PlanetMath Previews

Joe Corneli*, Raymond Puzio, Deyan Ginev, Aaron Krowne

September 23, 2013

Contents

Introduction

Previews of our ideas about things to develop for PlanetMath can serve several related purposes:

1. They can show members of the PlanetMath community what sorts of exciting and innovative features are in the works.

2. They can interest potential donors and volunteers by showing them what their contributions would help accomplish.

3. If the pace of progress on PlanetMath seems too slow, the previews can potentially indicate where we're stuck and what can be done to bring it back up to speed.

4. The previews can also help to explain to newcomers what PlanetMath "is" – and what it can become.

For example:

*holtzermann17@gmail.com

- PlanetMath is an peer-produced mathematics encyclopedia. ⋆**mature**⋆
- PlanetMath is a testbed for the Planetary software project. ⋆**complete**⋆
- PlanetMath is a collection of peer-produced exercise workbooks and user-created learning pathways. **evolving**
- PlanetMath is a place to ask questions and build a shared knowledge base. **evolving**
- PlanetMath is an incubator for websites using Planetary software. **evolving**
- PlanetMath is a place to get free or for-fee mathematics teaching and tutoring. **evolving**
- PlanetMath is a guide to the mathematics literature. **proto**
- PlanetMath is a collection of public domain or liberally licensed books. **proto**
- PlanetMath is a free/open research repository. **proto**
- PlanetMath is an integration hub for other free/open math projects. **proto**
- PlanetMath is an interactive mathematics MUD/MUSH. **proto**
- PlanetMath is a math wiki that blends ownership with universal editability using a distributed revision control system **proto**
- PlanetMath is the mathematics component of an "indie education bundle" **proto**
- PlanetMath.org, Ltd. is a nonprofit specializing in mathematical software services, mathematical communication, hypertext research, and consulting. **proto**

Several of the most interesting and important previews are described in the sections that follow.

1 "What's Next for PlanetMath/Planetary?"

Lifecycle stage: 1 —proto— 2 ⟶ [**3**]—evolving—4 ⟶ 5 —complete— 6 ⟶ 7 —mature— 8

The main platform (including NNexus, Planetary, etc.) has been doing quite well, but work is going slower now that the "lead dev" Joe Corneli has finished the work required for completing his thesis. Nevertheless, the trials in this thesis show that the user interface needs many improvements and a thorough usability review. Furthermore, board member Deyan Ginev pointed out that the system is not really as easy to install or use as, say, MediaWiki, so a reasonable effort should be put into making a quick-start dev / deployment strategy.

In order to progress to stage 4, we need a solid developer group making regular commits. If PlanetMath itself is going to remain relevant, we need to keep it as an integration platform for work.

See also

See planetary#63, planetary#381, planetary#88 There are a lot of other basic usability issues collected in this PM ticket: planetmath-docs#5 and these Planetary milestones, which are currently past due: PlanetMath Community and PlanetMath Community 2.

2 Books

Lifecycle stage: 1 —proto—[**2**] ⟶ 3 —evolving— 4 ⟶ 5 —complete— 6 ⟶ 7 —mature— 8

Sources of content: (a) Retrodigitization, (b) importing content from other CC-By-SA or more liberally licensed sources, (c) re-using PlanetMath's existing content.

In more detail:

(a) Library of Congress + US Copyright office + Archive.org + Infty + some manual labour yields newly typeset out-of-copyright works (we've successfully run this workflow with our Calculus book)

(b) Importing and enhancing material from websites such as Stackexchange and Wikipedia that use the same license is quite feasible.

(c) Reusing existing material from the PlanetMath encyclopedia by building "collections" (e.g., exploiting the NNexus autolinker or other automated tools to assist with content assembly)

We have done a lot of background research on this. In order to get things moving and progress to Stage 3, we need some hands-on-the keyboard time (mathematics background would be helpful throughout, programming experience helpful for b and c). By default, this will evolve slowly as we assemble new courses and make further experiments with NNexus. However, an influx of vounteer time (or funding) could make things progress more rapidly. (Further details on the Books preview are presented in an Appendix at the end of this document. With a bit more work, any one of these previews could be expanded in a similar fashion.)

3 PlanetMath Outline Series

Lifecycle stage: 1 —proto—[**2**] ⟶ 3 —evolving— 4 ⟶ 5 —complete— 6 ⟶ 7 —mature— 8

The idea of building approximate "feature parity" with the popular Schuam's Outline Series gives us a list of desirable topics to cover (see above).[1] As a rough measure of what it would take to establish "feature parity", let's imagine that each expository section of one of these Outlines is

[1] Abstract Algebra, Advanced Calculus, Advanced Mathematics for Engineers and Scientists, Astronomy, Basic Business Mathematics, Basic Electricity, Basic Mathematics for Electricity and Electronics, Basic Mathematics with Applications to Science and Technology, Beginning Calculus, Beginning Finite Mathematics, Bookkeeping and Accounting, Calculus, Calculus of Finite Differences and Difference Equations, College Algebra, College Mathematics, Differential Equations, Discrete Mathematics, Electronic Devices and Circuits, Elementary Algebra, Elementary Mathematics, Elements of Statistics, Essential Computer Mathematics, Financial Management, Formulas and Tables, Geometry, Intermediate Algebra, Introductory Surveying, Lagrangian Dynamics, Logic, Mathematics for Liberal Arts

equivalent to an encyclopaedia entry. One of the books selected at random has 97 sections and 877 problems, together with 420 worked solutions.

Assuming all the books are of roughly the same size, this comes to something like 6000 encyclopaedia entries and 60000 problems with 20000 solutions, in total. The number of entries is approximately half the number we currently have in the encyclopedia; however, I think we would have to add a few thousand entries to be really comparable, since our coverage in most areas is somewhat spotty.

If we can draw additional material from old out-of-copyright textbooks, that would help. We could also draw many problems and solutions from the questions and answers on StackExchange.

4 Centralized Bibliographic Database

Lifecycle stage: 1 —proto—[**2**] $\longrightarrow$ 3 —evolving— 4 $\longrightarrow$ 5 —complete— 6 $\longrightarrow$ 7 —mature— 8

Software: BibServer (Drupal and LaTeXML compatibility layers to be added)

Content: Initial content to be imported from the Library of Congress "QA" section (the library catalog is public domain since it is a US Government publication; we have the MARC records on disk and should hopefully have them imported into BibServer this summer)

In order to progress to Stage 3, we need some programmer time. We potentially have some help with BibServer lined up at the Open University.

See also

See planetary#80,

5 Virtual Classroom

Lifecycle stage: 1 —proto— 2 $\longrightarrow$ [**3**]—evolving— 4 $\longrightarrow$ 5 —complete— 6 $\longrightarrow$ 7 —mature— 8

Software for Version 0.1: Skype or Google Hangout + MathIM (a math enhanced chat) + PlanetMath forums, articles, collections, groups, and questions. (**This is working now.**)

Future versions:

- Improvements to PlanetMath to make it easier for teachers and students to use
- Tutorial booking and payments system
- More content (books, outlines, syllabi)

In order to progress to Stage 4, we need to run the course we are curently developing to see how that goes. We plan to start this in September and run the course for 10 weeks. This project has the

Majors, Mathematics for Nurses, Mathematics for Physics Students, Mathematics of Finance, Matrix Operations, Operating Systems, Partial Differential Equations, Physics for Engineering and Science, Precalculus, Principles of Accounting, Probability, Probability and Statistics, Software Engineering, Statistics, Statistics for Engineers, Thermodynamics.

benefit of having a built-in business model, so we're able to prioritize it. If it works, we can scale to meet demand by recruiting more teachers (and branching out to one-off tutoring).

See also

See planetary#350.

6 Re-orienting PlanetMath around "Projects"

Lifecycle stage: 1 —proto—[**2**] ⟶ 3 —evolving— 4 ⟶ 5 —complete— 6 ⟶ 7 —mature— 8

At the organizational level, the preview projects outlined here make a great start!

A full implementation would require significant implementation effort towards "PlanetMath 3.0." One reasonable technical step would be to add Git compatibility to Planetary, which will make it easier for collaborating groups of authors to build their own projects.

Further improvements, like issue tracking, etc., would come later. Some of the relevant research background is outlined in Joe's thesis, but the design specifics still need to be worked out. In the conclusion to this thesis, the following table is presented, which summarizes a tentative design (Entity-Relation diagram) for "PlanetMath 3.0," which would add support for projects, project updates, forks, and outcomes, conjectures and ephemeral discussions to the existing prototype.

Context	Feedback	Quality	Structure	Heuristic
$A \leftarrow A$ $A \overset{\ell}{\leftarrow} A$ $X \hookrightarrow \mathbb{X}$	$X \leftarrow T$ $S \leftarrow R$ $\mathbb{X} \to \mathbb{X}^\sharp$	$X \leftarrow Q$ $A \leftarrow C$ $X \to X'$ $\mathbb{X} \models \mathbb{X}^\star$	$A \leftarrow P \leftarrow \mathbb{J} \leftarrow S$ $L \leftarrow A, P$ $M \leftarrow A$ $Q \leftarrow A$	$G \hookleftarrow U$ $S \hookleftarrow H$ $Q, T \rightharpoonup C, W, P$ $G \hookrightarrow \mathbb{E}$
A article ℓ link X object	T post S solution R review	Q question C correction	P problem L collection M classification	G group U user W request H heuristic
$\mathbb{X}$ project	$\sharp$ update	$'$ fork $\star$ outcome	$\mathbb{J}$ conjecture	$\mathbb{E}$ ephemera

In order to progress to Stage 3, we need to get our preview series out to the public on PlanetMath, making it clear how they can add new projects or assist with one of these.

See also

See planetary#351.

7 Internationalization

Lifecycle stage: [**1**]—proto— 2 ⟶ 3 —evolving— 4 ⟶ 5 —complete— 6 ⟶ 7 —mature— 8

We're not working on this at the moment but we recognize that it would be worthwhile! There's a technical component (internationalizing the software) as well as a major content-porting initiative. This may intersect with the Books preview. There are possible connections with current work at KWARC towards building an internationalized glossary. It could be useful to mirror internationalization at the organizational level (with different computers + nonprofit orgs in different countries)

In order to progress to Stage 2 or 3 we would need some technical work, capacity building work, and hands-on-the-keyboard time. We can probably import a lot of content from other sources that would give us a boost once we have the platform ready.

See also

See planetary#309.

8 Computer Math and KRR

Lifecycle stage: [**1**]—proto— 2 ⟶ 3 —evolving— 4 ⟶ 5 —complete— 6 ⟶ 7 —mature— 8 ...

NB. this set of features is "proto" on PlanetMath, but for our collaborators at KWARC there are various projects in these areas, at all stages of completion!... accordingly, it would be great to integrate more KWARC tools into PM (e.g., make sTeX/OMDoc an optional format).

MathML that is produced by LaTeXML from regular TeX isn't particularly meaningful! `$f(x)$` reads as "f times x" in MathML. sTeX could be used to make a canonical (non-ad hoc) coding, to enable a regular TeX parse. These features would enable improved semantic searching and similar functionality. See http://kwarc.github.io/ for a partial list of other projects that build on this infrastructure.

In order to progress to stage 2 or 3, there are some specific Github tickets describing sTeX integration to deal with. There are benefits to using PM as an integration platform.

See also

See planetary#358, planetary#216, planetary#340, and eMath 3.0: Building Blocks for a Social and Semantic Web for Online Mathematics & eLearning.

9 Experimental Math

Lifecycle stage: [**1**]—proto— 2 ⟶ 3 —evolving— 4 ⟶ 5 —complete— 6 ⟶ 7 —mature— 8

We've done nothing about integrating SAGE or Maxima into PlanetMath so far, but we think it would be a good idea. SAGE has a nice web format already - maybe we could re-use it in some way on PlanetMath.

In order to progress to stage 2 or 3, we would want to make contact with some people in the SAGE community and find a concrete project to work on. Deploying SAGE notebooks as an alternative content type or sub-object that can be used within PlanetMath articles would be great.

See also

See planetary#417, WorkingWiki.

10 Hypertext and Metamathematics

Lifecycle stage: 1 —proto— 2 $\longrightarrow$ [**3**] —evolving— 4 $\longrightarrow$ 5 —complete— 6 $\longrightarrow$ 7 —mature– 8

Often distinct from but sometimes overlapping with the ideas in the Computer Math preview, we've been actively looking at and working on issues like these, over the past decade or so:

- Theoretical background work on representing mathematical knowledge as semantic hypergraphs and reasoning with it.
- Theoretical background work on linguistics and mathematics.
- Several versions of Arxana (an Emacs-based hypertext system) have been developed, which will eventually integrate with the above.

In order to progress to stage 4: (a) Round-tripping and exporting from LaTeX documents and Emacs to/from PlanetMath; (b) finishing some full interactive examples of network programming / network math inside Emacs.

An overview of a possible computational agents approach.

One approach would be to formalize the theory of peer learning that Joe Corneli developed in his thesis in a system of computational agents that can collaborate to solve mathematical problems.

There is some precedent for this sort of work within the knowledge-rich artificial intelligence tradition (e.g. Marvin Minsky's *Society of Mind*), although not typically with a mathematics focus.

In his thesis, Joe observed that paragogy has structural similarities to Imre Lakatos's description of mathematical argumentation, as embodied in the dialogs in his *Proofs and Refutations*. There are also parallels to Martin Nowak's work on the evolution of cooperation in a game theoretic setting. Previous research on Lakatos-style computational agents was carried out Alison Pease, a philosopher of mathematics, who Joe interviewed during the requirements-gathering phase of the thesis. However, this prior work does not fall within the knowledge-rich AI tradition, but instead focuses on theory construction from axioms.

The plan would connects the new theory of paragogy with my earlier writing on hypertext and knowledge representation for mathematics (Corneli & Krowne (2005), Corneli & Puzio (2013), Corneli (2003, unpublished), Corneli (2004, unpublished)). This work provides both technical and theoretical foundations, but full articulation of the project will take considerable time and effort. The results could transform mathematical practice, by offering automated assistance at many points in the learning and discovery process.

See also

http://arxana.net, planetary#416, and the old Hyperreal Dictionary of Mathematics wiki page.

Appendix: Further details for the books project

Mock-up/demo

Put in some artistic impression of what the books section might look like.

Detailed description

The purpose of the book project is to make mathematical books in the public domain accessible to the general public in the form of a collaborative digital library. To accomplish this goal, we plan to design and build a system comprised of three interoperating components.

The first subsystem is a retrodigitization toolchain. When complete, this system will allow one to start with a phyiscal book on a library shelf, scan it in to a computer, then subject the result to a series of processing steps which result in a TeX representaion of the book's contents. While the software for this already exists and has been tested, there is room for improvement; by introducing image preprocessing, clustering, and postprocessing, one should be able to significantly improve the accuracy of the process. Given that proofreading and correcting errors is a labor-intensive process, the labor saved by improving the OCR process justifies the effort.

Since, even with these improvements, this process is not 100% accurate, we need the next component, which is an editorial workflow. Based upon the CBPP approach which has been in use for the last decade to produce the PM encyclopaedia and inspired by predecessors such as the St. Pachomius Library and Project Gutenberg's Distributed Proofreaders, this system will coordinate the proofreading of mathematical works by members of the PM community. To participate, a member would start at a page which lists the various works which have been processed but not yet proofread. Upon picking a work, the member would be assigned a page. To work on the page, there would be a webpage which displays the original text, the computer output from the OCR suite, and the rendering of that output. The proofreader's job is to ensure that the rendered output agrees with the original text and, if not, to edit the output as appropriate. Once this is done, an editor will double-check the result and, once all pages have been satisfactorily edited, the system will collect the results and collate them into a hypertext edition.

The third and final component is a reading room which makes the results available to the reading public. To locate books, there will be a catalogue, search facility, and recommender. Once one has located a book, one can read it in several forms. The primary form is hypertext enhanced with links to the encyclopaedia, cross links to other books, notes, reviews, problem solutions, and the like. There will also be files of the book available for downloading and viewing on an e-book reader or printing out. In line with the philosophy of library as a social space, there will be plenty of opportunities for readers to interact with the text and each other by making notes, reviewing books, and participating in discussions.

In addition to these three components, there will also be an area for supporting the project and the PlanetMath organization by sponsoring books and purchasing hard copies.

Roadmap

- Install OCR program and process a first book.
- Conduct preliminary research on OCR techniques.
- Collect suite of samples for OCR evaluation.
- Examine effects of preprocessing strategies.
- Write utility to extract graphic images of individual characters from scans according to XML OCR output.
- Compare effects of different feature vectors, metrics, averaging techniques, and clustering algorithms.
- Determine statistical distributions of features and metrics and develop statistical models of identification.
- Study how to feed output of clustering and average back into training.
- Study distributions on lines and techniques for isolating characters and combining fragments of characters.
- Study postprocessing techniques.
- Study how to convert the "visual" TeX markup produced by Infty to more "semantic" TeX markup.
- Develop techniques for automatically extracting structure and metadata for books.
- Research techniques for combining symbols into equations such as, say, hierarchical clustering.
- Figure how to combine the various programs and techniques into a toolchain so as to maximize correctness.
- Improvise proofreading of first few books using the existing facility for editing encyclopaedia entries.
- Compare different strategies for presenting text to be proofread and highlighting questionable identifications.
- Revise 2005 specification from Noosphere to Planetary.
- Implement proper proofreading facility.
- Implement facility for keeping track of books and editorial workflow.
- Implement facility for outputting completed books.
- Test and document facilities for prooofreading books.

- Collect and write converters to produce versions of books in various file formats.
- Present the first few books using collections facility.
- Enter in math books from Project Gutenberg.
- Incorporate books into indexing and search.
- Study and compare algorithms for recommending books.
- Implement reading room.
- Test and document the reading room.

How to help

If you're a philanthropist, your donations will help move the research and development process along:

- $1000 will purchase an InftyOCR license.
- $2000 will purchase a high-end computer for OCR and related processing.
- $5000 will pay for an OCR research assistant.
- $10000 will pay to implement the books section on PM

If you're a Drupal dude, you can help implement the proofreading facilities and reading room.
If you're a script kiddie, you can help us build our toolchain.
If you're into statistics, you can help us with identifying characters by clustering.
If you're an proofreader, you can help us prepare the first few texts.

Acknowledgements

Thank people who have helped with the initial steps in the roadmap.