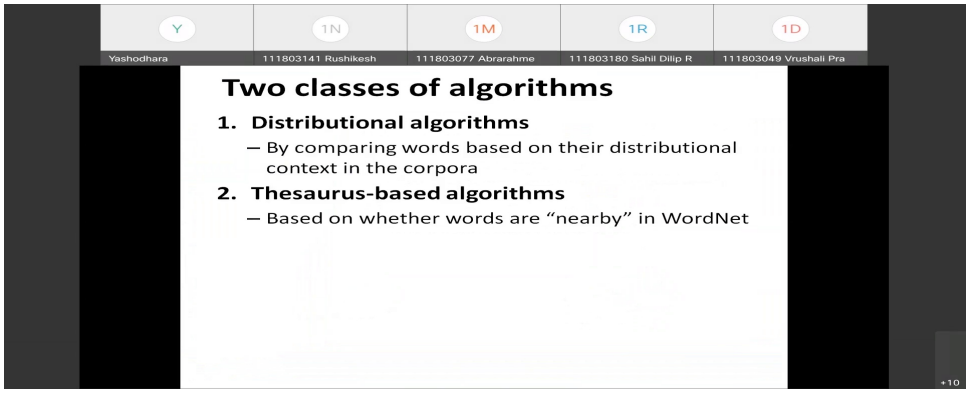
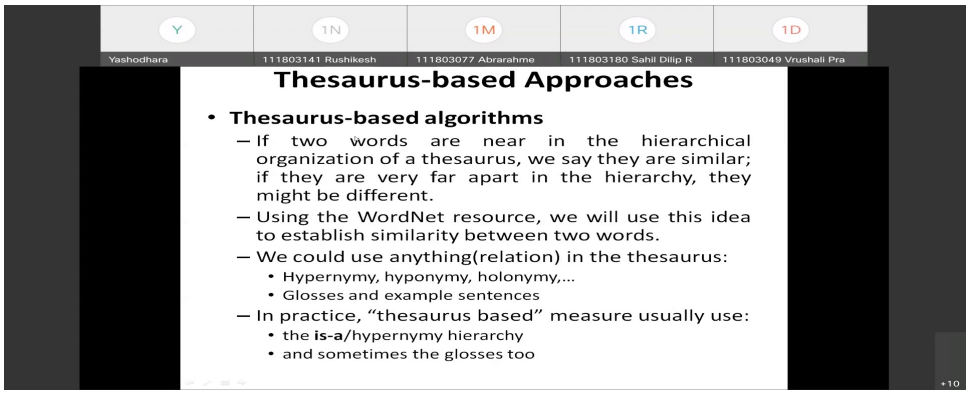
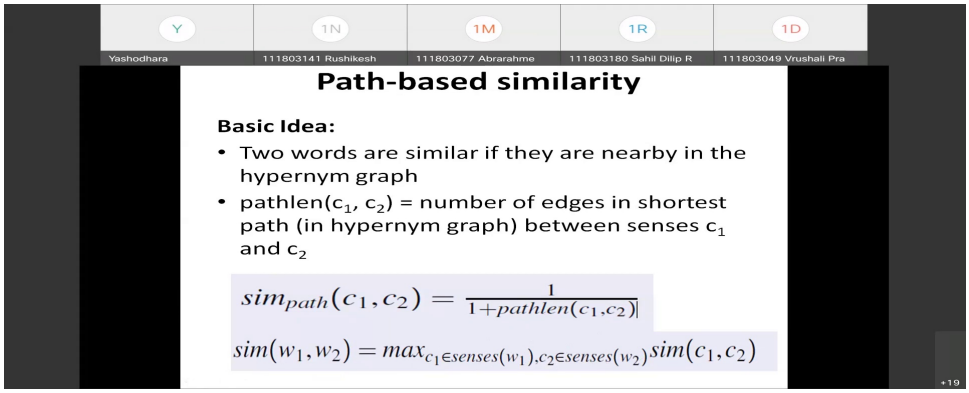


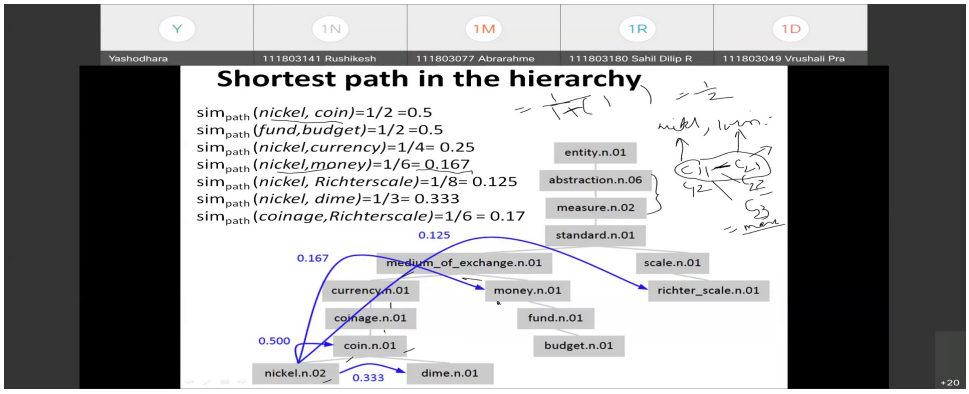
So if the words are nearby in the Iraqi organization, the thesaurus, we say they are similar.And if they are very far apart in the R key, they might be different. So the idea of finding the similarity using the knowledge based approach would be to find out the distances between the two words.And if you find out that the distance between the two words is minimum or smaller than we could predict that they are most similar. And when we say the two words are near nearby in the hierarchical organization, in the sense we are trying to find out, the distances between the two words, and in other terminology also we are trying to find out the similarity between the two words.



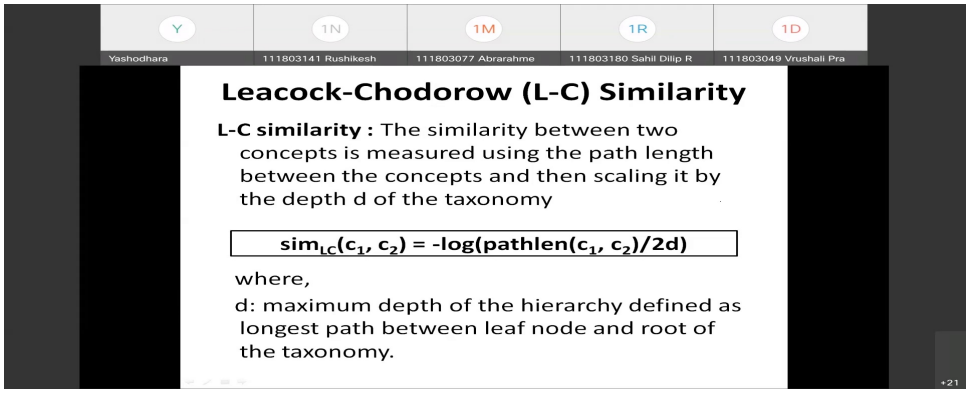
They are nearby. Logically we say that they are similar to each other and if they are far apart, then much dissimilar.So.We use this idea to find out the similarity distance idea, find out the similarity between the two words.No more for finding out the similarity using the knowledge base.Finding out the similarity between 2 words right so is area height or the glasses definitions. Are the two types of approaches you would find in practice being used for finding out the similarity between 2 words.No.Semantic similarity between words and word relatedness.Now what do we understand by semantic similarity or relatedness? We call semantic similarity or relatedness as the degree to which two concepts are related. But they are not similar.So I said similarity measures are limited to this error key OK, whereas if you look into relatedness measures can be applied to all kinds of relations. So when we go, we're talking for finding out the semantic similarity between 2 words. So typically we go for using the right the hypernymy hierarchy to find out the semantic similarity between 2 words. But when we talk about relatedness, it can be applied to all.Find soft relations, not typically hypernym.Some examples of related words is like car gasoline. Path based similarity measure.So the basic idea in part dissimilarity measure is.We are trying to find out the path between 2 words in the hypernymy graph.So we say 2 words are similar if they are nearby in the hypernymy graph, we are just trying to find out the parts between the two worlds in the hypernymy graph. OK, so we usually say that 2 words are similar if they are nearby in the hypernymy graph towards a similar.Similarity is nothing but.Closely proportional to similarity between the two and Sept, C1, and C2, using the path based measure is equal to 1 up on one plus part.Of the two concepts C1 and C2, so similarity is inversely proportional to park leg.So.How do we go for finding out the similarity between words using the pathlength based measure?

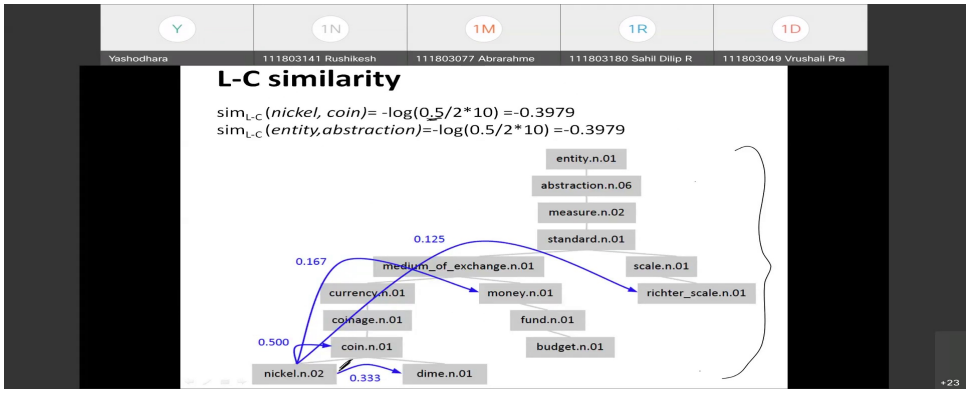


Then what I'm going to do? I'm going to find out the similarity between the sense one of word one with every sense of word two. So I'm going to find out similarity of South one one with two, One South, one one with two two South, one with two three similarly South 1-2 with South 22 S 1-2 with.2112 with S23 and so forth. OK, so we are trying to find out the similarity between every sense of first word with every sense of other word. And then you will be picking that particular combination which is giving a maximum similarity value. This is another way of estimating it, but if you are given this typical hypernymy graph directly, looking at this hypernymy graph currently because because we don't know how many sensors every word is having. So let us look at the hypernymy graph and with respect to the graph, let us go for finding of the shortest.Pot.Between any two concepts which will help us to find out the similarity using pathways measure. So let us look at Nikhil and money.

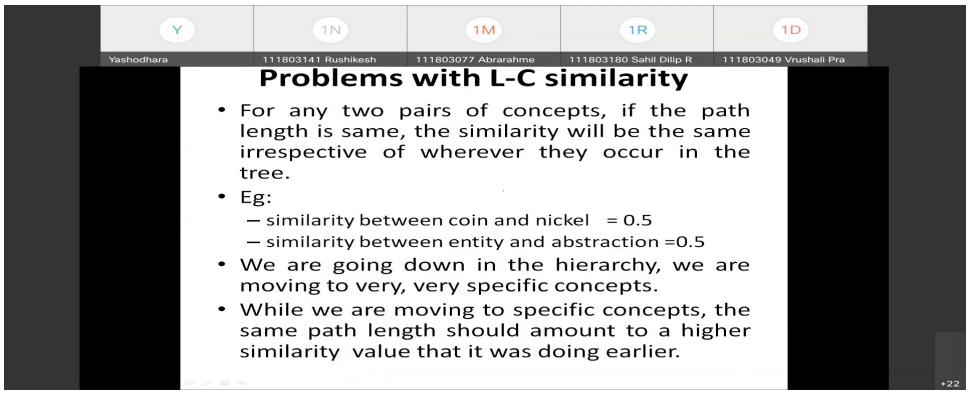


It's .5, right? Similarly, somewhere you go bottom up the air arkie and nickel and coin you have found it to be .5.Ah.So that's the way we are finding it.The similarity between two concepts or two words using the path based measure.No.

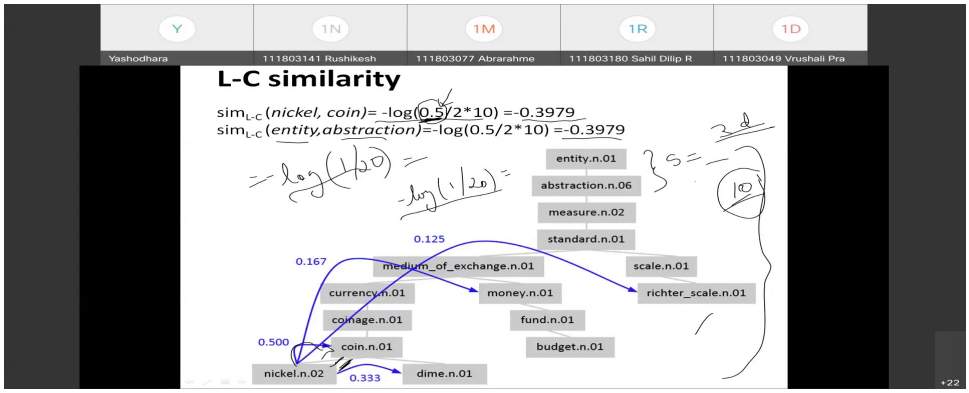




Is 1 but I have taken into consideration the path length we have found out by the Sparc based measure.OK, so like then while I found it as .5 so.I am thinking this is .5.So do not get confused with that also right? So you can take it as part length between the two as one.Similarly, entity abstraction that is entity and that is abstraction.Ogierthe beginning one.So again, it would be minus log of 1 by 20 whatever value you get OK. The similarity found out by LC measure would be the same.For nickel and coin for entity and abstraction, it is minus log of 1 by 20 or nickel and coin minus log 1 by 24 entity extraction.OK.No.As I was talking also when we were talking about part by similarity measure, there are problems.

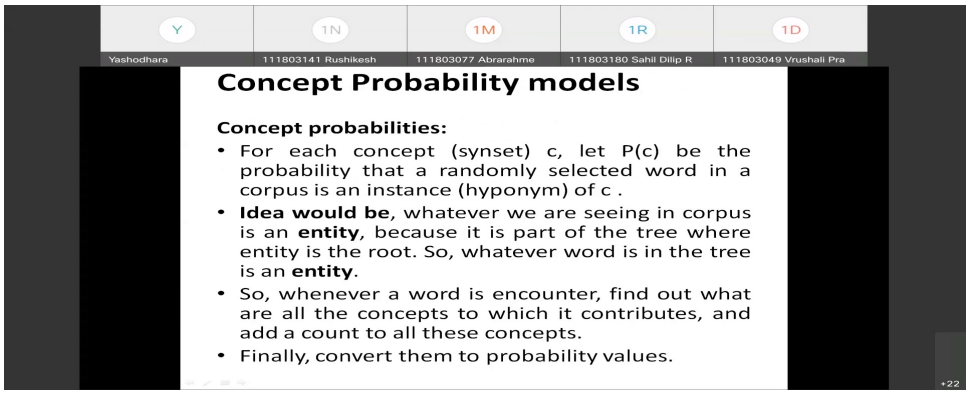


5 similarity between ND and entity and abstraction is .5.This this nickel in coin is with respect to you. I think part based yes, entity and abstraction is with respect to past base and if you look at LC measure or nickel and coin you are getting it as .3979 and also for entity abstraction you're getting .

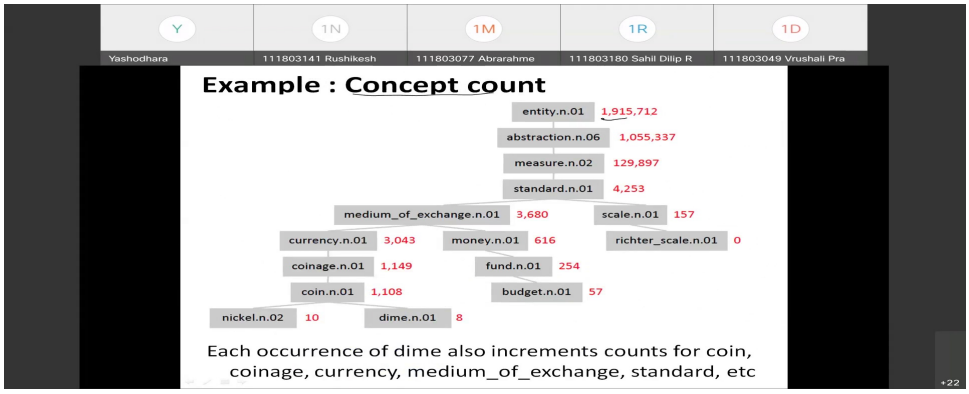


Dog won by 20 but.Whatever value we get it, but both this value would be the same value, OK?So as we are going down into the hierarchy, we are moving to very, very specific concepts while we are moving this specific concepts, the same path length should amount to a higher similar similarity value then it is doing in the.A year labels.But we are more interested in finding out the matrix which assigns different weightages OK two.Or different lens or different weight values to the edges which are at a higher level and to the edges which at a deeper level.So we want a matrix that represents the cost of each edge independently.And the words connected only through abstract nodes. So we want the matrix to give us any information where the nodes at a generic label, the cost of every edge between the two concepts, should be different as compared to cost of the edges are deeper

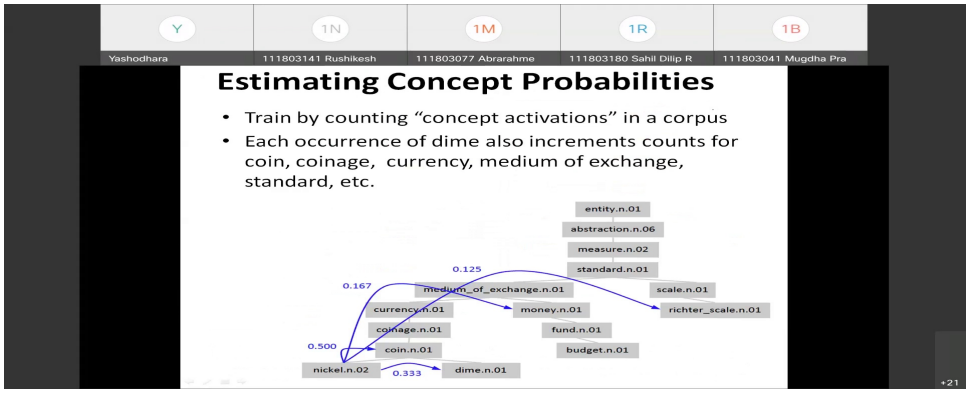
level.



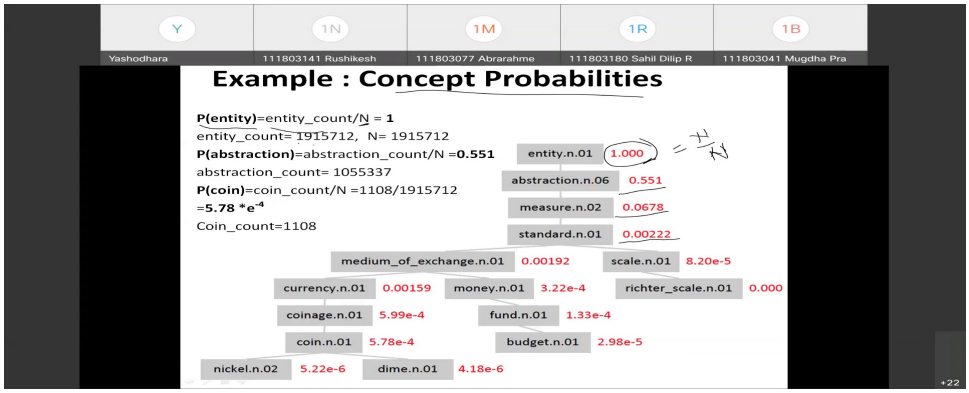
So the idea is we use the concept probability model.So every ward is represented as an entity, so every concept is represented as an entity. So for each concept we go for finding out the viability of the concept.Right?So when we talk about concept probability model, we are treating every concept or every word as an entity and then we go for finding out the probability for that concept.And how do we go for doing it? How do we go for finding find initially the count for every concept which will help to find out the probability for that concept.So the idea is whatever.VC.As I said.Is an entity. At this point of time will have a value of 1 and you will add 1 to entity. Then you have measure. Measure will have a count of 1, but all the concepts on the path from root to your dare count will be incremented by one. So will you will add 1 to abstraction 1 to entity.That you have standard the moment you have standard it will have a count of 1.OK.So by this what will happen?You get a concept. County of this form.At the moment you look at entity, you should immediately understand that these are the.



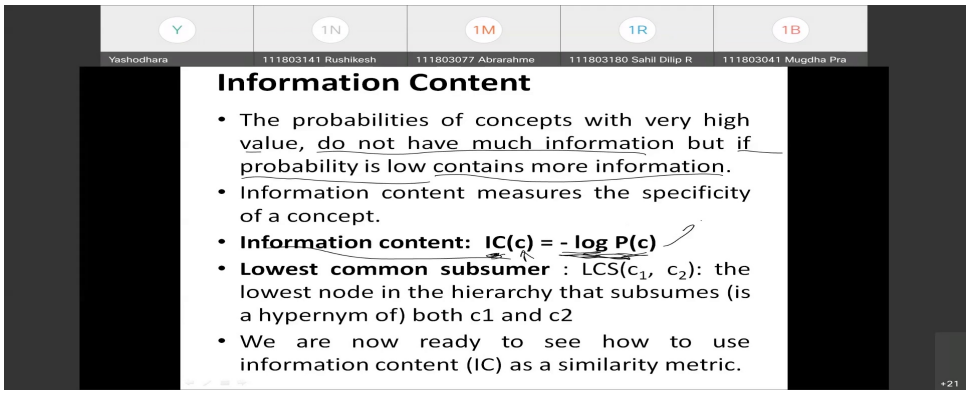
Total number of concepts existing in this.Tree.Goodnight, then immediately we can understand how we can go for finding out the liability.



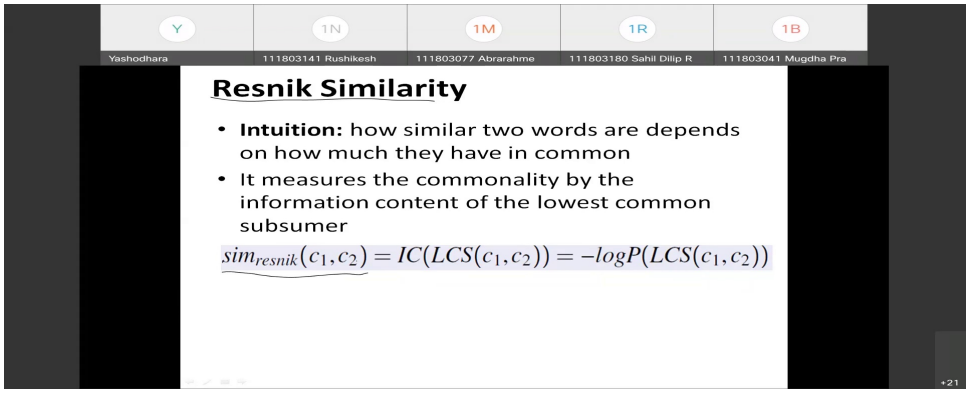
So again, to give you an example, each occurrence of dying.And so increments the count for coinage.So you could find out.The.Probability of entity is nothing but entity count on and where capital N is the concept count of the root node, which is 1915712.



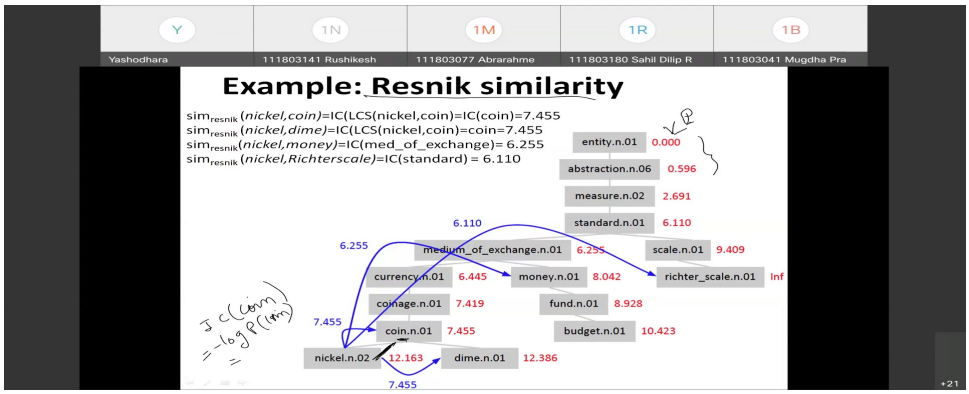
So that's why we have the probability of entity as one probability of abstraction is out of step count of extraction extraction divided by N, which comes out to be 5.51 point 551.And once we have the concept probability, what we are saying that the nodes higher in the hierarchy would be having higher probabilities as compared to the nodes or the concepts or the words lower in the dark.Now how this idea will help us to find out the similarity between 2 concepts?So if.Probability of a concept is very high than it does not contain much information.OK, so if probability of a concept is very high then it is not containing much information and if probability for concept is low then it contains some information or more information.So we say that the probabilities of concepts with very high value do not have much information.But probability of concept is low implies it contains more information.So information contents is a measure which measures the specificity of a concept. So if if.You are trying to measure the specificity using the information content then.More appropriately, for specific nodes and generic nodes, so you can look at.Minus log of probability of a concept.



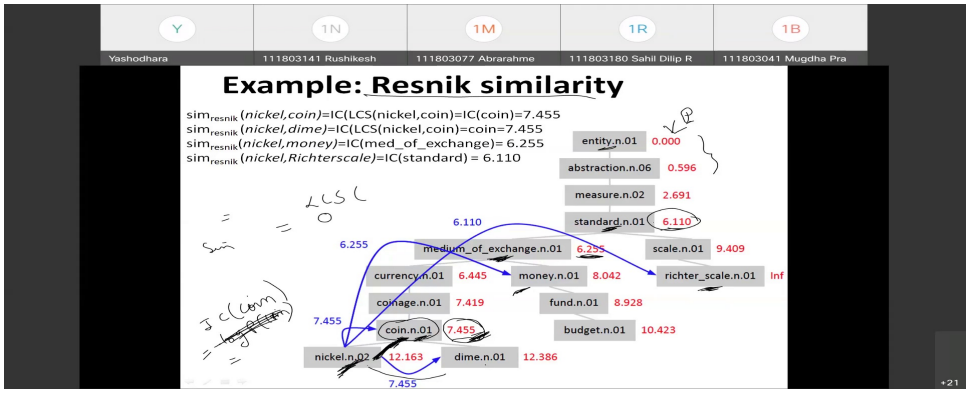
So the concepts which are up in there are key will be having less information content value as compared to concepts which are more deeper into the tree, more specific.OK.So remember this formulation. We have to find out the information of every node.So the nodes which are deeper into the tree are said to contain more information content as compared to the nodes which are higher in the R key.OK.No.What is lowest common sub zoomer?We write it as LCS of two concepts. Always subsume is found with respect to some concepts, at least two.We apply this formulation on every node.This is the information content value we are getting. We are slowly minus log of probability of the concept. So root is having the ability of 1. So obviously by this formulation entity will be having information, content is 0.Extraction by .596 measured as 2.691. Used if you start looking deeper into the tree.Having more information content as compared to the nodes at higher level.OK.So the most informative you know in the hierarchy subsuming the concept C1 and C2 would be called as lowest common subsume. So let us take an example with respect to the CR Key tree. What we can see.What are the sub zoomers of hill and rich?Summers of Hill and Richard the common anxious tears of both of them.It depends on how much they have in common.So the snake similarity measures the commonality by the information content of the lowest common subsume.So what we are saying that in Resnik similarity we are measuring the similarity between the two words by measuring how much common they have.OK, we are measuring the commonality between the two concepts, and the commonality is measured with reference to the information content of the lowest common sub servers.So the Resnik similarity between 2 concepts CC-1 and C2 is found out as information content of the lowest common sub zoomer of seven seat.



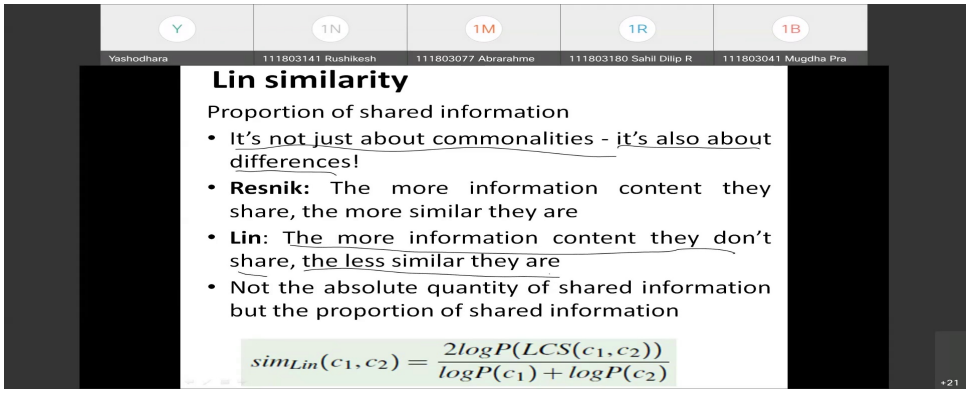
Let's look at nickel and coin and we will look at entity and abstraction.So what would be the?Similarity between nickel and coil using Resnik similarity.What does Reznik telling us it is information content of lowest common Subs Umarov, nickel and coin?Now what is this?This is.By ability value.



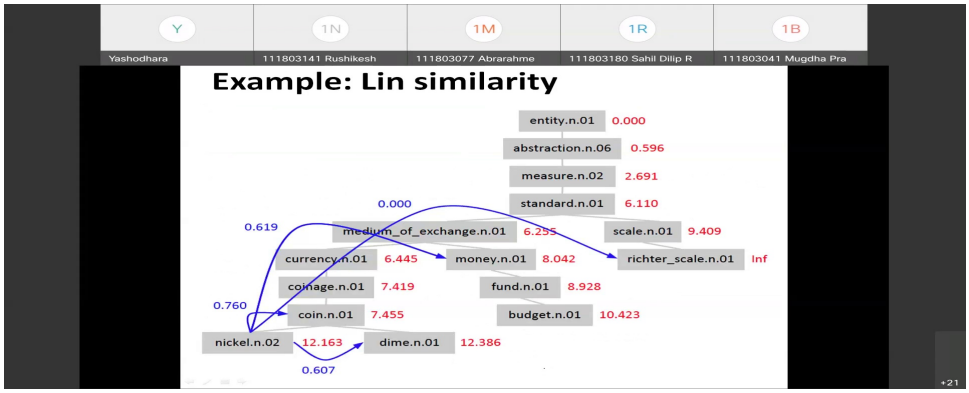
The concepts nickel and coin.If you look at the nickel and dime.Oh yeah, So what is the lowest common subsume of nickel and dime coin right now? What is the similarity between nickel and coin? Using this reasoning similarity, it is also 7.455.Right, if you look at nickel and money, the least common subsume of nickel and money is medium of exchange.



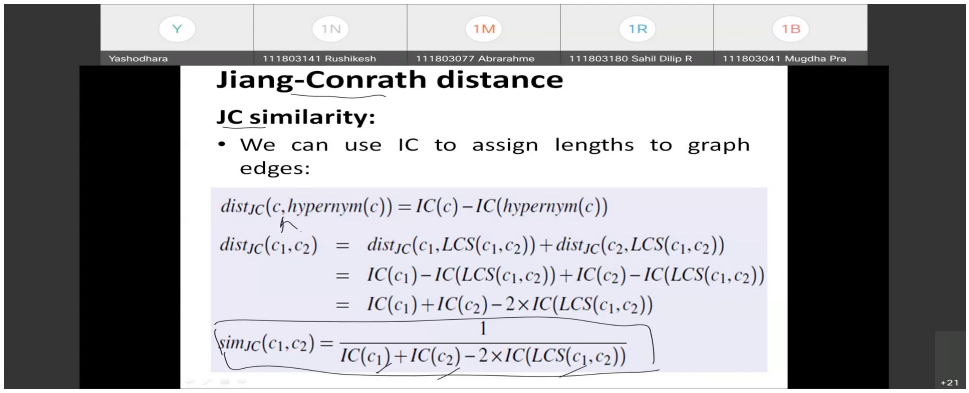
Was the same.Coinage and money.And.What I have taken.Klein agent budget.Or in other words.A year from the from the slide itself, what we can see is.And nickel and dime.They'd send coin and nickel and dime.But you could take or.Nickel and money.And.Nickel and budget both are having 6.Concerts.Like Google and coin.Orother examples for Nick and money, nickel and budget would be the same.So the Elks, the least common sub zoomer for nickel and money, is medium of exchange and budget. Is medium of change. OK, so for both of them, the similarity value would be same. Why we are getting it same because we are occurring how much information they share, how much information they have in common.



So.Similarity between let us say coinage and money would be.The information content common to coinage and money.Normalized by average information, content of coinage and money.So that's why we take it as. What is the information content of coinage and money? It is 6.255.The point of 6.255 upon.Information, content of coinage and money. So, which gives you this is nothing, but we are saying that.Normalized by average information content which gives us .8091.So the similarity between coinage and money bailing similarity. We get it as .8091 for coinage and budget. We get it as .7012 for nickel and money. We get it as .6192 and for nickel and coil we get it as .7600.OK.So putting that similarity on the information content graph, so where we were trying to find out between nickel and coin at this point 670 between nickel and dime, it comes out to be .



607. Similarly, between nickel and money. At this point, 619 and between nickel and Richter scale it is 0.Now even we can go ahead and we can go for using information content to assign lens to graph edges.So I will talk on this because how much time is remaining? Because I'll need at least 5 minutes. Yeah this time, so I'll talk on this. Yeah, I'll finish it on this.So.So we go to finding out JC similarity OK?And Jesus similarity is found out.By finding.No.By using this formulation.JC similarity between the concepts C1 and C2 is nothing but one upon information content of C1 plus information content of C2 minus twice information content of least common subsume of C1C2.And the distance of the concept.See.The graph has found out us.



Information content of C minus in information contact the hypernymy of C.I should say, and C1 and C2.163 as information content of money that is 8.042 minus twice information content of the least common subsume of nickel and money. So least common subsume of nickel and money is medium

of exchange, so it's minus 2 into 6.255 right? It comes out to be .1.That's the way you go for finding out is young cornett similarity. Similarity is nothing but one upon distance measure.Do you go for finding out distance between nickel and coin? It would be information content of nickel plus information content of coin minus twice information content of least common sub zoomer of nickel and coin.