IBM eDiscovery Analyzer

IBM

User Guide

Version 2.1

IBM eDiscovery Analyzer

User Guide

Version 2.1

Contents

ibm.com and related resources

Product support and documentation are available from ibm.com.

Support and assistance

Product support is available on the Web. Click Support from the product Web site at:

IBM eDiscovery Analyzer

   http://www.ibm.com/software/data/content-management/support/ediscovery-analyzer/

IBM eDiscovery Manager

   http://www.ibm.com/software/data/content-management/support/ediscovery-manager/

Information center

You can view the product documentation in an Eclipse-based information center that you can install when you install the product. By default, the information center runs in a Web server mode that other Web browsers can access. You can also run it locally on your workstation. See the information center at http://publib.boulder.ibm.com/infocenter/email/v1r0m0/topic/com.ibm.email.doc/welcome.htm.

PDF publications

You can view the PDF files online using the Adobe Acrobat Reader for your operating system. If you do not have the Acrobat Reader installed, you can download it from the Adobe Web site at http://www.adobe.com.

See the following PDF publications Web sites:

Product	Web site
IBM eDiscovery Analyzer	http://www.ibm.com/support/docview.wss?rs=3533&context=SSJKLP&uid=swg27013410
IBM eDiscovery Manager	http://www-01.ibm.com/support/docview.wss?rs=3506&context=SS8JHU&uid=swg27012725

How to send your comments

Your feedback is important in helping to provide the most accurate and highest quality information.

Send your comments by using the online reader comment form at https://www14.software.ibm.com/webapp/iwm/web/signup.do?lang=en_US&source=swg-rcf.

Contacting IBM

To contact IBM customer service in the United States or Canada, call
1-800-IBM-SERV (1-800-426-7378).

To learn about available service options, call one of the following numbers:
- In the United States: 1-888-426-4343
- In Canada: 1-800-465-9600

For more information about how to contact IBM, see the Contact IBM Web site at
http://www.ibm.com/contact/us/.

Chapter 1. Getting started with IBM eDiscovery Analyzer

Start here to learn how IBM® eDiscovery Analyzer enables you to search, analyze, and review archived e-mail documents.

IBM eDiscovery Analyzer helps you to collect e-mail and attached documents that might be relevant to a discovery request, so that senior counsel can understand which documents might require further review. eDiscovery Analyzer can help you to reduce the volume of documents that are typically sent to independent legal review services, which thereby reduces the costs of such review.

The eDiscovery Analyzer search engine searches the different parts of e-mail: the sender and recipient fields (to, from, cc, bcc), subject, body, and attachments. If you specify that the search engine search in one part of an e-mail, such as the subject line, the search engine does not search in any other parts of the e-mail.

You search a case to identify documents that are relevant to a discovery request, then further narrow the search before legal review. Your goals might be to:

- Better understand the content of the e-mail documents in the case
- Prioritize review of the e-mail documents
- Identify e-mail documents that do not require further review
- Identify and flag e-mail documents that might be responsive to a discovery order

The set of documents collected for further review is typically determined by flagging documents as **Responsive** or **Second-Level Review**, for example. After you flag relevant documents, an administrator can package the documents for further review by using eDiscovery Manager.

Search strategies

You can use one or more search strategies to identify the e-mail documents that might be relevant to a discovery request, or that require further review.

Modify search queries

The complete query is displayed in an editable search query box. You can save the search for reuse, move back or forward through queries to see the results from different queries, or edit the query and run the new query.

If you are familiar with the search query syntax, you can enter a query directly in the text box.

Reduce noise

Any set of search results is likely to have some irrelevant documents, such as spam, automatically generated notices, company announcements, and so on. You can reduce the volume of documents to review by searching for phrases or senders that frequently appear in the documents that you want to exclude. For example, if you know that automatically generated notices come from a particular e-mail

address, search for that address and flag the results as Non-Responsive. You can then exclude documents that are flagged as Non-Responsive from subsequent searches.

Frequently occurring phrases and sender domains can suggest which e-mails are noise. Use the categories, which list terms from the results by frequency, to identify these phrases.

Review query sets to reduce search terms

In addition to reducing noise, your senior counsel can negotiate with opposing counsel to establish queries that result in the smallest acceptable number of results. By importing queries that the parties to a case agree to, you can get a list of search results to start working with. Each search can be reviewed to identify the number of results and the scope of content returned. Use the categories display to reduce each specific query to a minimal set of results that satisfy a discovery request. These new queries can be saved and exported for further negotiation with outside parties.

Find relevant phrases

If you are looking for phrases, e-mail addresses, or names, for example, that most frequently occur in a case, use the categories display to include or exclude those terms in search queries. You can learn more about the content of the case by searching different sets of phrases and then narrowing the search results to relevant documents.

Limit the dates searched

You can limit the range of dates being searched by using the timeline. Examining the dates on which more or fewer e-mails were sent can help you gain insight into the content of the case. Iterate over different date ranges to gain more insight.

Reduce irrelevant words and phrases

You can limit the appearance of irrelevant and repetitive phrases in search results by specifying text that should be ignored (ignore text) in searches. For example, add common legal disclaimers to your ignore text. The search engine does not use common stop words, so you must specify any such words or phrases for each case.

View e-mail threads

From your search results, you can view e-mail threads (conversations) to get a better idea of the contents and conversations that appear in the e-mail documents that are being reviewed. Use the **View Thread** and **Sort by Thread** options in the results list.

Related concepts

"Interpreting the search results" on page 13

Related tasks
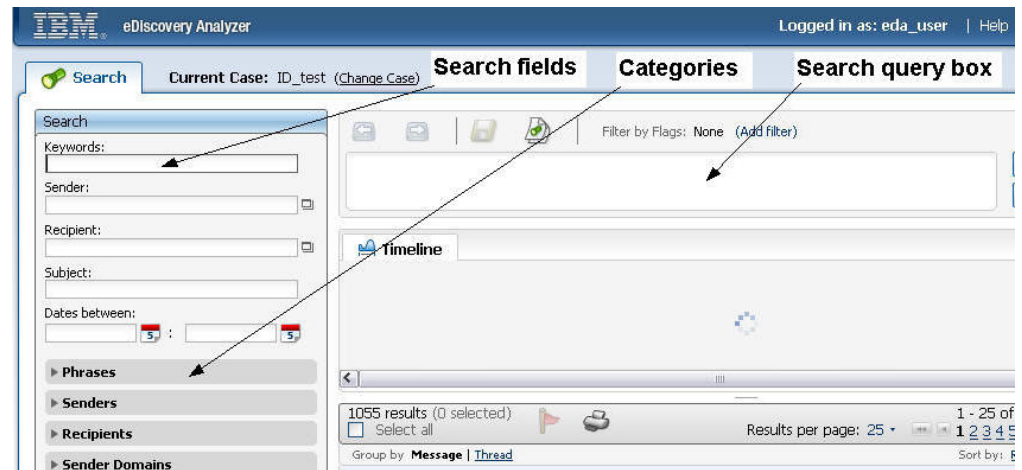
Chapter 2, "Searching a case," on page 3

Related reference

"Search tips" on page 4

Chapter 2. Searching a case

Search a case to identify relevant content, reduce the size of the case, and identify the documents that require further review.

The following image illustrates the location of the search fields, categories, and search query box in the search interface.

When you first open a case, the default search query runs automatically. Unless you define a different default query, the query returns all e-mails in the case.

To search a case:

1. Log in to eDiscovery Analyzer and select a case.
2. Enter any combination of keywords, senders, recipients, content, or dates in the search fields. You must enter at least one search term. The default query that returns all documents in a case is `*:*`, meaning all fields and all terms.
3. Click **Search**. Your query is displayed in the search query box, and results are displayed in the Results and Timeline sections of the search page.
4. Optional: To narrow your search, enter additional search terms in the search fields, select **Add to search**, and click **Search**. The new search terms are appended to the search query text box.
5. Optional: Try other search terms by editing the query in the search query box, selecting from terms in the categories, changing the date range by using the timeline or **Dates between** fields, or by adding or removing terms in the Search section.
6. Optional: Select one or more flags from the **Filter by Flags** menu to include or exclude documents with the selected flags for every subsequent query. Filter by flags settings are specific to the user.

   Related concepts

   "Search strategies" on page 1

   Related tasks

   Chapter 3, "Identifying document relevance with flags," on page 17

Search tips

You can refine your search by using search query operators.

The following rules apply to all searches:

- Punctuation is generally treated as a word separator. The query `case-sensitive` returns the same documents that the exact-match query "`case sensitive`" returns. However, some hyphenated terms are treated as single words, for example, `self-inflicted`.
- Search terms are not case sensitive. The query `IBM` matches documents that contain `ibm`.
- Search terms match other inflections of the same word (base form match). The query `stock option` matches `stock options`. The query `mouse` matches `mice`. Or, for example, if you enter `king`, the search engine returns all documents that contain the word `king` or `kings`. If you enter the query `king lear`, the search engine returns documents that contain the terms `king` or `kings`, and `lear`.

To do simple keyword search, you can enter one or more query terms (keywords) in the **Keywords** field. The search engine returns documents that contain all of those keywords or variations of the keywords in the subject, body, or attachments. The default Boolean operator is AND.

To see more precise results, use more specific keywords. For example, use `tax avoidance strategies` rather than `taxes`. Or use `California energy usage` rather than `energy`.

If a simple keyword search returns too many documents that are not what you are looking for, you can use operators or other search term fields to refine your search.

Excluding terms

Use the minus sign (-) or the Boolean NOT operator to exclude terms. For example, if you want to find documents with the term `lear` and you do not want to see documents with `edward`, enter the query `lear -edward` or `lear NOT edward`.

The minus sign (-) or NOT also applies to a term and its variants. For example, the query `-edward` will exclude documents that contain the word `edward`'s.

Matching terms exactly

If you want to ensure that terms appear in results exactly in the sequence in which you typed them, you can use double quotation marks ("). For example, if you want to see documents with the phrase `mouse trap` and you do not want matches on related phrases such as `trap a mouse` or `mice escaped the trap`, then enclose the query inside double quotation marks. Term variants, such as `mouse` and `mice`, are not considered matches for exact match search. The query "`mouse trap`" does not match `mouse traps` or `mice trap`. And no other terms can appear between any two terms in the phrase, so `mouse in trap` would not match. Note that the search is still case-insensitive, so it will match `Mouse Trap`. But punctuation is ignored, so `mouse, trap` would match, as would `See the mouse? Trap it.`

The following rules apply to exact match searches:

- The terms must appear in the exact sequence as they appear in the query. For example, the query "`life sentence`" does not match `sentence to life`.

- The terms must appear in the same form; term variants will not match. The query "stock option" does not match stock options.
- For fielded queries, multiple terms separated by white space must be grouped by enclosing the terms in parentheses.

Table 1. Example of exact match search

Find this	Sample query in the search query box
The phrase mouse trap without related phrases.	"mouse trap"

Searching within a specific proximity

To restrict results to terms that appear within the same sentence, paragraph, or within a range of words, specify the terms and include WITHIN *context* in the query. To require that the specified terms appear in the specified order, add INORDER. For example, the query WITHIN filed patents in December INORDER finds text that matches filed patents in December, but it does not match patents filed in February or patents filed in December.

- WITHIN SENTENCE matches documents that contain the specified terms in the same sentence.
- WITHIN PARAGRAPH matches documents that contain the specified terms in the same paragraph.
- WITHIN *n* matches documents that contain the specified terms within the specified number of words of each other. The value of *n* must be a positive number, and must be at least as large as the number of terms specified.

The terms that you specify in a proximity search must appear in the same part of the document for any proximity search query to match the document. For example, a proximity query will not return documents in which the specified terms appear in the subject and body fields, or within the body and attachment parts.

Table 2. Examples of proximity search

Find this	Sample query in the search query box
The three phrases article but, cancer dancer, and elegant funeral, in any order, within the same paragraph.	("article but" "cancer dancer" "elegant funeral") WITHIN PARAGRAPH
The three words and, begin, and court in any order, including variations of the words, within a range of five words	(and begin court) WITHIN 5
The three words and, begin, and court in that order, without variations of the words, within a range of five words.	("and" "begin" "court") WITHIN 5 INORDER

Searching with Boolean operators

The Boolean operator OR specifies that at least one of the terms in a query must appear in the returned document. For example, the query (othello OR otello) returns documents that contain the term othello or otello, or both. Boolean operators are not case-sensitive.

You can also use the Boolean operators AND, OR, or NOT in combinations by using parentheses. For example, the query cougar OR (jaguar AND NOT car)

returns any document with `cougar` and documents that do not have `cougar` but that have the word `jaguar` but not the word `car`.

Use parentheses for grouping. To search for a word that is also a Boolean operator, the word must be enclosed in double quotation marks, or preceded by an escape character (\).

For example:
- The query `subject:(hedge OR fund)` matches documents that contain either `hedge` or `fund`, or other forms of `hedge` or `fund`, in the subject.
- To search for reserved words (words that have meaning as an operator in search) such as AND, OR, or NOT, these words must be escaped if they are not enclosed in double quotation marks. For example, you can create a query such as `car AND \not`, `car AND "not"`, or `"car not"`, but the query `car AND NOT` will not work.

Table 3. Examples of search using Boolean operators

Find this	Sample query in the search query box
The words `hedge` or `fund`, or `hedging` or `funds`, in the subject.	`subject:(hedge OR fund)`
The words `car` and `not`	`car AND \not`, or `car AND "not"`, or `"car not"`
The alternate spellings of `othello` or `otello`, or both.	`(othello OR otello)`

Searching for unknown or variable characters or terms

The wildcard characters help you find documents when you do not know the full spelling, or if you want to find variations of the term. For example, the query `czech*` returns documents with the terms `czech`, `czechoslovakia`, `czechoslovakian`, and other words starting with the characters `czech`, including the word `czech` without any following characters.

Adding a wildcard character to the beginning of a query (for example, `*zech`) might cause the search engine to take longer to return results.

You can also use the wildcard characters in a phrase search. The wildcard character (*) matches none or multiple characters, but only one term. For example, the query `"John * Kennedy"` returns documents with the terms `John Fitzgerald Kennedy` and `John F Kennedy`, but not `John Kennedy`. However, the wildcard character (*) does not match across word separators, so `"Jo*ennedy"` will not match `"John Kennedy"` or any other variant of the name that contains white space. The wildcard character (*) matches any number of characters, including no characters, within a single term, but an asterisk will not match across a word separator (two or more terms), and there must be a term containing at least one character for it to match. For example, `"Jo*dy"` will not match `"John Kennedy"` (two terms), and `"John * Kennedy"` will not match `"John Kennedy"` (no middle term). Or, the query term `fin*int` matches `fingerprint` but not `fine print`.

Use a question mark (?) as the wildcard character for a single character. The search term `ra?or` matches `razor` or `rasor`, but not `raptor`. The query `"John ? Kennedy"` returns documents with the terms `John F Kennedy`, or `John F. Kennedy`, but not `John Fitzgerald Kennedy`.

Table 4. Examples of wildcard search

Find this	Sample query in the search query box
The nameJohn Fitzgerald Kennedy or John F Kennedy	person:"John * Kennedy"
The name John F Kennedy, or John F. Kennedy	person:"John ? Kennedy"
The name Mikhail Gorbachev	person:Mi*l Gorbachev

Searching for special characters

You can include special characters, such as " \ ( ) or Boolean operators (AND, OR, NOT), as literal terms in a search query by preceding the special character with an escape character: \. If the escape character (\) precedes a character that is not special, the escape character \ is ignored.

Single characters that require the escape character include Unicode characters, for example, \u3000, \t, \n, \r, +, -, (,), :, ^, @, [,], \, {, }, ~, *, ?, <, >, =, !. However, the only special characters that require the escape character when included in a query in double quotation marks are *, ?, ", \.

If you want to search for words or characters that are also operators in search queries, such as Boolean operators, proximity operators, or mathematical symbols such as <=, >=, !=, , insert a backslash (\) before the character or word. Special character sequences consisting of more than one symbol (for example, <= and >=) must escape each symbol; for example \<\=. Special character sequences consisting of alphabetical characters, such as SENTENCE, require only one escape character, for example, \SENTENCE.

Tip: For Lotus Notes® e-mail addresses, the special character / is retained and recognized without the escape character. For example, a query that includes "Raul Martinez San Jose" will return the Notes® address Raul Martinez/San Jose/Example.

Table 5. Examples of search with special characters or words

Find this	Sample query in the search query box
The symbols for less than or equal to (<=)	\<\=
The words car and not	car AND \not
The parenthetical phrase (but not here)	\(but not here\)

Searching by field or category

You can search by entering text into a search field, or by typing a valid query into the search query box. To search by field or category, type the field or category name, a colon (:), then the term. For example, to find documents that contain the phrase "tax evasion" in the subject line, enter "tax evasion" in the **Subject** field, or enter subject:"tax evasion" in the search query box. To find e-mails sent by joe.morales@example.org, in the **Sender** field enter joe.morales@example.org, or in the search query box enter from:joe.morales@example.org.

Restrict a query to match only in a certain field or category by prefixing the query term with the field or category name. If the query contains blank spaces that are not in the phrase, enclose the query in parentheses to group the appropriate parts

of the phrase. For example, `subject:(IBM Software)` returns only documents that contain the words IBM and `software` in the subject of an e-mail. The query `subject:"IBM Software"` returns documents that contain IBM Software in that order and without variants of the words.

The query term `subject:IBM Software` returns documents that contain IBM in the subject and `software` in the subject, body, or attachments. Or, the search query `senderdomains:example.org` returns all documents for which the domain name of the sender is `example.org` and the query `-senderdomains:example.com` returns all documents for which the domain name is not `example.com`.

Table 6. Examples of category search

Find this	Sample query in the search query box
The phrase `case sensitive`	`phrase:"case sensitive"`
E-mails sent from `john.smith@example.com`	`senders:"john.smith@example.com"`
E-mails sent to `jill.jones@example.com`	`recipients:"jill.jones@example.com"`
E-mails sent from the domain `example.com`	`senderdomains:"example.com"`
E-mails sent to the domain `example.org`	`recipientdomains:"domain.example.com"`
Any mention of the person named `John Smith`	`person:"John Smith"`
Any mention of the location `Silicon Valley`	`location:"silicon valley"`
Any mention of the company `IBM`	`company:IBM`
All e-mails flagged as `Confidential`	`flags:"Confidential"` Flags search is case sensitive.

Searching for e-mail senders and recipients

To search for e-mail senders or recipients, enter names or e-mail addresses in the **Sender** or **Recipient** search fields, or select from the **Senders** or **Recipients** categories. Any string that contains an at (@) sign in these fields is assumed to be an e-mail address. Incomplete e-mail addresses will not match the full addresses in indexed e-mails. An address that contains a wildcard is not necessarily incomplete. For example: John*@example will not match John.Smith@example.com, but John*@example.com will match. Or, Smith@example.com will not match all of the Smiths at example.com. Using wildcards in both parts of the address (identity and domain) will work: John*@example* will find John Smith@example.com. You can enter long lists of e-mail addresses into a larger text box by clicking the icon next to the **Sender** or **Recipient** field.

You can also enter names or e-mail addresses with the e-mail address fields `from:`, `to:`, `cc:`, `bcc:`, `toccbcc:` and `sentrepresenting:` in the search query box. These fields search only within the sender header, the to, cc, bcc, toccbcc lists, or principal (on-behalf-of) header, respectively.

Wildcards are allowed. Wildcards can match punctuation characters used as word separators in e-mail addresses, but wildcards do not match across word separators.

E-mail addresses are searched by taking into account the punctuation and special characters that can appear in them. E-mail addresses are not case sensitive. A word can match the e-mail address or the name of a sender or recipient. Alternative forms of terms are not supported in e-mail address search.

For example, `from:"John Doe"` does not match an e-mail from `john.doe@example.com`, but does match an e-mail from `John Doe <jdoe@example.com>`. A keyword or phrase search matches only within a single name or address. For example, the query term `to:(craig smith)` will match an e-mail that was addressed to `Craig Miller` and to `John Smith`. If the query contains an @ sign, the query is interpreted as requiring an exact match on the e-mail address. For example, the query `from:smith@example.com` matches only the e-mail address `smith@example.com` and does not match an e-mail from `john.smith@example.com`.

Table 7. Examples of search for e-mail addresses

Find this	Sample query in the search query box
E-mails sent from John Doe	`John Doe <jdoe@example.com>`
E-mails addressed to Craig Miller and John Smith	`to:(craig smith)`
E-mails sent from `smith@example.com`	`from:smith@example.com`
E-mails sent by John Smith	`from:"john smith",from:jsm?th@example.com,`
E-mails addressed to John Smith	`to:john.sm*th@example.com`
E-mails that Jill Jones was blind copied on	`bcc:jill.jo*es@example.net`
E-mails that John Smith or Jill Jones was copied on	`cc:"john smith" OR cc:"jill jones"`
E-mails that Jill Jones sent as a representative of another sender	`sentrepresenting:"Jill Jones"`

Searching for or between dates

The reserved word `docdate` can be used in a search query to restrict the search to documents that were sent within a range of dates. The search query `docdate < "2007-12-18"` returns only documents that were sent before the date December 18, 2007. The search query `docdate > "2007-12-18"` returns only documents sent after December 18, 2007. The search query `docdate >= "2006-05-16" AND docdate <= "2007-12-18"` returns only documents sent between and including May 16, 2006 and December 18, 2007.

You can also use the reserved words `postedtime` (the date the e-mail was posted) or `deliveredtime` (the date and time that the e-mail was delivered) instead of `docdate` (the date that the e-mail was sent). For example, `postedtime="2002–01–02 15:57"` searches for e-mail that was posted at 3:57 p.m. on January 2, 2002, and `deliveredtime>="2002–01–02 15:56" AND deliveredtime<"2002–01–02 15:58"` searches for e-mails delivered between 3:56 p.m. and 3:58 p.m. on January 2, 2002.

Specify dates in a fielded search with the ISO 8601 standard: `YYYY-MM-DD` or `YYYY-MM-DD HH:MM`.

You can use the operators >=, <= , =, > , or < operators. A colon (:) is not required after `docdate`, `postedtime` or `deliveredtime`. These are reserved words that can be used in query expressions, not field labels.

Table 8. Examples of date and time search

Find this	Sample query in the search query box
Documents that were sent before the date December 18, 2007.	`docdate < "2007-12-18"`
Documents that were sent after the date December 18, 2007.	`docdate > "2007-12-18"`
Documents that were sent between and including May 16, 2006 and December 18, 2007.	`docdate >= "2006-05-16" AND docdate <= "2007-12-18"`
E-mails that were delivered between 3:56 p.m. and 3:58 p.m. on January 2, 2002	`deliveredtime>="2002–01–02 15:56" AND deliveredtime<"2002–01–02 15:58"`

Tips for using the search query box

You can enter or edit search queries in the search query box that use specific field names and reserved words.

Search fields

Table 9. Keyword and subject syntax

Search field	Search query box syntax	Part of e-mail searched	Example to enter in search query box
Keywords	None	Subject, body, attachments	`stock option`
Subject	`subject:`	Subject only	`subject:401k`

- If the search term is not enclosed in double quotation marks then alternative forms are supported, unless alternate forms are not supported. The order of terms is not significant. Not case-insensitive. Special characters and reserved words used as query terms must be escaped. For field-labeled queries, multiple terms separated by white space must be grouped by enclosing the terms in parentheses.
- If the search term is enclosed in double quotation marks ("phrase" syntax) then alternative forms are disabled. The order of terms is significant, and consecutive terms in the query must be found in consecutive locations in a document to match. Case-insensitive. Only a few special characters must be escaped to be used as query terms (see detailed query-syntax page). In field-labeled queries, multiple terms separated by white space are grouped by double quotation marks.
- Wildcards are allowed in the **Subject** field. Wildcards do not match punctuation characters used as word separators, nor do wildcards match across word separators. See the detailed syntax rules regarding wildcards.

Table 10. Sender and recipient syntax

Search box field	Search query box syntax	Part of e-mail searched	Example to enter in search query box

Sender	`from:`	Sender, principal	`from:"john smith"`
			`from:jsm?th@example.com`
			`from:jjones@example.com`

Recipient	`toccbcc:`	to, cc, bcc lists	`toccbcc:"jill jones"` `toccbcc:jjones@example.com`

- Wildcards are allowed. Wildcards do not match punctuation characters used as word separators, nor do wildcards match across word separators. See the detailed syntax rules regarding wildcards.
- Alternative forms are not supported (not lemmatized).
- Any string that contains an at (@) sign in these fields is assumed to be an e-mail address. Incomplete e-mail addresses will not match the full addresses in indexed e-mails. An address that contains a wildcard is not necessarily incomplete. For example: John*@example will not match John.Smith@example.com, but John*@example.com will match. Or, Smith@example.com will not match all of the Smiths at example.com. Using wildcards in both parts of the address (identity and domain) will work: John*@example* will find John Smith@example.com.
- The same notes apply to terms entered in the dialog that appears when the icon to the right of the search field is clicked.

Table 11. Date field syntax

Search field	Search query box syntax	Part of e-mail searched	Example to enter in search query box
Dates between	`docdate`	Date sent	`(docdate>="2008—08—12" AND docdate<="2008—08—13")`

A colon (:) is not required after `docdate`. `docdate` is a reserved word used in query expressions, not a field label. Alternae forms are not supported.

Categories

Table 12. Categories

Search field	Search query box syntax	Part of e-mail searched	Example to enter in search query box
Phrases	`phrase:`	Subject, body, attachments	`phrase:"case sensitive"`
Senders	`senders:`	Sender, principal	`senders:"john.smith@example.com"`
Recipients	`recipients:`	To, cc, bcc lists	`recipients:"jill.jones@example.com"`
Sender Domains	`senderdomains:`	Sender, principal	`senderdomains:"example.com"`
Recipient Domains	`recipientdomains:`	To, cc, bcc lists	`recipientdomains:"domain.example.com"`
People	`person:`	Subject, body, attachments	`person:"John Smith"`
Companies	`company:`	Subject, body, attachments	`company:IBM`
Locations	`location:`	Subject, body, attachments	`location:"silicon valley"`

Alternate forms of terms are not supported.

Table 13. Flags

Search field	Search query box syntax	Part of e-mail searched	Example to enter in search query box
Flags	`flags:`	Flags table	`flags:"Confidential"`

Flag names are case-sensitive. Alternate forms are not supported.

Search query box

Table 14. Senders and recipients

Search field	Search query box syntax	Part of e-mail searched	Example to enter in search query box
None	`to:`	to list only	`to:john.sm*th@example.com`
None	`bcc:`	bcc list only	`bcc:jill.jo*es@example.net`
None	`cc:`	cc list only	`cc:"john smith" OR cc:"jill jones"`
None	`sentrepresenting:`	Principal only	`sentrepresenting:"Jill Jones"`

- If the search term is not enclosed in double quotation marks then alternative forms are supported, unless note 5 applies. The order of terms is not significant. Case-insensitive. Special characters and reserved words used as query terms must be escaped. For field-labeled queries, multiple terms separated by white space must be grouped by enclosing the terms in parentheses.
- If the search term is enclosed in double quotation marks ("phrase" syntax) then alternative forms are disabled. The order of terms is significant, and consecutive terms in the query must be found in consecutive locations in a document to match. Case-insensitive. Only a few special characters must be escaped to be used as query terms (see detailed query-syntax page). In field-labeled queries, multiple terms separated by white space are grouped by double quotation marks.
- Wildcards are allowed. Wildcards do not match punctuation characters used as word separators, nor do wildcards match across word separators. See the detailed syntax rules regarding wildcards.
- Alternate forms are not supported.

Table 15. Reserved words for date and time

Search field	Reserved word	Part of e-mail searched	Example to enter in search query box
None	`postedtime`	Date and time e-mail posted	`postedtime="2002–01–02 15:57"`
None	`deliveredtime`	Date and time e-mail delivered	`deliveredtime>="2002–01–02 15:56" AND deliveredtime<"2002–01–02 15:58"`

A colon (:) is not required after `postedtime` or `deliveredtime`. These are reserved words used in query expressions, not field labels.

Interpreting the search results

Use the search results to learn more about the contents of the case, and then narrow the results.

The search page displays:
- The search query
- A timeline of the search results
- A list of the search results
- The content of categories found in the search results

Search query box

The complete query is displayed in an editable search query box. You can save the search for reuse, move back or forward through sequential queries, or edit the query and run the new query.

If you are familiar with the search query syntax, you can enter a query directly in the text box.

Timeline

The timeline graphically displays the number of documents in the search results by the date, month, or year that it was sent. You can adjust the date range by moving sliders, which appear when the cursor is over the timeline. Click **Update Results** to change the date range in search results to the new range.

If you adjusted the date range by using the sliders, you can click in the new specified date range to zoom in. Click on one vertical bar in the timeline bar graph to display the results for just that time period.

The timeline shows both the actual number of documents and the expected number of documents for each date range displayed. The expected document count is calculated by assuming an even distribution of results across the entire case for the active date range. If the actual count is higher than the expected count, this might indicate a time period of greater activity and therefore of more interest for the reviewer. For example, if you search for the phrase `sexual harassment` and 5% of the five million e-mails in the case matched that phrase, you might expect that 5% of the e-mails in any selected time period would match the phrase. If, however, 20% of the e-mails sent in January of 2007 matched the phrase, then you might review the e-mails in that time period for discussion of a sexual harassment incident.

Categories

The search engine evaluates document content and classifies words and phrases according to text analytics rules. The words and phrases in search results are displayed in categories. The words and phrases for each category are sorted by frequency. You can add or exclude selected words or phrases in search queries.

In the extended list of each category that is displayed when you click **more**, small bar graphs display the correlation of the words and phrases in each category to the number of e-mails in the current search results. A higher number indicates that the word or phrase appears in many of the results. A smaller number indicates that the word or phrase appears in a smaller number of the results.

Related concepts

"Search strategies" on page 1

Related tasks

"Narrowing search results with categories"

Narrowing search results with categories

You can narrow search results by using the search categories.

To narrow search results by using categories:
1. In the Search box, expand one or more of the categories.
2. Modify your query:

To do this:	Do this:
Add the selected term to the current search query immediately	Click the term. A new search runs that narrows the results by including the selected term with the Boolean AND operator.
Include one or more terms in the current search query	Click the green plus (+) sign for the selected term and click **Search**. **Add to search** is automatically selected when you search by using categories. To run a new search with only the selected terms, select **New search** and click **Search**.
Exclude one or more terms from the current search query	Click the red minus (-) sign for the selected term and click **Search**. **Add to search** is automatically selected when you search by using categories. To run a new search with only the selected terms, select **New search** and click **Search**.

- Terms that are selected within the same category, such as **Phrases**, and with the same sign, are included in the search query with the Boolean OR operator.
- Terms that are selected within the same category with the red minus (-) sign are included in the search query with the Boolean AND NOT operator.
- Terms that are selected within different categories, such as **Phrases** and **Senders**, are included in the search query with the Boolean AND operator.

For example, if you select multiple green plus (+) signs in the same category, a query such as `phrase:("legal" OR "illegal")` is generated. If you select multiple red minus (-) signs in the same category, a query such as `phrase:(NOT ("acceptable" OR "legal"))` is generated.

A query `country` returns results in the **Phrases** category such as Mexico, Italy, Tanzania, and India, and in the **Companies** category such as IBM, Tata, and Toyota. If you click the green plus (+) sign next to India, red minus (-) sign next to Italy, and the green plus (+) sign next to Tata, then click **Search**, the resulting search query is `(country AND (phrase: (India AND NOT Italy)) AND (company: Tata)`.

Related concepts

"Interpreting the search results" on page 13

Saving searches

You can save search queries that you ran, edit saved searches, open and run saved searches, and import searches.

A search query must be displayed in the search query box before you can save the search. If you see no query in the search query box, run a search by entering terms in the search fields.

Saved searches are associated with the user and with the case. There are no system-wide saved searches. To share a saved search with another user, copy the search query and send it to the other user as text.

You can select a saved search to run as your default search each time you open the search page on a case.

Saved searches can be put into groups to keep related searches together.

To save a search query:

1. From the Search page, enter or run a search query.
2. Click **Save**.
3. Specify or select a group of saved searches into which the search query will be saved.
4. Click **Save**.

Managing saved searches

You can run, edit, or delete saved searches. You can also specify a saved search as your default search query.

To manage saved searches:

1. From the search page, click **Saved Searches**.
2. Choose an action:

To:	Do this:
Run a saved search	Select a saved search, then click **Run**, or double-click the saved search..
Edit a saved search	Select a saved search, edit the group or query, then click **Save**.
Delete a search	Click **Delete** for the saved search that you want to delete.

3. Optional: Select **Default Search** to specify the selected search query as your default, to be run automatically when you open the search page to a case or when you click **Reset** in the search interface.

Exporting saved searches

You can export saved searches to share with other reviewers or to use on another case.

You can import search queries from a text file, and export saved searches to a text file. The format of the saved searches text file requires one query per line.

To export saved searches:

1. From the Search page, click **Saved Searches**.
2. Select a saved search group to export.
3. Click **Export**.
4. Specify a name for the file of saved searches and a location.

Importing searches

You can import saved searches that have been exported from another reviewer or case.

You can import search queries from a text file, and export saved searches to a text file. The text file format is one search query that uses the supported query syntax per line.

To import a search:
1. From the Search page, click **Saved Searches**.
2. Click **Import Saved Searches**.
3. Specify or browse to a text file that contains the saved searches.
4. Specify or select a group into which the search query will be saved. If you do not specify a group, the search will be saved in the **Ungrouped** group.
5. Click **Import**.

Chapter 3. Identifying document relevance with flags

Set flags to identify selected documents by relevance to the requirements of the discovery request for each case. Documents can be grouped by flags when the case administrator transfers the case back to the archive.

A flag is metadata that is associated with one or more documents, but is not part of the document.

For example, flag an e-mail document as Responsive if you think it might be responsive to a discovery request. After you review all e-mail documents in a case, and flag a subset of documents, you can then search for all documents that are flagged Responsive and group them for review by counsel. Or you can flag e-mail documents that are obviously irrelevant and then exclude those documents from subsequent searches.

The flags that you set on documents ultimately determine which documents are sent to further review. Flagged documents can be stored in a new folder within the archive, as determined by the administrator.

A default set of common flags is included. An administrator can add flags to, or delete flags from, the default set of flags to meet your business requirements, or an administrator can add flags to a case. An administrator can also edit existing flags. Flags can be customized by a case administrator, so you should work with your administrator to determine which flags to use for a particular case.

The default flags are:
- Responsive
- Non-Responsive
- Privileged
- Hot
- Never Produce
- Confidential
- Second-Level Review
- Foreign Language

After an initial search, the flags that are set on documents in a result list are displayed in the Flags category and can be selected to further narrow search results. You can also select certain flags to include or exclude from all of the searches that you run.

To set flags:
1. In the Results List, select one or more documents. Select one document at a time, or multiple documents, or select the **Select all** check box.
2. Click **Flags**.
3. Select one or more flags, then click **Save**. If flags are already set on the selected e-mail, you can modify the selection. You can also view or set the flags for an individual document by clicking the *n* **flags** link for the document.

   Related tasks

   Chapter 2, "Searching a case," on page 3

17

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

IBM World Trade Asia Corporation
Licensing
2-31 Roppongi 3-chome, Minato-ku
Tokyo 106-0032, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows: © (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. 2004, 2006. All rights reserved.

Portions of this product are:
- Copyright © Outside In® Viewer Technology 1992-2005 Stellent Chicago, Inc. All rights reserved.

Trademarks

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both: http://www.ibm.com/legal/copytrade.shtml

Microsoft, Windows, and Windows NT are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Other company, product, and service names may be trademarks or service marks of others.

Glossary

Use the glossary to help you understand search system concepts.

access control list (ACL)
In computer security, a list associated with an object that identifies all the subjects that can access the object and their access rights.

annotator
A UIMA annotator analyzes unstructured information as part of the UIMA framework. See also UIMA

authentication
The process of validating the identity of a user or server.

certificate
In computer security, a digital document that binds a public key to the identity of the certificate owner, thereby enabling the certificate owner to be authenticated. A certificate is issued by a certificate authority and is digitally signed by that authority.

certificate authority
A trusted third-party organization or company that issues the digital certificates used to create digital signatures and public-private key pairs. The certificate authority guarantees the identity of the individuals who are granted the unique certificate.

character normalization
A process in which the variant forms of a character, such as capitalization and diacritical marks, are reduced to a common form.

clitic A word that syntactically functions separately but is phonetically connected to another word. A clitic can be written as connected or separate from the word it is bound to. Common examples of clitics include the last part of a contraction in English (*wouldn't* or *you're*).

crawler
A software program that retrieves documents from data sources or content repositories and gathers information that can be used to create search indexes.

crawling
The activity of software programs that retrieve documents from data sources such as content repositories and file systems. The retrieved documents are then processed by the search engine, analyzed, indexed, and made available for users to search. See also crawler and crawl space.

crawl space
A set of sources that match specified patterns (such as Uniform Resource Locators (URLs), database names, content repositories, file system paths, domain names, and IP addresses) that a crawler reads from to retrieve items for indexing.

credential
Detailed information, acquired during authentication, that describes the user, any group associations, and other security-related identity attributes. Credentials can be used to perform a multitude of services, such as authorization, auditing, and delegation. For example, the sign-on information (user ID and password) for a user are credentials that allow the user to access an account.

data source
Any repository of data from which documents can be retrieved, such as the Web, a file system, or a database.

diacritic
A mark indicating a change in the phonetic value of a character or a combination of characters.

directory depth
The length or number of subdirectories in a file system paths. For example, the directory C:\My Company\personnel\ private\records is deeper (has more subdirectories) than C:\My Company\personnel. See also URL depth.

featured links
Links with accompanying titles and

23

descriptions that can be configured to appear at the top of the search page whenever users enter specific queries. Featured links have four parts: queries, a Web address, a title, and an optional summary.

field An area into which a particular category of data or control information is entered.

fielded search
A query that is restricted to a particular field.

File system crawler
A type of crawler that retrieves documents from directories.

free-form text
Unstructured text consisting of words or sentences.

free text search
A search in which the search term is expressed as free-form text.

GET command
An HTTP command that requests a file from a Web server.

HTML form-based authentication
An authentication method that uses forms directly in an HTML page to provide user credentials.

HTML meta tags
HTML meta tags can be added to HTML documents, for example, to give instructions to crawlers.

HTTP basic authentication
A standard authentication scheme that is specified in the HTTP protocol and is designed to control access to Web sites. HTTP basic authentication uses only a user name and a password.

HTTP proxy server
A server that acts as an intermediary for HTTP Web requests that are hosted by an application or a Web server. A proxy server acts as a surrogate for the content servers in the enterprise.

ignore text
User-defined text, such as boilerplate disclaimers, that search queries will ignore. Similar to stop words. See also stop word

index A data structure that references data items

to enable a search to find documents that contain the query terms.

IP address
A unique address for a device or logical unit on a network that uses the IP standard.

Java Database Connectivity (JDBC)
An industry standard for database-independent connectivity between the Java platform and a wide range of databases. The JDBC interface provides a call-level API for SQL-based database access.

key ring
In computer security, a file that contains public keys, private keys, trusted roots, and certificates. See also keystore file.

keystore file
A key ring that contains both public keys that are stored as signer certificates and private keys that are stored in personal certificates.

keyword match ranking
For every result, the search engine determines how closely the result matches a query. The more occurrences of the query terms that a document has and the closer these occurrences of the keywords are to each other, the higher the result is likely to appear in the list of results. Keyword match is the most important factor for returning search results.

lemma
The base form of a word. Lemmas are significant in highly inflected languages such as Czech.

lemmatization
A process that identifies the root form and different grammatical forms of a word. For example, a search for mouse also finds documents that contain the word mice, and a search for go also finds documents that contain going, gone, or went.

lexical affinity
The relationship of search words in a document that are close to each other in meaning. Lexical affinity is used to calculate the relevancy of a result.

Lightweight Directory Access Protocol (LDAP)
An open protocol that uses TCP/IP to

provide access to directories that support an X.500 model and that does not incur the resource requirements of the more complex X.500 Directory Access Protocol (DAP). For example, LDAP can be used to locate people, organizations, and other resources in an Internet or intranet directory.

linguistic search
A search type that browses, retrieves, and indexes a document with terms that are reduced to their base form (for example, so that *mice* is indexed as *mouse*) or expanded with their base form (as with compound words).

masking character
A character that is used to represent optional characters at the front, middle, and end of a search term. Masking characters are normally used for finding variations of a term in an index. See also wildcard character.

metadata
Data that describes a particular piece of information and that helps that information be retrieved (by search), browsed (by category), or filtered (by interest). Metadata is often part of a taxonomy or classification scheme.

MIME type
An Internet standard for identifying the type of object that is being transferred across the Internet.

modification date ranking

Documents that are newer are more likely to be ranked higher than documents that are older.

Ranking by modification date can be important for many types of documents. However, document dates are typically unreliable for Web site documents because the last modification date that is returned by many Web servers does not reflect the true age of documents.

n-gram segmentation
A method of analysis that considers overlapping sequences of a given number of characters as a single word rather than using blank space or puncutation to delimit words as in Unicode-based white space segmentation. For example, if n=2,

the text ABEFD is segmented into the sequence AB BE EF FD.

no-follow directive
A directive in a Web page that instruct robots (such as the Web crawler) to not follow links found in that page.

no-index directive
A directive in a Web page that instruct robots (such as the Web crawler) to not include the contents of that page in the index.

normalization
See character normalization.

normalizer
A character normalization program that scans text and reduces the variant forms of a character, such as capitalization and diacritical marks, to a common form.

parametric search
A type of search that looks for objects that contain a numeric value or attribute, such as dates, integers, or other numeric data types within a specified range.

parser A program that interprets documents that are added to the enterprise search data store. The parser extracts information from the documents and prepares them for indexing, search, and retrieval.

POST command
An HTTP command that sends information to a Web server for processing. The POST method is widely implemented in HTML files for sending forms that contain typed-in data to the server.

query log
A record of query activity. You can review query logs to find out what users are searching for, how long the queries take, and other information.

ranking
The assignment of an integer value to each document in the search results from a query. The order of the documents in the search results is based on the relevance to the query. A higher rank signifies a closer match.

Robots Exclusion Protocol
A protocol that allows Web site administrators to indicate to visiting

robots (such as the Web crawler) which parts of their site should not be visited by the robot.

robots.txt file
To prevent crawlers from crawling a server, some Web site administrators create a file on the Web server that defines an access policy for crawlers. This file, called robots.txt, adheres to the Robots Exclusion Protocol.

search cache
A buffer that holds the data and results of previous search requests.

search engine
A program that accepts a search request and returns a list of documents to the user.

search page
A Web page that accepts user queries and displays a list of search results.

search results
A list of documents that match the search request.

Secure Sockets Layer (SSL)
A security protocol that provides communication privacy. With SSL, client/server applications can communicate in a way that is designed to prevent eavesdropping, tampering, and message forgery.

segmentation
The division of text into distinct lexical units. Nondictionary-based processing includes white space and n-gram segmentation, while dictionary-based support includes word, sentence, and paragraph segmentation, and lemmatization.

soft error page
A type of Web page that provides information about why the requested Web page cannot be returned. For example, instead of returning a simple status code, the HTTP server can return a page that explains the status code in detail.

starting directory
The starting point for a file system crawl. If you enter a starting directory such as C:\mydocuments, the crawler crawls all documents in that directory, including subdirectories.

starting Web address
The starting point for a Web crawl. If you enter a starting Web address such as www.example.org, the crawler crawls all Web pages (documents) at that site that are reachable by following HTML links from the starting page.

stemming
See word stemming.

stop word
A word that is commonly used, such as *the*, *an*, or *and*, that is ignored by a search application.

stop word removal
The process of removing stop words from the query to ignore common words and return more relevant results.

synonym
Different words with the same meaning. You can improve the chances that users will find the correct document by defining synonyms that the search engine will use to expand what it searches for when one of the synonyms appears in a query.

system log
A record of computer activity as the search engine runs. You can review the system logs to check for problems on your enterprise search system.

text segmentation
See segmentation.

token   The basic textual units that are indexed by enterprise search. Tokens can be the words in a language or other units of text that are appropriate for indexing.

tokenization
The process of parsing input into tokens.

tokenizer
A text segmentation program that scans text and determines if and when a series of characters can be recognized as a token.

UIMA   Unstructured Information Management Architecture, an open-source framework and SDK for developing applications that discover relevant knowledge in large volumes of unstructured information. See also Apache UIMA.

Unicode-based white space segmentation
A method of tokenization that uses Unicode character properties to distinguish between token and separator characters.

Uniform Resource Identifier (URI)
A compact string of characters that identifies an abstract or physical resource.

Uniform Resource Locator (URL)
The unique address of an information resource that is accessible in a network such as the Internet. The URL includes the abbreviated name of the protocol used to access the information resource and the information used by the protocol to locate the information resource.

URL depth

The length or number of slashes in a Web site address. For example, the Web site address www.example.org/personnel/ private/records is deeper (has more slashes) than the Web site address www.example.org/personnel.

Documents that have greater URL depth are typically less interesting than documents that have lesser depth. See also directory depth.

user agent
An application that browses the Web and leaves information about itself at the sites that it visits. In enterprise search, the Web crawler is a user agent.

User-Agent string
Identifies the Web crawler. One of its uses is in a robots.txt file, which can deny access based on the specific User-Agent strings.

Web crawler
A type of crawler that explores the Web by retrieving a Web document and following the links within that document.

Web links analysis ranking
A method of ranking where documents with many links pointing to them are ranked higher than documents with few links.

wildcard character
A character that is used to represent optional characters at the front, middle, or end of a search term.

word stemming
A process of linguistic normalization in which the variant forms of a word are reduced to a common form. For example, words like *connections*, *connective*, and *connected* are reduced to *connect*.

Index

IBM

Program Number: 5724-V36