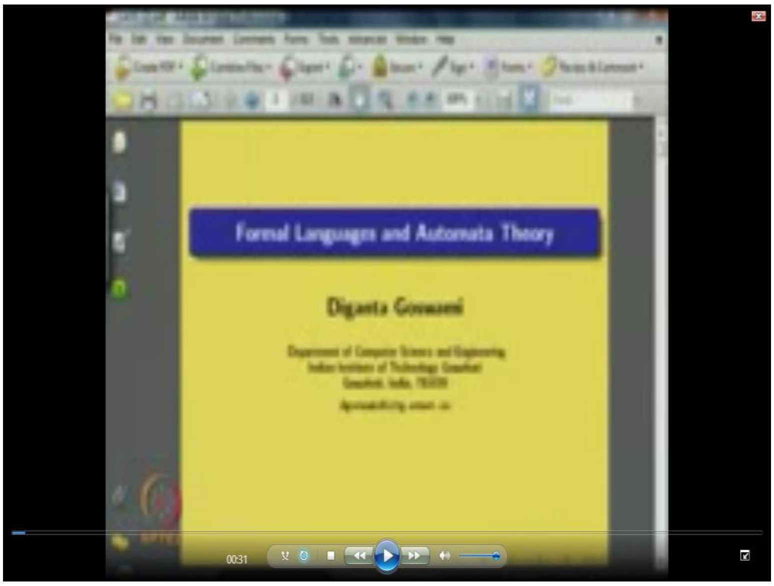
Formal Languages and Automata Theory
Prof. Diganta Goswami
Department of Computer Science and Engineering
Indian Institute of Technology, Guwahati

Module - 01
Languages and Finite Representation
Lecture - 02
Alphabet, Strings, Languages

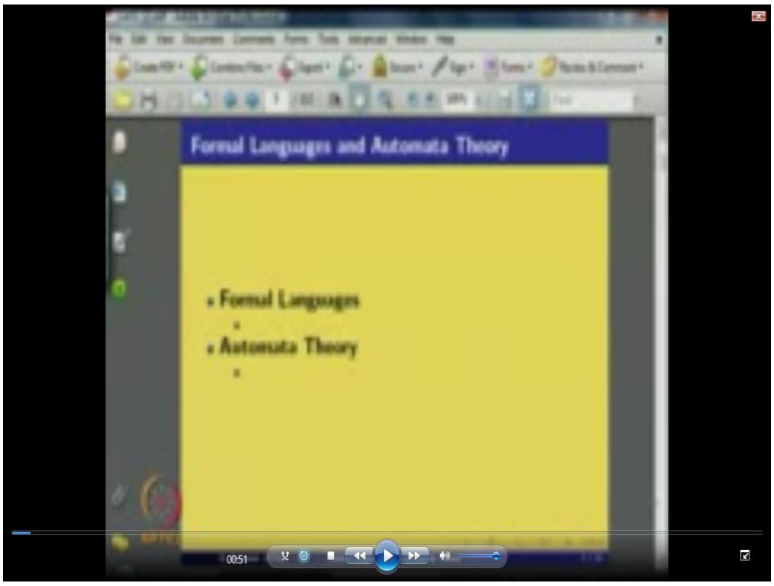
So, in this lecture we see some introduction concepts and define language formally. We will also see some language operations.

(Refer Slide Time: 00:31)



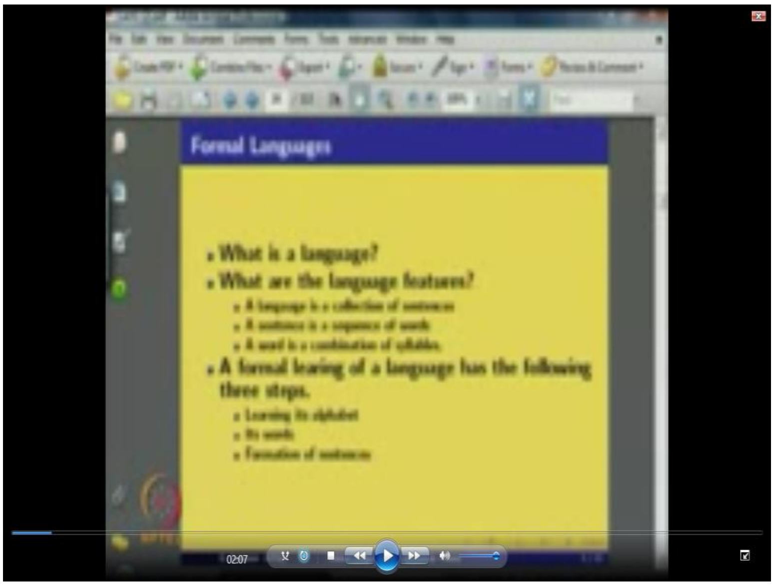
The language can be seen a system suitable for expression of certain ideas, sets and concepts.

(Refer Slide Time: 00:46)



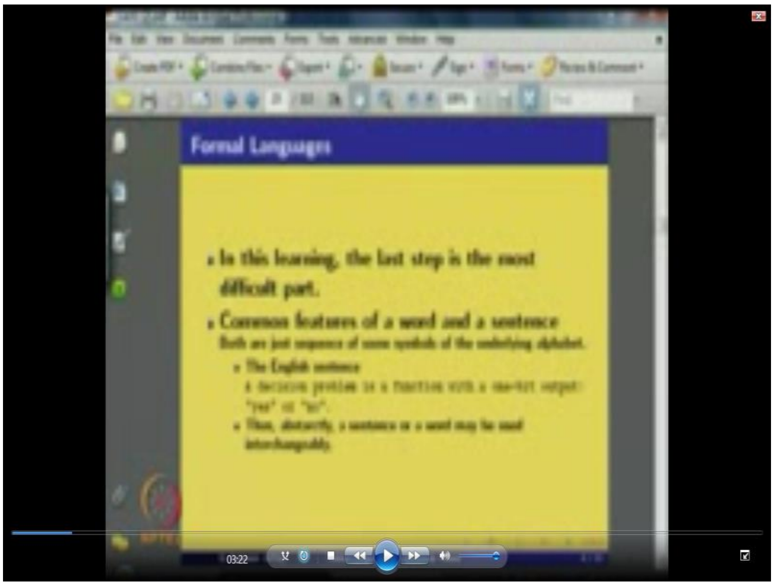
In this course, we have this formal languages and automata theory. We will first see the concepts of formal languages.

(Refer Slide Time: 01:01)



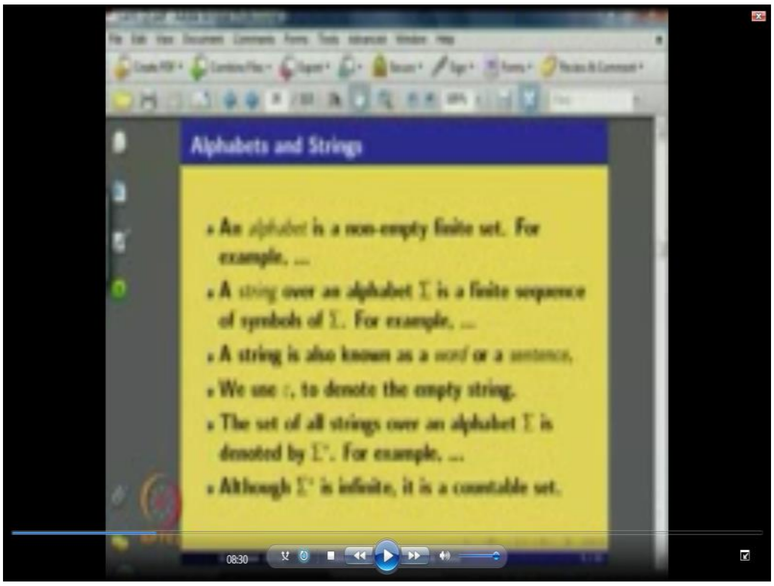
First we will see what are the various features common features across different languages? Because while discussing the languages we need to consider different kinds of languages like natural languages, recommended languages and so on. So, what are the various features of a language or common features across the various languages? We see that a language is a collection of sentences A sentence is a sequence of words and a word is a combination of syllables. So, while formal learning of a language we will be consider ((Refer Time: 01:40)) that there will be a therefore, it is necessary to understand the alphabet of the underlying languages. This is the first step for our learning of a language. Then what are the various words in this language and finally, how to found sentences from these words?

(Refer Slide Time: 02:14)



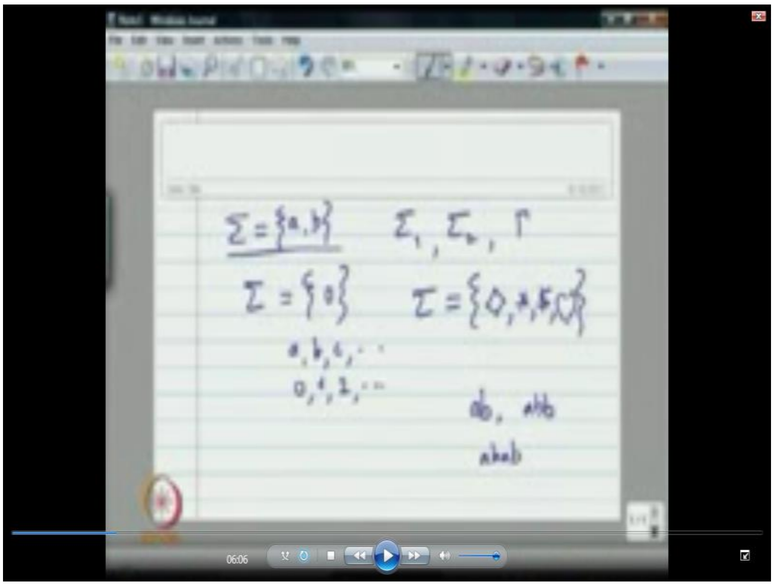
We will see that the last step in a learning of languages is the most difficult. Therefore we will postpone for time being part therefore, and we concentrate ((Refer Time: 02:28)) therefore, see what are common features of a word and a sentence. We find that both are just sequence of which of some symbols of the underlying alphabet. Consider the English sentence a decision problem is a function with a 1 bit output yes or no? We observe that nothing but sequence of symbols from the ((Refer Time: 03:00)) and some other special symbols like comma, question mark, full stop. And a special symbol is a blank space used separately of various words thus abstractly a sentence or a word may be used inter changed. So, here we distinguish between sentence and a word.

(Refer Slide Time: 03:31)



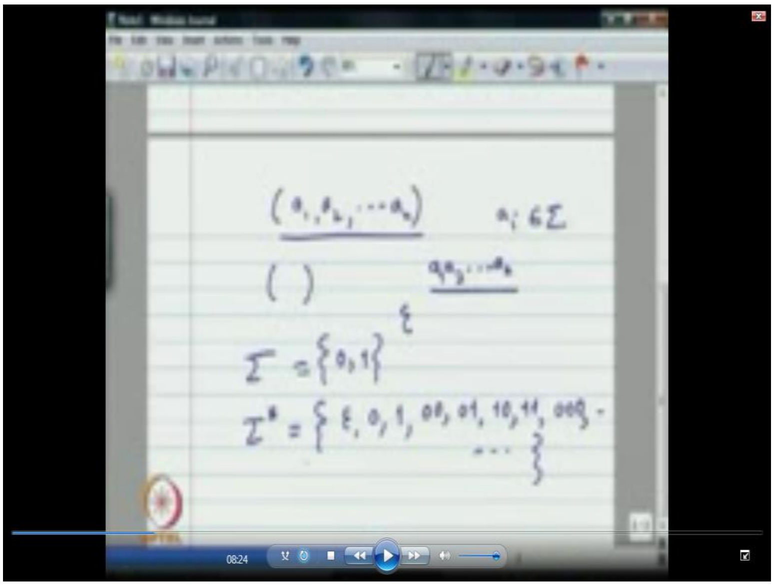
Now, this we can go for definitions of alphabets and strings we define an alphabet is nonempty finite set.

(Refer Slide Time: 03:44)



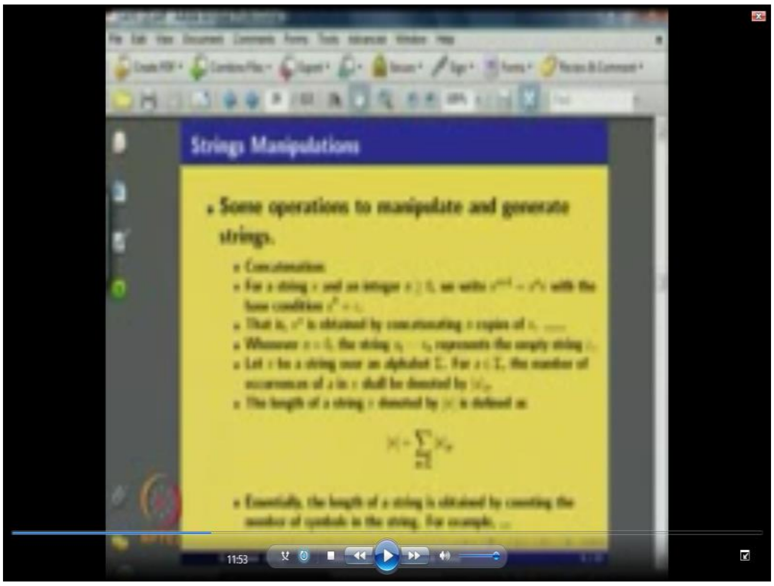
For example, the subset containing a b and consider an alphabet. We denote an alphabet by sigma we have ((Refer Time: 04:04)) alphabet by using a subsets sigma 1 sigma 2. To denote the alphabets sometimes we might also used some of the other as gamma and to denote alphabet. Now, for example, we can have an alphabet like this which contains that the single subset of or we have an alphabet like this starting from special symbols. So, this has 4 different symbols normally we use this alphabet such s alphabets a b c and so on. To denote the symbols an alphabets suppose we use an ((Refer Time: 05:10)) 0 1 2 and so on. To denote symbols, but we have to suppose now define string a string over an alphabet a string is a finite sequence of symbols of sigma. For example, just consider sigma is subset containing a b is string over this alphabets and word is alphabet. Similarly, a b b is also is string containing 3 symbols having then a b a b is also a string of this alphabet 4 symbols of this.

(Refer Slide Time: 06:15)



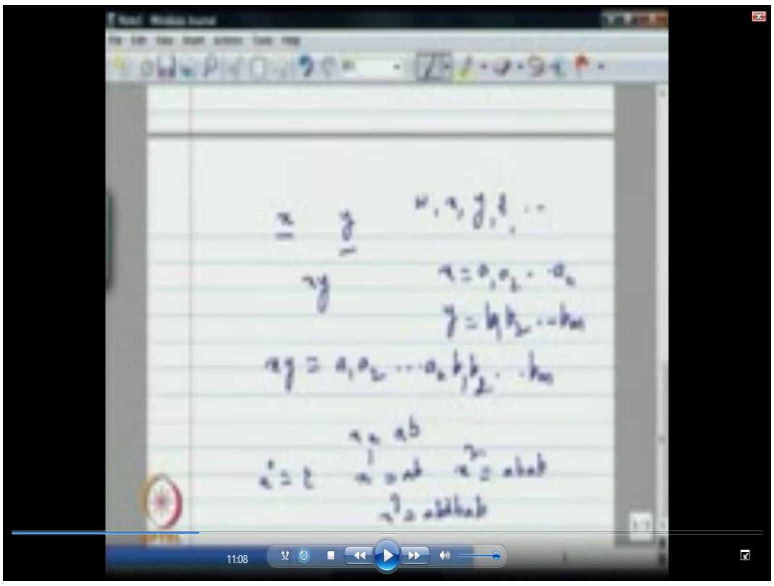
Normally to define we denote a sequence like we say a 1 a 2 a n for a i belong to sigma ((Refer Time: 06:35)) by defining the string the sequence by a 1 a 2 a n to and to enter ((Refer Time: 06:40)) just supporting the symbols. So, this sequence is denoted by this is 3 in the context we see that the empty string is denoted by ((Refer Time: 06:58)). So, in this context, we will be using the special symbol epsilon to denote the empty string, because this is bigger. Now the string is also known as a word or a sentence therefore, we will using the terms strings, word or sentence. We see that the set of all strings over an alphabet sigma is denoted by sigma string. For example, if sigma is say 0 1 then sigma star is say epsilon because epsilon is a string to 0 is a string 1 is a string 0 0 0 ((Refer Time: 08:10)) 1 1 0 1 1 0 0 0 and so on. So, ((Refer Time: 08:15)) by using symbols of sigma we get the sigma star similarly, now although sigma star is infinite it is a countable set.

(Refer Slide Time: 08:50)



Now, let us see some operations as much you can manipulate and we can general strings first let us see the operations, concentration.

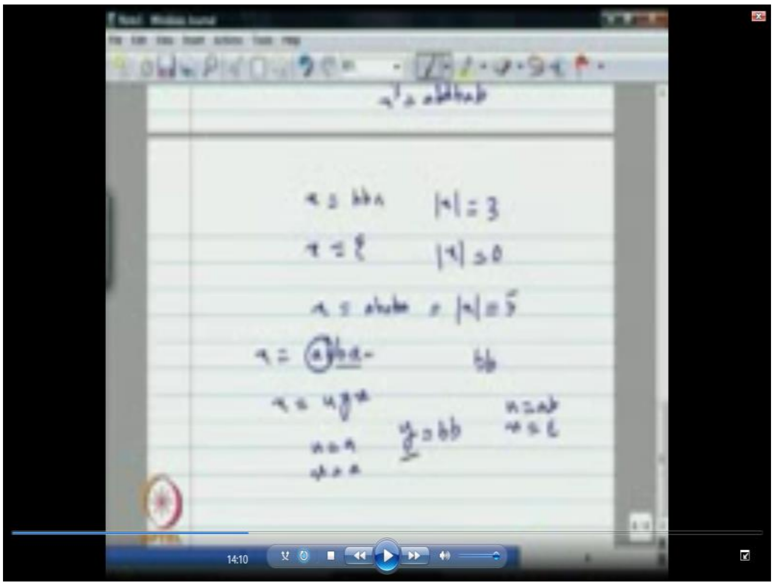
(Refer Slide Time: 09:06)



Now, we will define the x is a string normally we use the sigma ((Refer Time: 09:13)) n alphabet to denote for example, w x y z they are used to denote a string just consider string x and a n another string y. So, the concatenation of x and y is denoted as x y if x equal to a 1 a 2 say a n. So, y equal to b 1 b 2 say b m then x y is the symbol ((Refer Time: 09:45)) for this x y is the string that we can a 1 a 2 say a n b 1 b 2 b n ((Refer

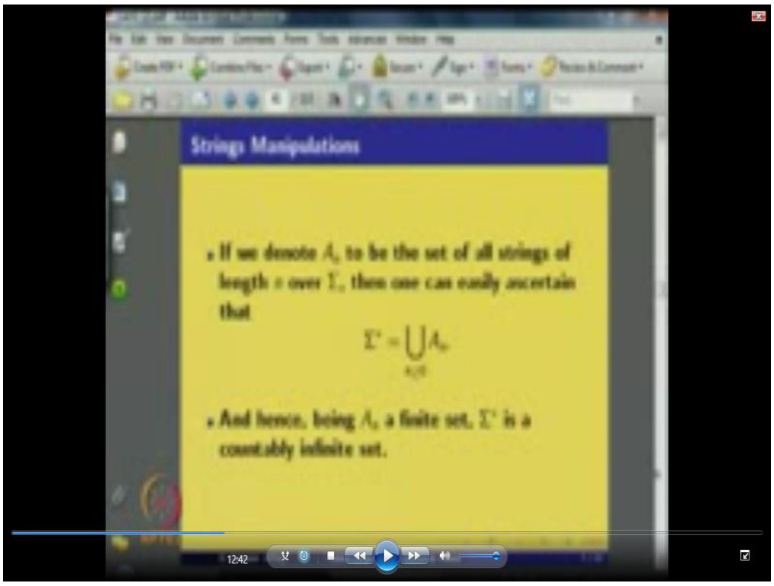
Time: 10:00)). Again x y we has for a string x and an integer x is greater than or equal to 0 then we write x to the power n plus 1 equals to x to the power n x with the base condition that is ((Refer Time: 10:27)) x power 0 equals to empty set. Then that is x n is nothing but string is obtained by concatenating n copies of x whenever n equals to 0 the string x 1 and so on equals to x n represents the empty string f. For example, if x equals to the string a b then x to the power 0 is sigma, x to the power 1 is a b x to the power 2 is a b a b x to the power 3 is a b a b a b and so on. Now, let x be a string over an alphabet sigma for every symbol belongs to sigma the number of occurrences of a in x shall denoted by x suffix n. The length of a string x denoted by suffix x is defined as summation of number of occurrence of suffix x n that is essentially the length of a string is obtained by counting the number of symbols in the string.

(Refer Slide Time: 11:57)



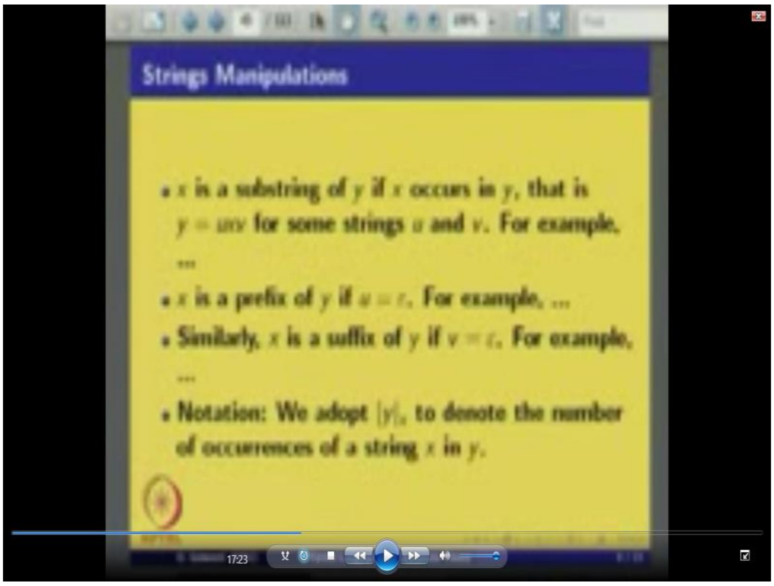
For example, if x equals to say b b a ((Refer Time: 12:00)) suffix x is equals to 3 ((Refer Time: 12:05)) if x equal to epsilon ((Refer Time: 12:09)) empty string. Then the ((Refer Time: 12:10)) string is 0 and then x equals to say a b a b a then suffix x is equals to pi.

(Refer Slide Time: 12:27)



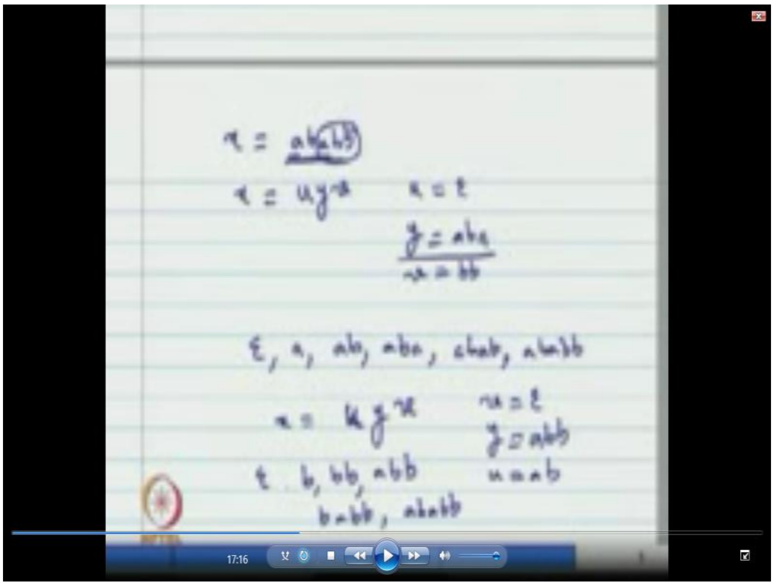
If we denote A n to be the set of all strings of length n over sigma then one can easily observe that sigma star is a union of A n for n greater than or equals to j. And hence since A n is a finite set sigma star is a union ((Refer Time: 12:47)) countable infinite set.

(Refer Slide Time: 12:55)



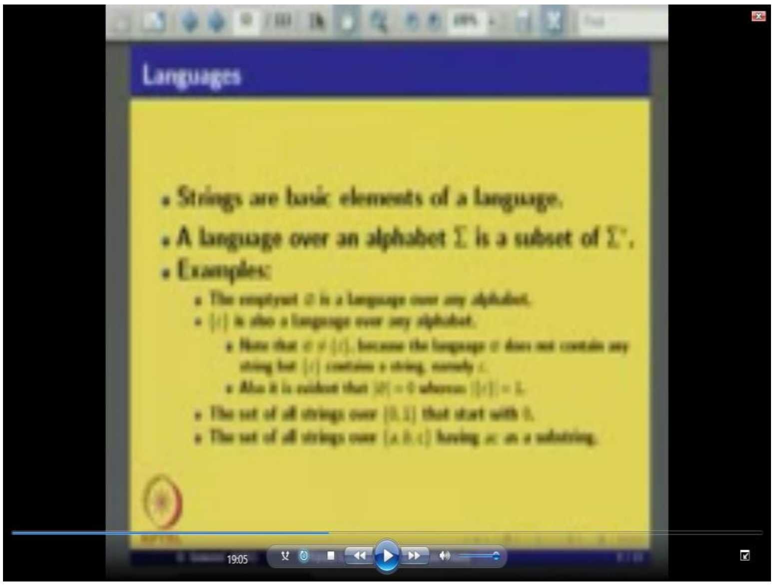
Let us see some more string manipulations we define a substring like this if x is a substring of y if x occurs in y that is we can write as ((Refer Time: 13:09)) y equals to u x v for some strings u and v belong to sigma star. For example, if x equals to a b b a then we say that b b is a substring of as ((Refer Time: 13:35)) x is equals to u y v where u equals to n v v equal to the string and y is the string bb. So. in this case the string u v y a sub string of x similarly, this b a is a substring of x in this case u equal to a b in this part and b equal to epsilon ((Refer Time: 14:10)) many other parts from this string x. Therefore, x is a prefix of y if this u equals to epsilon.

(Refer Slide Time: 14:32)



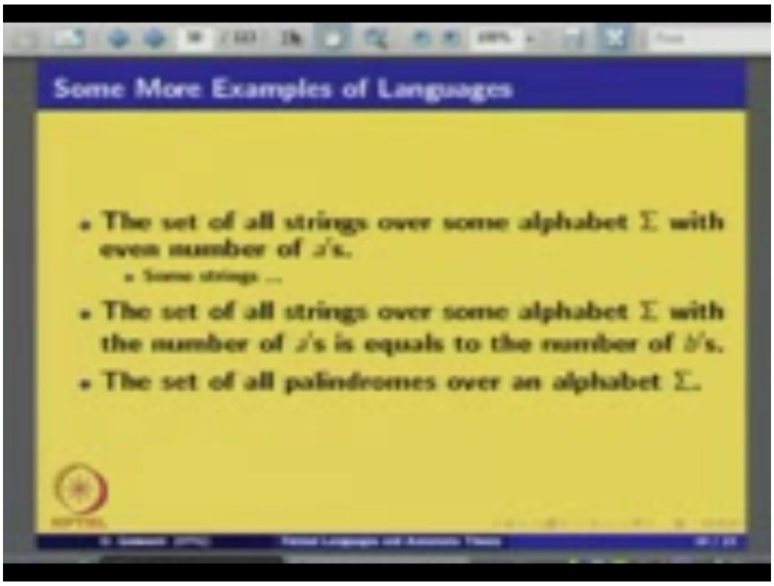
For example, if x equals to a b a b b then see that x can be written as u y v where u equal to epsilon y equal to a b a and u equals to b b. That means x can be written as u y v here a becomes epsilon y equal to a b a and similarly, see that the string a b is a prefix of x. So, similarly, ((Refer Time: 15:20)) a is prefix of x a b a is also a prefix of x a b a b is also prefix of x similarly, the whole string a b a b b is prefix of x. So, we can say that axiom also prefix of x interesting these are the only strings which are prefix of x. Similarly, ((Refer Time: 15:58)) x is suffix of y if v equal to epsilon, if we write for the self string a b a b b. We write x to the say a u y v where u equals to epsilon y equal to a b b and u equal to a b such a case we say that and y that means this is a b b is suffix of x. Similarly, we can find out all are the suffixes ((Refer Time: 16:55)), b is a suffix of x bb is a suffix x a b b is suffix of x. Similarly, b a b b is suffix of x ((Refer Time: 17:05)) a b a b b suffix of x say that f is a suffix f b string x ((Refer Time: 17:24)) this notation is we have some y with suffix x to denote the number of occurrences of a string x in y.

(Refer Slide Time: 17:35)



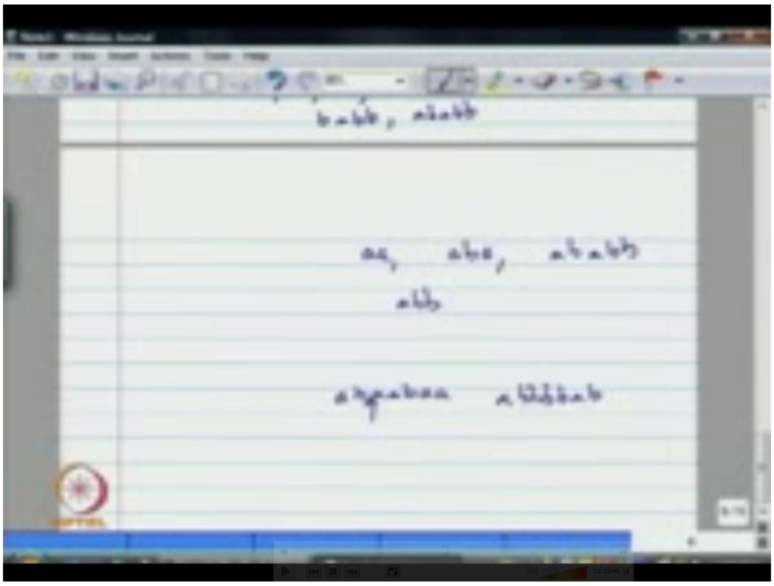
We see that strings are basic elements of a language let us define so many languages. We define a language over an alphabet sigma with a subset of sigma star a sigma star contains all the strings from the alphabet any subset of string consist way language. Let us see some various examples since the language may be any subsets. So, empty set pi is a language over an alphabet the single ton set containing an empty string ((Refer Time: 18:18)) is also a language over any alphabet. Please note that these 2 sets pi empty set and the single ton set containing a string or not identical or any string, because the language empty set does not contain any string but single ton set contains a string namely epsilon. Also we see that the cardinality of 0 the cardinality of the 1 another example of a language is for example, say that 0 1 similarly, the set of all strings over a b c having example of a language.

(Refer Slide Time: 19:13)



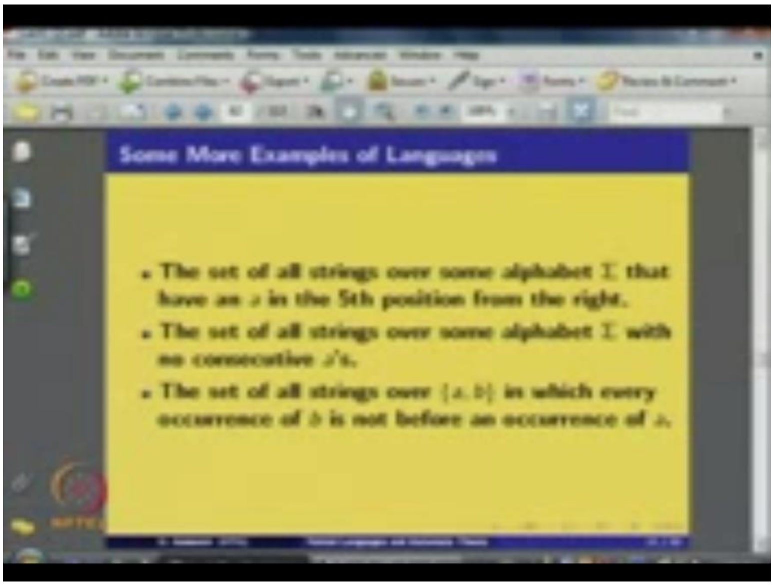
Let us see some more examples. So, in the set of all strings over some alphabets sigma with even number of a's let us consider what are the strings that may of the language since the language ((Refer Time: 19:34)) the string containing even number of a's.

(Refer Slide Time: 19:40)



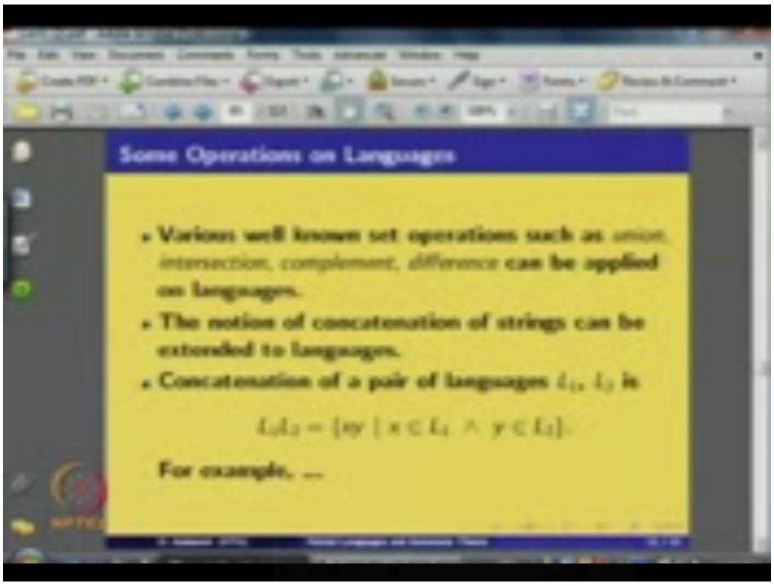
Therefore a a will be in the language the string a or a b a will be in the language then a b a b b will be in the language the numbers of strings. The numbers of a's on all the strings are, but a b b does not belong to this language because strings all are even ((Refer Time: 20:09)). Similarly, the set of all strings over some alphabet sigma a and b with the number of a s is equal to the number of b s we can an example of a language. We can found out many examples which are in the language and which are not in the language. Consider the set of all palindromes over an alphabet sigma similarly, the set of all palindromes over an alphabet sigma. Let us see an example always palindrome means all those of words or strings which deals from left and right.

(Refer Slide Time: 20:50)



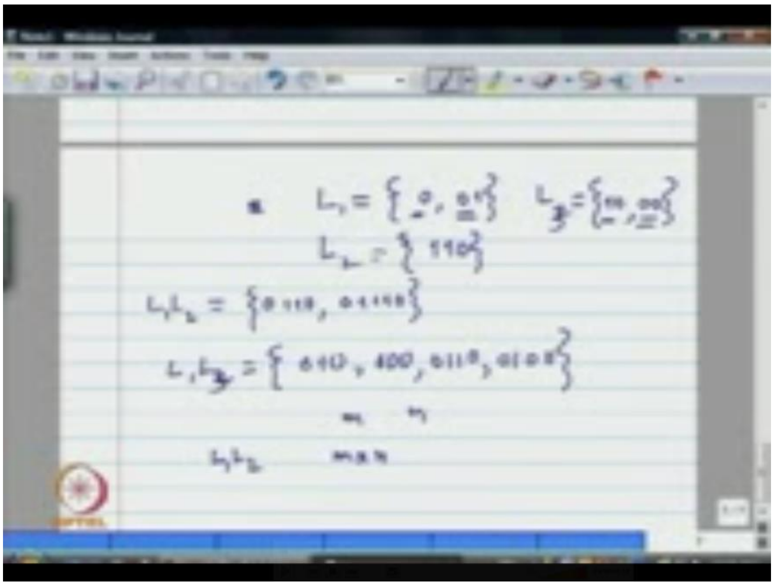
The set of all strings over some alphabets sigma that have a n a in the fifth position from the right. Just consider this language and see that the string an a b a a b a a belongs to this language, because the symbol a appears in the fifth position from the right whereas, a b b b b a b does not belong to the language, because in the fifth position of the right it does not appear. Similarly, the set of all the strings over some alphabet sigma with no consecutive a s and we can give any other example of a language. Similarly, the set of all strings over a b in which every occurrence of b is not before an occurrence of a.

(Refer Slide Time: 22:00)



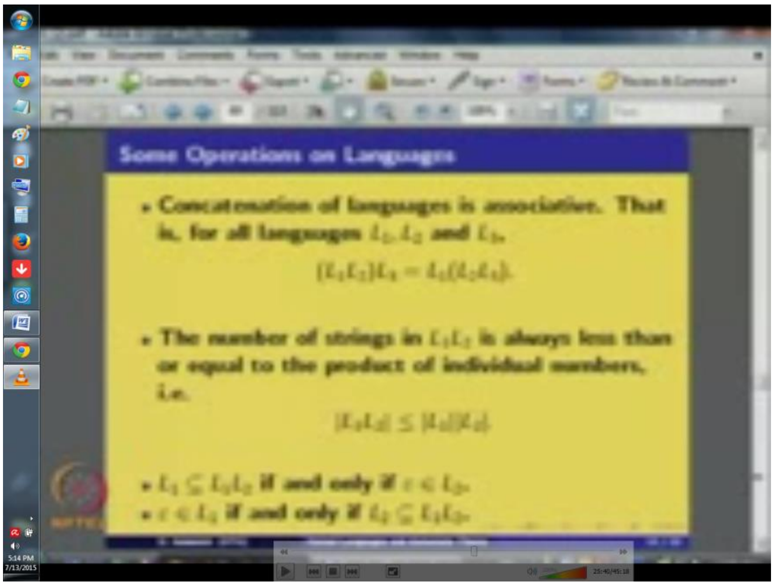
Now, since the language of sets we can apply a various set of operations such as union , intersection, complement, difference can be applied and so on languages. Similarly the notion of concatenation of strings can also be extended to language. We define the concatenation language of a pair of languages L 1 L 2 is lone L 2 is equals to set of all streams x y such that x belongs to L 1 and y belongs to L 2 let us see some examples.

(Refer Slide Time: 22:47)



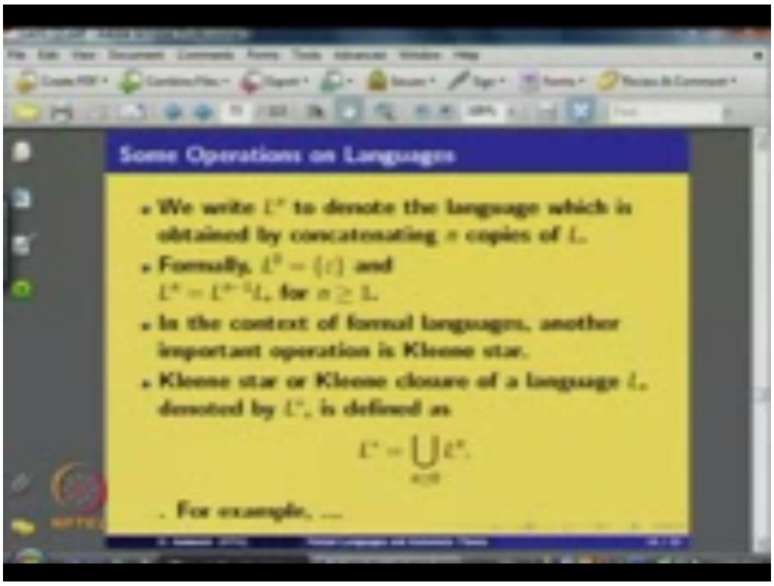
Say L 1 is a language contains set of strings of 0 0 1 that is L 2 is the language containing the string say of 1 1 0. Then L 1 and L 2 L 1 and L2 are the conclusion of L 1 and L 2 equals to set of all things that we have by concatenation first is from L 1 and second one from L 2. That means 0 1 1 0 then 0 1 1 1 0 this 2 things that we can get ((Refer Time: 23:38)). Now, suppose L 1 is the string language has set L 2 has same and L 3 has other language so it is 1 0 and 0 0. Then L 1 L 2 ((Refer Time: 24:05)) is set of all strings 0 1 0 then 0 0 0 and 0 1 1 0 and 0 1 0 0 it has 4 strings. In general L 1 has m strings and L 2 has n strings then we have L 1 L 2 is m into n.

(Refer Slide Time: 24:52)



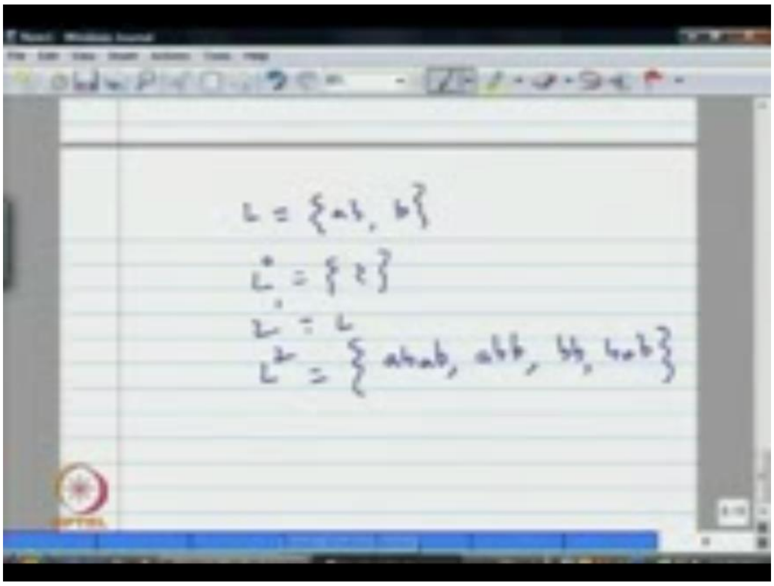
See that concatenation of languages is associative that is for all languages L 1 L 2 and L 3 L1 concatenation L 2 concatenation L 3 of L 1 L 2 is equals to L 1 is concatenation of L 2 and L 3 concatenation. So, in general we write it as L 1 L 2 L 3 ((Refer Time: 25:22)) then the number of strings in L 1 L 2 is always less than or equal to the product of individual numbers. We see that is L 1 L 2 is subset of L 1 if and only if belongs to L 2 and also the empty set belongs to L 1 if and only if L 2 subset of L 1 L 2.

(Refer Slide Time: 25:50)



We write L power n to denote the language which is obtained by concatenating n copies of L. That means L to the power 0 is epsilon ((Refer Time: 26:00)) and L to the power of n is concatenation of L to the power n minus 1 and L for all n greater than or equal to 1.

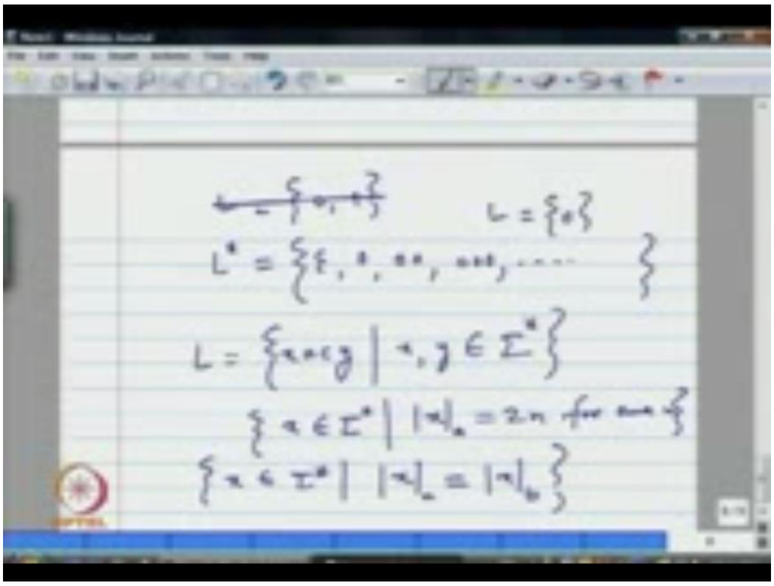
(Refer Slide Time: 26:23)



For example, if L is a a b and b and L to the power 0 is equals to empty set L to the power 1 is equals to 1 L to the power 2 we just concate the a b a b a b b b b b a b and belongs to by applying L with L. Similarly, in the context of formal languages another important operation in kleene star. So, define kleene star of L language L s union of or

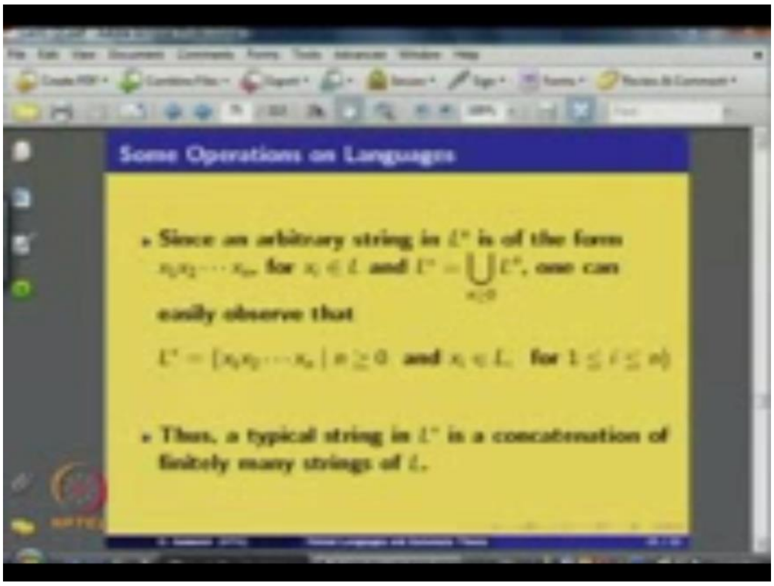
((Refer Time: 27:28)) kleene closure of a language L denoted by L power n is defined as union of n greater than or equals to 0 L power n.

(Refer Slide Time: 27:49)



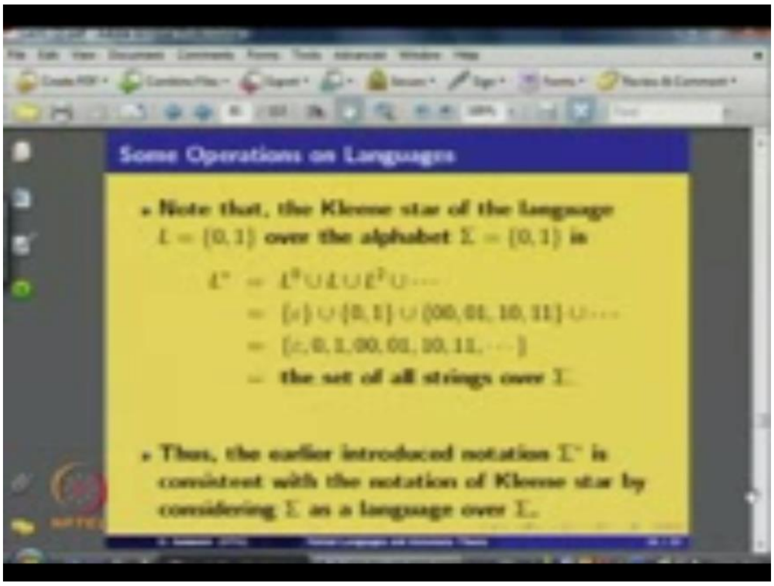
Consider for example, if L equal to 0 1 also if L equal to 0 then L star will be union of ((Refer Time: 28:10)) that means epsilon L 0 1 that means the 0 h f L 0 2 f is 0 0 L 0 3 is 0 0 0 and so on that means we get is all the strings that 0s.

(Refer Slide Time: 28:42)



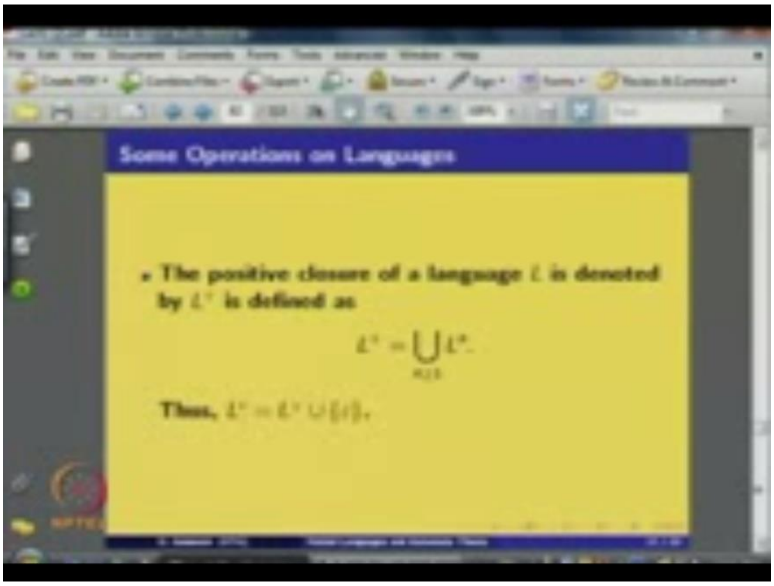
Since an arbitrary string in L n is of the form x 1 x 2 and so on x n for x 1 belongs to L and L star equals to union of all those for n greater than 0. We see that x 1 x 2 x n n is greater than or equals to 0 and x i belongs to L for 1 less than or equals to i greater than or equals to n. That means thus a typical string in L star is a concatenation of finitely many strings of l.

(Refer Slide Time: 29:06)



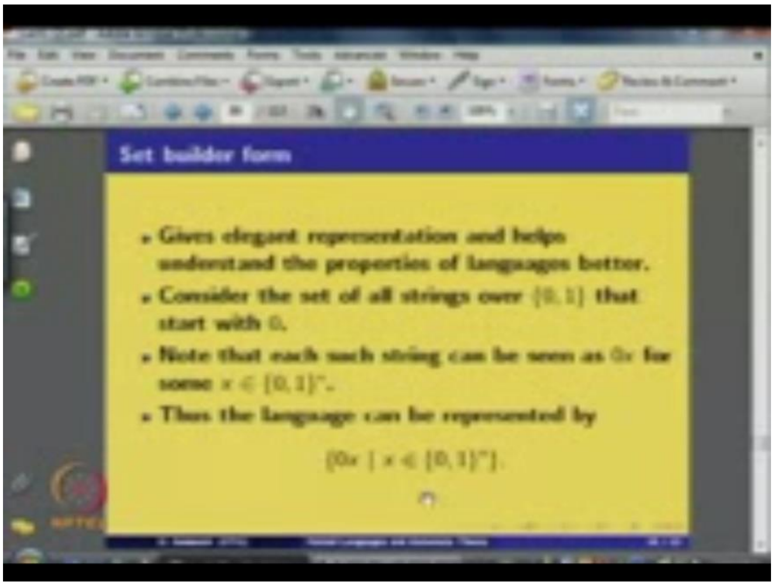
Note that the kleene star of the language L equal to of 0 1 over the alphabet sigma equals to 0 1 is L star equals to L 0 union L union L 2 and so on. So, it is nothing but, empty union 0, 1 union 0 0 0 1 1 0 1 1 and so on. Eventually we will get the set of all strings over sigma that means if we take sigma as a language over sigma then the notation of sigma star is consistent with the kleene star.

(Refer Slide Time: 29:52)



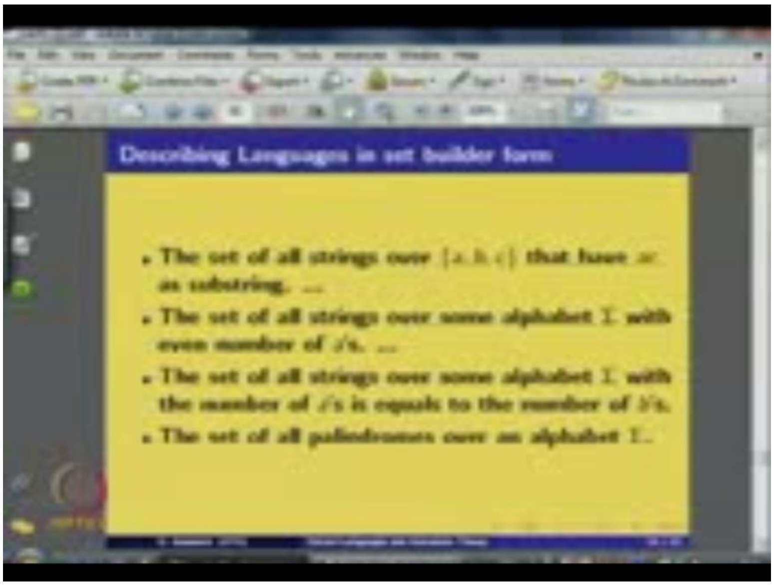
To define positive closure of a language L that means thus L star equals to L plus as union of all L to the power n that means L star is simply L plus union. So, it become epsilon now we are discussed about the different languages.

(Refer Slide Time: 30:20)



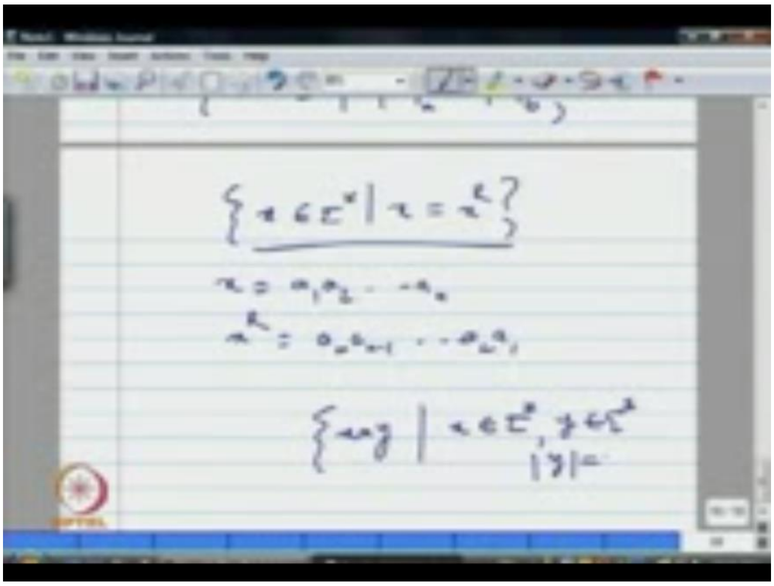
But if we describe the language using form gives elegant representation and helps us understand the properties of languages better. And also again thus consider that all strings 0 1 that starts with 0 in let us represent in different form of languages. Let consider the we can see that every string here in this language can be written as ox that means 0 giving a first place and x means any string over 0 1. Therefore, we can write it as set of all strings are 0 x such that x belongs to 0 1 star for the even languages.

(Refer Slide Time: 31:07)



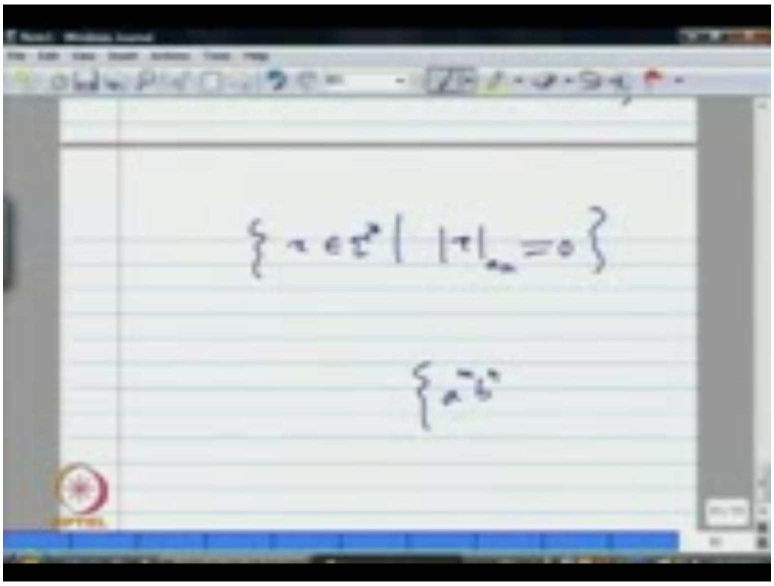
Just consider this set the set of all strings over a b c that have ac as substring and so on how do you referred using regular form any string in this language is written as x a c y. So, that x may be seen over a b c 1 any string over abc must be a distinguish therefore, we can represent this language as L equal to L x a c y. Such that x and y belongs to sigma star where sigma is a b c thus is the form for the given length. Similarly, the set of all strings over some alphabet sigma with even number of a’s the ((Refer Time: 32:00)) form by using. We adopted for example, we can write it as any string x in this case x that belongs to sigma star. This is basically a b such that the number of occurrence of a in x is 2 x n for some n it shows it how it present is even language containing even numbers of x. Again show so the set of all strings over some alphabet sigma with the number of a s is equals to the number of b s. Again the same notation comes all those things x belongs to sigma star and the number of occurrence of a equals to number of occurrence b consider this the set of all palindromes over an alphabet sigma to represent this we use the notation reverse.

(Refer Slide Time: 33:35)



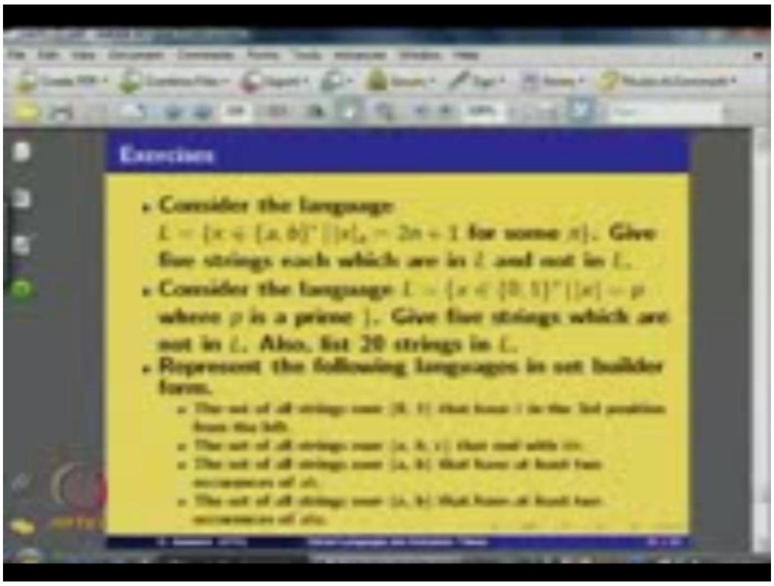
That means if x equal to the palindrome sigma to represent this the set of all strings x belongs to sigma star such that x equal to x power r where x to the r nothing but the reverse palindrome. That is if x equal to a 1 a 2 a n then the x r is simply a r a n equals to an minus 1 and so on a 2 a 1. Since the every string will be identical from left and right we had ((Refer Time: 34:13)). Similarly, one can find out that similar form for all these languages the set of all strings over some alphabet sigma that have a n a in the fifth position from the right. So, in this case we can write it as x ay since a must appear in the kleen position the set of all strings are the x ay such that x ((Refer Time: 34:48)) belongs to sigma star y may be string in the sigma from. But the condition that the length of 1 must be equal to a must be appear in the fourth position its fifth position. Similarly, the set of all stings over some alphabet sigma with no consecutive a‘s again we can representative each other form like.

(Refer Slide Time: 35:21)



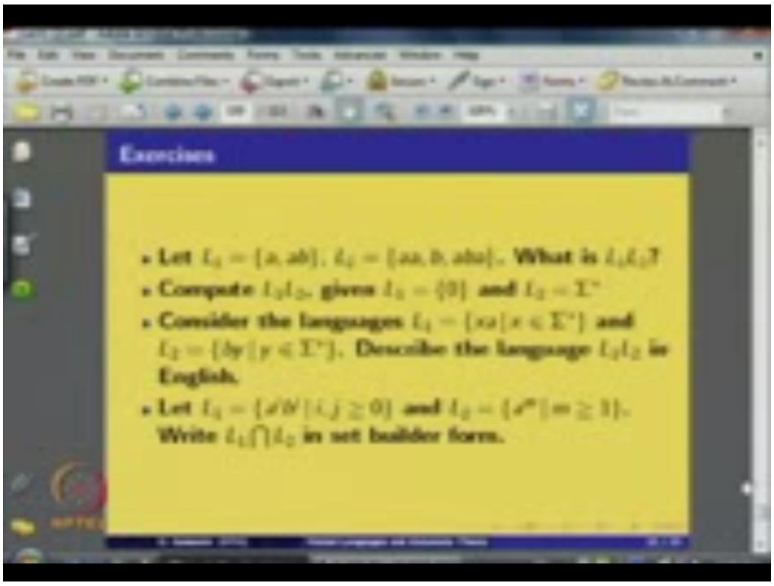
So, all strings x belongs to sigma star such that there is no occurrence of the string a a 0. So, this is equal to g the set of all strings over a, b in which every occurrence of b is not before an occurrence of a so occurrence of b should not be before an occurrence of a. So, every string here must be of the form a to the power of m b to the power of n. So, every occurrence of b must be precedence by ((Refer Time: 36:20)) if a occurs then b cannot be occurs before a. So, all those strings in this from for some m n is greater than or equal to 0.

(Refer Slide Time: 36:47)



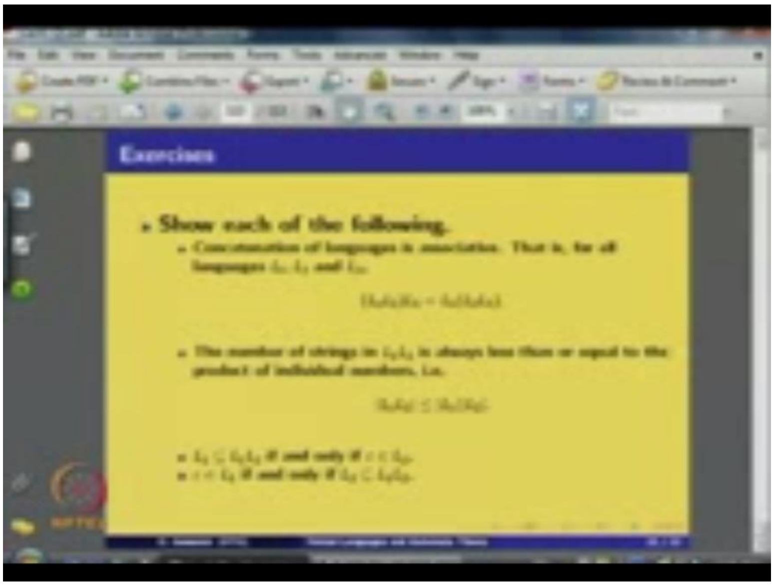
Now, let see some exercises consider the language L such that which contains all the strings over a b such that the number of a s is what, because it and give 5 strings ((Refer Time: 37:09)) which are in L and not in L consider the language L containing all those strings from 0 1. So, that the length of a string become every string of x is equals to p where p is a prime give 5 strings which are not in L also list 20 strings in L represent the following languages in set builder form the set of all strings over 0 1 that have one in the third position from the left. The set of all strings over a b c that and with b the set of all strings over a b that have at least 2 occurrence of a b the set of all strings over a b that have at least 2 occurrence of a b a.

(Refer Slide Time: 38:07)



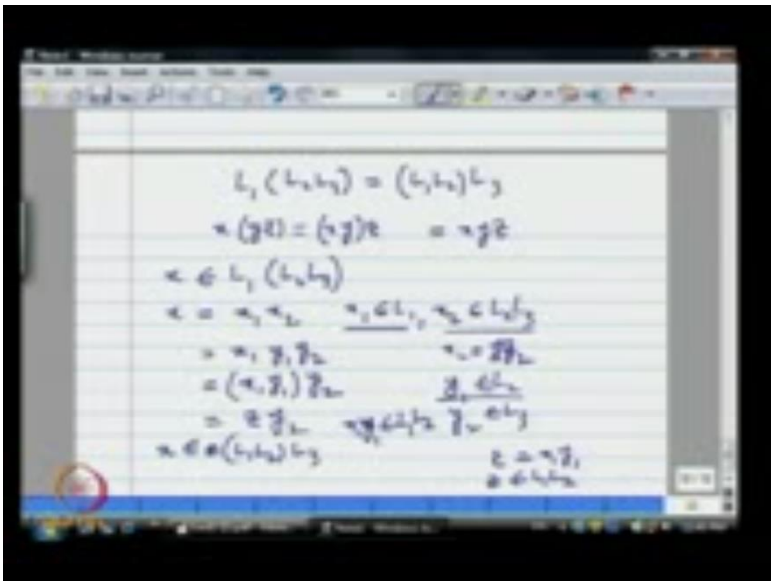
Let see some more exercises that L 1 is this language from a a b L 2 is a a b a b a lone L 2 ((Refer Time: 38:20)). Similarly compute L 1 L 2 given L 2 subset of 0 and L 2 sigma star consider the languages L 1 is x a x belongs to sigma star and L 2 by y belongs to sigma star. Describe the languages L 1 L 2 in English let L1 1 be and L l be show L1 instruction L 2.

(Refer Slide Time: 39:01)



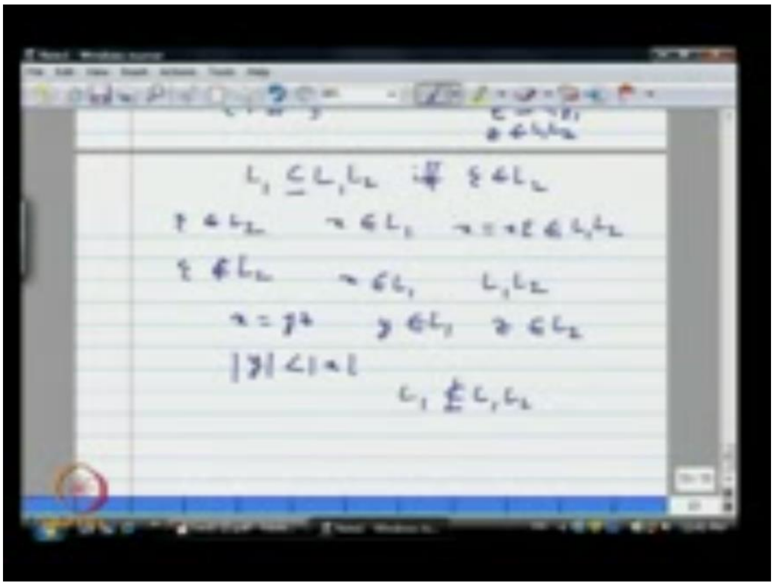
And properties that we discuss earlier so that concatenation of all languages is association that is for all languages L 1 , L 2 and L 3 L 1 at the concatenation, but L 2 concate at L 1 L 2 L 3. So, that the number of strings L 1 L 2 less than or equal to the product of even numbers L1 and L2 if and only if epsilon belongs to L2. Similarly, epsilon belongs to L 1 if and only L2 is a subset of L1 L2 let us show that concatenation of languages is associative.

(Refer Slide Time: 40:00)



That means if L1 L2 and L3 are languages then L 1 concatenation L 2 L3 is equal to L1 L 2 concatenation L3. To prove this ((Refer Time: 40:19)) the concatenation of string is also associative that means if x y z are strings then first concatenate y and z and then concatenation of x it will equivalent to x y concatenation z. So, we can write it as x y z now to show that the concatenation of string is considering language is associated. Just consider any string x which belongs to L 1 concatenation L2 L3. By definition we see that x can be written as x1 x2 where x1 belongs to L1 and x2 belongs to L2 L3. Therefore, since x2 belongs to L2 L3 we know that x2 can be written as some y z or y 1 y2 where y1 belongs to L2 and y 2 belongs to L 3. That means x equal to x1 x2 x1 y1 y2 y1 belong to L2 and y2 belong to y3 since conclusion of string is associative. We can write it as x1 y1 y2 now again here x1 belong to L1 and y1 belong to x2. Therefore, x1 y1 belongs to L1 L2 and we can write it as sum z y2 where z equal to x1 y1 and z belong to L1 L2. Therefore ((Refer Time: 43:16)) x belongs to L1 L2 concatenation L3 therefore, every string ((Refer Time: 43:28)) that belong to L1 L2 L3 x belong to L1 L2 L3 similarly, the conversion of sorry.

(Refer Slide Time: 43:45)



Now, let us prove a property that L1 subset of L1 L2 if and only if epsilon belongs to L2 ((Refer Time: 43:48)) for instance if epsilon belong to L2. Then for any x belonging to L1 we have x equal to x epsilon that belongs to L1 L2. On the other hand suppose epsilon does not belong to L2. Now, note that the string x belong to L1 of for this length in L1 cannot be L1 then this, because if x equal to yz for some y belonging to L1 z

((Refer Time: 44:42)) belonging to L2 then length of y is less than length of x. Now, this is the contradiction towards that x is of shortest length in L1 therefore, L1 is not a subset of L1 L2 hence it proved.