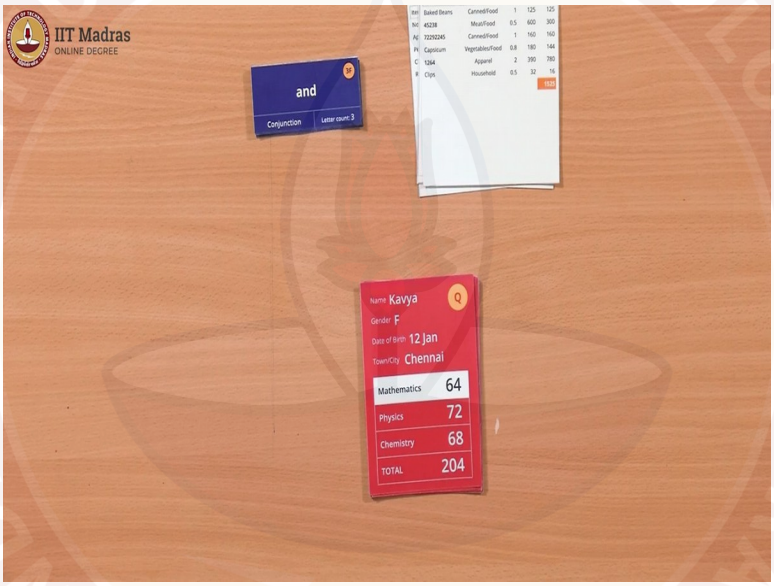
IIT Madras

ONLINE DEGREE

Computational Thinking
Professor Madhavan Mukund
Department of Computer Science
Professor G Venkatesh
Department of Electrical Engineering
Indian Institute of Technology Madras
Sanity of Data

Professor Madhavan Mukund: So we have been looking at these cards and looking at the data on the card but one thing I think we need to be sure about is that the information on the card is actually correct and this we have never been checking.

(Refer Slide Time: 00:28)



Professor G. Venkatesh: So if you take a card typical take this, student data set for example, you could, I mean there are many fields in this, right, so there is the name, there is the gender, there is date of birth, town/city then there are marks. Each of them presumably need to have, needs to have only some kind of data. For example, in the name you cannot put something else other than some characters which are representing a name, is not it?

Professor Madhavan Mukund: Correct, and there are even other things. For instance, we had this card number and look at this card for instance…

Professor G. Venkatesh: Q.

Professor Madhavan Mukund: So the card number is Q…

Professor G. Venkatesh: Supposed to be a…

Professor Madhavan Mukund: And we assumed it is a number between say starting with 0 or 1 and going up to the total number.

So we have to be careful, so in this for example we would expect as you said that the name represents a reasonable name it should not have semicolons or 3 or 5 or something like that. Gender should normally be only one of the categories allowed…

Professor G. Venkatesh: Which is M or F.

Professor Madhavan Mukund: M or F and various things like that but here for instance we see one mistake. Now this mistake of course we can check because we know that Q is not a number, but supposing it said 72…

Professor G. Venkatesh: That could be…

Professor Madhavan Mukund: But there are two cards which are numbered 72 or there is no 72 so we have to be little bit careful.

Professor G. Venkatesh: So there is presumably there is a valid range right, even if it is a number we know the number of, we know the range in which the number has to lie. So if it is like say 7400 then it is very unlikely that a class will have 7400, so you know basically it is out of range, so you can check.

Professor Madhavan Mukund: And for instance even the marks we are assuming that these marks are out of 100 but maybe there is some system where marks are out of 150 then you might see a number like 125 and you might not…

Professor G. Venkatesh: It still be okay.

Professor Madhavan Mukund: It might be okay. So we need to know that it is right type of value that it is number and that is within the legal range of values for that particular field.

Professor G. Venkatesh: For example, it cannot be minus 1, cannot be negative number…

Professor Madhavan Mukund: Minus 1 for sure is not…

Professor G. Venkatesh: Definitely not. Sometimes they put point in that, right so even 64.5 might be okay.

Professor Madhavan Mukund: Correct.

Professor G. Venkatesh: Because teachers tend to give half mark they like it, right.

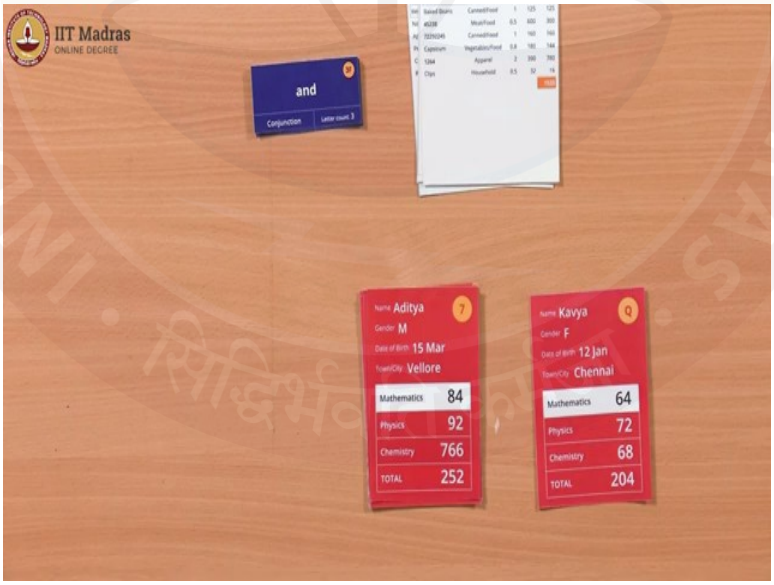
Professor Madhavan Mukund: That is right. That is right.

Professor G. Venkatesh: So two and a half out of 5 may be okay, 3 may not be okay like that, right so they give half marks. So when you add up you might get something and you may not round off, so it may you may get 64.5 that maybe okay. But if it is 64.43942 or something like that then it looks like a wrong number because…

Professor Madhavan Mukund: Absolutely.

Professor G. Venkatesh: Very unlikely that a, somebody will have mathematics marks which is 4 decimal places, it is very unlikely.

(Refer Slide Time: 2:55)



Professor Madhavan Mukund: So let us look at the next card maybe and see you some other things. So here we have a card…

Professor G. Venkatesh: This is okay, I mean this was Q this 7 okay, 7 is in the range so it seems sought out...

Professor Madhavan Mukund: So we have the range which is reasonable…

Professor G. Venkatesh: Aditya seems reasonable, gender is M…

Professor Madhavan Mukund: M, the date of birth seems reasonable…

Professor G. Venkatesh: 15th March.

Professor Madhavan Mukund: No complaints about the town, mathematics is reasonable, physics is reasonable, look at this…

Professor G. Venkatesh: 766. So very unlikely…

Professor Madhavan Mukund: And here in fact we can even tell assuming that the total is correct…

Professor G. Venkatesh: Yeah you can come back, right?

Professor Madhavan Mukund: We can already figure out that this total and this marks do not match, so…

Professor G. Venkatesh: So actually we can re-compute chemistry from, so if you assuming the 84 and 92 are correct from 252 minus 92 minus 84…

Professor Madhavan Mukund: And I think will get 60 so the probably this was something where somebody was typing it in a computer and by mistake type…

Professor G. Venkatesh: Type 6 twice I think

Professor Madhavan Mukund: Yeah or 7 twice I do not know, 76 was…

Professor G. Venkatesh: 76 might work because 628 plus 4 is 12, 178 plus 917 plus 8 is 24, so it looks like...

Professor Madhavan Mukund: It should means 76…

Professor G. Venkatesh: Should have been 76 yeah.

Professor Madhavan Mukund: So this is something again which is very common that is people are typing and they make a mistake and they do not always catch these mistakes because they are kind of going through quickly, right.

Professor G. Venkatesh: So actually this thing this business of checking whether the sum of these 3 is equal to the total, is a way of checking the sanity of the number.

Professor Madhavan Mukund: Correct, we may not know which number is wrong but at least we know that there is something wrong, yeah this number…

Professor G. Venkatesh: But again here because you know for example the 252 looks likely, 766 unlikely, yes, so you can even guess that it is in fact chemistry that is wrong in this case…

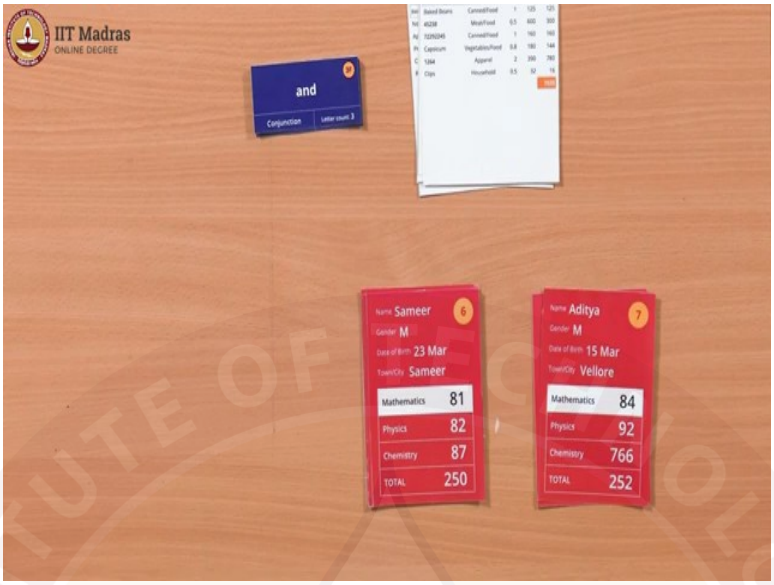
Professor Madhavan Mukund: But of course we would…

Professor G. Venkatesh: Sometimes you may not know…

Professor Madhavan Mukund: Yeah but we will need to have some just like we have some process to check. For example, adding marks and computing average. We would also need to have a process to check from the mistake how to derive what is the mistake because in each case we may not have the luxury of seeing it manually, so whatever we are doing it we might have to check it automatically, so that is another thing we have to think about it.

Professor G. Venkatesh: Let us look at the next card.

(Refer Slide Time: 04:47)



Professor Madhavan Mukund: So here marks look okay…

Professor G. Venkatesh: But look at that…

Professor Madhavan Mukund: So the town and the city…

Professor G. Venkatesh: They put the name in there.

Professor Madhavan Mukund: The mistake…

Professor G. Venkatesh: By mistake, yeah…

Professor Madhavan Mukund: So this could be for instance a copy paste kind of thing…

Professor G. Venkatesh: Could be data entry error you know means some person who is entering the data might have entered it wrongly.

Professor Madhavan Mukund: So very often form one line to the next they were trying to copy Vellore and then by mistake…

Professor G. Venkatesh: Or over wrote it or they wrote the city correctly and then sometime later editing problem. So it got over written by name.

Professor Madhavan Mukund: So here basically we have a case where it is not that the value is wrongly type in terms of mistake in the number but actually that I wrong sort of values there, we should not have a name for the town…

Professor G. Venkatesh: So how do you mean that can be found because you can check whether or not these two fields like this are the same that is one way. But also probably for town/city you may have a catalogue of names of towns…

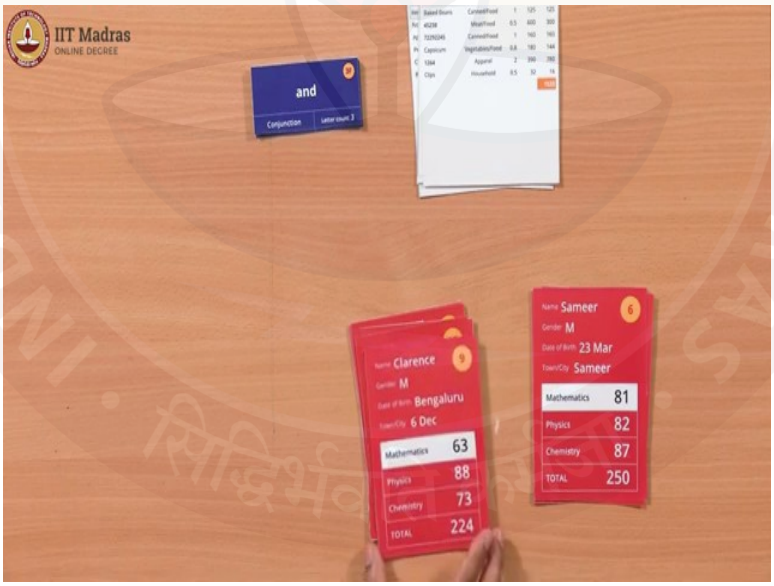
Professor Madhavan Mukund: Correct, you might have a list of all the towns…

Professor G. Venkatesh: List of all the towns…

Professor Madhavan Mukund: Towns and names which are legal…

Professor G. Venkatesh: Which are legal and then if it is not in that then you can flag that there is a problem, yeah.

(Refer Slide Time: 05:50)



Professor Madhavan Mukund: So let us look at this one…

Professor G. Venkatesh: This one seems the town/city, oh, what is going here?

Professor Madhavan Mukund: So here I think they have swapped the two right, so the date of birth says Bangalore the town/city says 6 December, so here it is a simple case of just making the

wrong entry in the wrong place, so the data is all correct but it is wrongly labeled. So again if we trying to do this in through some procedure the computational procedure we would make a mistake because when we pick up the date of birth will get some…

Professor G. Venkatesh: Nonsense.

Professor Madhavan Mukund: Some nonsense sequel thing which we cannot understand, so this is yet another type of mistake you could have, where the data actually correct but it is just in the wrong order, wrong place.

Professor G. Venkatesh: It is again we can find I mean you can because again if you check whether the town/city is valid you find out or even the date you should be able to…

Professor Madhavan Mukund: Exact…

Professor G. Venkatesh: The date is valid, right.

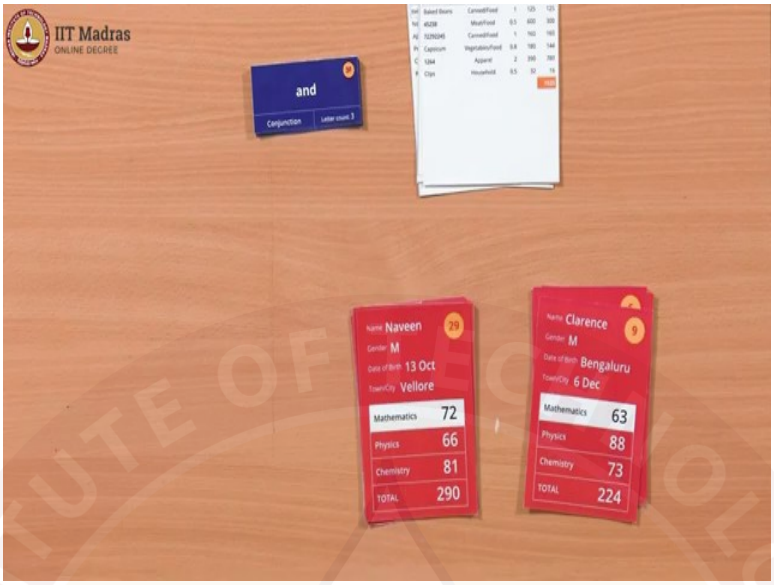
Professor Madhavan Mukund: We can check it has the date…

Professor G. Venkatesh: So date has to be, in this case it has a date and a month…

Professor Madhavan Mukund: And it should have a space for example, which is not there and so on.

Professor G. Venkatesh: Something like that.

(Refer Slide Time: 06:48)



Professor Madhavan Mukund: Let us look at this one, so here we have well the name looks okay, gender looks okay, the card number looks okay, date looks okay, town looks okay, the marks look reasonable they are between 0 and 100, so I wonder what is wrong here?

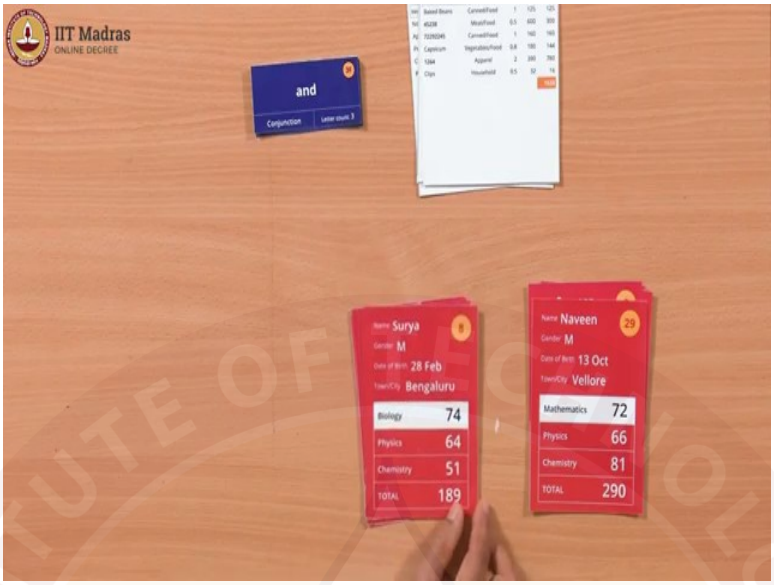
Professor G. Venkatesh: May be the total?

Professor Madhavan Mukund: Look at that, 6 plus 2 plus 1 is 9 and it ends with 0 so obviously the total is wrong, right.

Professor G. Venkatesh: So we do not know whether the total is wrong or one of the either marks…

Professor Madhavan Mukund: So either one of the marks is wrong or the total is wrong but something is wrong here because internally this thing, so there is a constraint that this total must be equal to the sum of these 3, somewhere something is gone wrong. So we have to figure out where that is. So this is yet another type of mistake, everything looks on its own correct but there is some relationship between the values which is not being maintained.

(Refer Slide Time: 07:37)

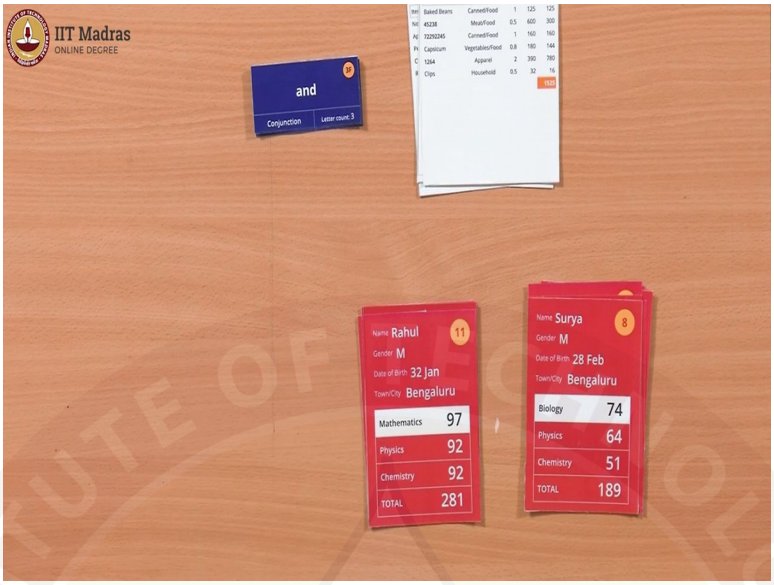


Professor G. Venkatesh: What does about this?

Professor Madhavan Mukund: Looks okay, the total is okay, everything is okay. Name, gender, date of birth, town/city, biology, oh it should be mathematics. So we have the wrong subject here, so this is either marks from a different class, group of students who are from a different stream or somebody has copied the subject wrong I think so, biology…

Professor G. Venkatesh: So presumably you will say which fields are allowed or not allowed, so Maths, Physics and Chemistry are the only fields allowed, so since it is of field which is Biology that is wrong and also is a missing field, it is mathematics, yes. So we can correct it presumably, right.

(Refer Slide Time: 08:19)



Professor G. Venkatesh: What about this one?

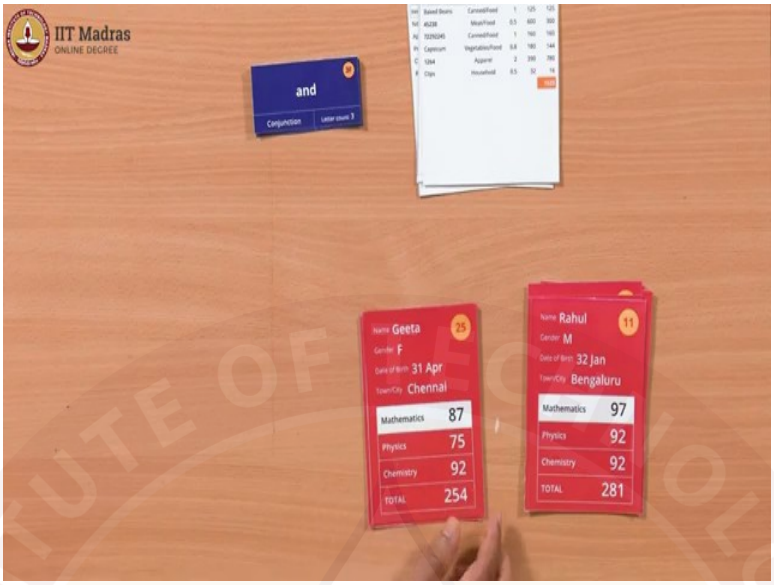
Professor Madhavan Mukund: So this one everything is, says 32 of January, so no month has 32 days. So clearly that is an invalid value…

Professor G. Venkatesh: So for a month it should be 31 or less.

Professor Madhavan Mukund: 31 or less and so 32 is…

Professor G. Venkatesh: 32 is definitely wrong.

(Refer Slide Time: 8:44)



Professor Madhavan Mukund: About this one?

Professor G. Venkatesh: This one is 31…

Professor Madhavan Mukund: So 31 or less but…

Professor G. Venkatesh: But it is April.

Professor Madhavan Mukund: Yes, so not all months are allowed 31 days…

Professor G. Venkatesh: For April it is only 30.

Professor Madhavan Mukund: April has only 30 days so we also have to check within the date…

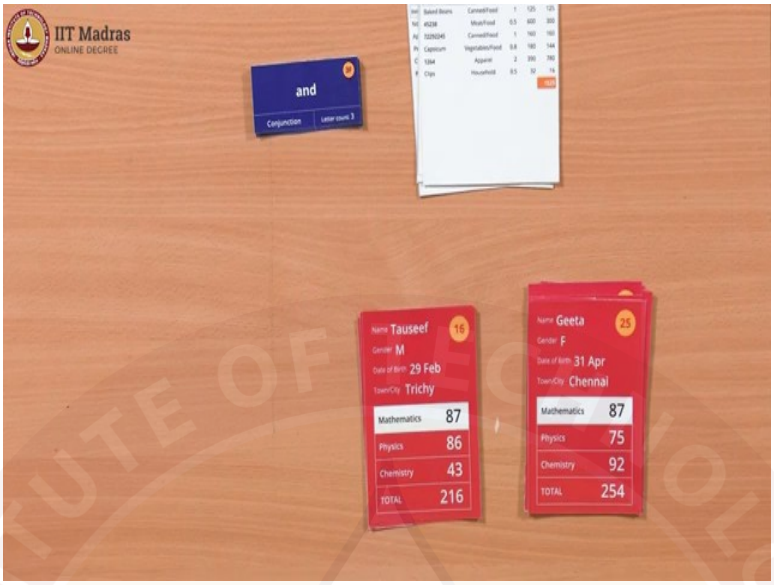
Professor G. Venkatesh: Between the two…

Professor Madhavan Mukund: Whether the month and the date are correct or not.

Professor G. Venkatesh: For month of April it has to be below 30.

Professor Madhavan Mukund: So we have to have a complicated thing for each month what are the legal dates.

(Refer Slide Time: 09:10)



Professor G. Venkatesh: This one is okay, right? What is wrong with this?

Professor Madhavan Mukund: Well, February may have 29 days but only if it is a leap year, so we do not have enough information actually to check whether this is correct or not because we have not…

Professor G. Venkatesh: We do not have the year…

Professor Madhavan Mukund: We do not have the year and unless we know the year of birth we do not know whether 29 is a mistake or not.

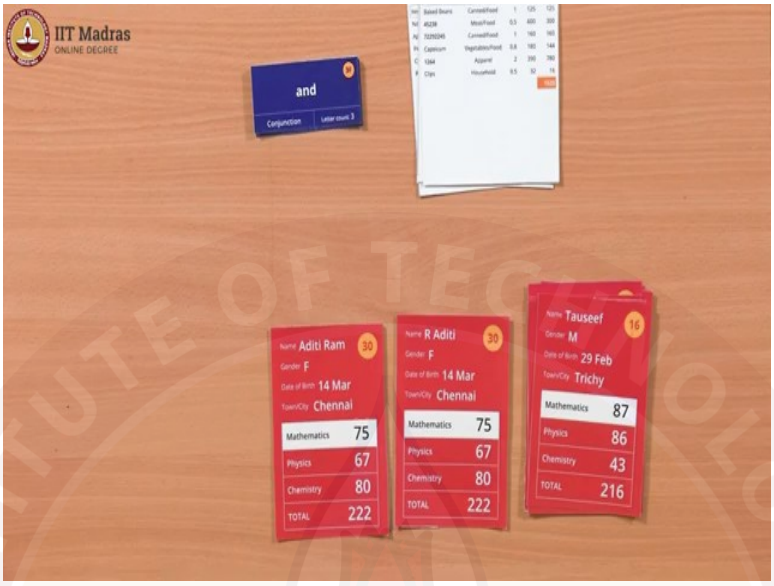
Professor G. Venkatesh: So presumably all the students have this, born in the same year, one assumes because there is no year given…

Professor Madhavan Mukund: But normally there is a range between say September and February, so they might be from one year or the next. So if both are not leap years then we are okay but if one of them is a leap year then…

Professor G. Venkatesh: Since the year is not given here one assumes that one has to actually input or determine the year from the context, right. So you got to figure out that is the batch that all joined in the year, some year, whichever year 2003 let us say and then from that you got to figure out basically that may be…

Professor Madhavan Mukund: Yeah this is not likely to be correct.

(Refer Slide Time: 10:06)



So here Aditi Ram as far as I make out everything is correct but now here is a common problem which we find especially I think in India where people write their names in many different ways, so we have essentially we have identical information…

Professor G. Venkatesh: Aditi Ram is the same as R Aditi.

Professor Madhavan Mukund: But R Aditi is a different way of writing Aditi Ram.

Professor G. Venkatesh: With only the initials.

Professor Madhavan Mukund: Yeah and this may happen and this often happens that we have a different name written in our school records and a different name written on say some other identity card and…

Professor G. Venkatesh: Could have been spelt slightly differently…

Professor Madhavan Mukund: If we go to catch a flight for example the booking in which the flight ticket is made may have a different name from the way it shows up on our identity card and then you have to sometimes explain to their security person that it is the same person. So it is easy when you are talking to a human being maybe to reason with that person but for a computer to understand that Aditi Ram is the same as R Aditi is quite complex.

Professor G. Venkatesh: So you have to find, so basically it could be Ram Aditi or Aditi Ram and then…

Professor Madhavan Mukund: Or it could be that they are coincidently two people who have the same details but the card number is should have been 30 and 31. So that also we do not know…

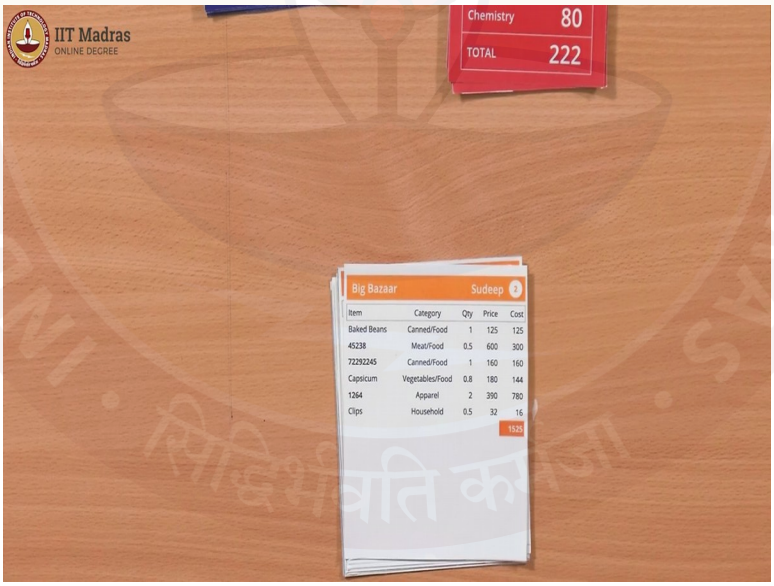
Professor G. Venkatesh: That also we do not know.

Professor Madhavan Mukund: So there are many possibilities but certainly there is this confusion which arises because the same person may be referred to in different ways. And very often when we are adding up combining information from multiple cards this kind of a problem would be important, okay. So that is one…

Professor G. Venkatesh: Data set.

Professor Madhavan Mukund: Data set.

(Refer Slide Time: 11:47)



So let us see, so let us look at same issue with our second data set which is the shopping bills. So remember that we have the item, the category, the quantity, the price which is the unit price and the cost and here we expect that the quantity multiplied by the price should be equal to the cost, so we have this kind of. So earlier in our school marks we only had to have a total, here we have two kinds of arithmetic going on which is one multiplication here and then an addition in the

final column so this last line should be the sum of everything here. So there are some quantities, some items which we have to check whether they are correct or not and then some internal relationship…

Professor G. Venkatesh: Of course there is a name of the shop which has to be valid name of the shop…

Professor Madhavan Mukund: Name of the shop…

Professor G. Venkatesh: So one could basically check in a database again to see whether the shop is a valid name.

Professor Madhavan Mukund: Of course we have no idea about who is customers are but at least we can check that the name is reasonable just like it and then we have to check that this card number is a valid number…

Professor G. Venkatesh: Card number is in the range, yeah, it is the correct range, is a number and so on.

Professor Madhavan Mukund: So this first one I think it is easy to spot that there are some problem with the items, because some of the items now are given as numeric codes rather than as item names, so I think…

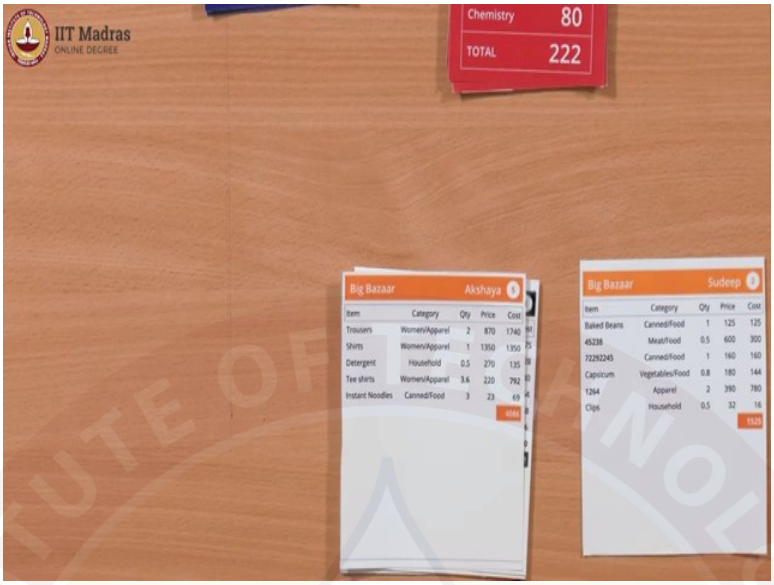
Professor G. Venkatesh: Just meaningful for the shop, may be but definitely not useful for the customer.

Professor Madhavan Mukund: I think somewhere in the shops system they have messed up this thing so it is printing out some kind of a code, item code instead of an item name.

Professor G. Venkatesh: So the shop can correct it…

Professor Madhavan Mukund: But the customer has no idea what 45, I mean of course they have a clue because they know if the category is right what kind of category it was but it is not very meaningful at this point.

(Refer Slide Time: 13:20)



So this is another bill and here for instance we can see that this person Akshaya has bought T-shirts and it claims that the number of T-shirts is 3.6…

Professor G. Venkatesh: Which cannot be right because…

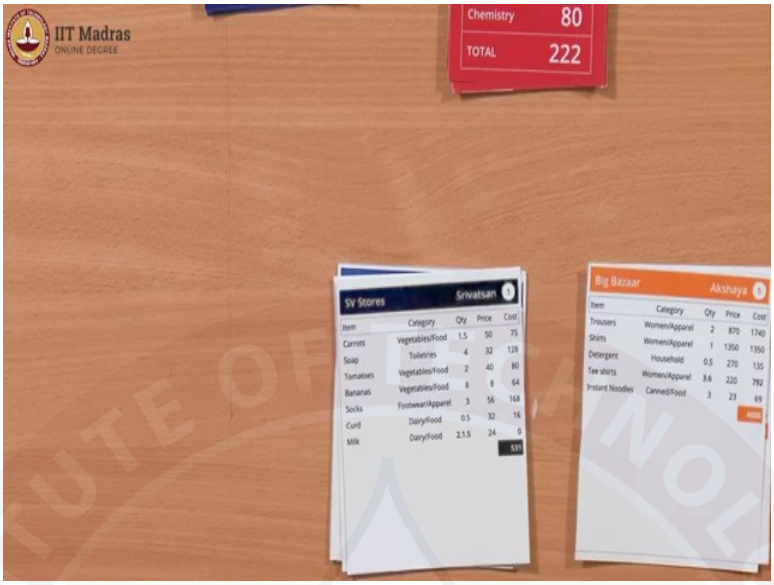
Professor Madhavan Mukund: So in now general you could have a fractional value like 0.5 may be 0.5 kilos of detergent is reasonable but 3.6 is not a reasonable quantity. So it is a bit context dependent because you have to look at what item it is and only then decide.

Professor G. Venkatesh: So an apparel category, clothes basically have to be whole numbers…

Professor Madhavan Mukund: Or anything which comes in unit for instance I would guess that if instant noddle is sold in cans, you cannot buy half a can that would also have to be a full number. Whereas something which is something sold loose or in which is the divisible like food items sugar or salt or something like that could presumably sold in smaller units. So here the problem is that we have some indivisible thing which is given as a fraction.

Professor G. Venkatesh: Even the price I guess once you know that the category like a T-shirt for example, you have the price says 220 it seems okay I think for T-shirts but if the price is say 1 or something like that, yeah seems unlikely your if it is 10000 also it is very unlikely, right. So there is a kind of a normal value for an item in the category.

(Refer Slide Time: 14:43)



Professor G. Venkatesh: What is wrong with this?

Professor Madhavan Mukund: So I think here we just have a nonsensical value…

Professor G. Venkatesh: It is a syntax problem…

Professor Madhavan Mukund: So we have something with two decimal points…

Professor G. Venkatesh: Which cannot be right…

Professor Madhavan Mukund: So and then you can also see now that there is problem with the operation because we would like so may be the last column is actually automatically computed in the system and so 0.5 into 32 is 16 but now when it encounters something like 2.1.5 into 24 there is no way you can take this 2.1.5 and think of it as a number. So this operation that we want to multiply these two does not work, so then you also end up as a result with cost which is 0 which cannot be I mean there is nothing that I mean unless the shop is going out of business and giving away its items there is not going to be…

Professor G. Venkatesh: In here because you have a total and you have all the other numbers you could presumably sum up all these, subtracting from 531 and get this so you can do…

Professor Madhavan Mukund: Assuming the regular cost for that but it could also be the 531 is taking…

Professor G. Venkatesh: Is also wrong…

Professor Madhavan Mukund: Yeah, is also wrong. But definitely you should not see a 0 on the last column and here it is because this error basically makes this operation invalid. So there is some legal value and there are also…

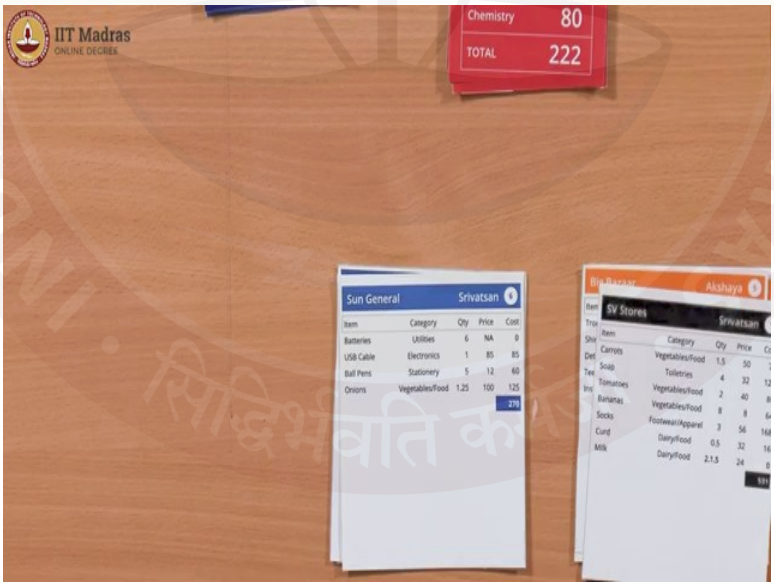
Professor G. Venkatesh: Typically, I guess in computer systems when you try to enter 2.1.5 it would not allow you to enter at even, right, should not.

Professor Madhavan Mukund: Should not allow you to enter it so maybe there should be the system should have been designed, the billing system should have been designed so you cannot do that.

Professor G. Venkatesh: So here we do not know that is 2.5 or it is 1.5 or 2.1…

Professor Madhavan Mukund: Because this is milk, unlikely to be 2.1 I think it is more likely be 1.5 because I do not think people buy 0.1 of anything but…

(Refer Slide Time: 16:24)



Professor G. Venkatesh: What is wrong with this?

Professor Madhavan Mukund: Look at that.

Professor G. Venkatesh: Oh NA.

Professor Madhavan Mukund: So this is another situation which is different from that here you are not able to compute the cost because you do not have the unit price, so some values actually missing and instead of telling us it is missing they I mean by instead of putting some random number they have written NA which usually means for not available I guess. But again our computational process should be able to identity that something is wrong and they say not to process this because then everything goes wrong because 270 is not a correct value…

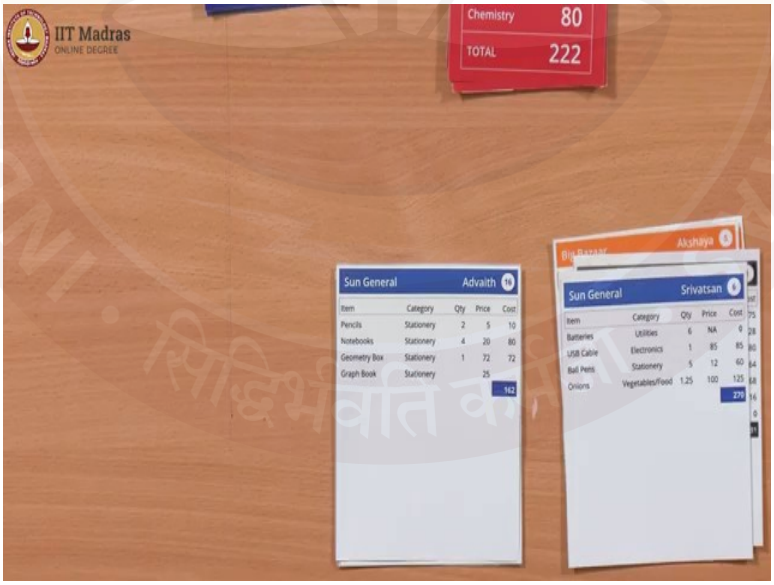
Professor G. Venkatesh: Or you basically you have some way of finding, you have a database where general prize of battery is available and you pick it up from there.

Professor Madhavan Mukund: Correct you could have some other source of, other source of informative…

Professor G. Venkatesh: Normative value of...

Professor Madhavan Mukund: Yeah so to fill-in that missing information, but first you have to recognize the information is missing.

(Refer Slide Time: 17:29)



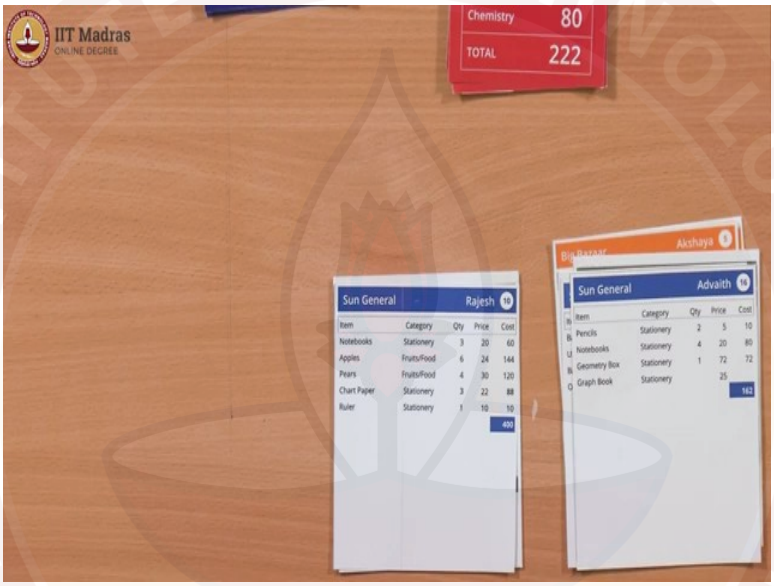
So here in the next one for instance the information is really missing and here there is no good way to fill-it in because here what is missing is not the price but the quantity. So somebody has bought graph books and this shop has forgotten to enter how many graph books were bought and here somebody might buy 1, somebody might buy 10, of course you may assume that on an

average people buy only 1. But so these are two different situations is missing information one where an explicit not available has been put and one where the value is has just been left blank and in both cases, so here the blank is propagated may be as a cost is also blank, whereas there the value has been calculated as…

Professor G. Venkatesh: It is also possible that this item was neither bought it, it got entered…

Professor Madhavan Mukund: Yeah, could have just been an incorrect entry and not deleted.

(Refer Slide Time: 18:15)



So in this particular bill the problem is little more settle I do not think that we have any illegal entries like missing entries or 2.1.5 or something but I think if you look at some lines for instance chart paper you find that…

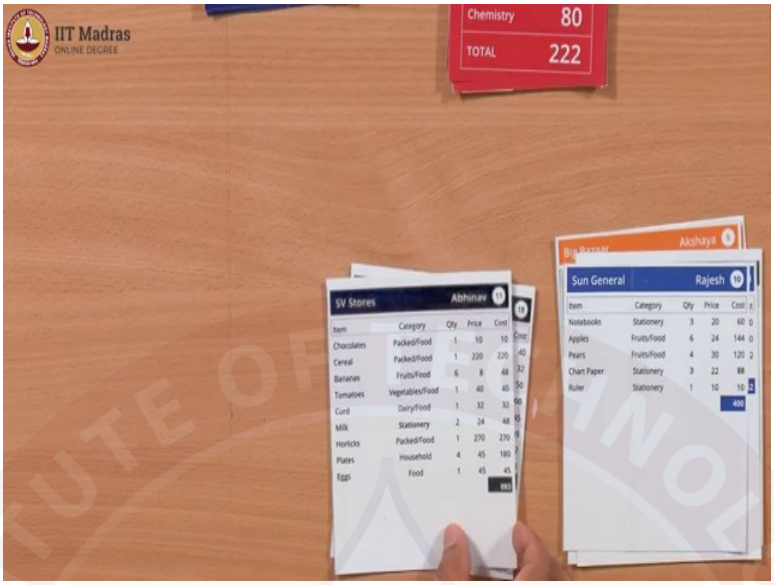
Professor G. Venkatesh: Oh I see…

Professor Madhavan Mukund: Total is wrong, so 3 times 22…

Professor G. Venkatesh: 4 times 22 is 88, 3 times when you do should be 66.

Professor Madhavan Mukund: So either this person has got 4 things and is been shown as buying 3 or they have bought 3 things and the bill has been wrongly calculated. So in any case there is an internal inconsistency so we know that at that line there is some mistake and we have to go back and fix that.

(Refer Slide Time: 19:06)



What about this one? So here I think the numbers look reasonable but the category is wrong…

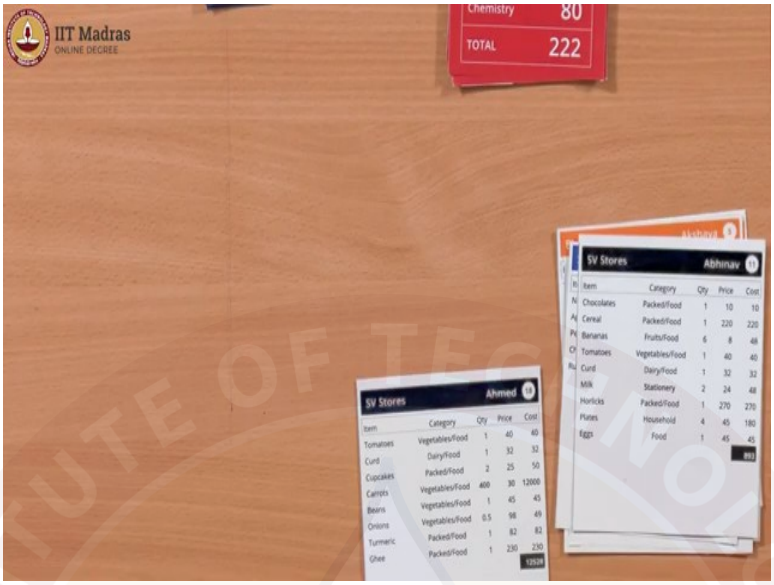
Professor G. Venkatesh: Milk cannot be stationary

Professor Madhavan Mukund: Milk has been labeled as stationary, so we again expect that the item and the category should be connected…

Professor G. Venkatesh: See there is a way to check whether or not that this consistency between this and this, right in some sense you can check because you could possibly take each item and some database could be there some place where you the category could be…

Professor Madhavan Mukund: So we could have an external way of looking up some table where you have these values there, the correct combinations, the ones which are allowed or present and we can check those.

(Refer Slide Time: 19:52)



So here there is a very strange, so the bill is very large we saw earlier that this SV stores is mainly a grocery store, their bills are typically a few hundred rupees and suddenly you see a 12000 rupees bill, which should already tell you something is unusual. And they have, we have an item of 12000 rupees for carrots now, it is very unlikely anybody has bought 12000 worth of carrots. And then you see that the quantity is 400 and the price is 30, now so it is reasonable to assume the 30 is say the price per kilo…

Professor G. Venkatesh: So 400 might be…

Professor Madhavan Mukund: So 400 is probably the wrong unit, so probably…

Professor G. Venkatesh: It should have been 0.4

Professor Madhavan Mukund: Yeah the person the shop assistant has probably typed in 400 instead of 0.4 and therefore the whole bill has become inflated. So here is a problem that the quantity is in a different unit and the price is in the different unit. So therefore the whole thing does not work, so you have to be careful so this is the very context sensitive thing so you need to have to understand that this price should not be like we had said earlier about say T-shirt should have an expected price.

So here by looking at the price you know that something is unusual given also the other quantities are very reasonably undergone 100, so to have 1 item which is 12000 rupees is very

unlikely, it is not like a wholesale buying for a restaurant or something like that. So this is a unit mismatch. So there are many different types of mistakes as we can see that you can make and some of them can be caught by looking at the operations that are performed some of them can be caught by looking at the expected values, some of them are just syntax mistakes which we can check.

Professor G. Venkatesh: Here we have not seen any missing values but presumably you know the category could be missing for example…

Professor Madhavan Mukund: Category could be missing, yeah so we saw some missing values in the quantity and price but many things could be missing.

Professor G. Venkatesh: So category could be missing, item name could be missing.

Professor Madhavan Mukund: Yes, it could be wrong also.

Professor G. Venkatesh: Could be wrong and in terms of operations we can see basically that we do not expect that we will do any operations in the name, these are kind of constants on which we… we may check basically whether given name whether, where is, how many tomatoes are there may be there are two lines with tomatoes on it…

Professor Madhavan Mukund: Correct, they should not add up.

Professor G. Venkatesh: So you can actually look here just compare whether this is tomatoes or whatever it is. Then similarly category you can check because it is has got two items in it vegetables slash food, you could check whether it is food item or you could check whether it is a under food either it is a vegetable, right, you could check like that. So comparison with just one of the sub-categories or with the major category which is food that could be one operation that may do.

Professor Madhavan Mukund: So this will be presumably an external table which we have to check whether it is correct or not.

Professor G. Venkatesh: And in terms quantity, these quantities we are expecting that they will be reasonable there will be small in number if it is and so these are kilograms, so it is very large then we can catch it I guess. And we should be able to add these quantities presumably these

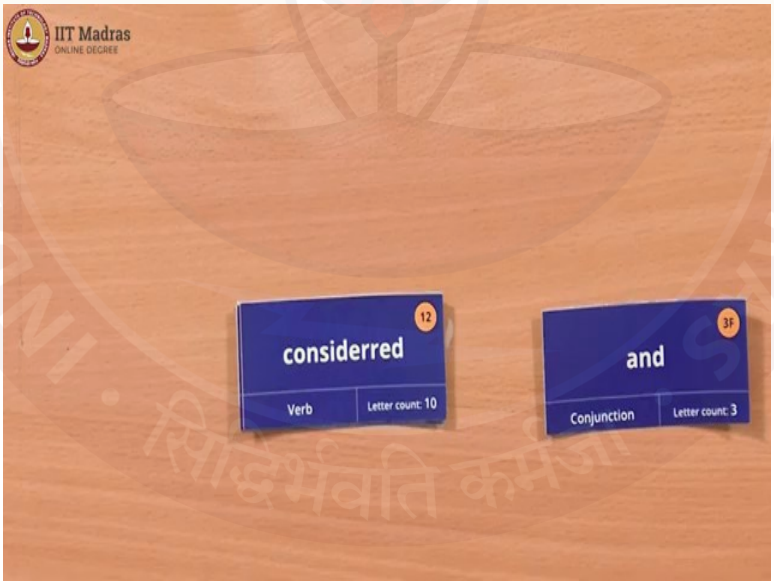
quantities can be added. Meaningfully it can be added in, for example you want to count how many articles were purchased, sometimes we do that right. We buy to check…

Professor Madhavan Mukund: So like for instance sometimes they have at the exit of a shop, they have some security checking the bill, that person is usually checking number of items in shopping cart and they just want to make sure that nothing is been unaccounted…

Professor G. Venkatesh: So you would guess I guess you would count the 0.5 as 1, because it will be in one packet and then you will count, so some operation basically you are doing. And similarly you can check whether this total is correct by adding up all these numbers, so the numbers can be added up, so they are all addable right, in some sense they can be total, then we can check whether this is same as a total.

Professor Madhavan Mukund: Okay, so that may be one set of, so similar set of data things for the shopping bills.

(Refer Slide Time: 23:58)



And the last data set which we have been looking at is the words in the paragraph. So we could have similar mistakes and here we have different, remember now we have the word itself then its position in the paragraph, its unique number, type of part of speech whether it is a noun, verb or something and how many letters it has. So here we have like we saw for the first marks thing that there is a card number 3F which is clearly not a valid number, so…

Professor G. Venkatesh: Because it is we expected that the number basically represents the number of the word in the paragraph, the position of the paragraph…

Professor Madhavan Mukund: And it should be a real…

Professor G. Venkatesh: Should be whole number,

Professor Madhavan Mukund: Whole number, yeah

Professor G. Venkatesh: So 3F is not okay.

Professor Madhavan Mukund: So here it is question of understanding the word so here the word is spelt wrong.

Professor G. Venkatesh: So we presumably have a dictionary and we can check, spell check…

Professor Madhavan Mukund: And the other thing that we get when we spell wrong is actually this letter count is actually for the original spelling, so there will be a mismatch between …

Professor G. Venkatesh: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

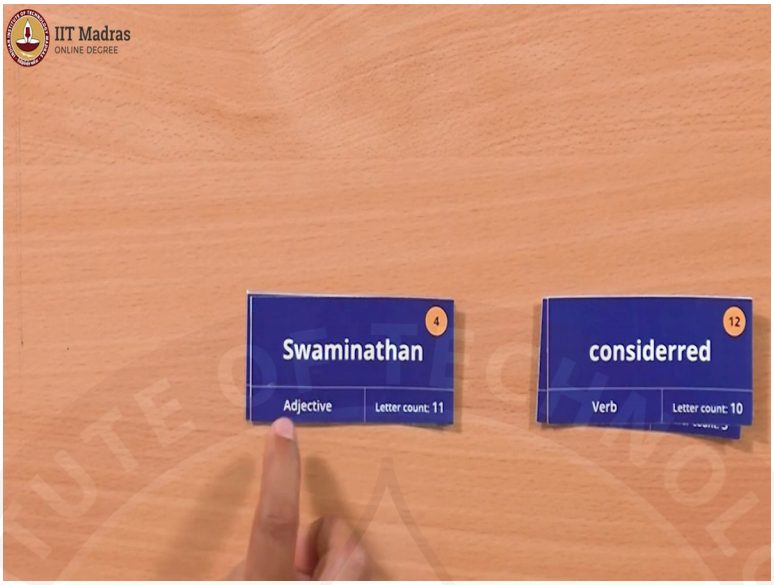
Professor Madhavan Mukund: So there is 11 letters and it says 10…

Professor G. Venkatesh: And it says 10.

Professor Madhavan Mukund: So one of the two is wrong. Either the word spelt wrong or the letter count is wrong. So, so this is a…

Professor G. Venkatesh: This case it is clearly the word that this wrong but if we do not know the word then maybe we can go back and check dictionary or something.

(Refer Slide Time: 25:19)

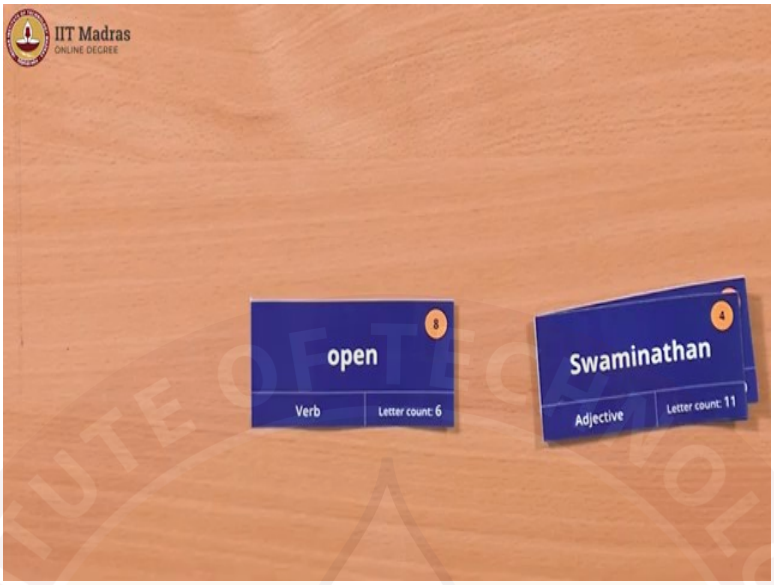


Professor Madhavan Mukund: So here…

Professor G. Venkatesh: Name seems okay.

Professor Madhavan Mukund: The name is okay, the letter count is I think correct, it is 5 plus 5 plus 1 but it is the labeling of the part of speech, so Swaminathan is clearly the name of a person so it should be proper noun, so it should be called noun and not an adjective, right. So in this particular case it is the type so this is again require some understanding of the language to know whether this is…

(Refer Slide Time: 25:50)



Professor G. Venkatesh: Presumably if it is starting with the capital letter it is proper noun…

Professor Madhavan Mukund: Yeah I think should be a proper noun, but look at this one for instance. The word is open now it has 4 letters, the letter count is 6 so that is a clear mistake but it is also labeled as a verb…

Professor G. Venkatesh: Is it correct?

Professor Madhavan Mukund: Now if you say the door is open then it is not a very actually because it is describing the state of the door…

Professor G. Venkatesh: On the other hand, if we say open the door…

Professor Madhavan Mukund: When we say open the door then it is…

Professor G. Venkatesh: It is a verb.

Professor Madhavan Mukund: Exactly. So here whether it is a verb or not we have to actually go back to the text…

Professor G. Venkatesh: Go to the full paragraph and see…

Professor Madhavan Mukund: And see how the word occurs in the text, in what context we see the word and so this part of speech actually is not a is ambiguous in that sense that mean there

could be the same word could have different interpretations in terms of verbs and adjectives and so on. So that is something that we again we cannot tell this, so we can tell that the letter count is wrong by just looking at this word but we cannot tell that the verb is wrong without going back and reading the paragraph.

Professor G. Venkatesh: So here we have clearly there is some operation we are doing when we do letter count, we are counting the number of letters in the word, so and then so there is some operation that we can, we just count the number of letters, right. And then this business of checking whether it is verb or a noun is by looking in a dictionary, we saw the spelling error also was looking in some kind dictionary.

Professor Madhavan Mukund: Yeah but in the dictionary the problem is that it could, dictionary might offer you two choices and you do not know whether you have picked up the right choice. So just looking in a dictionary…

Professor G. Venkatesh: Typically, a dictionary would also say, good dictionaries would also give sentence construct right, it will give a …

Professor Madhavan Mukund: But still, but I think by just looking at the card alone…

Professor G. Venkatesh: Alone you cannot do it…

Professor Madhavan Mukund: You have to go back…

Professor G. Venkatesh: To the sentence…

Professor Madhavan Mukund: Sentence and verify whether the correct choice is being made, so that is more complicated type of check, context is outside the card itself, you have to go to a…