# Table of Contents

## CHAPTER 1 - ABOUT THIS MANUAL

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview and Target Audience</td>
<td>5</td>
</tr>
<tr>
<td>Conventions Used in this Manual</td>
<td>6</td>
</tr>
</tbody>
</table>

## CHAPTER 2 - GETTING STARTED

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to the ArchivalWare Search</td>
<td>7</td>
</tr>
<tr>
<td>Program features</td>
<td>8</td>
</tr>
</tbody>
</table>

## CHAPTER 3 - NAVIGATING THE INTERFACE

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>9</td>
</tr>
<tr>
<td>Accessing ArchivalWare</td>
<td>9</td>
</tr>
<tr>
<td>Displays and Navigation</td>
<td>10</td>
</tr>
<tr>
<td>Query Display</td>
<td>10</td>
</tr>
<tr>
<td>To enter a full text query</td>
<td>10</td>
</tr>
<tr>
<td>To enter a metadata query</td>
<td>11</td>
</tr>
<tr>
<td>Results Display</td>
<td>12</td>
</tr>
<tr>
<td>Preferences Display</td>
<td>15</td>
</tr>
<tr>
<td>Libraries</td>
<td>16</td>
</tr>
<tr>
<td>Search Options</td>
<td>16</td>
</tr>
<tr>
<td>Results List Options</td>
<td>19</td>
</tr>
<tr>
<td>View Options</td>
<td>21</td>
</tr>
<tr>
<td>Language</td>
<td>22</td>
</tr>
<tr>
<td>Search History Display</td>
<td>23</td>
</tr>
<tr>
<td>Re-executing a search from the search history screen</td>
<td>24</td>
</tr>
</tbody>
</table>

## CHAPTER 4 - TYPES OF QUERIES

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>How the ArchivalWare Search Works</td>
<td>25</td>
</tr>
<tr>
<td>Semantic Network and Pattern Matching</td>
<td>25</td>
</tr>
<tr>
<td>The Query “Pipeline”</td>
<td>26</td>
</tr>
<tr>
<td>Query Precision and Recall</td>
<td>28</td>
</tr>
<tr>
<td>Understanding Ranking</td>
<td>29</td>
</tr>
<tr>
<td>Comparison of Query Types</td>
<td>31</td>
</tr>
<tr>
<td>Concept Search</td>
<td>32</td>
</tr>
<tr>
<td>Pattern Search</td>
<td>33</td>
</tr>
<tr>
<td>Boolean (AND, OR, NOT) Search</td>
<td>34</td>
</tr>
<tr>
<td>Fielded Search</td>
<td>35</td>
</tr>
<tr>
<td>Wildcard Search</td>
<td>36</td>
</tr>
<tr>
<td>Grouped Term Search</td>
<td>37</td>
</tr>
</tbody>
</table>
About This Manual

Overview and Target Audience

This manual covers the use of ArchivalWare 3.0. This application includes features to search, edit and browse the users electronic archive of images and other digital content.

The Query User’s Guide describes the operation of ArchivalWare’s web-based Search interface, a document retrieval and collection program. It is intended for end users of the system. This guide is one of several provided to customers of PTFS, Inc.

This guide explains how to use the ArchivalWare Search program for searching online libraries on your local intranet. It assumes that you are familiar with your web browser environment.

In this manual you will learn how to navigate the Archivalware interface. You will learn the types of searches and how to use them. You will learn how to change the user preferences to best search and find the results you are looking for. In chapter 4 you will learn how the query process works. Chapter 5 provides instructions for performing basic queries while chapter 6 provides information on more advanced search capabilities. This manual will provide instruction on how to use the browse hierarchies to view the documents available in the digital document system.
Conventions Used in this Manual

This manual is intended to be concise and comprehensive. There are specific notations and icons used in this manual which have a specific purpose.

Step-by-step procedures involving menus and commands will appear in a list with menus and commands in **bold**, as in the example below:

**To suspend a user:**

1. Click the **User Management** menu tab and select the **Edit User** sub-menu;

2. Click **Yes** for the **Is Suspended** setting

3. Click the **Save Changes** button to apply the setting

When you are instructed to input information, the sample user input appears bolded in a different type font, as in the query "**bank earnings.**"

In the left margin, you may find helpful notes or tips. These will usually be marked with an icon which puts the note in one or more of the contexts. The icons will appear with a gray backdrop to attract attention. It is important to consider these notes along with the primary text in this manual to provide a complete perspective.

---

**ICON KEY**

- 🌟 Valuable information
- 🔧 Advanced Feature
- ✏️ Useful Tip
- ☠️ Use caution!
Getting Started

Introduction to the ArchivalWare Search

ArchivalWare allows users on “internet” or local “intranet” systems to use the sophisticated Excalibur RetrievalWare™ search engine to access online document libraries. With ArchivalWare, you can perform Boolean, proximity, wildcard, truncation, date and date range, and other standard searches over existing libraries. ArchivalWare is very easy to use because you can also enter queries in plain English (Concept/Natural Language), without any special operators, complex nesting of statements, or rigid syntax. Quickly formulated queries can achieve highly accurate and complete results on the first try. Finding words with incorrect or missing characters is overcome by using pattern searching.

Concept querying uses a semantic network of word associations enabling the expansion of search terms by using variations, synonyms, antonyms, and other relationships to search the entire document text. The degree to which the user relies on those relationships is a preference option for each individual user.

While a Concept query, for example, will search a document’s full text and fielded metadata for matching terms, a metadata (key field) query operates by searching only the predefined fielded information. Performing a search using metadata fields narrows results by allowing you to enter specific document attributes (such as a full or partial title, an author, etc.) as search terms.

Because of these features, users do not have to build and maintain complex knowledge bases of their own in order to establish relationships.
between topics, nor do they have to meticulously formulate complex queries to find information that may be worded differently in the documents being searched.

Archivalware also has a Browse capability. Browse allows the user to peruse the libraries, without searching, by means of a hierarchical browse structure. You can go to one of the browse options and browse down through the hierarchy (or tree) to see what types of documents exist in different areas of the holdings.

**Program features**

The ArchivalWare Search interface allows you to do some or all of the following, depending on your system configuration:

- Choose which libraries to search
- Query using wildcards, number and date ranges, and exact phrases
- Query using Boolean operators
- Query using semantics
- Query using concept (natural language) searches
- Set a maximum number of documents to be retrieved
- Filter your query through data fields. For example, search for documents by certain authors, or in certain document categories (Purchase Orders, Correspondence, etc.)
- Browse summary information about each document, without opening it
- Launch documents in their native format, from within ArchivalWare
- Print or email retrieved documents and images.
Navigating the Interface

Overview

The ArchivalWare search can be run from any PC or workstation using a web browser such as Microsoft Internet Explorer (version 5.0 or later). The ArchivalWare search interface follows the standard web-based interface conventions for navigation, including the use of windows (which you may place where desired by clicking and dragging with the mouse), window control menu buttons, vertical and horizontal scroll bars, push buttons, dialog boxes, and pull-down lists.

Accessing ArchivalWare

First, open a web browser such as Microsoft Internet Explorer and point it to this location:

http://your host’s name or IP address/awweb/login.jsp

Figure 1 ArchivalWare Login Page
The ArchivalWare Search interface first opens to the Login Page. Provide your ArchivalWare username and password, be sure the Search Page button is selected and click the Log In button.

Next, ArchivalWare opens to the Welcome Page, from which you can perform queries.

Displays and Navigation

Query Display
The query box (search panel) is used to enter searches on full text or fielded metadata.

To enter a full text query
1. Select a query type (Pattern, Concept, Boolean) from the dropdown list (Pattern is usually the default query type).
2. Enter a search string in the top box of the ArchivalWare Search Panel.

3. Select the **Search** button to launch a full-text search of documents.

![ArchivalWare Query entry pane](image)

Example searches (these searches will yield results when searching on the PTFS ArchivalWare demo) include:

- **Pattern**: commemorative
- **Concept**: meteorology
- **Boolean**: Franklin Roosevelt

*See Chapter 4 for a further explanation of each query type.*

**To enter a metadata query**

1. Enter metadata terms (one or more), or select from the drop-down list in the metadata query boxes to prepare a search of metadata fields only.
2. Select the **Search** button to launch the search.

Example searches (these searches will yield results when searching on the PTFS ArchivalWare demo) include:

- **Creator/Author**: William Douglas
- **Source**: Library of Congress
- **Resource Type**: Photographs

*See Chapter 4 for a complete explanation of fielded searches.*

### Results Display

When you execute a search (by clicking on the **Search** button), the Results Page displays the documents that match your criteria.

![Figure 4 - Results List](image)

The columns shown in the Results list (Figure 4) are dependent upon preferences set by the user (see the next section, **Preferences Display: Result List Options**). The five columns that show as the default are defined in Table 1 below:
<table>
<thead>
<tr>
<th>Display</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>Displays the numerical position of the document within the Result list.</td>
</tr>
<tr>
<td>Type</td>
<td>Click on the icon to open the document in its native format. The different icons designate which library the document belongs to.</td>
</tr>
<tr>
<td>Rank</td>
<td>Indicates the document’s probable relevance to the search query.</td>
</tr>
<tr>
<td>Title</td>
<td>Displays the document title.</td>
</tr>
<tr>
<td>Size</td>
<td>Shows the size of the document in kilobytes.</td>
</tr>
</tbody>
</table>

Table 1 – Results list column description

Click on a document title or the **Type** icon in the results screen to open the document in its native format.

If the document contains text, the keywords successfully matching the search query are highlighted within the document.

The following buttons, which appear at the left of the window, may be used to navigate through the results:

<table>
<thead>
<tr>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Result set contains a list of documents successfully matching the search. The number of documents contained within a set is a preference set by the user (see the next section, <em>Preferences Display</em>). The default is 25 documents per Result set.</td>
</tr>
</tbody>
</table>

| First Result Set | jumps ahead to the first set of query results. A grayed -out icon indicates the current list is at the beginning of the set. |
| Previous Result Set | displays the previous set of query results in the series. A grayed -out icon indicates the current list is at the beginning of the set. |
| Next Result Set | displays the next consecutive set of query results in the series. A grayed -out icon indicates the current list is at the end of the set. |
| Last Result Set | jumps ahead to the last set of query results. A grayed -out icon indicates the current list is at the end of the set. |
| Result Sets | indicates how many Result sets were returned, with the current list shown in bold. Click on a number to jump directly to that list. |
Table 3 - Results buttons

If you are at the beginning or the end of the results list the previous/next icons will appear as an outline indication that there are no result sets to go to.

Figure 5 Document View

Document View

- **Show Document in a Separate Window** opens the document in a separate window.

- **Show Metadata in a Separate Window** displays the document’s metadata in a new window.

- **Hide Menu** removes the buttons and query pane at the left of the window and displays the document in a larger view.

- **Previous Document** displays the previous document from the list of returned documents, regardless of which set it is in.

- **Next Document** displays the next document from the list of returned documents, regardless of which set it is in.

- **Document Number** tells which document in the results set you are viewing.
Preferences Display

Use the Preferences button at the top left of the page to navigate to the Preferences display. The preferences are arranged in a tabular format. Just click on the tab at the top of the screen to go to that section of the preferences or click All to view all of the preferences on one screen. The Preferences page, set to show ALL preferences is displayed in Figure 6.

![Figure 6 Preferences Display](image)

Use this display at any time to choose preferences for the ArchivalWare Search interface operation. This display includes many settings which control the way in which the query program displays and processes queries. These include libraries to search, result summaries, word expansion, etc. Preferences must be set prior to conducting a search in order to get the desired results. There are five categories of options: Libraries, Search, Result List, View and Language defined below.
**Libraries**

A Library is a grouping of documents. Documents may be grouped together into libraries for security reasons or to keep certain types of documents together. You may search one or more of these libraries at a time. To choose which libraries to search, click on the checkbox next to the name of the library. A check in the box indicates that the library will be searched; an empty box means that the library will not searched. To turn all libraries on or off click on the checkbox next to the choice All Libraries.

![Figure 7 Libraries](image)

**Search Options**

The Fields List [Search Options] parameters constitute a list of the metadata fields available for searching. Use these fields as filters to narrow the search query. To add a metadata field to the search panel, select the checkbox next to the field in the search options section and click the Save Changes button. The metadata field will now appear in the search panel. To conduct a search, type the word or phrase into the appropriate field search box and click Search.
<table>
<thead>
<tr>
<th>Options</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Natural Language Expansion</strong></td>
<td>Controls the relationship between related words or concepts that may be relevant to your query. This parameter only takes effect when you have chosen Concept as the search mode on the main search panel. Relationship examples include synonyms, antonyms, and irregular forms. See Chapter 3, Query Precision and Recall, for a further explanation of each type of relationship.</td>
</tr>
<tr>
<td><strong>Word Expansion Limit</strong></td>
<td>Sets the maximum number of semantic expansion words that may be used for any concept query word. This parameter limits the number of expansion words that will be added to the query for each word in Concept mode, and for any word with the exclamation point (!) operator in Pattern or Boolean mode. This setting defaults to 20. See Chapter 3, Comparison of Query Types, for a further explanation of expansion words and the ! operator.</td>
</tr>
<tr>
<td><strong>Maximum Number of Spellings for Pattern Search</strong></td>
<td>Controls the degree to which a result can vary from the search query and still be considered a hit. Set the number lower to limit the variance, higher to allow for greater divergence. This feature is helpful in environments where documents have scanning and OCR errors, or when words have various or difficult spellings. This parameter only takes effect when in Pattern query mode. See Table 8 for examples.</td>
</tr>
</tbody>
</table>
Fuzzy Spell Half Words

This feature attempts to match words which may be only partially visible by best completing the visible pattern. A search for “Technology”, for example, would still return a match, even if only “Technol” were visible on the document.

<table>
<thead>
<tr>
<th>A setting of:</th>
<th>Returns matches such as:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>congress</td>
</tr>
<tr>
<td>5</td>
<td>congress, congressional, congressman</td>
</tr>
<tr>
<td>50</td>
<td>congress, congressional, congressman, concessions, countless, careless, greatness</td>
</tr>
</tbody>
</table>

Table 7

Table 8
## Results List Options

The **Results list** is displayed, with columns showing the document attributes selected in the **Result Options Fields List**.

<table>
<thead>
<tr>
<th>Fields List</th>
<th>Determines what attributes of the retrieved documents are displayed in the Results list. The listed attributes default to the document title, its ranking, and its position within the Result list.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hits to Browse at a time</td>
<td>Determines the number of document titles that will display, per page, in the Results Frame.</td>
</tr>
</tbody>
</table>
| Summary Type | Controls what information displays for each document in the Results display in addition to ranking numbers, document type, and titles. Displaying summaries is useful when your library has large documents, and the titles do not provide a good indication of the document contents. Note that you must set the Summary Type prior to executing the query for which you want to see the summaries.  

- **None**: Only ranking numbers, titles, and library names display. None is the default setting.  
- **Top Lines**: This summary type displays the top lines of the document.  
- **Query Based**: This summary type displays the highest-ranked hits from the document. |
| **Center Hit Mode**: Displays the highest-ranked hits from the document, but it ignores sentence boundaries. The hits are “centered” within each line, so part of the previous or following sentence might display, or only a single partial sentence (if it is a long sentence). |
| **Proper Names**: Choice of Person, Place or Entity. This setting displays any proper names that are located in the text of the document. Choosing proper name (person) might display: PERSON: George A. Custer USMA |
| **Max sentence length** | Sets the maximum number of characters to show in the summary section of the results screen. This option is only activated when a summary type is chosen. |
| **Number of sentences** | Sets the maximum number of sentences to be displayed in the results screen. This option is only activated when a summary type is chosen. |
| **Maximum number of documents to retrieve** | Limits the number of documents returned by a query. Set this number lower to increase the speed of the search query. |
| **Sort By** | Allows you to choose to sort the results list alphabetically by metadata field. The default is to sort by relevance ranking with the most relevant appearing first in the list. There are three sort levels. You may choose a primary sort only or add a secondary or up to a third level sort. To have the sort appear in descending order put a check in the box next to descending. To sort in ascending order, uncheck the descending box. |
| **Show File Size** | When turned on, this option shows the size of the file in kilobytes as the last column in the results list. The default setting is on. |
| **Result list display format** | There are three result list display options. The first is Column Format, which is the default setting and shows the search results in a columnar format as shown in figure 4. The Thumbnails Format option displays thumbnails of the returned documents (if available) in the Results list (see figure 10). The third option is Row Format. This choice displays the information in the results list for each document by row. |

Table 9 - Results Options
The View Options determine what information shows while viewing the opened document.

<table>
<thead>
<tr>
<th>View Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fields List</strong></td>
</tr>
<tr>
<td><strong>Highlight text inside document</strong></td>
</tr>
</tbody>
</table>

Table 10 View Options
Language options provide the ability to change international settings.

<table>
<thead>
<tr>
<th>International Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Query Language</strong></td>
</tr>
<tr>
<td><strong>Match Language</strong></td>
</tr>
<tr>
<td><strong>Expansion Language</strong></td>
</tr>
</tbody>
</table>
To reset the Preferences page to its original state and remove any changes that may have been made, select the restore default button.

**Search History Display**

Access the Search History display by selecting the Show Search History button.

The Search History window displays a record of all the previous searches performed during an ArchivalWare search session. Use this feature to access the parameters and stored results of previous ArchivalWare searches. The Search History feature makes it easy to edit and re-execute the query later, if you wish.

The information shown in the Search History includes the search string, the number of results returned, and the metadata field searched, if applicable. Searches are listed in the order they were performed. If a search within current results was executed, it is indicated by an outline number (example: 1.1) and is indented below the original search string.
Re-executing a search from the search history screen

1. Click on the underlined search string next to the number in the search history list.

2. The search will automatically be executed and the results list will appear.

Keep in mind, if the preferences are set differently than the first time the query was done, the results may be different.
Types of Queries

How the ArchivalWare Search Works

Semantic Network and Pattern Matching

ArchivalWare can analyze query terms as units of meaning. When you enter a query, ArchivalWare searches not only for exact word matches, but also for related words or concepts that may be relevant to your query (this is called “word expansion”). What makes this possible is ArchivalWare’s built-in “semantic network,” comprising approximately 400,000 word meanings and over 1.6 million expansion links between words, compiled from published electronic dictionaries and other lexical sources.

ArchivalWare can also analyze query terms as a pattern (instead of using semantic expansion), which tolerates spelling differences in either the body of the text or the queries. This is particularly useful in environments where documents have scanning and OCR errors.

ArchivalWare provides three primary search types: Pattern, Concept, and Boolean. You can even mix these types within a single search, if you wish.

In Concept search mode, ArchivalWare utilizes English-language dictionaries and thesauri as a knowledge base from which to process your queries. These sources provide information about word meanings, syntax, word variations, and relationships between words. These defined relationships between words make it possible to link them together in a “semantic network”. In the semantic network, each word meaning has an associated list of words and link strengths, indicating how closely each word is linked to that meaning. Individual words can be linked to multiple
meanings, at different strengths. You can control how many and which links are traversed by changing the expansion level, or by using no expansion at all.

The semantic network makes it possible for ArchivalWare to search for concepts, or units of meaning, instead of searching for just exact matches to your query words. For example, a query of “job seeker” may also locate the similar concepts of “applicants,” “candidates,” “hired,” and “opening.” Because ArchivalWare can search for concepts, it delivers a far more complete and relevant set of responses than other text retrieval programs.

In Pattern search mode, you can search for patterns that match your query; this can be useful in situations where you have “dirty” OCR data, or words with various or difficult spellings.

Boolean mode is a fast way to look up documents with (or without) a specific term or terms.

**The Query “Pipeline”**

Think of the ArchivalWare search process like a pipeline—at one end, you enter query words. During their journey down the pipeline, the query words undergo several phases of analysis and processing. This processing both contracts and expands the original list of query words until a final search list is created.

First, the words are tokenized (tokenization breaks strings of characters into words, including special forms such as dates or phone numbers). ArchivalWare then uses the dictionary for morphological analysis (reducing words to simpler forms by stripping off suffixes and re-spelling plurals) and idiom processing (translating phrases that have a meaning beyond that of the individual words added together, such as “real estate”). ArchivalWare also removes certain small-function words (like “the” or “of”) that provide little value in locating the information you are looking for.

As the query words travel farther down the pipeline, they are expanded via the dictionary and the links in the semantic network. When ArchivalWare expands words, it finds other terms and concepts related to the query words and adds them to the list of search terms. The list of words is ranked, so that exact query words are ranked highest, then closely related terms, then more distantly related terms. This ranked word list is used by the ArchivalWare program to search the document indexes in the library. During the search, the program determines:
• how many of the exact query words are contained in a document
• how many related terms there are
• what the relationships are (strong synonym, antonym, related-to, etc.)
• the proximity (physical distance from each other) of relevant words within a document

Using this information, ArchivalWare identifies and ranks the “hit words” in documents. Based on the strength and number of these hits, documents are found and ranked in order of probable relevance. By default, hit words are color-coded within the text so that you can quickly find the most relevant parts. A summary diagram of the search process follows.

![Query Pipeline Diagram]

Figure 14 Query Pipeline
• **Tokenizing** identifies strings of characters as words, dates, or numbers.

• Small-function **stop words** are marked within the query word list.

• **Morphology** reduces query words to their root forms.

• If any query word has a **pattern matching** operator (~), or if you are in Pattern search mode, the list of query words is expanded to include words in the library that match the pattern search word.

• Words enclosed in quotation marks (" ") are grouped so that they can be searched as **exact phrases**.

• **Idioms** (such as “real estate” or “ice cream”) are identified so that occurrences of the phrase are ranked higher than occurrences of the individual words.

• Numeric entries are normalized so that they can be searched on as **numbers or dates**, as appropriate.

• If any query word has a **wildcard** ( *, ?, [search expression], _, @, #, \, ^), the list of query words is expanded to include words in the library that match the wildcard.

• Using **semantic network expansion**, words related to the remaining query words are added to the query word list.

• Documents are **ranked by relevance** and displayed in a list.

**Query Precision and Recall**
The ArchivalWare system allows you to control those parts of the pipeline search that affect precision and recall. Recall is a measure of how well the system can find **all** of the relevant documents in the database. Precision is a measure of the system’s ability to return **only** relevant documents.

In most text retrieval systems, there is a trade-off between recall and precision, so that as one goes up, the other goes down. For example, assume you have a database of 10,000 documents, of which 1,000 are relevant to a particular query. If the system returns 200 of those 1,000 relevant responses, you would have 20% recall and 100% precision. If the system returns 5,000 responses (including all 1,000 relevant responses), then the recall is 100%, but the precision is only 20% (because it returned an extra 4,000 non-relevant responses).
ArchivalWare makes it possible for you to achieve both high precision and high recall. In ArchivalWare, recall is controlled by your ability to set the level to which words are expanded. Word expansion levels tell the system how "deep" into the semantic network to go in looking for terms related to your query words (and thus relevant to the search). Expansion levels (and examples for a single definition of the word "catch") are defined in table 12 below:

<table>
<thead>
<tr>
<th>Expansion Level</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>exact or slight variations</td>
<td>Exact word matches and morphological variants</td>
<td>catch, catches</td>
</tr>
<tr>
<td>other variations</td>
<td>Simple variations</td>
<td>caught</td>
</tr>
<tr>
<td>irregular form</td>
<td>Most strongly related concepts</td>
<td>catchable, catcher</td>
</tr>
<tr>
<td>strong synonyms / synonyms / related terms</td>
<td>Strongly related concepts</td>
<td>bag, arrest</td>
</tr>
<tr>
<td>strong antonyms / antonyms</td>
<td>Weakly related concepts</td>
<td>capture, secure, seize, nail</td>
</tr>
<tr>
<td>contrasted terms</td>
<td>Most strongly unrelated concepts</td>
<td>miss</td>
</tr>
</tbody>
</table>

Table 12

The higher you set the expansion level, the greater the recall, but also the less the precision, potentially. Sophisticated ranking algorithms that factor in completeness, contextual evidence, proximity, and hit density contribute to increased precision.

**Understanding Ranking**

In order to have a greater appreciation of how the overall search mechanism works and how adjusting various options affects that search, it is important to understand how ranking works.

During the query process, one of the last activities performed by ArchivalWare is "coarse grain" and "fine grain" ranking. It is during this process that documents are retrieved and ranked. (This process applies only to Pattern or Concept queries, not Boolean.)

In coarse grain ranking, ArchivalWare simply looks for the existence or absence of query words or related terms in the document. The program retrieves the top documents—up to the maximum number of documents
you set—based on the coarse grain rank. The program then performs fine grain ranking on those retrieved documents.

In fine grain ranking, ArchivalWare analyzes the retrieved documents to determine the exact document rank. This rank value can be used for the document list sort, if you want to see documents in order of probable relevance (or you may sort by a field such as date or title, if you wish).

The following further explains how each of these processes work. It should give you a better idea about why documents are returned on a particular query, and why some are ranked higher than others.

Coarse Grain Rank Calculation
The coarse grain rank calculation takes into account the following factors. Each factor adds a certain relative “weight” to the document. Added together, these weights determine a document’s relevance.

COMPLETENESS: The greater the number of query words (either exactly or by reference), the higher the weight.

A relevant document should contain at least one term or related term for each word in the original query. If the document contains only a fraction of the original words, then the maximum rank of the document is equal to this fraction. For example, if the document contains only 3 out of the 4 original terms in the query, then its maximum rank is 75%. Related terms contribute less weight than the original (exact) words. If you assign different weights to query words, these weights are factored into the completeness score.

CONTEXTUAL EVIDENCE: The greater the number of related terms, the higher the weight.

Words are supported by their related terms. If a document contains a word and its related terms, the word is given a higher weight because it is surrounded by supporting evidence. For example, the word “charge” near the words “credit,” “debt,” and “card” is more likely to mean “charge card” than to mean “ward,” “battery energy,” or “to assign a task.”

SEMANTIC DISTANCE: The more closely related the terms, the higher the weight.

The semantic network contains information on how closely two terms are related (for example, words that are synonyms of each other are more closely related than words that are defined as antonyms of one another). This is used to compute the amount of contextual evidence that supports a word. The closer the terms are in relationship to the query words, the
more weight they are given. Thus, ranking order is accordingly adjusted based on semantic distance.

The *semantic* distance of terms differs from the *physical* distance that separates hit words in a document. Physical distance is used in fine grain ranking, as discussed below.

**Fine Grain Rank Calculation**

The fine grain rank calculation looks at the physical location of query words and related terms within the document, as well as their total number, and takes into account the following “weighting” factors:

**PROXIMITY:** The closer together the query words and related terms within the document, the greater the weight.

A document is judged more relevant if it contains related terms that occur close together, preferably in the same sentence or paragraph. The system computes a factor for physical proximity, which is greatest for adjacent terms, and lessens as terms become increasingly distant (physically) from each other. Thus, documents with many hits close together are ranked higher than documents where those same hits are present, but scattered very far apart.

**HIT DENSITY:** The greater the ratio of query words and related terms to the total number of words in the document, the greater the weight.

A document is judged more relevant if a large number of the total number of words it contains are query words or related terms. Thus, short documents with many hits are ranked higher than longer documents where those same hits are present.

**COMPLETENESS, CONTEXTUAL EVIDENCE, SEMANTIC DISTANCE:** For any set of terms in a document, ArchivalWare computes the same factors of completeness, contextual evidence, and semantic distance, just as with coarse grain ranking.

**Comparison of Query Types**

All full-text queries are processed in one of the three main query modes (Pattern, Concept, and Boolean). There are other query types that may be used either in conjunction with the primary query, or after the initial query. Following is a comparison of ways to narrow or widen the search, ease of use, and general response time for each query, as well as conditions under which you may want to use that type of query.
The size of a response is always affected by the number you enter as the maximum number of documents to be retrieved (see Chapter 4); therefore, this option is not listed as a way to narrow or widen the search itself, but it will limit the number of documents returned by any given query.

*If any terms below are unfamiliar to you, look in the Glossary at the back of this guide for a definition. For detailed entry instructions on any of the following types of searches, see the applicable sections contained in Chapters 5 and 6.*

**Concept Search**
A primary query mode that lets you enter plain English queries with no operators; automatically does semantic expansion on all query words to the level you set; ranks returned documents for relevance

- **Search is narrowed by:**
  - setting a lower word expansion level
  - including an exact phrase in double quotes
  - using fields as filters

- **Search is widened by:**
  - setting the word expansion level higher
  - using a special operator to designate a specific query word for pattern (~), or wildcard (*, ?, [search expression], _, @, #, \, ^) expansion

- Query entry is easy—just enter it in plain English

- Search process is typically slower than Boolean searches, since ranking and semantic expansion are performed; however, recall and precision are higher, making response evaluation faster and easier

- Most effective when:
  - you are first learning the system (it is one of the easiest searches to do)
  - you need to perform a “quick and dirty” search
you know the words you are searching for exist in the database

- Examples:
  - German history
  - Nobel Prize winners

**Pattern Search**

A primary query type that also processes plain English queries, but tolerates spelling differences in either the body of the text or the queries; automatically does pattern expansion on all query words to the number of words you set; ranks returned documents for relevance

- Search is narrowed by:
  - setting a lower number of pattern expansion words
  - including an exact phrase in double quotes
  - using fields as filters
  - lowering the maximum number of word spellings

- Search is widened by:
  - setting a higher number of pattern expansion words
  - using a special operator to designate a specific query word for semantic expansion ( ! )
  - raising the maximum number of word spellings

- Query entry is easy—just enter it in plain English

- Search process is typically slower than Boolean searches, since ranking and pattern expansion are performed; however, recall and precision are higher, making response evaluation faster and easier

- Most effective when:
  - you have “dirty” OCR data
- you are looking for a word that is a proper noun with variant spellings
- you are looking for a specific term or phrase, but you are not sure of the spelling

Examples:
- German history
- Noble Prize winners

Boolean (AND, OR, NOT) Search

A primary query type that uses traditional Boolean operators to find exact matches for all query words you enter; no ranking of responses; no automatic expansion

- Search is narrowed by:
  - using fields as filters
  - using an exact phrase in double quotes
  - using certain Boolean operators (AND, NOT, WITHIN, ADJ, and nested statements)

- Search is widened by:
  - using an OR operator
  - using a special operator to designate a specific query word for semantic (!), pattern (~), or wildcard (*, ?, [search expression], _, @, #, \, ^) expansion

- Query entry is more difficult and time-consuming than plain English queries, since Boolean operators and correct syntax are required
- Search process is almost instantaneous since no ranking and no automatic expansion are performed; however, it may take you longer to sift through a large response since the most relevant documents can appear anywhere on the list
- Most effective when:
  - you are looking for a proper noun only
• you are searching for words (such as a title) or a document you have seen before

• Examples:
  o German history NOT Eastern
  o Nobel AND Prize AND winners

**Fielded Search**

A query performed against the document metadata, it limits responses by allowing you to enter specific metadata fields (such as a full or partial title, collection, author, etc.). A metadata field search can be done against metadata only (one or more fields at a time) or in conjunction with a full-text keyword query.

• Search is **narrowed** by entering a greater number of fields or very specific field information

• Search is **widened** by:
  o using fields in conjunction with a normal full-text keyword query
  o entering fewer fields or more general field information (such as one word of a title)

Query entry requires field entries (as many as you want) in addition to, or instead of, entry of the full text query

Search is generally faster because it quickly eliminates the areas of the database not matching your field entries

• **Most effective when:**
  o you want documents limited to a certain date range
  o you are familiar with the database and know certain titles, authors, publication numbers, etc. that you are searching for
  o the document field information is as important as (or more important than) the full text search terms
Wildcard Search

A query that substitutes part of a word, name, or number with a wildcard character (‘*’, ‘?’, [search expression], _, @, #, \, ^) to substitute for unknowns in the search terms or database, or to search for multiple terms; wildcards can be used in Concept or Boolean mode (not Pattern), in full text search or field entries, in multiple words, and even multiple times in one word.

- Search is narrowed by:
  - using a lower number of wildcards in the search expression

- Search is widened by:
  - using a higher number of wildcards to abstract the search

- Query entry is as easy as for the primary search type (Concept, Boolean)
• Search time is generally faster because no semantic expansion is performed on wildcard words; however, if wildcards are used too broadly (such as co*), the search time can be long and the response too large to be useful.

• Most effective when:
  ○ you are searching for a proper name that you can not remember how to spell
  ○ you are searching for multiple terms that have several similar characters (such as model “C1050,” “C1051,” and “C1052”)
  ○ you are not sure what form a particular word takes in the database, and want to ensure that all forms of it are found

<table>
<thead>
<tr>
<th>Wildcard</th>
<th>Description</th>
<th>Example</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>_</td>
<td>match one or zero characters</td>
<td>colo_r</td>
<td>color, colour</td>
</tr>
<tr>
<td>@</td>
<td>match exactly one alpha character</td>
<td>gr@y</td>
<td>gray, grey</td>
</tr>
<tr>
<td>#</td>
<td>match exactly one numeric character</td>
<td>#600</td>
<td>1600, 8600</td>
</tr>
<tr>
<td>\</td>
<td>take the next character literally, not as an operator</td>
<td>joe@ptfs.c om</td>
<td><a href="mailto:joe@ptfs.com">joe@ptfs.com</a></td>
</tr>
<tr>
<td>*</td>
<td>match anything or nothing</td>
<td>scholar*</td>
<td>scholar, scholarship</td>
</tr>
<tr>
<td>?</td>
<td>match exactly one character</td>
<td>la?er</td>
<td>later, lager</td>
</tr>
<tr>
<td>[ ]</td>
<td>search expression; can include a hyphen to indicate a range of letters or numbers; will match only one character within the brackets</td>
<td>199[1-4]</td>
<td>1991, 1992, 1993, 1994</td>
</tr>
</tbody>
</table>

Table 13 Wildcard Operators

**Grouped Term Search**
A query (Pattern or Concept) in which terms related to a common concept are grouped together with parentheses in order to improve search accuracy; the words within the parentheses are expanded, matched, and ranked for relevance as a group, instead of as individual words.
• Options for narrowing or widening the search are the same as for whatever query mode you are using.

• Query entry is as easy as for the primary search type.

• Response time will be the same as for whatever primary query type you are using.

• Most effective when:
  
  o you are searching for multiple related terms that are not in the dictionary (for example, (JFK “John Kennedy” Kennedy) election

  o you are searching for multiple terms that are not linked to one another in the dictionary (for example, (motorcycle boat RV trade shows)

“Mixed” Search
When you choose a query type (Pattern, Concept, or Boolean), all query words you enter are normally expanded that way. That is, all Concept query words are expanded to related terms via the semantic network. All Pattern words are pattern expanded, and no Boolean words are expanded at all. You can “mix” these query types by entering special operators on specific query words. This causes the terms with operators to be treated differently from the rest of the query terms for whatever query type you are using.

Concept Expansion in Pattern or Boolean Mode
To expand individual words via the semantic network when you are not in Concept mode, enter a semantic operator ( ! ) following the word. That word will be expanded to query terms with related meanings, up to the expansion level you set in the Preferences display.

For example, if you were to enter child! psychology in Boolean mode, the word “child” might expand to “youngster,” “kid,” and “children,” and the word “psychology” would not be expanded.

Pattern Matching in Concept or Boolean Mode
To expand individual words to matching patterns when you are not in Pattern mode, enter a pattern operator ( ~ ) preceding the word. That word will be expanded to similarly spelled words in the library, up to the number of pattern match words you set in the Query input field.

For example, in Concept mode, if you were not sure how to spell “psychology” you might enter child ~psychology to concept expand
“child” and still pick up the word “psychology,” even though it was not spelled correctly in the query.

**Boolean Words in Concept or Pattern Mode (Exact Phrase Search)**

To keep specific query words from being expanded when you are not in Boolean mode, enter double quotes (" ") around the terms. The words in double quotes are not expanded, and multiple words in quotes must be found in the order they were entered. This can be useful when you are looking for a specific name or phrase.

For example, if you were to enter “child psychology” magazine in Concept mode, the phrase “children’s psychology” would not match, because “child” would not expand to “children.” Likewise, “psychology of a child” would not match, even though the stop words “of” and “a” are ignored, because the terms “child” and “psychology” are out of order.

To simply prevent the expansion of multiple words—without the word order constraint—enclose each word in separate quotes (enter “Justice” “Department” cases to find both “Justice Department” and “Department of Justice”).

You can also use double quotes around phrases in Boolean mode, as a way to restrict the order of words you find. For example, child psychology would match “psychology of a child,” but “child psychology” would not because of the word order constraint.

**General Query Guidelines**

Because no two systems or databases are exactly alike, there are no hard and fast rules for entering queries and adjusting query options. The following general guidelines may be useful until you become more familiar with ArchivalWare’s capabilities:

- Start your search with a full-text query and narrow the search with fielded-queries.

- If your query is simple, use Concept mode with the “narrow” search style (expansion limit set to exact or slight var[iations]). About 50% of the time, a simple search such as this will return the results you need quickly and easily, especially if you are familiar with the database. If you are not familiar with the database, set expansion higher (strong synonyms) to make sure you do not miss relevant documents.
• **Expansion Level:** Set at strong synonyms or lower (synonyms)

• **To increase recall:**
  o set word expansion higher
  o set the number of documents to retrieve higher

• **To increase precision:**
  o choose word meanings, and pattern or wildcard expansions
  o use more search terms, and make them specific

• **To increase query speed on unfamiliar libraries:**
  o set word expansion lower
  o shorten the query

• **To increase query speed on familiar libraries:**
  o use metadata fields
  o use exact phrases
  o set the number of documents to retrieve
Performing Basic Searches

Performing a Basic Full-Text Query

Enter the query string and any additional search parameters you want to use in the query pane. There are three primary query types: Pattern, Concept, and Boolean (see Chapter 4 for details). Both Pattern and Concept queries allow you to enter queries in plain English (without special operators), and to take advantage of ArchivalWare’s unique retrieval capabilities. Boolean queries must be entered using the traditional Boolean operators (AND, OR, NOT, etc.) and make limited use of ArchivalWare’s capabilities.

Search times will vary (based on your query, the query type and options, the size of your database, and your computer’s physical specifications), but most are processed very rapidly. Rapid searching is possible because the Query program searches indexes created from the documents in your database, instead of searching the documents themselves. (These indexes are created by your system administrator, using ArchivalWare’s indexing program.)

Topics vs. Questions
When entering queries, keep in mind that ArchivalWare is not designed to answer specific questions, but rather to search for text relevant to a particular topic. If you type in a question, you will not get a factual response, but rather a number of documents that may contain answers to the question you asked. Normally, it is better to query on the topic you are interested in, instead of asking a question or entering instructions. Topic queries contain fewer irrelevant words (or “noise”) than do questions or instructions; therefore, they are easier to construct and are processed more quickly. Here are some examples:
OKAY: find articles about molecular physics

BETTER: molecular physics

OKAY: how do I plant tomatoes?

BETTER: planting tomatoes

Query Entry Steps
The following steps are provided to guide you through a basic full-text query. For more details about specific options, refer to the subsequent sections of this chapter. Follow these steps to perform a basic full-text query:

1. Select the type of query you wish to perform (Pattern, Concept, Boolean) by selecting from the drop-down list.

2. Position the insertion point (a blinking vertical bar) in the query entry panel where you want to begin inserting text. To search for full text, position the insertion point in the Full Text Box under the

Figure 16-Search Pane
words **Keyword Search.** To search by **Collection, Creator, Publisher** or **Title,** position the insertion point in the appropriately labeled fields. You may choose to query any or all of these parameters. To search by **Historical Time Period** choose from the drop-down list.

3. Type in your query.

   Capitalization is not important—the Query program is case-insensitive. If you make a mistake during entry, press **BACKSPACE** to delete what was typed and retype the correct letters.

4. Click on the **Search** button to execute the query.

   When the search is complete, the Results frame automatically opens to display the query results. If *none* of your query words are in the library, a message displays on the Results frame telling you that no documents were retrieved for the query. If this happens, simply enter another query, or change your query preferences.

5. To display the text of a document, click on its **Title** or click the **Document Type** icon (the actual icon displayed for the document type may change from one document type to another depending how the system is set up) to open the document in its native format.

   The interface displays document titles in groups (the default is 25 titles, but you can change this number in the Preferences display). To see the next full group of titles, click the **next Result Set** icon.

**Reviewing Query Results**

When the query is complete, the Results frame automatically opens to display the query results.
The document titles satisfying the query display in order of probable relevance or rank value starting from the top of the Results window. The documents are numbered sequentially. To the right of the document’s sequential number, the rank value of the document appears in brackets (such as [87]). This number refers to the degree of relevance that the document achieves within the parameters of the query. The higher the number, the stronger the degree of relevance.

Boolean searches do not have a relevancy score in the same sense that Concept and Pattern queries do, because where these queries are “gray”, a Boolean query is “black and white”. In other words, a Boolean query returns exact matches only. Due to this “hit or miss” functionality, all Boolean results have a ranking of 100.

Displaying a Document
Click on the document title or type icon and the text of the document will be displayed. When the text displays here, hit words which satisfy the query are highlighted in the text.

Navigating to Other Documents
The following buttons, which appear at the left of the window, may be used to navigate through the results:
RESULTS

A Result set contains a list of documents successfully matching the search. The number of documents contained within a set is a preference set by the user (see the next section, Preferences Display). The default is 25 documents per Result set.

**First Result Set** jumps ahead to the first set of query results. A grayed-out icon indicates the current list is at the beginning of the set.

**Previous Result Set** displays the previous set of query results in the series. A grayed-out icon indicates the current list is at the beginning of the set.

**Next Result Set** displays the next consecutive set of query results in the series. A grayed-out icon indicates the current list is at the end of the set.

**Last Result Set** jumps ahead to the last set of query results. A grayed-out icon indicates the current list is at the end of the set.

**Result Sets** indicates how many Result sets were returned, with the current list shown in bold. Click on a number to jump directly to that list.

**Browse Documents**

**Previous Document** displays the previous document from the list of returned documents, regardless of which set it is in.

**Next Document** displays the next document from the list of returned documents, regardless of which set it is in.

**Document Number** tells which document in the results set you are viewing.

Table 14

**Using Concept Mode**

In a Concept search, you enter the query in plain English. Query terms are expanded along the semantic network to other related terms, up to the expansion level you set in the Preferences display.

Word expansion levels determine which related concepts the system will seek. The higher you set the expansion level, the greater the recall, but the slower the query (because many more words may be added to the
query). In general, PTFS recommends that you set the expansion level on **strong synonyms** for a Concept query.

Use a Concept search if the terms in your query are very specific, if the library you are querying on is relatively small in scope and size, if you are familiar with the contents of the library, or if you use key fields along with the full-text query. As long as expansion is not set at a broad level, a Concept query should produce fast and accurate results.

Keep in mind the following when executing Concept queries:

- If you are not sure what word expansion level to set, err on the low side (expansion level **strong synonyms**) so that the query will be faster (you can always increase the level and re-execute the query if not enough documents are returned).

- Set the maximum number of documents to retrieve to at least 500 (500 is the default setting) to ensure good recall. This setting is dependent on the size of the library and the kind of results you want. The larger the library, the higher this value should be set.

- Use metadata fields if you are familiar with the contents of your library and want to limit results to a certain set.

**Special Operators**

You can use any of the following operators in Concept mode (singly or in combination). Terms with these operators **will not** be processed with normal concept expansion. *See Chapter 4 for details.*

- Tilde (~) for pattern match
- Double quotes (" ") for exact phrases
- Parentheses ( ) for nested statements
- Wildcards (?, *, _, @, #, \, ^, [search expression] )

**Using Pattern Mode**

Like Concept queries, Pattern queries are entered in plain English, and may be formulated and executed very quickly. Use a Pattern search if you are not sure how to spell something, or if you are searching over "dirty" OCR data (raw OCR-processed text).
Keep in mind the following when executing Pattern queries:

- If you are not sure what maximum number of pattern spellings to set, err on the low side (strong synonyms) so that the query will be faster (you can always increase the level and re-execute the query if not enough documents are returned).

- Set the number of documents to be returned to at least 500 to ensure good recall. This setting is dependent on the size of the library and the kind of results you want. The larger the library, the higher this value should be set.

- Use key fields if you are familiar with the contents of your library and want to limit results to a certain set.

**Special Operators**
You can use any of the following operators in Pattern mode (singly or in combination). Terms with these operators will not be processed with normal pattern expansion. See Chapter 4 for details.

- Exclamation point (!) for concept expansion
- Double quotes (" ") for exact phrase
- Parentheses ( ) for nested statements
- Wildcards (?,*,, @, #, \, ^, [search expression] )

**Using Boolean Mode**

Boolean queries must be entered using the traditional Boolean operators, instead of in plain English. Formulating Boolean queries is more difficult because word order and syntax can be critical to achieving the desired results, especially if you use nested statements.

Because there is no ranking, the default is to sort documents in chronological order (based on their addition to the database, not the date of the document). Be aware that the most relevant document could appear anywhere on the returned list. For this reason, do not set the number of documents to retrieve too low, or you could miss the most relevant responses.
When you enter a query, use Boolean operators (see the section on Boolean operators below for more information). The default operator is AND, so that if no operators are entered, AND is assumed between query terms. Entry of apples oranges will produce the same results as entry of apples and oranges.

Idioms (such as “first hand”) in Boolean queries are treated as the individual words, rather than as a phrase. A Boolean query on first hand simply looks for “first” and “hand” anywhere in a document. To query on the phrase—with the words adjacent to each other and in the order given—enter the phrase in quotes: “first hand”.

Like idioms, “stop words” are also handled differently in Boolean queries. Most stop words (such as “the” and “if”) will be removed from Boolean queries—however, if the stop word is also a Boolean operator (AND, OR, BUT, NOT, WITHIN), it will not be removed. (See your System Administrator if you need more information about stop words.)

Boolean queries are best at retrieving proper nouns, and words or phrases you know are in the database. For example, if you were searching only for the name of a company, a person, or a particular publication, a Boolean query would quickly return the most accurate list of hits. However, if you wanted to find a proper noun along with other search terms that should be expanded, you could use a Pattern or Concept search, enclosing the proper noun in double quotes to make it an exact phrase.

**Boolean Operators**

Queries in Boolean mode make use of the traditional Boolean operators, or their equivalent symbols, in the query:

<table>
<thead>
<tr>
<th>Boolean Operator</th>
<th>Equivalent Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>AND</td>
<td>&amp; (ampersand)</td>
</tr>
<tr>
<td>OR</td>
<td></td>
</tr>
<tr>
<td>NOT</td>
<td>^ (circumflex)</td>
</tr>
<tr>
<td>WITHIN</td>
<td>(no equivalent symbol)</td>
</tr>
<tr>
<td>ADJ</td>
<td>(no equivalent symbol)</td>
</tr>
</tbody>
</table>

The BUT operator is synonymous with the AND operator, and is typically used in conjunction with NOT (“this BUT NOT that”, or “this AND NOT that”).
AND/OR: In Boolean queries, the use of AND is assumed; the use of OR must be explicitly stated within the query. For example, if you enter electronic communications as your query, the Query program will search for documents containing both of those words, and will not return any document that does not contain both words. If you enter electronic or communications as your query, the Query program will search for documents containing either of those words.

NOT: If your query produces related responses you are not interested in, use NOT to eliminate responses you do not want. For example:

    bill clinton not hillary
    automatic or not manual
    not wsj and not ap and not reuters

WITHIN: You can use the WITHIN phrase in Boolean queries to increase precision through proximity constraints. The WITHIN phrase specifies that certain words must appear within so many words of each other. For example:

    network and security within 1
    finds network and security adjacent
    network and security within 2
    with one intervening word
    network and security within 3
    with two intervening words

The WITHIN number represents the number of “jumps” required to get from one word to the next.

The WITHIN operator will also work with a longer search string. For example:

    network and security and virus and password within 50

This query would return documents containing all four of these terms within a 50-word span.

Proximity constraints generally improve the accuracy of the search because fewer non-relevant documents (where the search words are all present but very far apart) are returned.
Be careful using parentheses when using WITHIN (especially in nested statements). The WITHIN and AND operators must be at the same level. For example:

- **Correct:** network and security within 1
- **Correct:** (general electric within 3) and (westinghouse electric within 3) within 40
- **Incorrect:** (network and security) within 1

**ADJ:** The adjacency operator, like the WITHIN operator, tests word proximity, but also checks that the two words are in order. For example:

```
faberge and egg adj 5
```

This query checks that the two words appear within 5 words of one another, and that “Faberge” comes first.

**NESTED STATEMENTS:** Enclosed within parentheses, nested statements can also improve accuracy:

```
(voting and record) and ("house of representatives" or house) and (members) within 50
```

This query returns documents containing the voting records of members of the House of Representatives. If you want the WITHIN operator to apply to only part of the search string, structure the query so that the word WITHIN falls inside the parentheses. For example:

```
radar and (terrain and tactical and symbology within 25)
```

This query would return documents containing the term “radar” anywhere in the document, but the terms “terrain,” “tactical,” and “symbology” must also appear in the document within a 25-word span.

**Special Operators**

You can use any of the following operators in Boolean mode (singly or in combination). Terms with these operators will not be processed as a normal Boolean term. See Chapter 4 for details.

- Exclamation point (!) for concept expansion
- Tilde (~) for pattern expansion
- Parentheses ( ) for nested statements
- Wildcards (?, *, @, #,, ^, \[, search expression ])

For more details on Boolean features used in specialized searches, refer to Chapter 6: Fielded Searches.
Performing Specialized Searches

Fielded Searches

Depending on what type of documents are contained in your library, it is possible for you to restrict (or “filter”) query responses by entering search terms in metadata fields, such as title, collection, date, and so on. These “fielded” queries can be used in combination with a full text query—or used alone—in order to limit the types of documents returned. You can perform a fielded search in conjunction with any primary query type (Pattern, Concept, or Boolean). When there are multiple fields set up for a library, you can use them in any combination for a given query.

Certain metadata fields appear by default on the query pane. Which fields appear depend on your particular setup. You may use any or all of these fields when performing a Fielded search. Alternate field options are found in ArchivalWare’s Preferences section (see the To perform a Fielded Search [Alternate Fields Method] section below for details). A second method for performing fielded searches (without using individual field entry lines) is discussed in the Using Field Tags section below.

To perform a Fielded Search [Default Query Pane Method]:

1. At the Query frame, position the cursor at the field entry line of any of the fields displayed such as Creator/Author or Subject and type in your entry. Some fields (Resource Type in our example) are made up of a predefined set of information. To choose from the set click on the drop-down list. An entry in a data field tells the query program to search for and return documents with data field(s) matching your entry. You may enter as few or as many fields as you wish.
2. Click the **Search** button to execute the search.

To perform a Fielded Search:

1. Navigate to the Preferences section by clicking the **Preferences** button at the top left of the page.

2. In the **Search Options** section of the **Preferences** page, select the checkbox next to the field you wish to search. Select as many fields as you wish.

3. Select **Save Changes** to confirm your choices and return to the query pane.

4. At the Query frame, the fields you selected appear on the screen, each with an entry line. Position the cursor at the entry line of the fields you selected and type in your entry. An entry in a metadata
field tells the query program to search for and return documents whose metadata match your entry.

5. Click on the Search button to execute the fielded search.

The system administrator can set up fields to process entry terms in various ways, depending on the type of data in the field. This means that an entry in one field may be processed differently from the same entry in another field, if the system administrator set the fields up differently. Check with the system administrator if you are not sure what type of entries are expected for a field. Field processing characteristics to be aware of when you enter fielded queries are described below.

**Boolean Filtering in Fields**

For more information on Boolean searching, refer to the Using Boolean Mode section of Chapter 4.

Most fields are set up to be “Boolean filters,” which applies Boolean search rules to the field. This means that your exact entry is required to be present in all returned documents. If you enter terms in two fields, both fields’ terms are required to be present in returned documents. Within a Boolean filter field, you can use the traditional Boolean operators (AND, OR, NOT, BUT, WITHIN, ADJ). If you enter no operators, AND is assumed between query terms. No concept or pattern expansion is done, unless you enter the concept expansion ( ! ) or pattern expansion ( ~ ) operators to expand specific terms in the field.

**Examples:** In a Boolean filtered title field, a search on army navy produces the same results as a search on army AND navy: documents with both "army" and "navy" in the title. But a search on army OR navy returns documents containing "army," with "navy," and containing both "army" and "navy" in the title. A search on plan! in the title field might return documents with “program” or “scheme” (concept expansions of “plan”) in the title, as well as documents containing “plan” in the title.

If a field is not set up as a Boolean filter, field processing is governed by which search mode you use.

- In **Boolean** search mode, processing is the same as if the field were a Boolean filter.

- In **Pattern or Concept** search mode, what you enter in the field is preferred, but not required in returned documents. Documents containing the entry are ranked higher on the Results tab display. Search terms are expanded according to the search mode, and you may enter concept or pattern expansion operators to create mixed searches.
Examples: A Concept search on child in the title field and psychology on the query entry pane (as a full text search) might return three documents: one with “psychology” in the document body, one with “kids” in the title, and one with both “child” in the title and “psychology” in the document body. The document with both “child” and “psychology” would be ranked higher than the other two. A Pattern search on child! psychology in the title field might return documents with "children’s” (a concept expansion) or “psychology” (a pattern expansion) in the title, as well as other expansions of those terms.

Number Searches

If you want to search the text of documents for numbers or ranges of numbers, you can do so using the following formats:

<table>
<thead>
<tr>
<th>Type of entry</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific number</td>
<td>100</td>
</tr>
<tr>
<td>Inclusive number range</td>
<td>100–200 (no spaces)</td>
</tr>
<tr>
<td>Greater than or equal to</td>
<td>&gt;100 (optional space between sign and number)</td>
</tr>
<tr>
<td>Less than or equal to</td>
<td>&lt;100 (optional space between sign and number)</td>
</tr>
</tbody>
</table>

NOTE: Be careful about searching for numbers that include dashes, such as telephone numbers or Social Security numbers. If a query has numbers separated by dashes, it will always be interpreted as a range of numbers to be matched (even if you use quotation marks to indicate an exact phrase).

You will normally want to search for numbers along with other query terms, as in:

law firms with > 200 employees

You may enter the number before or after the other query words; it makes no difference. You may also query on just numbers, if the contents of your library makes it practical.

The way number searches are processed can be affected by how the applicable fields are set up. If your results are not what you think they
should be, check with your system administrator to make sure the fields are configured properly for number searches.

**Floating Point Numbers**
Please be aware that a query of an integer such as 74 will not return hits on an equivalent floating point number value such as 74.0. However, if the query includes the floating point value (74.0) the search will return both 74.0 and 74 in the results. Also a number range query such as 74–74.99 will return hits on both floating point values and integer values.

**Date Searches**
This following section describes how ArchivalWare recognizes various date formats during indexing and querying. The following information applies to dates in indexed text and in queries and to date ranges as part of queries.

**Date Components and Formats**

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>The day is specified by one or two digits (leading zeroes are optional) comprising the whole numbers between 1 and 31, inclusive.</td>
</tr>
<tr>
<td>Month</td>
<td>The month is specified by one or two digits (leading zeroes are optional) comprising the whole numbers between 1 and 12, inclusive.</td>
</tr>
<tr>
<td>Year</td>
<td>The year is specified by one to four digits (leading zeroes are optional) comprising the whole numbers between 0 and 9999, inclusive.  If the year number is less than 100 (0 through 99), ArchivalWare interprets the number as an offset into the twentieth century; that is, it adds 1900 to the number internally. An optional property is available for the integrator to control the meaning of 2-digit years (whether to add 1900 or 2000 to them). This means that there is no way to specify dates in the years before A.D. 100. In order for any non-1900’s year (prior to 1900, or after 1999) to be recognized, all digits of the year must be present in the text or in the query, as in 7/10/1895, or 7/10/2001.</td>
</tr>
<tr>
<td>Separators</td>
<td>The character used to separate any of the above components in a fully-specified date may be any of the following: slash (/), hyphen (-), or period (.)</td>
</tr>
</tbody>
</table>

Use one of the three formats below to query on dates in either the body of the text or in fields. To inquire on dates in fields, however, your system administrator must have set up and indexed one or more Date fields for the library. Your system administrator sets up a numeric date
format for each library (depending upon how dates appear in the
documents for that library). Check with your administrator to see which
format (mm/dd/yy, dd/mm/yy, or yy/mm/dd) you should use for each
library.

The “format” indicates the order of the date components, not how many
digits, or which separators, will be used. As explained above, days and
months may be one or two digits, the year may be one to four digits, and
they may be separated by a slash, a hyphen, or a period.

- mm-dd-yy  month, day, year (default)
- dd-mm-yy  day, month, year
- yy-mm-dd  year, month, day

During indexing, the system recognizes dates that are in the format for
the library.

Example Date Entries
All of the following expressions can be recognized by ArchivalWare as the
date July 10, 1965 (both in documents and in queries):

- 07/10/65  (mm-dd-yy format)
- 10.7.65   (dd-mm-yy format)
- 1965-7-10 (yy-mm-dd format)

Any leading zeroes are ignored. Here are some other forms that are
recognized as dates, but we do not encourage you to use them due to
decreased readability:

- 007/10/65
- 7.0000010.001965

To search on a range of dates, use a hyphen between the dates. For
example (mm-dd-yy format):

- 7/10/65-7/10/95  July 10, 1965 to July 10, 1995
- 10.1.0-10.1.1   October 1, 1900 to October 1, 1901
You can also search for dates prior to or after a specific date. For example (mm-dd-yy format):

\begin{align*}
\text{} & <10.1.0 & \text{any date prior to October 1, 1900} \\
\text{} & >3/1/97 & \text{any date after March 1, 1997}
\end{align*}

**21\textsuperscript{st} Century Date Queries**

By default, the query program assumes all 2-digit dates refer to the 1900’s (1900-1999). For example, the query entry 9/25/01 would match "9/25/01" and "9/25/1901." However, to enable queries on dates in the 21\textsuperscript{st} century, an integrator may have modified your program to control the cutoff point at which 2-digit years in a query are assumed to be in the 21\textsuperscript{st} century (20\textsuperscript{xx} instead of 19\textsuperscript{xx}).

If this property is in effect, the integrator sets a cutoff value so that all queries on 2-digit dates up to that value are assumed to be 21\textsuperscript{st} century, and dates higher than that value are assumed to be 20\textsuperscript{th} century. For example, if the cutoff were set to 20, the query 1/1/20 would match the date “1/1/2020.” But the query 1/1/21 would match the date “1/1/1921.”

During indexing, 2-digit dates are currently always assumed to be 1900s dates. This means that queries on 4-digit years in the 20\textsuperscript{th} century will match 2-digit dates, regardless of any cutoff date set with this property. For example, even if the cutoff were set to 20, the query 9/25/1901 would match “9/25/1901” and “9/25/01.”

With this particular setup, when the year “01” appears in a query it means “2001,” but when the year “01” appears in a document it means “1901.”

**Exact Phrase Searches**

You can search for exact phrases within any primary query type (Concept, Pattern, or Boolean) by enclosing the phrase in double quotation marks. Exact phrase searches are useful when you know exactly how something is worded in the document library.

With an exact phrase, hits are only returned when the enclosed words occur in the same order and proximity as in the document library. This means you may inadvertently exclude some relevant documents if you are not sure of the exact phrasing in the library.
For example, if you included the phrase “united states department of justice” in your query, you would not get a hit on United States Justice Department. To eliminate this limitation, enclose each separate word in quotation marks (“united states” “department” “justice”).

Other query words may precede or follow exact phrases, and you may have multiple exact phrases within a query. Any stop words within the exact phrase will be removed. For example, if you searched for "phantom of the opera" you would also find phantoms in operas if those words happened to be in the library.

You can also use exact phrases with nested numbers, dates, wildcards, and pattern match operators:

“100 employees”
“100-200 employees”
“1/1/95-1/1/96 employees”
“execut* employees”
“~executive employees”

**Wildcard Searches**

Wildcards are useful when you are not sure what form a particular word takes in the library, when you cannot remember a name or a number in its entirety, or when you are searching for multiple terms that have several similar characters (such as model “C1050,” “C1051,” and “C1052”). You enter a portion of the word, name, or number, then use a wildcard to represent the rest.

When you execute a wildcard query, the Query program first checks the dictionary, then the document indexes for this library. If a qualifying word is in the dictionary, but not in the document indexes, it is thrown away. These wildcard operators are allowed:
<table>
<thead>
<tr>
<th>Wildcard</th>
<th>Description</th>
<th>Example</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>_</code></td>
<td>match one or zero characters</td>
<td>colo_r</td>
<td>color, colour</td>
</tr>
<tr>
<td><code>@</code></td>
<td>match exactly one alpha character</td>
<td>gr@y</td>
<td>gray, grey</td>
</tr>
<tr>
<td><code>#</code></td>
<td>match exactly one numeric character</td>
<td>#600</td>
<td>1600, 8600</td>
</tr>
<tr>
<td><code>\</code></td>
<td>take the next character literally, not as an operator</td>
<td>joe@ptfs.co m</td>
<td><a href="mailto:joe@ptfs.com">joe@ptfs.com</a></td>
</tr>
<tr>
<td><code>*</code></td>
<td>match anything or nothing</td>
<td>scholar*</td>
<td>scholar, scholarship</td>
</tr>
<tr>
<td><code>?</code></td>
<td>match exactly one character</td>
<td>la?er</td>
<td>later, lager</td>
</tr>
<tr>
<td><code>[ ]</code></td>
<td>search expression: can include a hyphen to indicate a range of letters or numbers; will match only one character within the brackets</td>
<td>199[1-4]</td>
<td>1991, 1992, 1993, 1994</td>
</tr>
</tbody>
</table>

Table 17 Wildcards

You can use wildcards in Concept or Boolean mode (not Pattern), anywhere within a search term, and even multiple times within a single search term. For example:

```
chair? la?er*
compu* AB430??Q
*classify la[sz]er
*ploy* AC[a-z][0-9][0-9][0-9]
```
Browse

What is Browse?

The ArchivalWare browse function is intended to provide users with a means to peruse the available documents via a hierarchical structure. Each structure has a number of levels. For example the top of the structure might be the cabinet with the second level being a drawer and the third level browse folder a file folder. The document itself resides at the bottom of its hierarchical structure.

Understanding browse
The user will begin to browse the documents from the topmost level of the hierarchy or tree. Each level of the hierarchy is called a “browse folder”. A browse folder with other folders underneath it is called a parent folder and the folders which fall below other folders are called children or child folders.

The browse structure for each individual document is built upon document import. The metadata of the document must contain the browse folder structure for that document.

At this point in time a browse hierarchy for a particular document may not be changed “on the fly”. To change the browse hierarchy for a document, the document must be reimported. The name of a particular browse folder may be changed through the editor and an empty browse folder can be deleted through the editor.
Navigating through Browse

The ArchivalWare Browse may be accessed through the main ArchivalWare page. Just click on the word **Browse** which is located at the top of screen in blue letters. You may toggle back and forth between the main search and browse by clicking on the appropriate words (see figure 21 below). When you click on Browse and go to that screen, then the word and **Browse** is highlighted in yellow to indicate what screen you are viewing.

Once you are looking at the main browse screen (figure 20), you may choose any one of the available browse hierarchies. The available hierarchies will vary from one installation to another. For illustrating our example we will use the PTFS ArchivalWare demo. To locate a document by browsing follow the steps in this example:

This example will show you how to dig down through a browse structure to locate a document. For this example we are looking for any Photograph images in the MHI Documents browse hierarchy.
1. Click on the **Browse MHI Documents by historical time period** link in the main browse screen.

2. Click on the “**1899-1917**” folder for the documents that belong to that source.

3. Click on the “**Allen, James > Photographs**” folder. You will now see the documents available in that folder.

Note that at this point you are viewing the “results screen.” The browse results screen is the same as the search results screen and the preferences can be changed through the preferences screen discussed in *Chapter 3 of this manual, Displays and Navigation*. To view a document click on the document title.

At all times during browse, the browse folder hierarchy in which you are browsing appears in the blue bar at the top of the ArchivalWare screen. After conducting the browse in the above example the structure that should appear is shown in figure 23.

![Browse>Time Period>1899-1917 >Allen, James >Photographs >](image)

Figure 22 browse folders are shown at top of screen

If at any time you want to go up the browse folder structure instead of down, just click the folder name in the blue bar (figure 22). It will take you back up the structure. To start over at the beginning browse screen click on the word Browse in the blue bar. Note: Do not use the back button to move back up the browse structure.

**More Browse Features**

**Browsing from within Search**

There is a capability which exists in the search results list that allows you to kick off a browse. When viewing the search results list note that the metadata in some of the columns are hyperlinks. This is true for collection which defaults to show in the results list but is also true for other fields if you choose to have them show in the columns. These include historical time period and folder. Clicking on one of the hyperlinked fields will launch a browse using that “browse folder.” For instance if you were to click on a collection name the browse would launch using that collection name and you would be able to browse any documents located in that collection.
Searching from within Browse
Searching from within the browse section is another useful feature. If while browsing a collection of documents you determine that you want to conduct a search then type the query into the full text search box and click Search. A search will be conducted within the context of the browse folders where you began the search.

Find All Occurrences of a Folder
The browse capability has a feature called “Find All”. This feature allows you to choose a browse folder and then find all of the other related browse folders. For illustrating our example we will again use the PTFS ArchivalWare demo. Follow these steps to try an example of showing a browse tree:

1. Go to the **Browse MHI Documents by historical time period** Hierarchy
2. Click on the **Late 20th Century** folder
3. Click on the **Find all occurrences of this folder Icon** located to the right of the **Clothing, Headgear** Folder (see figure 23 - Find All Folders below).
4. The “Find All” will find all of the related browse folders and show them on the screen as in **Figure 24** below.
Editor

The ArchivalWare Web Editor is a Java servlet that can be run from any PC or workstation using a web browser such as Microsoft Internet Explorer (version 5.0 or later). The program gives the user a single-record editing capability, allowing for the removal or assignment of additional metadata field information. You can also delete documents and edit browse folders (multi-record edit capability). The ArchivalWare Web Editor interface follows the standard web-based interface conventions for navigation, including the use of windows (which you may place where desired by clicking and dragging with the mouse), window control menu buttons, vertical and horizontal scroll bars, push buttons, dialog boxes, and pull-down lists.

Accessing the ArchivalWare Web Editor

Only authorized users have access to the Editing functions. Authorization is determined by login. Please check with your system administrator if you have questions regarding your system access. To login to the systems as an editor simply login to the ArchivalWare search screen with a username and password which has been given appropriate permissions.

In order to edit a document’s metadata, the desired document must first be retrieved from the Library (see Chapter 4 for details regarding different search query modes and their implementation). Once retrieved, the Editor interface is accessed through the Result list by selecting the edit icon in the second column, to the left of the desired document’s title. When you are logged in as an edit user the heading on the second column changes from Type to Edit.
### Results list with Edit Privileges

<table>
<thead>
<tr>
<th>#</th>
<th>Edit Rank</th>
<th>Title</th>
<th>Collection</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>87</td>
<td><em>Text file for boeing-747.djvu</em></td>
<td>Allen, James</td>
<td>53K</td>
</tr>
<tr>
<td>2</td>
<td>87</td>
<td><em>EDITING 747 ( DJVU FILE )</em></td>
<td>Allen, James</td>
<td>212K</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
<td><em>00000001.pdf</em></td>
<td></td>
<td>3K</td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td><em>Business Journal ( djvu format )</em></td>
<td>Allen, James</td>
<td>145K</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td><em>Proposal Repository System Requirements</em></td>
<td>Demos</td>
<td>18K</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td><em>ADMINISTRATIVE PROCEDURE ( DJVU FORMAT )</em></td>
<td>Allen, James</td>
<td>230K</td>
</tr>
</tbody>
</table>

Figure 25 - Results list with Edit Privileges

---

### Editor Screen

Figure 26 - Editor Screen
Editing Documents

Modifying Metadata
Each field appearing in the Editor window may be edited (with the exception of the modified date field, fields without a drop-down list and text boxes). To edit a field:

1. First position the cursor (a blinking vertical bar) in the textbox where you want to begin inserting, modifying or deleting text.

2. To add text, type in the box; to remove text, use the BACKSPACE or DELETE keys to remove the unwanted text.

3. Click on the Save Document button to save your changes to the database.

![Figure 27 - Metadata List](image)

Certain fields have lists associated with them. The Type field in the example above is linked to a list of possible countries already contained within the database.

1. To view a list select the pull down arrow to the right of the entry box.

2. To choose from the list use your mouse to select the desired type from the list.

![Figure 28 - Pull down list](image)
3. To modify the list please contact your system administrator. At this time lists are defined in the administrator tools and are not editable by users.

To make your changes a permanent part of the database, select the Save Document button at the lower left-hand side of the Editor frame.

A dialog box appears, confirming that all changes made in the Editor will be carried over into the database. Select OK to accept the changes you made, or Cancel to abort the save procedure and return to the Editor frame.

![Figure 29 - All changes submitted](image)

After selecting OK, a message is displayed confirming that the document has been updated with the changes made in the Editor.

![Update: Updated metadata succeed Back to Edit Mode](image)

**View Documents while Editing**

To view a document in another window while editing the metadata click on the Show Document icon in the upper right hand corner of the edit window.
Deleting Documents

To entirely delete a document and the document’s associated metadata from the document set, first follow the procedure described in Accessing the Editor, above, to retrieve the desired document. Select the Delete Document button at the top or bottom right-hand corner of the Editor frame to delete the document’s metadata records.

After selecting the Delete Document button, a dialog box appears, confirming that the document’s metadata will be deleted from the library.

![Figure 31 - Document will be deleted](image)

Select **OK** to remove the document’s metadata, or **Cancel** to abort the delete procedure and return to the Editor frame.

After selecting **OK**, a message is displayed confirming that the document has been deleted from the database.

![Figure 32 - Delete Successful](image)

Editing Browse Folders

To edit browse folders you must be logged into the digital document system with edit privileges. The browse editing features at this time include changing existing browse folder names and deleting browse folders.
When logged in with the appropriate authorization the browse folder will look slightly different than in the normal search interface. To the left of each browse folder there are now 2 additional icons. The first icon is the change folder name icon. The second icon is the delete folder icon.

To modify the name of a browse folder:

1. Click on the change folder name icon.
2. A window will appear with a text entry box (see figure 34). Type the new folder name into the text box.
3. Click OK to save the change. Click Cancel to abort the procedure and return to the browse window.

To delete a browse node:

1. Click on the delete folder icon.
2. A dialog box appears, confirming that you really want to delete this node. Select OK to delete the node, or Cancel to abort the delete procedure and return to the browse window.
Browse folders must be completely empty before they can be deleted. All documents that reside under a folder must be deleted and child folders that reside under a parent folder must be deleted. If a browse folder is not empty when a delete is attempted an error will occur denying the delete function.

---

**Editing Individual Documents**

Documents may be modified once they have been saved to your local drive. A document cannot be edited while it resides in ArchivalWare because of security issues. To edit an individual document:

1. Locate the document in ArchivalWare.
2. Save the document to your local or network drive.
3. Make your changes to the document
4. Upload the document to the appropriate library using the uploader. The upload feature is only available to authorized users. If you are not an authorized user please contact your system administrator.
administrator to have modified documents uploaded to your ArchivalWare library. If you are an authorized user, instructions for using the upload function can be found in the ArchivalWare Administrators Guide.

**Indexing Overview**

Indexing is the last step to be completed before the changes you made to the documents are reflected in the DDS database. The indexing process builds a file that stores information about all the words in the document library, including positional information about the words in each document. It is due to the presence of the index that ArchivalWare’s rapid searching is made possible, as ArchivalWare’s Query program searches the indexes created from the documents in your database, instead of searching the documents themselves.

For more information on the indexing process please go to the ArchivalWare Administrators guide.
Glossary

**Antonym** - A word that means the opposite of a given word; antonyms of query words will be included in a Concept search when the expansion level is set to **antonyms** or **strong antonyms**

**Boolean mode** - A query mode in which exact query terms are matched against the documents in the library (or against incoming documents, in the case of real time search agents); documents are not ranked for relevance; Boolean operators (AND, OR, NOT, WITHIN, ADJ) can be used to control matching; special operators used with individual query terms enable concept ( ! ), pattern ( ~ ), or wildcard ( *, ?, [search expression], _, @, #, ^) expansion

**Coarse grain ranking** - A query process during which documents in the library are retrieved and ranked for relevance according to how many query words are present, how many supporting (related) words are present, and the semantic distance (relationship) between the two; applies to Pattern and Concept modes, but not Boolean

**Concept expansion** - A query process that adds related terms to the original query word list; in Concept mode, all query words are concept expanded; in Pattern and Boolean modes, you can concept expand an individual word by entering an exclamation point ( ! ) after it

**Contrasted words** - A word that is dissimilar to a given word, but not as strongly as an antonym; contrasted words will be included in a Concept search when the expansion level is set to **contrasted terms**

**Database** - A collection of documents and structured fields against which your searches are performed; the documents may be organized into one or multiple libraries, and may or may not be integrated with a relational database; synonymous with the terms “library” and “document set”.

**Dictionary** - A list of meanings, where each meaning contains syntactic information and a group of words that share that meaning.

**Document Set** - A collection of documents and structured fields against which your searches are performed; the documents may be organized into one or multiple libraries, and may or may not be integrated with a relational database; synonymous with the terms “library” and database.
Exact phrase search - A specialized query in which you can search for specific words or phrases you know exist in the library; words enclosed in double quotes (" ") are not concept or pattern expanded, and must occur in the same order and proximity to be considered a hit; can be used in any query mode (Concept, Pattern, or Boolean)

Field - Document heading or other information (such as title, author, date, document type, etc.) that may be queried on separately from the body of the document

Fielded search - A specialized query in which you can limit the scope of the search by entering search terms in fields (author, title, date, etc.); the System Administrator must set up and index fields for each library before this type of query can be performed, and determine whether each field will be a statistical or Boolean filter

Fine grain ranking - A process, following coarse grain ranking, by which the retrieved documents are analyzed in order to determine the exact document rank. Applies to Pattern and Concept modes, but not Boolean

Fuzzy spelling - (see "Pattern expansion")

Grouped term search - A specialized query in which terms related to a common concept are grouped together with parentheses in order to improve search accuracy; the words within the parentheses are expanded, matched, and ranked for relevance as a group, instead of as individual words; applies to Pattern and Concept modes, but not Boolean

Hit - The words or phrases in a returned document that were matched to your search terms by the ArchivalWare program

Idiom processing - A task performed by the ArchivalWare program to identify phrases with meanings beyond their individual words (such as "real estate," "United States," and "ice cream")

Image - a single or multi page image. A photograph or a digitized picture of a page of text are both images.

Index - A file that stores information about all the words in the document library, including positional information about the words in each document; used by the ArchivalWare program during retrieval and ranking of documents

Inflected form - A variation of a word to reflect a distinction such as case, gender, number, tense, person, mood, or voice; inflected forms (such as "caught" is to "catch," or "mouse" is to "mice") will be included in a Concept search when the expansion level is set to irreg[ular] forms
Library - A collection of documents and fields which can be indexed and searched by ArchivalWare; your database may contain multiple libraries, which you may search over singly or simultaneously

Metadata - The indexed textual information that describes a document. Metadata would include the title, author, date, collection name or any other piece of descriptive information that is stored in the database and attached to an image or text document.

“Mixed” search - A specialized query in which you can mix different query modes (Concept, Pattern, and Boolean) within a single search by using special operators on individual query words

Morphological analysis - A process that removes word suffixes and changes spellings to reduce words to simpler forms found in the dictionary (such as “babies” to “baby,” or “highest” to “high”); morphological variants of query terms are always included in any search (including Boolean and exact phrase searches)

Concept mode - A query mode in which search terms are expanded to related terms via the semantic network, then matched against the documents in the library (or against incoming documents, in the case of real time search agents); each document has a rank indicating its probable relevance to your query; special operators used with individual query terms enable pattern ( ~ ) or wildcard ( *, ?, [search expression], _, @, #, ^ ) expansion, or disable expansion (“ ”)

Nested statements - A way of structuring Boolean queries that groups words or phrases within parentheses, for purposes of specifying conditions to narrow the search

Pattern expansion - A query process that adds similarly spelled terms to the original query word list; in Pattern mode, all query words are pattern expanded; in Concept and Boolean modes, you can pattern expand an individual word by entering a tilde ( ~ ) in front of it; also called “fuzzy spelling”

Pattern mode - A query mode in which search terms are expanded to words that are spelled similarly, then matched against the documents in the library (or against incoming documents, in the case of real time search agents); each document has a rank indicating its probable relevance to your query; special operators used with individual query terms enable concept ( ! ) or wildcard ( *, ?, [search expression], _, @, #, ^ ) expansion, or disable expansion (“ ”)

Precision - A measure of the text retrieval system’s ability to return only relevant documents
Query - The words or phrases you enter for a search, along with the various settings you choose, such as query mode and expansion level; queries can be stored with or without the list of returned documents, and even edited and/or re-executed

Query By Example - A specialized search in which you choose a single, highly relevant document and ask the system to find others like it

Query filter - The function performed by a field or group of fields when they are used to limit the results of a search to only those documents containing the fielded information, like author, title, etc.; fields may act as either statistical or Boolean filters

Ranking - A process performed by the query program to rate retrieved documents in order of their probable relevance to the query; applies to Pattern and Concept mode, but not Boolean

Recall - A measure of the text retrieval system’s ability to return all relevant documents

Related word - A word that is similar in meaning to a given word, but not strongly enough to be considered a synonym; related words will be included in a Concept search when the expansion level is set to related terms

Relational database (RDBMS) - A set of data organized into tables, with rules governing the relationships between the columns, rows, and tables of data; ArchivalWare provides full text searches of both structured and unstructured RDBMS data, explicitly a table-driven RDB

Semantic distance - A measure of how closely two terms are related to one another

Semantic link - Connections between words or concepts in the semantic network; used by ArchivalWare to find words and concepts related to your original search terms

Semantic network - A structure that links together related dictionary terms and concepts; each concept or word sense is a node that is linked to other nodes through word relationships (synonym, antonym, etc.)

Stemming - A less sophisticated form of morphological analysis that reduces words to their root forms

Stop words - A list of “small function” words and idioms that are not indexed and are automatically removed from queries prior to processing
**Synonym** - A word that has the same meaning as a given word; synonyms will be included in a Concept search if the expansion level is set to **synonyms** or **strong synonyms**

**Tokenizer** - A query process that divides a string of characters into words; it may include special processing for dates, phone numbers, hyphens, etc.

**Wildcard expansion** - A query process that adds terms matching a certain search pattern to the original query term list; in Concept and Boolean modes (not used in Pattern mode), you can wildcard expand an individual term by replacing part of the term’s letters, numbers, or punctuation with one or more wildcard operators ( *, ?, \[search expression\], _, @, #, ^); typically used to substitute for unknowns in the search terms or database

**Word expansion limit** - A query parameter that limits the number of expansion words that will be added to the query for each word in Concept mode, and for any word with the exclamation point (!) operator in Pattern or Boolean mode