PDNICKNAME LAYOUT AND DATA DEFINITIONS

Below are the complete layout specifications and data definitions of all files provided with *pdNickname*.

Each line below contains the following information: FIELD NUMBER: field position number. FIELD NAME: name of field. FIELD TYPE: field data type; "*Chr"* = alpha/numeric characters, "Num" = numbers. FIELD LENGTH: length of field. DECIMAL PLACES: number of decimal places (if any). START POSITION: field starting position. END POSITION: field ending position. DESCRIPTION: data definition of field contents.

LAYOUT OF PDNICKNAME NAMES DATABASE

Field Count: 11
Total Length: 485
Record Count: Pro and Standard: 397,847

FIELD NUMBER	FIELD NAME	FIELD TYPE	FIELD LENGTH	DECIMAL PLACES	START POSITION	END POSITION	DESCRIPTION
1	PEACOCK_ID	Chr	17		1	17	Unique identifier for each record
2	NORMAL	Chr	35		18	52	Normalized name spelling
3	STANDARD	Chr	35		53	87	Standardized name spelling
4	NAME	Chr	35		88	122	Stylized name spelling
5	GENDER	Chr	1		123	123	Gender flag: *M = Male* *F = Female*
6	GIVEN	Chr	1		124	124	Given name flag: G = Is a given name
7	NICK	Chr	1		125	125	Nickname flag: N = Is a nickname
8	RANK	Num	5	0	126	130	Name rank in the United States
9	ARCHAIC	Chr	1		131	131	Archaic name flag: *A = Archaic*
10	LANGUAGE	Chr	254		132	385	Language or languages of origin and use
11	SPECIAL	Chr	100		386	485	Special and unique origins

LAYOUT OF PDNICKNAME RELATIONSHIP FILE

Field Count: 13
Total Length: 82
Record Count: Pro: 51,391,682; Standard: 41,653,922

FIELD NUMBER	FIELD NAME	FIELD TYPE	FIELD LENGTH	DECIMAL PLACES	START POSITION	END POSITION	DESCRIPTION
1	GENDER1	Chr	1		1	35	Gender of the first name in the related name pair
2	NAME1	Chr	35		2	36	First name in the related name pair
3	GENDER2	Chr	1		37	37	Gender of the second name in the related name pair
4	NAME2	Chr	35		38	72	Second name in the related name pair
5	REL	Chr	1		73	73	Relationship flag: *1 = Close onomastic variant* *2 = Near onomastic variant* *3 = Distant onomastic variant* *S = Short form nickname* *D = Diminutive nickname* *P = Phonetic match* *X = Opposite gender match* *F = Fuzzy logic match* (*Pro* only)
6	SCORE	Chr	2		74	75	Match quality score: *01 (best) to 99 score; 99 is reserved for archaic matches; 00 is entered for opposite gender and fuzzy logic matches*
7	DMP	Chr	1		76	76	Double Metaphore: *P = Primary line match* *S = Secondary line match*
8	MP	Chr	1		77	77	Metaphone: *M = match*
9	NY	Chr	1		78	78	New York State Identification and Intelligence System (NYSIIS): *N = match*
10	CP	Chr	1		79	79	Caverphone: *C = match*
11	SX	Chr	1		80	80	Soundex: *S = match*
12	DMSX	Chr	1		81	81	Daitch–Mokotoff Soundex: *P = Primary line match* *S = Secondary line match*
13	DIR	Chr	1		82	82	Name pair direction flag: *A = Name pair is in standard order* *R = Name pair is in reverse order*

IMPORTING DATA INTO YOUR SYSTEM

pdNickname is designed to be compatible with any database system. It comes in multiple file formats, uses only the ANSI character set, and has a well-defined layout.

INCLUDED DATABASE FILES

pdNickname has two data sets, a names database and a relationship file.

Included files are:

NAMES DATABASE

This data set lists more than 397,000 given name and nickname formations. The relationship file is based on these names. The names database also includes a name ranking, archaic flag, and languages of origin and use.

RELATIONSHIP FILE

This data set has more than 40 million (*Standard* edition) or 50 million (*Pro* edition) name-pair records based on the given names and nicknames in the names database. It includes the pairs of related names and nicknames along with extensive information about their onomastic and phonetic relationships. A scoring system is provided to order matches from most likely to least likely. The relationship file is divided into sections for user convenience and to allow users more options when setting up their system.

FILE FORMATS

The database is available in three common file formats. Each format contains the same data.

Available file formats are:

CSV-COMMA SEPARATED VALUES

Files in Comma Separated Values (CSV) format (also known as Comma Delimited) separate fields with commas, and alpha/numeric character fields are usually delimited with double quotes (in case some of the field content includes commas). This format is the most commonly used. It is a native format for Microsoft Excel and is compatible with nearly all database management systems and spreadsheets.

TXT-FIXED LENGTH

Files in Fixed Length (TXT) format (also known as Standard Data Format or SDF) use constant field positions and lengths for all records. In other words, each field starts and ends at the same place in the text file and each record is on a separate line. While not as popular as comma separated values, this format is preferred by many due to its input precision and is widely used to transfer data between different software programs. It is compatible with most database management systems and spreadsheets.

DBF-DATABASE

Files in DBF database format (also known as xBase) are native to Microsoft FoxPro and Visual FoxPro, dataBased Intelligence dBase, Alaska Software XBase++, Apollo Database Engine, Apycom Software DBFView, Astersoft DBF Manager, DS-Datasoft Visual DBU, Elsoft DBF Commander, GrafX Software Clipper and Vulcan.NET, Multisoft FlagShip, Recital Software Recital, Software Perspectives Cule.Net, and xHarbour.com xHarbour. They are also compatible with any database management system that can import the DBF (xBase) format, such as Microsoft Access, Microsoft SQL Server, and numerous others.

CHARACTER SET

The ANSI character set is utilized for all database records. This includes ASCII values 0 to 127 and extended values 128 to 255. These are also known as the extended Latin alphabet. Some users may needs to configure their database system to import the extended values. In many cases the option will be labeled the “Latin-1” character set.

COMPATIBILITY

To ensure compatibility with any operating system and database platform, **pdNickname** is provided in multiple file formats and utilizes only the ANSI character set (ASCII values 0 to 127 and extended values 128 to 255).

USING PDNICKNAME WITH PDSURNAME AND PDGENDER

pdNickname, pdSurname, and *pdGender* make excellent partners. They have been developed to be fully compatible. The name pair format in *pdNickname* is very similar to the *pdSurname* database except *pdNickname* is used to match give names and nicknames while *pdSurname* matches last names. *pdGender* is based on the first name database and is designed to apply gender identification to first name records. Note that *pdSurname* and *pdGender* are not required to use *pdNickname* but they are highly attuned to work together.

COPYRIGHT NOTICE

pdNickname is Copyright © 2009-2016 Peacock Data, Inc. All Right Reserved.