INFOH417 Database System Architecture

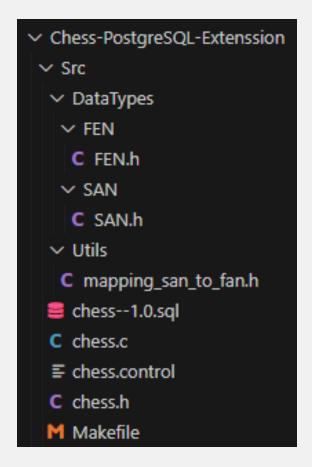
Chess Extension for PostgreSQL

Outline

- Project Structure
- Data types (SAN, FAN)
- Functions (getBoard, getFirstMoves, hasOpening, hasBoard)
- Indexes (Btree for has Opening, GIN for has Board)

Project Structure

- Project Structure
 - o Data types: FEN.h, SAN.h
 - o mapping_san_to_fen.h: A header containing a utility function san_to_fan written with Python embedded code
 - o chess—1.0.sql, chess.h, chess.c, chess.control: files for defining the Postgres extension named chess, version 1.0
 - o makefile: for installing Postgres extensions



Data Types

Standard Algebraic Notation (SAN)

1. Define the SAN type with struct

```
typedef struct {
    char data[MAX_PGN_LENGTH]; // Array to store the SAN string.
} SAN;
```

2. Define two functions parseStr_ToPGN and parsePGN_ToStr to convert a string (written in SAN format) to SAN data type and vice versa.

```
void parseStr_ToPGN(const char *pgn, SAN *game) {
    game->data[MAX_PGN_LENGTH - 1] = '\0'; // Ensure null-termination.
    strncpy(game->data, pgn, MAX_PGN_LENGTH - 1); // Copy the string.
}

void parsePGN_ToStr(SAN *game, char **result){
    *result = (char *)palloc(strlen(game->data) + 1); // Allocate memory for the string strcpy(*result, game->data); // Copy the SAN data.
}
```

3. Define a function truncate san to get N half moves of a SAN (to be used in

getFirstMoves function)
Idea:

- Loop over the input SAN
- Append the SAN token to result
- Count the half moves
- Stop when reaching the desired n half moves.

```
SAN *resultGame = (SAN *) palloc(sizeof(SAN)); // Allocate memory for the result SAN structure
int halfMoveCount = 0; // Counter for the number of half-moves processed.
char *token, *rest, *inputCopy; // Pointers for string tokenization.
bool isMoveNumber: // Flag to check if the token is a move number
inputCopy = palloc(strlen(inputGame->data) + 1);
strcpy(inputCopy, inputGame->data); // Copy the SAN string.
rest = inputCopy; // Set the rest pointer for strtok_r usage
resultGame->data[0] = '\0'; // Initialize the result SAN string
while ((token = strtok_r(rest, " ", &rest)) && halfMoveCount < nHalfMoves)</pre>
    rest = skipCommentsAndAnnotations(rest); // Skip any comments and annotations
    isMoveNumber = (strchr(token, '.') != NULL); // Check if the token is a move number.
    strcat(resultGame->data, token); // Append the token to the result SAN string.
    strcat(resultGame->data, " "); // Append a space after the token
   if (!isMoveNumber)
       halfMoveCount++;
pfree(inputCopy);
if (halfMoveCount < nHalfMoves){
    pfree(resultGame): // Free the result structure if not.
return resultGame; // Return the truncated SAN structure
```

Data Types

Forsyth–Edwards Notation (FEN)

1. Define the FANtype with struct

```
typedef struct
{
    char positions[MAX_FEN_LENGTH];
    char turn;
    char castling[5];
    char en_passant[3];
    int halfmove_clock;
    int fullmove_number;
} FEN;
```

 Define two functions parseStr_TofEN and parseFEN_ToStr to convert a string (written in FEN format) to FEN data type and vice versa.

3. Define a function isvalidFEN to check if the input string has the FEN format

```
int ret;
/* Regular expression pattern for FEN validation */
static const char *fen_pattern = "^[KQRBNPkqrbnp1-8]+(/[KQRBNPkqrbnp1-8]+){7} [wb] (-|[KQkq]+) (-|[a-h36]+) \\d+ \\d+$";

/* Compile the regular expression pattern */
regex_t re;
if (regcomp(&re, fen_pattern, REG_EXTENDED | REG_NOSUB) != 0)
    return false;

/* Execute the regular expression and check if it matches the FEN string */
ret = regexec(&re, fen, 0, NULL, 0);

/* Free the regex resources */
regfree(&re);

/* Temporary fix, always returns true for now */
ret = 0; // TODO: Address the temporary fix.

/* Return true if the FEN string matches the pattern */
return (ret == 0);
}
```

Utility function san_to_fen

Input: (SAN) gamegame Output: (FEN) chessboard

- Use a Python library named chess (installed with pip install chess) to convert SAN to FEN.
- Declare #include <python3.11/Python.h> to use Python embedded code in C code.
- Write the Python embedded code with PyRun_SimpleString, in which we define a function get fen from san.
- In the main C code, get the Python function using PyObject GetAttrString.
- Get the result using PyObject_CallObject, then convert to C string using PyUnicode AsUTF8

```
onst char* san to fan(SAN *gameTruncated)
 PyObject *pModule, *pFunc, *pArgs, *pValue;
 const char *result = NULL;
 Py_Initialize();
 // Define the Python function for SAN to FEN conversion.
 PyRun SimpleString(
     "import io\n"
      "def get fen from san(san):\n"
          if not san.strip():\n"
             return 'rnbqkbnr/pppppppp/8/8/8/8/PPPPPPPPPRNBQKBNR w KQkq - 0 1' # Default FEN for empty input\n
          game = chess.pgn.read_game(io.StringIO(san))\n"
           board = game.board()\n"
           for move in game.mainline_moves():\n"
          return board.fen()\n"
 pModule = PyImport_AddModule("__main__");
 if (pModule != NULL) {
     pFunc = PyObject_GetAttrString(pModule, "get_fen_from_san");
     if (pFunc && PyCallable_Check(pFunc)) {
         pArgs = PyTuple New(1);
         PyTuple_SetItem(pArgs, 0, PyUnicode_FromString(gameTruncated->data));
         pValue = PyObject_CallObject(pFunc, pArgs);
         Py_DECREF(pArgs);
         if (pValue != NULL) {
             result = PyUnicode AsUTF8(pValue);
             Py DECREF(pValue);
         } else {
             Py DECREF(pFunc);
             ereport(ERROR, (errmsg("Call to 'get_fen_from_san' failed")));
         if (PyErr_Occurred())
             PyErr Print();
         ereport(ERROR, (errmsg("Cannot find function 'get fen from san'")));
     Py XDECREF(pFunc);
     ereport(ERROR, (errmsg("Failed to load '__main__' module")));
 Py Finalize();
 return result;
```

getBoard(get_board_state)

Input:

- (SAN) chessgame
- (int) N: number of half moves

Output:

- (FEN) board state after N half moves

Idea:

- Use the function truncate_san (defined in SAN.h) to get the truncated SAN afer N half moves

- Convert to FEN type using the function san to fan (defined in mapping san to fan.h)

```
SAN *gameTruncated, *game;
int half moves;
const char *fenConversionStrResult;
if (PG ARGISNULL(0) || PG ARGISNULL(1))
   ereport(ERROR, (errmsg("get_board_state: One of the arguments is null")));
game = (SAN *) PG_GETARG_POINTER(0);
half moves = PG GETARG INT32(1);
   ereport(ERROR,(errmsg("get_board_state: Non-positive number of half moves")));
gameTruncated = truncate_san(game, half_moves);
if (gameTruncated == NULL)
   ereport(ERROR, (errmsg("Game is incomplete or shorter than the requested number of half-moves")))
fenConversionStrResult = san_to_fan(gameTruncated);
if (fenConversionStrResult == NULL) {
   ereport(ERROR, (errmsg("No FEN result returned from mapping san to fen")));
fen = (FEN *)palloc(sizeof(FEN));
parseStr ToFEN(fenConversionStrResult, fen);
PG RETURN POINTER(fen);
```

getFirstMoves (get_firstMoves)

Input:

- (SAN) chessgame
- (int) N: number of half moves

Output:

- (SAN) chessgame after N half moves

- Similar to the function get board state, but do not need to convert the result to FEN.
- Return the output right after using the function truncate san.

```
Datum get_FirstMoves(PG_FUNCTION_ARGS)
{
    SAN *result, *inputGame;
    int nHalfMoves;

    if (PG_ARGISNULL(0) || PG_ARGISNULL(1))
        ereport(ERROR, (errmsg("get_FirstMoves: One of the arguments is null")));

    inputGame = (SAN *) PG_GETARG_POINTER(0);
    nHalfMoves = PG_GETARG_INT32(1);

    if (nHalfMoves < 0)
        ereport(ERROR,(errmsg("get_FirstMoves: Non-positive number of half moves")));

    result = truncate_san(inputGame, nHalfMoves);

    if (result == NULL)
        ereport(ERROR, (errmsg("Game is incomplete or shorter than the requested number of half-moves")));

    PG_RETURN_POINTER(result);
}</pre>
```

hasOpening(has opening)

Input:

- (SAN) chessgame 1
- (SAN) chessgame 2

Output:

- (boolean) **True** if chessgame 1 starts with the same set of moves as chessgame 2.

- Use the C built-in function strncmp to compare chessgame 1 with chessgame 2 in terms of beginning moves, where the compared string length is the length of chessgame 2.
- Get True if chessgame 2 is included in the beginning moves of chessgame 1.

```
Datum has_opening(PG_FUNCTION_ARGS)
{
    int opening_length;
    SAN *game1, *game2;

    if (PG_ARGISNULL(0) || PG_ARGISNULL(1))
        ereport(ERROR, (errmsg("has_opening: One of the arguments is null")));

    game1 = (SAN*)PG_GETARG_POINTER(0);
    game2 = (SAN*)PG_GETARG_POINTER(1);

    opening_length = strlen(game2->data);

    if (opening_length > 0 && strncmp(game1->data, game2->data, opening_length) == 0)
        PG_RETURN_BOOL(true);

    PG_RETURN_BOOL(false);
}
```

hasBoard (has_Board)

Input:

- (SAN) chessgame
- (FEN) chessboard
- (int) N half moves

Output:

- (boolean) **True** if chessgame contains chessboard in its first N half-moves.

- Truncate the chessgame after N half moves with the function truncate san.
- Convert to FEN with the function san to fen
- Use the C built-in function strcmp to compare the converted FEN with input chessboard
- Get True if they are the same

```
FEN *current board, *input board;
SAN *gameTruncated, *input_game;
int input half moves;
bool positions_match;
const char *fenConversionStrResult;
if (PG ARGISNULL(0) || PG ARGISNULL(1) || PG ARGISNULL(2))
    ereport(ERROR, (errmsg("has Board: One of the arguments is null")));
input game = (SAN*) PG_GETARG_POINTER(0);
input_board = (FEN*) PG_GETARG_POINTER(1);
input_half_moves = PG_GETARG_INT32(2);
if (input half moves < 0)
    ereport(ERROR,(errmsg("hasBoard: Non-positive number of half moves")));
gameTruncated = truncate_san(input_game, input_half_moves);
if (gameTruncated == NULL)
    ereport(ERROR, (errmsg("Game is incomplete or shorter than the requested number of half-moves")));
fenConversionStrResult = san_to_fan(gameTruncated);
if (fenConversionStrResult == NULL) {
    ereport(ERROR, (errmsg("No FEN result returned from mapping san to fen")));
current_board = (FEN *)palloc(sizeof(FEN));
parseStr ToFEN(fenConversionStrResult, current board);
positions_match = strcmp(input_board->positions, current_board->positions) == 0;
PG RETURN BOOL(positions match);
```

B-Tree Index

6 operations: less than (san_lt), equal (san_eq), greater than (san_gt), comparator (san_cmp), like (san_like), not like (san_not_like)

Beforehand, define a utility function san_compare to compare two SAN strings Input: (SAN) a, (SAN), b
Output: (int) -1, 0, or 1

Idea: Simply use the C built-in function strncmp to compare two SAN strings, where MAX_PGN_LENGTH is a pre-defined variable.

```
static int san_compare(SAN *a, SAN *b)
{
   int cmp_result;

   cmp_result = strncmp(a->data, b->data, MAX_PGN_LENGTH);

   if (cmp_result < 0)
        return -1;
   else if (cmp_result > 0)
        return 1;
   else
        return 0;
}
```

This function will be used in B-tree operations.

B-Tree Index

```
Input: (SAN) a, (SAN), b
Output: (boolean) True if SAN string a is less than SAN string b
Get the result by checking if the returned value of
san_compare(a, b) is less than 0
```

```
The same as san_lt, except that the checking condition is equal to 0
```

```
The same as san_lt, except that the checking condition is greater than 0
```

```
Datum san_lt(PG_FUNCTION_ARGS)
{
    SAN *a, *b;
    if (PG_ARGISNULL(0) || PG_ARGISNULL(1))
        ereport(ERROR, (errmsg("san_lt: One of the arguments is null")));
    a = (SAN *) PG_GETARG_POINTER(0);
    b = (SAN *) PG_GETARG_POINTER(1);

PG_RETURN_BOOL(san_compare(a, b) < 0);
}
```

```
Datum san_eq(PG_FUNCTION_ARGS)
{
    SAN *a, *b;

    if (PG_ARGISNULL(0) || PG_ARGISNULL(1))
        PG_RETURN_BOOL(false);

    a = (SAN *) PG_GETARG_POINTER(0);
    b = (SAN *) PG_GETARG_POINTER(1);

PG_RETURN_BOOL(san_compare(a, b) == 0);
}
```

```
Datum san_gt(PG_FUNCTION_ARGS)
{
    SAN *a, *b;

    if (PG_ARGISNULL(0) || PG_ARGISNULL(1))
        ereport(ERROR, (errmsg("san_gt: One of the arguments is null")));
    a = (SAN *) PG_GETARG_POINTER(0);
    b = (SAN *) PG_GETARG_POINTER(1);

    PG_RETURN_BOOL(san_compare(a, b) > 0);
}
```

B-Tree Index

san_cmp
Get the result directly from the returned value of the function
san compare(a, b)

san like

Use Postgres built-in function DirectFunctionCall2 (defined in fmgr.h) to call the function textlike, taking

san_not_like
The same as san like, but return negation.

```
Datum san_cmp(PG_FUNCTION_ARGS)
{
    SAN *a, *b;
    int cmp_result;

    if (PG_ARGISNULL(0) || PG_ARGISNULL(1))
        ereport(ERROR, (errmsg("san_cmp: One of the arguments is null")));
        a = (SAN *) PG_GETARG_POINTER(0);
        b = (SAN *) PG_GETARG_POINTER(1);
        cmp_result = san_compare(a, b);
        PG_RETURN_INT32(cmp_result);
}
```

GIN Index

The idea behind this GIN index is indexing and querying on chess game positions by extending the GIN index for the chess game type.

8 functions to try to implement this index:

fens_from_san

Extracts FEN strings from a SAN representation of a chess game

Input: (SAN) chessgame Output: (FEN) chessboard

gin_compare

Compares two text values for GIN indexing

Input: Two text values (keys)
Output: (int) -1, 0, or 1 for the ordering of the two keys.

```
um fens_from_san(PG_FUNCTION_ARGS){
int32 *nkeys;
ArrayBuildState *astate;
SAN *san. *gameTruncated:
int i;
const char *fenConversionStrResult;
san = (SAN *) PG_GETARG_POINTER(0);
nkeys = (int32 *) PG_GETARG_POINTER(1);
*nkeys = 0;
astate = NULL;
while (true) {
   gameTruncated = truncate_san(san, i);
   if (gameTruncated == NULL)
   fenConversionStrResult = san_to_fen(gameTruncated);
   if (fenConversionStrResult == NULL)
        ereport(ERROR, (errmsg("No FEN result returned from mapping san to fen")));
   astate = accumArrayResult(astate, CStringGetTextDatum(fenConversionStrResult),
                              false, TEXTOID, CurrentMemoryContext);
    (*nkeys)++;
if (*nkeys > 0) {
   Datum *keys = (Datum *) palloc(*nkeys * sizeof(Datum));
    for (i = 0; i < *nkeys; i++) {
        keys[i] = astate->dvalues[i];
   PG_RETURN_POINTER(keys);
   PG_RETURN_NULL();
```

Indexes GIN Index

gin extract value

Prepares indexable keys for GIN indexing from a SAN

Input: (SAN) chessgame

Output: Array of keys (Datum)

gin extract query

Extracts a query key from a FEN type for GIN indexing searches

Input: (FEN) chessboard, array of

keys (Datum), search mode

Output: Array of keys (Datum)

gin consistent

Checks if indexed keys are consistent with a query key in GIN index searches

Input: (Boolean Array) check, (Query Key) query, (Number of Keys) nkeys, (Recheck Flag) recheck, (Array of Query Keys) query Keys Output: (Boolean) True if the result of the comparison between the query key and indexed keys is less than 0; otherwise, False

cum gin_extract_query(PG_FUNCTION_ARGS)

if (PG_ARGISNULL(0) || PG_ARGISNULL(1) || PG_ARGISNULL(2) || PG_ARGISNULL(3) ||

ereport(ERROR, (errmsg("gin extract query: One of the arguments is null")))

PG ARGISNULL(4) || PG ARGISNULL(5) || PG ARGISNULL(6)) {

FEN *itemValue;

Datum *keys, query;

int32 *nkeys, *searchMode:

query = PG_GETARG_DATUM(0);

PG_RETURN_POINTER(keys);

nkeys = (int32 *) PG_GETARG_POINTER(1);
itemValue =(FEN *) DatumGetPointer(query);
searchMode = (int32 *) PG_GETARG_POINTER(6);

*searchMode = GIN SEARCH MODE DEFAULT;

keys = (Datum *) palloc(*nkeys * sizeof(Datum));

keys[0] = CStringGetTextDatum(itemValue->positions);

```
tum gin_extract_value(PG_FUNCTION_ARGS)
 int32 *nkeys;
 bool **nullFlags;
 if (PG_ARGISNULL(0) || PG_ARGISNULL(1) || PG_ARGISNULL(2))
     ereport(ERROR, (errmsg("gin extract value: One of the arguments is null")))
 san = (SAN *) PG GETARG POINTER(0);
 nkeys = (int32 *) PG GETARG POINTER(1);
 nullFlags = (bool **) PG_GETARG_POINTER(2);
 keys = (Datum *) DirectFunctionCall4Coll(fens from san,
                                                 PG_GET_COLLATION(),
                                                 PointerGetDatum(san),
                                                 PointerGetDatum(nkeys),
                                                 BoolGetDatum(false),
                                                 PointerGetDatum(NULL));
 *nullFlags = NULL;
 PG RETURN POINTER(keys);
```

```
um gin consistent(PG FUNCTION ARGS)
bool *check = (bool *) PG_GETARG_POINTER(0);
Datum query = PG GETARG DATUM(2);
int32 nkeys = PG GETARG INT32(3);
bool *recheck = (bool *) PG GETARG POINTER(5);
Datum *queryKeys = (Datum *) PG GETARG POINTER(6);
char *queryFenStr = text to cstring(DatumGetTextP(query));
parseStr ToFEN(queryFenStr, &queryFen);
pfree(queryFenStr);
for (int i = 0; i < nkeys; i++) {
   if (check[i]) {
        char *keyFenStr = text to cstring(DatumGetTextP(queryKeys[i])
        parseStr ToFEN(keyFenStr, &keyFen);
        pfree(keyFenStr);
        if (strcmp(queryFen.positions, keyFen.positions) == 0) {
            *recheck = true;
            PG RETURN BOOL(true);
*recheck = true;
PG RETURN BOOL(false);
```

Indexes GIN Index

gin tri consistent

Performs a ternary consistency check for GIN index operations Input: (Array of Ternary Values) check, (Query Key) query, Number of Keys) nkeys, (Array of Query Keys) queryKeys Output: (Boolean) True if the result of **the** comparison between the query key and indexed keys is less than 0; otherwise, False.

has_board_fn_operator and fen_in_san_eq Both determine if a given FEN type matches any board state in a SAN type

Input: (FEN) a, (SAN) b.

Output: (Boolean) True if the result of the comparison between FEN string a and the SAN string from the indexed key (using san to fen) is less than 0; otherwise, False.