



NameFinder™



Proper Noun Identification System

Overview






AppTek's NameFinder™ is an advanced technology engine that scans text for proper nouns (such as human names) in various languages—even in writing systems that do not use capitalization. NameFinder™ can also accurately identify the ethno-linguistic origins of a person's name.

The system recognizes transliterations between the Roman (Latin) alphabet and other writing systems, and can identify transliterated names despite discrepancies, ambiguities, or simple misspellings. The system is designed to assist users with multilingual name lexicons. Proper nouns may label locations, persons, currencies, measurements, first or last names, etc.

The system features a large database of proper nouns, which may be associated with other nouns and types of events, regardless of their proximity in a given text; it also identifies nouns in free text, using shallow parsing.

Capabilities

AppTek NameFinder™ is comprised of the following components:

-  Proper noun normalization
-  Analysis and transliteration
-  Universal name identifier (UNI)
-  Rich proper nouns lexicon in several languages
-  Proper noun detection in free text













NameFinder™ operates on all Windows platforms. It has dynamic DLL and API calls for integration with other systems. The system can be deployed with various databases including B-tree, ODBC, SQL Server, Oracle, etc.

In addition, AppTek NameFinder™ supports multiple languages and scripts such as English/Latin, Cyrillic, Arabic, Farsi, etc.

Functionality

NameFinder™ uses local and remote context and statistical information to determine if a word is a potential proper noun. It looks for words that are flagged as "can be a name" in the lexicon, and context clues (such as "Mr." or "doctor") to help find other names. The remote context incorporates verbs and words in general that are associated (collocation) with proper nouns. These are determined statistically by using the running text corpora. If a word, or words, are flagged as "can be a name" in the lexicon and context information enforces the proper noun detection, the word is tagged as being a proper name. Words that do not meet both criteria are tagged as potential proper nouns.

AppTek NameFinder™ provides the following capabilities:

-  Utilizes special linguistic algorithms
-  Detects, analyzes, and transliterates proper nouns
-  Detects proper nouns in free-format text
-  Validates name entries and transliterations
-  Identifies possible similarity or match between various spelling of proper nouns
-  Identifies specific ethno-linguistic name groups
-  Handles name-spelling ambiguity, e.g., Roman (Latin) vs. non-Latin
-  Checks against standard (base-form) transliteration entry in the lexicon
-  Links discrepant entries with standard lexical entries
-  Produces reports on discrepant, rejected, and valid entries
-  Provides system administrator with maintenance utility for lexicon
-  Integrates with Machine Translation for Name-Context analysis



Bridge the Language Gap