

YAX: Yet Another Cross Referencer

Name: Xuheng Li(UNI: xl2784)

November 22, 2021

A cross referencing tool, or commonly known as cross referencer, is a software that indexes source code and provides information for symbols and definitions on a given code base such that the user can find where a symbol is defined or used in that code base. The cross referencer is widely used in software development and integrated into IDEs and editors like vscode or vim. A widely used cross referencer on Linux, Cscope [1], is single threaded and thus suffered on the performance when building its database on a large code base such as Linux kernel source code. Since parsing the symbols on a single source code file is usually independent from the rest of the files for the given code base, the performance of building the database can be improved by parallelizing the building procedure. Therefore, I propose **YAX: Yet Another Cross Referencer**, a parallelized version of Cscope written in Haskell. Given the time constraint, YAX will only work on C99 [3] source code, other languages, including assembly language and C in other standard, will not but supported.

Alex [5] and Happy [6] are the Haskell counterpart for Lex and YACC [4] for C, respectively. language-c [2] is a Haskell package provided by Hackage that leverages Alex and Happy to translate a C source code into a `CTransUnits` and form ASTs. YAX can then analyze `CTransUnits`, extract symbols and add symbols together with necessary information to the database. More specifically, the following symbols will be added to the database and indexed:

- Global variable definitions and references
- Function definitions and calls
- Composite data type definitions, including `struct`, `union` and `enum`
- String literals

Key words and local variables are omitted. The location of each reference of each symbol, including the file path and the line number where the symbol is referenced, is stored to a balanced search tree and stored to a file. YAX is able to load that file and look up symbols. The procedure for building the balanced search tree will be parallelized so performance improvement compared against Cscope is expected on YAX.

References

- [1] *Cscope Home Page*. 2012. URL: <http://cscope.sourceforge.net/>.
- [2] Joe Hermaszewski. *language-c: Analysis and generation of C code*. 2020. URL: <https://hackage.haskell.org/package/language-c-0.9.0.1>.
- [3] ISO. *ISO C Standard 1999*. Tech. rep. 1999. URL: <http://www.open-std.org/jtc1/sc22/wg14/www/docs/n1124.pdf>.
- [4] John R Levine et al. *Lex & yacc.* " O'Reilly Media, Inc.", 1992.
- [5] Simon Marlow. *Alex: A lexical analyser generator for Haskell*. URL: <https://www.haskell.org/alex/>.

- [6] Simon Marlow. *Happy: The Parser Generator for Haskell*. URL: <https://www.haskell.org/happy/>.