

RFC 7377 : IMAP4 Multimailbox SEARCH Extension

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 14 octobre 2014

Date de publication du RFC : Octobre 2014

<https://www.bortzmeyer.org/7377.html>

Nouvelle extension IMAP normalisée, pour effectuer une recherche dans plusieurs boîtes aux lettres simultanément, avec la nouvelle commande ESEARCH. Proposée initialement par le RFC 6237¹, à titre expérimental, cette extension accède désormais au statut de norme.

Par défaut, une recherche IMAP (commande SEARCH, RFC 3501, sections 6.4.4 et 7.2.5, puis étendue par les RFC 4466 et RFC 4731) ne cherche que dans la boîte aux lettres courante. Si on n'est pas sûr de la boîte où se trouve le ou les messages intéressants, et qu'on veut chercher dans plusieurs boîtes, la seule solution est de faire une boucle, effectuant séquentiellement des SELECT pour changer de boîte et un SEARCH à chaque fois. Ces commandes ne peuvent pas être exécutées en parallèle car il n'y a pas de moyen de distinguer leurs résultats (les réponses SEARCH n'indiquent pas quelle était la boîte). La nouvelle commande permet au serveur de faire les recherches en parallèle, s'il le peut, et évite des allers-retours entre client et serveur.

Donc, la nouvelle commande ESEARCH (à ne pas confondre avec la capacité ESEARCH du RFC 4731, même si les deux RFC ont des points communs, comme le format des réponses), présentée en section 2. Sa disponibilité est annoncée par la capacité MULTISEARCH dans la bannière du serveur IMAP. (Cette capacité figure dans le registre IANA <<https://www.iana.org/assignments/imap-capabilities>>.) Voici un exemple où on cherche des messages parlant du Tchad dans les boîtes folder1 et celles sous folder2 (C = client, et S serveur IMAP) :

```
C: mytag ESEARCH IN (mailboxes "folder1" subtree-one "folder2") subject "chad"
S: * ESEARCH (TAG "mytag" MAILBOX "folder1" UIDVALIDITY 1) UID ALL 3001:3004,3788
S: * ESEARCH (TAG "mytag" MAILBOX "folder2/salmon" UIDVALIDITY 1111111) UID ALL 50003,50006,50009,50012
S: mytag OK done
```

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc6237.txt>

La syntaxe pour indiquer les boîtes aux lettres où on cherche est celle de la section 6 du RFC 5465, avec l'ajout de `subtree-one` (utilisé ci-dessus pour `folder2`) qui, contrairement à `subtree`, s'arrête au premier niveau (donc la boîte `folder2/tuna/red` ne serait pas cherchée). Avertissement du RFC au programmeur : `subtree` peut aller très loin (pensez à une hiérarchie de boîte aux lettres mise en œuvre avec un système de fichiers hiérarchiques, et des liens symboliques qui vont se promener ailleurs) et il faut donc prêter attention au risque d'écroulement du serveur. Voir aussi la section 2.4, qui autorise le serveur à refuser les recherches trop gourmandes, avec le code `LIMIT` du RFC 5530, et la section 5, sur la sécurité, qui note que cette extension multi-boîtes peut être dangereuse pour le serveur. Un client méchant ou maladroit peut demander un gros travail au serveur, avec juste une commande.

Le RFC note que le mot-clé `personal` est le moyen le plus pratique de fouiller dans toutes les boîtes de l'utilisateur. Si celui-ci désigne au contraire des boîtes par leur nom, et que le serveur gère les ACL du RFC 4314, il faut avoir le droit de lecture sur ces boîtes (et celui de découverte, si les boîtes ne sont pas explicitement nommées, par exemple si on utilise `subtree`). Pas question de retourner des résultats pour des boîtes que l'utilisateur n'est pas censé lire.

À noter qu'un serveur qui accepte les recherches floues du RFC 6203 peut les accepter également pour les recherches dans des boîtes multiples.

La réponse est décrite dans la section 2.1 : elle consiste en lignes `ESEARCH` (cf. RFC 4731). Les messages doivent être identifiés par leur UID (RFC 3501, section 2.3.1.1) comme ci-dessus, pas par leurs numéros (qui n'ont de sens que pour une boîte sélectionnée). Le `ALL` (RFC 4731, section 3.1), indique qu'on retourne la totalité des messages qui correspondent au critère de recherche. Pour le premier dossier, le `folder1`, ces messages sont ceux d'UID 3001 à 3004 (une séquence, définie dans la section 9 du RFC 3501) et le message d'UID 3788.

La section 3 du RFC donne un exemple plus complexe, avec deux recherches multi-boîtes en parallèle, chacune identifiée par une étiquette différente :

```
C: tag1 ESEARCH IN (mailboxes "folder1" subtree "folder2") unseen
C: tag2 ESEARCH IN (mailboxes "folder1" subtree-one "folder2") subject "chad"
S: * ESEARCH (TAG "tag1" MAILBOX "folder1" UIDVALIDITY 1) UID ALL 4001,4003,4005,4007,4009
S: * ESEARCH (TAG "tag2" MAILBOX "folder1" UIDVALIDITY 1) UID ALL 3001:3004,3788
S: * ESEARCH (TAG "tag1" MAILBOX "folder2/banana" UIDVALIDITY 503) UID ALL 3002,4004
S: * ESEARCH (TAG "tag1" MAILBOX "folder2/peach" UIDVALIDITY 3) UID ALL 921691
S: tag1 OK done
S: * ESEARCH (TAG "tag2" MAILBOX "folder2/salmon" UIDVALIDITY 1111111) UID ALL 50003,50006,50009,50012
S: tag2 OK done
```

Des implémentations de cette norme ? Il y en a apparemment une, en logiciel non-libre, chez Oracle. Cette extension semble rare (et cela a été noté dans les débats <<http://www.ietf.org/mail-archive/web/imapext/current/msg05151.html>> à l'IETF : méritait-elle vraiment d'avancer sur le chemin des normes ?). J'ai testé quelques serveurs et je n'ai pas vu la capacité `MULTISEARCH` dans un Zimbra récent, ou dans la version de Courier actuellement chez Debian. Rien non plus dans des services en ligne comme `laposte.net` ou Gmail.

Depuis le RFC 6237, qui décrivait cette extension pour la première fois, peu de changements, les principaux sont (section 7) :

- Passage au statut de norme,
- Recherches floues,
- Et quelques autres détails.