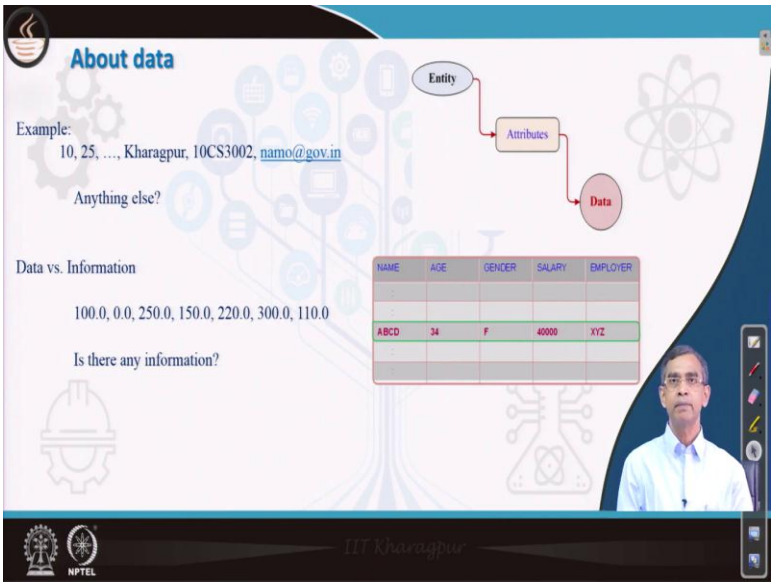
Data Structures and Algorithms Using Java
Professor Debasis Samanta
Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur
Lecture 1
Introduction and Course Plan

Welcome to the course. This is the first lecture in the series of lecture series. Now in this lectures I will try to give a brief introduction and plan of this course.

(Refer Slide Time: 0:42)

So basically you will be able to learn, as it is related to data structures and the programming, so we will try to give an idea about the data and then data whenever comes into the concern, so structure of data is important. I will give a brief about the importance of data structures. Different type of data structures are known at the moment, we will try to give a brief accounts of all those; objective of this course, plan of this course and finally the resources for learning apart from the, these video lectures or NPTEL course.

(Refer Slide Time: 1:32)



So let us first come to the discussion about the data. Now data is basically so common what nowadays we are very much known to this. Data usually represented in the form of with a very novice concept we can say in the form of numbers, in the form of strings like name, in the form of some mixed of numbers and string like email address, email ID and so on. There are numbers however different, some numbers related to the mobile and some numbers related to the roll number or numbers which indicates weight of a person are not necessarily same concept.

The different concept they (()) (2:26). Now whatever it be, we have the data. Now data is related to what? Actually data is related to what we call is an entity. Entity in other words can be simple way can be called as a record, for example, record of a student, record of an employer or employee, record of a course, record of an institution, and so many things are there. Now whenever an entity which represents a real life thing is basically characterized with a set of attributes.

For example, in this table as we see the attributes are namely name, age, gender, salary, employer; so this is basically is a employ, is a record or an entity, set of entities related to the different employees in an organization. So as we see the, in this table number of rows and the number of columns, each row represents the record, a particular record of a particular entity. In other words, what we can say – An entity better will be represented by a set of attributes.

Now, how an attributes can be represented? Attributes are of different type as we see in this example, for example, name is an attributes, which is basically is a collection of characters. We can say string. Age on the other hand is another attributes which is numeric type, gender is an another attribute, which is a special type called a Boolean type sort of thing, if it is a Boolean variables that means only true and false or male and female that kind of things are there or sometimes it will be represented by a single character M or F, etcetera.

Salary on the other hand is another type which basically store floating-point values, the real numbers. The employer is another example, it is similar to the name of an employee, it is basically string. Now as we see an entity is defined with set of attributes, where each attributes of different type and attributes stores the values, which we call the data. So here what is the data? So far an employee is concerned data indicates what is the name of employee, what is his or her age, what is the gender, and what is the salary and where the employee is working, the person is working.

So these are the several data is involved and as we see data are of various, data, those are there, possible for to represent that entity are in different form. So this is about the data. Sometimes there is a set of data belongs to the same category, for example, as we see here the, we have listed few data, those are in floating point number, 100, 250, 150, 220, 300, 110. So this is a collection or you can say set of data or we can say set of floating point values.

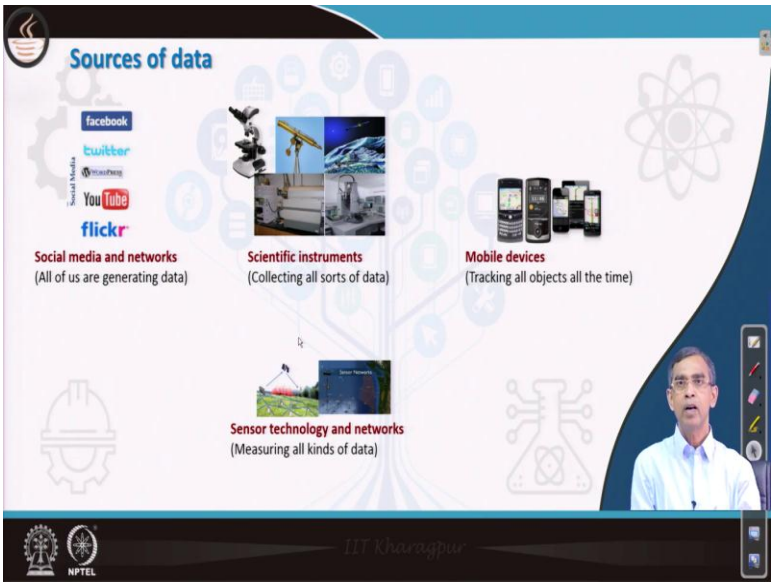
Now if I say these are the floating point values, in fact, it says that these are the set of five numbers or so, five values or so, but it does not give any meaning to anybody unless we extract meaningful fact from it. Now given a set of data, a collection, if we retrieve certain meaningful, what is called a concept from it, then that meaningful concept is called the information. So data and information are the two things.

Given a data, in fact, we are able to retrieve information. Now how this information can be retrieved? So, you have to process using certain algorithms or certain procedure and this processing is actually we do in our programming concept is by means of program. So if there is a program and data is fed to the program, program will return information. For example, if we say these are the values related to the amount of milk consumed by a person.

That is fine, so we can say that this is the amount of milk consumed by a person in maybe a week. Now, here is a question is that what is the information that how much money that he has spent for his milk consumption? We can obtain. So to do these things we have to do certain processing. For example, sum of all the values that is given there multiplied by cost of each unit, so it will return exactly an information.

More bigger, more in-depth information can be obtained, for example, what is the average protein consumption, protein intake for the person or fat intake for the person, that also can be obtained. More complex logic can be applied to find this information and so on. There are many-many that kind of examples that we can extend by which given a set of data we can retrieve many information thereafter.

(Refer Slide Time: 8:41)



Now, so data is an essential things that we have learned. Now we just want to give a brief discussion about what are the sources of such data? At the moment there are many such sources from where we receive the data. Our digital world is basically receiving the data, in fact, digital world really is growing like anything, everyday there are many data generated and then that data are stored and the data are consumed and like this only.

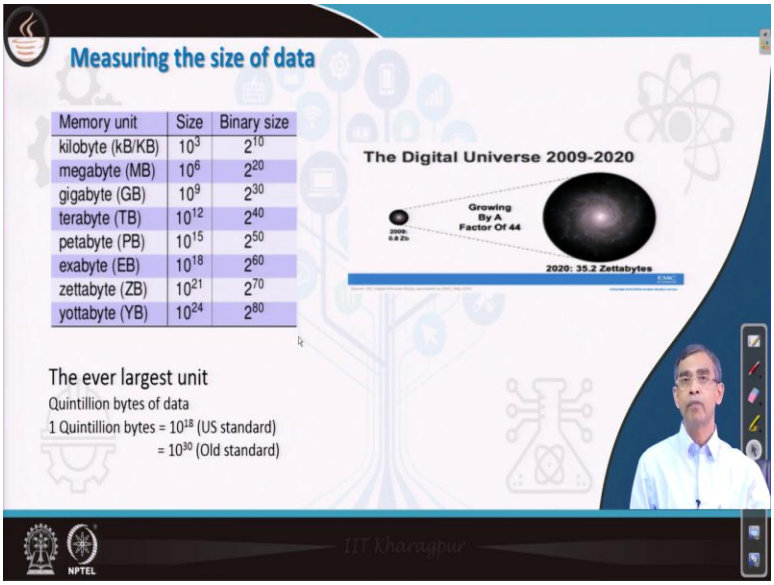
In brief, the source of data, major source of data are from the social media sites like Facebook, Twitter, YouTube, Flickr and so many things. Another source of data, there is a huge source of data from the different scientific instruments installed here and there, like from the satellites, like from the Earth Observatory, like from the different instruments installed in different research laboratory and so on.

So they are constantly pumping data, those data are stored in some machine, that machine is connected to the network, so all those data essentially linked to the network actually. Now another very popular source of data is the mobile handsets. From the mobile we are also capturing image, capturing audios, capturing many information, sending SMS, and so many things are there, all these data's are also populated and stored in some storage medium maybe the server in the service provider and so on.

There are other source of data also from many sensors, network of sensors like Internet, Internet of Things and other technologies; those are presently being used to solve many problems in different application areas. I have listed only few important source of data apart from this there are many other source of data in fact.

Now this source of data if you consider or sources of data from where we are getting, they are basically generating data in various formats, those format not necessarily only in text, this format maybe numbers, may be text, may be audio, may be video, may be image and what is not. So different forms, different types, different sources data is used and processing therefore, demands really a novel or a breakthrough technology.

(Refer Slide Time: 11:35)



Now data size is basically is enormous at the moment and that is why technology to cope with this huge volume of data, which basically we call Big Data. So data those are high in volume, high in varieties, high in what is called the requirement of high speed processing, then they are called Big Data.

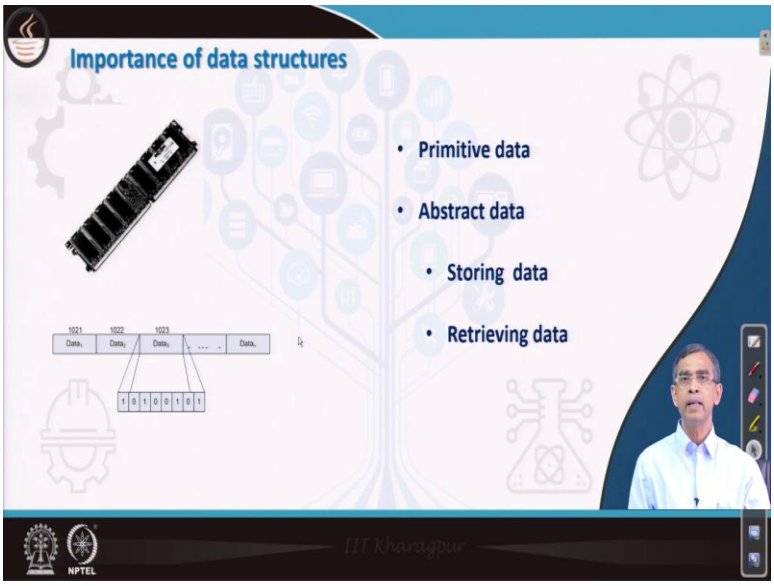
So at the moment the data as it is growing heavily, therefore in order to measure the amount of data different consortium proposes different unit, the largest unit known so far I have mention here is Quintillion. So one quintillion bytes, which is equal to 10 to the power 18th, it is basically

US standard. There is another old standard which was known as 10 to the power 30, so quintillion is the largest unit.

Now you also possibly know there are smaller units also there other than quintillion, for example, bytes, the KB, called kilobytes, MB, GB, GB for gigabytes, MB for megabytes, similarly TB - terabytes, PB – petabytes, and then EB, the Exabytes, ZB the zettabytes, and yottabyte YB, these are the largest unit as we can say. Now, so these are the different units evolved around the time in order to measure the data in this form.

Now, let us see what is the present size of our digital world. It is not long back, maybe 10 years down the line, in the around 2009 or so, the total size of the digital universe; that mean comprising all source of data those are linked in the internet, stored in a different laboratories, in the different organizations, whatever it is, it is around 0.8 zettabyte. Now this rate is every seconds, every moment is increasing at the rate of 44 percent. That means huge amount of what is called the explosion of data is taking place and as per the prediction it is says that in by the year 2020, the amount of data that digital world should hold it is around 35.2 zettabytes.

(Refer Slide Time: 14:38)



So the huge data, there is therefore, needs a better technology rather better way of solving the data or manipulating data and that is why we need a proper concept of data structure so that the huge volume, huge variety of data can be processed in real time, rather we can say Big Data can be handled in with a better concept of technology at the hand.

That is why the data structures are evolved, since the inception of computer technology, computer science, data structures is an well exercised topic, and data structures are known so far the different programming languages evolved and as you know the data structures means basically to define your data so that your program can handle them. So different programming languages, in fact, allow a programmer to define their data.

This type of defining data or the way the programming language can process the data is called primitive data. Primitive data really very simple and innocent like integer, real or Boolean or character, they are one sort of primitive data, but with this primitive data if you want to store some other complex data, for example, if you want to store an image, then definitely all these primitive data definitions supported by programming language developer is not the only way that you can do.

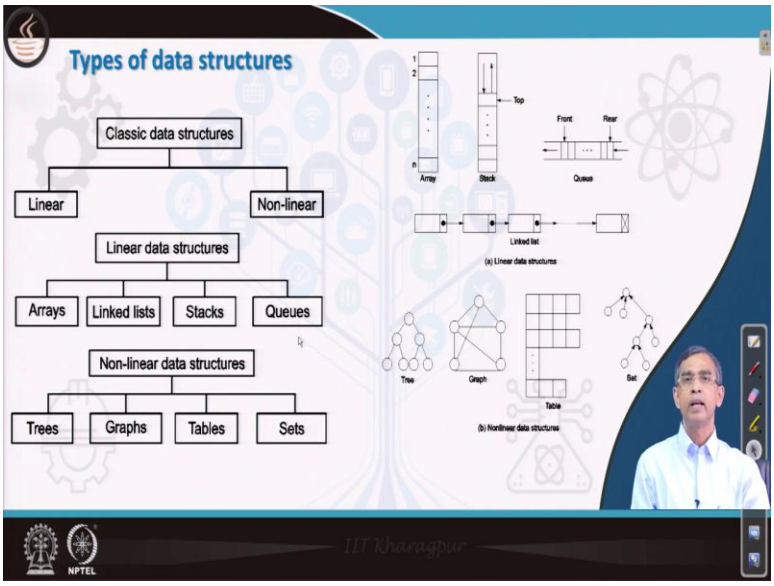
You can do, of course, but there is a need of a better concept of structuring data and that is why the recent programming languages like C, C++, Java, they proposed one concept called the Data Abstraction. What is the meaning of this data abstraction? The meaning is that using data

abstraction you can define your own data, so basically a programmer can define users data on his own way. Now it has many advantages. Advantage is that you can customize your data depending on your application.

This is obviously is an advantage, but problem is there. Unlike in case of primitive data, the different operations, rather different processing that you want to do, for example, how to store such a data, how to retrieve the data from memory, how to do certain operations on the data, for example, in case of image, how you can revert or how can zoom or how can remove the noise and so many things, it is of the users responsibility.

So defining users own data as the same time is advantageous as well as brings certain what is called the issues to the programmer that how all those data structures can be defined and not only it can be defined, how they can be stored, how they can be retrieved and how the different operations that can be performed on them. So, that is why the specials, specific studies which require in order to know your data structure.

(Refer Slide Time: 18:01)



Now there are many data structures known so far. I will quickly try to give a brief into the about the different structures of the different type of data, you can see one what is called a taxonomy of tree that I have mentioned here, so we can see these are the basic data structures. I will say what do you mean by basic data structures and then advanced data structure? First of all, let us learn what are the basic data structures?

Now the basic data structures are of two types, they are either linear or non-linear type. Now again the linear type of data structures can be divided into different categories like arrays, linked lists, stacks and queues. Likewise, nonlinear data structures are of different forms like trees, graphs, tables and sets. So, how all those data structure looks? Pictorially it is shown in this figure. For example, array look like this, a stack look like this, linked list like, this is queue, linked list like this, the tree, graph, tables, sets, like this one.

So this is a pictorial information, usually the teacher can teach about the different data structure with the help of figures to understand the concept easily. So these are the simple data structures or you can say basic data structures. Now what about the advanced data structures? What I want to say is that using the concept of all these data structures, any other type of data structures that you can realize. As an example, for example, you want to manipulate image.

Now which data structure that you can think for out of these basic data structures, yes, your answer is right, a simple array can be considered to store an image, to retrieve an image, to

manipulate an image, whatever the manipulation as I told you, zoom in or removal of knowledge or changing the color of images from one color to another and so many things are there, is not it interesting, right?

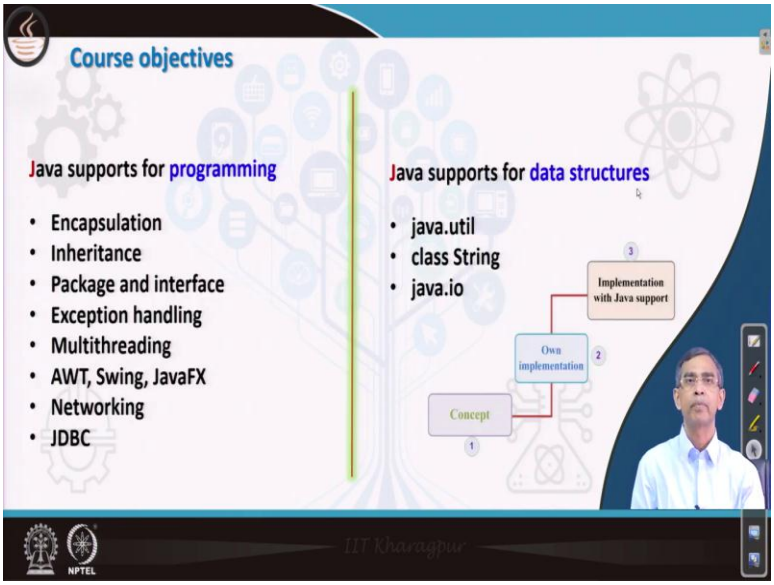
Yes, it is very interesting. Now, whatever the other data structures those are required, specifically a particular application can be thought of from this simple or basic data structures. So this is why in this course we will limit our discussion to all these basic data structures are there and finally give an idea about how the other data structures, those are we can say the advanced or custom data structures can be obtained with the help of all these basic data structures.

(Refer Slide Time: 21:00)

Now, I will come, quickly come to the course objective. Obviously by the time you have clear about idea that we want, we are going to learn about data structures. What are the different data structures are there? But I want to see that not only learning the different data structures, but at the same time we will also learn the programming for all these data structures. Our basic objective is that given a data how the program can retrieve the information from these data, where data can be in different structure.

So definitely to write the program what is necessary, necessity, necessary is that algorithms. For every type of data structure needs their own algorithms. So altogether data structures plus algorithms basically results a good programs or program to retrieve information given a set of data. So objective is basically learning data structures, algorithms and using them, how the program can be decided, can be, program can be written.

(Refer Slide Time: 22:19)



Now, so far this programming is concerned, more specifically we will consider about Java environment. Java is a versatile programming language. Java can provides different types of what is called our programming, different types in the sense that programming to handle different areas of applications, anyway all those application areas and the programming aspects in Java programming can be broadly classified into four major categories. First is called core programming.

The core programming basically helps you to learn about or know about the programming structure, control flow and essential programming features like encapsulation, inheritance, package and interface, exception handling and multi-threading. So this is basically core programming aspects. Other than these core programming Java as you know it is good for application software development and as for the application software development there are two parts, the front end of the application and back end of the application.

For designing front end of an application we need certain programming and Java support those kind of programming using what is called a graphical user interface concept. So, second component of Java programming is called the GUI programming. GUI programming can be obtained using AWT, swing, JavaFX like this. Now, the third part of this programming, because Java is heavily used at the moment for distributed programming or distributed computing like mobile computing network or Internet programming.

To support these things Java provides a huge set of features related to the networking, so this is the third part of the Java programming and the final part, it is also very essential part, it is dealing with database. Database somehow we are related to data structure or anyway, but database is a different concept and Java provides a good way of building a connection between your program to the database, which is stated as an external entity.

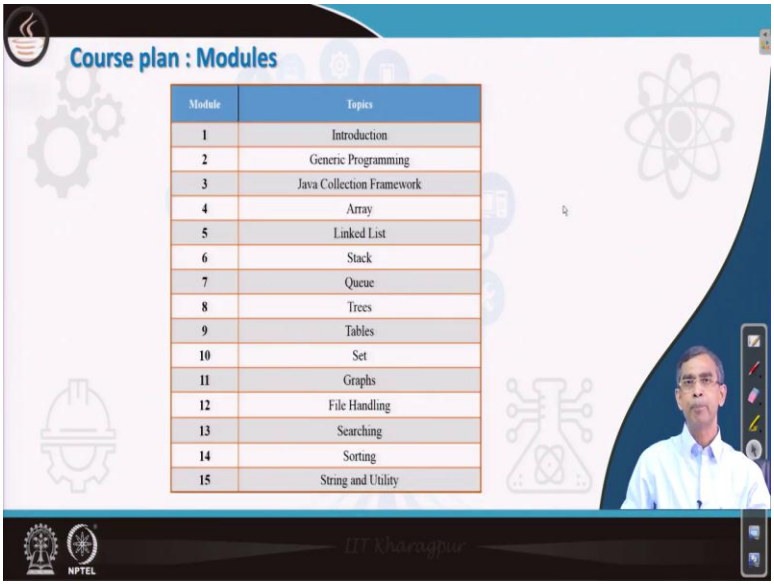
So this is called Java database connectivity supports, JDBC. So all those things are basically part of the Java supports for programming and other than the Java supports for programming, there is a data structure point of view. Java provides three unique what is called the features, they are stored in a, they are available in the form of a package like java dot util is a package, then java dot lang is a package and java dot io is another package.

The java dot util is basically the package which will give a very in-depth ways of dealing the different structure of data and so if you want to have the data structure in full sense, you have to be the master in with this package java dot util. There is a very default package always essential, without this default package no Java programmer can be written and it is the java dot lang package.

In this java dot lang there are several classes out of which one class called the string, this class is very important one class, so learning of this string class is very important and then java dot io, it is basically input, output related one package, which allows you to deal the flow of data from the different source, sources. So these are the three things those are required in order to understand about your data structures.

util managing the data structures, string is a special kind of data with which any program can be written and then input, output is a mechanism how the different form of data from the different sources can be managed. So this way the programming, mainly the core programming, and these are the three utility packages, java dot util, java dot lang and java dot io needs to be learned, so that is what a plan that we will prepare a comprehensive course for 12 weeks to cover these are the three packages, java dot util, java dot lang and java dot io.

(Refer Slide Time: 27:16)



Now our plan is basically here; all the topics those are related to this kind of knowledge gain is planned here how to in, and distributed among the 15 different modules as I mention here. So today the introduction session is going on, so introduction apart from introduction, the topics related to this will be covered for generic programming, where the, it gives an ace to a programmer to deal the data in a generic way, not the specific way.

Generic means in general way. I will discuss about generic programming concept and then generic programming facilities, which is there in Java programming and then Java collection framework, it is basically the huge chunk of java dot util package, will be covered here. This basically the JCF - Java Collections Framework, gives a thorough knowledge about dealing with different type of data structure.

Now different data structures those are there in the theory of computer science namely array, linked list, stack, queue, tree, table, set and graphs will be covered then. And finally towards the dealing with input, output mainly the file handling concept we have to study, apart from this file handling we will also study input stream, output stream, character stream form, so they are called the byte stream and then character stream concept.

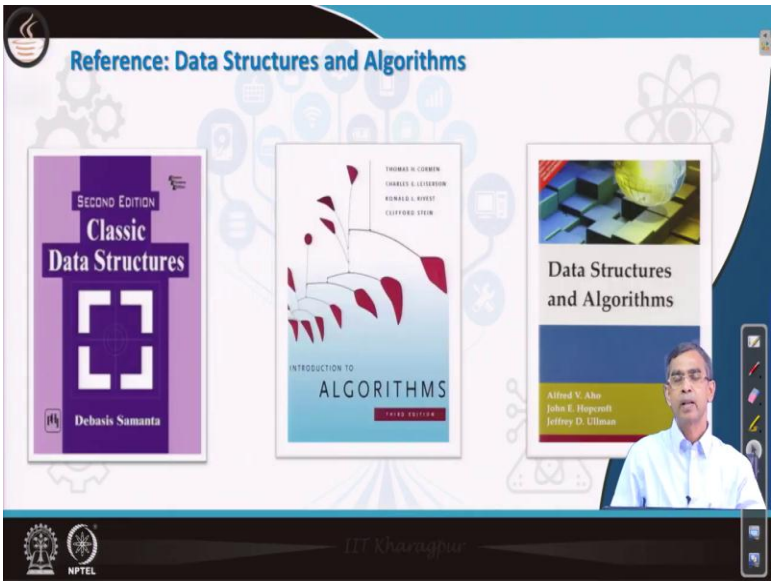
So those things will be discussed. Finally, we will discussed about the sorting and searching algorithms, because the data can be stored from the different sources of different type, but searching and sorting algorithm can vary from one type to another type, so different algorithms are there and not only the different types of algorithm dealing with different type of data structure, rather different type, different algorithms is required that to compare, which algorithms, for example, sorting or searching betas for what type of data.

So this needs one thorough study about the algorithms related to the data structures and finally, the utilities related to string and some miscellaneous utilities will be covered. So this is the plan of the course that we are going to cover in this course.

(Refer Slide Time: 29:38)

And here is the different lectures plan, week wise plan rather, how the (differ) 60 lectures can be spread and what are the different topics will be covered in different lectures will be given there.

(Refer Slide Time: 29:51)



So this is a bar view of the different concept. I will quickly come to a discussion about different resources for studies. As I told you here you have to cover three things the programming aspects, data structures and algorithms and finally their implementation. Now, if you are interested to learn data structure algorithm, there is a good set of books are available. I have listed three very popular books at the moment like Classic Data Structure by Debasis Samanta.

Then Algorithms by, it is basically Cormen and its associate editor and Data Structure and Algorithm which is the very first books written by Aho, Hopcroft and Ullman, these are the three, what is called the very good books in the field of data structures that you can follow. All books are available in Indian reprint.

(Refer Slide Time: 30:49)

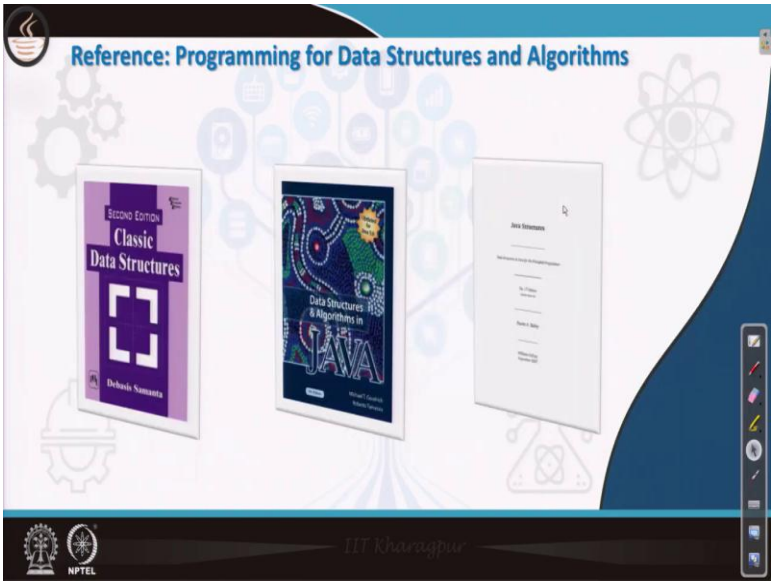
And for programming in Java there is obviously very good book called the Bible of Programming, this is called the Java The Complete Reference written by Herbert Schildt. Apart from the Herbert Schildt there is another book written by Ivor Horton's, it is also have a good coverage of different programming aspects and then there is one pedagogical book for very beginners, if you want to know about object-oriented programming, C++ and Java by D Samanta.

The all those books are available again in Indian reprint, for example, Herbert Schildt book can be from Oxford or it is a McGraw-Hill edition is also there and then this is basically the books from Willie and then D. Samanta books from (())(31:30).

(Refer Slide Time: 31:32)

There are many other books also known at the moment coming in the market they are basically programming related to data structures and algorithms, but they are, they mainly focus on data structures, rather than the programming, here are few set of books I have listed here. You can check it that will be good through learn all those books are there.

(Refer Slide Time: 31:53)

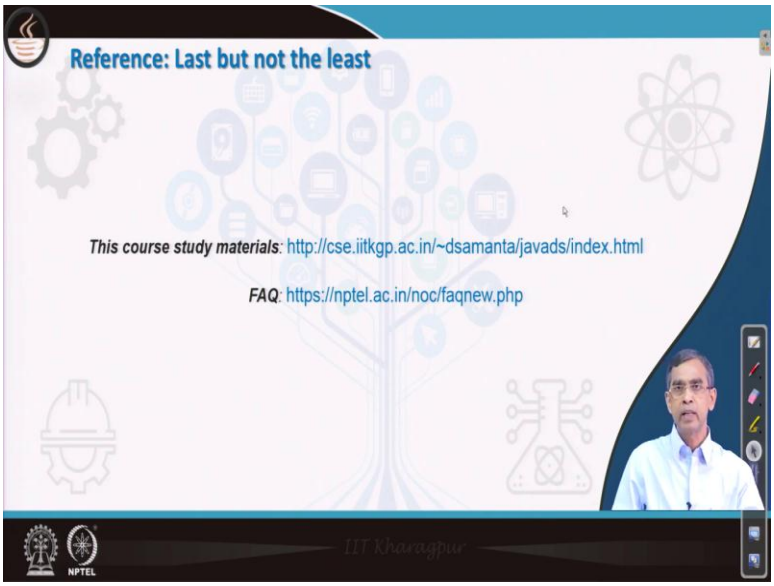


But so far the programming is concerned, I would like to narrow down three books. One books is Classic Data Structure, this books basically covers not only the data structures but implementation of different data structures in a neutral way that means language independent way. But there is one language specific book, Data Structures and Algorithms in Java, it is written by Goodrich and Tamassia, this also books have a lot of coverages you can consider. And another white book is there is a structure from David (()) (32:28) that also you can think about for your study further if you want to have.

(Refer Slide Time: 32:34)

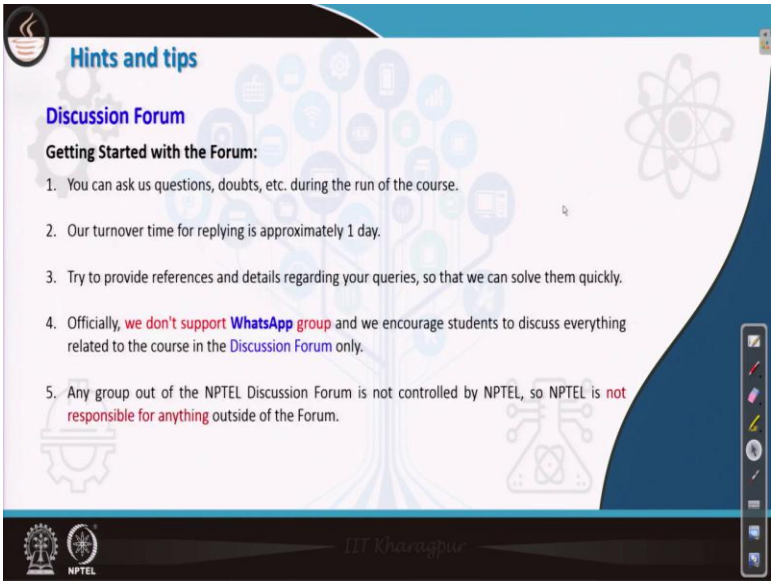
Now there are many resource materials available in the net. Here is, for example, three links I have given one is GeeksForGeeks, Javatpoint and Java Oracle website is the official website for Java programming, where you can get everything in details.

(Refer Slide Time: 32:51)



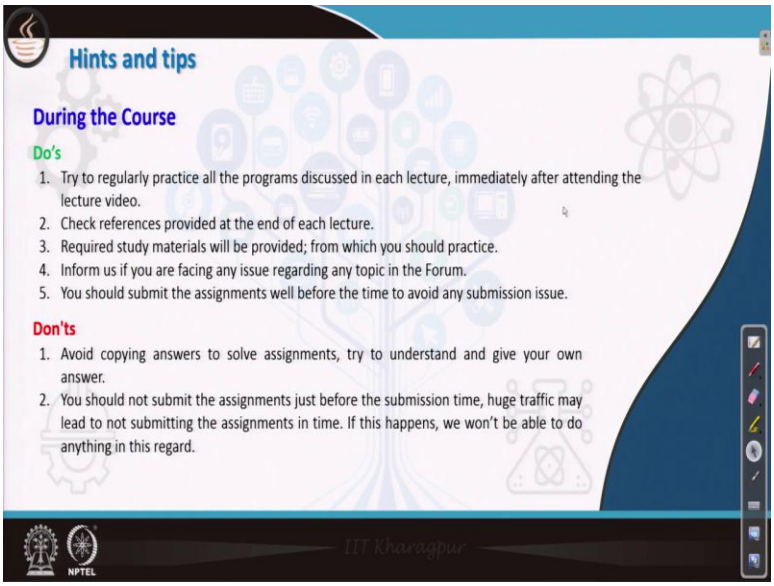
And in addition to this course video lectures and everything we have planned to have all the materials that will be covered related to these subjects you can have a link that you can get it network, all the programs that will be covered in this course also you will be able to get it from there.

(Refer Slide Time: 33:08)



And there will be a discussion forum, you can consider the discussion forum to raise your any questions during the running of this course and you can resolve any doubts. Here are few things that I have listed here how you can take part in the discussion forum, it is free for all the participants.

(Refer Slide Time: 33:27)



Now during the course, I want to see, I want to tell few do's and don't dos's. The first of all you should be regular, you try to write, be updated all the week wise coverage and then week wise assignments or and then try to solve them and another thing is that there will be a lot of programmings will be covered here, in the slides I will give some programs, all those programs will be available in the net given the link from my site, you have to download these programs and practice yourself.

This is very important and without these practicing and everything you will not be able to cope with this course actually. And obviously you have to learn of your own, this is your what is called a good advice, you should not copy, it is not that you want to have a good credit, only earning credit is not important, important is their learning.

This is about the different things that I want to tell about in this introductory session and then our next topics we will start about generic programming in the next lectures, you please join and then have a good fun of this programming language. Thank you.