

Endgame Study Database III

by Harold van der Heijden



Introduction

By the end of the 1980's I started my computerized collection of endgames studies. Since then I spent several hours each day to update my database. Such a project is an ongoing process. Not only new studies are being added, but I also update sources, solutions, etc.

The first version of my database (1991) contained 23,358 endgame studies. Ten years later the second version of my database was issued (2000) which had 58,801 studies. The present version has 67,691 studies. Not only almost ten thousand new studies were added, but at least the same number of studies in the database were updated. Especially in recent years many incorrections were reported to me probably due of the widespread availability of endgame table databases (EGTB), chess software and increasingly sophisticated hardware. I estimate the present collection to contain approximately 80% of all endgame studies that were ever published. In the 21st century with its “mega”-statistics the number of 67 thousand studies may not impress everyone. But if you would spend only 3 minutes per study, 4 hours each day, it would take 846 days to play through all studies in this database. Enjoy!

Files

In order to access Endgame Study Database III you'll need software that is able to open Portable Game Notation (PGN) – files. Almost every chess playing software and chess database software can handle this format. When you have access to internet and google “PGN“ and “format” you will find many websites in several languages explaining this format. Unfortunately, most chess programs only allow a limited number of characters for names (“White”) and sources (“Event”) to be entered. In such cases, for studies with many composers, the names appear truncated (although actually more information is entered in the database). And that's also the reason why sources are coded. But it should be noted that PGN is plain ASCII; i.e. you can also open this file in word processors.

The CD-ROM contains several files.

When you only want to access the new Endgame Study Database III (2005), you only need **hhdbIII.pgn** and you can ignore the rest of this paragraph.

hhdbIII.pgn: This is Endgame Study Database III (2005) with 67,691 endgame studies.

readme.pdf: This document.

hhdbII.pgn: Endgame Study Database II (2000) with 58,801 endgame studies (actually 58,796 since the first five are text lines).

hhdbI.pgn: Endgame Study Database I (1991) with 23,358 endgame studies.

xref.xls: A Microsoft® Excel file with crossreferences between **hhdbIII** and **hhdbII**. See file for more explanation.

codes.txt: A plain text file decoding the source codes.

hhdbIIIa.pgn: A version of Endgame Study Database III (2005) with source codes converted. As mentioned above, the source field information may be truncated (but, when accessed by word processor software, all the information is available). Unfortunately, also composer names are sometimes truncated (e.g. studies 11518 and 11519), but for obvious reasons I decided not to use codes for composers names. In both files (hhdbIII.pgn and hhdbIIIa.pgn) composer names are not truncated when accessed by word processors.

Codes

All sources in the database are coded (see codes.txt). Often the first letter of a code refers to the source type: “b” or “c” = book, “t”, “u”, or “v” = magazine, “j” = jubilee tourney, “m” = memorial tourney, “w” = match, “o” = unclear type. This is only a rule of thumb, since this classification is not error-prone (but of course it mainly matters that source-codes are unique). The code.txt file should only be used to recode sources of hhdbIII.pgn or hhdbIIIa.pgn. Probably, most codes also apply to hhdbII.pgn and hhdbI.pgn, but in a few instances the original codes were replaced. Codes that are not present in codes.txt (e.g. b001) are being used by myself for more elaborate coding of secondary sources but do not occur in the files on this CD-ROM.

Apart from source codes there are several additional codes:

@1: second solution (at move 1)

@2: cook; dual solution after move 1

@3: incorrect; White is unable to fulfill the stipulation (in a win study, Black draws or wins; in a draw study, White loses).

@4: “super-cook”. White even wins (in the intended main line) in a draw study.

(@1) or (@2): minor duals.

@1?, @2?, @3?, @4?: suspicious, or supposed bust reported without analysis.

{c}: correction; i.e. original study was incorrect.

{m}: modification; i.e. original study was correct, the improvement has another motivation.

{v}: version (perhaps a correction or a modification).

{tw}: twin study (also triplicates, quadruplicates, etc).

{pl}: plagiarism or accidental recomposition (it does happen!). These studies are “suspect”.

{ph}: posthumously published.

{te}: theoretical ending (i.e. probably not an endgame study).

{cr}: colours reversed (the original stipulation was, for instance: Black to play and win).

{ce}: computer ending (EGTB-derived ending).

{tt}: theme tourney.

{s}: corrected solution (without changing the position).

The GBR-code

GBR-code (after Guy, Blandford and Roycroft) denotes chessboard force. Here it is used to encode the initial position of a study. It's easy to learn and easy to code or decode a position.

The first four digits designate the sum of the number of Q, R, B and S respectively. A white piece counts as 1, a black piece as 3. All combinations can be coded: 0 = no white pieces, no black pieces, 1 = one white piece, no black pieces, 2 = two white pieces, no black pieces, 3 = no white pieces, one black piece, 4 = one white piece, one black piece, 5 = two white pieces, one black piece, 6 = no white pieces, two black pieces, 7 = one white piece, two black pieces, 8 = two white pieces, two black pieces, 9 = other combinations (i.e. promoted pieces in the initial position).

The two digits after the decimal point refer to the number of white pawns and black pawns, respectively.

The GBR-code is extended with a preceding “+” or “=” refering to a win or a draw study, and the positions of the white king and the black king as an extension.

For instance study #35838 has GBR-code [+0432.04a3a6] i.e. a win study with in the initial position no Queens, one white and one black Rook, one black Bishop, two white Knights, no white pawns and four black pawns. The white king is at a3 and the black King at a6.

Although this particular GBR-code has some drawbacks (e.g. identical positions but only vertically mirrored have a different GBR-code), it’s very convenient in simple queries.

Most chess database software allow e.g. to search “Black” for ?0300.10?6?1, which should result in two hits: the famous Saavedra-study and a modification of the same study. The last two question marks in the query are needed to find also mirrored versions.

Most chess database software have the possibility to perform more complex queries, e.g. to find a certain material force during the solution (or sub-lines).

Themes

The present database has no theme index. Classification of a large collection by hand is very complicated, and it would easily take thousands of hours to undertake this for the present database. In addition, themes are usually not well defined. And last but not least: the definition of new themes would make it necessary to do the whole classification all over again.

In my opinion *a priori* classification is therefore inferior to *a posteriori* classification. In the latter case we define a theme and then check the collection for examples. Lewis Stiller and Gady Costeff undertook to program a fantastic tool to query my endgame study database for material force, but also complicated manoeuvres of various kinds: Chess Query Language (CQL). In combination with the database it’s very useful to find examples of studies with a certain theme, e.g. for writing articles or anticipation testing. CQL is freely available on the web with extensive documentation: <http://www.rbnn.com/cql/contents.html>

News

For news about my database (and my latest e-mail address), please visit my homepage: <http://home.studieaccess.nl/heijd336/home.html>