

CMPS340 File Processing
HW #2 Fall 2008
Due: 12:00noon, Monday, October 20

1. Below is depicted a data file. NAME is a *key* field and MAJOR is the *ordering* field. (The values in the fields are strings; the usual alphabetic ordering is used in determining which is the lesser of two strings, including for the field BIRTH_MONTH.)

For *each of the three fields*, construct a flat (as opposed to tree-structured) index. For “references/pointers”, use bucket numbers in the indexes on NAME and MAJOR, but use NAME values in the index on BIRTH_MONTH.

(**Note:** The buckets are numbered to correspond to their logical order; this numbering has nothing to do with the physical locations of the buckets on disk.)

In constructing an index, make it a *primary* index (containing only one reference/pointer to each bucket), if possible. Otherwise, if it is possible, make it a *clustering* index (containing only one reference/pointer for each distinct value occurring in the indexing field). Otherwise, make it a *secondary* index. Assume that, in a secondary index based upon a non-key field, there is enough space for two references/pointers in each entry of the index. (Hence, if there exist more than two records with the same value in the indexing field, multiple entries for that value should exist in the index.)

NAME	MAJOR	BIRTH_MONTH
Joy	Art	Feb
Carol	Art	Mar
Bruce	Art	Dec
Mary	Bio	Mar
Joe	Bio	May

Bucket #1

NAME	MAJOR	BIRTH_MONTH
Jim	Bio	Jan
Kate	Chem	Jun
Marie	Chem	May

Bucket #2

NAME	MAJOR	BIRTH_MONTH
George	CIS	Apr
Chris	CIS	Sep
Renee	CS	Sep
Linda	CS	Jul
Kurt	Engl	Oct

Bucket #3

NAME	MAJOR	BIRTH_MONTH
Bill	Engl	Nov
Erin	Engl	Dec

Bucket #4

NAME	MAJOR	BIRTH_MONTH
Cathy	Engl	May
Bob	Hist	Aug
Tom	Hist	Apr
Tina	Math	Oct

Bucket #5

NAME	MAJOR	BIRTH_MONTH
Marge	Math	Jan
Mike	Math	Nov
Janice	Phys	Jul

Bucket #6

NAME	MAJOR	BIRTH_MONTH
Jean	Phys	Feb
Joyce	Psyc	Aug
Ernie	Psyc	Mar
Bert	T/RS	Apr
Julie	T/RS	Jul

Bucket #7