The problem of Examples at the Mathematics Research Frontier

towards a Catalogue of Mathematical Objects

Lars Hellström
Mälardalen University

21 March 2024, SNAG'24, Linköping

Important notice

This is not a talk of research in mathematics,
but a talk on research in mathematics:

- How is it conducted?
- What drives it? (Besides coffee.)
- How could it be furthered?

I believe I have identified an aspect which is not given the recognition it deserves, and have a proposal for how this could be remedied.

But let's start at the beginning: how I came to think about this.

Strong hom-associativity

Recall that a hom-algebra $(\mathcal{A}, \cdot, \alpha)$ is said to be hom-associative if

$$(x_1 \cdot x_2) \cdot \alpha(x_3) = \alpha(x_1) \cdot (x_2 \cdot x_3) \qquad \text{for all } x_1, x_2, x_3 \in \mathcal{A}.$$

I had proved, using rewriting methods, that it *does not* follow from the above that

$$\bigl(x_1 \cdot (x_2 \cdot x_3)\bigr) \cdot \bigl(\alpha(x_4) \cdot x_5\bigr) - \bigl(x_1 \cdot \alpha(x_2)\bigr) \cdot \bigl((x_3 \cdot x_4) \cdot x_5\bigr) = 0 \qquad \text{for all } x_1, x_2, x_3, x_4, x_5 \in \mathcal{A},$$

but it *does follow* that

$$\Bigl(\bigl(x_1 \cdot (x_2 \cdot x_3)\bigr) \cdot \bigl(\alpha(x_4) \cdot x_5\bigr) - \bigl(x_1 \cdot \alpha(x_2)\bigr) \cdot \bigl((x_3 \cdot x_4) \cdot x_5\bigr)\Bigr) \cdot \alpha\bigl(\alpha(x_6)\bigr) = 0 \qquad \text{for all } x_1, x_2, x_3, x_4, x_5, x_6 \in \mathcal{A}.$$

This prompted me to define the class of strongly hom-associative algebras.

Looking for hom-algebras

When you define a subclass of objects, you want examples that are in and examples not in the new class.

For hom-associative algebras, the standard examples are Yau twist algebras: associative algebras where you've deformed the multiplication.
Those turn out to all be strongly hom-associative.

Apart from some classifications of 2- and 3-dimensional algebras — which Per Bäck checked and found to also be strongly hom-associative — that seemed to be *all the examples known*!
That's not a satisfactory state of things.

Looking for hom-algebras

When you define a subclass of objects, you want examples that are in and examples not in the new class.

For hom-associative algebras, the standard examples are Yau twist algebras: associative algebras where you've deformed the multiplication.
Those turn out to all be strongly hom-associative.

Apart from some classifications of 2- and 3-dimensional algebras — which Per Bäck checked and found to also be strongly hom-associative — that seemed to be *all the examples known*!
That's not a satisfactory state of things.

My talk at last year's SNAG was about two new constructions I've found, one of which *can* produce a non-strongly hom-associative algebra. But that got me thinking.

Looking for an X

Suppose more generally you're looking for an example with property X. What can you do?

1. Sit down a think hard! (Sometimes surprisingly effective.)
2. Ask your colleagues.
3. Search the literature.

The last is what other sciences teach that a *researcher* should do. But it's surprisingly hard.
Sometimes you *can* type the obvious keywords into MathSciNet and out pops a paper with the sought example. But usually not.

And it's one thing if this is a distinct Research Question of yours—then you can afford sifting through a lot of tangential material.
It's another thing if you're just trying to make sense of a definition or need to clarify a step in a proof.

Knowing an X

At the other end of the stick, you may know an example of an X, perhaps from 1 or 2 above. What do you do with it?

Sergei tells me: Publish. In practice, there might be lots of reasons not to:

1. "In hindsight, it's obvious."
   I might make an exercise for an abstract algebra course out of it, but it's not enough for an article.
2. "I wasn't the one who originally came up with it."
3. "It's not in my areas of research."
4. "I don't know if it's new. Surely someone else must have thought of this already!"

But no doubt knowledge being stuck in people's hard drives or drawers is impeding progress.

Traditional options for publication

Separate article That would be very lightweight.

Risks rejection: too little substance, not enough reference to literature.

If you work on the article to make it more fleshed-out, the effort starts to grow unproportionally.

Add-on to article Include the example in an article under publication, whose main focus is something else, *just to get it out there*.

Obvious disadvantage is that nobody can find this example by searching the literature; you have to stumble upon it.

Wikipedia

At one point, I considered inserting my first example into a Wikipedia article, just so that I could *be done with it.*
This would have been the right amount of work for a random example, IMO.

However, doing so would have been wrong, because one of three core Wikipedia policies is that of No Original Research:

> *Wikipedia does not publish original thought. All material in Wikipedia must be attributable to a reliable, published source. Articles must not contain any new analysis or synthesis of published material that reaches or implies a conclusion not clearly stated by the sources themselves.*

For mathematics, there is a kind of loophole in that:

> *Simple calculations are not original research,*

but to publish things seemingly new is against the spirit of the policy.

Online databases

Another possibility might be to submit one's example to a suitable "online database" of appropriate mathematical objects. Examples of such are the:

- On-Line Encyclopedia of Integer Sequences (OEIS) `oeis.org` (canonical),
- Database of Ring Theory `ringtheory.herokuapp.com` (active),
- Manifold Atlas `www.map.mpim-bonn.mpg.de` (dormant),
- π-base `topology.pi-base.org` (very active).

Examining these exposes certain problems.

Limited and arbitrary scope

These sites all have a set scope—you can't link an object in one area to its image in another, because only one of those are within the scope.

For the OEIS this makes sense; the object is the sequence, and everything else is notes on that object.

For the Database of Ring Theory (DaRT) it becomes more problematic; the algebro-geometric correspondence is fundamental, but if it's not a ring or module, it's not welcome there.

Worse, the scope boundaries are often arbitrary. The DaRT requires all its rings to be *unital*, but the reasons given for why are dubious.

Moreover there is no space for extra structure, such as the twisting map α of a hom-algebra. But often this extra structure is what makes the example interesting!

Manual curation

A bottleneck for many of these databases is how new content gets added: most require admins to manually approve additions.
Frequently, adding material boils down to having these admins *manually edit files*.

The perceived *advantage* of this is quality control: you want to make sure that content is *correct*.

The disadvantage is that if the admins are busy, or drops off completely, then everything freezes.
This is likely the state of the *Manifold Atlas* (and I've come across collections even worse off).

Wikipedia wouldn't have become the resource it is today if there had been an editorial board that had to review every edit before it gets published!

Rectilinear thinking

Presentation may be more or less fancy, but many of these databases boil down to being huge tables: objects on one axis, properties on the other, every cell a tickbox yes/no/unknown.

This invites a kind of rectilinear thinking, where new objects are only welcome if all established properties make sense for it, and new properties are only considered if they make sense for all existing objects.

	P_1	P_2	P_3	P_4	...
Q_1		Yes			
Q_2	Yes	No	No	No	
Q_3	No	Yes			
Q_4			No	?	
Q_5		Yes	Yes		
Q_6					
Q_7	Yes				
Q_8				Yes	
$\vdots$					

Only half the story

Large parts of π-base reads like the "extra web content" for the book *Counterexamples in Topology*; you can search for examples alright, but when you look at the entry, more than half of them just say "see item so-and-so in the book".

To be useful, you want the database to show you the full construction.

Moreover, you also want to know the property definitions, because:

- Sometimes it's something unfamiliar you haven't heard of before.
- Sometimes they're using a *slightly different* definition of a familiar concept, which in unlucky cases changes the outcome.

How to do better?

Basically by avoiding the problems mentioned above.

- Don't segregate between the known and the original!
- Don't hardwire to one part of mathematics!
- Don't depend on a staff of moderators!
- Include the definitions!

You should be able go to a website, enter a search like 'I want something that is A, B, and D, but not C', and get back a report like ...

Searching for an object which is Lebesgue measure zero and isn't countable. There are 5 exact matches.

1 The Cantor set

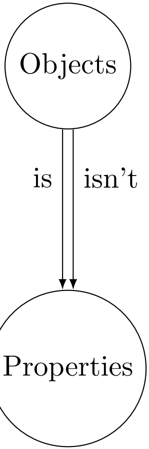
Construction. This is constructed by removing the open middle third $(\frac{1}{3}; \frac{2}{3})$ from the unit interval $[0,1] \subset \mathbb{R}$, then repeatedly the middle third from every remaining interval, an infinite number of times. Rigorously, an $x \in [0,1]$ belongs to this set if there is no $n \geqslant 1$ and $a_1, \dots, a_{n-1} \in \{0,2\}$ such that

$$\frac{1}{3^n} + \sum_{k=1}^{n-1} \frac{a_k}{3^k} < x < \frac{2}{3^n} + \sum_{k=1}^{n-1} \frac{a_k}{3^k}.$$

Construction. This is the set of points on the form $\sum_{n=1}^{\infty} a_n 3^{-n}$ where $\{a_n\}_{n=1}^{\infty} \subseteq \{0,2\}$.

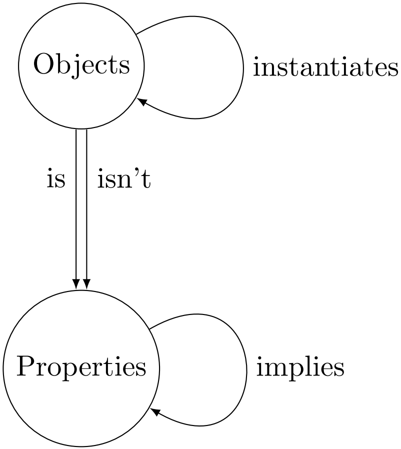
This is closed. This isn't connected. This is bounded, totally disconnected, and compact.

Backend: an RDF database

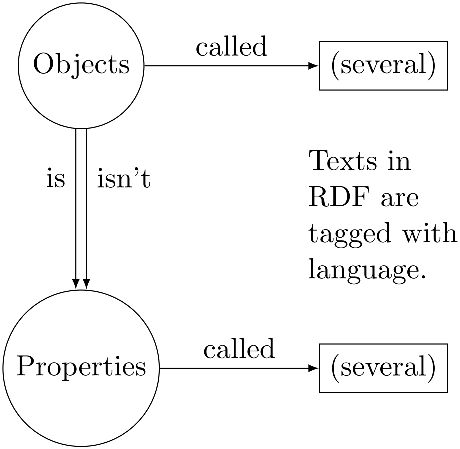


Backend: an RDF database

If *properties* are known to imply other properties, the implied ones need not be stated explicitly. Likewise some *objects* can be special cases of others.

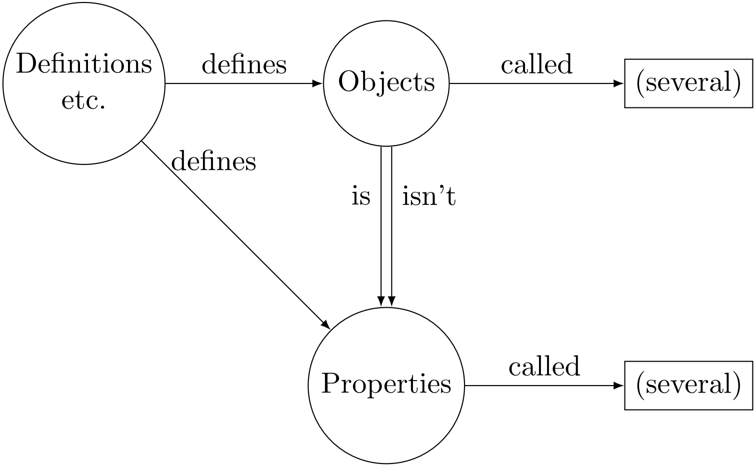


Backend: an RDF database



Texts in RDF are tagged with language.

Backend: an RDF database



Finding the property you want

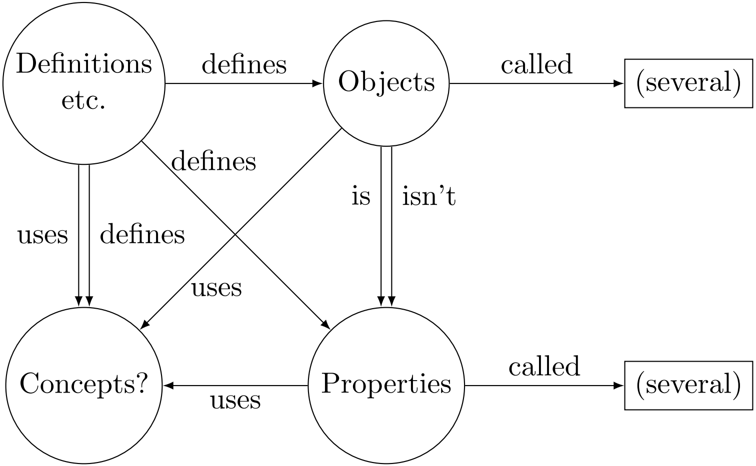
The name of a property is a good place to start.

But every now and then it will be necessary to look at the definition text to verify that it is the intended one.
Just consider how many different meanings there are of *regular*, *simple*, or *uniform*.

However, some properties don't have an established name, or maybe you don't know that name.
In that case, it can help to also classify things by which concepts they use: if you look for a property combining three concepts, then there is a chance the list is short enough that you can locate the one you want.

Another approach is to go via known *objects* with the property you seek. If it is in the database, then someone should have made use of it.

Backend: an RDF database



How can we know the contents are correct?

A good advice is to trust, but verify!

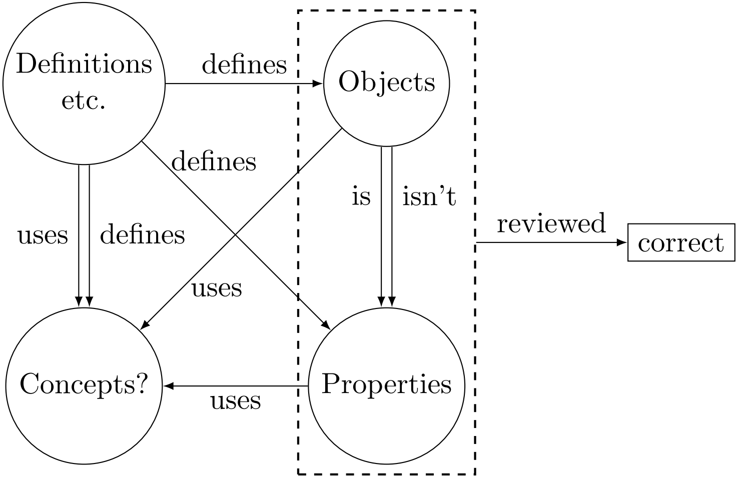
- Mathematicians rarely make claims of which they don't have a proof. If someone bothers to claim it, there's a good chance it really is true.
- If you're interested in an example, you're probably not content with seeing it claimed that it has a property—you will want to verify that it's true *yourself*.

  Hence it wouldn't be a disaster if some claim is wrong, because the stuff you rely on, you have checked.

  In this regard *examples* are different from *theorems*, which people are much more likely to use without reviewing the proof.

In addition, the database can store reviews of claims.

Backend: an RDF database



Reviews I

The report might actually look more like:

1 The Cantor set[reviewed]

Construction.[reviewed] This is constructed by removing the open middle third $(\frac{1}{3}; \frac{2}{3})$ from the unit interval $[0,1] \subset \mathbb{R}$, then repeatedly the middle third from every remaining interval, an infinite number of times. Rigorously, an $x \in [0,1]$ belongs to this set if there is no $n \geqslant 1$ and $a_1, \dots, a_{n-1} \in \{0,2\}$ such that

$$\frac{1}{3^n} + \sum_{k=1}^{n-1} \frac{a_k}{3^k} < x < \frac{2}{3^n} + \sum_{k=1}^{n-1} \frac{a_k}{3^k}.$$

This is closed[reviewed]. This isn't connected[review]. This is bounded[review], totally disconnected[reviewed], and compact[review].

Reviews II

Those review items would be links. *Clicking* on one of them would take you to a page where you can review the claim: is it correct, wrong, or something more complicated?

Matches based on claims reviewed as correct would appear higher in reports. Claims reviewed as wrong would by default be ignored in searches.

Reviews would *not* be anonymous; by clicking on the reviewed link, you would be able to see who made this review.

(I'll skip the matter of *whose* reviews may be trusted.)

How to add content

Well, what do mathematicians do with stuff they want to publish?

How to add content

Well, what do mathematicians do with stuff they want to publish?
They write it up in LaTeX!

1. Write a `.tex`-file on your own computer, with the content you want to add. You can typeset it yourself to see that everything looks fine.

   In that file, the searchable claims are metadata on the mathematics. These need to be encoded using *appropriate markup*, which doesn't have to be more complicated that used at the front of an ordinary article (if you consider complex author affiliations, keywords, and so on).

2. Upload the `.tex`-file to the catalogue website, to have its contents added.

   The file will *not* be used as-is, but rather parsed and chopped up into the individual claims. You would be shown the results and asked to confirm that they are as intended.

But you could be given remarks to address first. . .

Detecting duplicates

One issue for catalogues is duplicates of objects already included. But it is *easy to check* if there is an existing object with the same set of properties as the newly submitted one.

When there is, just ask the submitter to first check if the new object duplicates any of these existing ones.

- If it doesn't, then suggest the submitter to come up with some additional property that would distinguish the new object from the old ones.
- Even if the *object* is a duplicate, the submission may still contain useful material, e.g. a different construction of the same thing.

The site can be helpful with providing snippets of LaTeX code for speaking about old objects.

Project status

Rather early.

- *I'm* working on the basic encoding of the information, including the LaTeX aspects.
- In the near future Mälardalen University is going to announce *several* multidisciplinary doctoral student positions in mathematics, one of which can be for the Catalogue project, if we get a good applicant.
  I'm hoping this will help with the web side of the matter.

Can I help?

Sure, there's plenty of work to do. ;-)

If you're not into the programming side of things, you can still think about the following:

1. What nifty examples do *you* know, that others should learn about as well?
2. What properties of these examples might someone be looking for, who would be delighted to learn about them?
3. How would one *express* these properties, as something that can be searched for?

On the last matter, an example wouldn't have to be all about properties of *one* object. For example, an ideal with some properties and sitting within a ring with certain other properties (not referencing the ideal at all) is perfectly feasible to search for; graph databases are all about matching a *pattern*.

That's all!