

mizar-items: Exploring fine-grained dependencies in the Mizar Mathematical Library

Jesse Alama^{1*}

Center for Artificial Intelligence
New University of Lisbon
j.alama@fct.unl.pt

The MML is one of the largest collection of formalized mathematical knowledge that has been developed with various interactive proof assistants. It comprises more than 1100 “articles” summing to nearly 2.5 million lines of text, each consisting of a unified collection of mathematical definitions and proofs. Semantically, it contains more than 50000 theorems and more than 10000 definitions expressed using more than 7000 symbols. It thus offers a fascinating corpus on which one could carry out a number of experiments. This note discusses a system for computing fine-grained dependencies among the contents of the MML. For an overview of Mizar, see [3]; for a discussion of some successful initial experiments carried out with the help of mizar-items, see [1,2].

We say that a definition, or a theorem, ϕ *depends* on some definition, lemma or other theorem ψ , (or equivalently, that ψ is a *dependency* of ϕ) if ϕ “needs” ψ to exist or hold. The main way such a “need” arises is that the well-formedness or the justification of provability does not hold in the absence of ψ . Other senses of mathematical “dependency” are available that are related to what we describe here, but which mizar-items does not treat (at present). One might be interested, for example, in the space of all *logically possible* proofs of a certain result. Our interest is, to put it philosophically, intensional rather than extensional: we are interested in computing what minimally accounts for the success of a *specific* mathematical proof that has been formalized in the Mizar language. The extensional problem is what we are after, in the long run, but since we must work with specific formalizations of mathematical knowledge, we need to take an intensional approach.

The primary motivation behind mizar-items was the lack of a tool for giving a complete answer to the question of what, precisely, a Mizar text depends upon. This turns out to be rather non-trivial task. The difficulty stems primarily from various mechanisms (such as type inference) for making Mizar texts “smoother” for the author and human consumer because these mechanisms, by suppressing inferences—sometimes trivial, other times mathematically significant—can be “exposed” only through much computation.

* The author was supported by the FCT project “Dialogical Foundations of Semantics” (DiFoS) in the ESF EuroCoRes programme LogICCC (FCT Log-ICCC/0001/2007). The author thanks Artur Kornłowicz, Karol Pąk and Josef Urban for offering their Mizar expertise. The final publication is available at www.springerlink.com.

Naturally, not all items in the vast Mizar library are equally interesting. `mizar-items` and its accompanying website (see below) was motivated by the problem of discovering dependency information not for arbitrary Mizar items, but specifically for those with substantial mathematical or foundational value, such as the Jordan curve theorem, the axiom of choice, the existence of strongly inaccessible cardinal numbers, or Euler’s polyhedron formula (to name only a biased handful of examples). The fine-grained dependency data exposed by `mizar-items` could also be used in theory exploration and reverse mathematics [5] or Lakatos-style [4] investigations of necessary and sufficient conditions for mathematical theorems.

We compute the fine-grained dependency graph for the MML by starting with an over-approximation of what is known to be sufficient for an item to be Mizar-verifiable and then successively refining this over-approximation toward a minimal set of sufficient conditions. The method can be fairly characterized as brute-force: for each Mizar item, we successively hide implicit information normally kept hidden from a human Mizar formalizer, then see whether Mizar can still verify it. It turns out that this approach is rather slow; we needed to develop various heuristics to make the brute-force computation smarter.

`mizar-items` is accompanied by a website,

<http://mizar.cs.ualberta.ca/mizar-items/>

for exploring these dependencies. With the site one can view any particular Mizar item and see precisely what it depends upon (and what depends on the item). With the dependency graph, one can ask such queries as: *Is there a path between two given items? Do all paths from one item to another pass through a given intermediate node? Are there any paths between two given items that do not pass through a given node?*

To facilitate exploration, one can start by visiting a list of selected interesting entry points into the vast Mizar library.

`mizar-items` is a collection of programs in Common Lisp, Perl, Pascal, as well as shell scripts. The code is available online at

<https://github.com/jessealama/mizar-items>

(The Pascal sources are not included here: They are part of the Mizar code base, which, at present, is not publicly available.)

References

1. Alama, J., Kuehlwein, D., Tsvitsivadze, E., Urban, J., Heskes, T.: Premise selection for mathematics by corpus analysis and kernel methods. [preprint], submitted
2. Alama, J., Mamane, L., Urban, J.: Dependencies in formal mathematics. [preprint], submitted
3. Grabowski, A., Kornilowicz, A., Naumowicz, A.: Mizar in a nutshell. *Journal of Formalized Reasoning* 3(2), 153–245 (2010), <http://jfr.cib.unibo.it/article/view/1980/1356>

4. Lakatos, I.: Proofs and Refutations. Cambridge University Press (1976)
5. Simpson, S.G.: Subsystems of Second Order Arithmetic. Perspectives in Mathematical Logic, Springer, 2nd edn. (2009)