Generalizing theorems of the **mizar** Mathematical Library by type promotion and property omission

Jesse Alama
Center for Artificial Intelligence,
New University of Lisbon, Portugal

2011-09-27

Motivation: making implicit information of computer-assisted proofs explicit

- An experiment involving a large library of formal mathematics to bring out unarticulated inferences.
- Looking for cases where information that is implicitly supplied by an ITP is actually not needed.
- Applications:
  - Discovering necessary and sufficient conditions for theorems; reverse mathematics 'in the small'.
  - Library refactorings.
  - AI: Inventing generalizations of theorems/definitions.

Constructor properties in **mizar**

'Constructor' is a **mizar** term that covers the concepts **function/constant**, **relation**, **structure**, ...

Property	**Definition**	**Property**	**Definition**
reflexivity	$R(x,x)$	projectivity	$f(f(x)) = f(x)$
symmetry	$R(x,y) \to R(y,x)$	involutiveness	$f(f(x)) = x$
asymmetry	$R(x,y) \to \neg R(y,x)$	idempotence	$f(x,x) = x$
connectedness	$R(x,y) \vee R(y,x)$	commutativity	$f(x,y) = f(y,x)$
irreflexivity	$\neg R(x,x)$		

(Many more constructor properties could be implemented, e.g., transitivity, but currently are not.)

Eliciting dependence upon constructor properties

- Decompose the 1100+ articles of the **mizar** Mathematical Library, yielding about 100,000 self-standing 'microarticles'.
- Verification of an item is a two-step process:
  1. Construct the environment in which the verification proper will be carried out. (Run BibTEX, creating a `.bbl` file that will be used later.)
  2. Carry out the verification proper of the article with respect to the constructed environment. (Run LATEX with the `.bbl` file.)
- Step (1) imports all properties of all imported constructors. (`\nocite{*}`)
- Between step (1) and (2), omit a constructor property. (Delete an entry from a `.bbl` file 'behind the back' of LATEX, then see whether one's document compiles cleanly.)

Direct and indirect needs

Definition: A **mizar** item A directly depends on a **mizar** item B if **mizar**-verification of A (as a proof text) fails in the absence of B. **Indirect dependence** is the transitive closure of the direct dependence relation.

Definition: A **mizar** item I **directly needs** property P of constructor C iff **mizar**-verification of I (as a proof text) fails when P is detached from C. **Indirect needs** is the transitive closure of the direct need relation.

Most needed properties

Property	Directly Needed	Indirectly Needed
reflexivity	54113	102426
symmetry	29744	97220
asymmetry	256	82585
connectedness	5020	83083
irreflexivity	91	65951
projectivity	153	10002
involutiveness	533	67853
idempotence	535	70132
commutativity	14055	92580

Statistics: Reflexivity

- Reflexivity (of some constructor) is directly needed by nearly half of the items of the **mizar** Mathematical Library. Reflexivity (of some constructor) is indirectly needed by nearly the entire library.
- Reflexivity of **equality of sets** $=$ (a built-in notion in **mizar**) accounts for most of this.
- The next most important reflexive constructor is subset inclusion $\subseteq$, which is indirectly needed by 93284 items. A redefinition of subset inclusion for ordinals is indirectly used by 8279 items.
- The most important reflexive ``mathematical'' relation is the less-than-or-equal-to relation $\leq$ on (extended) real numbers, whose reflexivity is indirectly needed by 67196 items.

Statistics: Irreflexivity

- Irreflexivity is directly needed by only a handful of items in the library, but indirectly it supports about 2/3 of the library.
- The explanation is the proper subset relation $\subset$: the irreflexivity of this constructor is needed by 65546 items.
- The most important ``mathematical'' example is the relation of one element of a relation being strictly less than another.

Statistics: Asymmetry

- Asymmetry is attached to only five constructors in the entire mml:
  - $\in$ (set membership);
  - the proper subset relation;
  - a variant of $\in$, defined for many-sorted set structures;
  - the strictly-lexicographically-less-than relation on finite tuples of natural numbers;
  - the strictly-lexicographically-less-than relation on bags of ordinals.
- The asymmetry of $\in$ (which expresses a weak form of the axiom of foundation) accounts for essentially all items that need the asymmetry of any constructor: 82581 items indirectly need this weak form of foundation, whereas only 283 items directly depend on this property of $\in$.

Statistics: Projectivity

- Projectivity is rarely directly needed, but supports a substantial piece of the library.
- The projective constructor that accounts for nearly all of this: the closure operation defined on subsets of a topological space is indirectly needed by 7536 items.

Statistics: Connectedness

- Connectedness is attached to very few constructors of the **mizar** Mathematical Library, but the constructors to which this property is attached have a significant influence across the mml.
- The constructor whose connectedness is used indirectly by the greatest number of items: the subset relation $\subseteq$, restricted to ordinals.
  - The connectedness of this constructor expresses a rather significant fact about ordinals (any two ordinals are comparable).
  - The connectedness of the subset relation on ordinals is indirectly needed by 82490 items.
- The next most significant example is $\leq$ on rational numbers, indirectly needed by 71313 items.

Statistics: Involutiveness

- Two functions compete for being the most influential involutive constructor
  - The sign-changing operation $x \quad -x$ on real numbers is needed by 65501 items.
  - The reciprocal operation $z \quad 1/z$ on complex numbers is needed by 65105 items.
- The constructor with the next highest number of items that indirectly depend on its involutiveness is the relative complement operation $X\ Y$ on sets, which is indirectly needed by 8847 items.

Statistics: Idempotence

- The constructor whose idempotence is most frequently indirectly needed is the binary union $\cup$ of two sets, which is indirectly needed by 69184 items.
- The idempotence of binary set intersection $\cap$ takes second place: it is indirectly needed by 24249 items.

Maximally exploiting properties (indirectly)

- Several items in the **mizar** Mathematical Library indirectly depend on more than 100 constructor properties.
- The item with the greatest number of dependencies is taken from the proof of the Jordan curve theorem.
- This is perhaps no surprise, since the formal development leading to this theorem required a considerable amount of work and counts as a landmark of the mml.
- Many other items in the path toward the Jordan curve theorem likewise make heavy use of constructor properties: in the items leading to the Jordan curve theorem, on average several dozen constructor properties are indirectly needed.

Directly exploiting properties: example

- The **mizar** article `BROUWER`, devoted to the proof of Brouwer's fixed-point theorem for plane disks, contains a theorem that, among all theorems of the **mizar** Mathematical Library, implicitly exploits the greatest number of constructor properties:
- This theorem directly needs 14 constructor properties.

Brouwer example

```
for a being real number for r being non negative real number
for n being non empty Element of NAT
for s, t, o being Point of (TOP-REAL n)
for S, T, O being Element of REAL n st
  S = s & T = t & O = o
 holds
  (s is Point of Tdisk(o,r) & t is Point of Tdisk(o,r) &
   s <> t &
   a = ((-|(t-s,s-o)| + sqrt(|(t-s,s-o)|^2 -
        (Sum sqr (T-S)) * (Sum sqr (S-O)-r^2)))
        / Sum sqr (T-S))
  implies
  HC(s,t,o,r) = (1-a)*s+a*t
```

Using the information that a property of a constructor is (un)necessary

- When we learn that a property P of a constructor C is not needed for a theorem ϕ, we have, prima facie, a kind of generalization of ϕ.
- But what kind of generalization?
  - Suppose we find that a theorem does not require commutativity of $+$ on real numbers.
  - We may say, axiomatically, that we can prove ϕ from a set Γ of axioms that does not include commutativity of $+$.
  - We may or may not be interested in the range of models of Γ.
- It is one thing to find out that commutativity of $+$ on real numbers isn't needed for a particular theorem; it's another to find out that commutativity of group addition is not needed for a theorem about groups.

Future work

- Visualization/exploration of constructor property (in)dependencies.
- Design an advisor that informs one about unneeded constructor properties, and even suggests generalizations.
- Use ATPs to give another view of our data (as well as provide a way of testing our data), e.g., with countermodels.
- Carry out analogous experiments with other ITPs and libraries (Coq, Agda, Isabelle, etc.). The information uncovered would likely be quite different from the **mizar** data.

Literature

- A. Arana, "On formally measuring and eliminating extraneous notions in proofs", **Philosophia Mathematica 17** (2009), 189-207
- **The mizar homepage**
- **Grabowski, A. and Kornilowicz, A. and Naumowicz, A., "mizar in a Nutshell", in** *Journal of Formalized Reasoning* **3(2), pp. 153–245, 2010.**
- J. Urban, "MPTP—Motivation, implementation, first experiments", **Journal of Automated Reasoning**, **33**(3-4) (2004), pp. 319–339
- **The mizar-items site**