

Using the PGN File Format - R.C.C.C. Summary

I've copied, pasted, quoted, paraphrased, and in some cases, completely reworded the vital information from the "Portable Game Notation Specification and Implementation Guide" in order to create a short PGN guide. I hope you find it useful. -*Stef*

Quick & Dirty Instructions

Here's the quick explanation, using Fool's Mate as an example.

Using a text editor or word processor, type the game information and save it as a PGN file by creating the extension ".pgn" in the file name ("whatever.pgn"). If you're a beginner and creating PGN files from scratch, I recommend using Microsoft Notepad or something similar. With a program like Microsoft Word, you'll probably have to save the file with a ".txt" extension and change it to ".pgn" after you're done. Also, many chess programs will save games for you in PGN format, after you've entered the moves by playing them on an on-screen chess board. You can then export those PGN files to other programs.

Game Information - Fool's Mate

```
[Event "Sample Game"]  
[Site "Murfreesboro, TN USA"]  
[Date "2007.04.05"]  
[Round "-"]  
[White "Fool, Annie O."]  
[Black "Master, Mary D."]  
[Result "0-1"]
```

```
1. f3 e6 2. g4 Qh4# 0-1
```

First, list the seven headers (tag pairs) above and enter the appropriate data in quotation marks. The headers are always inside brackets. These particular seven headers must always be there, and they must always be the first seven listed. They should also appear in this order. Additional headers may appear afterward (the Implementation Guide has a complete list), but are not necessary. There should be no spaces or empty lines at the top of the page or before the first header. If the game isn't part of a match or a tournament, you can just put a single hyphen or a "1" in the "Round" tag. Also, you can use a question mark in any field that does not have data.

Second, skip a line.

Third, enter the moves (movetext) in SAN (Standard Algebraic Notation), with a space between the move number and white's move, a space between White's move and Black's move, and a space between Black's move and the next move number. There's always a period after the move number. Avoid lines with more than eighty characters. Leave no spaces after the result.

Fourth, skip another line and save the file. If you have multiple games in one PGN file, just skip a single line between each game.

That's it. You're now ready to use the PGN file in a program, offline or online. PGN files are very small text-only files, so you can easily send them to your friends as email attachments.

Here's More Detail

Basically, PGN is a simple way for anyone to record games that can be read by most database and playing software, and used internationally. According to its authors, "*PGN is 'Portable Game Notation,' a standard designed for the representation of chess game data using ASCII text files. PGN is structured for easy reading and writing by human users and for easy parsing and generation by computer programs. The intent of the definition and propagation of PGN is to facilitate the sharing of public domain chess game data among chess players (both organic and otherwise), publishers, and computer chess researchers throughout the world.*"

The standard criteria can be found in two documents: the "Portable Game Notation Specification and Implementation Guide" last revised on March 12, 1994, and the "Supplement," last revised on September 8, 2001. Please read those two documents for detailed information. They are both available from the R.C.C.C. website.

If you have a simple text editor (such as Crimson Editor or Microsoft Notepad, which comes with every version of Windows) or a word processor (such as Word, Wordpad, Wordperfect, Writer, Word Pro, Pages, etc.), then you can create PGN files. Just save the file with a ".pgn" extension. When using a word processor, you might have to save the file as a text file (".txt") and then change the extension to ".pgn" after the fact.

Here's a sample PGN file from the Implementation Guide. This includes only one game, but a single PGN file can include multiple games. There are two sections to any PGN file. The first is the "tag pair" section (the headers) and the second is the "movetext" section (the moves of the chess games).

[\[Event "F/S Return Match"\]](#)

[Site "Belgrade, Serbia JUG"]
[Date "1992.11.04"]
[Round "29"]
[White "Fischer, Robert J."]
[Black "Spassky, Boris V."]
[Result "1/2-1/2"]

1. e4 e5 2. Nf3 Nc6 3. Bb5 a6 4. Ba4 Nf6 5. O-O Be7 6. Re1 b5 7. Bb3 d6 8. c3 O-O 9. h3 Nb8 10. d4 Nbd7 11. c4 c6 12. cxb5 axb5 13. Nc3 Bb7 14. Bg5 b4 15. Nb1 h6 16. Bh4 c5 17. dxe5 Nxe4 18. Bxe7 Qxe7 19. exd6 Qf6 20. Nbd2 Nxd6 21. Nc4 Nxc4 22. Bxc4 Nb6 23. Ne5 Rae8 24. Bxf7+ Rxf7 25. Nxf7 Rxe1+ 26. Qxe1 Kxf7 27. Qe3 Qg5 28. Qxg5 hxg5 29. b3 Ke6 30. a3 Kd6 31. axb4 cxb4 32. Ra5 Nd5 33. f3 Bc8 34. Kf2 Bf5 35. Ra7 g6 36. Ra6+ Kc5 37. Ke1 Nf4 38. g3 Nxh3 39. Kd2 Kb5 40. Rd6 Kc5 41. Ra6 Nf2 42. g4 Bd3 43. Re6 1/2-1/2

Tag Pair Section

First, you must create a the headers. These headers are in brackets, and they tell the computer program who, when, where, etc. There are no spaces between headers. Also, you can choose to have no header information whatsoever.

There is a set of tags defined for mandatory use for archival storage of PGN data. This is the STR (Seven Tag Roster). For import format, the order of tag pairs is not important. For export format, the STR tag pairs appear before any other tag pairs (the STR tag pairs must also appear in order; this order is described below). Also, for export format, any additional tag pairs appear in ASCII order by tag name.

The seven tag names of the STR are (in order):

- 1) Event (the name of the tournament or match event)
- 2) Site (the location of the event)
- 3) Date (the starting date of the game)
- 4) Round (the playing round ordinal of the game)
- 5) White (the player of the white pieces)
- 6) Black (the player of the black pieces)
- 7) Result (the result of the game)

[Event "F/S Return Match"]
[Site "Belgrade, Serbia JUG"]
[Date "1992.11.04"]
[Round "29"]
[White "Fischer, Robert J."]
[Black "Spassky, Boris V."]
[Result "1/2-1/2"]

The same tag name should not appear more than once in a tag pair section. For PGN export format, a single blank line appears after the last of the tag pairs to conclude the tag pair section. This helps simple scanning programs to quickly determine the end of the tag pair section and the beginning of the movetext section.

The **Date** tag value field always uses a standard ten character format: "YYYY.MM.DD". The first four characters are digits that give the year, the next character is a period, the next two characters are digits that give the month, the next character is a period, and the final two characters are digits that give the day of the month. If any of the digit fields are not known, then question marks are used in place of the digits, as long as there are always eight characters (?????.??).

The **Result** tag value is the result of the game. It is always exactly the same as the game termination marker that concludes the associated movetext. It is always one of four possible values: "1-0" (White wins), "0-1" (Black wins), "1/2-1/2" (drawn game), and "*" (game still in progress, game abandoned, or result otherwise unknown). Note that the digit zero is used in both of the first two cases; not the letter "O".

All possible examples:

[Result "0-1"]

[Result "1-0"]

[Result "1/2-1/2"]

[Result "*"]

See section 9 "Supplemental Tag Names" of the "Portable Game Notation Specification and Implementation Guide" for a list of additional tag names.

See section 15 "International Olympic Committee Country Codes" of the "Portable Game Notation Specification and Implementation Guide" for all three-letter country codes.

Movetext Section

After the headers have been created, you must include the notation for the game itself. It can be a single line of text (maximum of 255 characters) or divided into multiple lines (as long as there's a space at the end of each line, if needed, in order to represent the space between the appropriate characters). However, lines with 80 or more printing

characters are strongly discouraged because of the difficulties experienced by common text editors with long lines.

The movetext section is composed of chess moves, move number indications, optional annotations, and a single concluding game termination marker. Because illegal moves are not real chess moves, they are not permitted in PGN movetext. They may appear in commentary, however.

SAN (Standard Algebraic Notation) is a representation standard for chess moves using the ASCII Latin alphabet. SAN as presented in this document uses English language single character abbreviations for chess pieces, although this is easily changed in the source. English is chosen over other languages because it appears to be the most widely recognized.

1. e4 e5 2. Nf3 Nc6 3. Bb5 a6 4. Ba4 Nf6 5. O-O Be7 6. Re1 b5 7. Bb3 d6 8. c3 O-O 9. h3 Nb8 10. d4 Nbd7 11. c4 c6 12. cxb5 axb5 13. Nc3 Bb7 14. Bg5 b4 15. Nb1 h6 16. Bh4 c5 17. dxe5 Nxe4 18. Bxe7 Qxe7 19. exd6 Qf6 20. Nbd2 Nxd6 21. Nc4 Nxc4 22. Bxc4 Nb6 23. Ne5 Rae8 24. Bxf7+ Rxf7 25. Nxf7 Rxe1+ 26. Qxe1 Kxf7 27. Qe3 Qg5 28. Qxg5 hxg5 29. b3 Ke6 30. a3 Kd6 31. axb4 cxb4 32. Ra5 Nd5 33. f3 Bc8 34. Kf2 Bf5 35. Ra7 g6 36. Ra6+ Kc5 37. Ke1 Nf4 38. g3 Nhx3 39. Kd2 Kb5 40. Rd6 Kc5 41. Ra6 Nf2 42. g4 Bd3 43. Re6 1/2-1/2

SAN moves can be as short as two characters (e.g., "d4"), or as long as seven characters (e.g., "Qa6xb7#", "fxg1=Q+"). SAN identifies each piece by a single upper case letter. The standard English values: pawn = "P", knight = "N", bishop = "B", rook = "R", queen = "Q", and king = "K". The letter code for a pawn is not used for SAN moves in PGN export format movetext. However, some PGN import software disambiguation code may allow for the appearance of pawn letter codes. Also, pawn and other piece letter codes are needed for use in some tag pair and annotation constructs.

A basic SAN move is given by listing the moving piece letter (omitted for pawns) followed by the destination square. Capture moves are denoted by the lower case letter "x" immediately prior to the destination square; pawn captures include the file letter of the originating square of the capturing pawn immediately prior to the "x" character.

SAN kingside castling is indicated by the sequence "O-O;" queenside castling is indicated by the sequence "O-O-O." Note that the upper case letter "O" is used, not the digit zero. The use of a zero character is not only incompatible with traditional text practices, but it can also confuse parsing algorithms which also have to understand about move numbers and game termination markers. Also note that the use of the letter "O" is consistent with the practice of having all chess move symbols start with a letter; also, it follows the convention that all non-pawn move symbols start with an upper case letter.

En passant captures do not have any special notation; they are formed as if the captured pawn were on the capturing pawn's destination square. Pawn promotions are denoted by the equal sign "=" immediately following the destination square with a promoted piece letter (indicating one of knight, bishop, rook, or queen) immediately following the equal sign. As above, the piece letter is in upper case.

If the move is a checking move, the plus sign "+" is appended as a suffix to the basic SAN move notation; if the move is a checkmating move, the octothorpe sign "#" is appended instead. There are no special markings used for double checks or discovered checks. There are no special markings used for drawing moves. Neither the checking or checkmating indicators are considered annotation as they do not communicate subjective information. Therefore, they are qualitatively different from move suffix annotations like "!" and "?."

In the case of ambiguities (multiple pieces of the same type moving to the same square), the first appropriate disambiguating step of the three following steps is taken: First, if the moving pieces can be distinguished by their originating files, the originating file letter of the moving piece is inserted immediately after the moving piece letter. Second (when the first step fails), if the moving pieces can be distinguished by their originating ranks, the originating rank digit of the moving piece is inserted immediately after the moving piece letter. Third (when both the first and the second steps fail), the two character square coordinate of the originating square of the moving piece is inserted immediately after the moving piece letter.

Each movetext section has exactly one game termination marker; the marker always occurs as the last element in the movetext. The game termination marker is a symbol that is one of the following four values: "1-0" (White wins), "0-1" (Black wins), "1/2-1/2" (drawn game), and "*" (game in progress, result unknown, or game abandoned). Note that the digit zero is used in the above; not the upper case letter "O". The game termination marker appearing in the movetext of a game must match the value of the game's Result tag pair (While the marker appears as a string in the Result tag, it appears as a symbol without quotes in the movetext).

Import format PGN allows for the use of traditional suffix annotations for moves. There are exactly six such annotations available: "!", "?", "!!", "!!?", "?!", and "??". At most one such suffix annotation may appear per move, and if present, it is always the last part of the move symbol. When exported, a move suffix annotation is translated into the corresponding Numeric Annotation Glyph as described in a later section of this document. For example, if the single move symbol "Qxa8?" appears in an import format PGN movetext, it would be replaced with the two adjacent symbols "Qxa8 \$2."

See section 10 "Numeric Annotation Glyphs" of the "Portable Game Notation Specification and Implementation Guide" for a list of NAGs.

More On The Movetext Section

In PGN import format, tokens in the movetext do not require any specific line justification. In PGN export format, tokens in the movetext are placed left justified on successive text lines each of which has less than 80 printing characters. As many tokens as possible are placed on a line with the remainder appearing on successive lines. A single space character appears between any two adjacent symbol tokens on the same line in the movetext. As with the tag pair section, a single empty line follows the last line of data to conclude the movetext section. Neither the first or the last character on an export format PGN line is a space (this may change in the case of commentary; this area is currently under development).

There are two export format move number indication formats, one for use appearing immediately before a white move element and one for use appearing immediately before a black move element. A white move number indication is formed from the integer giving the full move number with a single period character appended. A black move number indication is formed from the integer giving the full move number with three period characters appended.

All white move elements have a preceding move number indication. A black move element has a preceding move number indication only in two cases: first, if there is intervening annotation or commentary between the black move and the previous white move; and second, if there is no previous white move in the special case where a game starts from a position where Black is the active player. There are no other cases where move number indications appear in PGN export format.

PGN Files In General

According to the Implementation Guide, " PGN data is composed of characters; non-overlapping contiguous sequences of characters form lexical tokens. PGN data is represented using a subset of the eight bit ISO 8859/1 (Latin 1) character set ('ISO' is an acronym for the International Standards Organization). This set is also known as ECMA-94 and is similar to other ISO Latin character sets. ISO 8859/1 includes the standard seven bit ASCII character set for the 32 control character code values from zero to 31."

Comment text may appear in PGN data. There are two kinds of comments. The first kind is the "rest of line" comment; this comment type starts with a semicolon character and continues to the end of the line. The second kind starts with a left brace character and continues to the next right brace character. Comments cannot appear inside any token.

Brace comments do not nest; a left brace character appearing in a brace comment loses its special meaning and is ignored. A semicolon appearing inside of a brace

comment loses its special meaning and is ignored. Braces appearing inside of a semicolon comments lose their special meaning and are ignored.

A period character (".") is a token by itself. It is used for move number indications (see below). It is self terminating.

An asterisk character ("*") is a token by itself. It is used as one of the possible game termination markers (see below); it indicates an incomplete game or a game with an unknown or otherwise unavailable result. It is self terminating.

The left and right bracket characters ("[" and "]") are tokens. They are used to delimit tag pairs. Both are self terminating.

The left and right parenthesis characters ("(" and ")") are tokens. They are used to delimit Recursive Annotation Variations (see below). Both are self terminating.

The left and right angle bracket characters ("<" and ">") are tokens. They are reserved for future expansion. Both are self terminating.

A Numeric Annotation Glyph ("NAG", see below) is a token; it is composed of a dollar sign character ("\$\$") immediately followed by one or more digit characters. It is terminated just prior to the first non-digit character following the digit sequence.

A symbol token starts with a letter or digit character and is immediately followed by a sequence of zero or more symbol continuation characters. These continuation characters are letter characters ("A-Za-z"), digit characters ("0-9"), the underscore ("_"), the plus sign ("+"), the octothorpe sign ("#"), the equal sign ("="), the colon (":"), and the hyphen ("-"). Symbols are used for a variety of purposes. All characters in a symbol are significant. A symbol token is terminated just prior to the first non-symbol character following the symbol character sequence. Currently, a symbol is limited to a maximum of 255 characters in length.

See section 10 " Numeric Annotation Glyphs" of the "Portable Game Notation Specification and Implementation Guide" for a list of NAGs.

End

Stefanie Patterson
Rutherford County Chess Club
2007.04.09