

IIT Madras

ONLINE DEGREE

Computational Thinking
Professor Madhavan Mukund
Department of Computer Science, Chennai Mathematical Institute
&
Professor G. Venkatesh
Indian Institute of Technology, Madras
Introduction to Collections and List Data Structure

Professor G. Venkatesh: So welcome back. Congratulations on completing your qualifier successfully and enrolling in the program. So we are now going to start the second part of our Computational Thinking course, which has basically another 4 weeks and then, there will be a third part last 4 weeks alright.

So in the first 4 weeks, what we saw, means, we saw basically was that we introduced you to iteration and how to keep track of intermediate values in the iteration using variables. And using these variables to represent various things, it could be intermediate values; it could be accumulation of various things that you have found in the dataset.

And using that accumulated value, basically, to find a relationship between an element and the aggregate. Like, for example, you want to find all the students who are above average, below average, you want to award grades to students, things like that. So that is one iteration following another iteration.

And then, we saw that you could find the relationship between two objects, two cards, for example; two students or two customers or two shops, for example. Such relationships, you could track; two words. You could find these relationships by examining every card and comparing it with other cards, so which basically gave rise to nested iterations.

So this is what we did in the first 4 weeks. And...

Professor Madhavan Mukund: Yeah. So now, I think what we need to look at more carefully is how to keep track of these, some of the things that we have calculated. So when we were using variables in the first 4 weeks, we were typically using them to keep track of a number and then, comparing this number, later on. So we are accumulating a sum or we are keeping the max or we were computing the average and storing it. Or we were counting the third-highest mark in a subject.

But many of these situations, we actually came up with, for example, collection. So we might have the collection of the list or the set of students who have got the top-3 marks in a subject, or the set of students who were born in a month of March, for example. And we do not have a satisfactory way of, at the moment, remembering this kind of information in a variable.

So we need some way of keeping track of these collections, collections of information rather than just. So, of course, we cannot just keep talking about person 1, person 2, person 3 because there is no limit. So we cannot just keep creating new variables for each item in the collection. So we need a way to talk about the collection as a whole so that we can manipulate it and pass it around and use it in our further computations.

And similarly, when we were doing things like a relationships and we are looking at how do we keep track of the fact that a pair of students are born in the same day, so how do we keep track of this kind of, which students are connected to other students when we try to group them in various ways. Or even the shops and the customers, how do we keep track of these connections between the shops and the customers?

So this kind of more elaborate maintenance of data and manipulation of data will be one of the things that we will focus on in this second part.

Professor G. Venkatesh: So you know, I mean, in Maths, this goes on Maths, which also you teach in this program. Collections are sets presumable, right?

Professor Madhavan Mukund: Yes, collections are sets. So for us, a set will typically be a set of cards or a set of people, it like for the, if it is a shopping bill, it could be a set of customers who are part of the cards for instance.

Professor G. Venkatesh: So and relationships are relations in the.

Professor Madhavan Mukund: Relations, exactly as we did in the Maths course. Yes.

Professor G. Venkatesh: So relation, basically is creating an ordered pair.

Professor Madhavan Mukund: Yeah. So you take the, you take two sets and you take pairs from these two sets. And some of these pairs are in the relation and some of the pairs are not in the relation.

Professor G. Venkatesh: So this connection between, so but let us start with sets. The first one is set. So set is basically just a subset of the cards, some set? So if I take out some cards, that will be set.

Professor Madhavan Mukund: Yes, for instance, you could take out all the students who have scored more than 55 in physics, for example. So that will be one collection,

Professor G. Venkatesh: Collection of students.

Professor Madhavan Mukund: And that is the set and however it comes out, it is a set. It does not matter in what order you pull it out. The set says that is just the collection is important. It is not really important, how you sequence the things.

Professor G. Venkatesh: So what is the problem with just taking out those cards and keeping it aside, those are, so I could just, for example, if I want to find out all the.

Professor Madhavan Mukund: So supposing you pull out all the students who are say, born in May.

(Refer Slide Time: 05:11)



Professor G. Venkatesh: Born in May, I can do that. So let me do that. So this is December, February, July, this is May. June, February, March, March, Jan, September, October, May. September, December, October, January, March, March, December, November, March, May, May, April, July, November, July. So I put out all the May cards.

Professor Madhavan Mukund: So now we have a set.

Professor G. Venkatesh: So this is a set.

Professor Madhavan Mukund: So this is a set. This is a set of cards which are in of among this set. So it is a subset of the set of cards. And now, supposing I ask you as a different question to give me the set of all students who are from Chennai.

Professor G. Venkatesh: From Chennai?

Professor Madhavan Mukund: So now, what will you do?

Professor G. Venkatesh: Okay, so.

Professor Madhavan Mukund: So I do not want to forget that set. So that set is something; so this is what we do when we are doing our procedures. So we remember some information in a variable, and then we continue doing something else but we keep track of what we had done before so we do not have to go back and do it again and again

Professor G. Venkatesh: One way is to just keep track of these numbers and then go back to this card and do it again.

Professor Madhavan Mukund: So where will you, how will you do that? Where you will have to write it down somewhere or?

Professor G. Venkatesh: It would be nice if I did not have to pull these cards over, no?

Professor Madhavan Mukund: Yes.

Professor G. Venkatesh: If I could just keep the cards in the desk, that thing, but just note down; have a way to get to the card when I want, that would be nice.

Professor Madhavan Mukund: Yes. So for instance, now if you put these back, all the May things are at the top, which was not the case before. And so, each time you pull out a set, you are somehow disturbing the cards and we are creating a side effect.

Professor G. Venkatesh: You know, I mean, I was thinking about this. So it looks like we do in our natural daily life, we have way of doing this. So, for example, if I was reading this book. So

now, when I read this book, I read up to some page and then I go, I mean I have to do something else, so I put the book aside and go away.

But then, I do not remember where I am so usually, you have this interesting thing, the bookmark. So you put the bookmark like this and then, now, the next time when I come, know where I was. I can straightaway go like this, turn them and I am at the page.

Professor Madhavan Mukund: And if I am also reading the book?

Professor G. Venkatesh: Now, if you are reading the book.

Professor Madhavan Mukund: I could be at a different place.

Professor G. Venkatesh: You could be in a different place, so you could have a different one.

Professor Madhavan Mukund: So I can have a second bookmark and I put it somewhere else.

Professor G. Venkatesh: And since you know your bookmark and I know my bookmark, so you can start straight away from here. So any number of bookmarks we can put in and that way you can get to any page, you do not have to take out the page and put it aside, right?

Professor Madhavan Mukund: Exactly.

Professor G. Venkatesh: Here, you are taking out, you seem to be tearing out the page.

Professor Madhavan Mukund: Yes, that is more, exactly. So you are not exactly dividing the book between the two of us. And then, if you if you catch up with me, supposing I have read two chapters ahead of you and you catch up with me and I say no, no, I am still holding on, then you have no way to access the pages which I have.

Professor G. Venkatesh: Correct, so all those complications. Is there a way to do this bookmark thing here?

Professor Madhavan Mukund: So, yeah. So we could, what we can do is instead of pulling out, we can exactly do that. We can mark where we found the set. So let us put these May cards back maybe in some sort of random way and just to do it in a different order, we will first do this Chennai thing.

Professor G. Venkatesh: Okay, let us find the Chennai thing but we are not, the idea is now to not to disturb the cards. So we will just look at the card, Trichy, Madurai, Nagercoil; I am keeping the card like this. Bengaluru; Chennai we want to find. Bengaluru, Vellore, Erode, Trichy, Bengaluru, Ambur, Madurai, Chennai.

Professor Madhavan Mukund: Okay. So what I will do now is I will, for us to just to make sure we do not mess it up, I will write the card number just for our own reference. But the main thing is that we will put a bookmark here. So now, we have a bookmark at that point.

Professor G. Venkatesh: So without disturbing it, I have to hold it without disturbed.

Professor Madhavan Mukund: Yeah, somehow. Yeah.

Professor G. Venkatesh: So that is Chennai. This is Chennai and then, this is Vellore, this is Bangalore, Madurai, Erode, Vellore, Chennai.

Professor Madhavan Mukund: So which number is that? 10. So we have a 10.

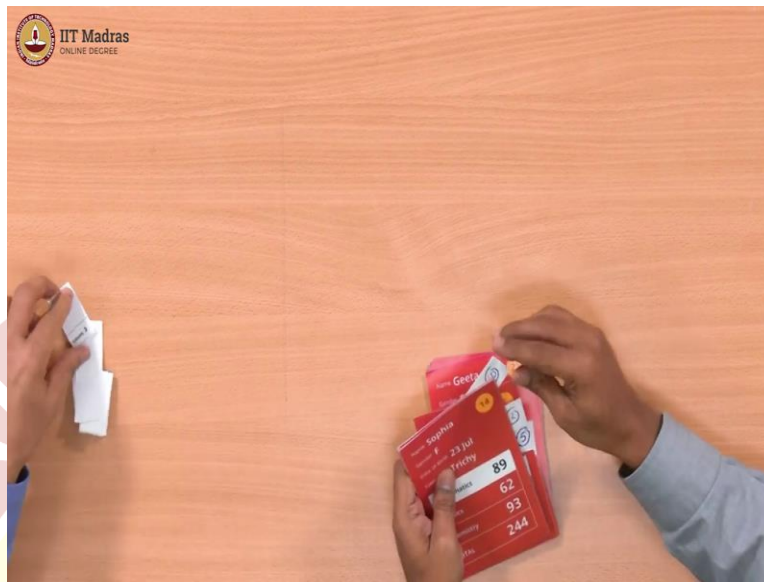
(Refer Slide Time: 09:20)



Professor G. Venkatesh: We have a second bookmark. We can write any number like this?

Professor Madhavan Mukund: Yeah.

(Refer Slide Time: 09:35)



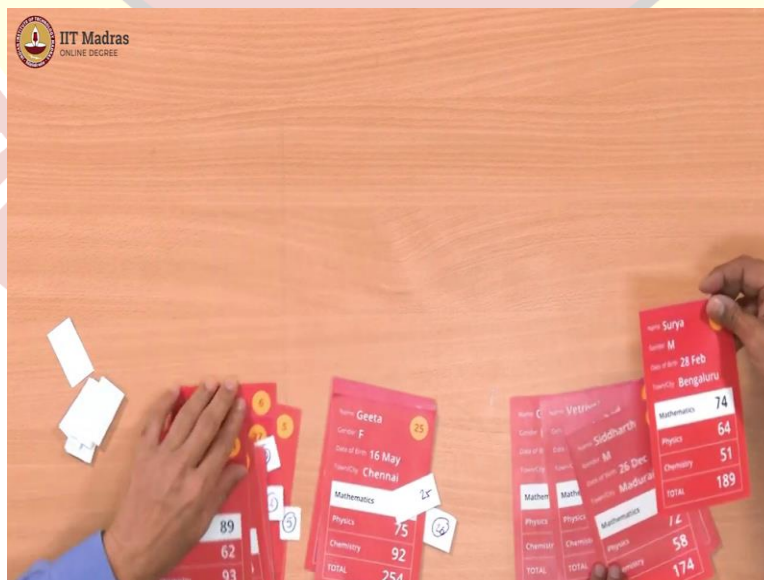
Professor G. Venkatesh: Chennai, Madurai, Ambur, another Chennai. So this is 5. So the advantage is that I can go directly to this card now.

Professor Madhavan Mukund: Yeah

Professor G. Venkatesh: Because I have the bookmark.

Professor Madhavan Mukund: Correct.

(Refer Slide Time: 09:52)



Professor G. Venkatesh: That is the idea. So now, you have 5. I have 5. Now, another Chennai.

Professor Madhavan Mukund: 25. So maybe you can just keep this stack known and start a new stack so that it is easier to hold on to this without.

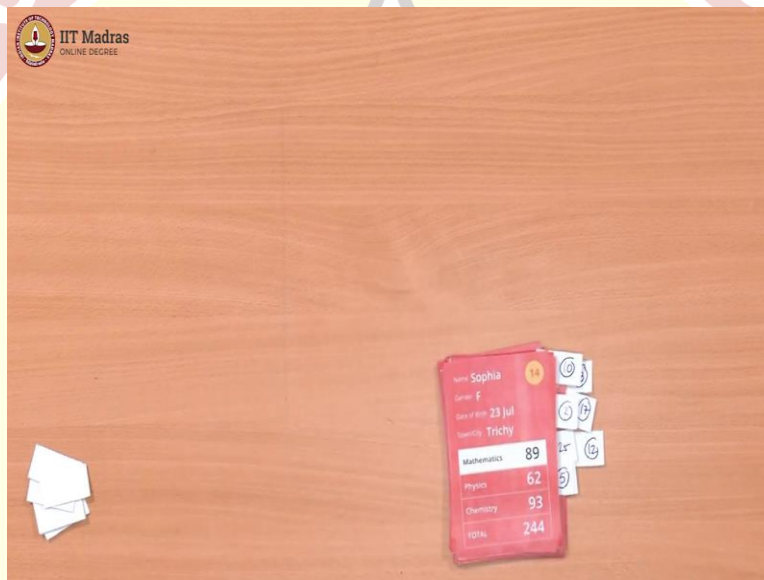
Professor G. Venkatesh: Salem, another Chennai.

Professor Madhavan Mukund: One second, let me just pull out that 5 here.

Professor G. Venkatesh: Chennai.

Professor Madhavan Mukund: So 26.

(Refer Slide Time: 10:28)



Professor G. Venkatesh: Bengaluru, Madurai, Chennai, one more. In fact, there are two more.

Professor Madhavan Mukund: So 17.

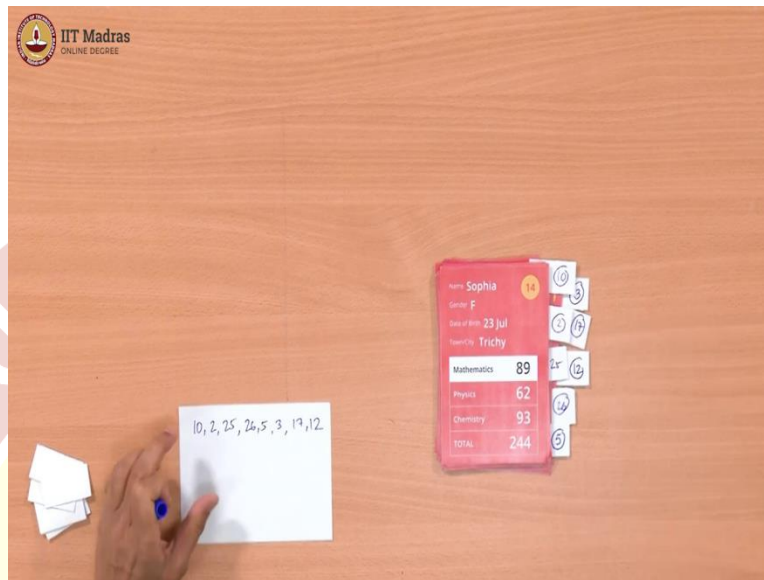
Professor G. Venkatesh: 17 and then 12. Then 12 then we have Trichy, and then we have one more Chennai. A lot of Chennai.

Professor Madhavan Mukund: 3 ok.

Professor Madhavan Mukund: So now, so if we did this carefully, and so it is a bit of a challenge just to manipulate these things physically but if you can imagine that if you are keeping track of it like this; now, we have the original deck of cards exactly as we started, except we have identified all the positions where we saw Chennai.

Professor G. Venkatesh: It is just like a bookmark.

(Refer Slide Time: 11:38)



Professor Madhavan Mukund: So we have created 1, 2, 3, I think roughly, 8, I think, if I remember right. Yeah. 1, 2, 3, 4, 5, 6, 7, 8 bookmarks indicating the positions where there are Chennai. And actually, we can if we wanted, separately, we could have written them down. So we could have written that these numbers were 10, 2, 25, 5, 6, or 16?

Professor G. Venkatesh: 26.

Professor Madhavan Mukund: 26, 5, 3, 17, and 12. Now, the reason why it is better to do that than this is because if I tell you that card number 25 is Chennai, you still have to go through the whole stack to look for card number 25, whereas here, in some sense, we can directly get to card number, whichever card number 3, I can go to the bookmark and pull out card number 3.

So it is easy for me to go from this to the actual card without worrying about how these cards are actually arranged because I have a direct way of doing this.

Professor G. Venkatesh: Now, this is not as going directly to a card, the computer memory is organized like that, right, you can go to any memory directly?

Professor Madhavan Mukund: Exactly. So we are kind of trying to give that same facility which a computer would allow you. So it is reasonable to assume that you have...

Professor G. Venkatesh: That is why they say, random access memory also. Because you can go to any point in the memory, any place in the memory.

Professor Madhavan Mukund: Yeah. So if you know where it is, if you know way of telling which place it is, then you can go to that place directly without having to go through the whole thing.

Professor G. Venkatesh: One by one.

Professor Madhavan Mukund: Yes.

Professor G. Venkatesh: So we are just basically mimicking what the memory of a computer does. We just putting the, directly the place where you have to go to. Okay.

Professor Madhavan Mukund: And we could do the same thing. For instance, without disturbing these white, these blue bookmarks. We could create, for example, red bookmarks for May. So we could keep the pile of cards the same.

So we had this earlier problem that when we created the set for May, that is children who are born in May, students born in May, and then we wanted to find a set for the students who are born in Chennai, we had to disturb that set and put it back into the thing so that we could go through the whole list again, a whole deck of cards again.

But now, whether we go through this deck of cards once, twice, 10 times, those bookmarks are in place and we are not disturbing the deck. So there is no side effect. The deck is as it is, each time we put, go through the deck, we put a different type of bookmark. So we will put bookmarks, maybe we can write it in red instead of blue and then, we will have two sets pointing into the same set of cards.

Professor G. Venkatesh: So in programming, this business of creating this list of what we are now calling bookmarks into a sequence like that it is called a list?

Professor Madhavan Mukund: Yes. So this is a list and this list is somehow indexing into this.

Professor G. Venkatesh: List is indexing, it is a list of positions into that.

Professor Madhavan Mukund: Yeah. So essentially, from the list, we can directly get to the same card whenever we want without incurring any extra effort. So we are not going to take more effort to suddenly look up 25 and...

Professor G. Venkatesh: So unlike a list of number, I mean this is, of course, are numbers but if you try then treating these as just a plain set of numbers, these are list of positions bookmarks. So very specific kind of list it is because it is a list of positions, which allow you to go directly into the set of cards and do something. So now, I have made a list of all the; what is it? The list of all the...

Professor Madhavan Mukund: Students from Chennai.

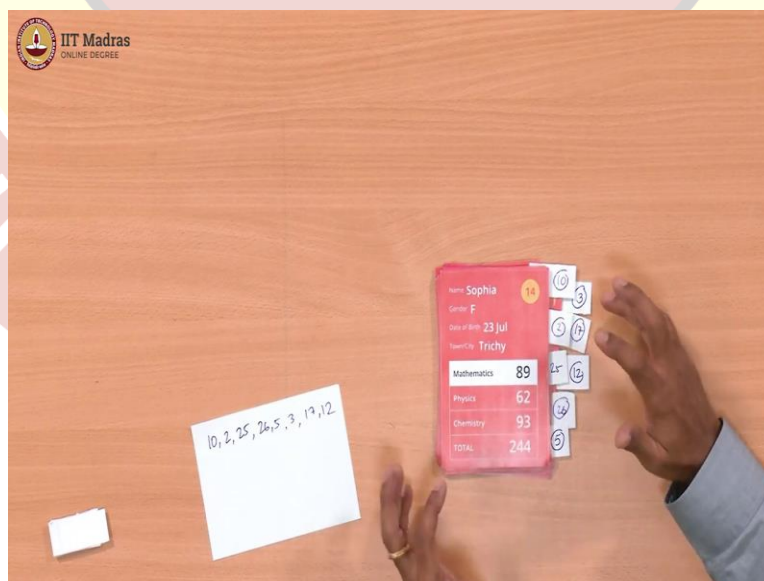
Professor G. Venkatesh: From Chennai, I could have similarly a list of all the students born in May that could another list. And then, I could do things with these two lists.

Professor Madhavan Mukund: Yes. I could compare for instance, which numbers appear in both. So those would be there.

Professor G. Venkatesh: Should we do that? I mean, should we want to find out.

Professor Madhavan Mukund: Yeah, let us try then.

(Refer Slide Time: 15:14)



Professor G. Venkatesh: How many students are there who are born in May from Chennai? Suppose I want to do. Normally this in sets, math, we would do this intersection. So now we find the set of all the people who are born in May, which is the set. And similarly, find a set of all the.

Professor Madhavan Mukund: This is Chennai, and then we will find this set.

Professor G. Venkatesh: We find all the born in May and then do you will do an intersection.

Professor Madhavan Mukund: Yeah. But then when you normally, when you write this intersection operation in maths, you do not describe the process of actually calculating the intersection.

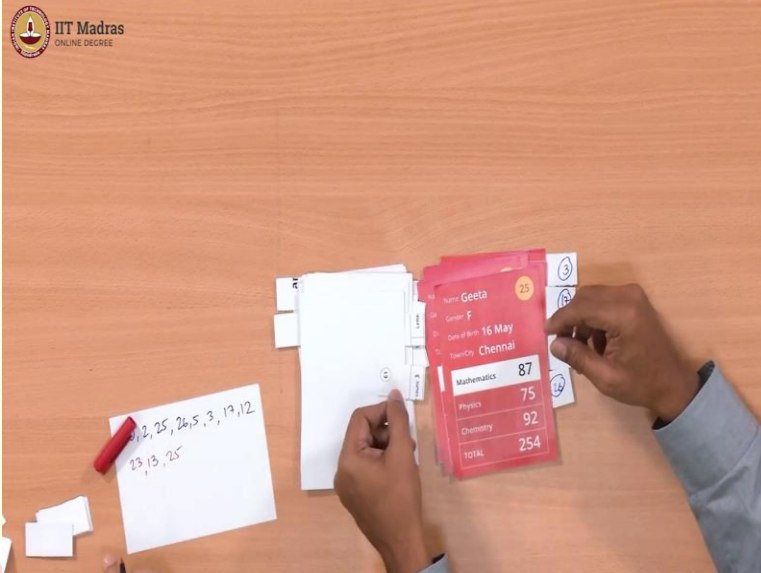
So you just kind of do it by some, for example, if I say what are all the prime numbers that also divides 75? Then you would kind of have many different ways of doing it but you do not explicitly specify a procedure to do it, whereas what we are doing is a kind of procedure. So we have done one iteration to find all the people who are born in Chennai,

Professor G. Venkatesh: And marked it

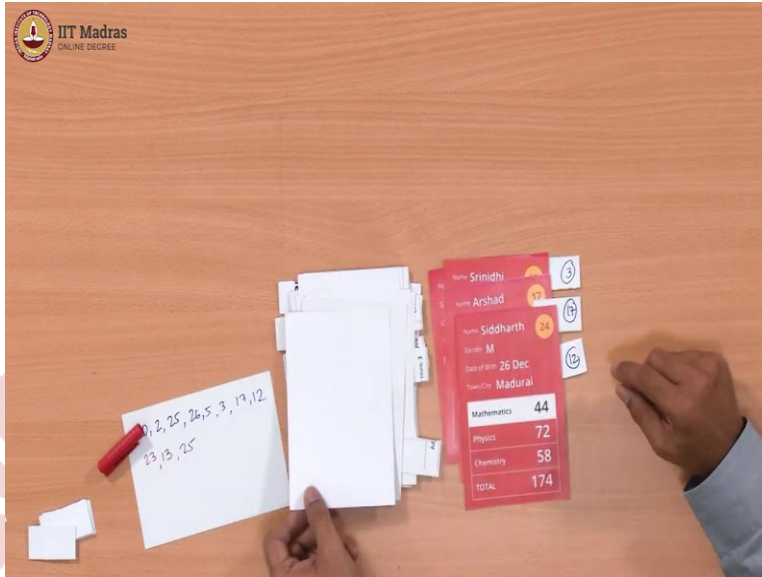
PPM: And marked it. Now, we are going to do one more iteration. Without disturbing the marks, we are going to do one more iteration, add new marks for those born in May. And then, we are going to compare those marks.

Professor G. Venkatesh: All right, shall we do that? I mean let us find all those people who are.

(Refer Slide Time: 16:21)



सिद्धिर्भवति कर्मजा



Professor Madhavan Mukund: So I have a suggestion. So to keep the order, we can just maybe flip the card so that it remains in the same order and not have to hold it.

Professor G. Venkatesh: Okay, so yeah. So we want to find all those born in May. And I have to move the bookmark also?

Professor Madhavan Mukund: Yeah, that is why I am saying that you do not have to worry about. So May, so 23 is a new one in May. So I give you a new bookmark 23 and I write it in red so that in case the bookmarks getting disturbed, so I will put the bookmark here.

Professor G. Venkatesh: So when you move it, I can move with the bookmark.

Professor Madhavan Mukund: March, November, December, March, March, another May, so this is 13.

Professor G. Venkatesh: May, we keep track of these bookmark, so I will keep this bookmark.

Professor Madhavan Mukund: And maybe you should keep it on the side. We can keep it somehow.

Professor G. Venkatesh: So this is Jan, so nothing. Then October, December, September, October, September, January, this is a bookmark. January, March, March again, so I keep the bookmark.

Professor Madhavan Mukund: February. Now, we have May and now, we have somebody who is actually in both. So this is 25 again, so I will just create a new red bookmark for 25. So we have actually two bookmarks for this.

Professor G. Venkatesh: Yeah. One either side. Then this June, this is another Chennai so I have to keep the bookmark. Then I have to keep, but we do not have a bookmark for this.

Professor Madhavan Mukund: Yeah, because we are looking for May. So as long as it is not May, we just pass it on with whatever, bookmark it already. So again, this is Chennai.

Professor G. Venkatesh: So we will keep track of it properly but again, it is not May so nothing to do. Again, another Chennai got it.

Professor Madhavan Mukund: Now, this is May.

Professor G. Venkatesh: This also is May.

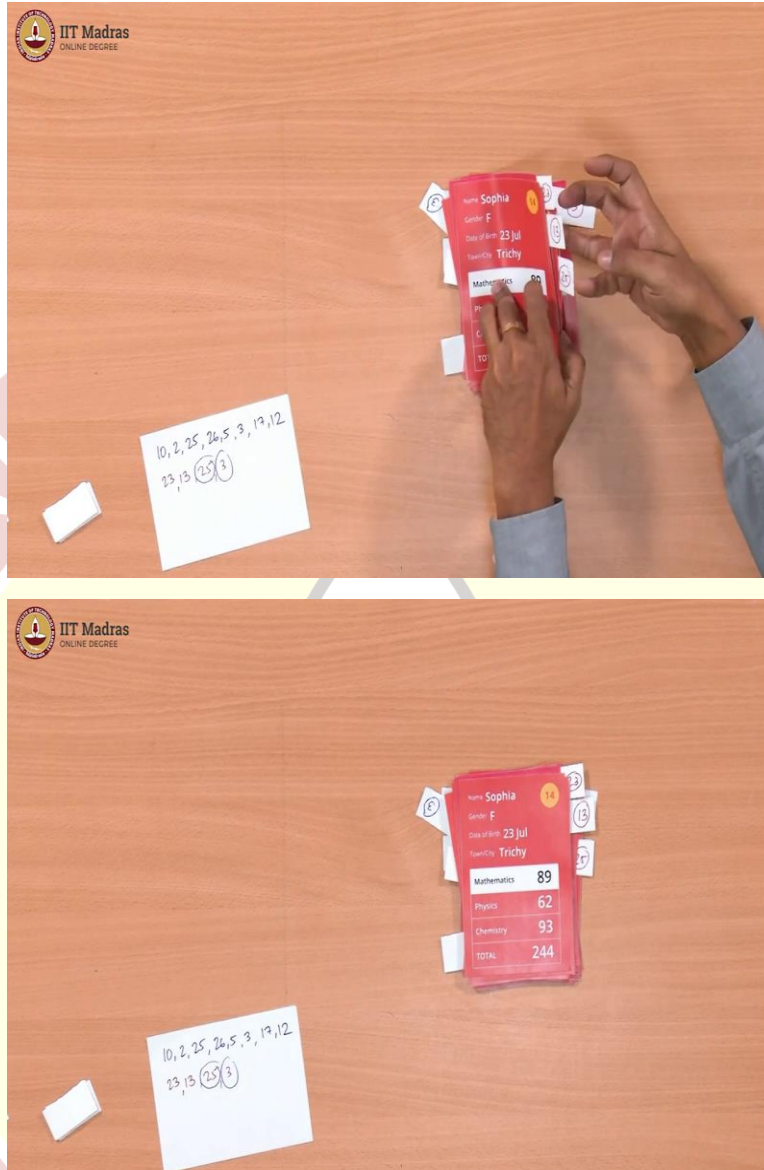
Professor Madhavan Mukund: So we will put a bookmark on this side also. And finally, this.

Professor G. Venkatesh: So we got two, so we got where to get to.

Professor Madhavan Mukund: So now, it is not; oops, it fell out so we have to. But the point is now, we have created two sets. So we have the old set on the left-hand side and the new set on the right-hand side. And these are these two lists. And so, now, if we look, for instance, for things which are in both lists.

Professor G. Venkatesh: So to do intersection, I can just, I do not need to look at the cards. I can just do this.

(Refer Slide Time: 19:26)



Professor Madhavan Mukund: Yeah. So I will say 10 to 25 and then, I see 25 is in this list. So for each of these lists, I will just check if it is here. So 26 is not here, 5 is not here, 3 is here, 17 is not here, 12 is not here.

Professor G. Venkatesh: So you are doing a nested iteration. First iteration is over this.

Professor Madhavan Mukund: Yes. For every element in this list, I am checking whether it is in this list.

Professor G. Venkatesh: So outer iteration is this list, inner iteration is this, and what you are checking for is equal to, right?

Professor Madhavan Mukund: Yes.

Professor G. Venkatesh: You found two elements which are equal.

Professor Madhavan Mukund: Yeah.

Professor G. Venkatesh: So this means that there are two students in this.

Professor Madhavan Mukund: Yes. And we know which ones they are because we have to look for 25 and this 3 which...

Professor G. Venkatesh: We can go straight like this.

Professor Madhavan Mukund: Yeah. And we can just go for straight to 25 and we can go straight to 3 and find it.

Professor G. Venkatesh: Here is 3. Fantastic.

Professor G. Venkatesh: So let us reflect what have we done? Basically, we said that we need to keep track of collections. Collections are sets. Sets are basically some subset of the cards. Now, one way was to take the cards out and keep it aside, but that means that you can only do one collection and cannot do two collections.

Professor Madhavan Mukund: Yeah. And also the cards get disturbed each time. Even if you put them back, the cards get disturbed. So if you have done something with one collection, then you put it back, then the order in which the cards were disturbed and you may...

Professor G. Venkatesh: So we did a simple trick, we used the bookmark trick to put a bookmark inside the collection. And kept track of only the bookmarks.

Professor Madhavan Mukund: And then we can work with only the bookmarks. For instance, if we want to find out the people who are from Chennai born in May, we only need to look at the bookmarks and do an analysis of the bookmarks. So any computation we do like nested loop or anything that we do needs to be done only here, we do not have to go back to the cards again and iterate through all the cards and do something.

So in some sense, we have done a filtering into these two sets.

Professor G. Venkatesh: Two filters, two separate filters.

Professor Madhavan Mukund: We have done two filters. Filtered out all the May, all the Chennai, and then we are comparing the Chennai and the May things using this.

Professor G. Venkatesh: Okay.

