PHP at Yahoo!

http://public.yahoo.com/~radwin/

Michael J. Radwin

October 20, 2005

Outline

- Yahoo!, as seen by an engineer

- Choosing PHP in 2002

- PHP architecture at Yahoo!

YAHOO!

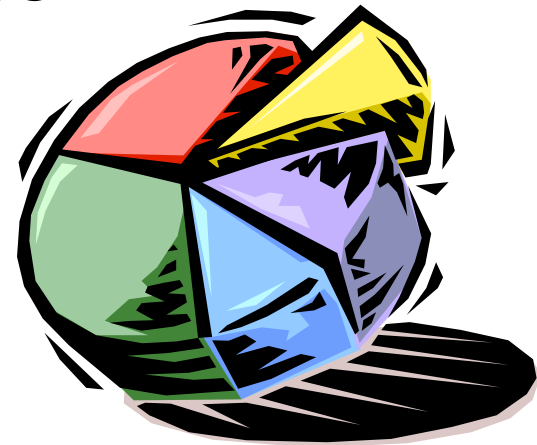# The Internet's most trafficked site

25 countries, 13 languages

Yahoo! by the Numbers

- 411M unique visitors per month

- 191M active registered users

- 11.4M fee-paying customers

- 3.4B average daily pageviews

October 2005

YAHOO!

YAHOO! We Value...

Our mission is to be the most essential global Internet service for consumers and businesses. How we pursue that mission is influenced by a set of core values - the standards that guide interactions with fellow Yahoos, the principles that direct how we service our customers, the ideals that drive what we do and how we do it. Many of our values were put into practice by two guys in a trailer some time ago; others reflect ambitions as our company grows. All of them are what we strive to achieve every day.

Excellence:

We are committed to winning with integrity. We know leadership is hard won and should never be taken for granted. We aspire to flawless execution and don't take shortcuts on quality. We seek the best talent and promote its development. We are flexible and learn from our mistakes.

Teamwork:

We treat one another with respect and communicate openly. We foster collaboration while maintaining individual accountability. We encourage the best ideas to surface from anywhere within the organization. We appreciate the value of multiple perspectives and diverse expertise.

Innovation:

We thrive on creativity and ingenuity. We seek the innovations and ideas that can change the world. We anticipate market trends and move quickly to embrace them. We are not afraid to take informed, responsible risk.

Community:

We share an infectious sense of mission to make an impact on society and empower consumers in ways never before possible. We are committed to serving both the Internet community and our own communities.

Customer Fixation:

We respect our customers above all else and never forget that they come to us by choice. We share a personal responsibility to maintain our customers' loyalty and trust. We listen and respond to our customers and seek to exceed their expectations.

Fun:

We believe humor is essential to success. We applaud irreverence and don't take ourselves too seriously. We celebrate achievement. We yodel.

YAHOO!

Engineering Values

1. Security & Privacy
    –   We must protect our customers' information

2. High Availability
    –   If the site is offline, we're missing the opportunity to serve our customers

3. Performance
    –   We serve billions of pageviews a day

4. Flexibility & Innovation
    –   Customize site for each market
    –   Rapid development of new features

From Proprietary to Open Source

94 95 96 97 98 99 00 01 02 03 04 05

Web Server

Apache

"Filo Server"

DB

Flat Files

php

Web Lang

yScript

YAHOO!
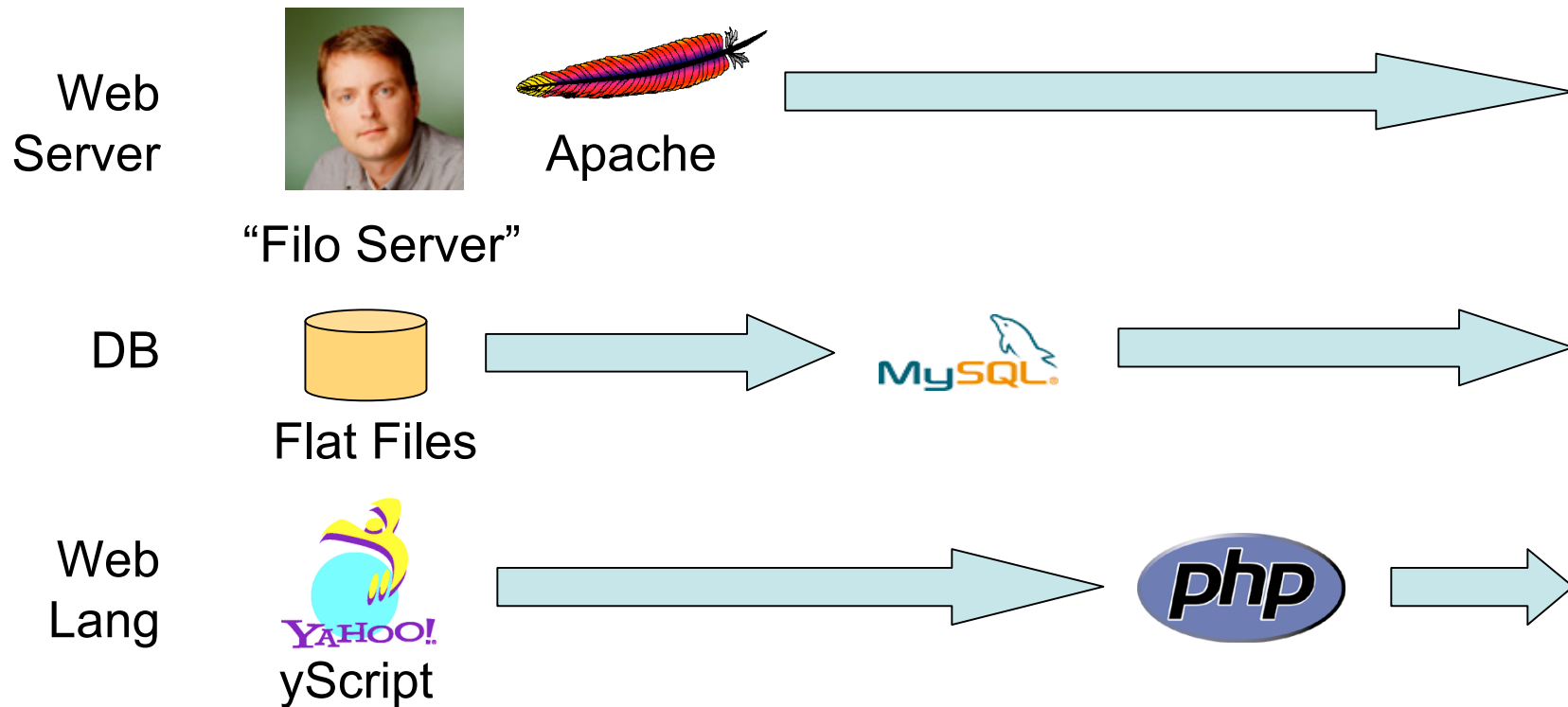
Choosing a Language

How and Why We Selected PHP

Choosing PHP: brief history

- October 2001: 3 proprietary languages
  - Costly to continue to maintain each
  - Limited features (no subroutines!)
- Committee began researching
  - Compare features, performance
  - Build vs. Buy vs. Open Source
- PHP selected May 2002

YAHOO!

Ideal Language Criteria

1. High performance
2. Robust, sand-boxed
3. Language features
   - Loops, conditionals
   - Complex data-types
4. C/C++ extensions
5. Runs on FreeBSD

8. Interpreted or dynamically compiled
9. i18n support
10. Clean separation of presentation/content/ app semantics
11. Low training costs
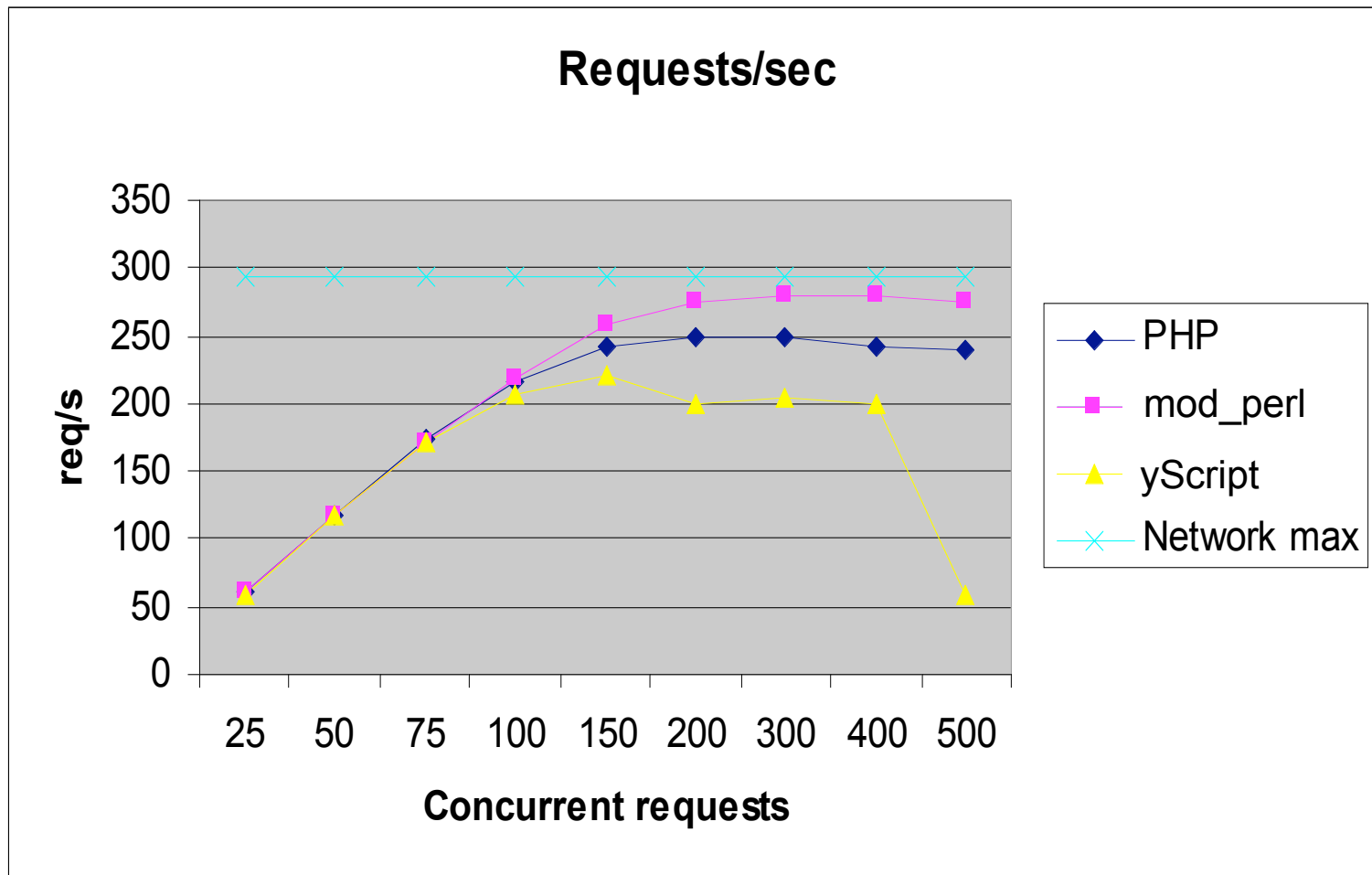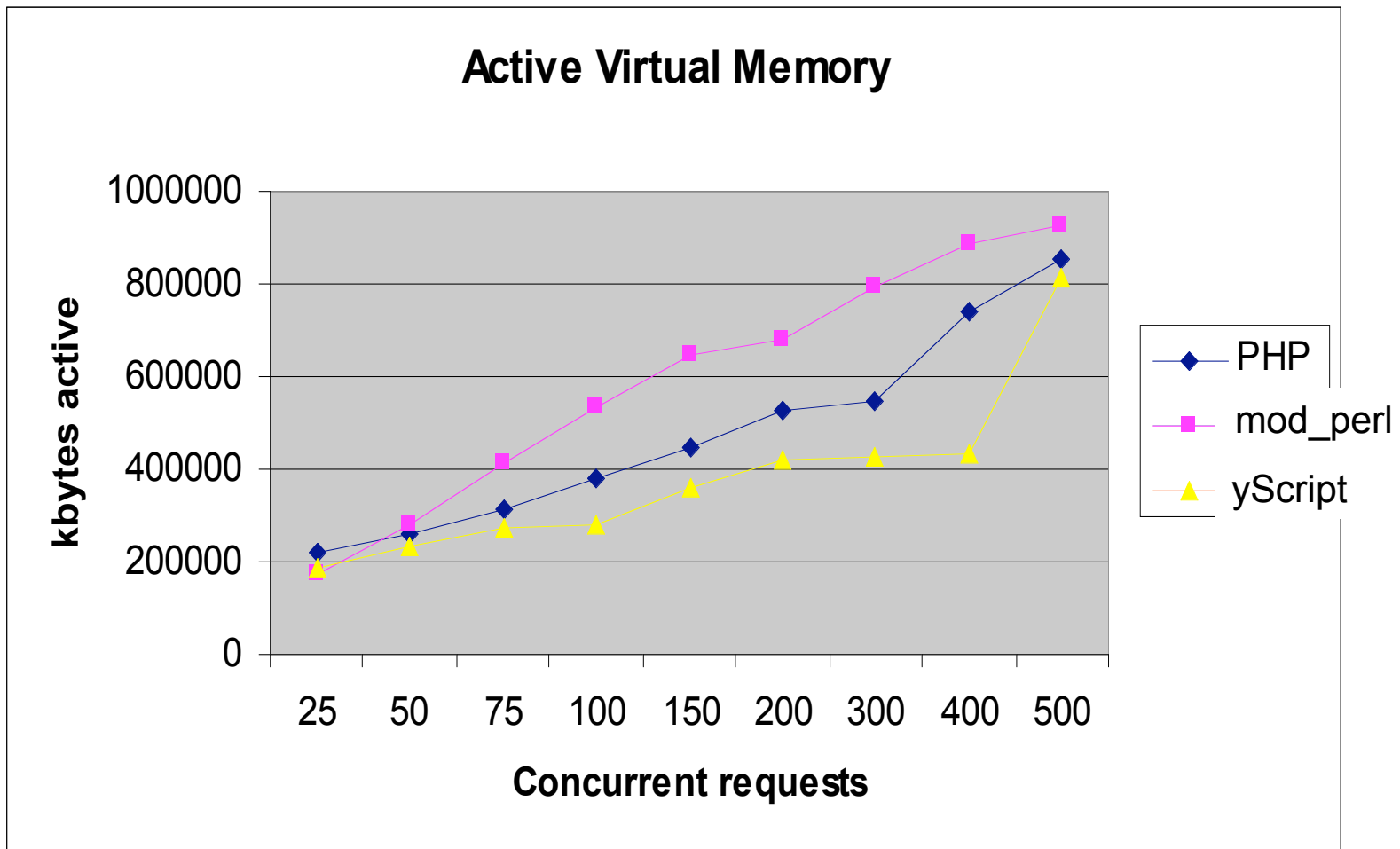12. Doesn't require CS degree to use

Top 10 Language Choices

yScript

mod_include

XSLT

Performance: Requests

Requests/sec

YAHOO!

Performance: Memory

Active Virtual Memory

Why we picked PHP

1. Designed for web scripting

2. High performance

3. Large, Open Source community
   - Documentation, easy to hire developers

4. "Code-in-HTML" paradigm

```
<html>
<?php echo "Hello World"; ?>
</html>
```

5. Integration, libraries, extensibility
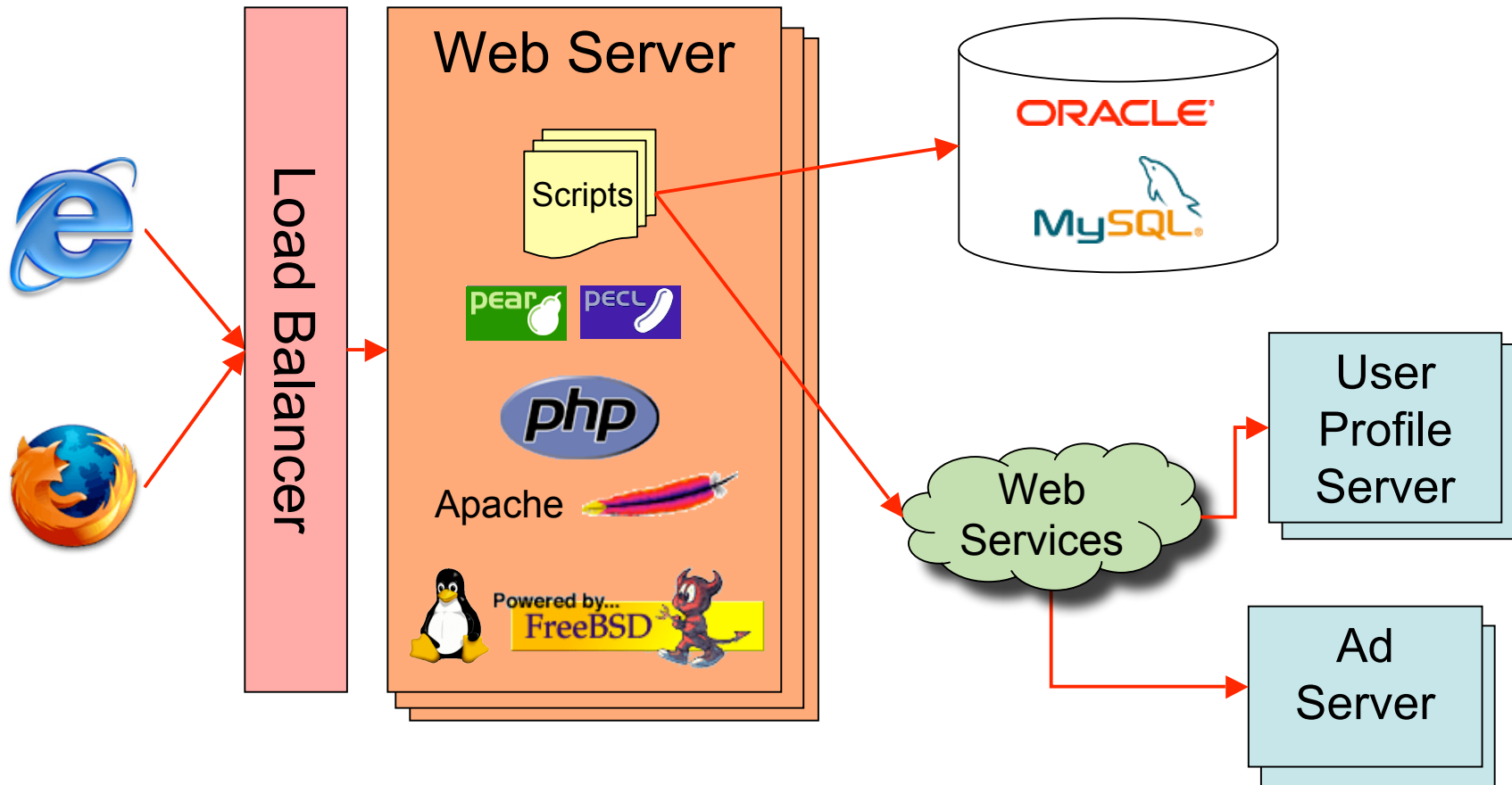
6. Tools: IDE, debugger, profiler

PHP at Yahoo! Today

Yahoo!'s Development Methodology

- Server Architecture

- File Layout

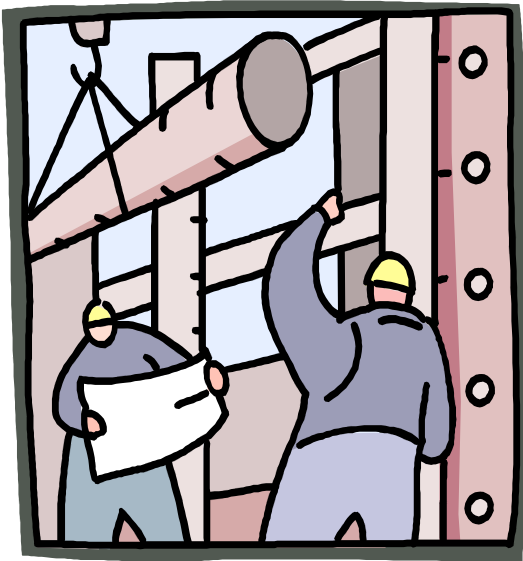- Dependency Management

- Security

- Performance

- Globalization

YAHOO!

Server Architecture

YAHOO!

File Layout

HTML Templates	95% HTML
/usr/local/share/htdocs/*.php	5% PHP

Template Helpers	50% HTML
/usr/local/share/htdocs/*.inc	50% PHP

Business Logic	0% HTML
/usr/local/share/pear/*.inc	100% PHP

C/C++ Core Code	0% HTML
Data access, Networking, Crypto	0% PHP

YAHOO!

Dependency Management

- Base PHP package depends only on XML parser

  ```
  ./configure --disable-all
  ```

- Self-Contained Extensions

  – mysql, dba, curl, ldap, pcre, gd, iconv

  – To enable

    1. Install `/usr/local/lib/php/20020429/mysql.so`

    2. Add "`extension = mysql.so`" to `php.ini`

  – Avoids unnecessary dependencies

  – Smaller Apache memory footprint

YAHOO!

Security: INI Settings

- `open_basedir`
  - Insurance against `/etc/passwd` exploits
- `allow_url_fopen = Off`
  - Use `libcurl` extension instead
  - Avoid open proxy exploits
- `display_errors = Off`
  - However, `log_errors = On`
- `safe_mode = Off`
  - Intended for shared hosting environment

YAHOO!

Security: Input Filtering

`http://search.yahoo.com/search?p=<script+src=http://evil.com/x.js>`

- Cross Site Scripting (XSS) most common attack
  - Also "SQL Injection"
- Normal approach
  - `strip_tags()`
  - `mysqli_escape_string()`
  - Examine every line code
  - Tedious and error-prone
- Use `input_filter` hook
  - Sanitize all user-submitted data
 - GET/POST/Cookie

YAHOO!

Performance: Opcode Caches

- Easiest performance boost
    - Cache parsed .php scripts in shared memory
    - Optimizations
    - No code modifications!
- Several products available
    - Zend Performance Suite
    - APC
    - Turck MMCache

YAHOO!

Performance: PHP Extensions in C++

- PHP ships with 80 extensions written in C/C++

- Yahoo! develops its own proprietary extensions

  – Fast execution speed

  – Access to client libraries

- Longer development cycle

  – Edit, compile, link, debug

  – Manual memory-management
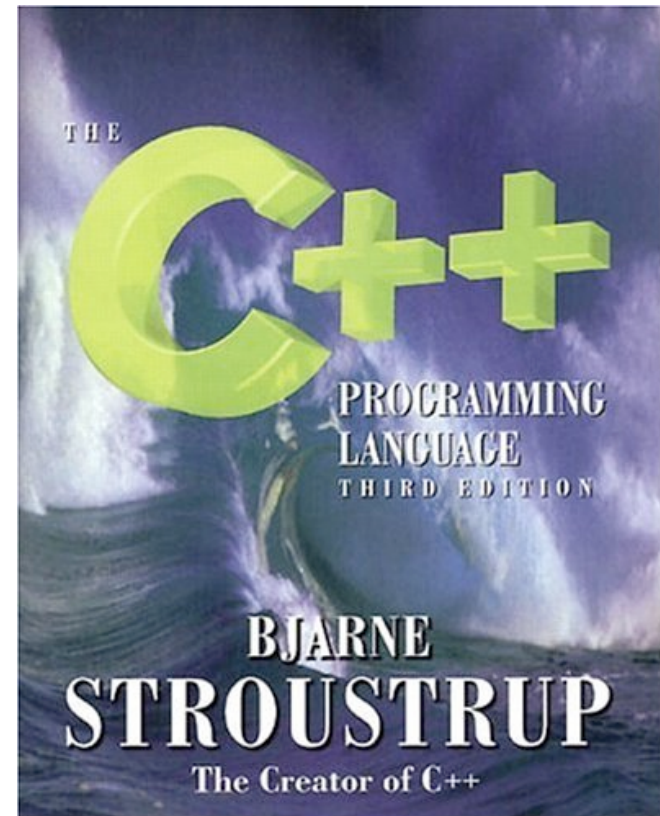
Globalization: PHP Unicode

YAHOO! + Zend *The php Company* + ICU = php 6

- Native Unicode support in 2006
- Collaborative effort
  - Andrei Zmievski (Yahoo!)
  - Andi Gutmans (Zend)
  - Many members of PHP Community

YAHOO!

Y! LIFE ENGINE™