PeacockData

WHAT IS FUZZY LOGIC?

If you typed "**Garfeild**" into a word processor, it would probably be underlined with a squiggly red line signifying a misspelling. It is the name "**Garfield**" with the "**IE**" reversed to "**EI**"—a common mistake.

Our fuzzy logic technology allows matching name data that has typographical errors. If you look at the fuzzy logic examples we have provided below, you are likely to see errors you have repeatedly made or seen. In many cases you will have to look close to see the difference, but they are different.

The **Pro** editions of our name software products, **pdNickname**, **pdGender**, and **pdSurname**, are fully incorporated with advanced fuzzy logic. This document illustrates the fuzzy logic technology in these products. Further information specific for these packages can be found in the product user documentation.

HOW FUZZY LOGIC WORKS

Fuzzy logic attempts to duplicate real errors created while entering names into databases. The most likely typographical errors are determined based on the number of letters, the characters involved, where they are located in the name, the language, and other factors.

The biggest advantage in our technology is in its ability to work with language rules that indicate how individual of various nationalities may hear and spell names.

Some fuzzy logic spellings have one typographical error while others have multiple issues, so the technology is suited for even the worst typists and transcribers. The algorithms have five layers:

PHONETIC MISSPELLINGS

These algorithms look at digraphs, trigraphs, tetragraphs, pentagraphs, hexagraphs, and even a German heptagraph, "SCHTSCH", used to translate Russian words with the "SHCHA" or "SHCH" (romanticized) sound. These are, respectively, two to seven letter sequences that form one phoneme or distinct sound. Most of letter sequences trigraph and above are Irish who have more language rules than you can shake a stick at.

Many misspellings occur as transcribers enter the sounds they hear. The character sequences and the sounds they produce are different for each language and situation, such as before, after, or between certain vowels and consonants, so our substitutions are language-rule based. Furthermore, our algorithms consider both how a name may sound to someone who speaks English as well as how it may sound to someone who speaks Spanish, which is often different. Take the digraph "SC". Before the vowels "E" or "I" it is most likely to be misspelled by an English speaker as "SHE" or "SHI" while a Spanish speaker may hear "CHE" or "CHI" and sometimes "YE" or "YI". Our library includes over 80,000 language-based letter sequence phonetic rules.

FIRST NAME EXAMPLES:

	Real name	**Fuzzy name**	**Gender**
Example 1	BARTHOLOMEW	BARTHOLOMUE	Male
Example 2	DAWNETTE	DAUNETTE	Female
Example 3	NATHANIEL	NATHANAIL	Male
Example 4	PHYLLIS	FYLLIS	Female
Example 5	SIGOURNEY	SIGOURNI	Female
Example 6	XAVIER	XAVAR	Male

LAST NAME EXAMPLES:

	Real name	Fuzzy name
Example 7	AGLIANO	ALLANO
Example 8	GUALTIERREZ	GUALTIEREZ
Example 9	HEATHFIELD	HEATHFALD
Example 10	AAGARD	OUGHGARD
Example 11	YOUNGMAN	YONGMAN

REVERSED DIGRAPHS

These algorithms look for misspellings due to reversed digraphs (two letter sequences that form one phoneme or distinct sound) which are a common typographical issue, such as "IE" substituted with "EI". The character sequences and the sounds they produce are different for each language and situation, such as before, after, or between certain vowels and consonants, so our substitutions are language-rule based.

FIRST NAME EXAMPLES:

	Real name	Fuzzy name	Gender
Example 12	ANNABETH	ANNABEHT	Female
Example 13	CAETLIN	CEATLIN	Female
Example 14	EUGENE	UEGENE	Male
Example 15	FRIEDRICH	FREIDRICH	Male
Example 16	RAQUEL	RAUQEL	Female
Example 17	VICKTOR	VIKCTOR	Male

LAST NAME EXAMPLES:

	Real name	Fuzzy name
Example 18	ANGLES	ANLGES
Example 19	DIELEMAN	DEILEMAN
Example 20	OLEARY	OLAERY
Example 21	RODREGUEZ	RODREUGEZ
Example 22	SCHUMACHER	SCHUMAHCER

DOUBLE-LETTER MISSPELLINGS

These algorithms look for misspellings due to double letters typed as single letters and single letters that are doubled. The most common typographical issues occur with the characters, in order of frequency, "SS", "EE", "TT", "FF", "LL", "MM", and "OO".

FIRST NAME EXAMPLES:

	Real name	Fuzzy name	Gender
Example 23	EMANNUEL	EMMANNUEL	Male
Example 24	KASSANDREA	KASANDREA	Female

LAST NAME EXAMPLES:

	Real name	Fuzzy name
Example 25	HUMBER	HUMBEER
Example 26	ZWOLLE	ZWOLE

MISSED LETTERS

These algorithms look for missed keystrokes and provide fuzzy logic matches with missing letters. Unlike the other algorithms, these are not language specific. Keystrokes can be missed in any language.

FIRST NAME EXAMPLES:

	Real name	Fuzzy name	Gender
Example 27	ABDUL	ADUL	Male
Example 28	MARGARET	MRGARET	Female

LAST NAME EXAMPLES:

	Real name	Fuzzy name
Example 29	HUNTER	UNTER
Example 30	TAMERON	TAMRON

STRING MANIPULATIONS

These algorithm changes letters and syllables in a variety of ways. They are less guided by language rules and more guided by randomness.

FIRST NAME EXAMPLES:

	Real name	Fuzzy name	Gender
Example 31	CYNTHIA	CYNTTHA	Female
Example 32	GERALD	GERLLD	Male

LAST NAME EXAMPLES:

	Real name	Fuzzy name	Gender
Example 33	ELWORTHY	ELWROTHY	Female
Example 34	PEOPLE	POEPLE	Male

COPYRIGHT NOTICE