

EECS 498 W26 System Design of a Search Engine

06 HashBlob and HashFile

Due Sunday, February 8, 2026, 11:59 pm

In this assignment, you will convert your hash table of words and occurrences from the previous assignment into what I've called a hash blob, a contiguous blob of memory that still works as a hash table but can be relocated anywhere in memory, making it suitable for memory-mapping as a file.

You will build two utilities. HashBlob will build a hash table from a text file, convert it to a hash blob, search it, and optionally write the blob out to a file. HashFile will search a memory-mapped hash blob stored in a file.

Starter files

The starter files in HashTableStarter.zip are the same as for the previous assignment.

The assignment

You are to complete the code in Common.h, Common.cpp, HashTable.h, and HashBlob.h, allowing the HashBlob and HashFile utilities to be built.

HashBlob

Usage: HashBlob [-L] wordsin.txt [blobfile]

Reads words from wordsin.txt, creates a hashblob, then searches it for words read from cin.

If blobfile is specified, it writes the hashblob into the blobfile.

-v means verbose output

-L means read whole lines as words

Here is an example of how it behaves. (User input is in bold.)

```
% ./HashBlob Hamlet.txt
now is the time for all good men
36  now
299  is
929  the
32   time
162  for
83   all
71   good
9    men
```

If you give HashBlob a second argument, it will write the blob into a file.

```
% ./HashBlob BigJunkHtml.txt HashBlob.bin < /dev/null
```

HashFile

```
Usage: HashFile [ -v ] blobfile
```

Map the blobfile into memory and then search it for words read from cin.

-v means verbose output -v means verbose output

Here is an example of how it behaves. Notice you get the same result searching the hash file as you did searching the hash blob or the hash table.

```
% ./HashFile HashBlob.bin
now is the time for all good men
36  now
299  is
929  the
32   time
162  for
83   all
71   good
9    men
```

Autograder

These are the files you will submit to the autograder at <https://autograder.io/web/project/3491>.

1. Common.cpp
2. Common.h
3. HashTable.h
4. HashBlob.h