



## NameSearch - News & Information

### New MerlinMerge®SpeedPro

MerlinMerge® SpeedPro is a high performance self contained, graphical duplicate detection, householding and merge/purge utility. By leveraging our unique knowledge and experience gained through the development of our highly acclaimed search and matching software tool NameSearch®, MerlinMerge® offers remarkable performance while overcoming significant variations without matching errors.

MerlinMerge® SpeedPro is extremely versatile and can process data from a multitude of sources including fixed width and delimited text files, MS-SQL, Oracle, DB2 and Microsoft access. Users can vary the degree of fuzzy searching by the selection of match criterion and threshold scores.

Competitively priced, MerlinMerge® SpeedPro offers unsurpassed quality, accuracy and speed without incurring the overhead associated with other enterprise matching solutions.

MerlinMerge® SpeedPro runs under Windows and Unix based platforms.

Supported Platforms		
Windows NT/2000/XP	Windows 95/98/ME	Net
Solaris	AIX	UNIX
Linux	HP-UX	ZOS, OS/390

## Northrop Grumman & NameSearch®

### “State of Maine’s Criminal Identification System Intelligently Retrieves Information”

Formerly TRW, Northrop Grumman utilized the Advanced Searching and Matching Technology of NameSearch® in their Criminal History solutions.

Thanks to the power and accuracy of the NameSearch® software toolset Maine’s Criminal History system overcomes variations due to misspellings, the use of nicknames, shortened forms, transpositions, phonetic differences, abbreviations, and sequences variations.

Criminals often bastardize their names and other identifying information to undermine electronic detection. The ability to overcome name variations and accurately find criminal history records enables law enforcement agents to construct better criminal profiles. NameSearch® not only aids criminal investigations but significantly mitigates the risk to law enforcement personnel by identifying criminals that could be armed and dangerous.

When reliability matters and life and death decisions depend on finding information top law enforcement systems employ NameSearch®.

MerlinMerge®SpeedPro	.... 1
Northrop Grumman	..... 1
MS-SQL 2000 & 2003	..... 2
Business Objects	..... 2
Corporate Name Searching	3

## NameSearch and SQL Server 2000

### MerlinMerge® for SQL Server 2000&2003

NameSearch® includes MerlinMerge®, a sophisticated database management tool that performs duplicate detection (deduping), merge/purge operations and household link determination.

SQL and Oracle. MerlinMerge® automatically creates extended stored procedures and user defined functions.

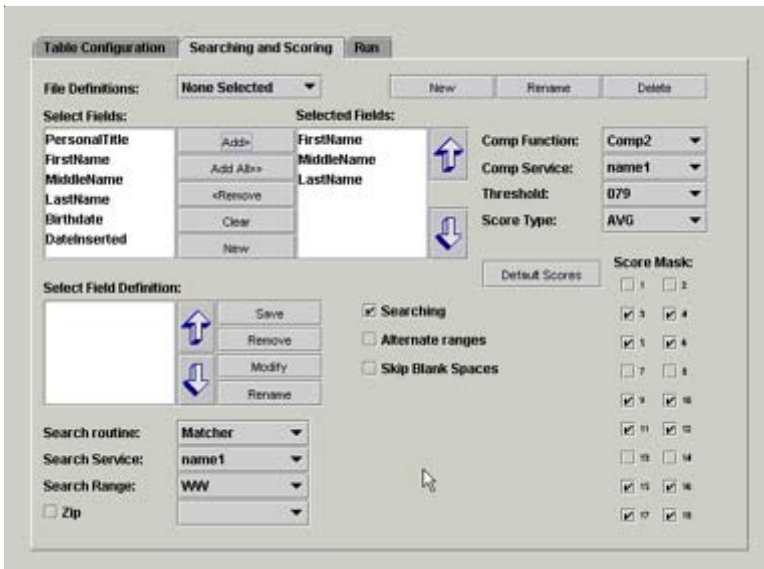
The intuitive graphical user interface enables the selection of search and matching criterion, scoring thresholds and fuzzy search depth. MerlinMerge® constructs a SQL script that uses NameSearch® to perform the batch matching process.

Essentially MerlinMerge® writes a database program based on the selected options and configuration settings.

Unlike most machine generated programs, MerlinMerge® scripts are very easy to read and understand providing unsurpassed flexibility.

### Business Objects Utilize NameSearch®

Brewster, NY and Paris, France - Intelligent Search Technology, Ltd, a leading provider of name and address searching software announced that Business Objects will utilize IST's flagship product, NameSearch® to streamline its internal operations and database management. Business Objects, the world's leading provider of business intelligence (BI) solutions chose NameSearch® after careful evaluation of other products available on the market. Business Objects found that NameSearch® adds highly intelligent and advanced searching and matching capabilities to their global customer management systems.



Like its sister product MerlinMerge® SpeedPro, MerlinMerge® harnesses the advanced search and matching facilities provided by the "NameSearch" engine to achieve unparalleled accuracy and speed while overcoming variations due to misspellings, transcriptions, transpositions, acronyms, phonetics, sequence differences, nicknames and many other common errors found in data.

Where the SpeedPro product extracts data, MerlinMerge® proper works directly with MS-

# Corporate Name Searching

Data retrieval, the most fundamental requisite for information systems is seemingly uncomplicated yet extraordinarily elusive to implement. Boolean expressions accurately record digital information yet obscure the inherent symbolism and patterning readily achieved through inherent human cognitive understanding. The inability to reconstitute ideas, images and memories from generalized concepts and utilize the resulting abstractions for hypothesizing, categorization and problem solving is the single greatest impediment facing the evolution of intelligent systems.

Routinely, technologists are unknowingly faced with the “abstraction-categorization” barrier when asked to construct applications that query text based identity information.

The failure of search systems to satisfy an end user’s expectations is caused by people’s innate ability to extrapolate meaning from fragmented data where computerized systems require exacting precision in order to retrieve information found in detailed binary data.

Corporate name searching concretely illustrates the pragmatic difficulties in developing solutions that find correct information without missing likely candidates. People readily understand the similarities between “Triple A towing” and “AAA towing” yet computerized systems would need to employ a knowledge based algorithm to recognize the relationship between Triple A and AAA.

The deployment of intelligence through knowledge based systems greatly benefits search and matching algorithms by identifying nicknames, shortened forms, noise words and other circumstances that require experience to return a more comprehensive result set. However, knowledge based systems are limited by the breadth and depth of their lexicon. Contrary to names such as IBM and AT&T, the vast majority of acronyms lie outside the scope of knowledge base processing. For example, our clients often used the IST acronym interchangeably with Intelligent Search Technology yet it would be

unreasonable to expect the inclusion of IST in a knowledge based system.

Exacerbating the company name search problem is the use of elements that have varying implications based on their context necessitating the need for contradictory rules. The use of “and” in the name “Judy Ann and Richard Scott Wagner LLP” as compared to “Kohler and Barnes LLP” demonstrate a typical contradictory scenario. In the first instance, the “and” denotes two distinct name units (Richard Scott Wagner and Judy Ann Wagner) as compared to the later where the “and” should be ignored.

Similarly, the most robust knowledge based systems fail to account for the infinite deviations created by combinations of misspellings, transpositions and phonetic variations.

Sequence variation, extra or missing information further complicates the successful retrieval of relevant data. Many systems fail to return “Bush, Cheney, Powell and Rumsfeld” when only the “Rumsfeld and Cheney” name elements were used to initiate a search. Regrettably the inability of the airline’s electronic detection systems to overcome this class of variations resulted in the failure to identify the 9/11 hijackers.

“Regrettably the inability of the airline’s electronic detection systems to overcome this class of variations resulted in the failure to identify the 9/11 hijackers.”

Through decreasing distinctiveness, generalizing transformation functions combined with the removal and masking of variations limit accuracy and introduce increased potential for invalid matches. The consequence of increasing variation tolerance is the expansion of result sets. Not only are performance penalties incurred as result sets get larger but the inclusion of irrelevant information increases.

Full table scan intelligent string based comparison solutions may achieve high levels of accuracy but are computationally intensive and limited to small data sources. Until highly parallelized computers with thousands of processing nodes become commercially available non indexed based solutions will be limited to fairly modest sized data stores. Any enhancement made to this class of solutions increases computational load by multiplying the execution time of the additional instructions with the number of comparisons needed to scan the data



## ISTwatch - News & Information

source. Intelligent string based comparison solutions become impracticable as the comparison functions become more robust and data stores increase in size.

The NameSearch® software with its corporate search algorithms and acronym recognition subroutines significantly advances an information system's ability to seek and match corporate name data.

NameSearch® enables systems to make and apply abstractions with out ignoring discrepancies. The searching function casts an abstraction net that captures likely candidates. Utilizing rule based expertise, knowledge enhanced phonetic recognition, sanitization, acronym recognition and multiple pathing technologies the NameSearch® routine over comes discrepancies while minimizing the inclusion of irrelevant records.

The NameSearch® match engine is employed to intelligently rank or eliminate unlikely candidates. The corporate name scoring routine derives results through its evaluation of a neural network. The neural network is created by comprehending the degree of similarity between two names. The advanced heuristic pattern recognition subprogram receives information from the raw input combined with observations made by the knowledge base and phonetic routines. The pattern recognition facility digests the information and constructs weighted nodes that are placed on the neural net. Once the neural net is completed it is optimistically evaluated and intelligent scores rendered.

NameSearch® achieves unsurpassed performance through returning only relevant data reducing both computationally intensive matching activities and the utilization of expensive input/output operations.

With NameSearch® information systems will attain unsurpassed levels of matching performance, accuracy and reliability.

### **Trial and Demonstration Software**

*For a free 60 day trial call*

**(800) 287-0412**

*or*

**(845) 278-8989.**