



CRM DEDUPLICA

Deduplica e Normalizzazione dei clienti doppi ... o simili

Validità: Giugno 2019

Questa pubblicazione è puramente informativa.
impresoft non offre alcuna garanzia, esplicita od implicita, sul contenuto.
I marchi e le denominazioni sono di proprietà delle rispettive società.

SOMMARIO

1. INTRODUZIONE	3
1.1 A COSA SERVE LA DEDUPLICA?	3
1.2 PRINCIPALI FUNZIONALITÀ	3
1.3 INSTALLAZIONE	4
1.4 AVVIO DEL MODULO DI DEDUPLICA	4
1.5 PUNTI DI FORZA	6
2. ESECUZIONE DELLA DEDUPLICA	7
2.1 CRITERIO DI RICERCA DEI DUPLICATI	7
2.2 IMPOSTAZIONE DI CRITERIO DI FILTRO	8
2.3 SELEZIONE DEI DUPLICATI DA ELABORARE	8
2.4 CRITERIO DI DEDUPLICA	9
2.5 DEDUPLICA DELLE INFORMAZIONI DI BACKOFFICE	10
2.6 RICERCA DEDUPLICATI	11
2.7 ESECUZIONE DEDUPLICATA MANUALE O AUTOMATICA	13
2.8 RISULTATO FINALE	15

**Il modulo Deduplica e Normalizza
individua e risolve facilmente
il problema dei clienti doppi ...o simili !**

1. Introduzione

1.1 A cosa serve la Deduplica?

All'interno di una **piattaforma CRM** un elemento fondamentale consiste proprio nella **pulizia dei dati e delle anagrafiche** che in esso vengono gestite.

L'efficienza e l'efficacia delle attività di Marketing che **il CRM** può sviluppare sono direttamente collegate infatti con il livello di correttezza di queste informazioni. La possibilità di gestire in modo certo e preciso ogni singola opportunità (prospect) garantisce di evitare le classiche "figuracce", come invii multipli di telefonate, email, fax, ecc... evitando tra l'altro anche spreco di tempo e denaro.

Il processo di "**Deduplica**" è quindi un'attività essenziale e fondamentale all'interno di una piattaforma CRM professionale.

Questo permette di analizzare i dati salienti delle anagrafiche e di evidenziare, ed eventualmente eliminare, i prospect doppi, o anche soltanto sospettati tali.

Con il CRM è possibile garantire **la perfetta pulizia dell'archivio di marketing** individuando tutti i clienti/prospect doppi o simili e quindi intervenire per **sistamarli automaticamente**.

1.2 Principali funzionalità

La **deduplica viene applicata all'intero sistema CRM**, integrando tutte le informazioni anagrafiche, i contatti, gli ordini, gli appuntamenti, ... presenti sulle anagrafiche duplicate sull'anagrafica corretta rimasta.

- Analizza i dati anagrafici secondo diversi criteri
- Determina i doppi certi o sospetti in base alle chiavi certe o di prossimità scelte
- Selezione dei gruppi di duplicati certi o sospetti da elaborare
- Aggiorna automaticamente il database marketing mantenendo i clienti in base ai criteri impostati oppure in base a selezioni manuali
- Le anagrafiche individuate come sospette durante una fase di analisi possono essere anche "segnate" come diverse, in modo da istruire il sistema di deduplica affinché le consideri tali nelle successive analisi.

Il modulo di **DEDUPICA AUTOMATICA** permette di verificare se all'interno del database sono presenti dei **clienti doppi**. Il modulo è in grado di riconoscerli, spostare tutti i dati dalle schede dei doppi sulla sola scheda che si decide di mantenere ed eliminare le anagrafiche "svuotate".

NOTA BENE: si consiglia di eseguire la deduplica quando nessuno è collegato al sistema CRM e senza interrompere il processo. Si consiglia ad esempio di lanciare il processo alla sera dopo il termine del lavoro o durante il weekend. E' vivamente consigliata la **copia di backup** degli archivi prima di eseguire la procedura.

1.3 Installazione

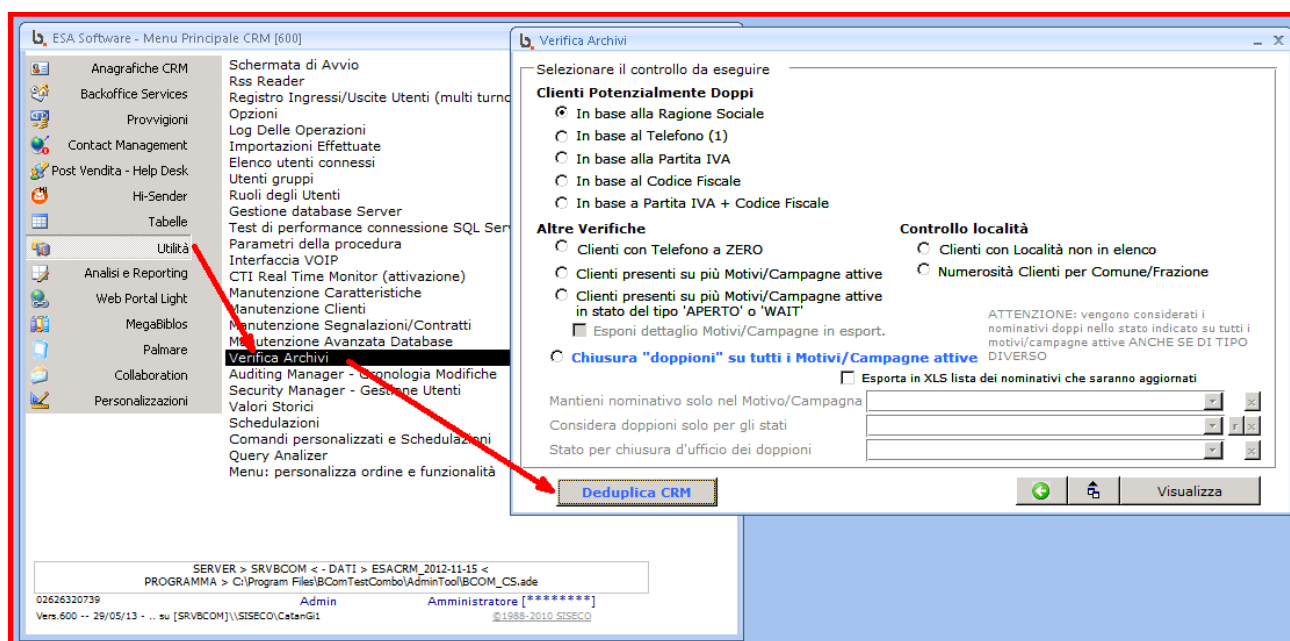
Il modulo di **DEDUPLICA AUTOMATICA** può essere **scaricato gratuitamente** da tutti i clienti di SISECO tramite:

- la pagina: <http://support.siseco.com/pro/gat/clienti/aggiornamenti.htm>
- dal link: <http://www.siseco.it/download/SetupCRMDeduplica.exe> (4Mb ~)

ed installato su una postazione (es. quella dell'amministratore o del supervisore) in cui è già presente **CRM C/S**: il client residente su PC per compiere le operazioni di impostazioni base e le manutenzioni complesse.

E' necessario che sulla stessa postazione siano anche installati **Windows Installer 3.0** e **Framework .Net 2.0**, liberamente reperibili sul sito Microsoft.

Una volta installato il modulo di **DEDUPLICA AUTOMATICA**, la funzionalità si troverà poi nel menù **Utilità / Verifica Archivi di CRM C/S**.



Il modulo prevede un wizard guidato che di volta in volta chiede al cliente i parametri per poter eseguire la deduplica (es. quale anagrafica tenere, quella più recente o quella più vecchia, oppure quella con più contratti ecc).

Una volta stabilite le impostazioni di lavorazione il sistema CRM mantiene una sola scheda anagrafica, scelta automaticamente o manualmente e vi riporta tutti i dati precedentemente collegati ai "doppioni", cancellando poi le anagrafiche in eccesso.

E' anche possibile riportare sull'unica anagrafica rimanente tutte le informazioni relative all'indirizzo, cellulare, telefoni ed indirizzi email degli altri doppioni eliminati.

1.4 Avvio del modulo di Deduplica

Il modulo di **DEDUPLICA** vi guida alla scelta dei nomi da normalizzare con un semplice wizard, composto da diverse pagine. Vediamo insieme un esempio del processo di deduplica.

Per eseguire la deduplica procedere come segue:

- Avviare **CRM C/S** (il client residente su PC)
- Selezionare **Utilità / Verifica Archivi di CRM**
- Nella maschera Verifica Archivi premere **“DEDUPLICA CRM”** e si aprirà la seguente maschera che già presenta le potenzialità del modulo.



Da questa pagina è possibile selezionare se operare su database CRM in formato GAT oppure BCOM.

Il Wizard evidenzia in basso a sinistra:

- la **versione del motore** di deduplica
- il server e il **database** su cui sta operando
- il **numero totale di anagrafiche** presenti prima della deduplica

Selezionato il formato di database è possibile procedere in modo guidato con le altre pagine.

1.5 Punti di forza

Le novità ed i miglioramenti della versione attualmente in linea sono:

- **merge delle informazioni dei duplicati** sul record risultante
- selezione del record da mantenere in base all'ultima modifica e in caso di uguaglianza all'IDCliente maggiore
- visualizzazione di tutti i campi dell'anagrafica Clienti nella finestra di "Dettagli"

Uno dei punti di forza del modulo DEDUPLICA è la possibilità di poter modificare i dati presenti sul database anche mediante **un file esterno di correzione** (un file ad esempio di formato TXT). Il file deve riportare tutti i duplicati dello stesso cliente su un'unica riga separati da un carattere a scelta (per default, punto e virgola, ma può essere indicato in fase di importazione). Ogni riga è un cliente differente.

Selezionandolo e proseguendo con la procedura, alla fine viene presentata questa maschera dove indicare il file da utilizzare:

Importazione lista duplicati da file

File con la lista dei duplicati

Nome del file: C:\Documents and Settings\lorenro1\Desktop\calls.txt Sfoggia

Separatore dei campi

☐ Tabulazione ☐ Punto e virgola ☒ Virgola ☐ Spazio ☐ Altro

Anteprima dei dati:

LAN | ISDN | 218 | 3482878361 | 29/11/08 19.49.43 | 00.00.12 | 218 | 3482878361 | 3482878361 | 202084

OK Annulla

2. Esecuzione della Deduplica

2.1 Criterio di ricerca dei duplicati

Dalla seguente maschera si impostano i criteri di ricerca dei nomi duplicati. E' possibile inserire dei filtri per la ricerca, sia semplici (ragione sociale, o telefono o codice fiscale, ecc.) che avanzati (campi diversi combinabili tra loro, es. città + telefono, in questo caso le condizioni si sommeranno).

ver. 1.2.2.5
GAT-DEMO482
Clienti prima della deduplica:
2.365

Per esempio selezionando come indicato nell'immagine si ricercheranno i "doppioni, simili" in base al campo anagrafico "Ragione Sociale".

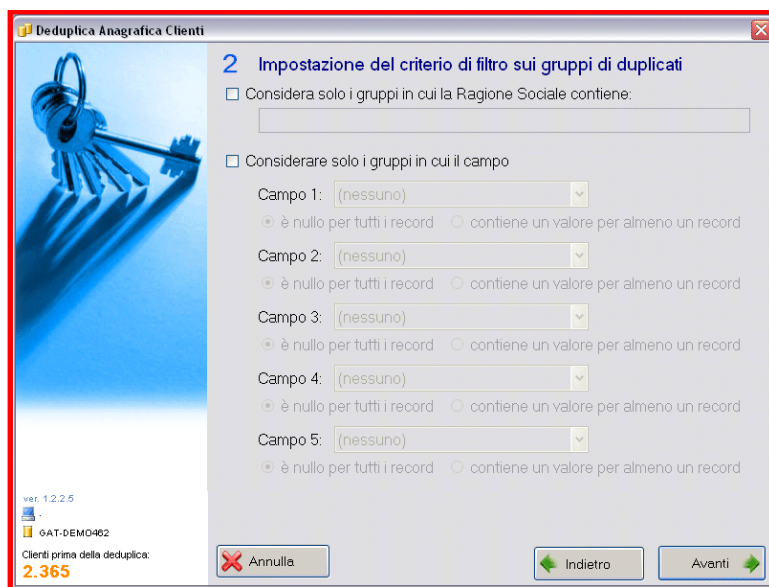
E' molto importante indicare anche la modalità di corrispondenza che potrà essere:

- **ESATTA** - ovvero la chiave di deduplica dovrà essere "identica" tra i vari doppioni affinché questi vengano segnalati
- **SOMIGLIANTE** – in questo caso la ricerca dei duplicati avviene mediante un concetto di similitudine implementato da una funzione in C#, derivata dagli studi di Siderite Zackwehdex, denominata: "Super Fast and Accurate string distance algorithm: Sift3". La stessa funzione per comodità è disponibile anche come funzione sul Database (dbo.GatSimilFx). Questa funzione accetta:
 - Un parametro di **somiglianza**, livello di confidenza per considerare "simili" i clienti, il cui valore predefinito suggerito è pari a 75. Il valore massimo è 99, che tendenzialmente troverà soltanto i clienti "praticamente identici". Il valore minimo 1, che però è privo di utilità pratica.
 - Un parametro di **precisione** di lavoro il cui valore predefinito suggerito è pari a 100. Per la funzione di deduplica occorre usare questo valore.

Utilizzando il criterio di ricerca per SIMILITUDINE dovrà compiere controlli molto più complessi e l'operazione potrebbe richiedere tempi molto più lunghi.

2.2 Impostazione di criterio di filtro

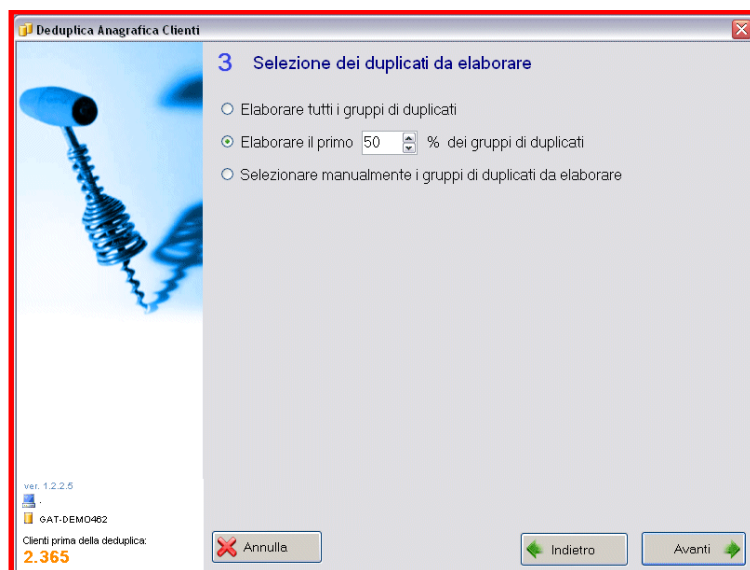
Si imposta un criterio di filtro “speciale” sui gruppi duplicati, es. considerare soltanto i gruppi dove compare un dato preciso.



Questo criterio di filtro limita le attività su un sottoinsieme, agevolando la velocità complessiva di elaborazione.

2.3 Selezione dei duplicati da elaborare

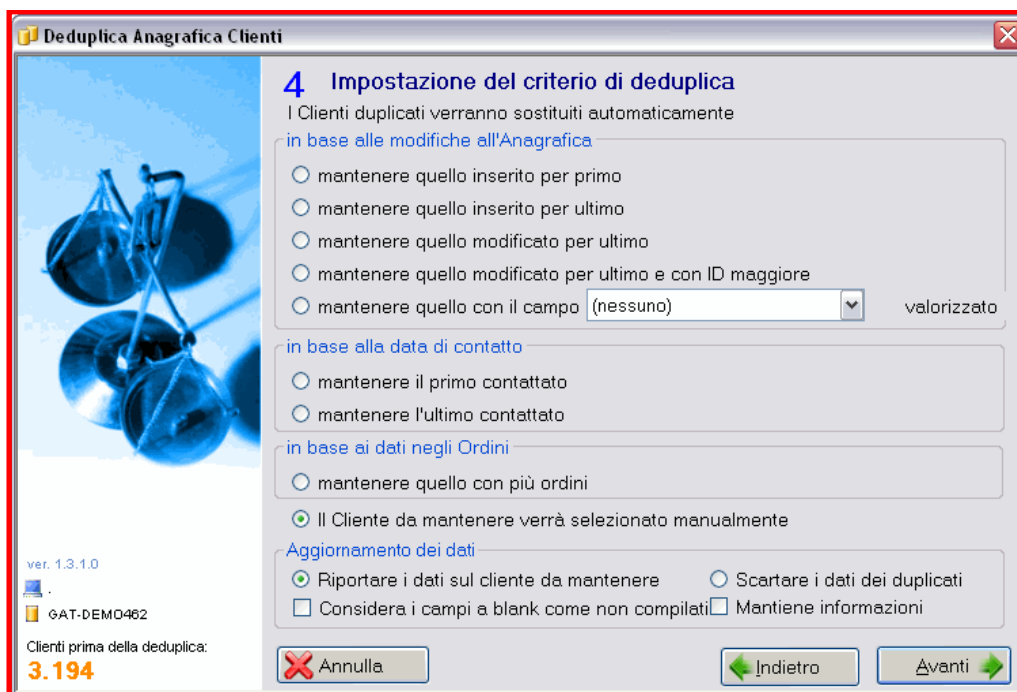
La pagina tre chiede come elaborare i duplicati trovati: se elaborarli tutti, solo una parte dei gruppi trovati o selezionarli manualmente.



NB: Nel caso di ricerca per similitudine, al fine di evitare ragionamenti automatici “dubbi” viene automaticamente imposta la selezione manuale dei dopponi individuati.

2.4 Criterio di deduplica

Serve per impostare il criterio di deduplica. E' possibile mantenere il nominativo che ha un determinato campo valorizzato, oppure aggiornare i dati sul cliente da mantenere o di scartarne i duplicati:

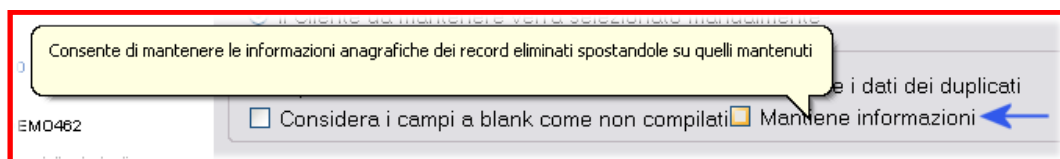


The screenshot shows a window titled "Deduplica Anagrafica Clienti" with a version number of 1.3.1.0 and a demo ID of GAT-DEMO462. It displays 3,194 clients before deduplication. The main section is "4 Impostazione del criterio di deduplica", which explains that duplicate clients will be automatically replaced based on certain criteria. The criteria are grouped into three sections: "in base alle modifiche all'Anagrafica" (with options to keep the first, last, or most modified record, or the most modified with the highest ID), "in base alla data di contatto" (with options to keep the first or last contact), and "in base ai dati negli Ordini" (with options to keep the record with the most orders or to select manually). There is also an "Aggiornamento dei dati" section with options to report data to the record to be kept, discard duplicate data, consider blank fields as non-completed, and a new option "Mantiene informazioni". At the bottom are buttons for "Annulla", "Indietro", and "Avanti".

La modalità di aggiornamento è molto importante per scegliere come verrà eventualmente aggiornato il nominativo che verrà mantenuto come definitivo.

E' possibile ad esempio riportare su questo nominativo tutte le informazioni rilevate sulle altre anagrafiche qualora i campi di destinazione siano nulli (o anche vuoti se viene selezionata la relativa opzione).

La nuova opzione "Mantiene Informazioni" consente di mantenere le informazioni anagrafiche dei record eliminati spostandole su quelli mantenuti riutilizzando campi liberi compatibili.



This close-up shows the "Mantiene informazioni" checkbox, which is currently unchecked. A yellow tooltip bubble points to it, containing the text: "Consente di mantenere le informazioni anagrafiche dei record eliminati spostandole su quelli mantenuti". Other visible options include "Considera i campi a blank come non compilati" (unchecked) and "Scartare i dati dei duplicati" (unchecked).

Es. se il cliente da eliminare ha compilato l'indirizzo EMAIL1 e il cliente che verrà mantenuto anche al termine il valore EMAIL1 del cliente eliminato finirà nel campo EMAIL2 del cliente rimasto. Questo ragionamento viene applicato per i campi: indirizzo, email, telefoni e cellulare.

2.5 Deduplica delle informazioni di Backoffice

Si tratta di un passo opzionale. Completa la deduplica delle anagrafiche e delle entità collegate: contatti, appuntamenti, ... precedente con la possibilità di deduplicare anche gli ordini dei clienti.

Con la seguente maschera è possibile indicare come trattare gli ordini collegati ai clienti.

Deduplica Anagrafica Clienti

Deduplica degli ordini collegati ai clienti

☒ Deduplicare gli ordini eventualmente associati ai clienti
Gli articoli presenti negli ordini relativi ai clienti duplicati verranno a loro volta deduplicati in modo che lo stesso articolo non compaia in più di un

Considerare diversi gli ordini con date differenti per più di giorni

Criteri di scelta dell'ordine da mantenere

☒ L'ordine associato a una campagna
Verrà mantenuto di preferenza l'ordine collegato a una campagna, altrimenti verrà mantenuto l'ordine modificato per ultimo.

☐ L'ordine modificato per ultimo

Azioni di deduplica dei dettagli ordini

☐ Riportare i dettagli degli ordini eliminati sull'ordine mantenuto
Tutte le righe degli ordini con dettagli duplicati da eliminare verranno collegate all'ordine che verrà mantenuto e l'ordine sarà eliminato.

☒ Eliminare i dettagli duplicati, ma non compattare gli ordini
Verranno eliminate le righe con gli articoli duplicati, ma non l'ordine a cui erano collegate se questo contiene altre righe.

ver. 1.3.1.0
GAT-DEMO462
Clienti prima della deduplica:
3.194

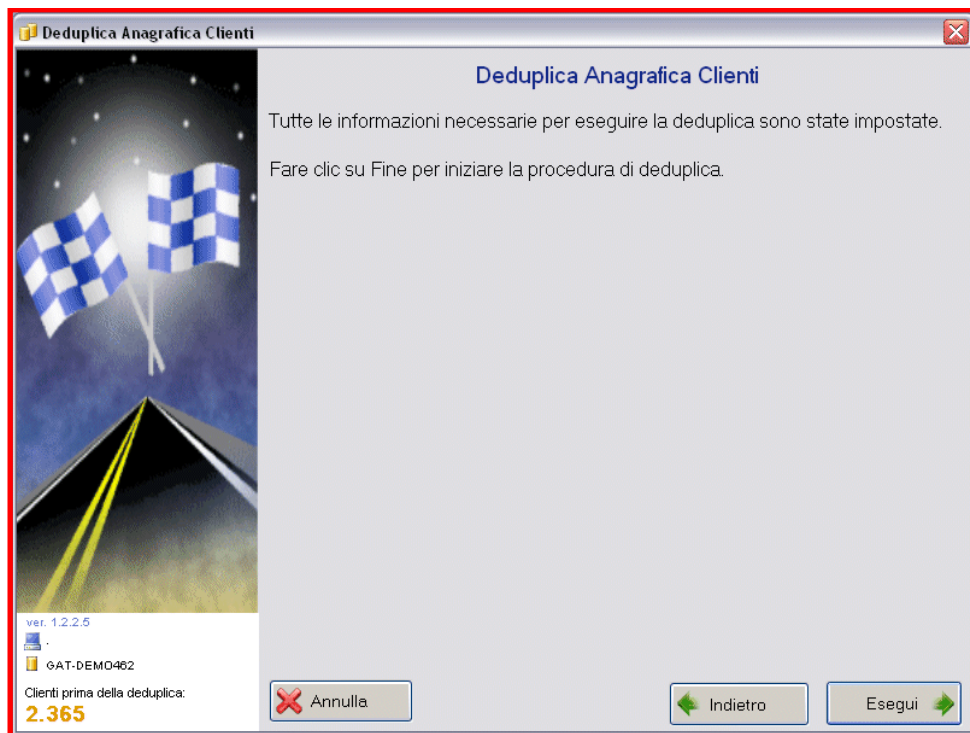
Vengono ricercati i prodotti contenuti in più di un ordine per tutti i clienti duplicati che si stanno elaborando; prima di eliminare gli articoli o gli ordini duplicati, i dati di riga e testata dell'ordine da eliminare vengono riportati nei campi non compilati dell'ordine da mantenere.

Si deve indicare:

- Quanti giorni di differenza sulla Data Segnalazione (è l'unico campo data sicuramente compilato) fanno sì che due ordini con lo stesso articolo siano considerati diversi;
- Il criterio di scelta dell'ordine da mantenere (quello associato a una campagna, oppure quello modificato per ultimo)
- Cosa fare degli ordini che contengono articoli duplicati: se spostare anche gli eventuali altri articoli sull'ordine che si manterrà, oppure se lasciare gli altri articoli collegati all'ordine originale, limitandosi a cancellare le righe duplicate.

2.6 Ricerca Deduplicati

Dopo aver impostato tutti i criteri suddetti, il sistema è pronto per eseguire la deduplica, presentando questa pagina.



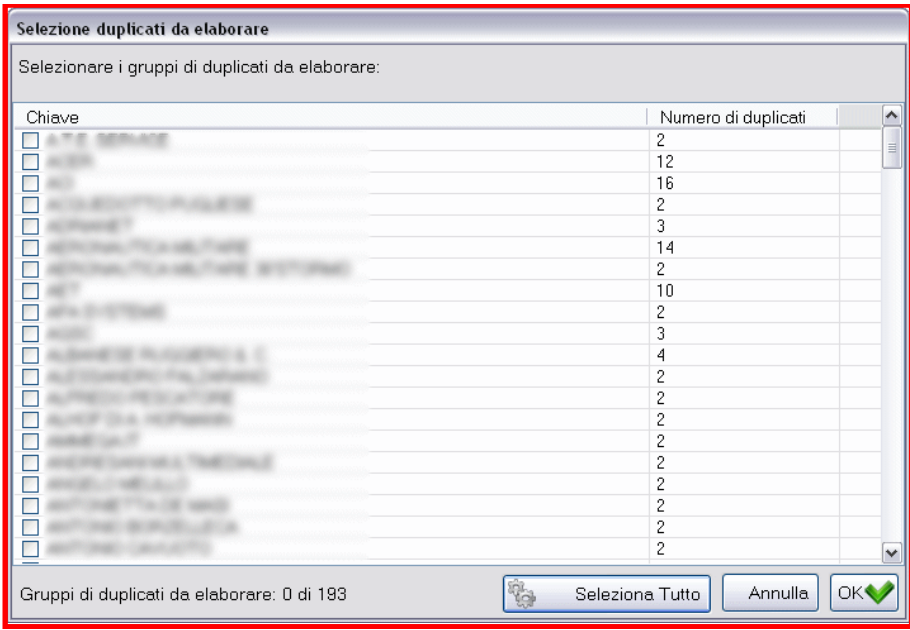
Premendo il tasto Esegui inizia l'elaborazione.



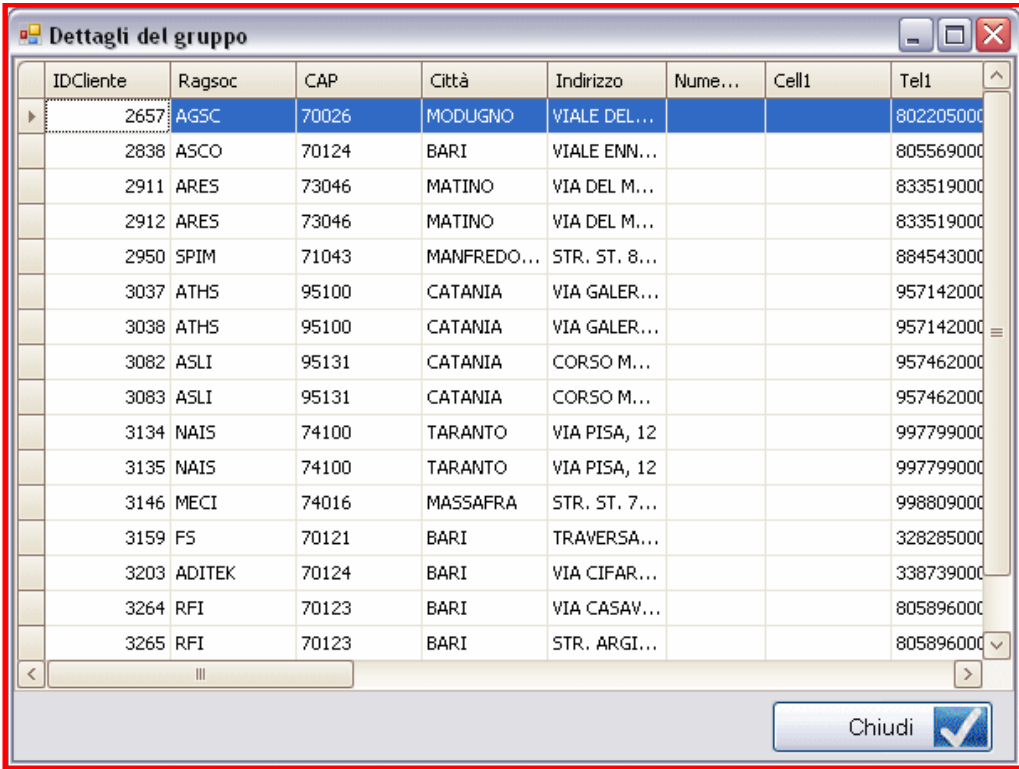
Durante l'elaborazione viene chiaramente indicato lo stato di avanzamento.

NB: in caso di ricerca doppioni per similitudine il sistema viene eseguito più lentamente, ma aumenta la sua velocità progressivamente avvicinandosi al termine.

Il CRM elabora i duplicati come da impostazioni e visualizza i gruppi duplicati, esempio:



Mediante un click con il **tasto destro** del mouse su uno dei nominativi trovati, è possibile aprire il dettaglio, visualizzando l'elenco dei doppioni collegati:



2.7 Esecuzione Deduplicata manuale o automatica

Una volta selezionati i gruppi da deduplicare manualmente (oppure automaticamente) si procede al processo di deduplica.

Selezione cliente da mantenere

Ragsoc	Mantenere	Diversi	IDCliente	CAP	Città	Indirizzo	Numero
AERONAUTICA MILITARE	☐	☐	2498	70057	BARI PALESE	VIALE G. D'ANNUNZIO, 3	
AERONAUTICA MILITARE	☐	☐	2499	70057	BARI PALESE	VIALE G. D'ANNUNZIO, 3	
AERONAUTICA MILITARE	☐	☐	2500	70057	BARI PALESE	VIALE G. D'ANNUNZIO, 3	
AERONAUTICA MILITARE	☐	☐	2640	185	ROMA	VIALE DELL'UNIVERSITA', 4	
AERONAUTICA MILITARE	☐	☐	2672	70023	GIOIA DEL COLLE	VIA GIANRIZZI, 1A	
AERONAUTICA MILITARE	☐	☐	2674	70023	GIOIA DEL COLLE	PROVINCIALE PER TARANTO, 1533	
AERONAUTICA MILITARE	☐	☐	2855	70128	BARI	VIA D'ANNUNZIO, 36	
AERONAUTICA MILITARE	☐	☐	2856	70128	BARI PALESE	VIALE G D'ANNUNZIO, 36	
AERONAUTICA MILITARE	☐	☐	2859	70057	BARI PALESE	VIALE G. D'ANNUNZIO, 3	
AERONAUTICA MILITARE	☐	☐	2860	70128	PALESE	VIALE D'ANNUNZIO, 36	
AERONAUTICA MILITARE	☐	☐	3187	70023	GIOIA DEL COLLE	VIA GIANRIZZI, 1A	
AERONAUTICA MILITARE	☐	☐	3188	70023	GIOIA DEL COLLE	VIA GIANRIZZI, 1A	
AERONAUTICA MILITARE	☐	☐	3189	70023	GIOIA DEL COLLE	VIA GIANRIZZI, 1A	
AERONAUTICA MILITARE	☐	☐	3256	70057	BARI PALESE	VIA G.D'ANNUNZIO, 3	

< ||| >

Dettagli Ignora OK Annulla

Nel caso di selezione manuale del nominativo da mantenere compare la maschera seguente, dove sarà possibile selezionare il nominativo da mantenere (uno solamente) ovvero confermare la diversità di alcune anagrafiche tra loro. In quest'ultimo caso il sistema si ricorderà questa diversità per le successive elaborazioni.

Premendo il tasto "DETTAGLI" è possibile confrontare il valore di tutti i campi.

Immagine	VIALE G. D'ANNU...	VIALE G. D'ANNU...	VIALE G. D'ANNU...	VIALE DELL'UNIV...	VIA GIANRIZZI, 1A	PROVINCIALE F
Indirizzo	VIALE G. D'ANNU...	VIALE G. D'ANNU...	VIALE G. D'ANNU...	VIALE DELL'UNIV...	VIA GIANRIZZI, 1A	PROVINCIALE F
Indirizzo2						
InterventiCondizi	0	0	0	0	0	0
InterventiDove	0	0	0	0	0	0
InterventiTipo	0	0	0	0	0	0
Note						
Notestampa						
Notestrada						
Numdip						
NumeroCC						
NumeroCivico						
NumeroFax1	80586400000000	80584000000000	80584000000000			
NumeroFax2						
Paese						
Paese2						
Piva						
Prov	BA	BA	BA	RM	BA	BA
Prov2						
Ragsoc	AERONAUTICA ...	AERONAUTICA ...	AERONAUTICA ...	AERONAUTICA ...	AERONAUTICA ...	AERONAUTICA
SettAtt	Pubb. Amm. Cen...	Pubb. Amm. Cen...	Pubb. Amm. Cen...	Pubb. Amm. Cen...	Pubb. Amm. Cen...	Pubb. Amm. Ce
Tel1	80584000000000	80584000000000	80584000000000	64986600000000	80343400000000	803487000000
Tel2					33466800000000	

Chiudi

Selezionando il nominativo da mantenere si dovrà premere "OK".

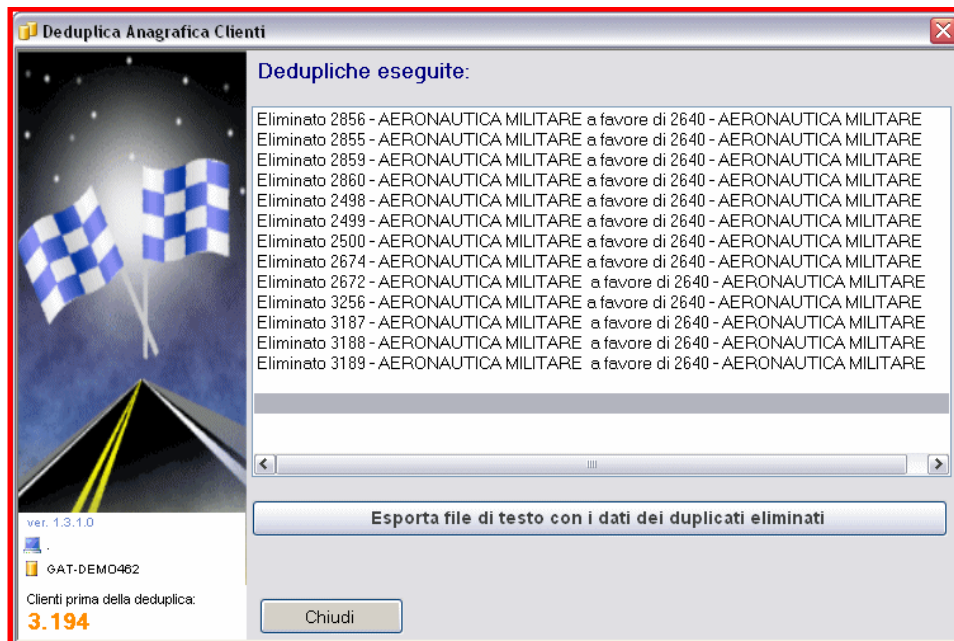
In caso di selezione manuale si procede a questa scelta per ogni gruppo di doppioni che si è scelto di deduplicare.

Fatto questo il sistema procederà all'aggiornamento del database ...



2.8 Risultato finale

Il CRM indica le dedupliche eseguite, evidenziando il cliente mantenuto e quelli eliminati:



NOTA BENE: si consiglia di eseguire la deduplica quando nessuno è collegato al sistema CRM e senza interrompere il processo. Si consiglia ad esempio di lanciare il processo alla sera dopo il termine del lavoro o nel weekend. E' vivamente consigliata la **copia di backup** degli archivi prima di eseguire la procedura.