

自動メール分類ツール POPFile

いいむらなおき@POPFile Core Team



<http://getpopfile.org/>

自己紹介

- 名前 : いいむらなおき (amatubu)
- 所属 : POPFile Core Team
- メール : amatubu@mac.com
- 日記 : <http://d.hatena.ne.jp/amatubu/>

本日のメニュー

- POPFile の概要
- POPFile の動作原理 (ベイズ理論)
- 精度を高めるために
- POPFile Core Team と私の活動
- おわりに

POPFile の概要

POPFile の概要 -1-

- オープンソース(GPL v2)で開発されている自動メール分類ソフトウェア
- メールの分類には、メールに含まれている単語を統計的に処理して分類先を決める、単純ベイズ(Naive Bayes)法が用いられている
- メールの分類を学習させていくことで、自動的にメールを分類してくれる
- Perlで書かれており、様々なプラットフォームで動作する
 - Windows、Mac OS X 向けにはインストーラあり

POPFile の概要 -2-

- もともと spam 対策ソフトとして開発されたわけではなく、いくつかの分類に自動的に振り分けるために作られた
- このため、「spam とそれ以外」という分類だけでなく、さまざまな目的に使用可能
- 統計情報によれば、バケツ(分類)の数の平均は 6個で、66.33% のユーザは 3個以上のバケツ(分類)を作成している
- 中には 200近くのバケツ(分類)を作成しているユーザも!!
- 私個人は、自宅では 3個、職場では 6個作成

POPFile の機能

- プロキシサーバモジュール

プロキシサーバとして稼働し、通過するメールに分類結果に応じた「印」をつける

- POP3 – POP3 サーバとメールクライアントの間
- NNTP – NNTP サーバとニュースクライアントの間
- SMTP – SMTP サーバ同士の間

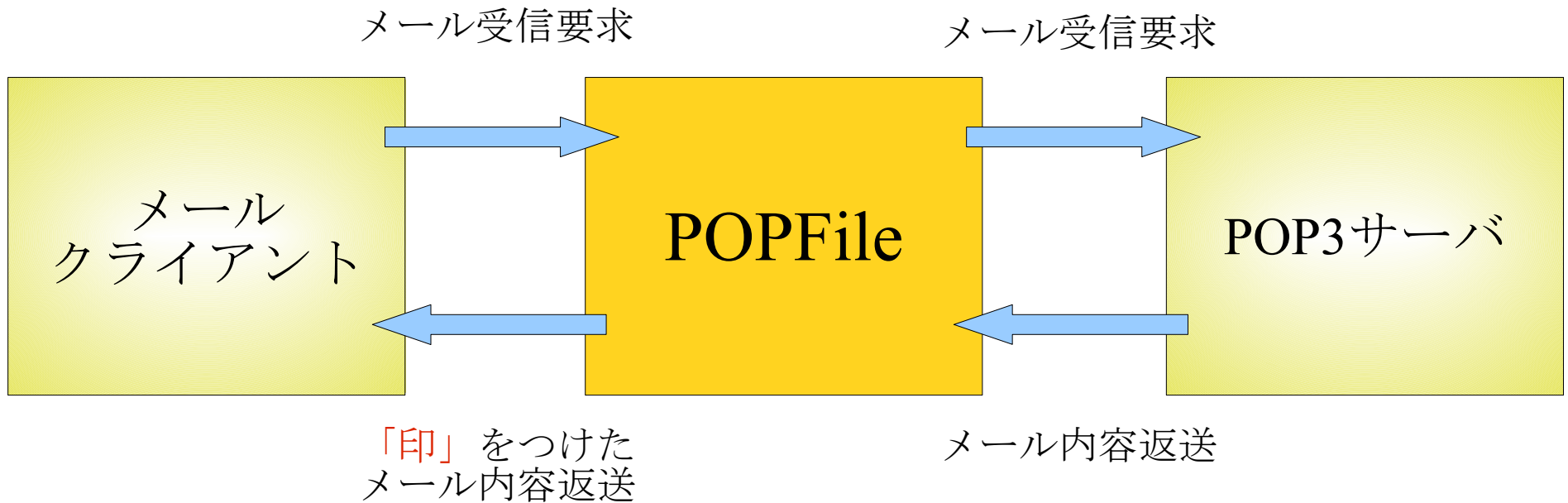
- サービスモジュール

- IMAP – IMAP サーバに接続し、分類結果により指定されたフォルダに移動させる

- インターフェース

- HTTP – Web ベースのユーザインターフェース
- XMLRPC – 他のソフトウェアから POPFile を操作

POPFile の動作イメージ (POP3)



```
Subject: [spam] Test mail
From: sender@example.com
To: receiver@example.com
X-Text-Classification: spam

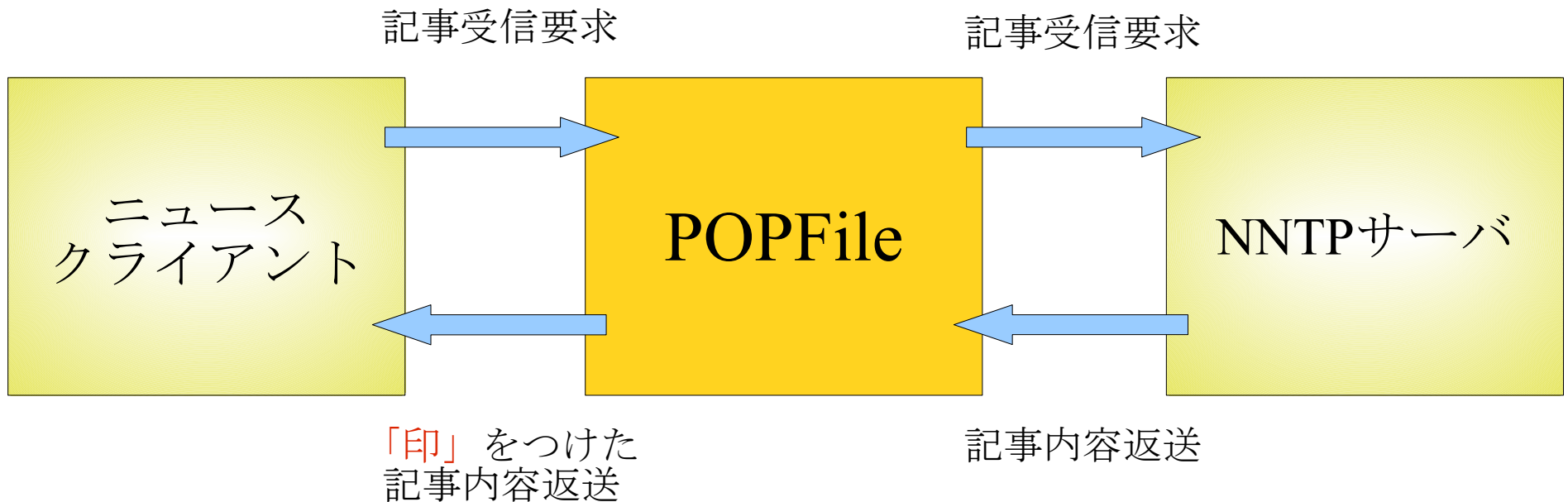
This is a test mail.
```

```
Subject: Test mail
From: sender@example.com
To: receiver@example.com

This is a test mail.
```

※ つけられた「印」をもとにメールをどう処理する (特定のフォルダに移すとか) は、メールクライアントに設定されたルールによる

POPFile の動作イメージ (NNTP)



```
Subject: [spam] Test mail
From: sender@example.com
To: receiver@example.com
X-Text-Classification: spam

This is a test mail.
```

```
Subject: Test mail
From: sender@example.com
To: receiver@example.com

This is a test mail.
```

POPFile の動作イメージ (SMTP)



メール転送

「印」をつけて
メールを転送

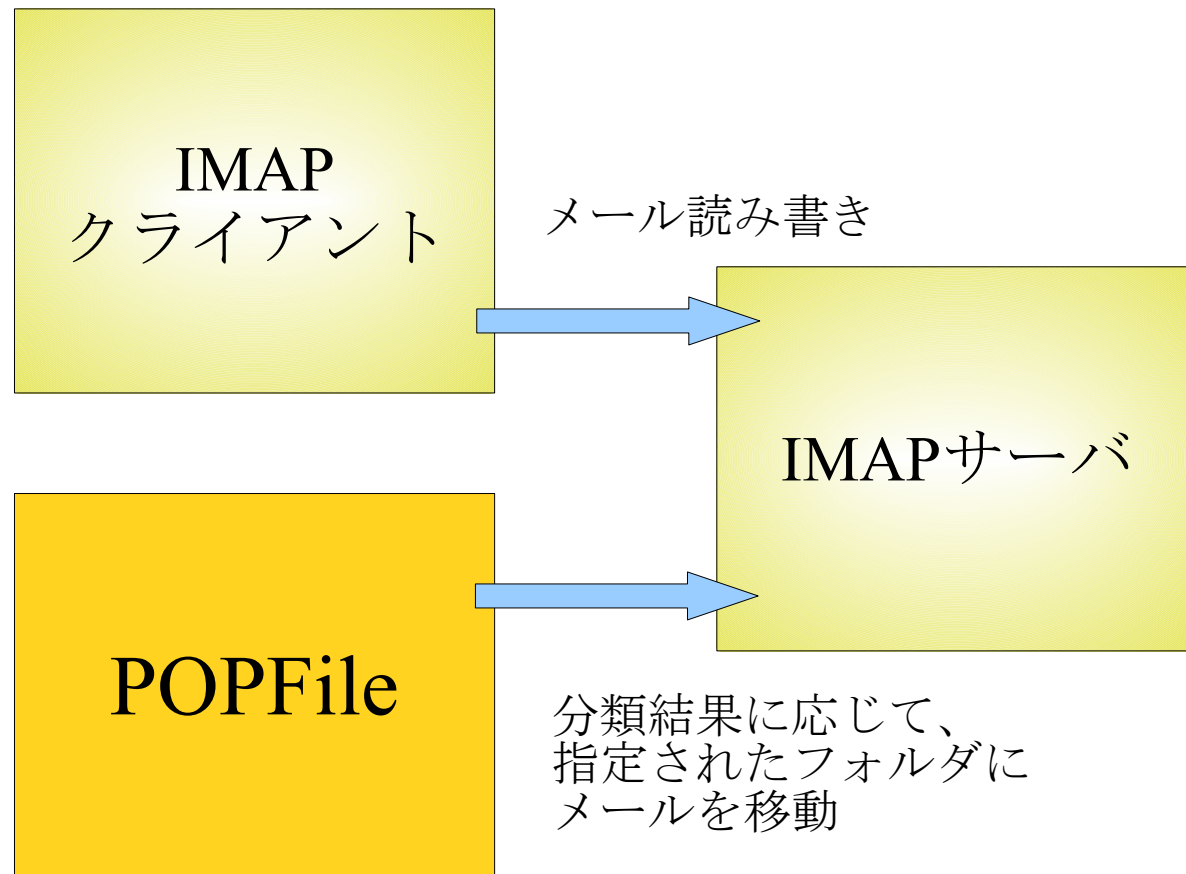
```
Subject: Test mail
From: sender@example.com
To: receiver@example.com
```

This is a test mail.

```
Subject: [spam] Test mail
From: sender@example.com
To: receiver@example.com
X-Text-Classification: spam
```

This is a test mail.

POPFile の動作イメージ (IMAP)



※ IMAP モジュールはプロキシではなく、クライアントとは関係なく動作する

POPFile のインターフェース

- POPFile の設定を変更するためには、Web インターフェース(コントロールセンター)を使用
- POPFile が分類を間違った場合、このインターフェースに接続して正しい分類を学習させる



POPFile の動作原理

ベイズの定理から

$$P(B_i|E) = \frac{P(E|B_i) \times P(B_i)}{P(E)}$$

<http://getpopfile.org/docs/jp:glossary:bayesian> より

- $P(B_i|E)$ は、メールEがバケツ(分類) B_i に含まれる確率 (=知りたいもの)
- $P(E|B_i)$ は、バケツ(分類) B_i からメールEが取り出される確率
- $P(B_i)$ は、バケツ(分類) B_i のメールが届く確率
- $P(E)$ は、メールEが届く確率

P(B_i)とは?

- バケツ(分類)B_iのメールが届く確率
- 言い換えれば、受信するメールのうち、バケツ(分類)B_iのメールの割合
- POPFileでは、この確率として、バケツ(分類)B_iに含まれる単語数 ÷ 全体の単語数を採用している

$$P(B_i|E) = \frac{P(E|B_i) \times P(B_i)}{P(E)}$$

P(E)とは?

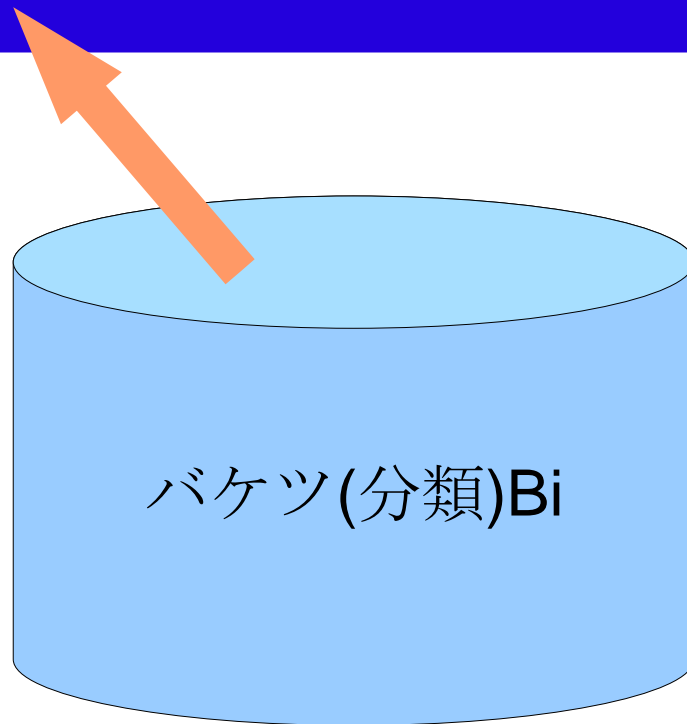
- メールEが届く確率
- この確率を計算するのは困難だが、この項目はどのバケツ(分類)の計算において共通に存在している
- 知りたいことは「どのバケツ(分類)に含まれる確率が最も高いか?」であるため、この項目は無視してよい

$$P(B_i|E) = \frac{P(E|B_i) \times P(B_i)}{P(E)}$$

P(E|Bi)とは? -1-

- バケツ(分類)BiからメールEが取り出される確率
メールE

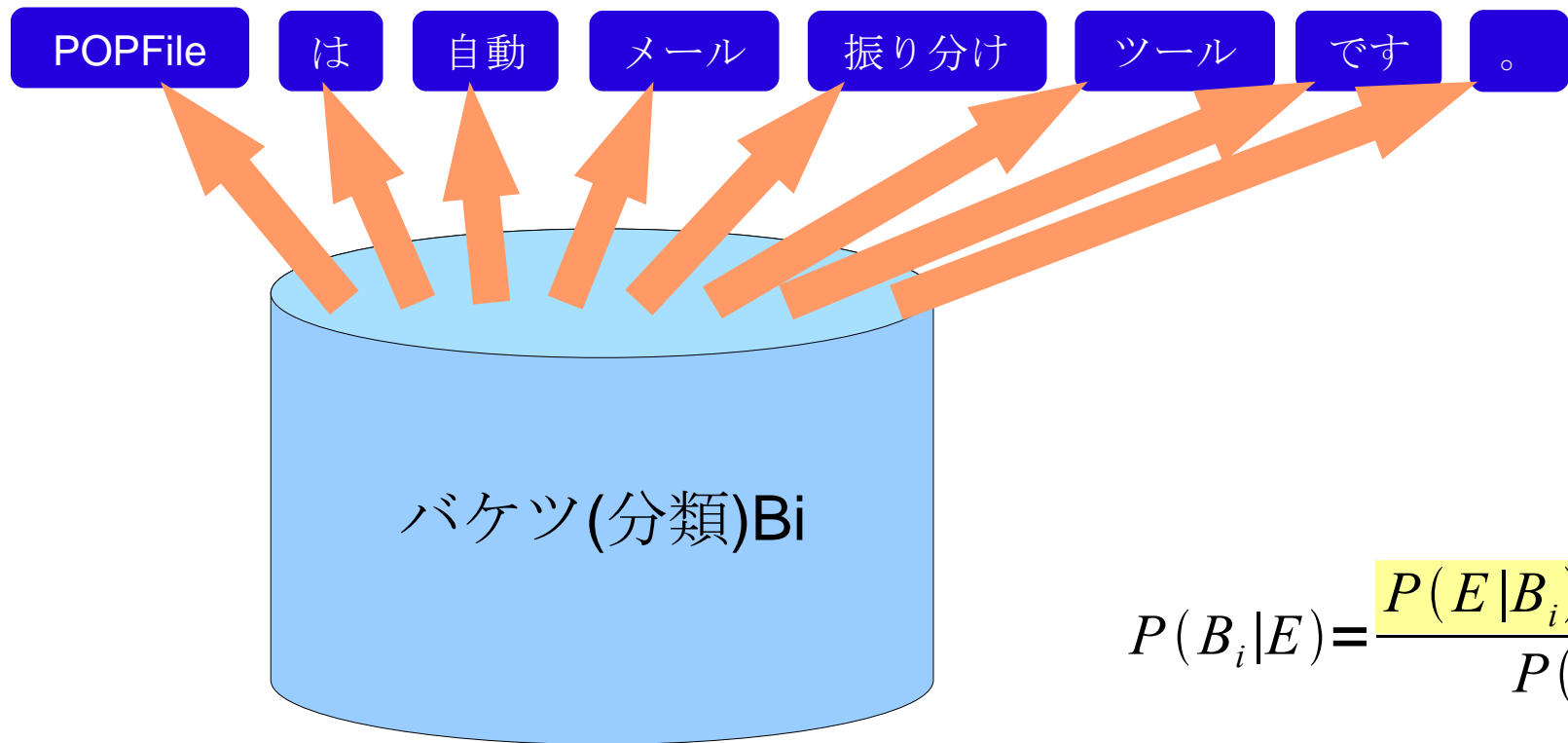
POPFile は 自動 メール 振り分け ツール です 。



$$P(B_i|E) = \frac{P(E|B_i) \times P(B_i)}{P(E)}$$

P(E|Bi)とは? -2-

- メールEを単語ごとに分割して、それぞれの単語がバケツ(分類)Biから取り出される確率、P(Ej|Bi)の積を計算することにする



$$P(B_i|E) = \frac{P(E|B_i) \times P(B_i)}{P(E)}$$

こんな単純計算でよいの？

- それぞれの単語が現れる確率は、他の単語には影響されず、独立であると仮定している（ここが「単純=naive」のゆえん）
- 実際にはこの仮定は正しくなく、ある単語の後にでてきやすい単語、出てきにくい単語がある
- この「単純」なプロセスでも、実用上問題のない分類精度を実現している
- 延べ約35,000人の利用者から収集した統計情報によれば、分類精度の平均値は97.25%。500通以上のメールを受信した後では98.23%

http://getpopfile.org/popfile_stats.html

結論 (計算すべきもの)

$$P(B_i) \times \prod_{j=1}^o P(E_j | B_i)$$

- $P(B_i)$ は、バケツ B_i が選ばれる確率 (=バケツ B_i のメールが届く確率)
- $P(E_j | B_i)$ は、バケツ B_i から単語 E_j が取り出される確率 (=バケツ B_i 内の単語 E_j の出現回数 / バケツ B_i の総単語数)

※バケツに存在しない単語の出現確率については、
1 / (全体の総単語数 × 10) とする

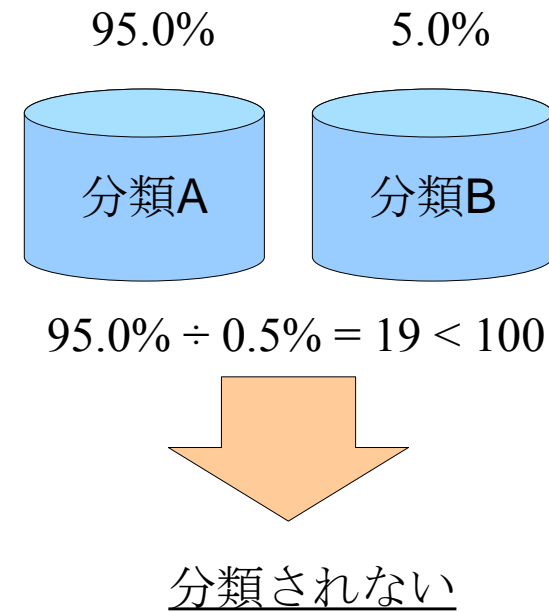
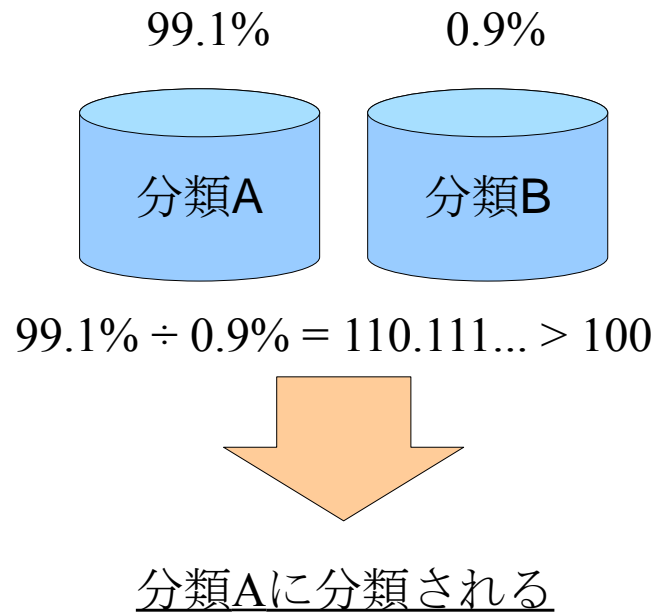
精度を高めるために

疑似単語

- 疑似単語 (pseudoword)
 - spam 送信者がフィルタを混乱させようとするために使うトリックなどを特別に扱う機能 (トリックを見つけると、ある特定の単語の数を加算する)
 - trick:spacedout – 単語の文字と文字の間にスペースを挟んだもの (ex. M_O_R_T_G_A_G_E)
 - html:emptypair – 中身が空のHTMLタグ (ex.)
 - 疑似単語の一覧
 - <http://getpopfile.org/docs/jp:faq:pseudowords>
 - 他にどのようなトリックが使われているかについては、Spammers' Compendium の情報などが参考になる
 - <http://www.virusbtn.com/resources/spammerscompendium/index>

分類するしきい値

- 分類するしきい値 (bayes_unclassified_weight)
 - メールを、特定のバケツ(分類)に分類するか、未分類(unclassified)とするかを判断するしきい値



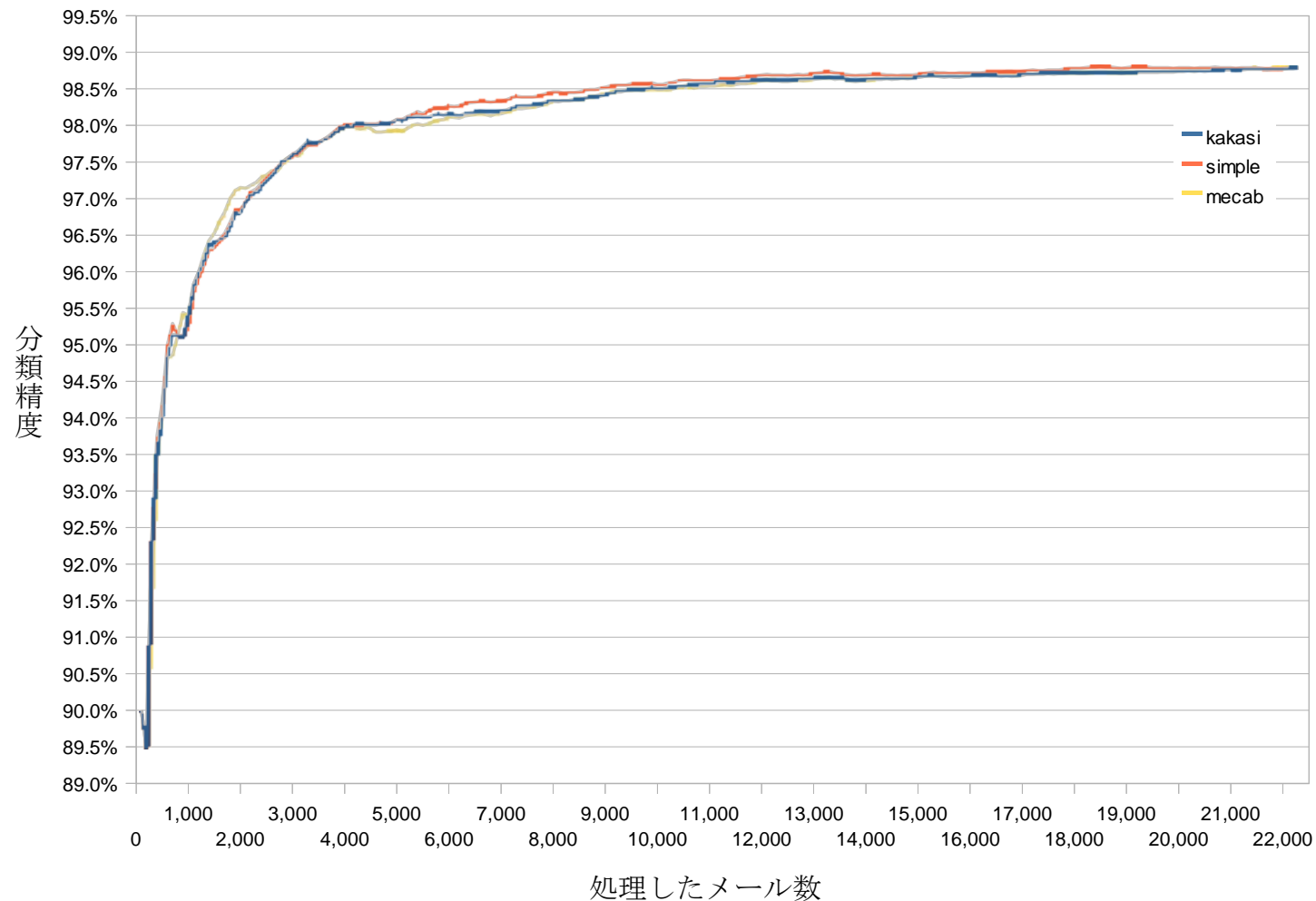
<http://d.hatena.ne.jp/amatubu/20040429#p1>

日本語特有の処理 -1-

- 分かち書き
 - 日本語の文章には単語と単語の間にスペースが含まれていないため、まず文章を単語ごとに区切る必要がある
 - POPFileでは、デフォルトでは Kakasi を使用
 - 好みにより、MeCab や内蔵パーサ (ひらがな、カタカナ、漢字などの文字種による分割) を選択可能
 - 22,340通のメールを 17のバケツ(分類)に振り分けるテストを行ってみたところ、いずれもほぼ同じ分類精度 (98.791%~98.796%) であった
 - <http://amatubu.skr.jp/index.cgi?POPFile/Accuracy>

日本語特有の処理 -2-

- 分かつ書きのプログラムによる分類精度の比較



日本語特有の処理 -3-

- 文字コードの変換
 - 日本語のメールでは、ISO-2022-JP、UTF-8 など、いろいろな文字コードが使われている
 - POPFile では、Encode モジュールを用いて文字コードを統一している
- 単語とみなす文字数
 - 欧文においては、2文字の単語は一般的な単語が多いため、POPFile では 3文字以上の単語のみを分類に使用している
 - 日本語では 2文字の単語が非常に多いため、2文字以上の単語を分類に使用している

POPFile Core Team と私の活動

POPFile Core Team

- POPFile の開発を行っているチーム
 - メンバー
 - Brian Smith – Windows 版のインストーラ作成
 - Joseph Connors - インターフェースやスキンの作成
 - Manni Heumann – IMAP モジュールの作成、プロジェクトのサーバの管理
 - Naoki IIMURA (私) - 上記以外のもろもろ...
 - As chairman
 - John Graham-Cumming (オリジナルの開発者)
- 詳しくは . . .
 - <http://getpopfile.org/docs/Glossary:POPFileCoreTeam>

私の活動（これまで）

- これまで
 - ドキュメント(Wiki)の翻訳 (2004.2-)
 - バグ修正や日本語処理のためのパッチ提供 (2004.2-)
 - Mac OS X 版の独自公開 (2004.3-)
 - プロジェクトのメンバーに (2004.12)
 - MacPeople に原稿を書く (2005.1)
 - プロジェクトのリポジトリに初コミット (2005.8)
 - POPFile Core Team 発足。メンバーに (2008.5)
 - Mac OS X 版をオフィシャル化 (2008.11)

私の活動 (最近の主なもの)

- 最近の主なもの
 - v1.0.0 (2007.12)
 - 分かち書きに使用するプログラムを選択可能に
 - メール分析のパフォーマンス向上
 - v1.1.0 (2008.11)
 - SQLite 3.x へのバージョンアップ対応
 - Windows 版におけるタスクバーアイコンの改善
 - Mac OS X 版のインストーラをオフィシャルに
 - NNTP モジュール向けのパッチをマージ
 - v1.1.1 (予定)
 - さまざまな細かい機能追加とバグ修正

私の活動 (現在とこれから)

- 現在

- Mac OS X 版のメンテナンス
- 日本語処理関係のメンテナンス
- その他の機能追加、バグフィックス
- インターフェースのローカライズ
- フォーラムでのサポート

- これから

- バージョン 2 に向けて... ..
 - マルチユーザ対応 (ほぼ完成)
 - SSL 接続の受付 (完成。要テスト)
 - 英語でのドキュメント書き (一番の難関... ..)

おわりに

- ありがとうございます
- 現在、次のバージョンである POPFile v1.1.1 のリリース候補版 (RC6) をリリースし、テスト中です。
テストにご協力いただける方を募集中です。
興味のある方は POPFile 日本語フォーラムへ!

<http://getpopfile.org/discussion/4>

参考文献

- POPFile はどのようにしてメールを分類しているのか (How POPFile does classification)
<http://getpopfile.org/docs/jp:Glossary:Bayesian>
- POPFile Real Time Statistics
http://getpopfile.org/popfile_stats.html
- POPFile が解釈することができる'疑似単語'にはどんなものがあり、それらはどのように働きますか?
<http://getpopfile.org/docs/jp:faq:pseudowords>
- バケツに振り分けるしきい値
<http://d.hatena.ne.jp/amatubu/20040429#p1>
- POPFile の精度について - 分かち書きに使用するプログラムによる精度の違い
<http://amatubu.skr.jp/index.cgi?POPFile/Accuracy>

質疑応答