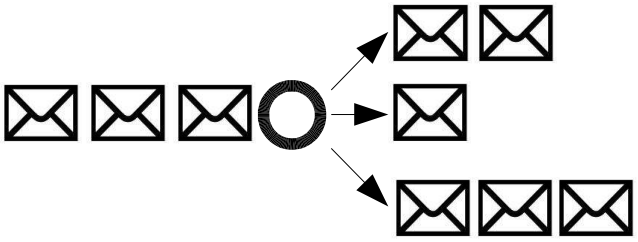
polymail

fast, accurate message classification



About

polymail is a C library that performs fast, accurate message classification. It is designed for anti-spam, message prioritization or message routing applications.

Uses

- Intelligent email routing
- Spam detection
- Message prioritization
- Article selection

Lightning Fast

polymail is very fast. On standard single CPU 2.8Ghz PC hardware polymail handles more than 500 messages per second with ease in a single thread. The average message is classified in 1.7ms.

polymail's unique 'in-situ' classifier requires minimal data copies, and its application-specific hashing algorithms ensure high-speed learning and classification.

polymail is thread-safe and supports unlimited classification threads.

Train on 6,047 messages	*Classify 6,047 messages*
10.4s	13.4s
581 msgs/second	450 msgs/second
1.7ms per msg	2.2ms per msg

Accurate

polymail is accurate. It incorporates a fast message parser, an optimized naïve Bayesian classifier and an inverse chi-squared test to provide accuracy of >99%.

Unlimited Message Classes

polymail supports unlimited message classes allowing classification into as many categories as desired. A simple application might involve sorting messages into spam or not spam; a complex implementation would be sorting customer service enquiries into appropriate categories for automatic response.

Compact and Embeddable

polymail is highly optimized and compiles to under 50k of code making it suitable for inclusion in any type of application.

polymail's data storage requirements are application dependent and can be tuned at compile time. polymail stores its data in a compact binary format for saving to non-volatile memory or disk.

Linux static library	*Windows 2000 DLL*	*Sun Solaris static library*
49k	40k	47k

Understands Multi-language Messages

polymail has built-in support for RFC2822 messages, multipart MIME, quoted printable encoding, base64 encoding, Unicode, Latin-1, HTML entities and encoded URLs.

Anti-spam Features

When used as an anti-spam library polymail enhances a fast naïve Bayesian text classifier with automatic or supervised training, database compaction and SURBL integration.

Since spam detection is an adversarial environment with spammers actively attempting to evade filtering, polymail incorporates many spam trick detection techniques and is backed by the research into The Spammers' Compendium.

Cross Platform

polymail compiles on Linux, Microsoft Windows and Sun Solaris. It is available as a shared (.so or .dll) or static library (.a or .lib), and can be built using gcc, Microsoft's CL or Sun's cc.

Complete C source code is included for customization, and includes a cross-platform build environment that supports parallel compilation. The source is fully documented and a unit test suite covering the code is provided.

polymail includes extensive debugging, logging and profiling facilities.

Licensing

polymail is licensed as source code on a per-organization basis. A single license fee is charged per organization with an annual support fee for ongoing maintenance and updates.

polymail is updated on a quarterly basis as email technologies, and spammer trickery, evolves.

Contact extravalent, llc

To learn more about polymail, or to evaluate its capabilities, please contact extravalent, llc through John Graham-Cumming by email:

info@extravalent.com.