

Programming assignment – Sparse Vector

Traditionally vectors are implemented using an array data structure (static or dynamic). In this assignment you will use linked-list data structure to obtain memory-efficient implementation of vectors. The assignment should give you an additional practice with pointers and dynamical memory allocation. The goal is to implement several functions that manipulate nodes in a linked list. These functions include the traditional list operations like inserting, removing, searching, as well as several vector-related functions – scalar product and addition.

Here is a structure to represent a linked list node

```
struct ElementNode {
    int    data;
    int    pos;
    struct ElementNode *next;
};
```

Integer **data** is the value which is stored in the node. Compared to the traditional node ours has an additional field – **pos**, which represents element's index in the vector. Since the nodes are to be used in a singly linked list, we have a pointer to the next element **next**.

Opaque pointers

First of all, if you were using regular C-style pointers, you'd put the above definition of **ElementNode** in the beginning of **spvector.h** and your functions would look like:

```
int scalar_product(ElementNode const *p_e1, ElementNode const *p_e2);
```

For this assignment you are required to use opaque pointers – here is a roadmap to convert existing code that uses pointers to opaque pointers. I'll also explain why regular pointers are bad.

1. move struct **ElementNode** definition into **spvector.c** (thus it is hidden from the client)
2. in **spvector.h** add

```
typedef struct ElementNode* ElementNode_handle;
typedef const struct ElementNode* ConstElementNode_handle;
```

(see source code for explanation) note that when compiler reads the above line, it DOES NOT know the definition of **ElementNode** – **spvector.c** has not been seen yet (make sure you understand why). Therefore **ElementNode** is an incomplete type. The rule in C/C++ is that pointers to incomplete types are legal (but dereferencing of those pointers is not)
3. rewrite all functions to use **ElementNode_handle** instead of **ElementNode*** and **ConstElementNode_handle*** instead of **ConstElementNode****. Note that inside **spvector.c** you may use **ElementNode** type and **ElementNode** structure fields, since by that time compiler will have full information about **ElementNode** structure. Note that inside **spvector.h** you may NOT use **ElementNode*** and **ElementNode_handle** interchangeably.

Example: this is legal in **spvector.c**, but not **spvector.h** and **driver.c**:

```
ElementNode p = ...;
p.pos = 5;
```

So what?

As we mentioned, since **->next** is illegal in the driver, client cannot modify keys and linked list is safe. To make modification of **next** value possible, implement **setNext(...)** function. Notice how much similar opaque pointers to private data in a C++ class!

Opaque pointers are more than just safety. They also provide encapsulation. Here is a typical example: in a big project we have a struct **ElementNode** in a file **spvector.h** – so its definition is visible to everyone. **ElementNode** uses **next** and **pos** names for its fields. Since **next** and **pos** names are visible to the client (like public data in a C++ class) other files may use them (and will do that, even though methods like **getNext()** and **getpos()** are implemented). Then the maintainer of **spvector.h** decides to change the names to **Next** and **Pos** (or may be change their types or eliminate them completely). Now every file that uses **spvector.h** is broken and has to be updated, which may be a lot of work. Opaque pointers help to avoid this kind of problems by **forcing** the clients to use available getters and setters **getNext()** and **getpos()** instead of directly accessing the data. So now the change will proceed like this:

- implementor changes `next` to `Next`
- updates `getnext()` from

```
... getnext() return next;
```

to

```
... getnext() return Next;
```
- recompiles the project

that's all, no other code is effected since everyone was using `getnext()` instead of `next`.

back to sparse vectors

These nodes may be used to represent a very long vector of integers most of which are 0's. Example vector (0,0,0,1,0,0,0,0,0,0,0,2) is a vector with 1 at position 3 and 2 at position 11 (counting from 0). When stored in an array, the array size should be at least 12 (which gives $12 * 4 = 48$ bytes). When stored as a list, the vector looks like (1,3)->(2,11), where the first number in the pair is the value and second is position. So that the total amount of memory used is 2 nodes, which is $2 * (4 + 4 + 4) = 24$ bytes only.

The idea is to save storage space by never storing 0's. Make sure that all functions that modify the vector test for 0 before actually writing into it. Note that insert function may be used to overwrite a value, that is `insert_element(list,pos,val)`

- creates a new node if no node at position `pos` exists
- just modifies the data if a node at position `pos` exists
- deletes a node at position `pos` exists if `val=0`

This type of vector is usually known as *sparse vector*.

List representation has only one drawback – it doesn't remember the length of the original vector, e.g. both (0,0,0,1,2,0,0,0) and (0,0,0,1,2,0,0,0,0,0,0,0,0,0) have the same list representation (1,3)->(2,4). Functions that require size information will have an additional parameter – dimension of the vector, see `printf_elements`.

Functions to be implemented are insert (ordered by position), delete at position.

Because lists with positions may be viewed as vectors, there are additional functions that may be useful:

- `get` the value at the given position – similar to index operator.
- vector addition
- scalar multiplication

To submit

You must submit electronically through submission

- implementation file `spvector.c`

you do not have to zip if there only one file.