COMP 212 Spring 2015
Lab 08

1 Introduction

In this lab, you will experiement with the module system. In lecture, we discussed the module system as a way to clearly mark and enforce abstraction boundaries. In this lab, you will use modules and abstract types from the perspective of both a client and an implementer.

2 Dictionaries

A *dictionary* is a datastructure that acts as a finite map from *keys* and to *values*. We represent a dictionary by a type

```
('k, 'v) dict
```

`dict` is a type constructor that takes two type arguments (unlike e.g. `'a list`, which takes only one). The first, `'k`, represents the type of keys, whereas the second, `'v` represents the type of values. For example, an `(int,string) dict` maps integers to strings.

There are many possible implementations of dictionaries, including using lists, trees, and functions. Since there are so many different ways to implement dictionaries, it would be nice if we could have an abstract interface to them that is the same regardless of the underlying implementation. This is where the module system comes in.

In `LabDict.sig`, we have provided the following signature for you to implement for dictionaries (see Figure 1).

The type and values in it have the following specifications:

- `('k, 'v) dict` is an abstract type representing the type of the dictionary. Note that it is parametrized over two different types—`'k`, the type of keys, and `'v`, the type of values.
- `empty` is a dictionary that contains no mappings.
- `insert` is a function that takes a comparison function for keys[1], a dictionary, and a key-value pair and returns the dictionary with the mapping added. If the key is already in the dictionary, the new value supersedes the old one.

[1]You may feel a bit uncomfortable adding the comparison function as an argument to each of these functions, instead wondering why we cannot just abstract over it somehow. If so, good! We will go over this more in lecture tomorrow.

```
signature LABDICT =
sig

  (* We model a dictionary as a set of key-value pairs written k ~ v:
     (k1 ~ v1, k2 ~ v2, ...) *)
  type ('k, 'v) dict

  (* the empty mapping *)
  val empty : ('k, 'v) dict

  (* insert cmp (k1 ~ v1, ..., kn ~ vn) (k,v)
        == (k1 ~ v1, ..., ki ~ v,...) if cmp(k,ki) ==> EQUAL for some ki
     or == (k1 ~ v1, ..., kn ~ vn, k ~ v) otherwise
     *)
  val insert : ('k * 'k -> order) -> ('k, 'v) dict -> ('k * 'v) -> ('k, 'v) dict

  (* lookup cmp (k1 ~ v1,...,kn ~ vn) k == SOME vi
                                           if cmp(k,ki) ==> EQUAL for some ki
                                        == NONE otherwise
     *)
  val lookup : ('k * 'k -> order) -> ('k, 'v) dict -> 'k -> 'v option

end
```

Figure 1: Dictionary Signature

- `lookup` is a function that takes a comparison function, a dictionary, and a key, and returns `SOME v` if that key maps to the value `v` in the dictionary, or `NONE` if there is no mapping from the key.

We have called this signature `LABDICT` because it is a version of dictionaries that is small enough for you to implement in lab; a real dictionary library would provide more operations.

3 Implementation: Binary Search Trees

First, you will implement dictionaries using a binary search tree (BST). Recall the discussion of binary search trees from when we implemented mergesort on trees: the key invariant is that, for every `Node(l,x,r)`, everything in `l` is less than or equal to `x`, and everything in `r` is greater than or equal to `x`.

To implement dictionaries, we will store both a key and a value at each node, using the following datatype:

```
datatype ('k, 'v) tree =
    Leaf
  | Node of ('k, 'v) tree * ('k * 'v) * ('k, 'v) tree
```

In `Node(l,(k,v),r)`, `k` is the key and `v` is the value. Every key in `l` should be less than or equal to `k`, and every key in `r` should be greater than `k`. That is, the keys satisfy the BST invariant; the values are just along for the ride.

Task 3.1 Create a file `TreeDict.sml` in which you implement a structure `TreeDict` matching the signature `LABDICT`. You should use the datatype above as the internal representation of a dictionary. Make sure you ascribe the signature LABDICT to make the type `dict` abstract!

To test your implementation, you can run the command

```
- CM.make "sources.cm";
```

from the REPL. Note that your code will not compile until you make `TreeDict.sml` and put a module in it.

To create tests for your code inside the SML file, you can do them normally as we have been doing all semester. You should put them inside the `TreeDict` structure.

To test your code from the REPL, you will need to refer to functions inside your `TreeDict` structure as components of the module. (i.e. as `TreeDict.<function_name>` where `<function_name>` is the name of the function you want to run). Recall that you can only refer to functions that have been defined in the signature.

Some notes about the compilation manager: If you compile your code using `CM.make`, the compilation manager will compile all of the files specified in the `.cm` file. One way to work is to use `CM.make` every time you want to compile.

However, this has the disadvantage that none of the project loads if there is a compilation problem anywhere, which can make debugging harder—you can't use the REPL to play around. An alternative is to `CM.make` when you first start working on a module, assuming your initial state is one where the `CM.make` succeeds (this is generally true for the support code we hand out). Then you can reload the file containing the module you are currently working on with `use` after you make updates to it. Note that this will shadow the modules in that file, and thus **not** update the modules "downstream". However, it is useful if you are using emacs, and like to use the emacs command to load the current buffer: you can load one module repeatedly and use the REPL to test it. It's also useful if your current implementation of the module has a bug that causes later files in the `.cm` file to fail to compile. But when when you are done working on the module, you will want to run `CM.make` again to reload all the downstream modules, so that they refer to your new implementation.

Have the TAs check your code before continuing!

4 Client: Word counts

Now you will write some client code in the module `Count` in `count.sml`.

We use a `(string,int) TreeDict.dict` sorted according to `String.compare` to map words to numbers.

Task 4.1 Write a function

```
val increment : (string,int) TreeDict.dict
                -> string
                -> (string,int) TreeDict.dict
```

such that `increment d w` increments the number associated with the word `w` if `w` occurs in `d`, or maps `w` to `1` if it does not already occur.

Task 4.2 Why should or shouldn't `increment` be in the client code, as opposed to the implementation of dictionaries?

Task 4.3 Redo the Twitter problem from COMP 211 Homework 3: write a function

```
count : string list
       -> (string,int) TreeDict.dict
       -> (string,int) TreeDict.dict
```

The input to `count` is a list of tweets, where each tweet is a string, and a dictionary mapping words to the frequencies with which they have appeared. The function should return an updated dictionary, which includes the counts for all of the words in any tweet in the input. Use the provided `words` function to divide a string into a list of strings, where each string in the result is a single word.

For example:

```
val test_dict = Count.count ["hi there",
                              "how are you today",
                              "I'm fine how are you"]
                             TreeDict.empty
val SOME 1 = TreeDict.lookup String.compare test_dict "hi";
val SOME 1 = TreeDict.lookup String.compare test_dict "there";
val SOME 2 = TreeDict.lookup String.compare test_dict "how";
val SOME 2 = TreeDict.lookup String.compare test_dict "are";
val SOME 2 = TreeDict.lookup String.compare test_dict "you";
val SOME 1 = TreeDict.lookup String.compare test_dict "today";
val SOME 1 = TreeDict.lookup String.compare test_dict "I'm";
val SOME 1 = TreeDict.lookup String.compare test_dict "fine";
```

One way to do this is to define a helper function or two recursively. Or, if you're feeling fancy, use `List.foldr` to define the function in one line of code.