Identity Resolution & Data Quality Solutions
Matching, Address Validation, Profiling, Screening, Compliance

NameSearch®
Product Data Sheet

I. INTRODUCTION

Intelligent Search Technology's (IST) flagship product, *NameSearch®*, uses advanced fuzzy searching & matching, multicultural intelligence to search and match data such as, personal and company names, account numbers, Street addresses, phone numbers, cities, email addresses, zip codes, dates of birth, countries, custom alphanumeric terms(e.g. Part Numbers).

This powerful searching and matching technology overcomes data variations due to misspellings, transpositions, acronyms, abbreviations, nicknames, etc. The result is more accurate search results while virtually eliminating false positives. *NameSearch* comes with an extensive nickname rulebase which includes not only English, but also multi-cultural names. Users are also given the opportunity to import other rule sets in order to fully customize the matching process.

In addition to the rulebase system, *NameSearch* relies on a set of sophisticated and highly efficient phonetic algorithms and comparison routines to provide the most accurate and powerful searching and matching technology available today.

II. FEATURES

Other searching and matching engines often use exact matching, wildcard matching, or phonetic matching such as that used in the *Soundex Algorithm*. Research has shown that a single approach to fuzzy matching is not sufficient to overcome the numerous variations possible in data. *NameSearch* utilizes a combined approach that identifies matches often overlooked using standard methodologies. It features:

- Cleansing and standardization to remove noise and extraneous information.
- Rule-based expertise to provide context-specific knowledge.
- Phonetic algorithms and rules to identify sound-alike scenarios.
- Key-based matching to improve performance and to address multiple typographical errors.
- Neural net technology to "intelligently" process the data.
- Advanced pattern matching to calculate similarity levels between matching candidates.

NameSearch will increase the quality of matches while minimizing I/O expense. By providing intelligence to applications requiring efficient and accurate searching and matching, the search can successfully enable organizations to retrieve and identify information regardless of variations. The technology uses searching and matching intelligence 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. For example:

When looking for...	NameSearch identifies potential matches such as...
"Robert Smith at 223 East Main Street"	"Rob Smith", "Roby Smith", "Smith, Robert", "Bob Smith", "Bobby Smith", "Bob Smyth", "Rob Smeeth", "Bob Smythe", or "223 East Main Str", "223 Main Street East", "223 E. Mein Str", "223 East Main Street Apt 4"

Identity Resolution & Data Quality Solutions
Matching, Address Validation, Profiling, Screening, Compliance

A. ADVANCED FEATURES

NameSearch Generation Shell - The *NameSearch* Generation Shell is a graphical user interface that can be used for customizing, testing and/or exporting the product's Shared Library. It enables users to create new services, specify option settings, customize rulebases, perform frequency analysis, create phonetic scripts, and create statistical information for prediction routines. The Generation Shell includes powerful testing facilities that enable organizations to prototype and test *NameSearch*.

NameSearch SDKs - Software Development Kits (SDKs) make it very easy to implement *NameSearch* within various types of applications. The product integrates into almost any environment including the Windows, UNIX, Linux, and Mainframe operating systems, SQL Server, Oracle, DB2, Sybase, MySQL database environments, web-based applications and almost any other type of server or user applications. *NameSearch* exposes its functionality through the use of a shared library. The product comes with a set of SDKs that enable calls to *NameSearch* from almost any programming environment i.e. Java, VB.NET, ASP, ASP.NET, C#, C, C++, PHP, Perl, COBOL, PL/SQL, T-SQL, DB2, Sybase, etc.

MerlinMerge SpeedPro - *NameSearch* includes a full version of our *MerlinMerge*® *SpeedPr*o duplicate detection engine/data management software. This application is capable of interfacing with database systems such as SQL Server, Oracle, DB2, Microsoft© Access, MySQL and Teradata. A customized built-in engine retrieves and reloads database records quickly and efficiently.

SQL Scripter Database Tool - This sophisticated application analyzes databases and table structures and creates complex SQL scripts to be used within the database environment. Scripting capabilities include: key building, searching, deduping, merge/purge, and triggers. Accessing *NameSearch* functionality directly from SQL scripts provides great flexibility and efficiency, because the scripts are fully customizable and the respective database environments are optimized for running scripts. The created scripts can be executed right from the application or can be accessed from user applications as stored procedures.

III. BENEFITS

Accuracy - NameSearch consistently outperforms the competition in terms of accuracy and performance. Intelligent search key and range building is used for the retrieval of records regardless of variation caused by phonetics, transcription or keyboarding errors, nicknames, short forms, missing words, extra words, noise words, sequence, and multi-cultural variations.

Performance - Thanks to its highly sophisticated key building mechanism, *NameSearch* retrieves likely matches quickly and without full table scans thereby considerably minimizing I\O expense. Performance is further enhanced by our multi-threaded architecture. Applications that utilize multiple CPUs will see an increase in performance. *NameSearch* has been integrated into numerous real-time applications due to its ability to retrieve information instantaneously.

Scalability – *NameSearch's* search and comparison routines are highly efficient. There are no limitations to the number, size of files, or database tables used in the matching process. Some of IST's larger clients use *NameSearch* to search and match datasets in excess of 1 billion records. Significant increases in data may cause a minimal increase in response time. This is also due to *NameSearch's* database integration model, which uses the database to do the work related to search key indexing and querying, instead of expensive in-memory alternatives.

Flexibility – *NameSearch's* matching process is comprised of two phases: searching and comparison. For searches, users may choose how restrictive or loose a particular search should be. For comparisons, users may choose form a selection of comparison algorithms in order to achieve optimal results. Users can also adjust score thresholds to broaden or restrict match results. In addition, they can specify custom weight factors, or define multiple combinations of match criteria.

Identity Resolution & Data Quality Solutions
Matching, Address Validation, Profiling, Screening, Compliance

Ease of Customization - *NameSearch* is extremely powerful and highly adaptable, with the flexibility required to meet the needs of a very large number of business applications. The architecture allows for customization of search and match criteria, phonetic algorithms, search and comparison algorithms, rulebases, integration procedures, and more. All of the customization can easily be performed and tested within the Generation Shell, or dynamically via client and database applications.

Ease and Flexibility of Integration - The *NameSearch* search and comparison functions are exported into shared libraries for different OS platforms, including Windows, UNIX, Mainframe, etc. The product can be integrated in both web-based and client applications via API calls or directly within database systems via custom stored procedures, external and user-defined functions.

NameSearch also facilitates real-time data integration operations through the use of database triggers, which enable quick synchronization between user and search look-up tables. This not only simplifies development tasks but also allows for faster data updates.

Multi-platform Support – Currently runs on all versions of Windows, UNIX, and Linux; and on AS/400, IBM Mainframe systems. It delivers the same results across all platforms.

IV.   HOW OUR FUZZY SEARCHING AND MATCHING ENGINE WORKS

In production environments, *NameSearch*'s fuzzy searching and matching functionality is accessed through a specialized shared library - **DLL** on Windows, **shared object** on Linux/UNIX, or **Link Load Library** for the mainframe. The shared library is a collection of callable software components.

NameSearch uses sophisticated **search key and range building** mechanisms to add meaning to data and **intelligent scoring** to determine the likelihood that the examined records are indeed the data that users requested.

A. SEARCH KEY AND RANGE BUILDING

Retrieval of information is achieved by the insertion of search keys. The search keys are used to select a subset of the data that contains only relevant records to be further examined by advanced comparison routines. This approach finds relevant records regardless of variation caused by phonetics, transcription or keyboarding errors, nicknames, short forms, missing words, extra words, and noise and sequence variations. The goal of producing an intelligent search key is to enhance the quality of records returned, while improving performance, by avoiding a complete table scan.

1. SEARCH KEY BUILDING

The key building process includes three steps: **sanitization**, **word pattern recognition**, and **phonetic tokenization**.

Sanitization - Sanitization is the first step to manufacture the search keys. This step removes noise characters, extra spaces, and control characters and converts lower case letters to uppercase.

1. **Word Pattern Recognition** –There are two sub-functions used by the word pattern recognition routines. The first, p*hrase replacemen*t, recognizes multiple word phrases and replaces them with one common phrase. For example, "I.B.M" will be equated with "International Business Machines" or "aka" will be equated with "also known as." The second, the *rulebase expert system,* processes an input string and identifies nicknames, noise words, common prefixes, diminutives, etc.

2. **Phonetic Tokenization** - Discrepancies caused by phonetic errors account for 20-25% of all input variations. *NameSearch* addresses phonetic issues by employing sophisticated analysis routines to determine the extent of phonetic tokenization. Phonetic issues can then be overcome without negative consequences. The result is more accurate searches with more meaningful results.

2. SEARCH RANGE BUILDING

When conducting an inquiry, *NameSearch* accepts an input string as a parameter and returns search ranges. The search ranges are used to find records whose search keys lay within them. A number of search ranges are provided: from narrow to broad with each successive range getting bigger and including the previous set.

B. INTELLIGENT SCORING

Matches are established using *NameSearch*'s advanced scoring functions that utilize neural net technology, rule-based intelligence and advanced heuristic pattern recognition. The scoring functions use the strength of the key building routines to intelligently calculate numeric values (scores) which indicate the likelihood of a match. The matching functionality can deliver scores that approximate values similar to an individual possessing significant linguistic expertise.

V. TECHNICAL REQUIREMENTS & COMPATIBILITY

In production environments, *NameSearch*'s functionality is accessed through a specialized *NameSearch* shared library - DLL on Windows, shared object on Linux/UNIX, or Link Load Library for the mainframe. The shared library is a collection of callable software components.

NameSearch needs to be compiled using an ANSI C compiler. *The* Generation Shell runs on all versions of Windows supported by Microsoft. It runs successfully on all Windows Service Packs. 64MB of RAM is required but 128MB is recommended, however, it can successfully run on computers with less memory. 70MB of available hard-disk space is required.

NameSearch easily integrates into many platforms including Windows (NT, 2000, XP, 2003, Vista, 7), UNIX, Linux, Solaris, HP-UX, AIX, AS/400, z/OS, OS/390, VMS, and VSE. It can seamlessly integrate with SQL Server, Oracle, DB2, Teradata, Sybase, MySQL and other commercially available database systems. In addition, it comes with a set of APIs that enable applications to call the product from programming environments like Java, VB.NET, ASP, ASP.NET, C#, C, C++, PHP, Perl, COBOL, PL/SQL, T-SQL, DB2, Sybase, PowerBuilder, etc.

VI. ABOUT IST

Founded in 1993, Intelligent Search Technology, Ltd. (IST), a privately held company, has devoted its resources to the development of high performance, multicultural identity search, and matching software. Its highly acclaimed identity search and matching engine, *NameSearch®*, adds intelligence to applications requiring efficient and accurate retrieval of identity data regardless of variations. With the development of *MerlinMerge® SpeedPro* (duplicate detection, merge-purge software), *CorrectAddress®* (address verification and standardization software), and *IST Watch*© (OFAC compliance and watch list searching software), IST continues to bolster its position as a true and independent leader in overall data quality.