Searching the Space of Mathematical Knowledge

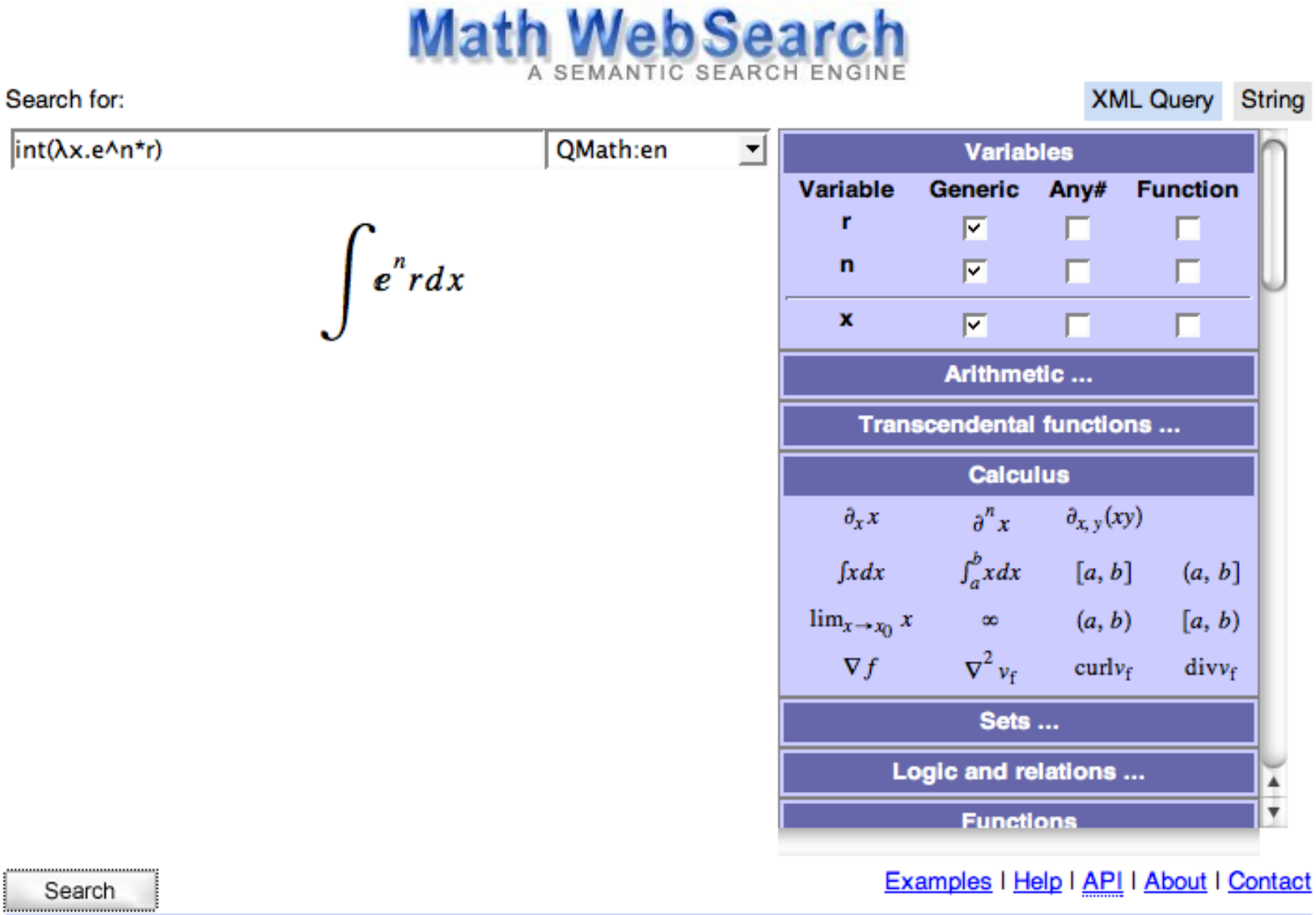
Michael Kohlhase & Mihnea Iancu

http://kwarc.info/kohlhase
Center for Advanced Systems Engineering
Jacobs University Bremen, Germany

8. July 2012, Math Information Retrieval Symposium

Classical Math Search Engines

Instead of a Demo: Searching for Signal Power



Instead of a Demo: Search Results

Other integrals (5 formulas) (Source)

Other integrals (5 formulas)

Matched term:

$$\int \frac{e^{3z/4}}{\left(-2+e^{3z/4}\right)\sqrt{-2+e^{3z/4}+e^{3z/2}}}\,dz == \frac{2}{3}\left(\log\left(-2+e^{3z/4}\right)-\log\left(4\sqrt{-2+e^{3z/4}+e^{3z/2}}+5e^{3z/4}-2\right)\right)$$

Rank: 100%

XML Source

Used substitution:

$$\mathbf{n} \to 3z4^{-1}$$

$$\mathbf{r} \to \left(\left((-2)+e^{3z4^{-1}}\right)\left((-2)+e^{3z4^{-1}}+e^{3z2^{-1}}\right)^{1/2}\right)^{-1}$$

$$\mathbf{x} \to z$$

Instead of a Demo: LaTeX-based Search on the arXiv

Questions Activity Sign In Books Articles MWS Engine BETA

\lim_{\qvar{x}\rightarrow 0}\qvar{y}

$\lim_{x \to 0} y$

```
<m:apply>
      <m:apply>
        <m:csymbol
cd="ambiguous">subscript</m:csymbol>
        <m:limit/>
        <m:apply>
          <m:ci>→</m:ci>
```

Search

Examples - LaTeX queries
Generic subscript search
Specific subscript search
Specific integral search
Physical constant search
All limits approaching zero
Text in math search

1 2 next

$$\chi(t,t_w) = \lim_{h_0 \to 0} \frac{m[h](t)}{h_0}.$$

Generalized off-equilibrium fluctuation-dissipation relations in random Ising systems

Author: Federico Ricci-Tersenghi <ricci@chimera.roma1.infn.it>

$$\lim_{\mu,\mu_0 \to 0} I_1^t(\mu,\mu_0,\phi-\phi_0) = \frac{aF_0}{4(c+1)},$$

Behavior of the reflection function of a plane-parallel medium for directions of incidence and reflection tending to horizontal directions

Author: Daphne Stam <d.m.stam@sron.nl>

$$\lim_{\mu,\mu_0 \to 0} I_1^t(\mu,\mu_0,\phi-\phi_0)$$

Behavior of the reflection function of a plane-parallel medium for directions of incidence and reflection tending to horizontal directions

Instead of a Demo: Applicable Theorem Search in Mizar

```
definition
 let k, n be Ordinal;
 pred k divides n means :Def3: :: MTEST1:def 3
 ex a being Ordinal st n = k *^ a;
 reflexivity
 proof
  let n be Ordinal; :: thesis:
  thus ex a being Ordinal st n = n *^ a ;
```

ATP Proof not found

status: Timeout
Suggest hints, Unification query,

Suggested hints

t73_card_2, t39_ordinal2,

Try SPASS, Export problem to SystemOnTPTP

```
  :: thesis:
 end;
end;
```

Searching the Math Knowledge Space

- Classical Setup: they all work more or less the same:
  - crawl the resources (the Web or a corpus)
  - index the search-relevant information (formulae, words, structures,...)
  - process user queries (via tf/idf, unification, ...)
  - rank/process the hits (needs work!)
- Question: Is this enough for the working Mathematician?
- Answer: depends on what you want.
  - Yes, if we restrict ourselves to what is explicitly written in books, papers, etc.
  - No, if we are looking for "Mathematical Knowledge"! (and I claim we should be)
- **Observation 1** *Mathematical knowledge is induced by combinations of explicitly represented facts.* *(that's why we usually ask humans)*
- **Example 2** Combine mathematical facts (no, we don't need theorem proving!)
  - **Theorem 3.1**: *Idempotent monoids are Abelian.* (from course Algebra I)
  - **Lemma 2**: $(\mathbb{S}, \sharp)$ *is an associative, untial, idempotent magma.* (you just found out)
  - Search for $x\sharp y = y\sharp x$ (Find it as an instance of Theorem 3.1)

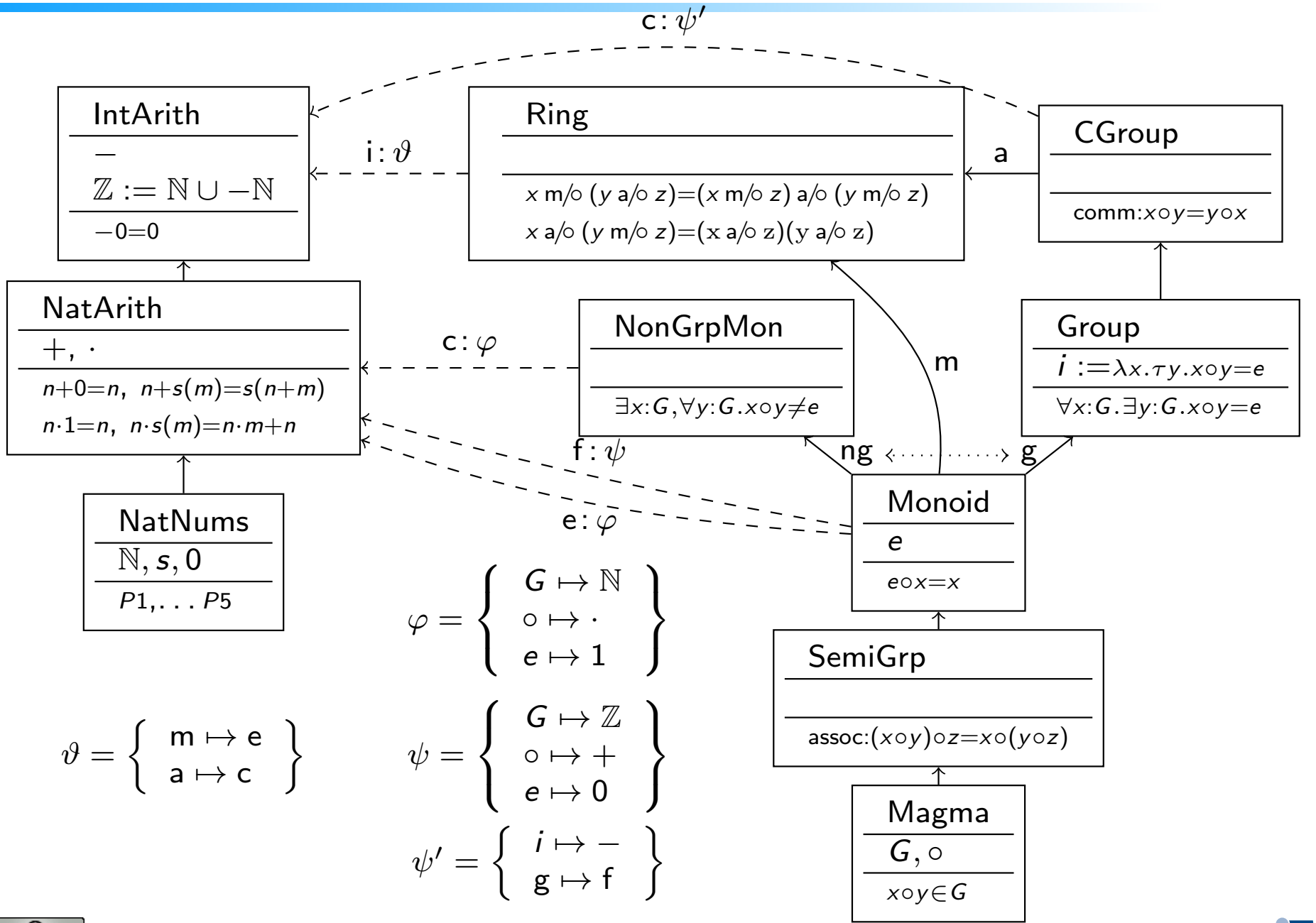
Modular Representation of Mathematics

Modular Representation of Math (Theory Graph)

- Idea: Follow mathematical practice of generalizing and framing
  - framing: If we can view an object a as an instance of concept B, we can inherit all of B properties (almost for free.)
  - state all assertions about properties as general as possible (to maximize inheritance)
  - examples and applications are just special framings.
- Modern expositions of Mathematics follow this rule (radically e.g. in Bourbaki)
- formalized in the theory graph paradigm (little/tiny theory doctrine)
  - theories as collections of symbol declarations and axioms (model assumptions)
  - theory morphisms as mappings that translate axioms into theorems
- **Example 3 (MMT: Modular Mathematical Theories)** MMT is a foundation-indepent theory graph formalism with advanced theory morphisms.

  Problem: With a proliferation of abstract (tiny) theories readability and accessibility suffers (one reason why the Bourbaki books fell out of favor)

Modular Representation of Math (MMT Example)



The MMT Module System

- Central notion: theory graph with theory nodes and theory morphisms as edges
- **Definition 4** In MMT, a theory is a sequence of constant declarations – optionally with type declarations and definitions
- MMT employs the Curry/Howard isomorphism and treats
  - axioms/conjectures as typed symbol declarations (propositions-as-types)
  - inference rules as function types (proof transformers)
  - theorems as definitions (proof terms for conjectures)
- **Definition 5** MMT had two kinds of theory morphisms
  - structures instantiate theories in a new context (also called: definitional link, import)
    they import of theory S into theory T induces theory morphism $S \to T$
  - views translate between existing theories (also called: postulated link, theorem link)
    views transport theorems from source to target (framing)
- together, imports and views allow a very high degree of re-use
- **Definition 6** We call a statement t induced in a theory T, iff there is
  - a path of theory morphisms from a theory S to T with (joint) assignment σ,
  - such that $t = \sigma(s)$ for some statement s in S.
- In MMT, all induced statements have a canonical name, the MMT URI.

Searching for Induced statements

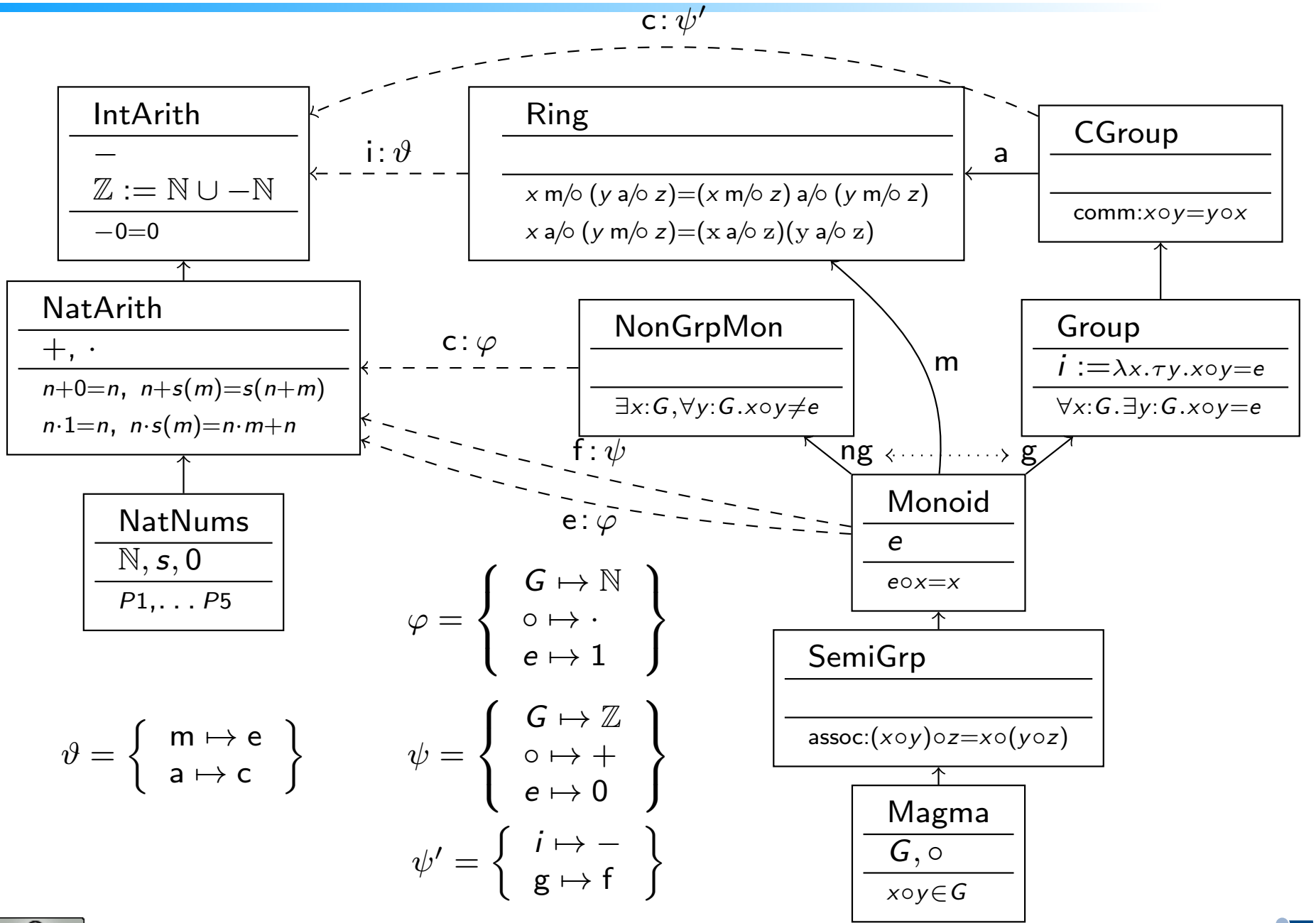
♭search: Indexing flattened Theory Graphs

- Simple Idea: We have all the necessary components: MMT and `MathWebSearch`
- **Definition 7** The ♭search systen is an integration of `MathWebSearch` and MMT that
  - computes the induced formulae of a modular mathematical library via MMT (aka. flattening)
  - indexes induced formulae by their MMT URIs in `MathWebSearch`
  - uses `MathWebSearch` for unification-based querying (hits are MMT URIs)
  - uses the MMT to present MMT URI (compute the actual formula)
  - generates explanations from the MMT URI of hits.
- Implemented by Mihnea Iancu in ca. 10 days (MMT harvester pre-existed)
  - almost all work was spent on improvements of MMT flattening
  - `MathWebSearch` just worked (web service helpful)

♭search User Interface: Explaining MMT URIs

- Recall: ♭search (`MathWebSearch` really) returns a MMT URI as a hit.
- Question: How to present that to the user? (for his/her greatest benefit)
- Fortunately: MMT system can compute induced statements (the hits)
- Problem: Hit statement may look considerably different from the induced statement
- Solution: Template-based generation of NL explanations from MMT URIs. MMT knows the necessary information from the components of the MMT URI.

Modular Representation of Math (MMT Example)



Example: Explaining a MMT URI

- **Example 8** ♭search search result *u*?IntArith?c/g/assoc for query $(\boxed{x} + \boxed{y}) + \boxed{z} = \boxed{R}$.
  - localize the result in the theory *u*?IntArithf with

    Induced statement $\forall x, y, z : \mathbb{Z}.(x + y) + z = x + (y + z)$ *found in* `http://cds.omdoc.org/cds/elal?IntArith` (*subst*, *justification*).
  - Justification: from MMT info about morphism c (source, target, assignment)

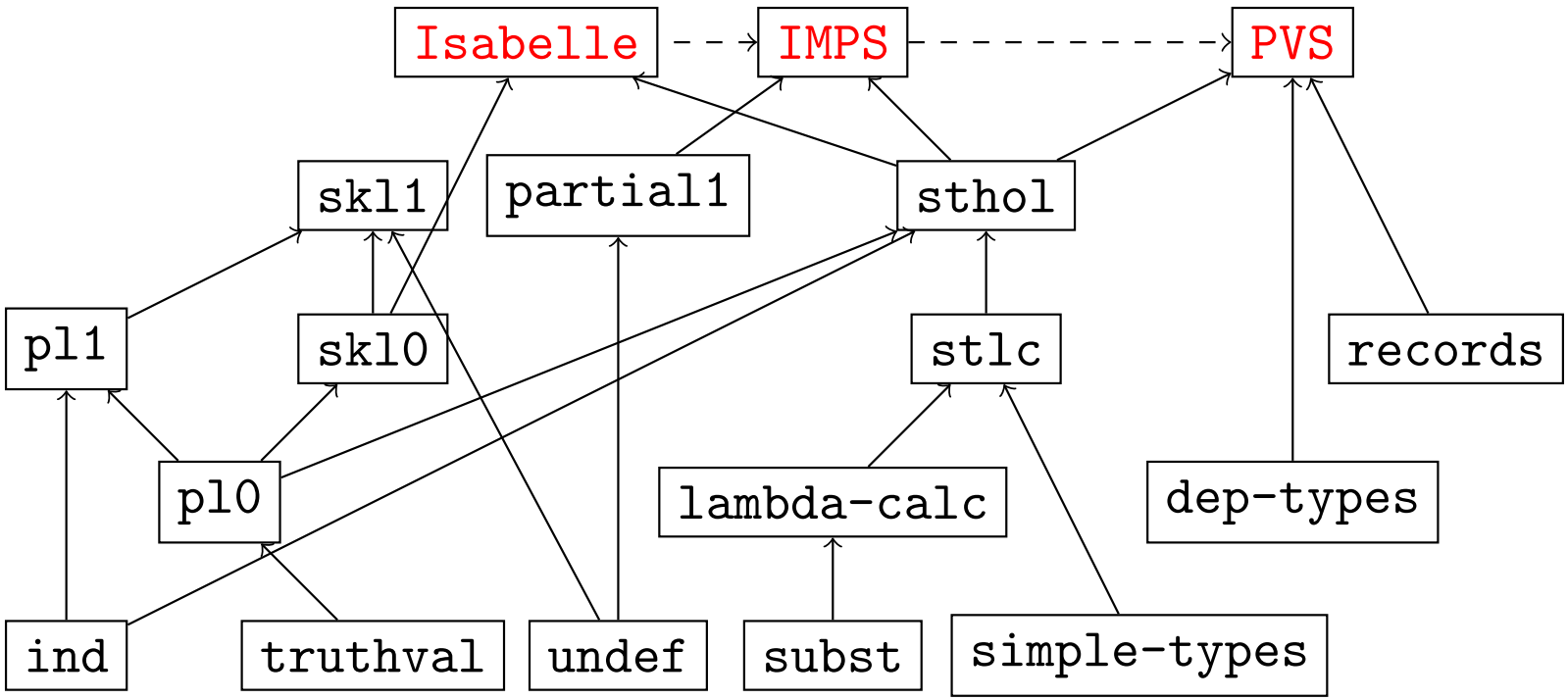
    IntArith is a CGroup if we interpret $\circ$ *as* $+$ *and* G *as* $\mathbb{Z}$.
  - skip over g, since its assignment is trivial and generate

    CGroups are SemiGrps by construction
  - ground the explanation by

    In SemiGrps we have the axiom assoc : $\forall x, y, z : G.(x \circ y) \circ z = x \circ (y \circ z)$

The LATIN Logic Atlas

- **Definition 9** The LATIN project (Logic Atlas and Integrator) develops a logic atlas, its home page is at `http://latin.omdoc.org`.
- Idea: Provide a standardized, well-documented set of theories for logical languages, logic morphisms as theory morphisms.

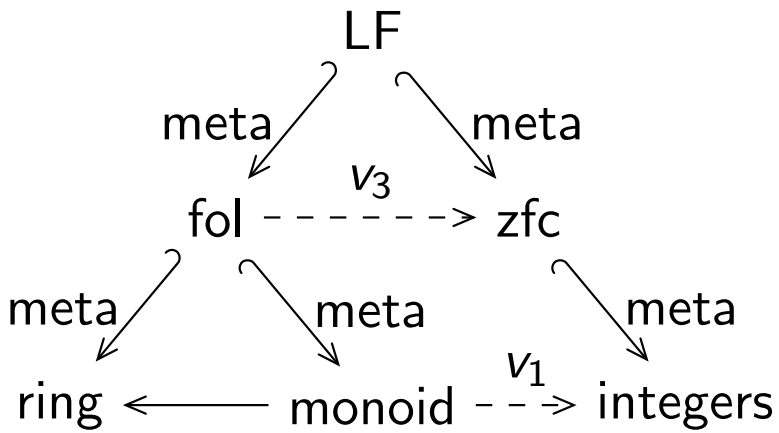


- Technically: Use MMT as a representation language logics-as-theories
- Integrate logic-based software systems via views.

LATIN: Representing Logics and Foundations as Theories

- Logics and Foundations as Theories:

- Logics and foundations represented as theories
- Meta-relation between theories
- Models represented as theory morphisms
- e.g. v_1 interprets monoid in integers using meta-morphism v_3



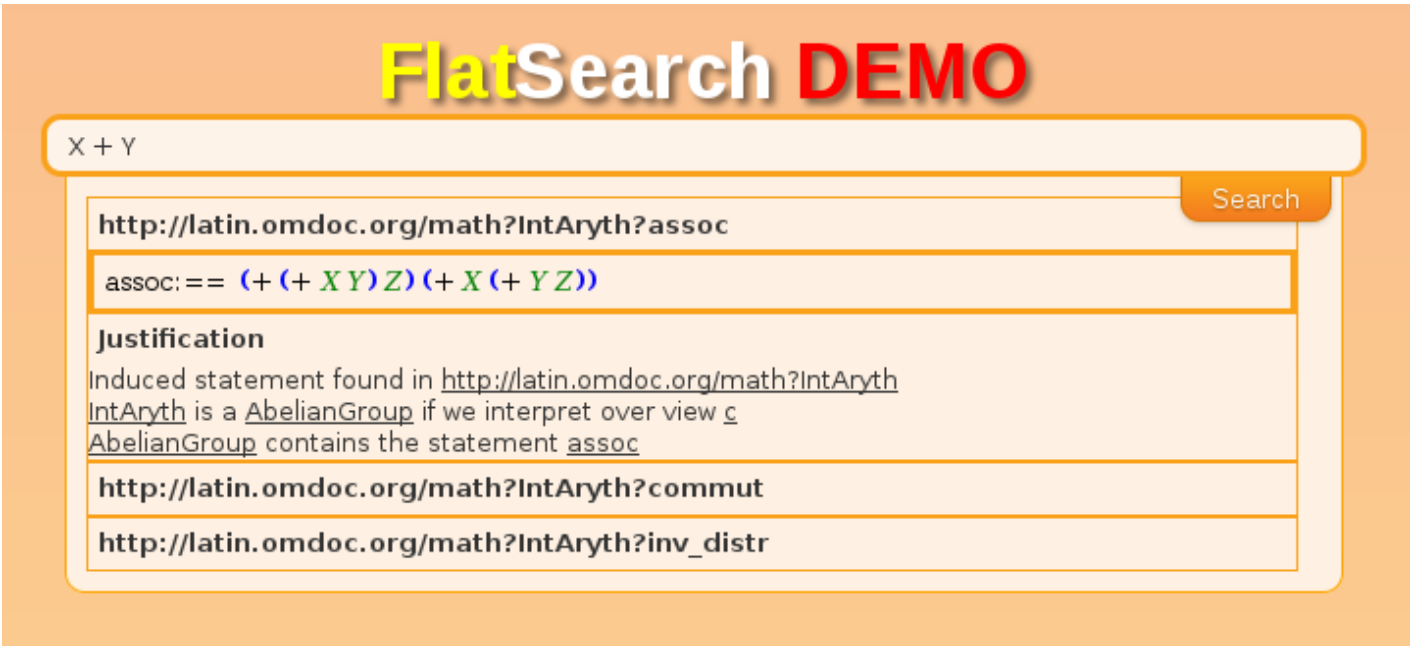
- The LATIN atlas in numbers: it currently contains (tiny theories approach)
  - 449 theories with 2310 symbol declarations (avg. = 5.14 declarations/theory)
  - and 1072 direct imports (including metas) (avg = 2.39 imports/theory)
  - 382 views between theories.
  - Size: 123.9 MB in native OMDoc format

♭search on the LATIN Logic Atlas

- Flattening the LATIN Atlas (once):

type	modular	flat	factor
declarations	2310	58847	25.4
library size	23.9 MB	1.8 GB	14.8
math sub-library	2.3 MB	79 MB	34.3
`MathWebSearch` harvests	25.2 MB	539.0 MB	21.3

- simple ♭search frontend at `http://cds.omdoc.org:8181`



Conclusions and Recap

- From searching documents to searching knowledge spaces!
- ♭search implemented from existing components
  - MMT for modular representations of mathematical knowledge
  - MMT URIs name induced statements
  - flattening to compute all induced statements
  - generate human-oriented explanations of induction paths
- Prototypical implementation for the LATIN logic atlas
- Future work: we have only just begun (most work in MMT though)
  - Flattening away other language features, e.g. patterns ($\rightsquigarrow$ F. Horozal)
  - Avoiding duplication from structures.
  - Integrating graph structure constraints into `MathWebSearch`
  - Extending MMT (and flattening) to informal Math! (redo Bourbaki)