



DATA MINING YOUR DOCUMENTS

*A practical discussion of easily-implementable methods for
using your existing system and data to
identify and contact the people you want to reach.*

Originally published: 2009-11-29
Updated: 2012-07-25

Contents

Overview

Clearing through misrepresentation and obfuscation

What are you looking for?

Start Simple

Search Items

Keyword

Title

Categorization [General]

Item-Specific Indexing

Timing Indicators

Quantitative Parameters

Behavioral Indicators

Associative Structures

Search Methods

Single Element Search

Single Value Search

Multiple Elements [multiple simultaneous criteria]

Multiple Elements [multiple combined criteria]

Selected Contacts Search

Sequential Elements [Sub-Search/drill-down]

Excluding items to narrow a search

Conclusion

DATA MINING YOUR DOCUMENTS

Overview

One of the most *valuable* assets of a company is the information it processes every day throughout its normal business activities. Making that information *useful* is a key function of your Enterprise Content Management system.

Due to the truly unique architecture of GoldVisionPro, documents and other business content are integrated with Customer Information in a single system. This is not only convenient; more importantly, it is *valuable*, as it creates the basis for using this information in your marketing efforts.

At this point you are looking beyond the primary value of document storage and retrieval in your GoldVisionPro system. You are now seeing the *additional* value of using GoldVisionPro as an easy, yet powerful, analytical tool for sophisticated search and Contact identification/Group-building.

Clearing through misrepresentation and obfuscation

Although sometimes misrepresented in negative terms, **data mining** in this sense **is simply** using all your existing information in **identifying the exact/most-appropriate people you want to reach for a specific purpose**. GoldVisionPro enables you to identify customer activity/characteristics by the extensive information it associates with their documents. The following discussion will show examples of information that can be searched, returning a broad range of document records.

From these document records, GoldVisionPro will then automatically create a GoldMine Group of the associated Contact Records for these documents, for use in your marketing campaign. [See the details on the VisionPak-Plus tool set, in the GoldVisionPro Operations manual.] **Although initially a document search, the ultimate purpose is not document retrieval, but rather identification of the associated Contacts**. As of v6.8.5, GoldVisionPro provides an Intelli-Search window to directly return Contacts based on document information, for stream-lined Group creation.

In this way, information about your documents is accessed for what it reveals about specific *activity*, details and purchase characteristics. From this previous activity you can imply potential interest in your current offering/objective. You use the document-search to derive the most appropriate candidates for the new activity you are presenting in your marketing campaign, with GoldVisionPro assembling them as a target marketing group. And as we will see, the most appropriate candidates may be the ones that DO NOT match the criteria you evaluate. Let's take a look.

What are you looking for?

While businesses know who their customers are, they don't always know who they should be contacting *now*. Constantly flooding all your contacts with unwanted/inappropriate information is more likely to be viewed as harassment than helpful! What you want to do is identify the people that would be most receptive to the information you intend to present.

Another purpose can be people you need to identify because of an upcoming or time-sensitive event. These might be business offers that have a limited "shelf-life", or the expiration of an existing item, such as a contract or legal requirement. In this case, the objective is finding the candidates "missing" the information that would exist if they had already purchased/processed/renewed.

Identifying likely candidates for timely contact requires broader knowledge of their interests, preferences and needs beyond simple contact information. Particular activity, and how recently it has occurred, are informative aspects of customer behavior/interest/need. However, tracking customer behavior can be complex and time consuming. Devising database elements for these purposes can be confusing and frankly, a lot of work to maintain.

This is where your GoldVisionPro system provides a flexible, easy-to-implement capability for capturing a wide variety of information.

Start Simple

As our objective is to make the information in and about each file useful in future GoldVisionPro searches, let's look at some of the items that can be captured, and methods, to "index" each file. At the same time, knowing how items can be retrieved helps you determine what items to store.

Keep in mind that one of the advantages of a database solution is that it allows indexing a single file by multiple criteria. This is particularly useful as **the most relevant indices may be information added/provided by the user at the time of storing the item, rather than some element already contained in the file.** Not only does this allow searching by a broader range of criteria, but it also allows for more-specific identification by requiring satisfying more than one criterion in the search.

Realizing that GoldVisionPro will allow you to easily "drill-down" through data, further refining additional criteria as you search-within-your-search-results, also allows you to more broadly design what information you will capture/use. We will explore both general and custom indexing in the following discussion.

SEARCH ITEMS

Key-Word [Contents]

A good starting point is to incorporate GoldVisionPro's AutoSummarize capability in your initial file-storage process [also available through the GoldVisionPro Link Center after a file is stored].

This automatic process:

- captures all available file text,
- determines the most likely words/phrases to be requested, and
- stores them in GoldMine.

This makes key-word searching for files available at database speed.

Data Mining Usages: Find all contacts whose files mention

- "opening your new account" in a letter
- "MBA" in a resume
- "appreciate your referral" in a thank-you

Title

File titles are best used for categorizing documents by type; for example, distinguishing between a Contract, Invoice, Letter, Order, Receipt, Resume, Sales Quote, Statement, etc. As these are items that can have multiple occurrences, you may want to add an additional identifier to easily distinguish them: a date is a common choice ["Statement – 2008-10-31"], or some business category ["Application – Life Ins"]. GoldVisionPro provides customizable drop-down lists of your usual document types, for easy user-selection when storing a file. This minimizes typing, and promotes consistency in adhering to your corporate naming conventions. It also preserves "business knowledge" from one employee group to the next, by incorporating standardization in terms.

Data Mining Usages: Find all contacts whose file titles mention

- "Order - 2008"
- "Application – Retirement Plan"
- "Claim – Auto"

Categorization [General]

Grouping files into general categories is the familiar method used in paper filing cabinets. A typical hanging folder for a customer in a file cabinet might contain the classic manila-folders for "General Correspondence", "Accounts Receivable", "Agreements", "Notes To File", "Orders", etc. Such a

categorization represents an easy selection for the user when storing an item. [It also preserves familiar terminology for users when converting from paper to electronic files.]

On retrieval, this form of indexing provides both an easy identifier for assembling a large category of files, as well as a first-of-multiple criteria for “drill-down” searches of only specific file categories [e.g., “General_Correspondence – then -> Sales” vs. “Accounts_Receivable – then -> Sales”].

Data Mining Usages: Find all contacts whose file notes mention

- “Accounts_Receivable”
- “Agreements”
- “Orders”

Item-Specific Indexing

A particularly useful feature of GoldVisionPro is its ability to add, and therefore retrieve, indexing information specific to each file. This feature allows for *unlimited indices*, providing a powerful addition to your information system. Through GoldVisionPro’s easily customizable Intelli-Entry+ feature, or free-form user input, searchable data is captured that subsequently can be used to “mine” your system.

While every type of item in the following examples would not be included with every document, GoldVisionPro makes it easy to create a list of items that pertain to a specific document type. Each of your most frequent document types can have its own list of items, and GoldVisionPro will prompt the user to complete those specific items, as the document is stored. The Intelli-Entry+ feature even allows you to prompt the user with an appropriate/calculated default entry for each item to facilitate processing. In subsequent searching, all the Intelli-Entry+ items are selectable from a drop-down list, making retrieval that much easier.

Timing Indicators

For example, tracking multiple **date-sensitive items** for your contacts can be very useful, such as

Expiring:

- certifications
- loan rate/reset
- checks
- product offerings
- contracts
- licenses
- eligibility
- legal forms
- discounts

Quantitative Parameters

Other examples are **purchase descriptions**, where targeting sales criteria can identify prospects for a new/modified/renewal sale.

- product name
- total value
- product quantity
- price/rate
- product type
- location-serviced

Behavioral Indicators

Particularly useful items are those that directly indicate **behavior and/or activity**, such as

- purchase
- sale
- residential move
- job move
- education degree
- financial default
- sport
- renewal
- cancellation
- marriage
- divorce
- separation
- travel destination
- club
- product/service inquiry
- refund
- return
- accident
- crime/violation
- child born
- death

Associative Structures

In a business environment, documents are often related to more than one contact. For example, for a legal department preparing a contract, each of the following contacts has an involvement/activity:

- Attorney Preparing the document
- “Selling” party/provider Providing the product or service
- “Buying” party/consumer Receiving the product or service
- Financing party Funding the transaction [e.g., Bank]
- Contributing party Source of info in transaction [e.g., Accountant – tax docs]

GoldVisionPro makes it easy to add an existing link to another Contact [Add New Link], carrying all the existing information to associate with the new contact. In addition, another entry can be added, indicating this record is associated with a secondary- or related-contact, rather than being the primary contact for the item. In this way, identifying all the parties related to an activity is easily accomplished in your search.

SEARCH METHODS

Again, knowing how items can be retrieved helps you determine what items to store. In this section we will discuss some of the search methods that can be used to return the records you want.

Single Element Search [single criterion]

GoldVisionPro allows you to perform a primary search based on the values or contents of 19 items in your database. These items are:

Contact fields:

- Company
- Contact
- Last Name
- City
- State
- Zip Code
- Country
- Phone 1
- Key 1
- Key 2
- Key 3
- Key 4
- Key 5

Text fields:

- Title
- Notes
- Filename

User field:

- UserID
- Inactive UserID
- User-Group

Single Value Search

GoldVisionPro allows a user to search for the same value that may be contained in more than one field. This search method specifically applies to all the Text Fields selected by the user: Title, Notes, File Name. It returns the result of a logical “OR” search, meaning the requested search-value is contained in the Title OR in the Notes OR in the File Name. Generally, this provides a larger number of records than searching by one field.

Text fields:

- Title
- Notes
- Filename

Multiple Elements [multiple simultaneous criteria]

GoldVisionPro allows users to perform a search using multiple criteria to narrow the initial search results. Even the basic application allows for two or three criteria to be designated: one Contact-field value, one Text-field value, one User-field value.

- Contact field
- Text fields
- User field

As of v6.8.5, the **Intelli-Search** window expands the multiple element search capability to allow for creating and saving search criteria based on ALL of the primary Contact1 fields plus the primary document fields, *simultaneously*. This also facilitates matching multiple values for a particular field all in the same search. As an added feature, while the user adds each element of the search criteria, GoldVisionPro displays the number of records that match each item, both individually and inclusively. This provides a real-time indication of how broadly/narrowly the search is constructed/will respond, to facilitate user modification. . [See the “GoldVisionPro V6 Operations” manual for more details/examples of this and other search capabilities.]

Multiple Elements [multiple combined criteria]

Another GoldVisionPro capability allows the user to retrieve records based on one set of criteria, and then combine those records with the result of subsequent new searches based on other criteria. This allows you to expand the number of records, while still being specific in your criteria. For example, a series of searches could be run, selecting a different specific State in each search.

Selected Contacts Search

When performing a search where it is easier to initially limit the included contacts by user-selection than defining field-values, use the List+Find option in the Advanced tab of the GoldVisionPro Link Center. This will split the search process into two searches. The first search returns all the contacts that match the selected Contact field value. In this list, the user can manually select as many of the displayed contacts as desired, to be included in the second search. The second search returns all the links for the selected contacts that match the Text and User field criteria [if applicable].

Sequential Elements [Sub-Search/drill-down]

Another approach to searching consists of refining the number of records, one criterion at a time. This “Sub-Search” or “drill-down” method repeatedly applies new search criteria to the records remaining from a previous search. This method can be applied after performing one or more of the previously mentioned methods for retrieving a set of records. It is particularly useful in repeatedly searching the Notes field, which may contain many stored custom items, which can be searched individually.

Contact fields:

- Company
- Contact

Text fields

- Title
- User
- Notes
- File Name
- File Type
- File Path

Dates

- Link Date
- File Created Date
- File Last Modified Date

Excluding items to narrow a search

Often the target audience is actually the one that does not match certain criteria. It may be that a particular item that should have been stored is either absent entirely, or does not have a current entry.

Because documents represent a many-to-one association [i.e., many documents to each Contact] there are multiple opportunities to capture items of information that represent the customer's activity/preferences/current state. These can be used not only to identify "who-bought-Item-A", but also "who-has-NOT-bought-Item-A-yet". Or "who-has-not-bought-Item-A-recently".

Today's businesses often deal with multiple different items that expire, renew or need to occur in a certain way/time. Most require monitoring to assure revenues are generated for the business. Creating separate record fields for each one is impractical, unnecessary, and overall poor data-design.

When used methodically, the Intelli-Entry+ indexing capabilities facilitate easily searching for Contacts who do not have a particular entry:

For example, annual contracts could be indexed with "Contract_ExDate=2013-01-01". The search could then locate all active customers who did not have a Intelli-Entry+ value of "Contract_ExDate=2013".

To only assess customers that HAVE a contract but HAVE NOT renewed, the search could INCLUDE "Contract_ExDate=2012" and EXCLUDE "Contract_ExDate=2013".

Even the basic application allows the user to enter Sub-Search criteria and then return only those records that do *not* match/contain the search value. [In the GoldVisionPro Link Center, this feature is toggled by pressing Alt+right-click on the Sub-Search button, activating the red "exclude" indicator above the button.]

NOTE: When GoldVisionPro creates a Group from the selected document records, it applies a "unique contact" logic to assure that a contact record only appears once in the created Group.

There are additional methods of constructing specific search-syntax in GoldVisionPro, following popular Boolean-syntax. These are covered in more depth in the Operations manual for GoldVisionPro.

As of v6.8.5, the **Intelli-Search** window greatly expands this capability to enable more specific exclusionary searching, even with multiple criteria. The above example could be executed in a single search, even with additional criteria for refining the result. The results can be used to create a marketing group with one-click, and then accessed directly in GoldMine for processing. As previously mentioned, the Intelli-Search criteria can be saved for re-use on an ongoing basis, as part of a periodic process. [See the "GoldVisionPro V6 Operations" manual for more details/examples of this and other search capabilities.]

Conclusion

In summary, a business can expand the use of its existing information system to gain additional value from its data. Documents are often a part of important activity, and can be used to identify that activity as a historical data-store. GoldVisionPro enables a business to not only associate valuable detail along with normal file storage, but also retrieve and actively use that information in marketing efforts, as well as in normal file retrieval.