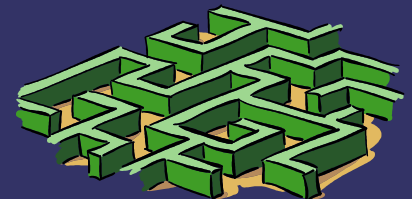


Swish-e Lightning Talk

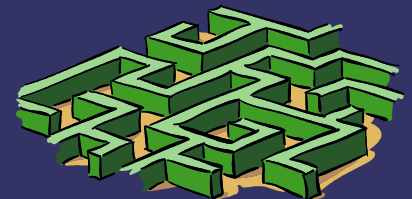
For YAPC::NA 2008

Josh Rabinowitz



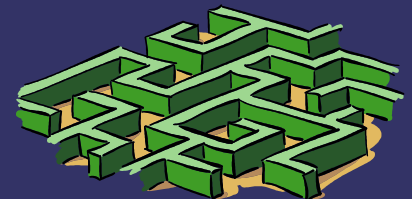
Swish-e: Overview

- ➔ We'll cover basic usage and mention the Perl API
- ➔ We'll also mention current and future development
- ➔ (For the Future: Can you say Swish3?)



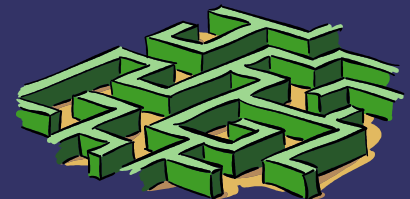
Swish-e: What It Is

- ➔ Makes flat-file indexes for performing full-text searches
- ➔ Searches are very fast
- ➔ Building indexes... Not so fast



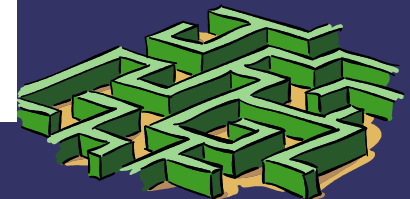
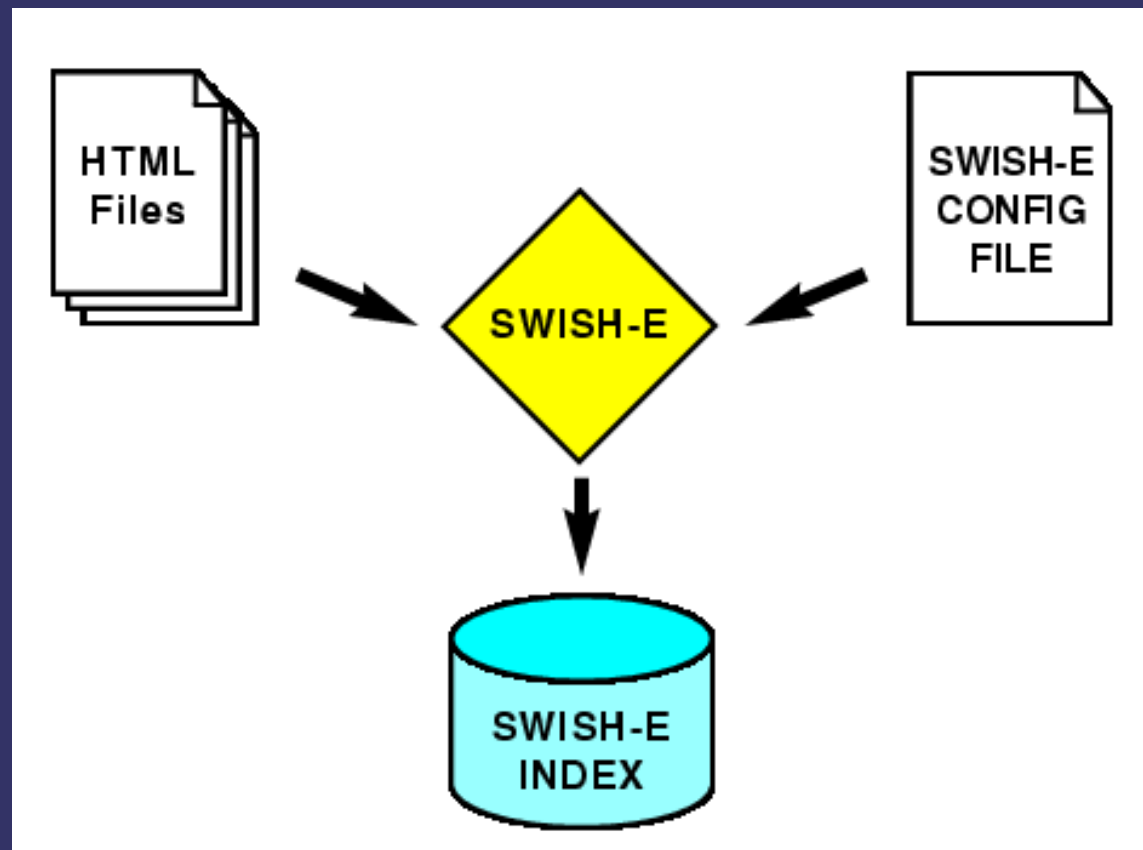
Swish-e: Basic Usage

- ➔ Assuming howto html files to index in ./html
- ➔ # howto-html.conf
- ➔ IndexDir ./htmls/
- ➔ IndexOnly .html
- ➔ IndexFile ./howto-html.index
- ➔ Then,
% swish-e -c howto-html.conf



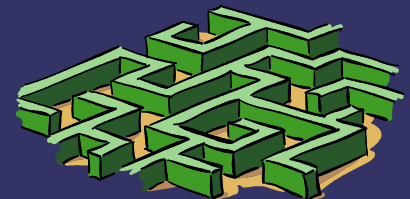
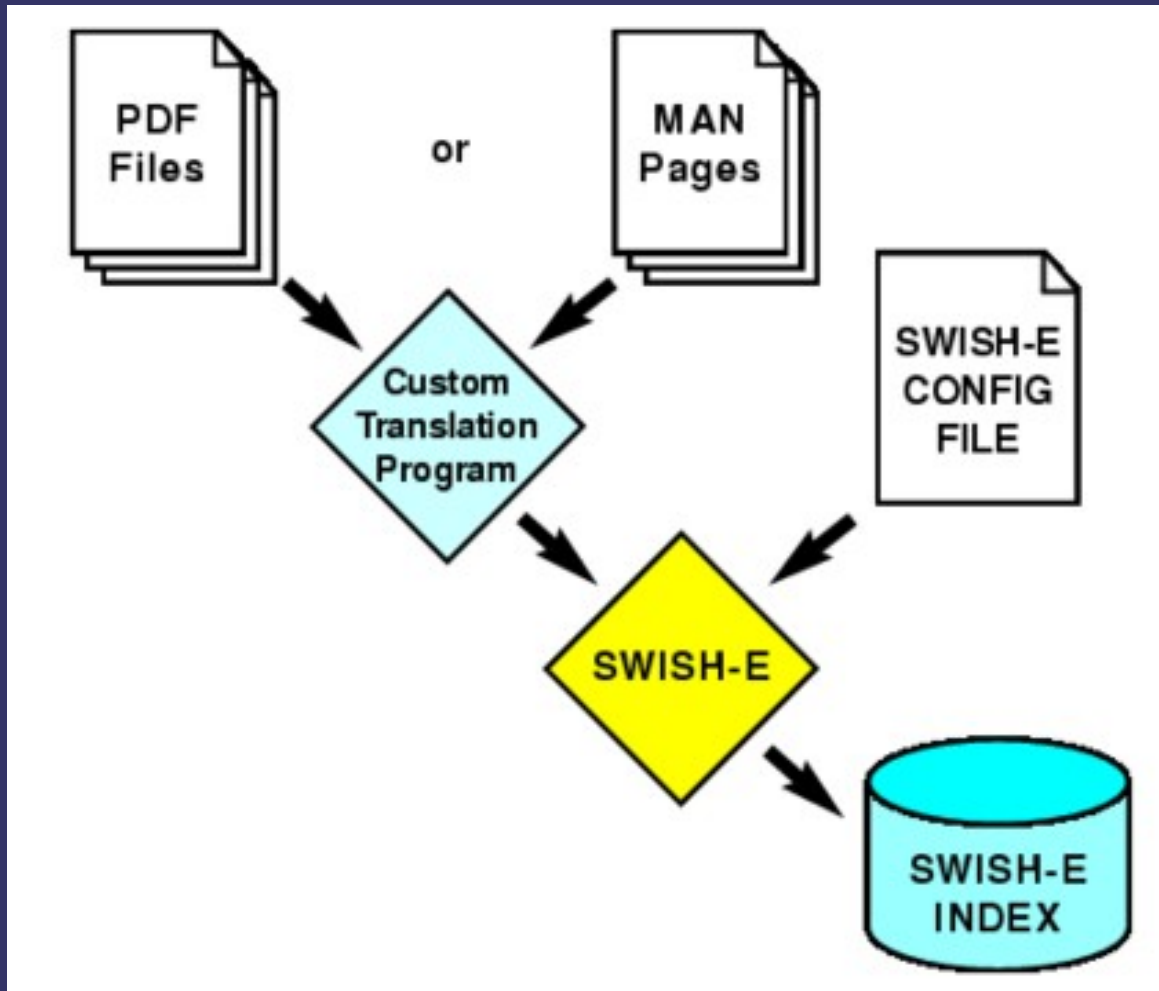
Swish-e: A Pretty Picture of Building a Basic Index

- ➔ A pretty picture!



Swish-e: Building Indexes With “External Programs”

➔ Another Pretty Picture:

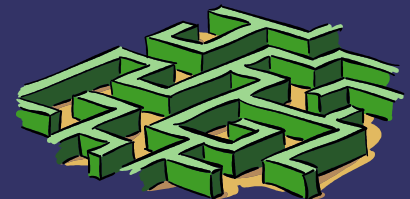


Swish-e: Searching from the Command Line

➔ `% swish-e -f howto-html.index -m 3 -w nfs`

➔ This returns (abridged and reformatted):

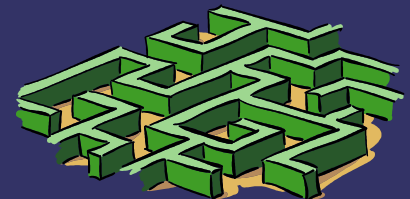
➔ `1000 ../HOWTO-htmls/NFS-HOWTO/performance.html`
➔ `"Optimizing NFS Performance" 33288`
➔ `998 ../HOWTO-htmls/NFS-HOWTO/intro.html`
➔ `"Introduction" 10966`
➔ `993 ../HOWTO-htmls/NFS-HOWTO/security.html`
➔ `"Security and NFS" 35968`



Swish-e: Query Language

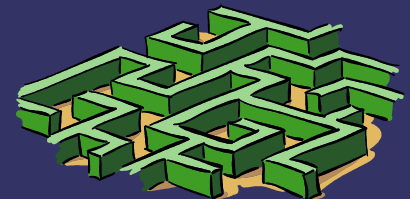
- ➔ SWISH-E supports a full-featured text retrieval search language with syntax including AND, OR, NOT and parenthetic grouping that all work predictably.

- ➔
- ➔ `% swish-e -f howto-html.index -w nfs AND tcp`
- ➔ `% swish-e -f howto-html.index -w nfs OR tcp`
- ➔ `% swish-e -f howto-html.index \`
- ➔ `-w '(gandalf AND frodo) OR (lord AND rings)'`



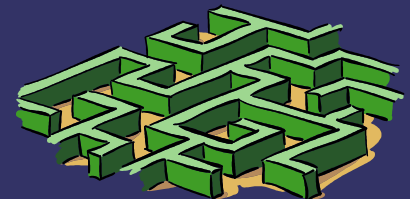
Swish-e: Perl API

- ➔ Just like you'd expect.
- ➔ Open, Search, Iterate over results
- ➔ Note that you cannot create indexes through the API. You must use the Swish-e exe (and perhaps your own perl external program too)



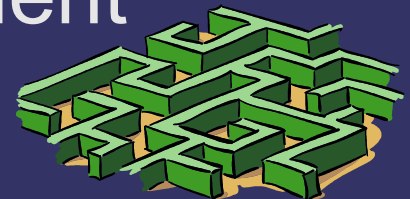
Swish-e: Shortcomings

- ➔ Yes, this means Swish-e doesn't handle these things now. But wait for the next slide:
- ➔ No UTF8
- ➔ No index updates (you must rebuild the whole index)
- ➔ Not platform (nor even word-size) independent



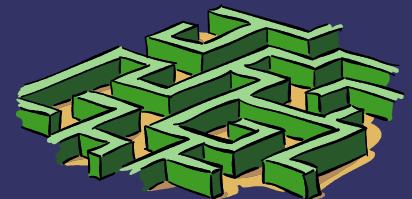
Swish-e: The Future

- ➔ Peter Karman has taken up the gauntlet
- ➔ Swish-e? No, Swish3!
- ➔ UTF8! Multiple backends!
- ➔ I've been working on getting it into Fedora. Did get it into fink and debian! Also wrote Swishetest. See rpmbuild.joshr.com
- ➔ Still probably not platform independent indexes (wah)



Swish-e:

- ➔ Peter Karman has taken up the gauntlet
- ➔ Swish-e? No, Swish3!
- ➔ UTF8! Multiple backends!
- ➔ Still probably not platform independent indexes (wah)



Swish-e: Thanks!

- ➔ I've written two papers on this subject, one for USENIX and one in Linux Journal.
- ➔ Both can be found at josh.com/src/docs
- ➔ *Thanks for attending!*

