

# HP Integrated Archive Platform Version 2.1

## User Guide

Includes information about using the Integrated Archive Platform (IAP) Web Interface (Content Search and Retrieve interface).

For additional user information on Email Archiving software for Microsoft Exchange and IBM Lotus Domino, see the *HP Email Archiving software for Microsoft Exchange User Guide* and *HP Email Archiving software for IBM Lotus Domino User Guide* contained in those products.



**Legal and notice information**

© Copyright 2004–2010-NaN Hewlett-Packard Development Company, L.P.

Confidential computer software. Valid license from HP required for possession, use or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Microsoft®, Windows®, Windows XP®, and Windows Vista® are U.S. registered trademarks of Microsoft Corporation.

Adobe and Acrobat are trademarks of Adobe Systems Incorporated.

---

# Contents

About this guide .....	9
Intended audience .....	9
Prerequisites .....	9
Related documentation .....	9
Document conventions and symbols .....	9
Support .....	10
Subscription service .....	11
1 IAP overview .....	13
About the IAP .....	13
Understanding document archiving .....	14
Understanding searching and document indexing .....	14
What the IAP indexes .....	15
Indexed file types .....	15
Message MIME types .....	15
What the IAP does not index .....	15
Document markup and formatting .....	15
Non-indexed file types .....	16
Limitations for Microsoft Office items .....	16
2 Introducing the IAP Web Interface .....	17
Logging in and out .....	17
Understanding the user interface .....	17
Using the toolbar .....	18
Search basics .....	18
Common tasks .....	19
Changing your password .....	20
Changing your language .....	20
3 Searching for email or files .....	21
Completing simple searches .....	21
Completing advanced searches .....	22
Email Content Type .....	24
Document Content Type .....	26
4 Working with search results .....	29
Displaying search results .....	29
Query results navigation bar .....	31
Result retrieval batches .....	31
Saving search criteria .....	32
Saving search results .....	32
Accessing saved criteria .....	34
Accessing saved results .....	34

Sending search results .....	35
Exporting search results .....	35
Using quarantine repositories .....	36
Copying saved results to a quarantine repository .....	36
Deleting the contents of a quarantine repository .....	37
<b>5 Using the AuditLog .....</b>	<b>39</b>
The AuditLog repository .....	39
Searching the AuditLog .....	40
Completing the Search for field .....	41
AuditLog advanced search fields .....	42
Viewing AuditLogs .....	44
User actions in Web Interface .....	45
Duplicate Manager jobs .....	46
Administrative Delete actions .....	46
<b>6 Troubleshooting problems in the Web Interface .....</b>	<b>47</b>
Unable to display saved results .....	47
Corrupted documents .....	47
Wildcard errors in queries .....	47
Query and display issues in Japanese, Chinese, and Korean documents .....	48
Problems exporting email .....	49
Problems exporting files .....	49
<b>7 Query expression syntax and matching .....</b>	<b>51</b>
Query expressions .....	51
Word characters .....	51
Word characters and separators .....	52
Regular expression definition of English word characters .....	52
Letters and digits in different character sets .....	52
Letters and digits defined .....	52
Letters and digits in files .....	53
Matching words .....	54
Matching similar words .....	54
Fuzzy words .....	54
Measuring word similarity .....	55
Matching word sequences .....	55
Simple word sequences .....	55
Proximity word sequences .....	56
Matching word sequences in files and email attachments .....	56
Separators are ignored .....	56
Sequence is not intuitive .....	56
Boolean query expressions .....	58
Using the AND and OR operators .....	58
Using the NOT operator .....	59
Example of queries using NOT operator .....	59
Separators .....	60
Sub-expressions .....	60
Nested query expressions .....	60
Query expression examples .....	61
<b>A Indexed file types and MIME types .....</b>	<b>63</b>

<b>B Additional indexing detail for Microsoft Office .....</b>	<b>65</b>
Microsoft Office supported features .....	65
Microsoft Office supported file properties .....	66
<b>Index .....</b>	<b>69</b>

---

# Figures

1 IAP Web Interface toolbar .....	18
2 Simple Search page .....	21
3 Advanced Search page (email content type) .....	24
4 Advanced Search page (document content type) .....	26
5 Query Results page (email content type) .....	29
6 Query Results page (document content type) .....	30
7 Query results navigation bar .....	31
8 Save Criteria page .....	32
9 Save Results page .....	33
10 Saved Criteria view, Query Manager page .....	34
11 Saved Results view, Query Manager page .....	34
12 AuditLog advanced search .....	43
13 AuditLog query results .....	44
14 AuditLog for user actions in Web Interface .....	45
15 AuditLog for Duplicate Manager job .....	46
16 AuditLog for Administrative Delete actions in PCC .....	46
17 AuditLog entry for email deletion .....	46
18 Extra space in email Subject field .....	48

---

# Tables

1 Document conventions .....	9
2 Applications for users .....	13
3 Toolbar buttons, IAP Web Interface .....	18
4 IAP Web Interface tasks .....	19
5 Additional advanced search query fields (email) .....	24
6 Additional advanced search query fields (documents) .....	27
7 Query results navigation bar .....	31
8 Logged actions and descriptions .....	41
9 Additional fields for AuditLog repository searches .....	43
10 Supported character sets .....	53
11 Excel spreadsheet .....	57
12 Boolean AND and OR operators .....	58
13 Boolean NOT operator properly used in query .....	59
14 Query expression examples .....	61
15 IAP indexed file types and MIME types .....	63
16 Microsoft Office supported features .....	65
17 Microsoft Office supported properties .....	66





---

# About this guide

This guide provides information about using the IAP Web Interface to search for and retrieve archived documents.

For additional information on using Email Archiving software for Microsoft Exchange and IBM Domino, see the *HP Email Archiving software for Microsoft Exchange User Guide* and *HP Email Archiving software for IBM Domino User Guide* contained in those products.

## Intended audience

This guide is intended for users of the IAP Web Interface.

## Prerequisites

Prerequisites for using this product include knowledge of using Web browsers.

## Related documentation

In addition to this guide, HP provides the following on the IAP documentation CD:

- *HP Integrated Archive Platform Administrator Guide*
- *HP Integrated Archive Platform Release Notes*

Online help is also available for the IAP Platform Control Center (PCC) and IAP Web Interface. The IAP Web Interface online help is a subset of this guide. The IAP PCC online help is a subset of the *HP Integrated Archive Platform Administrator Guide*.

## Document conventions and symbols

**Table 1 Document conventions**

Convention	Element
Blue text: <a href="#">Table 1</a>	Cross-reference links and email addresses
Blue, underlined text: <a href="http://www.hp.com">http://www.hp.com</a>	Web site addresses
<b>Bold</b> text	<ul style="list-style-type: none"><li>• Keys that are pressed</li><li>• Text typed into a GUI element, such as a box</li><li>• GUI elements that are clicked or selected, such as menu and list items, buttons, tabs, and check boxes</li></ul>
<i>Italic</i> text	Text emphasis

Convention	Element
Monospace text	<ul style="list-style-type: none"> <li>• File and directory names</li> <li>• System output</li> <li>• Code</li> <li>• Commands, their arguments, and argument values</li> </ul>
<i>Monospace, italic</i> text	<ul style="list-style-type: none"> <li>• Code variables</li> <li>• Command variables</li> </ul>
Monospace, <b>bold</b> text	Emphasized monospace text

 **CAUTION:**

Indicates that failure to follow directions could result in damage to equipment or data.

 **IMPORTANT:**

Provides clarifying information or specific instructions.

 **NOTE:**

Provides additional information.

 **TIP:**

Provides helpful hints and shortcuts.

## Support

You can visit the HP Software Support Web site at <http://www.hp.com/go/hpsupport>.

HP Software Support Online provides an efficient way to access interactive technical support tools. As a valued support customer, you can benefit by using the support site to:

- Search for knowledge documents of interest
- Submit and track support cases and enhancement requests
- Download software patches
- Manage support contracts
- Look up HP support contacts
- Review information about available services
- Enter into discussions with other software customers
- Research and register for software training

Most of the support areas require that you register as an HP Passport user and sign in. Many also require a support contract.

To find more information about access levels, go to [http://support.openview.hp.com/new\\_access\\_levels.jsp](http://support.openview.hp.com/new_access_levels.jsp).

## Subscription service

HP strongly recommends that customers register online using the Subscriber's choice Web site: <http://www.hp.com/go/e-updates>.

Subscribing to this service provides you with email updates on the latest product enhancements, newest driver versions, and firmware documentation updates as well as instant access to numerous other product resources.



---

# 1 IAP overview

This section describes the HP Integrated Archive Platform (IAP) and the IAP Web Interface, which you will use to search for email or files archived on the system.

Major topics include:

- [About the IAP](#), page 13
- [Understanding document archiving](#), page 14
- [Understanding searching and document indexing](#), page 14
- [What the IAP indexes](#), page 15
- [What the IAP does not index](#), page 15

## About the IAP

The HP Integrated Archive Platform (IAP) is a fault-tolerant, secure system of hardware and software that archives files and email messages for your organization, and lets you search for archived documents. IAP provides the following main functions:

- Automatic, active data archiving (email and specific types of documents) to help your organization meet regulatory requirements.
- Interactive data querying to search for and retrieve archived data according to various criteria.

To interact with the system, you can use the following applications.

**Table 2 Applications for users**

Application	Tasks
Outlook Plug-In (customer option)	Search for, view, and work with archived email using Microsoft Outlook. Export email from the IAP to a PST file.
OWA Extension (customer option)	View and work with archived email using Outlook Web Access.
Notes Client Plug-In Local Cache Export Search (customer options, Windows users only)	View and work with archived email using IBM Lotus Notes. Access archived email offline while you are traveling. Export email from the IAP to a Notes database file.
DWA Extension (customer option)	View and work with archived email using Domino Web Access (iNotes).
IAP File Export (customer option)	Export files from the IAP to your computer or a network folder.

Application	Tasks
IAP Web Interface	<p>Use your Web browser to search for and view archived email and files on the IAP. Send email to your email account or download files to a local computer or network folder.</p> <p>Compliance officers: Place a legal hold on documents. View the compliance system log (AuditLog).</p> <p>(Customer option) Export email and files from the IAP when the appropriate export utility is installed on your system.</p>

The IAP Web Interface is available to all users. Outlook Plug-In, Local Cache, Export Search, OWA Extension, DWA Extension, and IAP File Export are independent customer options. Depending on the configuration of your system, each may or may not be available to you.

## Understanding document archiving

The IAP archives files and email messages associated with registered users.

Archiving involves physically storing copies of a document (file or email message), but also virtually storing it in one or more repositories. A repository is a virtual collection of documents associated with routing rules (for archiving) and user access control lists (for retrieving).

- Documents associated with a given user are archived to a given set of repositories. User-repository associations are defined by routing rules.
- A user has query and retrieval access to a given set of repositories. This is controlled by access control lists associated with each repository.

Most users have search and retrieval access only to their own documents, which are archived in their individual repositories.

In email archiving, the system can automatically archive to your individual repository all messages that you send or receive. In addition to being automatically routed to your individual repository, your email is probably also routed to other repositories established by your organization. For example, a company audit repository might be used to keep track of all company email. Some users have access to other repositories besides their own. For example, your manager or supervisor might have access to your repository.

## Understanding searching and document indexing

You can search for any email message or file archived in your repository or other repositories to which you have access. However, you can search the contents of a document only if the contents have been indexed. Indexing a document's contents involves cataloging the document words to prepare them for later searching. When you search for a document, your query is checked against an index of words that is updated each time a document is archived.

---

### NOTE:

There is a delay from the time that documents are archived to the time they are indexed. Depending on the system configuration, documents may not be available for search for at least two hours after they are archived.

---

# What the IAP indexes

The following sections explain what is indexed by the IAP and can be searched for in the Web Interface.

## Indexed file types

You can search the contents of email messages and the file types listed below. For a complete list of the file types that are indexed by the IAP, see [“Indexed file types and MIME types”](#) on page 63.

- Plain text files
- Rich text files (.rtf)
- Documents in TNEF format (for Microsoft Outlook and Exchange)
- HTML (HyperText Markup Language) files
- Files used by the following Microsoft Office programs: Word, Excel, and PowerPoint  
We support indexing of these file types for Microsoft Office 2007 and prior releases of Office. See [“Additional indexing detail for Microsoft Office”](#) on page 65 and [“Limitations for Microsoft Office items”](#) on page 16 for more information on what is and is not indexed in Microsoft Office documents.
- Files used by the following Corel WordPerfect Office programs: WordPerfect, Quattro Pro, Presentations  
We support indexing of these file types for WordPerfect Office X3 and prior releases of Office.
- PDF (Portable Document Format) files viewed with Adobe Acrobat Reader
- ZIP files
- Embedded messages (RFC 822 messages)  
For ZIP files and embedded messages, the content inside the files is expanded and indexed.

## Message MIME types

An email message can contain message parts of different MIME (Multipurpose Internet Mail Extensions) content-types. The IAP indexes the content-types shown in [“Indexed file types and MIME types”](#) on page 63. Each content-type corresponds to one of the indexed file types.

An email message that is entirely plain text, not MIME, is also indexed.

# What the IAP does not index

The following sections explain what the IAP does not index.

## Document markup and formatting

Separators (such as punctuation) between words are ignored during indexing. Invisible source-code words, such as HTML markup tags, are also ignored in indexing.

Document formatting usually has no bearing on indexing, and only the words you see in an email or file are indexing candidates. However, when a dropped cap (a large initial capital letter) is used in a Microsoft Word document, the word with the dropped cap is indexed as two separate words. This is because Word puts a dropped cap into a text box to set it off from the surrounding paragraph.

In the example below, XYZcorp would be indexed as “X” and “yzcorp.”

**X**YZcorp has posted strong results in the second quarter. Chief Executive Officer Bob Brown announced today that the company is on track to beat

## Non-indexed file types

The following file types are not indexed:

- Graphic files
- Music files
- Video files

Depending on the way your IAP system is configured, these types of files might be archived, but they are not indexed. You can search for them only by using external identifying information, such as the file name or file extension.

## Limitations for Microsoft Office items

### Non-indexed Microsoft Office files

The following types of Microsoft Office files are not indexed by the IAP:

- Microsoft Access files
- Microsoft Project files
- Microsoft Visio files

If these files are archived, you can search for them only by using external identifying information, such as the file name or file extension.

### Limitations for Microsoft Word items

Some items in Microsoft Word documents are not indexed by the IAP:

- Macros: In Microsoft Word 2003 and 2007 documents, macros are not indexed. You cannot search on text in macros.
- Microsoft Word 2003 documents that use the versioning feature: Only the text in the latest version of the document is indexed. You cannot search on text that is included solely in an earlier version of the document.

### Limitations for Microsoft Office 2007 items

Office 2007 documents that were archived before IAP 1.6.1 or 2.0 was installed are not indexed or content searchable.

In addition, some documents converted to Microsoft Office 2007 by the Office File converter might not be properly indexed.



---

## 2 Introducing the IAP Web Interface

Use the Web-based search tool to search for email and documents archived in the HP Integrated Archive Platform (IAP) system.

This section explains the basics of the IAP Web Interface.

- [Logging in and out](#), page 17
- [Understanding the user interface](#), page 17
- [Common tasks](#), page 19
- [Changing your password](#), page 20
- [Changing your language](#), page 20

### Logging in and out

Before logging in for the first time, see your system administrator for the URL to use.

The IAP Web Interface can be accessed with any of the following supported Web browsers:

- Microsoft Internet Explorer for Windows versions 7.x and 8.x
- Mozilla Firefox version 3.0.11 and above

The IAP Web Interface can be accessed from your computer using regular HTTP protocol as well as secure HTTP protocol (HTTPS), for better protection during authentication. The IAP Web Interface supports HTTPS by default. If you use regular HTTP, you will be redirected automatically to HTTPS protocol. Once logged in, you can use regular HTTP.

To access the IAP Web Interface:

1. In the Address field of your Web browser, enter the URL (Web address) that was provided to you by your system administrator.  
The Web browser displays a login screen.
2. Enter your user name and password (provided by your system administrator).  
Passwords are case-sensitive. If you need to start over, click **Reset** to clear the Username and Password boxes.
3. Select your language.
4. Click **Login**.  
The Simple Search page is displayed.
5. To log out, click **LogOut** in the toolbar.

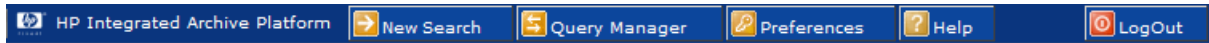
### Understanding the user interface

User interface topics include:

- [Using the toolbar](#), page 18
- [Search basics](#), page 18

## Using the toolbar

Each page of the IAP Web Interface has a toolbar at the top.



**Figure 1 IAP Web Interface toolbar**

The following table describes each button:

**Table 3 Toolbar buttons, IAP Web Interface**

Button	Description
New Search	Click to display the Simple Search page, where you can submit a query. See <a href="#">“Completing simple searches”</a> on page 21. To display the Advanced Search page, point to this button and click <b>Advanced search</b> from the menu. See <a href="#">“Completing advanced searches”</a> on page 22.
Query Manager	Click to display the Query Manager page, where you can display saved queries (see <a href="#">“Accessing saved criteria”</a> on page 34) and search results (see <a href="#">“Accessing saved results”</a> on page 34).
Preferences	Click to display the Preferences page, where certain users can change their password for accessing the IAP Web Interface. See <a href="#">“Changing your password”</a> on page 20. You can also use the Preferences menu to change the language of the user interface. See <a href="#">“Changing your language”</a> on page 20. Supported languages include English (United States), English (United Kingdom), French, German, Spanish, Portuguese, Chinese (Traditional), Chinese (Simplified), Korean, and Japanese.
Help	Click for online help about the IAP Web Interface.
LogOut	Click to log out of the IAP Web Interface.

## Search basics

### Using the Search for field

The Search for field is available when using the Simple Search and Advanced Search pages. Use this field to search for specific words in a document (email messages or files).

The query syntax allowed is described in [“Query expression syntax and matching”](#) on page 51. You can enter simple words, words with wildcards, or a more sophisticated query involving Boolean expressions or word sequences.

---

 **NOTE:**

To narrow the search to the documents you want to find, make your search text as specific as possible. In general, the more information provided in the Search for field, the narrower the search. If the field is blank (empty), all documents within the specified date range of the query are returned.

---

## Searching indexed contents

The Search for field is checked for a match against the indexed contents of documents.

- For email messages, the Search for field searches for words in the message body, but *not* in other message fields such as From or To. The Search for field can also be used to search indexed message attachments; for example, the text in a Microsoft Word file.
- For files, the Search for field applies only to indexed document files. (Files such as graphics or music files do not have indexed contents and cannot be searched.)

## Common tasks

Use the following table as a quick reference for performing common tasks in the Web Interface.

**Table 4 IAP Web Interface tasks**

<b>Task</b>	<b>Reference</b>
Search for archived documents	<a href="#">“Completing simple searches”</a> on page 21 and <a href="#">“Completing advanced searches”</a> on page 22
Display or print the search results	<a href="#">“Displaying search results”</a> on page 29
Save search criteria	<a href="#">“Saving search criteria”</a> on page 32
Save the results of a search	<a href="#">“Saving search results”</a> on page 32
Display or delete saved search criteria	<a href="#">“Accessing saved criteria”</a> on page 34
Display the saved results of a search	<a href="#">“Accessing saved results”</a> on page 34
Send items to your email account	<a href="#">“Sending search results”</a> on page 35
Export email or files	<a href="#">“Exporting search results”</a> on page 35
Save archived documents to a quarantine repository (legal hold)	<a href="#">“Copying saved results to a quarantine repository”</a> on page 36
Delete a quarantine repository	<a href="#">“Deleting the contents of a quarantine repository”</a> on page 37
Search the AuditLog repository	<a href="#">“The AuditLog repository”</a> on page 39
Change the Web Interface language	<a href="#">“Changing your language”</a> on page 20
Change your Web Interface password (system administrators and compliance users only)	<a href="#">“Changing your password”</a> on page 20
Troubleshoot problems	<a href="#">“Troubleshooting problems in the Web Interface”</a> on page 47

# Changing your password

---

 **NOTE:**

This option does not appear if you cannot change your password from the Web Interface.

---

Depending on how your system is configured, your password is usually the same as your Windows or your Lotus Notes password. Users who have a local account created in the IAP, such as system administrators or compliance officers, have the option of managing their password through the Web Interface.

To change your password for accessing the IAP Web Interface:

1. Click **Preferences** in the toolbar.
2. Click **Change Password**.
3. Enter your current password (Old Password), enter the New Password twice, and then click **Change**.

Change your password immediately after you log in for the first time. For security reasons, change your password periodically.

# Changing your language

You can change the language of the Web Interface by following these steps:

1. Click **Preferences** in the toolbar.
2. Click **Change Language**.
3. Enter the language you prefer.

Supported languages are English, French, German, Spanish, Portuguese, Chinese (Traditional), Chinese (Simplified), Korean, and Japanese.

---

# 3 Searching for email or files

This section explains how to conduct simple searches or more advanced searches for documents that are archived in the IAP.

- [Completing simple searches](#), page 21
- [Completing advanced searches](#), page 22

## Completing simple searches

Use the Simple Search page to search for email or files using a pre-set time frame and words entered in the Search for field. The Simple Search page is simpler than the Advanced Search page because there are fewer fields you can search on.

To complete a simple search:

1. Click **New Search** in the toolbar. The Simple Search page is displayed.  
(The Simple Search page also appears after you log in to the Web Interface.)

---

Content Type:  ?

Timeframe:  ?

Where to Search:  ?

Search for:

---


**Figure 2 Simple Search page**

2. Search using the following fields on the Simple Search page:

- Content Type:
  - Use **email** to search for email messages.
  - Use **document** to search for files that were migrated using the HP File Archiving software (formerly known as FMA).  
Using “document” does not search for attachments in email sent by Microsoft Outlook or Lotus Notes. Use “email” to search for email attachments.  
Compliance officers: For instructions on searching the AuditLog repository, see [“The AuditLog repository”](#) on page 39.
- Timeframe: The time period to search. Select the time period from the list.
- Where to Search: The repository to search. Only the repositories to which you have access are displayed. For example, you can search for your own archived email, but you cannot search for John Smith's email unless your system administrator has given you access to John Smith's repository. The repositories are listed in alphabetical order.
- (Optional) Search for: Searches for words in the message body, message attachment, or file.



**NOTE:**

For information about a search field, momentarily hold your mouse pointer over  to display a tooltip.

---

3. When you have finished defining your query, click **Find Now** to start the search.

(If you find that you need to perform an advanced search instead, click Advanced Search next to the Find Now button.)

The Query Results page is displayed when the search is complete.

Your search criteria is preserved until you create another query in the Web Interface session.

## Completing advanced searches

For more specific searches, use the Advanced Search page. The Advanced Search page provides added query fields to help refine your search.

To complete an advanced search:

1. Point to **New Search** in the toolbar and click **Advanced search** from the menu.


The Advanced Search page is displayed.

2. Search using the following fields on the Advanced Search page:

- Content Type:
  - Use **email** to search for email messages.
  - Use **document** to search for files that were migrated using the HP File Archiving software (formerly known as FMA).  
Using “document” does not search for attachments in email sent by Microsoft Outlook or Lotus Notes. Use “email” to search for email attachments.  
Compliance officers: For instructions on searching the AuditLog repository, see “[The AuditLog repository](#)” on page 39.
- Search for: Searches for words in the message body, message attachment, or file. You can enter simple words, words with wildcards (\*), or a more sophisticated query involving Boolean expressions or word sequences.
- By TimeFrame: The time period to search. This includes the last-modified date of a file or the date an email message was sent.  
You can choose a predefined time period, or define a specific time period to search by entering the Start and end (To) dates. For example, to search for documents dated between March 8, 2009 and March 23, 2009, enter **03/08/2009** in the Start field and **03/23/2009** in the To field.
- Where to Search: The repository to search. Only the repositories to which you have access are displayed. For example, you can search for your own archived email, but you cannot search for John Smith's email unless your system administrator has given you access to John Smith's repository. The repositories are listed in alphabetical order.

---

 **NOTE:**

For information about a search field, momentarily hold your mouse pointer over  to display a tooltip.

---

3. To refine the search, use one or more of the additional query fields, as shown in [Email Content Type](#) or [Document Content Type](#).

The available fields depend on the Content Type you select.

4. When you have finished defining your query, click **Find Now** to start the search.

(If you find you do not need to perform an advanced search, click Simple Search next to the Find Now button to switch to the Simple Search form.)

The Query Results page is displayed when the search is complete.

Your search criteria is preserved until you create another query in the Web Interface session.

---

 **NOTE:**

You can use wildcards (\* or ?) in the Advanced Search fields, but do **not** use a wildcard for the first character in a field. For more information on using wildcards, see “[Matching words](#)” on page 54.

For more information on composing IAP queries, see “[Query expression syntax and matching](#)” on page 51.

If you receive an error message after submitting a query, see “[Wildcard errors in queries](#)” on page 47.

---

# Email Content Type

## Step #1 - Search Type:

Content Type:  ?

## Step #2 - Field Search Criteria:

Search for:  ?

Subject:  ?

From:  ?

To / Cc:  ?

Address:  ?

Folder Name:  ?

Attachment Name:  ?

MessageID:  ?

## Step #3 - Time Search Criteria:

By Timeframe:  ?

By Date:

Start:  [mm/dd/yyyy] ?

To:  ?

## Step #4 - Search Location:

Where to Search:  ?

Find Now

Simple Search

Figure 3 Advanced Search page (email content type)

Table 5 Additional advanced search query fields (email)

Query field	Matches
Subject	Enter one or more words from the email's Subject field.
From	Enter the email address of the person who sent the message. You can use a partial email address with a wildcard. The From field is disabled if the Address search field is completed.
To / Cc	Enter the email address of a person who received the message or was copied in the message. You can use a partial email address with a wildcard. The To/Cc field is disabled if the Address search field is completed. If you are a user with "compliance" authorization, this value will match recipients in the Rcpt To, To, Bcc, Cc, or Apparently-To message fields. If you are not a legal or compliance officer, this value will match recipients in the To, Cc, and Apparently-To message fields.



Query field	Matches
Address	<p>Search for messages that were sent to or from a person by entering the person's email address. You can use a partial email address with a wildcard.</p> <p>The Address field is disabled if either the From or To/CC search field is completed.</p> <p>If you are a user with "compliance" authorization, this value will match addresses in the From, Rcpt To, To, Bcc, Cc, or Apparently-To message fields.</p> <p>If you are not a legal or compliance officer, this value will match recipients in the From, To, Cc, and Apparently-To message fields.</p>
Folder Name	<p><b>NOTE:</b></p> <p>This field only appears if your IAP system records the name of the Outlook folder where an email was located. The Folder Name field is not available for messages archived from Lotus Notes.</p> <p>To complete the field, enter the name of the Outlook folder.</p> <p>Use quotation marks to search for a folder that has spaces in the name. For example: <b>"my folder"</b> or <b>"Sent Items"</b>. (This does not apply to Japanese, Chinese, or Korean email folder names. See <a href="#">"Query and display issues in Japanese, Chinese, and Korean documents"</a> on page 48.)</p> <p>Example queries:</p> <p><b>2006</b> searches all folders and subfolders with the name 2006; for example, \Inbox\2006 and \Inbox\test\2006, but not \Inbox.</p> <p><b>\Inbox\2006</b> searches only folder 2006 in path \Inbox\2006.</p> <p><b>\Inbox\*</b> searches Inbox and all its subfolders and nested folders.</p> <p><b>\2006</b> searches only the root folder \2006.</p> <p><b>\Inbox\2006\personal</b> searches only the folder personal in the path \Inbox\2006\personal.</p> <p><b>2006\personal</b> searches the folder 2006\personal anywhere in the path.</p> <p><b>2006\personal\*</b> searches the folder 2006\personal and all its subfolders and nested folders.</p> <p><b>sent*</b> searches all folders that start with sent, for example, Sent Items and SentEmail. Search is not case-sensitive.</p> <p><b>bugs\20*</b> searches all folders in "bugs" that start with 20, for example \bugs\2006 and \bugs\20data.</p> <p>Searching in core folders (Outlook hard-coded folders like Inbox, Sent Items, or Sync Issues\Conflicts):</p> <p>If the core folder only has one level (like Inbox), there is no search limitation. If the core folder has more than one level (like Sync Issues\Conflicts) and you want to search the emails in the second level (Conflicts), enter both levels or use wildcards. Searching for <b>"Sync Issues\Conflicts"</b> will find email in Sync Issues\Conflicts. Searching for <b>"Sync Issues\Con*"</b> will find email in Sync Issues\Conflicts. However, searching for <b>Conflicts</b> will not find email in Sync Issues\Conflicts.</p>

Query field	Matches
Attachment Name	<p>Enter the file name of an email attachment. (The contents of indexed email attachments are searched using the Search for field.) You can use a partial file name with a wildcard.</p> <p><b>NOTE:</b></p> <p>If an email attachment contains other files, you can only perform a file name search on the “top level” attachment.</p> <p>For example, if a ZIP file is attached to an email, you can search on the ZIP file name but not the name of a Word file contained in the ZIP. (The contents of the Word file can be searched using the Search for field.)</p>
MessageID	<p>Enter the MessageID (message identification number) from Outlook. (Not all messages have MessageIDs.) This field is used mainly for compliance searches.</p> <p>To display the MessageID field in Outlook:</p> <ol style="list-style-type: none"> <li>1. Double-click to open the message in its own window.</li> <li>2. Select <b>View &gt; Options</b>. The Message Options dialog box is displayed.</li> </ol> <p>If the message has a MessageID, the field is shown in the Internet headers field of the Message Options dialog box. Example: Message-ID: &lt;LISTMANAGER-115380-9228-2003.03.04-17.34.24-user#hp.com@lists.FrameUsers.com&gt;.</p>

## Document Content Type

### Step #1 - Search Type:

Content Type:  ?

### Step #2 - Field Search Criteria:

Search for:  ?

Document Name:  ?

Document Path:  ?

Extension:  ?

Title:  ?

Author:  ?

### Step #3 - Time Search Criteria:

By Timeframe:  ?

By Date:

Start:  ?  
[mm/dd/yyyy]

To:  ?

### Step #4 - Search Location:

Where to Search:  ?

Figure 4 Advanced Search page (document content type)

**Table 6 Additional advanced search query fields (documents)**

<b>Query field</b>	<b>Matches</b>
Document Name	Enter the file name, not including the file extension.
Document Path	<p>Enter the document's original file path.</p> <p>As for any other text query field, separators such as slash ( / ), backslash ( \ ), and colon ( : ) are ignored, and the query words are searched in any order. For example, query text <code>c:\abc\xyz</code> will match path <code>abc:\xyz\c</code>, as well as path <code>c:\abc\xyz</code>.</p> <p>To ensure that path components are searched in order, enclose the text in double quotation marks. For example, "<code>c:\abc\xyz</code>." See "<a href="#">Query expression syntax and matching</a>" on page 51 for more information on the IAP query language.</p>
Extension	Enter the file extension. Example: doc for a Microsoft Word file.
Title	Enter the title of the document. Only some files have associated titles. For example, to see the title of a Word document, select <b>File &gt; Properties</b> in Word. The Title field is shown on the Summary panel of the displayed Properties dialog box.
Author	Enter the author of the document. Only some files have associated authors. For example, to see the author of a Word document, select <b>File &gt; Properties</b> in Word. The Author field is shown on the Summary panel of the displayed Properties dialog box.



# 4 Working with search results

After completing a search, there are many ways to work with the results of the search. This section explains how to save the items you find, save the criteria for the search, send message copies to your email account, and create quarantine repositories for legal holds.

- [Displaying search results](#), page 29
- [Saving search criteria](#), page 32
- [Saving search results](#), page 32
- [Accessing saved criteria](#), page 34
- [Accessing saved results](#), page 34
- [Sending search results](#), page 35
- [Exporting search results](#), page 35
- [Using quarantine repositories](#), page 36

## Displaying search results

The Query Results page is displayed after you perform one of the following actions:

- Submit a simple or advanced search (see [“Completing simple searches”](#) on page 21 or [“Completing advanced searches”](#) on page 22).
- Submit a search from previously saved criteria (see [“Accessing saved criteria”](#) on page 34).
- Access previously saved results (see [“Accessing saved results”](#) on page 34).

The examples below show sample Query Results pages. (Note that in email results, the Folder column only appears if your system records the Outlook folder in which the email was located.)

You are logged in as JoeUser

Query results complete  
Number of hits: 7 [ 1 - 7 ]

**More Options:**

**Repository :** JoeUser Repository

<input type="checkbox"/>	From:	Subject:	Folder:	Date:
<input type="checkbox"/>	"wendtest013000@itchyexch.com" <wendtest013000@itchyexch.com>	Mail #13000 from E-Sender @bianca-sender1	\Inbox	03/18/2009 07:04:04
<input type="checkbox"/>	"wendtest014000@itchyexch.com" <wendtest014000@itchyexch.com>	Mail #14000 from E-Sender @bianca-sender1	\Inbox	03/11/2009 03:57:24
<input type="checkbox"/>	"wendtest014500@itchyexch.com" <wendtest014500@itchyexch.com>	Mail #14500 from E-Sender @bianca-sender1	\Inbox	03/07/2009 13:24:04
<input type="checkbox"/>	"wendtest017000@itchyexch.com" <wendtest017000@itchyexch.com>	Mail #17000 from E-Sender @bianca-sender1	\Inbox	02/17/2009 17:37:24
<input type="checkbox"/>	"wendtest019500@itchyexch.com" <wendtest019500@itchyexch.com>	Mail #19500 from E-Sender @bianca-sender1	\Inbox	01/30/2009 21:50:44
<input type="checkbox"/>	"wendtest020000@itchyexch.com" <wendtest020000@itchyexch.com>	Mail #20000 from E-Sender @bianca-sender1	\Inbox	01/27/2009 08:17:24
<input type="checkbox"/>	"wendtest023500@itchyexch.com" <wendtest023500@itchyexch.com>	Mail #23500 from E-Sender @bianca-sender1	\Inbox	01/02/2009 09:24:04

Figure 5 Query Results page (email content type)

You are logged in as JoeUser

Query results complete  
 <<<<<< [Progress Bar] >>>>>>  
 Number of hits: 10 [1 - 10]

**More Options:**

Repository : JoeUser Repository

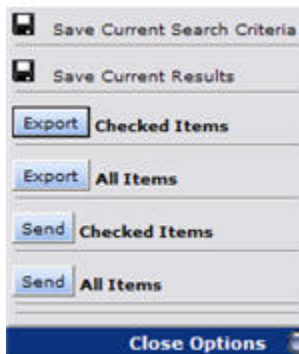
<input type="checkbox"/>	File Name:	File Size:	File Path:	Date:
<input type="checkbox"/>	asftest1.txt	23	E:\DOCS\bxt\	04/09/2009 15:21:53
<input type="checkbox"/>	asftest10.txt	24	E:\DOCS\bxt\	04/09/2009 15:21:53
<input type="checkbox"/>	asftest2.txt	23	E:\DOCS\bxt\	04/09/2009 15:21:53
<input type="checkbox"/>	asftest3.txt	23	E:\DOCS\bxt\	04/09/2009 15:21:53
<input type="checkbox"/>	asftest4.txt	23	E:\DOCS\bxt\	04/09/2009 15:21:53
<input type="checkbox"/>	asftest5.txt	23	E:\DOCS\bxt\	04/09/2009 15:21:53
<input type="checkbox"/>	asftest6.txt	23	E:\DOCS\bxt\	04/09/2009 15:21:53
<input type="checkbox"/>	asftest7.txt	23	E:\DOCS\bxt\	04/09/2009 15:21:53
<input type="checkbox"/>	asftest8.txt	23	E:\DOCS\bxt\	04/09/2009 15:21:53
<input type="checkbox"/>	asftest9.txt	23	E:\DOCS\bxt\	04/09/2009 15:21:53

**Figure 6 Query Results page (document content type)**

You can complete any of the following tasks from the Query Results page:

- To display the contents of an email in the viewing pane, click the item from the list once. Clicking the item twice will open the preview pane as a new window.
- To display the contents of a document (file), click the item in the list, click **Download**, and then click **Open** in the File Download dialog box.
- To download a file, click the item in the list and then click **Download**. Click **Save** in the File Download dialog box, browse to a location on your computer or network, and then click **Save**.
- To display a different group of results, click the different symbols in the query results navigation bar. See “[Query results navigation bar](#)” on page 31 for more information.
- To select all displayed results or clear all selected results, select or clear the check box in the query results header.
- To print the Query Results page, use the print button in your Web browser.

Click **More Options** on the Query Results page to perform the following tasks:



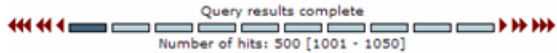
- To save the current criteria , see “[Saving search criteria](#)” on page 32.
- To save the current results, see “[Saving search results](#)” on page 32.
- To mail all or selected results to your mailbox, see “[Sending search results](#)” on page 35.
- To export all or selected results, see “[Exporting search results](#)” on page 35.

## Query results navigation bar

When search results are retrieved, the most recent documents are displayed first.

Documents that were archived within the past two hours might not appear in the results, depending on your system's configuration.

















Fifty results (maximum) are shown on the Query Results page. You can use the query results navigation bar to display different groups of 50 results.




**Figure 7** Query results navigation bar

The query results navigation bar shows the following information:

**Table 7** Query results navigation bar

Item	Description
bars: 	<p>From left to right, the five bars represent subsequent pages of 50 results (maximum). Click a bar to display its page of results. The dark bar represents the currently displayed results.</p> <p>To see which documents a given bar represents, momentarily hold the mouse pointer over the bar to display a tooltip.</p>
arrows: <ul style="list-style-type: none"> <li>• </li> <li>• </li> <li>• </li> <li>• </li> <li>• </li> <li>• </li> </ul>	<p>Click an arrow to display a different page of results:</p> <ul style="list-style-type: none"> <li>• Move the results display forward by 50 () , 100 () , or 500 () documents.</li> <li>• Move the results display back by 50 () , 100 () , or 500 () .</li> </ul> <p>For example, if the current page shows results 1-50, clicking the right double-arrow () displays results 100-150.</p>
status	<p>Text indicating the current status of results retrieval:</p> <ul style="list-style-type: none"> <li>• Query Still In Progress: Searching is in progress. Less than 50 results have been found so far.</li> <li>• Click Above: Searching is in progress. At least 50 results have been found and displayed. You can click a navigation bar () to display a set of 50 results while the search is in progress.</li> <li>• Query Results Complete: Searching is complete if the query produces no more than 500 query results. If there are more than 500 results, the search for the first 500 results is complete. To search for additional results for the query, click the right triple-arrow symbol () to retrieve all possible results.</li> </ul>

## Result retrieval batches

When you submit a query or search, the first 500 results are retrieved. To see more results for a query that returns more than 500 results, you must click the right triple-arrow symbol () . When there are more than 500 results, the Query Results Complete status message means that the first batch of 500 results has been retrieved. It does not mean that all results have been retrieved.

## Saving search criteria

After you submit a search, you can save the search criteria.

To save criteria:

1. Display the Query Results page by completing one of the following tasks:
  - Submit a simple or advanced search (see [“Completing simple searches”](#) on page 21 or [“Completing advanced searches”](#) on page 22).
  - Access previously saved results (see [“Accessing saved results”](#) on page 34).
2. From the Query Results page, click **More Options**, and then click **Save Current Search Criteria**. Or right-click and select **Save criteria**. The Save Criteria page is displayed.

Save Query Criteria as:

---

**Where:**

**Type:**

**Search For:**

**Subject:**

**From:**

**To:**

**Address:**

**Folder:**

**Attachments:**

**MessageID:**

**Timeframe:**  to

**Figure 8 Save Criteria page**

3. Enter the name of the criteria you are saving in the Save Query Criteria as field. (To erase text entered in the Save Query Criteria as field, click **Clear**.)  
Special characters @ \$ % ^ & \* # ( ) [ ] / \ { + } ` ~ = | are not allowed.
4. Click **Save Now!**
5. To access the saved criteria, see [“Accessing saved criteria”](#) on page 34.

## Saving search results

After you submit a search, you can save the results. The search results are saved for one week and then deleted. The Web Interface also contains a button for manually deleting saved search results. It is possible that items in saved search results might be deleted from the IAP before the week is over if their retention period has expired. See [“Unable to display saved results”](#) on page 47.

The one week period does not apply to search results that are placed in a legal hold. See [“Using quarantine repositories”](#) on page 36 for more information.



---

 **NOTE:**

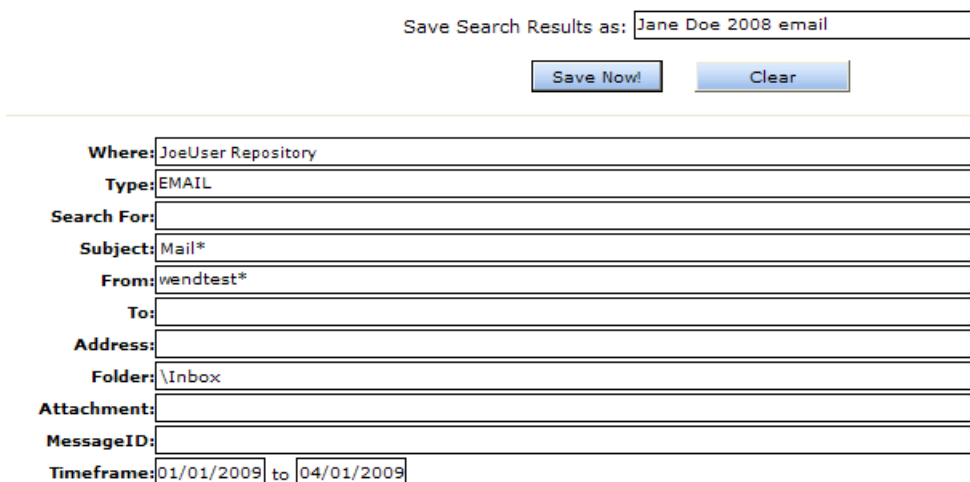
Deleting search results does not delete the items on the IAP. The actual items remain on the IAP according to the retention period set by your system administrator.

---

If a search locates a large number of documents, saving the results is useful. The query is resubmitted as a background process. Because the query runs in the background, you can continue to use the IAP Web Interface (for example, to submit other queries) and then retrieve the results at a later time.

To save results:

1. Display the Query Results page by completing one of the following tasks:
  - Submit a simple or advanced search (see “[Completing simple searches](#)” on page 21 or “[Completing advanced searches](#)” on page 22).
  - Submit a search from previously saved criteria (see “[Accessing saved criteria](#)” on page 34).
2. From the Query Results page, click **More Options**, and then click **Save Current Results**. Or right-click and select **Save results**. The Save Results page is displayed.



Save Search Results as:

Where:

Type:

Search For:

Subject:

From:

To:

Address:

Folder:

Attachment:

MessageID:

Timeframe:  to

**Figure 9 Save Results page**

3. Enter the name of the results you are saving in the Save Search Results as field. The name should not exceed 60 characters if you plan to place the results in a legal hold (quarantine).  
(To erase text entered in the Save Search Results as field, click **Clear**.)  
Special characters @ \$ % ^ & \* # ( ) [ ] / \ { + } ` ~ = | are not allowed.
4. Click **Save Now!**
5. To access the saved results, see “[Accessing saved results](#)” on page 34.

---

 **NOTE:**

The results are not retrieved in chronological order, but are sorted chronologically after they have all been retrieved (query processing is finished). If you access saved results before the query is finished, the results are not sorted.

---

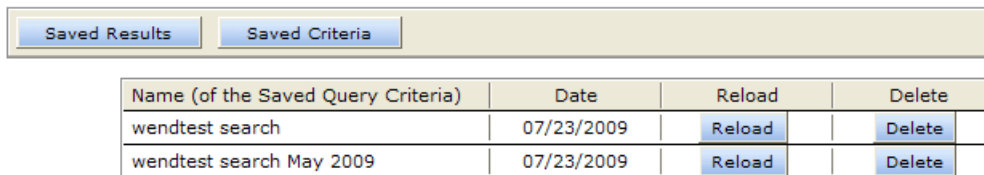
## Accessing saved criteria

If you save the criteria used to define a query, you can access the saved criteria from the Query Manager page. Each item listed shows the name of the saved criteria and the date when you saved the criteria.

Twenty sets of query criteria can be saved at any one time. After that, a set must be deleted before you can save a new set of query criteria.

To access saved criteria:

1. Click **Query Manager** in the toolbar.
2. On the Query Manager page, click **Saved Criteria** to display all previously saved searches.



Name (of the Saved Query Criteria)	Date	Reload	Delete
wendtest search	07/23/2009	Reload	Delete
wendtest search May 2009	07/23/2009	Reload	Delete

Figure 10 Saved Criteria view, Query Manager page

3. Complete any of the following tasks:
  - To resubmit the query, click **Reload**. The Advanced Search page with the saved criteria already entered is displayed. You can then resubmit the query by clicking **Find Now**.
  - To delete the saved criteria, click **Delete**. You are not prompted to confirm the deletion.
  - To switch to the Saved Results view, click **Saved Results**. Or right-click and select **Saved Query Results**.

## Accessing saved results

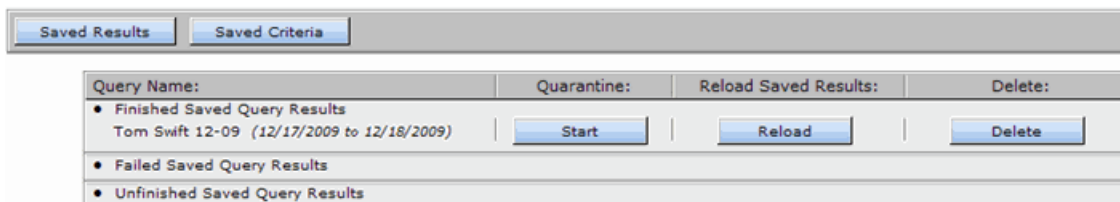
You can access saved results from the Query Manager page. This includes results that have been placed in a quarantine repository. (See “Using quarantine repositories” on page 36 for information about those repositories.)

Each item listed shows the name of the result, its status (if unfinished), and the start and end dates of the search.

To access saved results:

1. Click **Query Manager** in the toolbar.

The default Query Manager page displays all saved results. You can also access this view by clicking **Saved Results** on the Query Manager page.



Query Name:	Quarantine:	Reload Saved Results:	Delete:
• Finished Saved Query Results Tom Swift 12-09 (12/17/2009 to 12/18/2009)	Start	Reload	Delete
• Failed Saved Query Results			
• Unfinished Saved Query Results			

Figure 11 Saved Results view, Query Manager page

2. Complete any of the following tasks:
  - To display the results, click **Reload**. The Query Results page is displayed.
  - To copy the saved results to a quarantine repository, click **Start**. A completed message appears in the row when the contents are copied to the quarantine repository.
  - To delete the saved results, click **Delete**.
  - To switch to the Saved Criteria view, click **Saved Criteria**. Or right-click and select **Saved Query Criteria**.

## Sending search results

You can send copies of messages that are found in a search to your email account.

To send results:

1. Display the Query Results page by completing one of the following tasks:
  - Submit a simple or advanced search (see “[Completing simple searches](#)” on page 21 or “[Completing advanced searches](#)” on page 22).
  - Submit a search from previously saved criteria (see “[Accessing saved criteria](#)” on page 34).
  - Access previously saved results (see “[Accessing saved results](#)” on page 34).
2. From the Query Results page, select the check box next to each item you want to send. Skip this step if you are sending all items on the page.
3. Click **More Options** to open the Options menu.
4. To send all results on the Query Results page, click **Send All Items**. To send the selected items, click **Send Checked Items**. A confirmation message is displayed when the items are sent.

## Exporting search results

An export utility is required to export archived email or files from the IAP.

### Email Content Type

- Exporting email to Microsoft Outlook: The HP EAs Exchange Outlook Plug-In must be installed on your computer. Microsoft .NET Framework version 2.0 or later is required for the utility to be installed.
- Exporting email to IBM Lotus Notes: Two methods are supported. The HP EAs Domino Export Search utility can be installed on your computer, or a Web application can be used to export email. Your system administrator will advise you of the method that your organization supports. Java Runtime Environment version 1.6 or later is required for the Export Search utility to be installed on your computer.

Information on exporting archived email to Outlook or Lotus Notes is detailed in the *HP EAs for Microsoft Exchange User's Guide* and *HP EAs for IBM Lotus Domino User's Guide*. The user guides are included on the documentation CD in those products.

## Document Content Type (files)

Several requirements must be met to export archived files from the IAP:

- File export must be configured on the IAP system.
- The IAP File Export utility must be installed on your computer. (Microsoft .NET Framework version 2.0 or later is required for the utility to be installed.)
- The destination drive for the exported files must have at least 5 GB of free space available.

For information on exporting archived files from the IAP, click **Help** in the IAP File Export utility.

## Using quarantine repositories

Documents are kept in the IAP according to the retention policy specified by your organization. After an item's retention period has expired, it is automatically deleted from the system.

Documents that are placed in a special repository called a quarantine repository are not subject to automatic deletion. As long as items are in this repository, they are preserved in a legal hold and cannot be deleted when their retention period expires.

When the hold is no longer necessary, the contents of the quarantine repository can be deleted. This action does not delete the items from the IAP. It releases documents from the legal hold and makes them subject once again to the retention period that was set for them.

This section describes how to create a quarantine repository and how to delete the repository when items no longer need to be saved.

## Copying saved results to a quarantine repository

Use the instructions below to create a quarantine repository from saved query results. You can create as many quarantine repositories as you need.



### NOTE:

A saved query name should not have more than 60 characters. If the limitation is exceeded, you will not be able to create a quarantine repository for the results.

If this situation occurs, reload the saved results, re-save them with a name less than 60 characters, and then copy the saved results to a quarantine repository using the procedure below.

---

To create a quarantine repository from saved query results:

1. Click **Query Manager** in the toolbar to access the saved results.
2. In the Query Name column, locate the row with the saved results.
3. Click **Start** in the Quarantine column to copy those saved results to a quarantine repository.
4. Click **OK** in the dialog box that appears.

"Quarantine in Progress" appears after the results name. When the copy is complete, "Quarantine Completed" appears after the results name and the quarantine repository is shown in the Where to Search list on the search page.

You have access automatically to the quarantine repositories you create, and you can work with quarantine repositories as you would with any repository. For instance, you can select a quarantine repository from the Where to Search list on the search page. The repository is listed under the name you gave the saved results. For example. "John Doe results 6-22-09."

## Deleting the contents of a quarantine repository

Once the legal hold is not required, you can delete the quarantine repository. The documents in the repository return to the control of the retention manager. If an item's retention period has not expired, it remains archived in the IAP.

To delete a quarantine repository:

1. Click **Query Manager** in the toolbar to access the saved results.
2. In the Query Name column, locate the row with the saved results.
3. Click **Delete** in the Quarantine column to remove access to the quarantine repository and release its contents from legal hold.

The quarantine repository is removed from the Where to Search list after you click Delete.



---

# 5 Using the AuditLog

This section describes how to work with the AuditLog, the IAP's compliance system log. Major topics include:

- [The AuditLog repository](#), page 39
- [Searching the AuditLog](#), page 40
- [Viewing AuditLogs](#), page 44

## The AuditLog repository

---

 **NOTE:**

AuditLog repositories are usually available only to legal or compliance officers. If the AuditLog repository does not appear in the Where to Search list on the search page, you do not have access.

---

When the AuditLog feature is enabled on the IAP, an AuditLog repository is created automatically for each domain. The AuditLog repository provides a method for creating a compliance system log. This is useful for demonstrating your company is adhering to compliance processes. The AuditLog records details about the following types of actions:

- Actions that users perform from the Web Interface, such as searching for an email, downloading a file, or viewing a message.
- Actions that are performed to manually delete email from the IAP. Logged actions include granting Administrative Delete privileges to a user from the IAP's Platform Control Center (PCC), and deleting emails from the IAP via the Administrative Delete feature in the Web Interface.
- Actions that are performed to merge duplicate copies of an email in IAP's Duplicate Manager.

For a complete list of the actions that are logged, see "[Logged actions and descriptions](#)" on page 41. For examples of AuditLogs, see "[Viewing AuditLogs](#)" on page 44.

# Searching the AuditLog

You can complete either a simple or advanced search when searching the AuditLog repository.

To perform a search:

1. Click **New Search** in the toolbar and select the simple or advanced search form.

---

 **TIP:**

If you need to search for multiple criteria, complete an advanced search.

---

2. From the Content Type list, select **document**.

---

 **NOTE:**

An email search on an AuditLog repository will produce no results. You must perform a document search.

---

3. From the Timeframe list, select the time period to search.

*Advanced searches only:* As an alternative to the predefined time periods, you can specify a time period to search by entering the Start and end (To) dates. For example, to search for documents dated between March 8, 2009 and March 23, 2009, enter **03/08/2009** in the Start field and **03/23/2009** in the To field.

4. From the Where to Search list, select **<domain>.userauditlog**.
5. Complete the Search for field by entering one or more of the search criteria in “[Completing the Search for field](#)” on page 41. Leave the field blank to see all logged actions performed during the time period.
6. *Advanced searches only:* To refine the search, use one or more of the additional query fields shown in “[AuditLog advanced search fields](#)” on page 42.
7. Click **Find Now** to start the search.

The Query Results page appears. An AuditLog is generated for each session in which a logged action occurs.



## Completing the Search for field

 **TIP:**

You can use Boolean operators AND, OR, and NOT to enter multiple criteria in this field. For example, `jdoe AND Search; Mail OR Export; jsmith* NOT jsmithson`. For more information on Boolean expressions, see [“Boolean query expressions”](#) on page 58.

You can also use wildcards (\* and ?). See [“Matching words”](#) on page 54.

To complete the Search for field when performing a search of the AuditLog repository:

- Enter a specific user name, in one of the following forms:
  - User ID: Enter the login name of the user, such as **jdoe**.
  - First Name: Enter the user's first name, such as **John**, as shown in the LDAP directory (address book).
  - Last Name: Enter the user's last name, such as **Doe**, as shown in the LDAP directory.

The search results will include the user's logged sessions in the Web Interface, plus any logged actions on email that was sent to or from the user.

The results might also include logged actions on email where the user's name was in the Subject header. This would depend on the form of the name you use in the search.

To search only for actions the user performed, use the Author field in the Advanced Search form.

- Enter the type of action performed, from the table below.

**Table 8** Logged actions and descriptions

Logged action	Description
<i>Actions performed by users in the Web Interface</i>	
Search	Simple or advanced searches performed by the user.
Query Result	Information about the query results returned or displayed.
Navigation	Navigation through the query results.
View Message	Email displayed from the query results.
View Document	Files displayed from the query results.
Download	Files downloaded by the user.
Mail	Email sent to the user.
Export	Email or files exported by the user.
Save Query	Query criteria saved by the user.
Save Query Result	Query results saved by the user.
Start Quarantine	Saved query results placed in a quarantine repository by the user.
Delete Quarantine	Quarantined query results deleted by the user.

Logged action	Description
<i>Actions performed to delete email</i>	
Login	IAP Admin login to the Platform Control Center (PCC), for the session in which Administrative Delete authority is granted.
Admin Delete Authority	Administrative Delete authority granted to a user (from the PCC).
Delete by Admin	Email deleted from the IAP by the Administrative Delete user. (Performed from the Web Interface.)
<i>Action performed to merge duplicate copies of a message</i>	
Duplicate BitFiles Merged	Messages with the same unique ID (hash) that were merged in an IAP Duplicate Manager job.

A search on “View Message,” for example, would return:


- A list of all user sessions in which email messages were viewed during the specified time period
- The login of the user who viewed a message
- The date and time of the action

Opening an AuditLog session would display all actions performed during that session, plus additional information about the email that was viewed.

## AuditLog advanced search fields


To refine an advanced search of the AuditLog repository, complete one or more of the following fields.


---


**Step #1 - Search Type:**Content Type:  


---


**Step #2 - Field Search Criteria:**


Search for:  

Document Name:  


Document Path:  

Extension:  



Title:  

Author:  


---

**Step #3 - Time Search Criteria:**By Timeframe:  

By Date:

Start:    
[mm/dd/yyyy]To:  

---

**Step #4 - Search Location:**Where to Search:  

---

**Figure 12 AuditLog advanced search****Table 9 Additional fields for AuditLog repository searches**

Query field	Matches
Document Name	Name of the component generating the AuditLog. There are two options available: <ul style="list-style-type: none"><li>• <b>Web Interface</b> For logged actions performed by users in the Web Interface</li><li>• <b>PCC Interface</b> For logged actions performed by the IAP administrator in the PCC to create an Administrative Delete user</li></ul>
Document Path	Host or IP address of the server where the AuditLog was generated (the HTTP server or the PCC server.) For example: <b>hp-s0-1-93.hp.com</b> .
Extension	File extension. Not used because the AuditLog is always an XML file.
Title	Not used because it is not applicable to AuditLog files.

Query field	Matches
Author	<p>Search for:</p> <ul style="list-style-type: none"> <li>• Actions performed by a specific user. Use one of the following forms to identify the user: <ul style="list-style-type: none"> <li>• User ID: Enter the login name of the user, such as <b>jdoe</b>.</li> <li>• First Name: Enter the user's first name, such as <b>John</b>, as shown in the LDAP directory (address book).</li> <li>• Last Name: Last Name: Enter the user's last name, such as <b>Doe</b>, as shown in the LDAP directory.</li> </ul> </li> <li>• Messages merged during a Duplicate Manager job: Enter <b>Merge Manager</b>.</li> </ul>

## Viewing AuditLogs

The following information about an AuditLog session is listed on the Query Results page:

- User: Login ID of the user performing the logged actions.
- Session Start: Start time of the user's session.
- Session End: End time of the user's session.
- File Size: Size of the session AuditLog file.
- Server: Server (HTTP portal or PCC server) on which the logged session was captured.
- Date: Date the AuditLog file was archived.

The screenshot shows a web interface with the following elements:

- Top left: "You are logged in as bianca@itchyexch.com"
- Top center: "Query results complete" with navigation arrows and "Number of hits: 5 [1 - 5]"
- Top right: "More Options" button
- Repository: itchyexch.com.userauditlog
- Table with columns: User, Session Start, Session End, File Size, Server, Date.

<input type="checkbox"/>	User:	Session Start:	Session End:	File Size:	Server:	Date:
<input type="checkbox"/>	tnance@itchyexch.com	Fri, 17 Jul 2009 23:08:49	Fri, 17 Jul 2009 23:10:55	2226	hp-s0-71-2.persistcorp.com (10.0.71.2)	07/17/2009 23:58:13
<input type="checkbox"/>	tnance@itchyexch.com	Fri, 17 Jul 2009 23:08:49	Fri, 17 Jul 2009 23:10:55	2226	hp-s0-71-2.persistcorp.com (10.0.71.2)	07/17/2009 23:48:11
<input type="checkbox"/>	tnance@itchyexch.com	Fri, 17 Jul 2009 23:08:49	Fri, 17 Jul 2009 23:10:55	2226	hp-s0-71-2.persistcorp.com (10.0.71.2)	07/17/2009 23:38:08
<input type="checkbox"/>	tnance@itchyexch.com	Fri, 17 Jul 2009 23:08:49	Fri, 17 Jul 2009 23:10:55	2226	hp-s0-71-2.persistcorp.com (10.0.71.2)	07/17/2009 23:28:03
<input type="checkbox"/>	tnance@itchyexch.com	Fri, 17 Jul 2009 23:08:49	Fri, 17 Jul 2009 23:10:55	2226	hp-s0-71-2.persistcorp.com (10.0.71.2)	07/17/2009 23:18:00

Figure 13 AuditLog query results

From the Query Results page, you can save the search criteria you entered to generate the AuditLog. See "Saving search criteria" on page 32.

To display the contents of a logged session in the viewing pane, click the item in the Query Results page. Double-click the item to display the AuditLog content in a new window. You can print the contents of an AuditLog by clicking **Print this page**.

Each logged session contains the following additional information:

- User Name: Name of the user performing the logged actions. For a Duplicate Manager job, the user is shown as MergeManager.
- Session ID: The ID of the logged session.
- Session Start Time: Start time of the session.
- Session End Time: End time of the session.
- User ID: Login ID of the user performing the logged actions. (Correlates with User Name.)
- Operation: Name of the logged action.
- TimeStamp: Date and time of the logged action.
- Details: Specific information about the logged action. This information varies by action.

## User actions in Web Interface


AuditLogSession preview for " IAP Web Interface "			Print this page 
User Name:		Tim Nance	
Session ID:		CADA2A43E717A75C9088CF62D7F7ED8B	
Session Start Time:		Fri, 17 Jul 2009 23:08:49	
Session End Time:		Fri, 17 Jul 2009 23:10:55	
User ID:		tnance@itchyexch.com	
Operation	Time Stamp	Details	
<b>Search</b>	Fri, 17 Jul 2009 23:08:57	Document Type:	EMAIL
		Repository Name:	tnance@itchyexch.com Repository
		Starting Date:	Fri, 17 Apr 2009
		Ending Date:	Fri, 17 Jul 2009
<b>View Message</b>	Fri, 17 Jul 2009 23:09:15	Viewed Message	Mail #10000 from E-Sender @bianca-sender
		To	tnance@itchyexch.com
		From	wendtest010000@itchyexch.com
<b>View Message</b>	Fri, 17 Jul 2009 23:09:40	Viewed Message	Mail #1 from E-Sender @bianca-sender1
		To	tnance@itchyexch.com
		From	wendtest000001@itchyexch.com
<b>View Message</b>	Fri, 17 Jul 2009 23:09:58	Viewed Message	Mail #1 from E-Sender @bianca
		To	tnance@itchyexch.com
		From	wendtest000001@itchyexch.com
<b>Save Query Result</b>	Fri, 17 Jul 2009 23:10:39	Document Type:	EMAIL
		Query Name:	results 7-17
		Repository Name:	tnance@itchyexch.com Repository
		Starting Date:	Fri, 17 Apr 2009
		Ending Date:	Fri, 17 Jul 2009

Figure 14 AuditLog for user actions in Web Interface

## Duplicate Manager jobs


AuditLogSession preview for " Duplicate Merge Manager "		Print this page 	
User Name:	MergeManager		
Session ID:	1246507201423		
Session Start Time:	Thu, 2 Jul 2009 01:00:04		
Session End Time:	Thu, 2 Jul 2009 01:00:13		
User ID:	MergeManager		
Operation	Time Stamp	Details	
<b>Duplicate BitFiles Merged</b>	Thu, 2 Jul 2009 01:00:11	Hash Used To Merge	bigfoot
		Primary Master BitFile	10.0.204.4:/archive/TP_0000000000
		Secondary Master BitFile	10.0.172.3:/archive/TP_0000000000
		Master BitFile Group	02000c298679fc121d6c34d970
		BitFile Merged	10.0.172.4:/archive/TP_0000000000
		BitFile Merged	10.0.204.2:/archive/TP_0000000000

Figure 15 AuditLog for Duplicate Manager job

## Administrative Delete actions

### IAP administrator actions in PCC


AuditLogSession preview for " IAP PCC Interface "		Print this page 	
User Name:	testFirst000001 testLast000001		
Session ID:	C44B8FD911D802B2A173390F239D9571		
Session Start Time:	Wed, 10 Jun 2009 15:26:39		
Session End Time:	Wed, 10 Jun 2009 16:22:59		
User ID:	test000001@itchyexch.com		
Operation	Time Stamp	Details	
<b>Login</b>	Wed, 10 Jun 2009 15:26:39	IAP Admin User	test000001
		Action Taken	Login
<b>Admin Delete Authority</b>	Wed, 10 Jun 2009 15:27:03	IAP Admin User	test000001
		Delete Admin User	test000002@itchyexch.com
		Action Taken	Delete Admin Authority granted.
		Date	06/10/2009 15:27:03

Figure 16 AuditLog for Administrative Delete actions in PCC

### Administrative Delete actions in Web Interface

Operation	Time Stamp	Details	
<b>Delete by Admin</b>	Thu, 2 Jul 2009 00:41:54	From	test000003@itchyexch.com
		Subject	test
		Date	07/01/2009 17:27:00
		Smart cell IP addresses	10.0.172.3 / 10.0.204.4

Figure 17 AuditLog entry for email deletion

---

# 6 Troubleshooting problems in the Web Interface

Use the topics in this section to troubleshoot problems that can occur when using the Web Interface.

- [Unable to display saved results](#), page 47
- [Corrupted documents](#), page 47
- [Wildcard errors in queries](#), page 47
- [Query and display issues in Japanese, Chinese, and Korean documents](#), page 48
- [Problems exporting email](#), page 49
- [Problems exporting files](#), page 49

## Unable to display saved results

Search results are saved for one week and then deleted. If you save the results of a query, but the IAP retention settings delete the files before the end of the week, the results still appear in the Web Interface. However, they cannot be reloaded from the saved search results. Clicking the saved results displays an error because the Web Interface cannot find the saved results on the IAP.

## Corrupted documents

Email and files that are corrupted cannot be displayed in the Web Interface.

## Wildcard errors in queries

You can use wildcards in the Search for field in simple and advanced searches and several advanced search fields. (See [“Completing advanced searches”](#) on page 22.)

However, if you use a wildcard incorrectly, an error message will appear when you submit your query.

### Leading wildcards

Do not use a wildcard for the first character in a field. For example, \*smith or ?smith are not valid queries.

### Fuzzy searches

Wildcards cannot be used in fuzzy word searches. (A fuzzy word is a word that is textually similar to a literal query word.) For example, defin\*~ or defin?~ are not valid queries.

For more information on fuzzy searches, see [“Fuzzy words”](#) on page 54.

## Wildcards in phrase searches

Wildcards are not allowed in phrase queries (queries with spaces between words). For example, accounts\* department or accounts? department are not valid queries.

## Query and display issues in Japanese, Chinese, and Korean documents

### Extra space appears in email Subject text

When the text in an email Subject field includes non-ASCII characters, the Web Interface may sometimes display a space in the text. The space is not present in the archived email.

The following example shows how the Subject text appears in an archived email, and how the same text appears when viewing the email in the Web Interface.

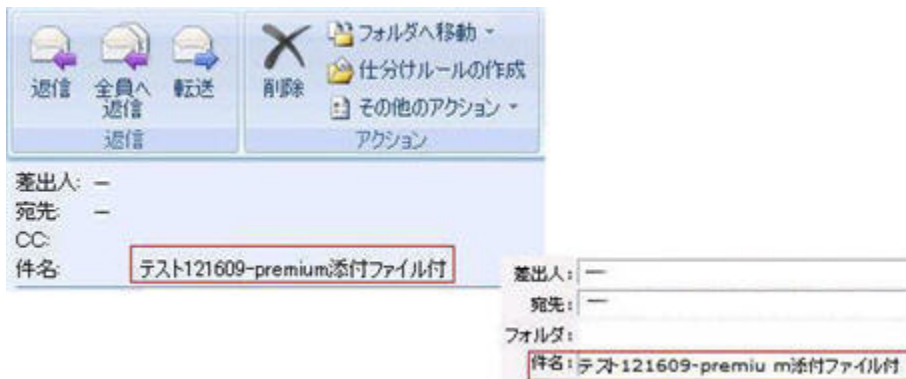


Figure 18 Extra space in email Subject field

### Query on words with furigana returns no results

In Japanese documents, furigana are phonetic readings that are printed above or beside kanji (Chinese characters). When the IAP archives email and files, it cannot index words with furigana. Because of this limitation, you cannot search on words with furigana in the Web Interface.

### Query on words in Hankaku-kana returns no results

The IAP cannot index Japanese words written in Hankaku-kana (half-width Katakana) in WordPerfect Office files (WordPerfect, Quattro Pro, and Presentations). For that reason, you cannot search on Hankaku-kana text in WordPerfect Office files and email attachments.

### Query on email folder name returns no results

Spaces in the name of an email folder are not indexed in Chinese, Japanese, and Korean. Searching for a folder name that contains spaces in these writing systems will fail, even if you surround the name in quotes.

### Japanese underscore character not displayed

If your computer runs Microsoft Windows XP Service Pack 2 or later, the underscore (" ") Japanese multi-byte character might not be visible in the Web Interface under the following conditions:



- Your computer's language option for non-Unicode languages (**Control Panel > Regional and Language options > Advanced**) is set to English and the language in the Web Interface is set to Japanese
- The non-Unicode language is set to Japanese and the Web Interface language is set to English
- The non-Unicode language is set to Japanese and the Web Interface language is set to Japanese

## Problems exporting email

For information on problems that may occur when you attempt to export email, see your *HP EAs for Microsoft Exchange User's Guide* or *HP EAs for IBM Domino User's Guide*, as appropriate. The user guides are included on the documentation CD in those products.

## Problems exporting files

For information on problems that may occur when you attempt to export files, click **Help** in the IAP File Export utility and then click **Troubleshooting** in the Help contents. The IAP File Export utility must be installed on your computer in order to export files from the IAP.



---

# 7 Query expression syntax and matching

This section describes the IAP Web Interface syntax used to search for archived documents (files or email messages), and explains how queries are matched against documents.

Major topics include:

- [Query expressions](#), page 51
- [Word characters](#), page 51
- [Letters and digits in different character sets](#), page 52
- [Matching words](#), page 54
- [Matching similar words](#), page 54
- [Matching word sequences](#), page 55
- [Boolean query expressions](#), page 58
- [Query expression examples](#), page 61

## Query expressions

Query expressions can be as simple or as complex as needed. The essential idea behind document retrieval is that query words are compared with document words to find a match. You can also:

- Look for document words that are textually similar, but not necessarily identical, to query words. See [“Matching similar words”](#) on page 54.
- Look for word sequences in a document: words that are near each other, and in a particular order. See [“Matching word sequences”](#) on page 55.
- Combine query words using logical (Boolean) operators (AND, OR, NOT). See [“Boolean query expressions”](#) on page 58.

Together, these query constructs provide considerable power to find what you need, provided you learn to use them well.

The way query expressions are interpreted is similar to the way documents are indexed when archived. Text is parsed (broken down) into words. Remaining characters are considered separators and ignored. Query expressions are fundamentally composed of words, no matter how complex the expression.

For indexing and searching, a word need not belong to a natural language, such as English. For example, `wt6_ht3` is a valid document word or query word. Query words can contain wildcards, such as in `f??t`.

## Word characters

When the system examines a query expression to determine its words, some characters are not included in query words, but are treated as word separators. When a document is archived, indexing determines which document words are available for searching in the same way.

Learning the rules of creating query words means also learning the rules of document indexing and, therefore, what words you can search for.

Topics include:

- [Word characters and separators](#), page 52
- [Regular expression definition of English word characters](#), page 52

## Word characters and separators

Word characters include all uppercase and lowercase letters, digits, and the following additional characters:

- `_` (underscore)
- `#` (number/pound/hash sign)
- `&` (ampersand)

All other characters are separators (except in queries, wildcards `?` and `*`, and special query characters `~`, `"`, `-`, and `!`).

However, `&&` by itself is not a word. It is a Boolean operator. When combined with at least one more word character, `&&` can be part of a word. For example, `a&&b` is a word.

Query analysis and document indexing are not case-sensitive. Uppercase and lowercase letters are treated the same.

## Regular expression definition of English word characters

The following regular expression provides, in succinct form, a complete specification of English word characters (except for treatment of `&&` as a non-word):

```
[ A-Za-z0-9_#& ]+
```

## Letters and digits in different character sets

Topics include:

- [Letters and digits defined](#), page 52
- [Letters and digits in files](#), page 53

## Letters and digits defined

All letters and digits are word characters. What IAP considers a letter or digit depends on the character set encoding used. For US ASCII encoding, letters are uppercase and lowercase English letters (A-Z, a-z). For ISO 8859-1 (Latin-1) encoding, used for Western European languages, accented letters are included. Most ideographic characters, such as those used in Asian languages, are also considered letters.

Whatever the language and encoding used for a particular document (file or email message), IAP maps encoded characters to the Unicode 2.0 standard. The Unicode 2.0 standard is then used to determine if a given character is a letter or a digit (or neither):

- A letter is any Unicode character in one of the following Unicode categories: Ll (lowercase letter), Lu (uppercase letter), Lt (title case letter), Lm (modifier letter), or Lo (other letter).
- A digit is any Unicode character whose Unicode name contains the word DIGIT, provided it is not in the range `\u2000` (en quad = en space) through `\u2FFF` (ideographic description - future).

## Letters and digits in files

Although all letters and digits are word characters, their treatment in files (including email message attachments) depends on the character encoding used. You can search for any words in email message bodies and headers, regardless of the encoding.

You can search for words in files (including email body, header, attachments, and indexed documents) provided the character encoding is shown in the following table.

The IAP does not perform automatic character set detection. If the character set is not provided in the email or file, the IAP uses a default of ISO-8859-1 (Latin 1).

**Table 10 Supported character sets**

<b>Supported character set</b>	<b>Description</b>
ISO-LATIN-1	Western European, extended ASCII
ISO-8859-1	Western European, extended ASCII
ISO-8859-15	Western European, extended ASCII
WINDOWS-1252	(Code pages supported by Windows) Latin 1
US-ASCII	7-bit American Standard Code for Information Interchange
UTF-7	Universal (all languages)
UTF-8	Universal (all languages)
ISO-8859-2	Eastern European
KOI8-R	Cyrillic (Russian and Bulgarian)
WINDOWS-1251	(Code pages supported by Windows) Cyrillic
WINDOWS-1253	(Code pages supported by Windows) Greek
WINDOWS-1254	(Code pages supported by Windows) Turkish
WINDOWS-1255	(Code pages supported by Windows) Hebrew
WINDOWS-1256	(Code pages supported by Windows) Arabic
WINDOWS-1257	(Code pages supported by Windows) Baltic: Estonian, Latvian, Lithuanian
ISO-8859-9	Turkish
WINDOWS-1258	(Code pages supported by Windows) Vietnamese
GB18030	Chinese (Mainland)
GB2312	Chinese (Mainland)
GBK	Chinese (Mainland)
EUC-TW	Chinese (Taiwan)
BIG5	Chinese (Taiwan)

Supported character set	Description
BIG5–HKSCS	Chinese (Hong Kong)
EUC-KR	Korean
ISO-2022–KR	Korean
Johab	Korean
KS_C_5601-1987	Korean
ISO-2022-JP	Japanese
EUC-JP	Japanese
SHIFT-JIS	Japanese

## Matching words

Matching words is not case-sensitive: *cat*, *Cat*, *cAt*, and *CAT* all match. Corresponding uppercase and lowercase letters, such as *A* and *a*, are treated the same in all respects.

There are two kinds of query words:

- Literal words that do not contain wildcards
- Words containing occurrences of one or both wildcard characters `*` and `?`

A literal word in a query expression matches the same word, character for character (case ignored), in an archived document. A word with wildcard characters (`*` or `?`) matches a document word in the same way, character by character, except for the following:

- A `?` matches any single character in a document word. For example, `b??t` matches *beat*, *beet*, *boat*, *blot*, *best*, *bust*, *bout*, and so on.
- An `*` matches any sequence of characters in a document word, including a sequence of no characters. For example, `f*t` matches the document words *foot*, *feet*, *fit*, *fault*, and *ft*; and `f*` matches any document word beginning with *f*.

You can use any number of wildcard characters (`*` or `?`) in a query word, but you cannot use a wildcard at the beginning of a query word. An error message results. For example, `*ion` is not a valid query.

## Matching similar words

Topics include:

- [Fuzzy words](#), page 54
- [Measuring word similarity](#), page 55

## Fuzzy words

You can search for document words that are textually similar to a given literal query word (that is, one containing no wildcards). To do this, append a tilde (`~`) character to the word, creating a fuzzy

word. For example, the fuzzy word `define~` matches the similar words *defined* and *definite*, but does not match *defining*, *definition*, *indefinite*, or *pine*. It also matches *define* itself.

---

 **NOTE:**

Do not use wildcards in fuzzy searches. For example, `foo*~` or `foo?~` is not a valid query

---

## Measuring word similarity

The edit distance (also called Levenshtein distance) between two words is the number of single-character operations (deletion, replacement, or insertion) required to change one word into the other word.

For example, the edit distance between *define* and *pine* is three: two deletions (*d* and *e*) and one replacement (*f* by *p*). The distance between *define* and *definite* is also three (*e* replaced by *i*; *te* inserted).

The search engine considers *define* more similar to *definite* than to *pine*, even though the edit distances are the same (three), because the edit distance (number of character changes) is compared to the word length (of the shorter of the query and document words). Two words are closer, for querying purposes, if it takes less to change one word into the other word relative to their lengths.

The similarity ratio used by the search engine is  $d/\min(\text{query}, \text{doc})$ , where  $d$  is the edit distance,  $\min$  is a function that returns the lesser of its arguments, and *query* and *doc* are the lengths of the query word and document word, respectively. A fuzzy word matches a document word if this ratio is no more than 0.5.

Examples:

Words compared	Similarity ratio	Match?
define, definite	$3/\min(6, 8) = 3/6 = 0.5$	yes
define, pine	$3/\min(6, 4) = 3/4 = 0.75$	no ( $0.75 > 0.5$ )

## Matching word sequences

You can use word sequences to find documents with words that occur in a specified order and are separated by a specified maximum distance.

Topics include:

- [Simple word sequences](#), page 55
- [Proximity word sequences](#), page 56

### Simple word sequences

To search for an ordered sequence of words, use a simple word sequence, which is a list of literal query words (no wildcards) separated by spaces (or other separators) and enclosed in quotes (""). A document matches a simple word sequence if all words occur in the document in the same order, with no intervening words.

For example, the sequence "like a rolling stone" does not match a document with the text *like a large rolling stone* because of the intervening word *large*.

## Proximity word sequences

You can use simple word sequences to search for words separated by separators but not by other words. To search for document words that are in an ordered sequence, but might be separated by other words, use a proximity word sequence.

To write a proximity word sequence, use the same syntax as a simple word sequence, but append a tilde (~) character to the second quote, and follow that with a numeric proximity value. The proximity value represents the maximum number of other document words that can occur between any two successive words of the sequence. A document matches a proximity word sequence if all words occur in the document in the same order, with at most N intervening words, where N is the proximity value.

For example, the sequence "bird garden stone"~3 matches any document that has these three words in this order, with *bird* and *garden* separated by no more than three words, and *garden* and *stone* separated by no more than three words. This sequence matches a document with the text *a bird in the rose garden is near a stone* because there are at most three words between successive sequence words. This sequence also matches *a bird garden with a stone* for the same reason.

Simple word sequences are a special case of proximity word sequences: ". . ." is the same as ". . ."~0. Any documents found by ". . ."~N are also found by ". . ."~M, when M > N.

## Matching word sequences in files and email attachments

IAP renders files and email attachments (like spreadsheets and PDF files) into text words. *When the IAP renders a document, it follows the document application's internal representation of the file.*

Certain file types, for example spreadsheets, look very different internally than they do externally. This means that word sequence in the external application representation which the end user sees may differ from the internal application representation. IAP query matching uses the internal application representation.

## Separators are ignored

IAP renders text into words. Remaining characters such as periods, commas, spaces, and newlines are considered separators and are ignored. Phrase queries ignore all formatting elements and non-word characters. The following original plain text of:

"This was news to Mr. Smith. Johnson, however, knew better."  
matches the phrase query of:

"Smith Johnson"

This is because internally, the two plain text sentences are represented as one long string of continuous words: "This was news to Mr Smith Johnson however knew better."

## Sequence is not intuitive

Internally in the file's original application, a large multi-page document or a single page spreadsheet equates to a long text sequence. *Text may not appear in the same sequence internally as it appears externally. Also, multiple instances of the same text in certain file types are represented as a single instance.*



## Spreadsheets

Look at the external representation of the following spreadsheet example.

**Table 11 Excel spreadsheet**

United States Presidents named John		
John	Adams	1797-1801
John Quincy	Adams	1825-1829
John Fitzgerald	Kennedy	1961-1963
John	Tyler	1841-1845

The specific order in which the text in the cells is stored internally depends on:

- The version of the product, for example Excel or Quattro Pro, used to generate the spreadsheet
- The insertion order for the spreadsheet text

For the spreadsheet above, assuming the cell text for names were entered in the displayed order of top left to bottom right (John Adams was entered first), and the title and dates were entered after the names were entered, most versions of spreadsheets store the text internally as follows:

John  
Adams  
John Quincy  
John Fitzgerald  
Kennedy  
Tyler  
United States Presidents named John  
1797-1801  
1825-1829  
1961-1963  
1841-1845

Note the following features of the internal representation:

- Text sharing: Where certain text appears in more than one cell in the spreadsheet, the text *may* appear only once in the internal representation. In this example, this is the case with the text "John" and "Adams". (Note that not all versions of Excel consistently share text in exactly this way.)
- This text sharing only occurs at the level of the entire text of a cell, and never occurs within cells. Thus, "John Quincy" and "John Fitzgerald" remain whole and independent.
- Even accounting for text sharing, the specific ordering of various cell text in the internal representation does not necessarily follow presentation order, and instead often follows insertion order.

Because of these factors, text sequence matches in an Excel spreadsheet (for example) are only consistent with the spreadsheet as viewed in Excel if the matched text appears wholly within a cell. However, it is possible for sequences to match in inconsistent ways across cell text depending on the precise version and editing history of that spreadsheet.

For the spreadsheet and order of insertion shown above, the following queries would match:

"John Adams"  
"Adams John"  
"Quincy John"  
"John Fitzgerald Kennedy"  
"Presidents named John"

And the following queries would **not** match:

"John Tyler"  
"Quincy Adams"  
"John Quincy Adams"  
"John Adams 1797–1801"

## PDF documents

PDF documents are another case where the internal text representation can vary widely from the visible presentation in PDF readers. The following issues can arise:

- Text sequences can appear out of order on the same page depending on how the page was composed.
- Text can appear doubled or can have spacing inserted into or removed from the internal representation to assist some specific visual presentation.

In general, PDF documents generated via print drivers are far more susceptible to these issues than PDF documents generated directly using Acrobat and other composing tools. However, because of the nature of PDF itself, even they are not immune.

## Boolean query expressions

You can combine words, fuzzy words, and word sequences using the Boolean (logical) operators AND, OR, and NOT. These operators must be uppercase.

- [Using the AND and OR operators](#), page 58
- [Using the NOT operator](#), page 59
- [Separators](#), page 60
- [Sub-expressions](#), page 60
- [Nested query expressions](#), page 60

## Using the AND and OR operators

AND is the default connector in query expressions. You do not need to supply AND explicitly. It is assumed, if neither AND nor OR is used explicitly. For example, the query `peace quiet` is equivalent to the query `peace AND quiet`.

The following table describes Boolean AND and OR operators, where `exp1` and `exp2` represent a word, word sequence, or other Boolean query expression.

**Table 12 Boolean AND and OR operators**

Syntax	Matches
<code>exp1 OR exp2</code>	all documents that match either <code>exp1</code> or <code>exp2</code>
<code>exp1 AND exp2</code> alternative syntax: <code>exp1 &amp;&amp; exp2</code> alternative syntax: <code>exp1 exp2</code>	all documents that match both <code>exp1</code> and <code>exp2</code>

## Using the NOT operator

The Boolean NOT operator excludes every term after NOT in a query.

For example, the query `beta* NOT beta2` would look for `beta1` or `beta05`, but not `beta2`.

For IAP searches, the NOT operator must always connect at least two terms or sub-expressions (such as `yes NOT no`), but the query cannot consist solely of negative criteria (such as `(NOT yes) OR (NOT no)`).

The following table shows examples of legal and illegal IAP queries using the NOT operator:

<code>NOT quiet</code>	illegal
<code>(NOT quiet) AND (NOT blue)</code>	illegal
<code>NOT NOT quiet</code>	illegal
<code>blue NOT quiet</code>	legal
<code>nois? NOT noisy</code>	legal

## Example of queries using NOT operator

In the example below, five email messages are archived that contain the following text:

Message 1      John V. Brown  
Message 2      Jim Brown  
Message 3      Benefits for Brown, Jim  
Message 4      Jim Smith  
Message 5      Benefits for Brown, Sally

The following queries in the Search for field use the NOT operator correctly and successfully locate the relevant messages:

**Table 13 Boolean NOT operator properly used in query**

Query	Result
<code>brown NOT jim</code>	Message 1 Message 5
<code>(John OR Jim) NOT smith</code>	Message 1 Message 2 Message 3
<code>(benefits AND Brown) NOT Jim</code>	Message 5
<code>brown NOT (jim OR sally)</code>	Message 1

## Separators

Boolean operators must be surrounded by one or more separators, typically white space. For example, the query `peas&&carrots` is not equivalent to the query `peas && carrots`; `peas&&carrots` is a single word.

Negation operators (`-` and `!`) are exceptions to this rule. They must be preceded by a separator, but they need not be followed by a separator. For example, `carrot-a6` is a single query word, but `carrot -a6`, like `carrot (- a6)`, is equivalent to the Boolean expression `carrot NOT a6`.

## Sub-expressions

You must connect at least two sub-expressions with the operators NOT (`-` or `!`), AND (`&&`), and OR.

The following table shows examples of legal and illegal queries using sub-expressions:

<code>omega NOT alpha</code>	legal
<code>alpha AND omega</code>	legal
<code>alpha OR omega</code>	legal
<code>NOT alpha</code>	illegal
<code>NOT NOT alpha</code>	illegal
<code>OR alpha</code>	illegal
<code>OR OR alpha</code>	illegal
<code>AND omega</code>	illegal
<code>AND AND omega</code>	illegal

## Nested query expressions

You can nest Boolean query expressions using Boolean expressions as arguments of Boolean expressions. For example, the following query searches for documents containing `alpha` and `beta` but not `omega`:

```
(alpha AND beta) NOT omega
```

## Query expression examples

The following are examples of query expressions.

**Table 14 Query expression examples**

Query expression	Finds documents with ...
peace OR quiet	Either <i>peace</i> or <i>quiet</i> , or both, in either order.
peace quiet peace AND quiet peace && quiet	Both <i>peace</i> and <i>quiet</i> , in either order.
peace&&quiet	The single word <i>peace&amp;&amp;quiet</i> .
peace or quiet	The three words <i>peace</i> , <i>or</i> , and <i>quiet</i> , in any order. "or" is a word. The OR operator must be uppercase.
not quiet	The words <i>not</i> and <i>quiet</i> . The NOT operator must be uppercase.
NOT quiet	Illegal. A query cannot be purely negative and must have some positive expression.
peace & quiet	The three words <i>peace</i> , <i>&amp;</i> , and <i>quiet</i> , in any order. <i>&amp;</i> is a word character. The AND operator is implied.
peace   quiet	Both <i>peace</i> and <i>quiet</i> . <i> </i> is a separator. The AND operator is implied.
peace NOT quiet peace -quiet -quiet peace	The word <i>peace</i> but not <i>quiet</i> .
peace AND quiet OR silence	Avoid using. Parentheses are needed: <code>peace AND (quiet OR silence)</code> .
quiet OR silence AND peace	Avoid using. Parentheses are needed: <code>quiet OR (silence AND peace)</code> .
pea*	Any word starting with <i>pea</i> such as: <i>pea</i> , <i>peas</i> , <i>peace</i> , or <i>peach</i> .
pea*c*	Words such as: <i>peace</i> , <i>peach</i> , or <i>peaches</i> .
"peace quiet"	Both <i>peace</i> and <i>quiet</i> , in that order, with no intervening words. Examples: <i>peace quiet</i> or <i>peace \$^%+{} quiet</i> .
"peace quiet"~1	Both <i>peace</i> and <i>quiet</i> , in that order, separated by at most one word. Examples: <i>peace and quiet</i> ; <i>peace, now;%^\$ quiet</i> ; <i>peace quiet</i> ; or <i>peace george quiet</i> .
peace~	Words similar to <i>peace</i> such as: <i>peaches</i> , <i>piece</i> , <i>place</i> , or <i>plate</i> .
(peace quiet) NOT "peace quiet"	Both <i>peace</i> and <i>quiet</i> , but not <i>peace</i> followed immediately by <i>quiet</i> . Examples: <i>quiet peace</i> ; <i>quiet blue peace</i> ; or <i>peace, water, land, quiet</i> .



# A Indexed file types and MIME types

The following file types and MIME content-types are indexed by IAP 2.1. You can search the contents of archived files or email attachments if their file type is listed in this table.

**Table 15 IAP indexed file types and MIME types**

File extension	File type	MIME content-type
.xml	XML document	text/xml
.txt	Plain text file; treated as ISO-8859-1 unless otherwise specified	text/plain
.htm, .html, .stm	HTML document	text/html, rtf/html
.rtf	Rich text format	rtf/text, application/rtf
.dat	TNEF (for Microsoft Exchange)	ms/tnef
.mht, .mhtml, .nws, .eml	Email message	message/RFC 822
.doc, .dot	Microsoft Word 97-2003 document	application/msword
.xla, .xlc, .xlm, .xls, .xlt, .xlw	Microsoft Excel 97-2003 document	application/vnd.ms-excel, application/ms-excel
.pot, .pps, .ppt	Microsoft PowerPoint 97-2003 document	application/vnd.ms-powerpoint, application/vnd.mspppt
.pdf	Adobe Portable Document format	application/pdf
.zip	ZIP archive	application/zip
.docx	Microsoft Word 2007 document	application/vnd.openxmlformats-officedocument.wordprocessingml.document
.docm	Microsoft Word 2007 macro-enabled document	application/vnd.ms-word.document.macroEnabled.12
.dotx	Microsoft Word 2007 template	application/vnd.openxmlformats-officedocument.wordprocessingml.template
.dotm	Microsoft Word 2007 macro-enabled document template	application/vnd.ms-word.template.macroEnabled.12
.xlsx	Microsoft Excel 2007 workbook	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet

<b>File extension</b>	<b>File type</b>	<b>MIME content-type</b>
.xlsm	Microsoft Excel 2007 macro-enabled workbook	application/vnd.ms-excel.sheet.macroEnabled.12
.xltx	Microsoft Excel 2007 template	application/vnd.openxmlformats-officedocument.spreadsheetml.template
.xltm	Microsoft Excel 2007 macro-enabled workbook template	application/vnd.ms-excel.template.macroEnabled.12
.xlam	Microsoft Excel 2007 add-in	application/vnd.ms-excel.addin.macroEnabled.12
.pptx	Microsoft PowerPoint 2007 presentation	application/vnd.openxmlformats-officedocument.presentationml.presentation
.pptm	Microsoft PowerPoint 2007 macro-enabled presentation	application/vnd.ms-powerpoint.presentation.macroEnabled.12
.ppsx	Microsoft PowerPoint 2007 slide show	application/vnd.openxmlformats-officedocument.presentationml.slideshow
.ppsm	Microsoft PowerPoint 2007 macro-enabled slide show	application/vnd.ms-powerpoint.slideshow.macroEnabled.12
.potx	Microsoft PowerPoint 2007 template	application/vnd.openxmlformats-officedocument.presentationml.template
.potm	Microsoft PowerPoint 2007 macro-enabled presentation template	application/vnd.ms-powerpoint.template.macroEnabled.12
.wpd	Corel WordPerfect for Windows – versions through Version 12.0, X3	application/wordperfect, application/wpd
.qpw, .wb1, .wb2, .wb3	Corel Quattro Pro for Windows – versions through Version 12.0, X3	application/qpw, application/wb1, application/wb2, application/wb3
.shw	Corel Presentations – versions through Version 12.0, X3	application/presentations



---

# B Additional indexing detail for Microsoft Office

The tables below list the Microsoft Office features and document properties that are indexed (and not indexed) by IAP 2.1. Indexed features and properties can be used in Web Interface searches.

The tables cover Microsoft Office versions 97–2003 and 2007.

## Microsoft Office supported features

**Table 16 Microsoft Office supported features**

<b>Feature</b>	<b>Microsoft Word</b>	<b>Microsoft PowerPoint</b>	<b>Microsoft Excel</b>
Contents	Yes	Yes	Yes
Table	Yes	Yes	Yes
Textbox	Yes	Yes	Yes
Header/Footer	Yes	Yes	Yes
Comment	Yes	Yes	Yes
FootNote/EndNote	No	No	No
Signature	Yes	Yes	No
Chart	Yes	Yes	Yes
Object (Microsoft Office, WordPad ...)	Yes	No	Yes
Embedded Objects (OLE)	Yes	Yes	Yes
Notes	N/A	Yes	N/A
WordArt	Yes	Yes	Yes
SmartArt	No	No	No
Sheet's name	N/A	N/A	Excel 2007: No Excel 97-2003: Yes

# Microsoft Office supported file properties

Table 17 Microsoft Office supported properties

Type	Property	Microsoft Word	Microsoft PowerPoint	Microsoft Excel
Document Properties	Author	Yes	Yes	Yes
	Title	Yes	Yes	Yes
	Subject	Yes	Yes	Yes
	Keywords	Yes	Yes	Yes
	Category	Yes	Yes	Yes
	Status	Yes	Yes	Yes
	Comments	Yes	Yes	Yes
	Location	No	No	No
Advanced Properties: General	Type	No	No	No
	Location	No	No	No
	Size	No	No	No
	MS-DOS name	No	No	No
	Created	No	No	No
	Modified	No	No	No
	Accessed	No	No	No
	Attributes	No	No	No
Advanced Properties: Statistics	Created	Yes	Yes	Yes
	Modified	Yes	Yes	Yes
	Accessed	Yes	Yes	Yes
	Printed	Yes	Yes	Yes
	Last saved by	Yes	Yes	Yes
	Revision number	Yes	Yes	Yes
	Statistics	Yes	Yes	Yes

Type	Property	Microsoft Word	Microsoft PowerPoint	Microsoft Excel
Advanced Properties: Custom	Checked by	Yes	Yes	Yes
	Client	Yes	Yes	Yes
	Date completed	Yes	Yes	Yes
	Department	Yes	Yes	Yes
	Destination	Yes	Yes	Yes
	Disposition	Yes	Yes	Yes
	Division	Yes	Yes	Yes
	Document number	Yes	Yes	Yes
	Editor	Yes	Yes	Yes
	Forward to	Yes	Yes	Yes
	Group	Yes	Yes	Yes
	Language	Yes	Yes	Yes
	Mailstop	Yes	Yes	Yes
	Office	Yes	Yes	Yes
	Owner	Yes	Yes	Yes
	Project	Yes	Yes	Yes
	Publisher	Yes	Yes	Yes
	Purpose	Yes	Yes	Yes
	Received from	Yes	Yes	Yes
	Recorded by	Yes	Yes	Yes
	Recorded date	Yes	Yes	Yes
	Reference	Yes	Yes	Yes
	Source	Yes	Yes	Yes
	Status	Yes	Yes	Yes
Telephone number	Yes	Yes	Yes	
Typist	Yes	Yes	Yes	

<b>Type</b>	<b>Property</b>	<b>Microsoft Word</b>	<b>Microsoft PowerPoint</b>	<b>Microsoft Excel</b>
Advanced Properties: Contents	Document Contents	Yes	Yes	Yes
Advanced Properties: Summary	Title	Yes	Yes	Yes
	Subject	Yes	Yes	Yes
	Author	Yes	Yes	Yes
	Manager	Yes	No	Yes
	Company	Yes	No	Yes
	Category	Yes	Yes	Yes
	Keywords	Yes	Yes	Yes
	Comments	Yes	Yes	Yes
	Hyperlink base	Word 2007: Yes Word 97–2003: No	Yes	Excel 2007: Yes Excel 97–2003: No
	Template	Word 2007: Yes Word 97–2003: No	Yes	No

---

# Index

## Symbols

&&  
in Boolean query expressions, 58

, 20

## A

access control list (ACL)  
definition, 14

AND queries, 58

archiving  
definition, 14

audit queries, 33

AuditLog repository  
accessing, 39

## B

Boolean queries  
AND and OR operators, 58  
characters, 51  
expressions, 58  
nested, 60  
NOT operator, 59  
separators, 60  
sub-expressions, 60

## C

cannot display saved results, 47

case sensitivity  
Boolean queries, 58  
matching words, 54  
word characters, 51

changing  
IAP Web Interface language, 20  
IAP Web Interface password, 20

Chinese query issues, 48

content-type indexing, 15, 63

contents  
indexing, definition, 14

control list  
definition, 14

conventions  
document, 9  
text symbols, 10

copying  
results to a quarantine repository, 36

Corel Presentations, 15, 63

creating  
quarantine repository, 36

## D

default Boolean connective (AND), 58

definition  
access control list (ACL), 14  
archiving, 14  
document, 14  
IAP, 13  
indexing documents, 14  
repository, 14  
routing rule, 14

deleting  
quarantine repositories, 37  
saved criteria, 34

digits, 52

displaying  
results, 29  
saved criteria, 34  
saved results, 34

document  
conventions, 9  
definition, 14  
related documentation, 9

document archiving  
explanation, 14

dropped caps, 16

DWA Extension, 13

## E

Email Archiving software (EAs)  
applications for users, 13

embedded messages, 15

Excel, 15, 65, 66

export problems  
email, 49  
files, 49

- exporting
  - email, 35
  - files, 35
- expressions, query
  - about, 51
  - Boolean, 58
  - examples, 61
  - languages, 52, 53
  - letters and digits, 52
  - matching words, 54
  - separators, 51
  - sequences, 55
  - sequences, matching, 55
  - word characters, 51

## F

- File Export, 13
- file type indexing support, 15, 63
- furigana, 48
- fuzzy words, 54

## H

- half-width Katakana, 48
- Hankaku-kana, 48
- help, obtaining, 10
- HP
  - Subscriber's choice Web site, 11
  - technical support, 10
- HTML formatting, 15

## I

- IAP
  - definition, 13
- IAP Web Interface, 14
  - advanced searches, 22
  - basics, 17
  - changing language, 20
  - changing password, 20
  - display problems, 47
  - query problems, 47
  - Query Results page, 29
  - requirements, 17
  - simple searches, 21
  - supported Web browsers, 17
  - toolbar, 18
  - troubleshooting, 47
- implicit Boolean connective (AND), 58

- indexing documents
  - definition, 14
  - indexed file types, 15, 63
  - indexed MIME content-types, 63
  - what IAP does not index, 15
  - what IAP indexes, 15
- Integrated Archive Platform
  - See IAP

## J

- Japanese query issues, 48

## K

- Korean query issues, 48

## L

- language, changing
  - IAP Web Interface, 20
- languages, query expressions, 52, 53
- letters, 52
- Levenshtein distance, 54
- logging in and out
  - IAP Web Interface, 17
- Lotus Notes options, 13

## M

- Message ID
  - displaying, 26
- Microsoft Office
  - indexing limitations, 16
- Microsoft Office applications, 15, 65, 66
- Microsoft Word, 15, 65, 66
- MIME content-type indexing, 15, 63

## N

- New Search button
  - IAP Web Interface, 18
- non-indexed file types, 16
- NOT queries, 58

## O

- OR queries, 58
- Outlook Plug-In
  - description, 13
- OWA Extension, 13

## P

- PDF files, 15, 63

PowerPoint, [15](#), [65](#), [66](#)  
Preferences button  
IAP Web Interface, [18](#)

## Q

quarantine repositories, [32](#), [36](#)  
  creating, [36](#)  
  deleting, [37](#)  
Quattro Pro, [15](#), [63](#)  
query criteria  
  deleting, [34](#)  
  displaying saved criteria, [34](#)  
  saving, [32](#)  
query expressions  
  about, [51](#)  
  Boolean, [58](#), [60](#)  
  examples, [61](#)  
  languages, [52](#), [53](#)  
  letters and digits, [52](#)  
  separators, [51](#)  
  sequences, matching, [55](#)  
  word characters, [51](#)  
query issues  
  Japanese, Chinese, Korean documents, [48](#)  
Query Manager button  
  IAP Web Interface, [18](#)  
query problems  
  wildcards, [47](#)  
query results  
  displaying, [29](#)  
  displaying saved results, [34](#)  
  saving, [32](#)  
  sending, [35](#)  
Query Results page, [29](#)

## R

related documentation, [9](#)  
repository  
  definition, [14](#)  
requirements  
  IAP Web Interface, [17](#)  
results  
  displaying, [29](#)  
  displaying saved results, [34](#)  
retrieving large number of documents, [33](#)  
RFC 822 messages, [15](#), [63](#)  
routing rule  
  definition, [14](#)

## S

saving  
  query or search criteria, [32](#)  
  query results, [32](#)  
search criteria  
  deleting, [34](#)  
  displaying saved criteria, [34](#)  
  saving, [32](#)  
Search for field  
  IAP Web Interface, [18](#)  
search results  
  displaying, [29](#)  
  displaying saved results, [34](#)  
  exporting, [35](#)  
  quarantine repository, [36](#)  
  saving, [32](#)  
  sending, [35](#)  
searching  
  advanced searches, [22](#)  
  AuditLog, [40](#)  
  query syntax, [51](#)  
  search basics, [18](#)  
  simple searches, [21](#)  
sending  
  query or search results, [35](#)  
separators  
  characters, [51](#)  
  matching word sequences, [55](#)  
similarity, matching words, [54](#)  
simple word sequences, [55](#)  
Subscriber's choice, [11](#)  
symbols in text, [10](#)

## T

technical support  
  HP, [10](#)  
text symbols, [10](#)  
TNEF, [15](#), [63](#)  
toolbar  
  IAP Web Interface, [18](#)  
troubleshooting  
  IAP Web Interface, [47](#)

## U

Unicode standards, [52](#)

## W

Web Interface  
  See IAP Web Interface  
Web sites  
  HP Software support, [10](#)  
  HP Subscriber's choice, [11](#)

- wildcard characters, [54](#)
- Word, [15](#), [65](#), [66](#)
- word sequences
  - matching, [56](#)
  - proximity, [56](#)
- WordPerfect, [15](#), [48](#), [63](#)
- WordPerfect Office applications, [15](#), [63](#)
- words, query
  - characters and separators, [51](#)
  - fuzzy, [54](#)
  - letters and digits, [52](#)
  - literal, [54](#)
  - matching, [54](#), [55](#)
  - sequences, [55](#)

## Z

- ZIP files, [15](#), [63](#)