



INTELLIGENT DATA CLEANSING

IT'S ACCURATE DATA THAT ENABLES SUCCESSFUL BUSINESS DECISIONS

Consider for a moment how new contact records are added to your database, the number of touch points and the propensity for error. Now multiply this by the number of new records being created each day, and the number of individuals in your company looking up, modifying, updating or adding new contact records to your database or CRM system.

Now consider a solution: the matchIT API.

The matchIT API is a customizable plug-in component for multiple hardware platforms, for state of the art fuzzy matching, data formatting and data cleansing. Its most common uses are duplicate prevention, inquiry, batch deduplication and merge/purge, and it can be built into your system at any stage of the process.

matchIT API



matchIT is an effective tool for:

- point of capture
- web entry
- search/lookup
- batch

N.B. Fields in red are compulsory fields

Score	Contact	Company	Zip/Pos...	Address1	Addre...	Ad...	Address4
80	Robert Breigh	Siebel Systems, Inc.	94404	2207 Bridgepointe Parkway			San Mateo

Example:
Duplicate detection at record entry

helpIT systems is...

...recognized as a leading expert in the field of data cleansing. Established over 15 years ago, helpIT systems software is used by more than 1000 companies worldwide. Our diverse client list, which ranges from Fortune 500 corporations to forward-thinking start-ups, relies on helpIT systems for their data quality management needs.



The Problem

Data quality - duplicate records, incomplete records, the lack of standardization, the accessibility to data, the inability to resolve data across disparate sources, and their impact on data quality. Incomplete and duplicate contact records will infiltrate your systems from every touch point and from every person accessing your data. The reason is simply the failure to identify potential duplicate records at the point of capture, and the inability to aggregate new data as it is introduced.

The fact is most CRM, ERP, web and data warehouse systems fail and in many cases are simply incapable of addressing the problem of data quality. There are a myriad of touch points to both individual systems and across an entire enterprise, making the control of data extremely difficult and the task of creating a single accurate view of your customer seem insurmountable.

Attaining a Single Accurate View of your Customer

Whether your data quality needs stem from Business Intelligence (BI), data warehousing, Customer Data Integration (CDI), direct marketing or merging data residing across the company, the process will all start with the same requirement – a single accurate view of customer data.

The matchIT API is your key component in developing a broad systematic process to unify customer data. It can be called to resolve a single data quality need, or as an integral component of a larger data management process for tackling not only the quality but also the accessibility of data. The matchIT API is designed for deployment in both batch and real time processes to address a multitude of data quality issues.

With the matchIT API plugged into your systems, you can:

Identify Prevent Standardize Enhance Enrich

- prevent the addition of duplicate records
- remove existing duplicate records
- match records across systems
- intelligently search/lookup records in a database
- properly case names and addresses
- populate salutations/addressee fields from freeform names
- improve structure and relocate data elements to correct fields
- detect gender from forenames
- screen for garbage and offensive entries
- match phonetic, miskeyed and abbreviated variations
- standardize titles and qualifications
- intelligently split names into component parts

The case of Bill Dayton vs. William Deighton

Your system contains separate records for:

- Bill Dayton, Cable News Network
- William Deighton Jr, CNN

The matchIT API knows that 'Dayton' sounds like 'Deighton', and recognizes that Bill is short for William. As a result, matchIT API flags these entries as potential duplicates.

The value of high accuracy matching and data cleansing:

The benefits from having the matchIT API fully deployed throughout your company are incalculable. They range from: increased efficiency in the call center; improvement in direct marketing performance, usable data aggregated from multiple sources, and business & financial reporting that is based on data that is far more accurate.

The matchIT API will finally allow you to gain control of data flowing through your enterprise. In general your data will be trusted.

Point of Capture Duplicate Prevention & Data Formatting

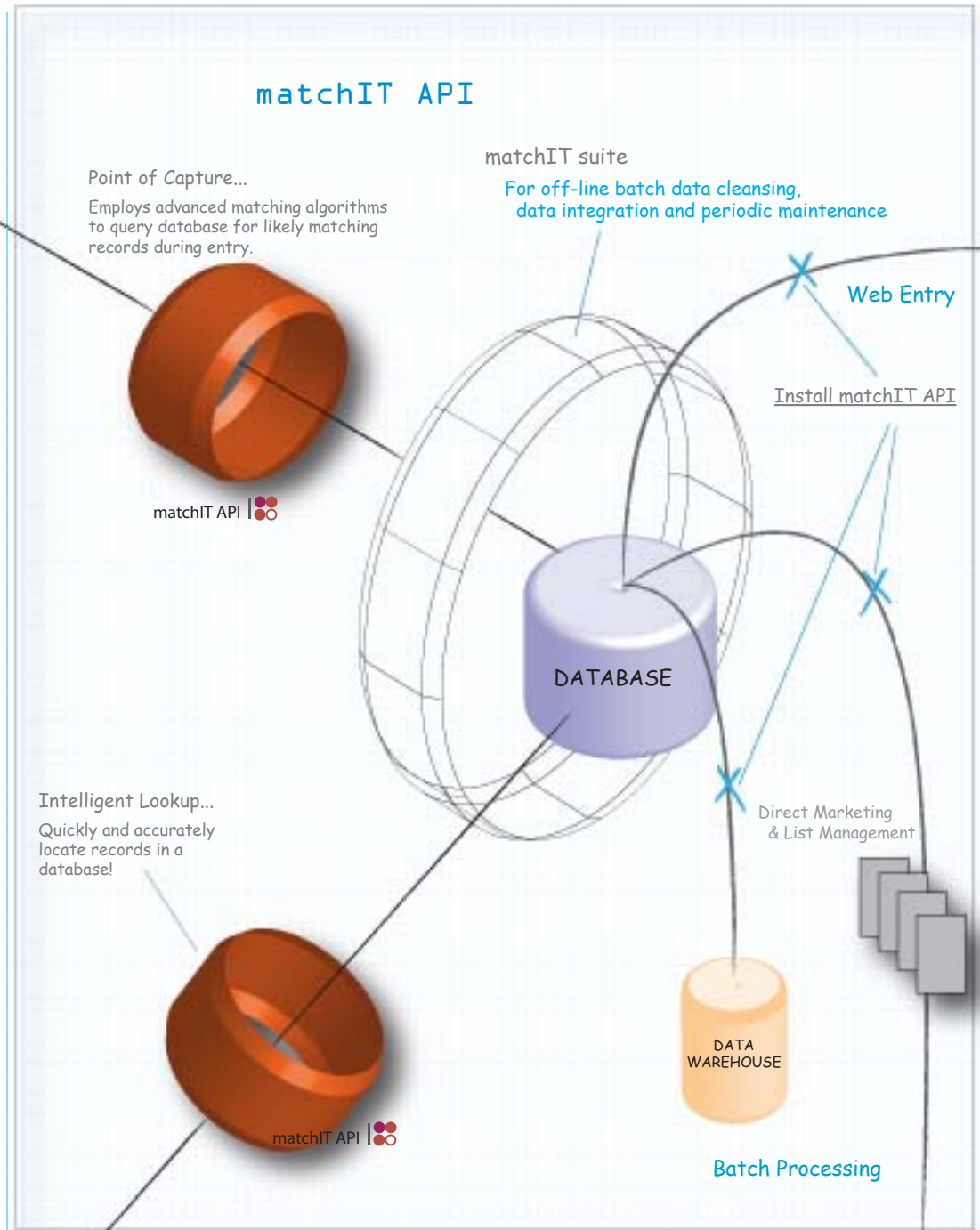
The accurate capture of names and addresses is a time consuming, mistake filled process that leads to duplicate records eroding the quality and accuracy of your data. The matchIT API can eliminate duplicate contact entries at the point of entry by performing a fuzzy matching lookup as the record is keyed to check whether the record already exists.

The matchIT API intelligently splits and standardizes the customer record to improve structure and simplify loading to the database.

Further improving data entry, is an available toolkit for the matchIT API to incorporate rapid addressing. This technology is designed to greatly reduce the number of keystrokes required to complete an address (typically requiring fewer than 15 keystrokes) and validate the address against the U.S. Postal Service® (USPS®) data file.

Intelligent Lookup

matchIT API can be plugged in at each workstation, enabling users to retrieve a record by using advanced match routines on any element of name, company, postal code, street, and city/state. This search will identify records that match the inquiry, even if they are misspelled, miskeyed, or simply entered differently. If there are several matching records, the matchIT API enables you to see the best matches first. Just as importantly, it ensures that you do not have to view records which are obviously not matches. The result is greater efficiency and accuracy.



Supplied Input	
Name	BILL DAYTON

Returned Records							
Name	company	address1	address2	address3	state	zip	telephone
Bill Dayten Jr	Palmer Air Charters	7350 Airport Road #106	Wilmington		DE	19801-4273	
Mr W R Dayton	Big Oak Ranch	2330 Diana Ave		Morgan Hill	CA	95037	408 778 5479
William R Deighton	McDonnell Douglas Inc	100 Washington Ave S	Ste 1248	Minneapolis	MN	55401-2121	612-845-2172



Web Entry

matchIT API can help you to make the most of your web data by screening for garbage/offensive entries. You can specify your own parameters for this, and instruct the matchIT API on how to deal with such data.

Not only can the matchIT API work directly with web data as it is entered, but it allows you to simplify web forms. It can correctly case the data, expand abbreviations, split it into the correct fields and create salutations to personalize the dialog. For web data entry, matchIT API is suitable for both Internet and corporate Intranet users.

You can reject or flag records without a full and proper name or sufficient address details, or any which contain inappropriate terms (such as expletives). matchIT API will enable you to check that records don't already exist, either in real time or later on as an off-line process.

With matchIT API, the data you collect will be of a much higher quality.

Batch Processing

The matchIT API allows you to automatically clean and convert data from a variety of different sources. The same routines that the matchIT API uses in data capture can be used to standardize new batch data and match it effectively against the main database.

As well as the usual phonetic and other keying errors, the data is usually structured differently across different systems and keyed using different standards. This can compromise the quality of your data warehouse or marketing database, to the extent that it seriously undermines its value, use, and organizational acceptance. The matchIT API, unlike most competing products, deals effectively with poorly structured data.

Contact helpIT systems today

Call helpIT systems at 866.628.2448,
or go to www.helpit.com.

Direct Marketing

The matchIT API allows you to automatically clean and convert data from a variety of different sources. The same routines that the matchIT API uses in data capture can be used to standardize new batch data and match it effectively against the main database. You perform the database updates in your program that calls the matchIT API, so you remain in complete control.

With the matchIT API, you can:

- discard duplicates
- create subsidiary records
- export matching records to a separate file
- place flags within the global file

CDI - Customer Data Integration as defined by Gartner Research:

"the combination of the technology, processes and services needed to create and maintain an accurate, timely and complete view of the customer across multiple channels, business lines and enterprises, where there are multiple sources of customer data in multiple application systems and databases."

Several Matching Techniques - One API

The matchIT API encompasses several matching algorithms and methodologies to accurately locate and match records. The matchIT API has the ability to locate potential duplicate records with misspellings typically caused by keying errors and mishearing during data entry. It also can identify suspect records even if the information is incomplete or invalid.

Phonetic

-to match "sounds like" names such as Dayton and Deighton

Acronym & Initial Matching

-to match inconsistencies like John and J., "Cable News Networks" and "CNN"

Character Occurrence

-to match keying errors & transpositions like "Wilson & Wislon" and 95037 and 93507

Table Based

-to match names such as Bill and William

Element Matching

-to match names with elements missing or reversed, e.g., Mr. J R Smith, John Smith and Smith John

Data Warehousing/ CRM & ERP Management

Often, companies have to deal with large quantities of name and address data in different files, systems and departments across the organization. With the matchIT API you can aggregate data from the multiple source systems in which it currently resides and intelligently merge and consolidate data from these sources. Establishing a single "truth" in data across the enterprise

Typical uses:

- Loading batches of new data into a data warehouse
- Merging data from different corporate systems
- Conversion of data from old systems
- Company acquisitions & mergers

matchIT API Overview

The matchIT API is a component written by helpIT systems for state of the art fuzzy matching, data formatting and data cleansing – amongst its most common uses are duplicate prevention, inquiry, deduplication and merge/purge. The matchIT API splits and cases names and addresses, generates match keys, and grades matching records.

The ActiveX components use Microsoft's ATL (Active Template Library) to provide a compact and efficient solution to the problems of data quality and duplication on any Windows based system. On Unix and Linux systems, the matchIT API is available as shared libraries, enabling the same intelligent logic to be applied to meet the data quality challenge.

The matchIT API is designed to enable the developer to easily implement it within your application. Properties and settings are accessed via an intuitive and hierarchical interface using simple object-dependent methods.

Sample applications are available in toolkits written in a variety of languages for both interactive and batch processing. These provide a suite of interrelated classes that can be easily reused in your own Windows application, with little or no modification.

Each toolkit:

- provides a simplified higher level interface to the matchIT API
- provides classes for interacting with a database, using ADO or ADO.NET to provide functionality for data retrieval, insertion, update, and deletion
- supports relational tables or a single table
- provides a shared interface to different database systems, such as SQL Server, MySQL, and Access
- provides classes that implement data entry, data merge, and address lookup dialogs
- optionally integrates with rapid addressing / address verification software
- is scalable for volume of records and number of users.

The C++ toolkit:

- uses ADO
- is built using the lightweight and standard STL and ATL
- uses ATL for dialogs
- is compatible with Visual Studio 6.0 and up.

The VB.NET and C#.NET toolkits:

- use ADO.NET
- comprise pure .NET classes (no unsafe code)
- use Windows Forms for dialogs
- are compatible with the .NET Framework 1.1 (Visual Studio .NET 2003) and up.

Data cleansing is the first step, and most critical, in a Business Intelligence or Data Warehouse project - yet easily the most underestimated. Internal processes, manual labor, queries and scripts will never return anywhere near the results helpIT systems software delivers.

The net result in undervaluing data cleansing is: cost overruns, data inconsistencies, data quality problems and, in many instances, outright failure of the Business Intelligence project.