

InterMail Post.Office 3.9.0J 補遺マニュアル

マニュアル・バージョン3.9.0

2004年10月

Open
Technologies

目次

1. 新しく追加された機能について.....	1
2. QuattroJ ジャンク判定サーバとの連携機能.....	2
2.1. QuattroJ とは.....	2
(1) ベイズ理論を適用した日本語対応のジャンク判定機能.....	2
(2) メールのトークンを蓄積するためのデータベース機能.....	2
(3) 管理用の Web インタフェース機能.....	2
(4) InterMail Post.Office と連携するジャンク判定サーバ機能.....	2
2.2. QuattroJ の設定.....	3
2.3. Post.Office と QuattroJ の振る舞い.....	4
2.4. ジャンク判定時の付加情報について.....	5
(1) 元のメッセージを添付ファイルとしてラッピングするかどうか.....	5
(2) 独自の X ヘッダを付加するかどうか.....	6
(3) Subject(題名)の先頭に、ジャンクを明示するプレフィックスを付加するかどうか.....	7

1. 新しく追加された機能について

InterMail Post.Office 3.9.0J では、次の機能が新たに追加されました。本マニュアルでは、これらの機能を順次、簡単に説明します。

- QuattroJ ジャンク判定サーバとの連携機能

2. QuattroJ ジャンク判定サーバとの連携機能

2.1. QuattroJ とは

QuattroJ(クアトロJ)は、InterMail Post.Office と連携する迷惑メールフィルタです。QuattroJ には、次のような特徴があります。

- (1) ベイズ理論を適用した日本語対応のジャンク判定機能
- (2) メールトークンを蓄積するためのデータベース機能
- (3) 管理用の Web インタフェース機能
- (4) InterMail Post.Office と連携するジャンク判定サーバ機能

(1) ベイズ理論を適用した日本語対応のジャンク判定機能

ベイズ理論とは、数学者のトーマス・ベイズが提唱したといわれる、過去に起きた事象から、その事象が未来に起こる確率を求める確率論です。

過去に届いたジャンクメールの集合と、ジャンクでないメール(ノーマルメール)の集合から、メールに含まれるトークン(語彙)を抽出し、それぞれのメールにどのようなトークンが多く含まれるかを蓄積して、ジャンク判定を行います。アルファベットなどで構成されている英語圏からメールだけでなく、日本語のメールに対してもジャンク判定できます。

蓄積したメール数や、トークンの数にもよりますが、判定率 90%以上の確率でジャンク判定が可能です。

(2) メールトークンを蓄積するためのデータベース機能

QuattroJ では、メールから取り出したトークンが、ジャンクメールや、ノーマルメールに現れた回数を蓄積するためのデータベースを内蔵しています。メールを処理した数や、ジャンク判定の履歴や、新規の学習履歴なども蓄積します。

また、データベースには、予めジャンク判定として扱わないトークンを「判定対象外トークン」として蓄積でき、それらのトークンをジャンク判定に含めずに判定を行いますので、効率的な判定が行えます。

(3) 管理用の Web インタフェース機能

管理者が QuattroJ 全体を制御、管理するための機能で、インターネットブラウザを使って簡単に管理できる仕組みになっています。

さらに、トークンの情報や、蓄積した学習の結果、履歴などを参照したりできます。

QuattroJ がジャンク判定した結果、ジャンクメールをノーマルと「見逃し」たり、ノーマルメールをジャンクメールと「誤認」したりした場合、QuattroJ に対して再学習させることもできるので、非常に簡単に管理できます。

また、QuattroJ 管理下のサーバの起動/停止も、Web インタフェースで運用できます。

(4) InterMail Post.Office と連携するジャンク判定サーバ機能

InterMail Post.Office と連携動作するためのジャンク判定サーバ機能を提供します。

InterMail Post.Office がメールを受け取ると、メールボックスに配信する前に、ジャンク判定サーバに接続し、メール全体を QuattroJ に渡します。QuattroJ では、Post.Office から受け取ったメールからトークンを抽出し、ジャンク判定を行い、判定結果を、Post.Office に返します。Post.Office はジャンク判定結果を、メールに付加して、メールボックスに配信します。

ただし、これらのジャンク判定は、いわゆる迷惑メールや、ウィルス付きメールを完璧に判断し、除外するものではありません。

2.2. QuattroJ の設定

Post.Office の管理画面に、QuattroJ との連携のための[QuattroJ の設定]画面が追加されました。

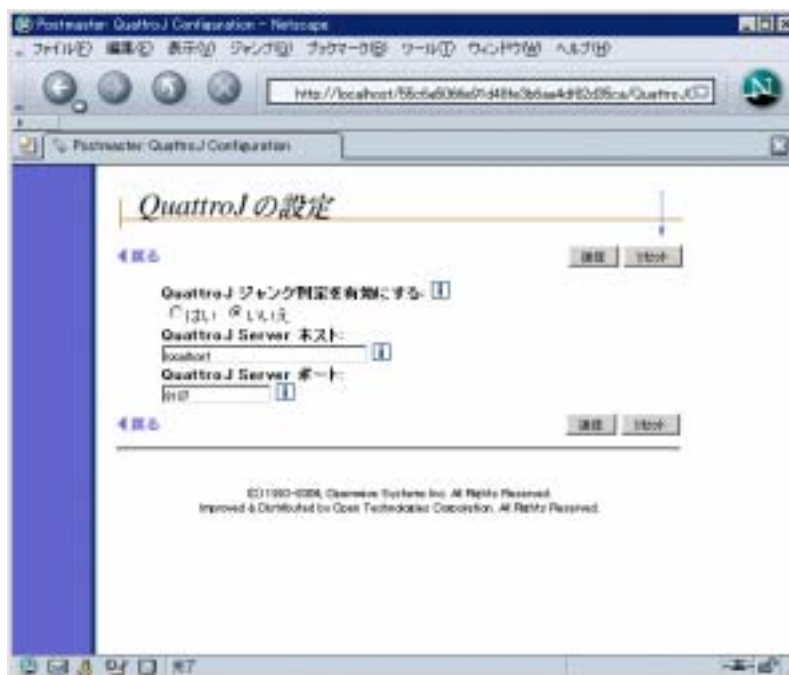


図 1 [QuattroJ の設定]画面

[QuattroJ ジャンク判定を有効にする]の項目で、[はい]を選択して[送信]ボタンを押すと、QuattroJ のジャンク判定サーバと連携して、メールに対してジャンク判定を行うようになります。[いいえ]を選択して[送信]ボタンを押すと、QuattroJ のジャンク判定サーバと連携が無効になります。[QuattroJ Server ホスト]には、QuattroJ が稼働しているサーバ名を入力してください。Post.Office と同一マシン上で稼働している場合は、[localhost]で構いません。[QuattroJ Server ポート]は、QuattroJ がジャンク判定サーバをサービスしているポート番号を入力します。デフォルトは、9107 番です。QuattroJ の設定で、判定サーバのポート番号を変更している場合は、そのポート番号を入力してください。特に理由がない限り変更する必要はありません。

2.3. Post.Office と QuattroJ の振る舞い

QuattroJ ジャンク判定が有効になると、Post.Office と QuattroJ は、次の図のようなフローで振る舞います。

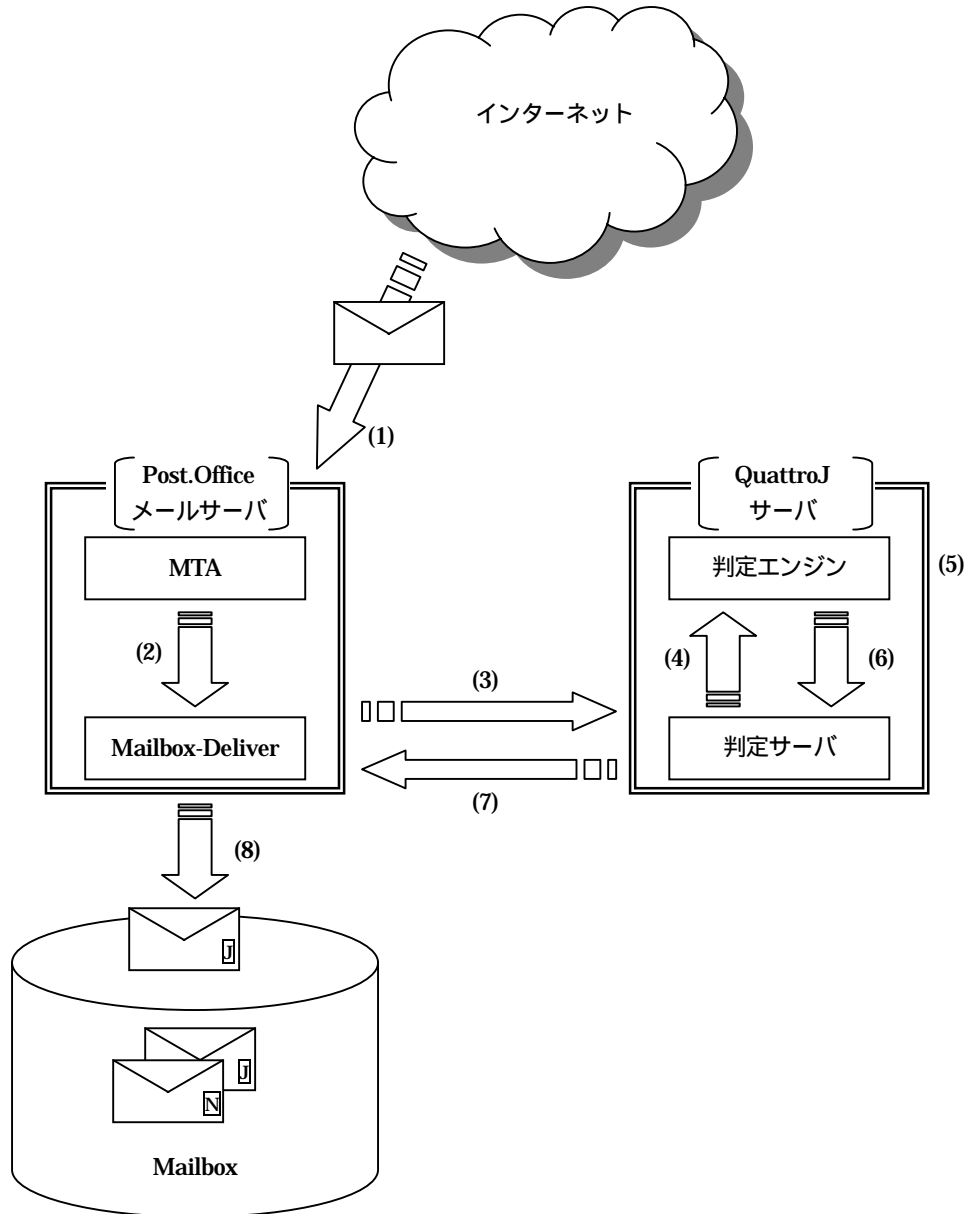


図 2 [Post.Office と QuattroJ の連携フロー]

- (1) インターネットからのメールを、Post.Office の MTA が受け取ります
- (2) MTA はルーティングの設定に基づいてメールをどこに配ればよいのかを判断し、ローカルドメイン宛のメールを、Mailbox-Deliver に渡します
- (3) ローカルドメイン宛のメールを受け取った Mailbox-Deliver は、QuattroJ の判定サーバに接続してメールをそのまま、判定サーバに渡します
- (4) 判定サーバは、メールを判定エンジンに渡します
- (5) 判定エンジンでは、次のような処理が行われます
 - (5-1) メールを、形態素解析器にかけて、トークンに分割します
 - (5-2) 分割したトークン単位に、それまでに蓄積した出現回数を、内蔵データベースから検索します
 - (5-3) トークン単位に出現回数からジャンク率を算出し、ジャンク率の高いトークンの上位いくつかから、さらに複合確率を算出します。

(5-4) 複合確率が、ジャンク判定のボーダーポイントよりも大きい場合に「ジャンクメール」、そうでない場合に「ノーマルメール」として判断します。

- (6) 判定エンジンは、ジャンク判定の結果を、判定サーバに返します。
- (7) 判定サーバは、ジャンク判定の結果と、メールに追加する付加情報を、Mailbox-Deliver に返します。
- (8) Mailbox-Deliver は、判定サーバからの返事に、メールに追加する付加情報があれば、その追加情報を付加して、配信先のメールボックスにメールを配信します。

2.4. ジャンク判定時の付加情報について

QuattroJ のジャンク判定サーバは、ジャンク判定の結果を Mailbox-Deliver に返す際、付加情報を一緒に返すようになっています。その追加情報には、次のようなものがあります。

- (1) 元のメッセージを添付ファイルとしてラッピングするかどうか
- (2) 独自の X ヘッダを付加するかどうか
- (3) Subject(題名)の先頭に、ジャンクを明示するプレフィックスを付加するかどうか

(1) 元のメッセージを添付ファイルとしてラッピングするかどうか

ジャンク判定されたメールは、ただの迷惑メールだけではなく、もしかすると、ウィルス付きのメールかも知れません。そのようなメールをメーラでそのまま開いてしまうことは、例え中身を覗くだけでも大変危険です。QuattroJ では、ジャンクメールを、ジャンク報告メールの添付ファイル化することができ、直にメールを開いてしまう危険を回避できます。ジャンク報告メールの本文には、元のメッセージの差出人(From)、題名(Subject)、日付(Date)記載されますので、添付を開かなくてもある程度判断できます。また、仮にそのメールのジャンク判定が誤認で、本当は必要なメールだった場合は、添付ファイルを保存して、メーラで開けばそのままメールとしても利用できます。

QuattroJ で、ジャンク判定時に元のメッセージを添付ファイルにするには、QuattroJ の管理画面の[基本設定] - [判定設定]で、[ジャンク判定時、元のメッセージを添付にする]を[する]にして、[更新]ボタンを押してください。

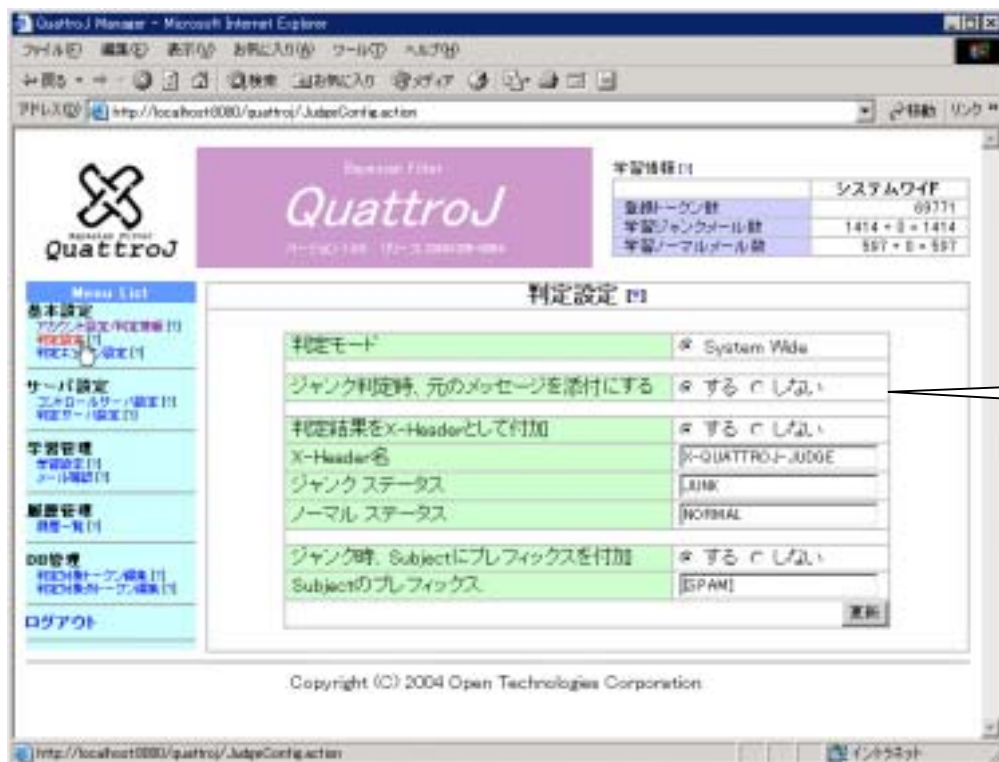


図 3 [QuattroJ : 判定設定]画面

(2) 独自の X ヘッダを付加するかどうか

ジャンク判定の結果を、独自の X ヘッダに付加するかどうかを、設定できます。QuattroJ では、独自の X ヘッダ名と、ジャンク判定された場合のステータス、ノーマル判定された場合のステータスのそれぞれを設定できるようになっています。この付加情報を使うと、この独自の X ヘッダの有無や、ステータスの内容で、メーラに付属する振り分け機能の条件に利用でき、ジャンク判定されたメールを、特定のフォルダに自動的に移動させたりすることができます。

QuattroJ で独自の X ヘッダおよびそれぞれのステータスを設定するには、QuattroJ の管理画面の[基本設定] - [判定設定]で入力します。

[判定結果を X-Header として付加]を[する]にします。

[X-Header 名]に、設定したい独自の X-Header 名を入力します。

[ジャンクステータス]に、ジャンク判定された場合に、X-Header に付くステータスを入力します。

[ノーマルステータス]に、ノーマル判定された場合に、X-Header に付くステータスを入力します。

内容を入力したら、[更新]ボタンを押してください。



図 4 [QuattroJ : 判定設定]画面

(3) Subject(題名)の先頭に、ジャンクを明示するプレフィックスを付加するかどうか

ジャンク判定の結果、ジャンク判定された場合に限り、Subject の先頭に、ジャンクメールであることを明示するプレフィックスを付加できます。これは、先ほどの独自の X ヘッダと同じ考え方ですが、メーラとして Outlook をお使いの場合は、独自の X ヘッダに対して振り分け処理を定義できません。この替わりとして、Subject の先頭が、ジャンクメールであることを示すプレフィックスで始まっているかどうかで、振り分け処理を定義できます。

QuattroJ で Subject のプレフィックスを設定するには、QuattroJ の管理画面の[基本設定] - [判定設定]で入力します。

[ジャンク時、Subject にプレフィックスを付加]を[する]にします。

[Subject のプレフィックス]に、設定したいプレフィックスを入力します。

内容を入力したら、[更新]ボタンを押してください。

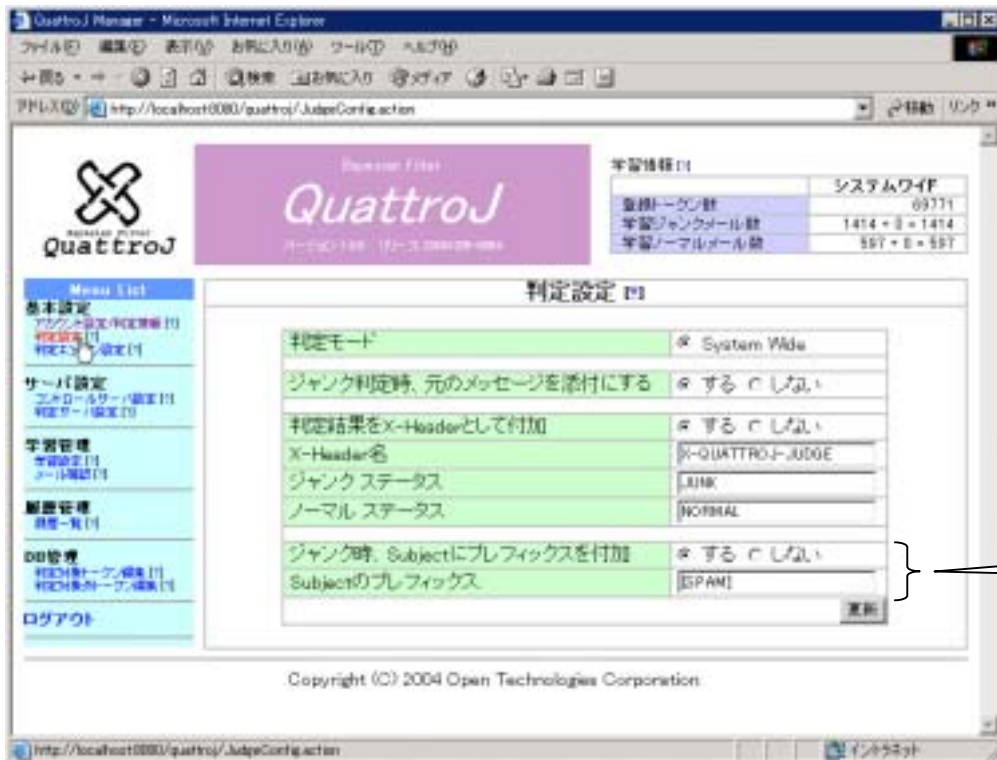


図 5 [QuattroJ : 判定設定]画面

(C) 1993-2002, Openwave Systems Inc. All Rights Reserved.

(C) 2002-2004 Open Technologies Corporation. All Rights Reserved.

Improved & Distributed by Open Technologies Corporation.