called *office suites*. Moreover, depending on the discipline, each institution has its own focus and teachers expect homework to be edited using a specific software instance.

For this reason, our system tries to, in the first phase, convert each document type to a unified LaTeX representation. This is not always possible due to technical reasons (converter segmentation fault, faulty output, etc.). Several third party tools (both open source and commercial) exist and operate with greater or lesser degrees of success. Writer to LaTeX (`writer2latex.sf.net`) and Word to LaTeX (`wordtolatex.com`) are some examples. LyX has the advantage of being able to perform a clean LaTeX export [14].

A better tool to translate between LaTeX and traditional office formats is the `latexmlc` introduced above, which has been developed in recent years by the KWARC group. Finally, the tool that has recently been attracting significant focus in the computer language research community is Pandoc, described in [20] and [19]. Pandoc can convert documents in markdown, HTML, LaTeX, MediaWiki markup, TWiki markup, Microsoft Word docx and EPUB (among others) to other formats, such as DocBook, Adobe InDesign, LaTeX, PDF and many others, through the application of external drivers written in the Lua computer language.

Anecdotally, recent efforts are even trying to directly translate mathematical handwritten expressions to LaTeX. A nice summary can be found in [25] and an example of such an application can be tested online thanks to Detexify [15], available at `detexify.kirelabs.org`.

As a next step, the LaTeX source is parsed and transformed to LaTeXML, which already contains the necessary knowledge companion information to be harvested by MathWebSearch. On the other side, the teacher pulls a list with *N* wildcard expressions to the classification platform. Finally, an instance of MathWebSearch performs these *N* searches on each homework document and screens which of them provides some degree of equivalence. Our platform is responsible for filtering teachers' templates and student homework in a coordinated fashion.
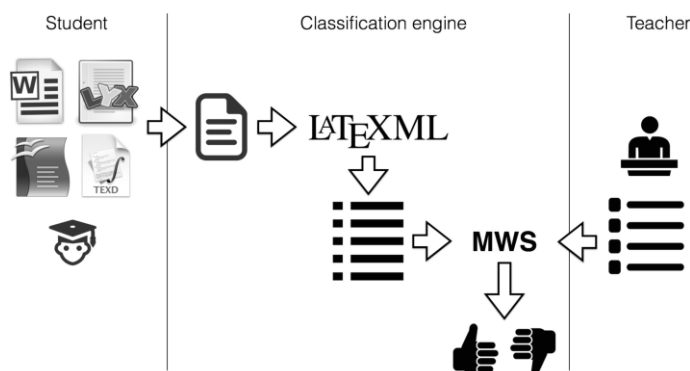

Fig. 2. Overview of our platform. Students submit their homework and a conversion process to LaTeXML takes place. On the other side, teachers feed the system templates with mandatory mathematical expressions.

IV. PRACTICAL EXPERIENCE WITH PHYSICS STUDENTS' HOMEWORK BULLETINS

As a proof of concept, we have carried out a practical experience with 300+ homework assignments from 50+ students enrolled in a basic Physics course in the degree of Computer Science at the School of Engineering at Universidad International de La Rioja (UNIR, ingenieria.unir.net).

We have configured our classification engine based on MathWebSearch together with teachers' templates in order to pre-distribute assignments, before they are finally delivered to the teacher/assistant for an in-depth (and manual) conventional correction phase.

A. Experimental Setup

The online campus platform deployed at UNIR is an instance of the Apereo Sakai CLE. Students submit their homework to this platform digitally, using the assignments tool. Usually, documents are formatted using Microsoft Word®, WML or OpenOffice ODT, though some students have used LaTeX or LyX for their submissions. A very small percentage of students submitted bulletins in other office suite formats, such as Apple Pages® or Microsoft PowerPoint®, which were easily translatable to WML or ODT.
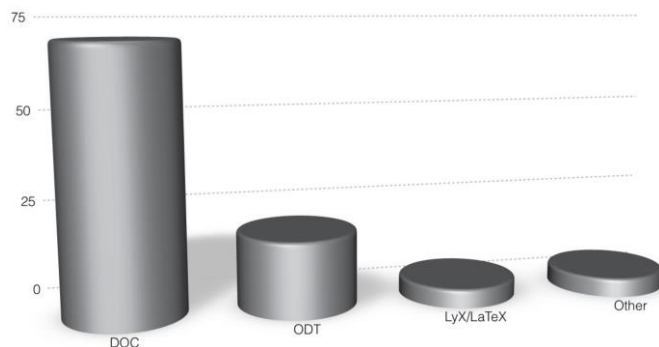

Fig. 3. Percentage of document types used by students.

The rate of conversion success to LaTeX and LaTeXML from this range of commonly available office suites is summarized in Table I. After running each of the conversion tools, further refinement can take place if the source office documents are pre- or post-manually processed.

TABLE I.
CONVERSION LEVEL OF SUCCESS FROM OFFICE DOCUMENTS SUBMITTED BY STUDENTS TO LATEX AND LATEXML

Source *office* format	% of success	after manual edition
WML (recent MS Word ver.)	75 %	90 %
.doc (legacy MS Word ver.)	70%	90%
ODT (OpenOffice)	80 %	95 %
Other office suites	50 %	70 %
LyX	100 %	100 %

The conversion tool most used in our setup, given its success ratio, was Pandoc, as described above. Fig. 4 shows a real example of the result of the conversion of a MS Word-submitted homework file to its LaTeX twin. PDF output (from LaTeX) is also shown as a proof of the fidelity of the file-translation process.
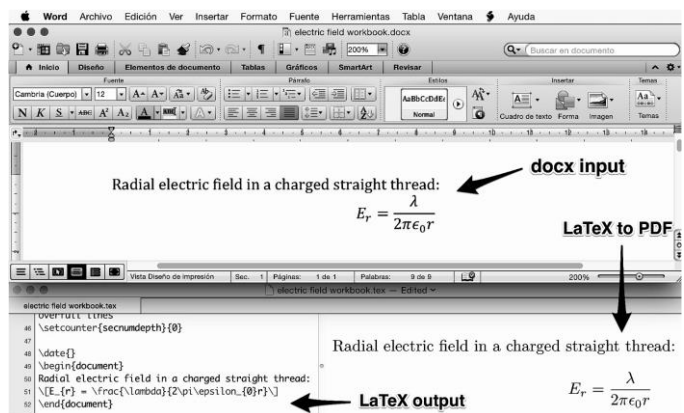

Fig. 4. Example of the conversion process performed with Pandoc.

Fig. 5 shows the ratio of success in the process of translating to LaTeX, of some of the file-conversion tools that are mentioned above and were used in this project.
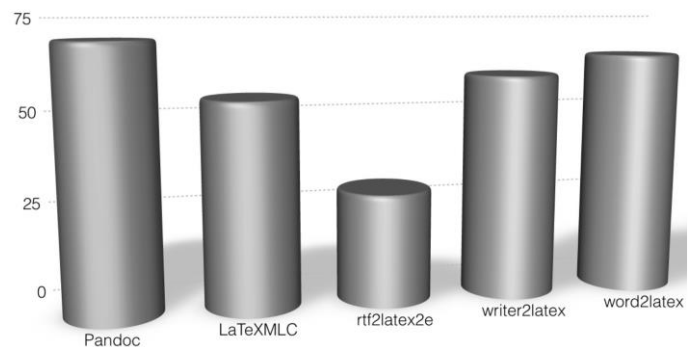

Fig. 5. Rates of success for some of the converter tools (to LaTeX).

B. Methodology

The physics course mentioned above, as it is part of the Computer Engineering degree's curriculum, is mainly based on areas related to electromagnetism. Most required homework exercises should include at least some of the mathematical expressions appearing in table II –depending on the specific topic being studied– in order to be considered suitable for further analysis by the teacher and manually assigned a score. This mathematical content has been agreed with the academic staff. The corresponding set of simplified queries (introduced above) has also been defined and has been made available to the system.

In Table III, there is another example of how our implementation can also handle more complex formulae, for instance those related to quantum theory and thermodynamics, which could prove useful in a Physics MSc.

Our solution has been tested offline (no real feedback has been sent to students or teachers) with pre-existing homework bulletins from an already concluded semester. A batch process, similar to that described in Fig. 2, has been implemented and executed.

Besides taking into account specific mathematical content related to the topic *electromagnetism*, we have also established a special and separate realm devoted only to pure mathematical transversal correctness. This means that our solution can separately test for the exactitude of common mathematical statements, like the ones listed in Table IV

TABLE II.
SOME MATHEMATICAL EXPRESSIONS RELATED TO THE TOPIC
ELECTROMAGNETISM TO BE TESTED.

Electric field by an infinite plane (perpendicular)	$E_x = 2\pi k\sigma$
Electrostatic force (vector form)	$\vec{F} = k\dfrac{|q_1 q_2|}{r^2}\hat{r}$
Charge linear density	$dq = \lambda dx$
Capacitor equation	$C = \epsilon\dfrac{S}{d}$
Electric field by a charged ring	$E_x = \dfrac{kQx}{(x^2 + R^2)^{\frac{3}{2}}}$
Electric field by a charged circle	$E_x = 2\pi k\sigma\left(1 - \dfrac{x}{\sqrt{x^2 + R^2}}\right)$
Electric field by a charged thread	$E_y = k\lambda\dfrac{L}{y\sqrt{\frac{1}{4}L^2 + y^2}}$
Electric potential by a set of isolated charges	$V = \sum_i \dfrac{kq_i}{r_{i0}}$
Relation between potential and electric field	$dV(r) = -\mathbf{E}(r)\cdot dr$
Electric field outside a charged a spherical cavity (radial component)	$E_r = \dfrac{1}{4\pi\epsilon_0}\dfrac{Q}{r^2}$

TABLE III.
EXTENDED PHYSICS-RELATED EXAMPLES.

Heat diffusion law	$\dfrac{dQ_x}{dt} = -kA\dfrac{\partial T}{\partial x}$
Entropy	$S = \sum_i p_i \ln p_i$
Tunnel effect	$T = \dfrac{e^{-2\int_{x_1}^{x_2} dx\sqrt{\frac{2m}{\hbar^2}(V(x)-E)}}}{\left(1 + \frac{1}{4}e^{-2\int_{x_1}^{x_2} dx\sqrt{\frac{2m}{\hbar^2}(V(x)-E)}}\right)^2}$
Momentum by a photon	$p_f = \dfrac{h\nu}{c}$
Gyro dynamics	$\omega_p = \dfrac{MgD}{\mathbf{I}_s \omega_s}$
Charged particle in accelerator	$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right)$

TABLE IV.
TRANSVERSAL MATH EXPRESSIONS.

Two-component vector definition	$\vec{V} = (V_x, V_y)$
Vector sum	$\vec{V_1} + \vec{V_2} = (V_{1_x} + V_{2_x}, V_{1_y} + V_{2_y})$
Vector length	$|\vec{V}| = \sqrt{V_x^2 + V_y^2}$
Vector dot-product	$\vec{V_1}\cdot\vec{V_2} = V_{1_x}\cdot V_{2_x} + V_{1_y}\cdot V_{2_y}$
Scalar-vector product	$N\cdot\vec{V} = N\cdot(V_x, V_y) = (N\cdot V_x, N\cdot V_y)$

With this external test, our system allows teachers to filter bulletins based only on pure mathematical fidelity, ignoring topic-specific inaccuracies or errors

V. RESULTS AND DISCUSSION

After running a batch process with the 300+ homework bulletins and specific rule sets, results show that around 63% of the documents that could be safely converted to LaTeX satisfied the formulae template requirements (both for the topic electromagnetism and for the transversal one related to mathematics). Of these homework assignments, 78% were given a positive score by the teacher at the moment of the reviewing process. The remaining 22% of documents that were classified as incorrect, though encapsulating the required mathematical expressions, contained inaccuracies and/or were poorly developed by the student.

VI. CONCLUSIONS

Our simplified and relatively quick set-up proves that semi-automated correction processes may represent an acceptable compromise between the pure self-assessment approach –typically present in MOOCS and courses with a large enrolment rate– and the more conventional scenario in which the teacher manually reviews assignments for each student.

ACKNOWLEDGMENT

REFERENCES

[1] John M Aiken, Shih-Yin Lin, Scott S Douglas, Edwin F Greco, Brian D Thoms, Michael F Schatz, and Marcos D Caballero. The initial state of students taking an introductory physics mooc. arXiv preprint arXiv:1307.2533, 2013.

[2] Ron Ausbrooks, Stephen Buswell, David Carlisle, Giorgi Chavchanidze, Stéphane Dalmas, Stan Devitt, Angel Diaz, Sam Dooley, Roger Hunter, Patrick Ion, Michael Kohlhase, Azzeddine Lazrek, Paul Libbrecht, Bruce Miller, Robert Miner, Murray Sargent, Bruce Smith, Neil Soiffer, Robert Sutor, and Stephen Watt. Mathematical Markup Language (MathML) version 3.0. W3C Recommendation, World Wide Web Consortium, 2010.

[3] Marcos Cramer, Peter Koepke, and Bernhard Schröder. Parsing and disambiguation of symbolic mathematics in the Naproche system. Proceedings of the 18th Calculemus and 10th International Conference on Intelligent Computer Mathematics, 2011.

[4] J.H. Davenport, W.M. Farmer, F. Rabe, and J. Urban. Intelligent Computer Mathematics: 18th Symposium, Calculemus, and 10th International Conference on Artificial intelligence, 2011.

[5] Luis de-la Fuente-Valentín, Aurora Carrasco, Kinga Konya, and Daniel Burgos. Emerging technologies landscape on education. a review. International Journal of Interactive Multimedia and Artificial Intelligence, IJIMAI, 2(3), pp.55, 2013.

[6] Mohan Ganesalingam. The Language of Mathematics. PhD thesis, Cambridge University, 2009.

[7] Deyan Ginev. LaTeXML: A LATEX to XML converter, arXMLiv branch. Web Manual. 2011.

[8] Deyan Ginev. The LaTeXml daemon: Editable math for the collaborative web.

[9] Deyan Ginev. The structure of mathematical expressions. PhD thesis, Master Thesis, Jacobs University Bremen, 2011.

[10] Deyan Ginev, Constantin Jucovschi, Stefan Anca, Mihai Grigore, Catalin David, and Michael Kohlhase. An architecture for linguistic and semantic analysis on the arXMLiv corpus. Applications of Semantic Technologies Workshop, 2009.

[11] Helmut Horacek and Magdalena Wolska. Handling errors in mathematical formulas, Lecture Notes in Computer Science, 2006.

[12] Muhammad Humayoun and Christophe Raffalli. Mathnat: mathematical text in a controlled natural language. 9th International Conference on Mathematical Knowledge Management, 2010

[13] Fairouz Kamareddine. Motivations for mathlang, T-O-U Workshop, 2005.

[14] David Kastrup. Revisiting WYSIWYG paradigms for authoring LaTeX. Proceedings of the 2002 Annual Meeting, TUGboat, 2002.

[15] D. Kirsch. Detexify: Erkennung handgemalter LaTeX-symbole. PhD thesis, Westfälische Wilhelms-Universität Münster, 2010.

[16] Lukas Kohlhase and Michael Kohlhase. System description: A semantics-aware LaTeX-to-office converter. Intelligent Computer Mathematics, 2014

[17] Michael Kohlhase, Bogdan A. Matican, and Corneliu-Claudiu Prodescu. Mathwebsearch 0.5: Scaling an open formula search engine. Lecture Notes in Computer Science, 2012.

[18] Michael Kohlhase and Ioan Sucan. A search engine for mathematical formulae, Artificial Intelligence and Symbolic Computation, 2006.

[19] Jacco Krijnen, Doaitse Swierstra, and Marcos O.Viera: Towards an extensible Pandoc, Lecture Notes in Computer Science, 2014.

[20] John MacFarlane. Pandoc user's guide. http://johnmacfarlane.net/pandoc/README.html, 2013.

[21] Arnold Neumaier and Peter Schodl. A framework for representing and processing arbitrary mathematics. KEOD, 2010.

[22] Jiri Rakosnik and Radoslav Pavlov. European digital mathematics library. Digital Presentation and Preservation of Cultural and Scientific Heritage, 2013.

[23] Lucía Romero, Milagros Gutiérrez, and María Laura Caliusco. Conceptualizing the e-learning assessment domain using an ontology network. International Journal of Interactive Multimedia and Artificial Intelligence, IJIMAI, 1(6), pp. 20, 2012.

[24] Heinrich Stamerjohanns and Michael Kohlhase. Transforming the arxiv to XML. In Intelligent Computer Mathematics, 2008.

[25] Martin Thoma. On-line Recognition of Handwritten Mathematical Symbols. PhD thesis, National Research Center, 2014.

[26] Edward Rolando Núñez Valdez, Juan Manuel Cueva Lovelle, Óscar Sanjuán Martínez, Carlos E Montenegro Marín, and Guillermo Infante Hernandez. Social voting techniques: a comparison of the methods used for explicit feedback in recommendation systems. International Journal of Artificial intelligence, IJIMAI, 1(2), pp.79, 2011

[27] Magdalena Wolska and Ivana Kruijff-KorBayova. Analysis of mixed natural and symbolic input in mathematical dialogs. 42nd Meeting of the Association for Computational Linguistics, 2004.

[28] Richard Zanibbi and Dorothea Blostein. Recognition and retrieval of mathematical expressions. International Journal on Document Analysis and Recognition, 2012.

[29] ClausZinn. A computational framework for understanding mathematical discourse. IGPL, 2003.

[30] L. de-la Fuente-Valentín, D. Burgos, R. González Crespo. A4Learning: Alumni Alike Activity Analytics to Self-Assess Personal Progress. IEEE International Conference on Advanced Learning Technologies, 2014

Alberto Corbi works as a senior researcher at the Technology-enhanced Learning & Social Networks (TELSOCK) research group and as a teaching assistant at the School of Engineering, both part of the International University of La Rioja. With a background in physics (M.Sc. in ocean-atmosphere interaction), computing and education, he is currently involved in research fields around recommender systems, eLearning standards and systems interoperability. Simultaneously, he is working on his PhD thesis around medical imaging at the Spanish Council for Scientific Research (CSIC).

Daniel Burgos works as Vice-Chancellor for Research & Technology and UNESCO Chair on eLearning at the International University of La Rioja (unir.net, research.unir.net). Previously he worked as Director of Education Sector and Head of User Experience Lab in the Research & Innovation Department of Atos, Spain, and as an Assistant Professor at Open Universiteit Nederland before that. His research interests are mainly focused on Adaptive and Informal eLearning, Learning & Social Networks, eGames, and eLearning Specifications. He is or has been involved in a number of European-funded R&D projects, such as Intuitel, Hotel, Edumotion, Inspiring Science Education Stellar, Gala, IntelLEO, Go-MyLife, Grapple, Unfold, ProLearn, TenCompetence, EU4ALL, NiHao, Kaleidoscope, Sister, and ComeIn. Prof. Dr. Burgos holds degrees in Communication (PhD), Computer Science (Dr. Ing), Education (PhD), and Business Administration. Furthermore, he is a member of a number of Executive Boards of associations and professional clusters focused on educational technology and eLearning innovation, for example Telearc, Telspain, Menon Network, Adie, and others.