nlab_clean

June 7, 2024

1 nLab in plain text

The aim of this notebook is to convert the nLab page source into plain text so that it can be fed into standard NLP tools.

[1]:
```
import JSON

skip_pages = (
    "Timeline of category theory and related mathematics",
    "AUTOMATH",
    "IGNORE",
    "DELETE",
    "EMPTY",
)

cd("/NetMath/nLab/data")
pages = open(JSON.parse,"2024/nlab_scrape.json","r")
print("All pages: ", length(pages),"\n")
filter!(page -> page["name"]   skip_pages, pages)
pages = Dict(pop!(page, "name") => page for page in pages)
print("Filtered pages: ", length(pages),"\n")
```

```
All pages: 19246
Filtered pages: 18516
```

We will use pandoc to convert the Markdown to plain text.

[2]:
```
function markdown_to_plain(s)
    io = IOBuffer()
    open(`pandoc --quiet --from markdown --to plain`, io, write=true) do pd
        write(pd, s)
    end
    String(take!(io))
end

markdown_to_plain("A **sentence** with _some_ [Markdown formatting](https://www.
 ↪markdownguide.org)")
```

[2]:
```
"A sentence with some Markdown formatting\r\n"
```

Besides Markdown formatting, the nLab supports LaTeX math and wiki, specifically Instiki, syntax. We strip out all LaTeX math which is not converted by Pandoc and all wiki syntax besides page links.

[3]:
```
""" Strip LaTeX math, display and inline.
"""
function strip_latex_math(s)
    s = replace(s, r"\$\$(.*?)\$\$"s => "")
    s = replace(s, r"\$(.*?)\$"s => "")
    s = replace(s, r"\\\[(.*?)\\\]"s => "")
end

""" Strip Instiki commands such as includes, redirects, and ToCs.
"""
function strip_wiki_commands(s)
    s = join(filter(split(s, "\n")) do line
        !any(startswith(lstrip(line), prefix)
             for prefix in ("+--", "=--", "{:", "{#", "[[!"))
    end, "\n")
    s = replace(s, r"{#(.*?)}" => "")
end

""" Replace wiki page links with plain text.
"""
function replace_page_links(s)
    # Links of form [[page name|displayed text]].
    s = replace(s, r"\[\[([^\]]*?)\|(.*?)\]\]" => s"\2")
    # Links of form [[page name]]
    s = replace(s, r"\[\[(.*?)\]\]" => s"\1")
end

nlab_to_plain(source) = source |> strip_wiki_commands |> replace_page_links |>
    markdown_to_plain |> strip_latex_math
```

[3]:
```
nlab_to_plain (generic function with 1 method)
```

Run this pipeline on the nLab corpus.

[4]:
```
using ProgressMeter
ProgressMeter.ijulia_behavior(:clear)

prog = Progress(length(pages))
for (name, page) in pairs(pages)
    ProgressMeter.next!(prog, showvalues=[(:name, name)])
    page["plain"] = nlab_to_plain(page["source"])
end
```

```
Progress: 100%|                         | Time:
0:15:41
  name:  formal deformation quantization
```

[15]:
```
open("nlab_clean.json", "w") do clean
    println(clean, "[")
    for (name, page) in pairs(pages)
        println(clean, "  {")
        rec = page["plain"]
        rec = replace(rec, r"(\r\n|\n|\r)" => " ")
        rec = replace(rec, r"\s+" => " ")
        rec = replace(rec, r"-" => " ")
        println(clean, "    \"context\": \"$rec\",")
        println(clean, "    \"title\": \"$name\"")
        println(clean, "  },")
    end
    println(clean, "]")
end
```

1.1 Examples

[6]:
```
pages["rig"]["plain"] |> println
```

```
###Context### #### Algebra

#Contents# * table of contents

Idea

In algebra, by a rig one means a mathematical structure much like a ring
but without the assumption that every element has an additive inverse,
hence without the assumption of negatives (whence the omission of "n"
from "ring" [Schanuel 1991 p. 379, Lawvere 1992 p. 2])

Definition

A rig is a set R with binary operations of addition and multiplication,
such that

-   R is a monoid under multiplication;
-   R is a commutative monoid under addition;
-   multiplication distributes over addition, i.e. the distributivity
    laws hold:
    x  (y + z) = (x  y) + (x  z)
    (y + z)  x = (y  x) + (z  x)
    , and also the absorption/annihilation laws, which are their nullary
    version:
```

$$0 \ x = 0 = x \ 0$$

In a ring, absorption follows from distributivity, since $0 \ x + 0 \ x = (0 + 0) \ x = 0 \ x$ and we can cancel one copy to obtain $0 \ x = 0$. In a rig, however, we have to assert absorption separately.

More sophisticatedly, we can say that, just as a ring is a monoid object in abelian groups, so a rig is a monoid object in commutative monoids, where abelian groups and commitative monoids have suitable monoidal structures (they are not the cartesian ones).

Equivalently, a rig is the hom-set of a category with a single object that is enriched in the category of commutative monoids.

Rigs and rig homomorphisms form the category Rig.

Further weakening

As with rings, one sometimes considers non-associative or non-unital versions (where multiplication may not be associative or may have no identity). It is rarer to remove requirements from addition as we have done here. But notice that while R can be proved (from the other axioms) to be an abelian group under addition (and therefore a ring) as long as it is a group, this argument does not go through if it is only a monoid. If we assert only distributivity on one side, however, then we can have a noncommutative addition; see near-ring.

Properties

Many rigs are either rings or distributive lattices. Indeed, a ring is precisely a rig that forms a group under addition, while a distributive lattice is precisely a commutative, simple rig in which both operations are idempotent (see (Golan 2003, Proposition 2.25)). Note that a Boolean algebra is a rig in both ways: as a lattice and as a Boolean ring.

Any rig can be "completed" to a ring by adding negatives, in generalization of how the natural numbers are completed to the integers. When applied to the set of isomorphism classes of objects in a rig category, the result is part of algebraic K-theory.

More formally, the ring completion of a rig R is obtained by applying the group completion functor to the underlying additive monoid of R, and extending the rig multiplication to a ring multiplication by exploiting distributivity; this gives the left adjoint $F : Rig \to Ring$ to the forgetful functor $U : Ring \to Rig$. Note however that the unit of the adjunction $R \to UF(R)$ is not monic if the additive monoid of R is not cancellative, despite an informal convention that "completion" should usually mean a monad where the unit is monic.

Matrices of rigs can be used to formulate versions of matrix mechanics.

Every rig with positive characteristic is a ring.

Examples

Some rigs which are neither rings nor distributive lattices include:

- The natural numbers.

- The nonnegative rational numbers and the nonnegative real numbers.

- Polynomials with coefficients in any rig.

- The set of isomorphism classes of objects in any distributive category, or more generally in any rig category.

- The tropical rig, which is $\{\infty\}$ with addition $x \;\; y = \min(x, y)$ and multiplication $x \;\; y = x + y$.

  Tropical rigs are among the important class of idempotent semirings.

- The ideals of a commutative ring form a rig under ideal addition and multiplication, where the unit and zero ideals are the unit and zero elements of the rig, respectively. They also form a distributive lattice and therefore a rig in another way; note that the addition operation is the same in both rigs but the multiplication operation is different (being intersection in the lattice).

Related concepts

- semiring

- rng (nonunital ring)

- ring, near-ring

- tropical geometry, tropical semiring

- idempotent semiring

- A categorification of the notion of rig is the notion of rig category, or more generally colax-distributive rig category. See also 2-rig and distributivity for monoidal structures.

- semi-monoid/semi-group

- semi-category, semi-Segal space

- semi-simplicial set

- Burnside rig

- division rig

- multiplicatively cancellable rig

References

The terminology "rig" is due to:

- Stephen H. Schanuel, p. 379 of: Negative sets have Euler characteristic and dimension, in: Category Theory, Lecture Notes in Mathematics 1488 (1991) 379-385 [doi:10.1007/BFb0084232]

- William Lawvere, pp. 1 of: Introduction to Linear Categories and Applications, course lecture notes (1992) [pdf, Lawvere-LinearCategories.pdf:file]

as recalled in:

- F. William Lawvere: The legacy of Steve Schanuel! (2015) [web]

"We were amused when we finally revealed to each other that we had each independently come up with the term ‘rig’."

Discussion under the name semirings:

- Jonathan S. Golan, Semirings and their applications. Updated and expanded version of The theory of semirings, with applications to mathematics and theoretical computer science, Longman Sci. Tech., Harlow, 1992, MR1163371. Kluwer Academic Publishers, Dordrecht, 1999. xii+381 pp.

- Jonathan S. Golan, Semirings and affine equations over them: theory and applications (Vol. 556). Springer Science & Business Media, 2003.

- M. Marcolli, R. Thomgren, Thermodynamical semirings, arXiv/1108.2874

- wikipedia semiring

- J. Jun, S. Ray, J. Tolliver, Lattices, spectral spaces, and closure operations on idempotent semirings, arxiv/2001.00808

[7]:
```
pages["locally posetal 2-category"]["plain"] |> println
```

```
#Contents# * automatic table of contents goes here

##Definition

A 2-category C is locally posetal or locally partially ordered or
Pos-enriched if every hom-category C(x,y) is a poset - an object of the
category Pos of partial orders. One can also consider a locally
preordered 2-category, where every hom-category is a proset (a
preordered set); up to equivalence of 2-categories, these aren't any
more general.

Locally posetal 2-categories are the usual model of 2-posets, aka
(1,2)-categories. Just as the motivating example of a 2-category is the
2-category Cat of categories, so the motivating example of a 2-poset is
the 2-poset Pos of posets. If you interpret  as a full
sub-2-category of , then it is indeed locally posetal. Similarly,
the 2-category of prosets is a locally preordered 2-category that is
equivalent to Pos.

Compare the notion of partially ordered category. A locally partially
ordered category is a category enriched over the category Pos of posets,
while a partially ordered category is a category internal to Pos.
Similarly, a locally partially ordered category is a special kind of
2-category, while a partially ordered category is a special kind of
double category.

Examples

-   Pos
-   Rel
```

[8]:
```
pages["hypergraph category"]["plain"] |> println
```

```
Context

Category theory

Graph theory

Hypergraph categories

-   table of contents

Idea
```

A hypergraph category is a monoidal category whose string diagrams are hypergraphs. Recall that in general the vertices of a string diagram correspond to morphisms in a category, and its edges to objects. An ordinary string diagram is a directed graph, where the inputs and outputs of a vertex describe objects appearing in a tensor product-decomposition of the domain and codomain of a morphism; each edge is connected to only one vertex as input and one vertex as output because of how morphisms in a category are composed. A hypergraph category allows edges to connect to many vertices as input and many vertices as output, which category theoretically means that we may compose many morphisms containing an object in their codomain with many morphisms containing that object in their domain.

Hypergraph categories have been reinvented many times and given many different names, such as "well-supported compact closed categories" (Carboni and RSW), "dgs-monoidal categories" (GH), and "spidered/dungeon categories" (Morton). The name "hypergraph category" is more recent (Kissinger and Fong).

Definition

A hypergraph category is:

- a symmetric monoidal category in which
- each object is equipped with the structure of a special commutative Frobenius algebra, such that
- the Frobenius algebra structure of any tensor product X Y is induced in the canonical way from those of X and Y.

Note in particular that we do not require the morphisms of the category to be Frobenius algebra morphisms.

Examples

- The category Rel of sets and relations (with cartesian product of sets as the monoidal product) is a hypergraph category, with the Frobenius algebra on X given by the "deleting and copying" comonoid {(x,*) x X}, {(x,(x,x))|x X} together with its opposite.

- More generally, categories of spans, cospans, relations and corelations in any category (with the appropriate structure) can be made into hypergraph categories by choosing the correct monoidal structure. For example, the category FinRel is hypergraph when this category is given the × monoidal structure (beware: this is not the categorical product in FinRel; it comes from cartesian product in FinSet). The same is probably true of relations in any regular category. The category FinRel is not hypergraph when given the + monoidal structure. The category FinCorel is hypergraph when this category is given the + monoidal structure (coming from disjoint unions).

- Categories of decorated cospans and decorated corelations are hypergraph categories.

Remarks

- The reason for the definition is that if X is a special commutative Frobenius algebra, then there is a unique morphism induced by the Frobenius algebra structure. It can of course be defined as the m-ary multiplication followed by the n-ary comultiplication; the real point is that the special commutative Frobenius axioms ensure that any composite of two such morphisms is again another such morphism. This is what enables the hypergraph string diagrams described informally above. (Some authors refer to these morphisms as "spiders" due to their appearance in string diagrams, as a black node with $m + n$ legs.)

- The free hypergraph category on one object is the category of finite sets and isomorphism classes of cospans. This is a decategorification of the fact that the free monoidal category containing a (non-special) commutative Frobenius algebra is the category of 1-dimensional manifolds and isomorphism classes of 2-dimensional cobordisms. More general free hypergraph categories can be constructed using labeled cospans.

- Note that the special commutative Frobenius algebras are not required to be "extra-special", meaning that the morphism $I = X^{(0)} \to X^{(0)} = I$ need not be the identity. Thus, the relevant sort of "hypergraphs" can contain "edges not incident to any vertices". If we add the extra-special condition, the cospans are replaced by "co-relations", i.e. jointly surjective cospans.

- A hypergraph category can be equivalently defined as a symmetric monoidal category that supplies special commutative Frobenius algebras.

Properties

As monoids in a presheaf category

Let Cospan_(Δ) be the free hypergraph category on generators Δ. The objects are lists of elements of Δ, or equivalently pairs (m, l) of a natural number m and a labelling $l : [m] \to \Delta$. The morphisms are compatible cospans of functions (up to isomorphism).

Theorem

Fong, Spivak. To give a hypergraph category with chosen objects Δ that generate it is to give a lax monoidal functor Cospan_(Δ) → Set. (Formally, this can be made into an equivalence between the category of objectwise-free hypergraph categories and the lax monoidal functors; and every hypergraph category is in some sense equivalent to an objectwise-free one.)

A hypergraph category is thus a presheaf H: Cospan_(Δ) → Set that has lax monoidal structure [1H()H(a)H(b)H(a,b)] which is strongly reminiscent of the mix rule. A lax monoidal functor is the same thing as a monoid for the Day convolution. Thus hypergraph categories are monoids in the presheaf category [Cospan_(Δ), Set].

References

- Aleks Kissinger, Finite matrices are complete for (dagger-)hypergraph categories. (arxiv:1406.5942)

- Brendan Fong, Decorated cospans, Theory and Applications of Categories, Vol. 30, 2015, No. 33, 1096-1120. (arxiv:1502.00872)

- Brendan Fong, The Algebra of Open and Interconnected Systems, Ph.D. Thesis, Department of Computer Science, University of Oxford, 2016. (arxiv:1609.05382)

- Aurelio Carboni, Matrices, relations and group representations J. Algebra, 138(2):497-529, 1991.

- Robert Rosebrugh, N. Sabadini, and R. F. C. Walters. Generic commutative separable algebras and cospans of graphs. Th. App. Cat. 15(6):164-177, 2005. online.

- F. Gadducci, R. Heckel. An inductive view of graph transformation. In "Recent Trends in Algebraic Development Techniques", Lecture Notes in Computer Science 1376:223-237. Springer-Verlag, Berlin Heidelberg, 1998.

- Jason Morton, Belief propagation in monoidal categories In Bob Coecke, I. Hasuo, P. Panangaden (eds.) Quantum Physics and Logic 2014 (QPL 2014), EPTCS 172:262-269.

- Brendan Fong and David Spivak, Hypergraph categories. (arxiv:1806.08304)