

Mail Guardian Neo v1.0

☆特徴

- ・常駐させるだけで広告メールを自動駆除
- ・強力なスパムメールフィルタ搭載
- ・わかりやすく配慮したインターフェース
- ・最新版が常にチェック出来る機能搭載

☆製作

所属：千葉県立国府台高等学校 普通科
氏名：中野 裕樹



☆Mail Guardian Neoについて

広告メールなどの受信者が必要としていない無駄なメールを常駐させているだけで自動的に削除できるソフトウェアです。このタイトルの最後につくNeoとは、以前からウェブサイト上で公開されていたMail Guardianの新バージョンという意味があります。これは旧バージョンMail Guardianとは互換性はなく、全く違う視点からスパムメールを排除するという事を追究し、新たに製作したのがこのMail Guardian Neoです。

☆動作環境

OS : Windows 98系 Windows 2000系 確認済み(Windows 2000/XP)
メモリ : 256MB以上推奨 (使用領域 : 20MB前後) HDD : 空き 3 MB 程度

☆”強力なフィルタ”について

このプログラムのフィルタの部分には”ベイズフィルタ”という強力なフィルタを実装しています。この方式は、届いたメールの文章の特長をしらべ、スパムである確率を計算するものです。ユーザはまずは、届くメールをフィルタに学習をさせます。学習は、メールを単語分割し、それらの単語がスパムか否かを記録します。ある程度学習が済むと、新たに届いたメールがその学習データから、メールに含まれている単語1つ1つにどれぐらいのスパム指標確率を持っているかが計算され、その中の代表的ないくつかの確率からメール全体のスパム指標の確率が計算されます。その確率がユーザの指定した値を超えていればスパムと判断され、下回ってれば、ノンスパムと判断されます。

☆単語分割方法

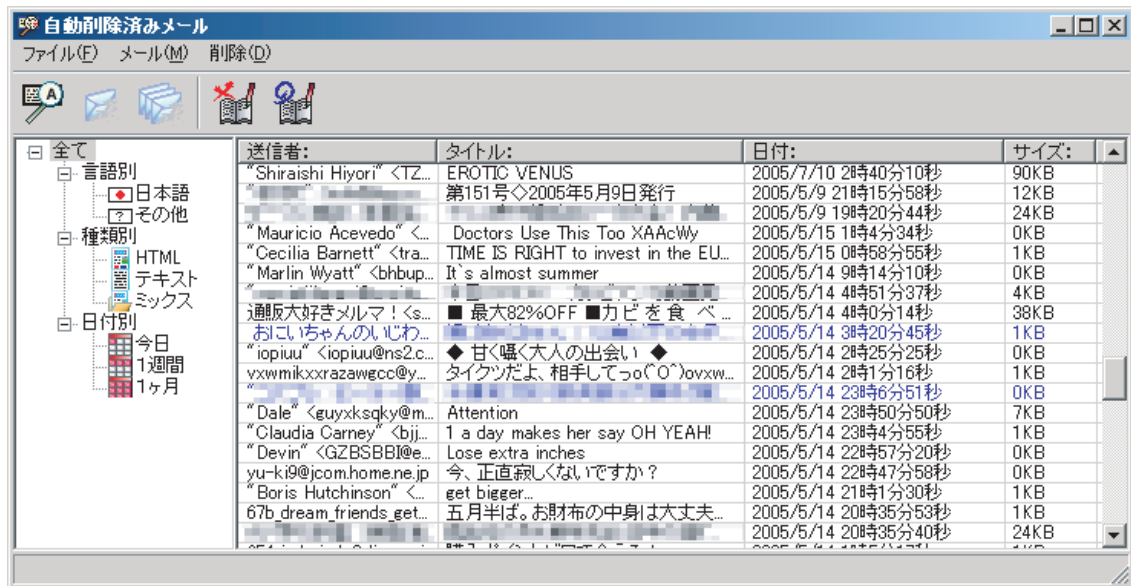
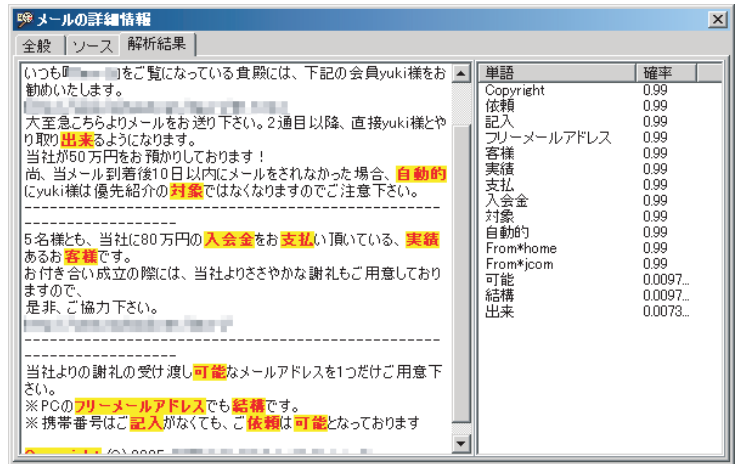
トークン分割に関しては、日本語も対応できるようにしました。日本語からトークン分割するのは容易ではなく、一番悩んだところでもあります。”bi-gram”を実装しているという方もおられますが、MG Neoでは単純に漢字とカタカナでつながった文字を抽出するという方法を用いました。

☆解析結果の表示

手動メールチェックのときは、到着したメールの判定結果の詳細を表示する事ができます。

これは届いたメールの1例です。本文中に反転された文字がその文章の重要な単語であり、判定に実際に使われた単語です。右側にはその使われた単語の確率が示されています。

この右に表示されている確率から、結合確率を算出します。



☆ゴミ箱フォルダ

ゴミ箱フォルダは非常に便利な機能です。常駐メールチェック時、ゴミ箱に入れる設定しておくとお見の事が出来ます。常駐している時に、削除されたメールを見るには、「システム」から「ゴミ箱のメールを閲覧」をクリックします。

検索フォルダで分かれています。また、ゴミ箱に入ったメールを1度だけ学習する事ができます。これにより、常駐による自動化の状態でも誤検出の修正が可能になります。一度学習したメールは青く表示され、学習ができなくなります。

☆今後の予定

今後の課題として、デコード機能があります。。一般的なメールはほとんどがデコード可能にはなりましたが、現段階でも、JISコードをQuoted-Printableでエンコードしたメールなどがデコードに失敗します。また、破損したメッセージにいかにも未知のエラーを起こさせないかも必要だと思います。

しかしそれよりも優先すべき課題は、POPプロキシとして動作させる事だと思います。これは僕の技術不足で、今回はとりあえずメールチェッカとして完成させました。ベイズフィルタに重きを置いたのが理由です。POPプロキシとして動作できるようになるまでは、β版として公開する予定です。