# 27.09.21

## Data structures

### Python

As reported on [Faster Lookups In Python. Comparison of dictionaries and lists | by Seyma Tas | Towards Data Science](https://towardsdatascience.com/faster-lookups-in-python-1d7503e9cd38), the length of the dictionary doesn’t affect the lookup time, indeed, with the lists do.

See examples.

### C++

I can see this question on StackOverflow: [c++ - Is there a more efficient implementation for a bidirectional map? - Stack Overflow](https://stackoverflow.com/questions/21760343/is-there-a-more-efficient-implementation-for-a-bidirectional-map). In particular:

Immagine che contiene testo

Descrizione generata automaticamente

See examples.

*Questions*

1. What do I have to store in the data structure?
   1. Involutions
   2. Destination dart, involution
   3. source dart, destination dart, involution

## Analysis of the state of the art

Study of the code provided.

# 28.09.21

During a discussion between Majid and Kropatsch, I listened that, maybe, there is a way to apply contraction and removal operations so that I can skip several levels of the pyramid. In this way, I will store less information than saving the involutions through all levels.

*Questions*

1. How can I choose which edge has to be removed/contracted at each level?

## About n-dimensional generalized Gmap pyramids

Consulting the file [..\Paper\n\_generalized\_map\_pyramid.pdf](../Paper/n_generalized_map_pyramid.pdf), in chapter 4 is mentioned the construction of a pyramid for nD images.

In practice, it is not possible to describe the initial image by an n-G-map since it would take too much memory space. We can represent the initial n-G-map implicitly by using a simple matrix. Another idea consists in starting directly with a first segmentation, the corresponding n-G-map being level 0 of the pyramid.

Three possible representations are equivalent: explicit, hierarchical, and implicit (see Fig. 7). The first one follows immediately the definition; the two others are proposed to reduce the cost in memory space.

*Hierarchical*

This kind of representation has the purpose that all involutions are explicitly for all levels.

*Implicitly*

This representation contains the map of level-0 with additional information to compute the other levels.

Chapter 5.2 of the same file have explained the differences between the three representations in terms of space complexity. In particular:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Explicitly | Hierarchical | Implicitly |
| Space complexity | O(mnp) | O(mnp) | O(np) |

Where:

* p
  + number of the darts of the level-0
* n
  + dimension of the space
* m
  + number of the pyramid’s levels

What I have understood is that implicitly encoding is needed to store the information about the first level of the pyramid (level-0) and only another level with the explicit information. This makes me realize that there is a way to not store the information in memory but compute the interested level of the pyramid when we want. The question is: the explicitly one is the last one of the pyramid?

# 29.09.21

Now the question is: what do I have to save to have the right links between darts and involutions?

Considering that the data structure I want to use is the *bidict* for the python script.

On this website there is a pdf doc of the bidict library:

* https://bidict.readthedocs.io/\_/downloads/en/v0.13.0/pdf/

I am writing a toy example for using the above-mentioned library and I am noting that there are some properties I have to respect to use it. In particular:

* I cannot use a list as a value parameter because it is not a hashable value
* To solve it, I have read I can use a tuple value
  + In this case, every time I will want to add a new element in the tuple, I have to redefine the bidict with the new tuple created.
  + **UPDATE**
    - I have now seen that there is a method *update*. I will read about it and I will check if it is useful and works. OK, IT WORKS!
* If will be an item (for instance, a tuple) that is inserted more times in the bidict, the structure automatically ignores it and print it just one time.
* The use of a tuple is the same as a list, i.e. in order to access an element, I have to know the index. Should I have to sort it?

### About implementation

As I said before, I am trying to use the bidict library. I have some doubts about when I have to introduce in the script this data structure. In detail, should I introduce it when I apply the operations of removal and contraction, should not?

What do I need?

* A variable that traces the information about the levels of the pyramids
* The information is available at the same time?
  + I think yes because when I apply a removal or contraction operation I know what is the dart involved, at which level and also the type of the involution applied.

At the moment I was focusing on the pixel map, but maybe I should focus on the dict\_gmap script.

Immagine che contiene testo

Descrizione generata automaticamente

I have found it on: [Bidirectional Dict In Python | Nakoblog (s-nako.work)](https://s-nako.work/2020/10/bidirectional-dict-in-python/).

However, this website considers also a custom version of the bidict where we can decide to allow duplicate values or not. That is done by building a custom bidict as a subclass of dict.

* IMPORTANT THING
  + I have to consider the special and time complexity of this data structure
    - I can make tests considering the time taken to execute a cell as well as a *mathematic* proof (it is useful if I can find something online).
  + Where can I find the time complexity/results that the PRIP lab has reached with the default dict? I would want to use them to have a benchmark.
  + Is right to modify the dict version of the Gmap and use the bidict? Or do I have to use the bidict in another part of the script?
  + The purpose is that of modifying the removal and contraction in dict\_gmaps? Of course, using bidict instead of dict.

*During a dialogue with Carmine*

Maybe I don’t have to modify the main implementation but I can take the information about the dart that is removed/contracted when is made the call at the remove/contraction function.

# 30.09.21

Today I have problems with the classes in python. The methods of my custom class aren’t seen by it that see only the methods of the main class.

I also have extended the main class and made override of the interested methods, but nothing change.

*Questions*

When and how can I know I have arrived at the top level of the pyramid? I have to store the information of the last level, so, this information is important for me.

An idea can be checking if the removal/contraction operations are possible, but maybe I am not interested in it. In the case that both of them cannot be done, I suppose I have arrived at the top level of the pyramid. Otherwise, do I have to store all the information for each level? **Not clear yet**. After a discussion with Carmine and Luca, we suppose that I have to store all the information, about each level.

Always during the discussion with them, Carmine pointed out to me that if I use a tuple, as I was doing, I will occupy more memory than if I would use n different bidicts (one for each kind of involution) and as key, I will have the starting dart and as value, I will have the destination dart.

Another thing is I have to ask professor Jiri how I have to manage the information about labels, considering that we should work with segmented images.

*Problems*

One of the problems is that the class nGmap provided is private and I cannot use its methods.

# 01.10.21

Looking at the class nGmap provided by the prof Jiri, I notice that almost all the methods are private. Consequently, some public methods recall the private ones.

I have changed my way to implement the data structure, using *n*-bidict for each alpha. In this way, however, when an *i-*cell will be removed, I will have to insert the source dart and the destination one in the correct bidict.

I think I have to better understand what happens when a removal/contraction operation is applied.

Moreover, I need to know how I will manage the selection of the right alpha bidict based on the call of remove\_i\_cell/contraction\_i\_cell. I need something that makes it easy to know.

*Problems*

I have some problems with the implementation of the switch case in python due to the indices, I don’t know why. I would want to solve to not have more if cases.

I am implementing it so that I can have in input the dart and the i-cell the user wants to remove/contract and I can use it to manage my data structure chosen.

*Questions*

Can I modify the file gmap.py?

This question borns because I need to be able to have the information about d2 that is in *\_i\_remove\_contract* and then is given as a parameter at the *set\_ai* method. The last one does not do anything, over that assign the dart d2 to the d1 one and alpha.

The involutions involve two i-cell are always different? I am raising this question because, if the answer is yes I don’t have problem with the key and value if have the same *value*. I will study it.

*Note*

When I use the bidict I have to be careful because this data structure is bidirectional and it means that also the values cannot be duplicated, like the keys. Maybe, the better way to store this kind of information is to have a sorted list by level as value.

*Initial assumption: the level-0 is store implicitly.*

What remains to solve?

* Recover the information about d2 (the destination dart) and try to put it into the data structure
* Invent some tests (benchmark) to try the work of the data structure
* The problem with also the values of the data structure
  + Understand what type of data structure I have to use to not occupy more memory space

# 02.10.21

I think that today I won’t do anything. We have to prepare some dishes for the international dinner.

# 03.10.21

Nothing also today.

# 04.10.21

The idea is that today I am working on the implementation of bidict or two different dictionaries. The first one with a tuple, and the second one with a list as the value of the structure. The important thing is the spatial complexity I have to check for both implementations.

In this way, for the implementation with two dictionaries I have to store the information within 2\*(n+1) ones (2 dictionaries times alpha\_i). Maybe, intrinsically, is the same for the bidict because it is composed of two dict.

I have found a naïve implementation of the bidict with the same value online, but if I want to insert a value with a key that already was into the data structure, it will be overwritten. In this case, I can think of a solution that provides a tuple as value, always in this, I found online.

*Questions*

1. What does my solution have to take in input? A sequence of the operations? It is not pretty clear. Right now I am implementing a naïve solution that takes in input a single dart to remove and a single i-cell.
2. Does the prof have an implicit representation of level-0?
3. Do I have to modify the function set\_ai of the class gmaps?



The image above represents a possible solution. The implementation consists of two dict where we can have the same value also for the keys.

* [Bidirectional Dict In Python | Nakoblog (s-nako.work)](https://s-nako.work/2020/10/bidirectional-dict-in-python/)
* Another useful reference [How to implement an efficient bidirectional hash table? (py4u.net)](https://www.py4u.net/discuss/11773)

In this way, I have solved the problem relative to the values of the data structure, at least initially.

# 05.10.21

Happy birthday, Mum!

OK, now let’s study😂

I want to visualize the code of the prof and understand where he has implicit memorization, assuming there is.

What I see in the code of the prof:

1. There is a branch with some code that I can use in the future to test the performances of my solution, compared to the prof ones. In particular, with that code, I can also estimate the memory occupied by the stored information.

*Questions*

* are there implicit memorization in your code?
* Do I have to implement a script like *dict\_gmap.ipynb* using my data structure?
* All the LUTs will have to be removed? I mean, all the n-Gmaps will be based on my data structure? In the case of an affirmative answer, do I have to rewrite all the script?

On the branch *voxelmap* there is an encoding of implicit alpha:

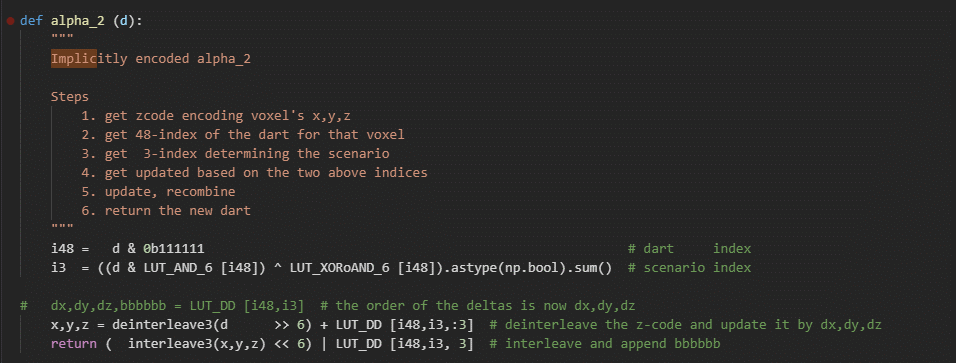


Figure 1: implicit alpha2 using Morton Code

# 06.10.21

Summary of question to ask Jiri:

About INPUT

* How can I choose which dart has to be removed/contracted at each level?
  + I mean, What does my solution have to take in input? A sequence of the operations? It is not pretty clear.

About IMPLEMENTATION

* Do I have to implement a script like *dict\_gmap.ipynb* using my data structure? Or I can easly use my data structure in another part of the project?
* Can I modify the function *set\_ai* of the class gmaps? So, I have overrided it in my custom class to check the working of the basic implemented data structure.
* All the LUTs will have to be removed? I mean, all the n-Gmaps will be based on my data structure? In the case of an affirmative answer, do I have to rewrite all the script?

About DATA TO STORE

* Do I have to store all information at each level, don’t I?
* If you will work with segmented images, do I have to store also labels?

About IMPLICITLY

* Does the prof have an implicit representation of level-0?

TO DO

* Find a way to create a dynamic custom name of a function given a parameter *i*.
* Find information about computational and spatial complexity.
* Add comments to my custom implementation, if it is not pretty clear.
* Looking for a solution for a switch statement that is more elegant than if one.
  + We don’t need switch statement because we can use the custom name function

Immagine che contiene testo

Descrizione generata automaticamente

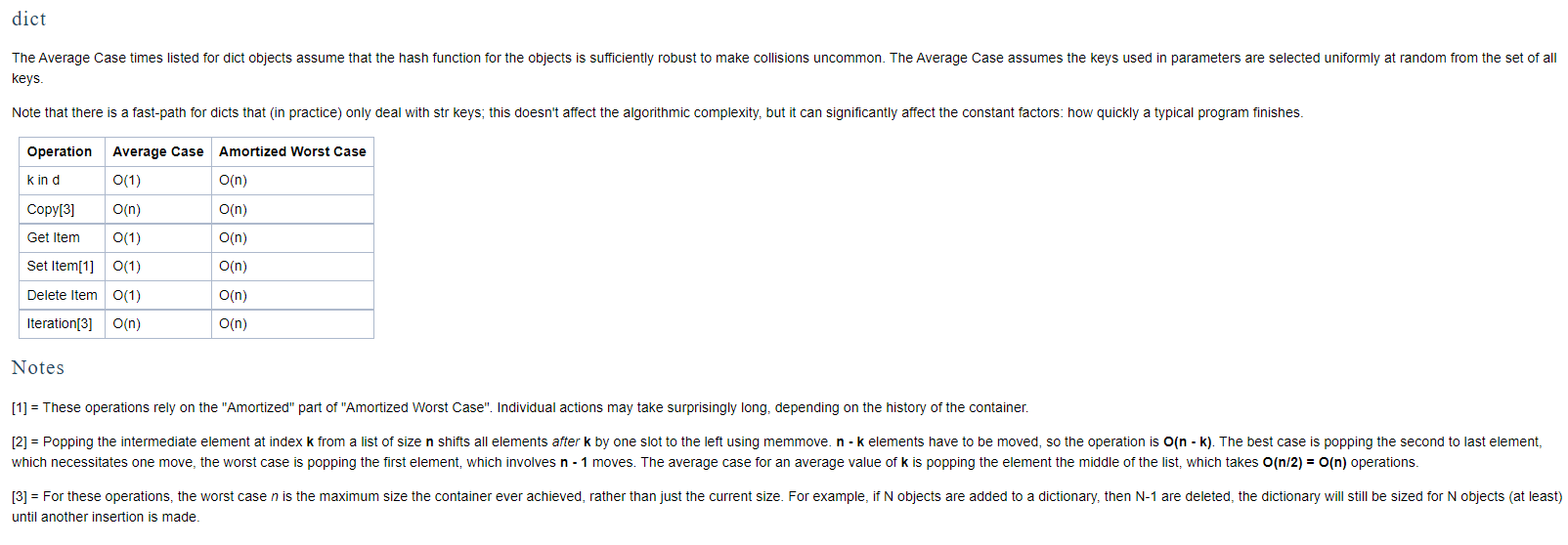
Figure 2: Performance bidict - reference pdf manual

The memory occupied with an implementation in C is much less than one in Python. I mean that *int* and *string* variables, for example, would occupy more space in memory using Python. For that reason is better, then, convert that part of the script in C++ using the CGAL library.

About SPACE COMPLEXITY

**Auxiliary Space** is the temporary space allocated by your algorithm to solve the problem, with respect to input size.

**Space Complexity** is the total space used by your algorithm to solve the problem, with respect to input size. Note that the Space Complexity includes input size.



# 07.10.21

I have a meeting with prof. Jiri.

What I should check:

* When I insert as dart input a certain value, i.e. 2, and as i-cell input, i.e. 0, in the data structure of alpha\_0 will be stored the value 0 as a key.
  + **UPDATE**
    - The problem is that when we call the remove/contract function to compute the operation, it calls the *set\_ai* method. That method will store the involution between a *new dart* and another one that is not the dart I want to remove, but the dart that replaces it. So, the set of the involution not involved the dart we want to remove, but only the survive ones.

What I observed this morning is useful because I caught an error I was doing in storing data in the structure.

* I should solve this problem and store the needed information in the data structure.

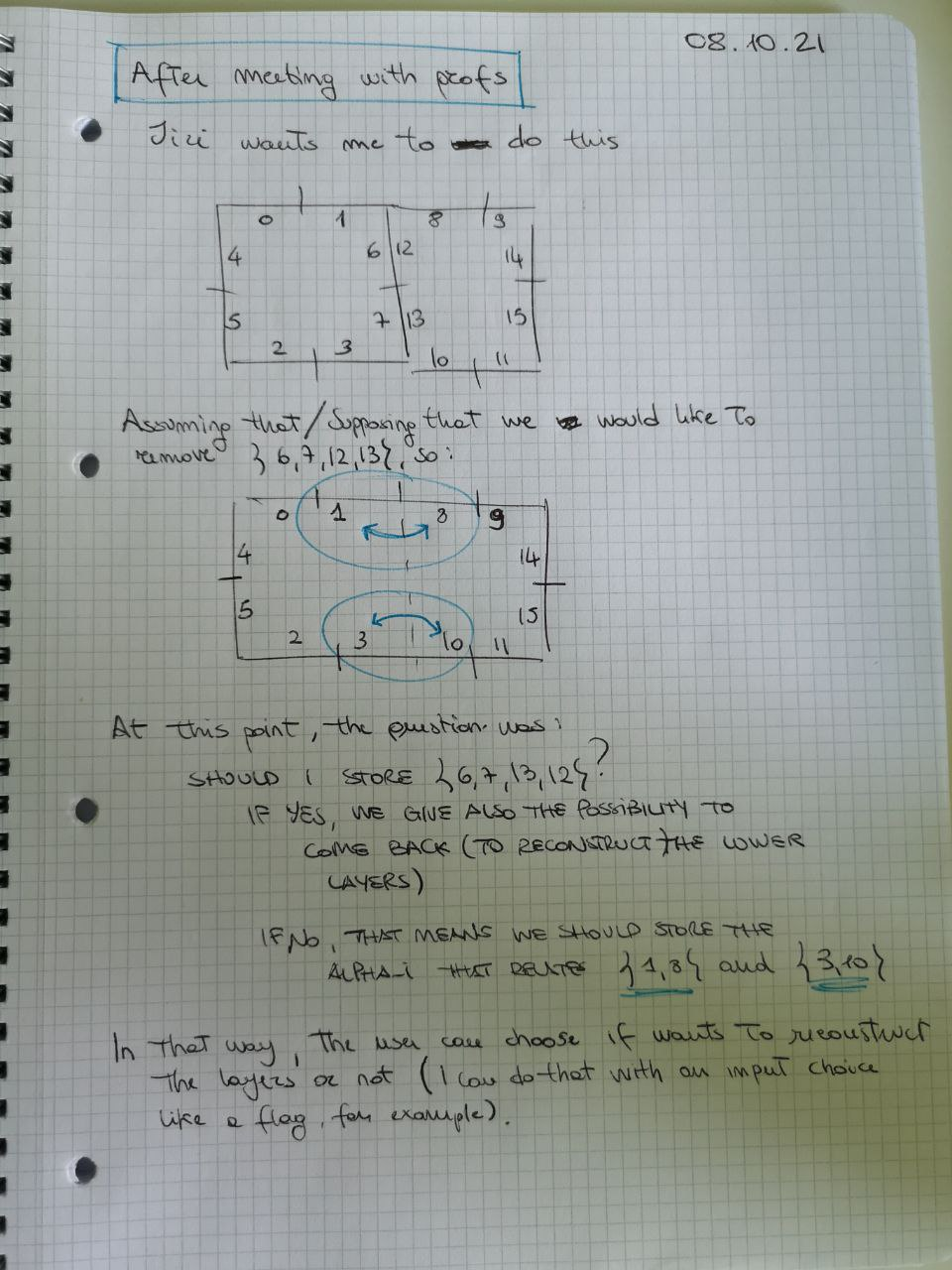
Another observation I did, is that I should not use the *set\_ai* method, but *\_remove\_dart* one. In that case, though, I don’t have the information about the *new\_dart* I should associate with the dart I want to remove.

* Can I modify the gmaps class and create a new method in order to obtain interesting information for the scope?

# 08.10.21

*After the meeting with profs*

* What they don’t understand is that the bidict is implemented with two dicts and the dicts are based on hash tables. So, in that way, if I want to access, with the key at the value, it takes O(1). The same worths if I want to access with the value at the key.
* I can modify the *gmaps* class (that class starts from line 72 of the *custom\_gmaps.py* script)
* I should check if is possible to store the self-loops, i.e. {100, 100}
* See *indirect indexing* for Matlab
* See the canonical representation (doc sent from Luca on Telegram)
* As input, we have a set of darts (that we want to remove/contract) and an involution kernel (as said Walter)
* I have to work only on gmaps with no label. So, I can put an implementation of the gmap with labels in future work.



Today I tried to think of a possible solution for my problem. In particular, I concentrated most of the time thinking about how I should implement a new nGmap class following advice given by the profs. I copied the class that was in the dict version of a nGmap. So, now, I should modify it to have my custom solution for the topic.

The most important work starts now!

I should focus on it thinking that I have as input only the set of darts and I will not know which darts are removed or contracted. That is my task.

The commitment for half tomorrow day is to better understand the dict implementation and modify it to use bidict. Eventually, I should also found a solution to modify the *from\_string* method, which is static somehow, and assign a value to the *reconstruction* variable.

# 09.10.21

Saturday…

# 10.10.21

Sunday…

# 11.10.21

Let’s go, the week starts now!

Yesterday prof Kropatzch send us two papers we should read for our topic con Gmap:

* [..\Paper\Fourey2010.pdf](../Paper/Fourey2010.pdf)
* [..\Paper\Implicit\_encoding\_combinatorial\_maps.pdf](../Paper/Implicit_encoding_combinatorial_maps.pdf)

# 13.10.21

Some slides of Florian’s work:

Immagine che contiene testo

Descrizione generata automaticamente

Immagine che contiene testo

Descrizione generata automaticamente

He has implemented a prototype with Python language and then has rewritten the code in C++ using the CGAL library. I should do the same thing using a different data structure, my custom data structure, so I could take your code, study it and using it to implement my solution.

*Problem to solve*

During the first execution of custom bidict implementation an error occurred:

