WikiMathDB Idea Description

Dzmitry Lazerka

2007-11-01

1 Introduction

Idea of the WikiMathDB project was born while trying to understand a mathematical proof while lot of theories were involved and lot of "obvious" facts were omitted. Then I imagined: was there any way to store the proof so detailed, that even computer can understand it? The more I thought about it, the more I convinced of technical possibility to build a computer system which goal is to contain proofs in very detailed form; so detailed, that even computer can check it (there already are such systems, see below). So, everyone can view at any proof in more detail, as detailed as they need. And allow everyone to add his own theories without interference with others. And provide possibility of different representations on one proof for various languages and terminologies. So, the idea was born.

Who needs in the WikiMathDB system? What is its the purpose?

At first, for educational purposes: if a student cannot understand something "obvious "or have just forgot a well-known thing, he always can look in the WikiMathDB for explanation. Also, teacher can add parts of his course to help students understand the course. A lot of universities distribute synopses and/or their video recordings, the system will be also useful to avoid mistakes and better lectures understanding.

At second, for science purposes: one can always have the most up-to-date version of his work, which surely does not contain mistakes. This will take away a need in continuously distributing and verbose explanation the work to other mathematicians. Of course, the system will not replace science magazines and conferences, but will appreciably help them.

Yet another artificial intelligence? A theorem prover?

No. The system aims to be a DataBase. There is a wide range of computing tools, automated theorem provers. WikiMathDB is not one of them, its responsibilities is storing and explaining, but not evaluation. There also is a huge number of computer-aided math libraries, WikiMathDB differs from them by containing the logical consistency, in the way how an author understands it. The most closest system is a Mizar project (www.mizar.org), born in 1973 and is now containing more than 8000 math definitions and 45000 theorems.

WikiMathDB mainly differs from the Mizar project by allowing to work with usual typographic-processed article (like in DVI format), not with an ASCII computer language. Moreover, a common theoretical description and discussion has taken place in a QED project (http://www-unix.mcs.anl.gov/qed), where the idea of unified computerized mathematical knowledge repository was born.

2 Examples

Example 1 (elementary).

Given: $B \subset A, b \in B$, so $b \in A$.

Why "so"? Maybe in some theories it is not true? Here you can ask the WikiMathDB to explain you why (by, for example, ctrl-clicking on the word "so"). And get the answer based on definition of the subset operator:

$$(B \subset A) \Leftrightarrow (\forall b \in B \Rightarrow b \in A)$$

This answer is an interpretation of already inputted definition of *subset*:

$$\forall X, Y - sets \Rightarrow [X \subset Y) \Leftrightarrow (\forall x \in X \Rightarrow x \in Y)]$$

Where B can be placed on the place of X, A — on the place of Y, and b — on the place of x.

Example 2.

Given: **Linear basis** is a *linearly independent spanning set*.

Having read this, one may want to know, what do they mean while talking "linearly independent "and "spanning set"? By ctrl-clicking on these terms he can get the definitions of the terms. Here WikiMathDB can make an easy translation from English to math: "is a" → " $\in''$, so, this definition will be in form:

$$\forall b \in B \Rightarrow (b \in L \wedge b \in S)$$

Where L is a set of linearly independent vector sets, and S — a set of spanning sets. Asking the system what is "L "gives us a definition of a set of linearly independent vector sets.

But, you should notice that here we omitted the field in which "basis "is defining: subspace and vector space. Because "basis" will be used if form "basis of the subspace M in vector space V". Omitting these details is not allowed in the WikiMathDB. The details like that make possible to fully understand matter by computer. So, writing a definition like for "basis", author has to consider these details while adding anything to the WikiMathDB. It is obvious, that there are a number of such details, and they slow down process of adding new data to the system significantly, but only this way allows to make data logically clear. And, as the Mizar project shows, it is solvable problem.

Example 3 (more comprehensive).
Let's look at one solution in field of probability theory, characteristic functions:

Given: Random variable ξ has the Couchy distribution

$$p_\xi(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}.$$

Problem: Find the characteristic function for ξ.
Solution:

$$f_\xi(t) = \frac{a}{\pi} \int_{-\infty}^{\infty} e^{itx} \frac{dt}{x^2 + a^2} = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{cos(tx)}{x^2 + a^2} dt.$$

Notice that $f_\xi(t)$ – even function, so we only need to know its values for $t > 0$. Let us differentiate both sides of the last equation, we'll get:

$$f_\xi(t) = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{-x \, sin(tx)}{x^2 + a^2} dt. \qquad (1)$$

It is known from the math analysis that

$$\int_{-\infty}^{\infty} \frac{sin(tx)}{x} dx = \pi \ \ (t > 0) \quad , \text{so} \quad a = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{sin(tx)}{x} dx. \qquad (2)$$

Summarizing (1) and (2), we'll get

$$f_\xi' + a = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{a^2}{x^2 + a^2} \frac{sin(tx)}{x} dx.$$

Let us differentiate both sides by t. We get $f_\xi''(t) = a^2 f_\xi(t)$. So, $f_\xi(t) = c_1 e^{at} + c_2 e^{-at}$, $t > 0$. Because of $f_\xi(t)$ – bounded function on $\mathbb{R}$, then $c_1 = 0$, and, basing on property $f_\xi(0) = 1$, get $c_2 = 1$. So, while $t > 0$ $f_\xi(t) = e^{-at}$, but taking consideration of the function eveness, finally get $f_\xi(t) = e^{-a|t|}$, $t \in \mathbb{R}$.

Here we see a lot of hidden things: starting from definition of characteristic function and ending with differentiation of improper parametrized integral by parameter, along with on-the-fly solve of Euler's differential equation. There are wasting a reader's time if the reader wants to check (or understand) the solution. It is good if reader is always a genius, but if he isn't? With the help of WikiMathDB, author can waste his own time but save readers' time. Also, explaining something in detail to the stupid computer can eliminate all mistakes that human eye may not notice.

3   Global Design

Version 1.0 should provide basic functionality for **explaining** and **viewing** existent content, along with **adding** new content. These include:

1. **Common**. All development must use the portable technologies and tools for compatibility, easier maintaining and development.

2. **Client side**. Basic client should be **crossplatform**, has a **WYSIWYG math editor**. Added documents should be sent and stored too, not only parsed and transformed versions. A web-application form sounds to be a good idea.

3. **Server side**. Server side should provide an open, documented API, to enable creating different clients for the same servers. The SOAP technology sounds to be a good idea. Structure should consider further features described below for future upgrades.

Furter versions will point on:

- **internationalization** – distinguish different languages, along with different terminologies for the same fields.

- **distributed server side** – allow anyone to install their own clones of the database, work with them, and send changes to our repository.

- **versioning** – track changes made to the database, allowing reverting changes made. Forking and merging will be a good point too.

- **authorization** – close some parts of database from public edit.

- **heuristics** – help editors to add new content by performing evaluating and computing like it done by a lot of computing environments like Mathematica®.