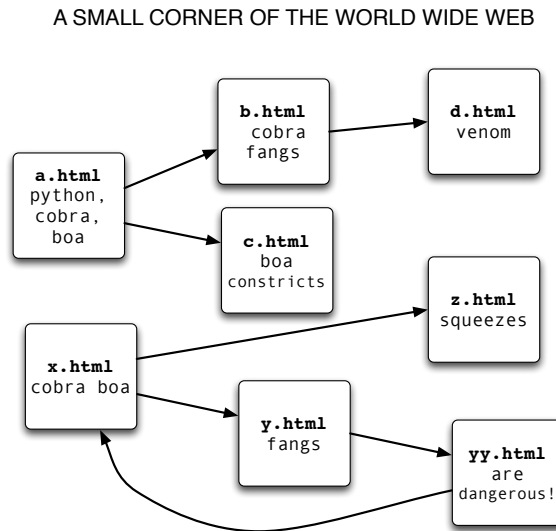


# 1 Crawling the web

In class today we will implement a *version* of the `crawl_web` function that is included on your homework. The code will be available on the moodle version of this handout.

We will focus on a tiny corner of the web that looks like this. These pages exist on the web. For example, `a.html` can be found at `http://cs.colgate.edu/~mhay/cosc101/webdemo/a.html`. This is a **different** set of web pages than the ones that you will crawl for the homework assignment.



The version we will write in class is **not** complete. To complete it for your homework, you will need to modify it in a couple of ways:

- Replace the stub functions, which only work for the toy example in class, with properly defined functions that *actually* download the page and extract links.
- Modify `crawl_web` to not only build the web graph but to also build the reverse index. See the homework description for details.

## 2 Exercises

1. Implement the `crawl_web` function.