Assignment 2
Chengyu Jia, Zesheng Ye

size of bit-slices: 2 bytes
size of tuple and page signatures: tuple signature: 9bytes; page signature: 746bytes
(The formula of the tuple signature is "tm=(int)(log2*log2 * nattrs * logF);". The formula of the page signature is "pm=(int)(log2*log2 * nattrs*capacity * logF);")

1. not using indexing (use x as the signature type):
Query: "./select X 1000555,?,?,?,? x "
Query Stats:
total pages read: 13

Analysis: Because we don't use any signature, So the "scanAndDisplayMatchingTuples"function scan through the whole data file to find the match tuples. The result is there only one tuple match the query, and total data pages are 13. So false match pages number is 12. And this method need to read whole data file, so the efficiency is not good.

2. open query (?,?,...,?):
Query: "./select X ?,?,?,?,? t"
Query Stats:
total pages read: 16

Query: "./select X ?,?,?,?,? p"
Query Stats:
total pages read: 16

Query: "./select X ?,?,?,?,? b"
Query Stats:
total pages read: 13

Analysis: Becuase the query is "?,?,?,?,?", so every tuple signature is the subset of the query's signature. So the "scanAndDisplayMatchingTuples" function need to scan the whole tuple file to find the match tuple. And include scan the signature pages, the scanning page number is bigger than the data page number. So this open query method is bad, but bit-slices signature need reading a slightly less pages than other two signature methods.

3. query with one solution (100001,?,...,?)
Query: "./select X 1000001,?,?,?,? t"
Query Stats:
total pages read: 4
Analysis: This signature strategy need scan whole tuple signature(3 pages), then find one tuple signature match the query signature. Then scan this particular data page to find there are match tuple or not. The time efficiency is quit good.

Query: "./select X 1000001,?,?,?,? p"
Query Stats:
data pages read: 4
Analysis: This signature strategy need scan whole page signature(3 pages), then find one page signature match the query signature. Then scan this particular data page to find there are match tuples or not. The time efficiency is quit good.

Query: "./select X 1000001,?,?,?,? b"
Query Stats:
data pages read: 4
Analysis: This signature strategy need find the query page signature's none zero positions, and then use a Bits type "matches" to do the "or operating" with those positions' bit-slices signature in the "bsigf". In this case the program need to read 3 bit-slices signature pages to find the final match data pages(there only have one). Then scan this particular data page to find there are match tuples or not. The time efficiency is quit good.

4. query with many solutions (?,?,a3-001,...,?) :
Query:"./select X ?,?,a3-001,?,? t"
Query Stats:
total pages read: 9
Analysis: Analysis: This signature strategy need scan whole tuple signature(3 pages), then find one tuple signature match the query signature. Then scan this particular data page to find there are match tuple or not. In this query case, the time efficiency is not as good as the other two method, because there is one false match page.

Query:"./select X ?,?,a3-001,?,? p"
Query Stats:
total pages read: 8
Analysis: This signature strategy need scan whole page signature(3 pages), then find one page signature match the query signature. Then scan this particular data page to find there are match tuples or not. In this query case, because the page signature has long enough bits to be putted the hash values. So there are few chances to be false matched. The time efficiency is good.

Query:"./select X ?,?,a3-001,?,? b"
Query Stats:
total pages read: 8
Analysis: The signature strategy need to scan "k" bit-slices signature, which are distributed in 3 bit-slices signature pages. Then the program need to scan particular data pages(5 pages) which match the query page signature. The time efficiency is good.

5.query with multiple values (100001,?,a3-001,...,?):
Query: "./select X 1000001,?,a3-001,?,? T"
Query Stats:
total pages read: 4

Query: "./select X 1000001,?,a3-001,?,? p"
Query Stats:
total pages read: 4

Query: "./select X 1000001,?,a3-001,?,? b"Query Stats:
total pages read: 4

Analysis: In this query case, the situations are same like previous one. And generally, the program need read whole pages for the tuple signature file and page signature file, and bit-slices signature may need read less pages. And some times, the page signature file may bigger than the tuple signature file, So we need reading more page signature pages.