

## Eliminate Duplicate Records & Reduce Waste

Achieve a single, comprehensive view of your customers when you purge duplicate records from your database. MatchUp Object's 12 matching algorithms dedupes your files so you can improve sales, marketing, and analytics.

## MatchUp® Object

### Identify & Dedupe Duplicate Records

For nearly three decades, Melissa Data has been a world leader in Contact Data Verification partnering with numerous Fortune 500 companies. Our domain expertise makes MatchUp® a powerful data and address deduplication (merge/purge) API featuring multiple ways of identifying duplicate records. MatchUp dedupes files in different formats with different name, address, and city/state/ZIP™ structures, without taking the time to manipulate files and get them into the same format.

### Fuzzy Algorithms

Fuzzy matching is an advanced mathematical process that identifies “non-exact matching” duplicate records that are difficult to uncover. MatchUp employs over a dozen different algorithms in the fuzzy matching process:

- **Phonetic Matching.** Utilizing the phonetic algorithm for fuzzy matching – also known as the “double metaphone” method – detects “alike-sounding” relationships between words.
- **N-gram or Q-gram-based Algorithms.** The linear n-gram or q-gram-based algorithm models are primarily used in statistical natural language processing. An n-gram is a subsequence of "n items" from a given sequence – which can be phonemes, syllables, letters, words, or base pairs.
- **Jaro-Winkler Algorithm.** The Jaro-Winkler distance is a measure of similarity between two strings. It is mainly used in the area of record linkage for duplicate detection.
- **MD Keyboard.** An algorithm developed by Melissa Data which counts keyboarding mis-hits with a weighted penalty based on the distance of the mis-hit and assigns a percentage of similarity between the compared strings.
- **Containment.** Matches when one record's component is contained in another record. For example, “Smith” is contained in “Smithfield.”

### Benefits:

- Reduce costs associated with duplicate records
- Gain a more accurate, single view of your customer data
- Identify multiple records at the same address
- Fast processing, 10-50 million records per hour
- Easy-to-Use -- eliminates "rules-based" matching

## Try MatchUp for Free Today!

MatchUp was built with contact- domain specific knowledge. We have leveraged our nearly three decades of experience working with addresses and names to build in contact-specific rules so you don't have to.

>> GET A FREE TRIAL: [www.MelissaData.com/match](http://www.MelissaData.com/match)



### Fuzzy Algorithms - Continued

- **Frequency.** Matches the characters in one record's component to the characters in another without any regard to the sequence. For example "abcdef" would match "badcfe."
- **Fast Near.** A typographical matching algorithm. It works best in matching words that don't match because of a few typographical errors. Exactly how many errors is specified on a scale from 1 to 4 (1 being the tightest).
- **Accurate Near.** This is a typographical matching algorithm. The Accurate Near algorithm produces better results than the Fast Near algorithm, but is slower.
- **Frequency Near.** Similar to Frequency matching except that the algorithm lets you specify how many characters may be different between components.
- **Vowels Only.** Only vowels are compared. Consonants are removed from the search criteria.
- **Consonants Only.** Only consonants are compared. Vowels are removed from the search criteria.
- **Alphas Only.** Only alphabetic characters are compared.
- **Numerics Only.** Only numeric characters are compared. Decimals and signs are considered numeric.

### Matching Scenarios

- **Householding.** Group data by pre-defined criteria (i.e., all members of a household count as one group)
- **Proximity Matching.** Group records that are geographically close together.
- **List Intersection/Suppression.** Find all the common data between multiple lists & find data that is unique to each individual list.
- **Flexible Matching.** Handle over 35 pre-defined data types (like address, email, SSN, etc.) and allows up to 16 cascading rule sets.
- **Point Of Entry/Batch.** Real-time monitoring of incoming records allowing you to accept unique records while rejecting duplicate records.
- **Speed/Scalability.** MatchUp is multi-thread safe and scalable for data sets of any size.

### Domain Specific Knowledge

MatchUp relies on Melissa Data's years of experience working with numerous contact data idiosyncrasies. This work has helped us build rules and logic to handle such things as Address Obscurities; Nicknames and Abbreviations; and Different Formatting.

### Features:

- **Multiplatform (Windows, Linux, Solaris, AIX, HPUX)**
- **Process files with different field structures**
- **Native, Java, Ruby, Perl, Python, PHP integration**
- **Assign priority for eliminating duplicates**
- **Split name, address, and city/state/Zip® fields**

### Three Ways to Dedupe Your Contact Data

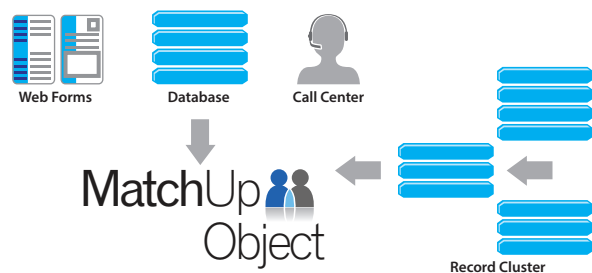
**1. Read/Write Deduping** – Compares records and one or more databases at once. Each unique group will have one record that receives an “output” status; the other matching records receive a “duplicates” status. Ideal for batch merge/purge/suppressing existing data.



**2. Incremental Deduping** – Compares each record as they come in and against a database of already processed records. Ideal for real-time data entry. If the incoming record is new, it can be added to the existing database Match.



**3. Hybrid Deduping** – Gives you the flexibility to customize the process if your environment requires internal key storage or comparisons to smaller clusters of records. Ideal for real-time data entry or batch processing entire lists.



### Matchcode Editor

Matchup also features a Matchcode Editor, a separate Graphical User Interface tool that allows users to maintain a personal matchcode database. Once you develop custom matchcodes that work well with your data, you can easily retrieve them for continuous use.

### System Requirements

**Multiplatform API** - Library available for Windows 2003/2008 7/8.x; Red Hat; Solaris; SPARC; AIX; Power, rs/6000, PPC; HPUX: PA-RISC; and Itanium.

### About Melissa Data Corp.

Since 1985, Melissa Data has been a leading provider of data quality and address management solutions. Melissa Data's data quality software, Cloud services and data integration components verify, standardize, consolidate, enhance and update U.S. and global contact data, including addresses, phone numbers and email addresses for improved communications and ROI. More than 10,000 companies worldwide rely on Melissa Data to gain and maintain a single, accurate and trusted view of critical information assets.

Melissa Data Corp.  
22382 Avenida Empresa  
RSM, CA 92688-2112

1-800-MELISSA  
[www.MelissaData.com](http://www.MelissaData.com)

### Find us on:

Twitter: @melissadata  
Facebook.com/melissadata  
[linkedin.com/company/melissa-data](https://www.linkedin.com/company/melissa-data)