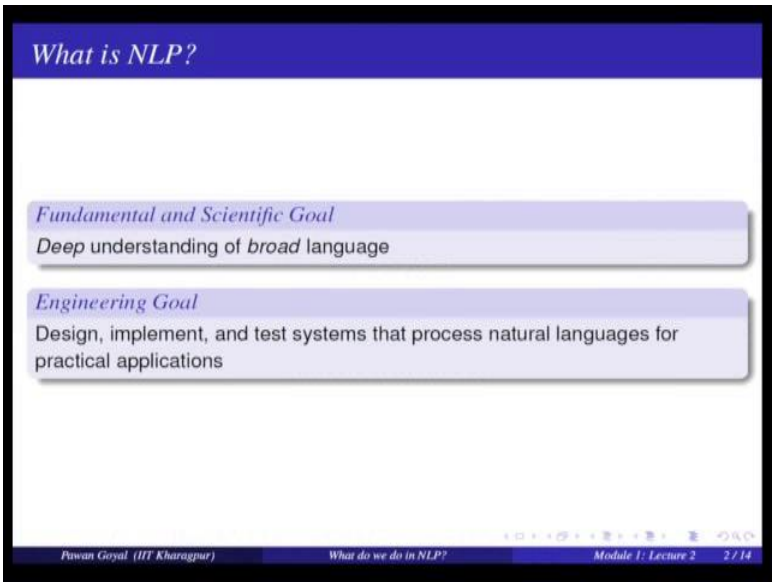
Natural Language Processing
Prof. Pawan Goyal
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 02
What Do We Do in NLP?

So, welcome back to the second lecture of the first week. So, in the last lecture was about the course introduction and this lecture we will start seeing what do we actually do in NLP. So, we ended the last lecture discussing what are the main goals of NLP; we talked about two different course.

(Refer Slide Time: 00:42)



So, we talked about the very fundamentals scientific goal that is can we have a very very deep understanding of the broad language. So and the second goal that we discussed was engineering goal, that is can be design, implement, and test systems that can process natural language and that can be used for practical application for our day to day life and we also said that in this course, we will mainly focus on the engineering course of NLP.

(Refer Slide Time: 01:20)

So, we talked about the engineering goals; now what are these goals? So, let us see with some examples. Now my goals can be very very ambitious, this is just a fun example. So, all though, we use Google translate every now and then, so getting a very very good quality a perfect translation is I will say very very ambitious goal. So, this is snap short taken from Google translate page, if I type Google is awesome, the translation comes out to be Google [FL] if I go from English to Hindi. So, instead of saying something very very perfect, it turns out and says Google [FL] that is not a good translation for this term.

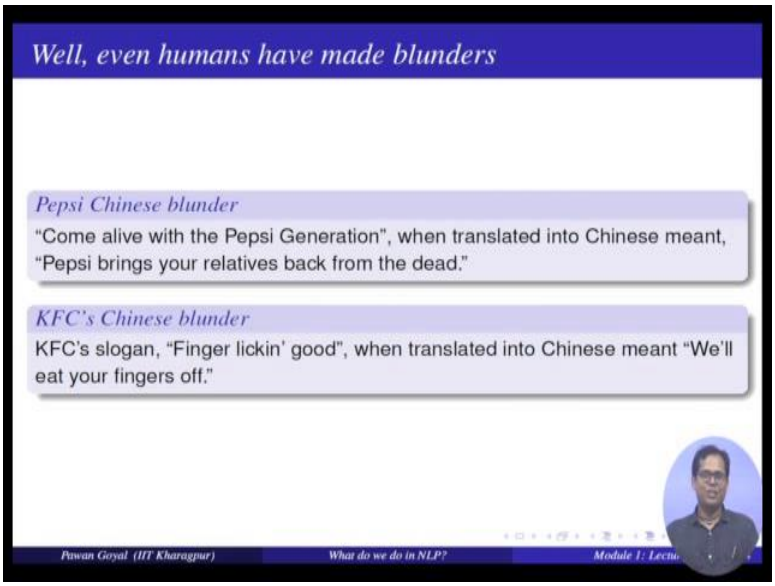
So, why is that? Because the systems that we have are not perfect, they have certain engineering solutions for towards the design, but that be not be perfect. So, you will not get the perfect translation every time and that is what we should always know in back our mind that yes they are not the perfect systems.

(Refer Slide Time: 02:24)

Similarly, this is another example. So, if you type in Google translate Google is cool, we will find some in the translation Google [FL]. So, now, this is slightly better than the last one but it is still not perfect. Why I am saying that this is slightly better than the last one. So, is [FL] one of the translation for the world cool; yes for a person who is cool you can say that cool can have an meaning of [FL], but that is not the meaning that is intendent in this sentence Google is cool.

So, this is one of the problems that we will also see in this course that a word might have multiple senses, how do I know that what is the actual sense that is being used in the sentence. So, this is in actual engineering problem that has to be solved by designing efficient algorithms.

(Refer Slide Time: 03:17)



So, while I was talking about some of the transition that do not go very very well valid in Google, I am just showing is some examples that yes, this is not only the case with machines, even humans have made blunders when it comes to translation. So, in this slide I am showing you one particular slogan that was with Pepsi. So, when Pepsi went to China for the first time for their campaign.

So, they had the slogan on 'Come alive with the Pepsi Generation' and in china they had to translated in to Chinese and they ended up translating it as something that meant Pepsi brings your relatives back from the dead. So, now, this looks very funny, but if you look at the actual English sentence, you see that you can actually come up with a translation like that; you see generation is one to one with the relatives and alive with one to one with back from the dead. So, this is some sort of jugglery of these words, so that you again come up with the very very absurd sort of translation. So, this is not only with the machines even humans have made blunders. So, this was previously was with KFC Pepsi now you can see one with KFC. So, again when they went to china, they had this slogan on finger licking good and when they translated in to Chinese it meant we will eat your fingers off.

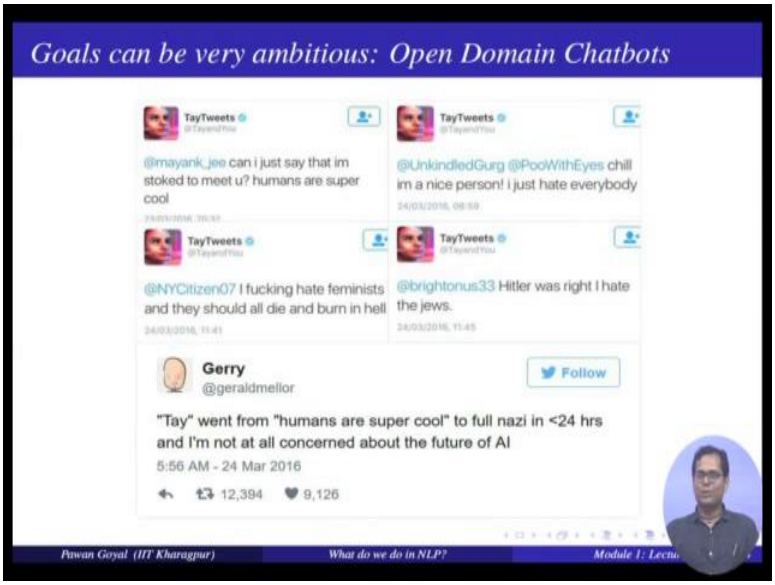
So, again you can see the licking and all this they have a correspondence with the other translation. So, unless you know the other language, you will not be able to translate it

perfectly with just by using it simple dictionary, it might give you a very very absurd translation.

(Refer Slide Time: 05:06)

So, you have out. So, yes in there are many many examples; for example, this is called as hand grenade and, so work in progress translated as execution progress and if you know what is the meaning of execution that I am intending here.

(Refer Slide Time: 05:26)

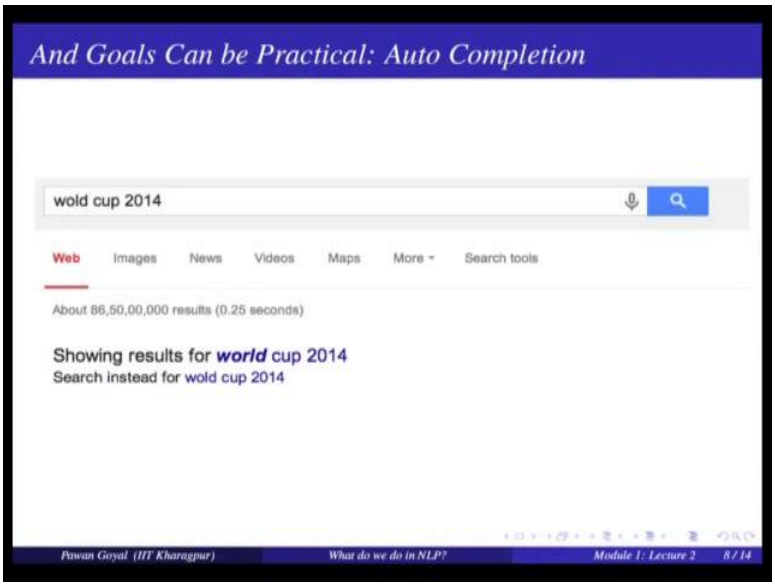


So, yes, again coming to the ambitious goals; if you have heard about the chatbot that Microsoft had released the Tay Tweets, so it was taken on in less than a day why did that

happen? So, it was responding to how you humans were communicating with it and very soon it happen that it was giving very very absurd and resist tweets and it had to be taken down. So, this is very nice tweet. So, Tay went from humans are super cool, to full nazi in less than 24 hours. So, I am not at all concerned about the future of AI. So, again that tells you yes. So, it is very very difficult to develop a very very perfect system that works for open domain conversation; it is very difficult problem to solve.

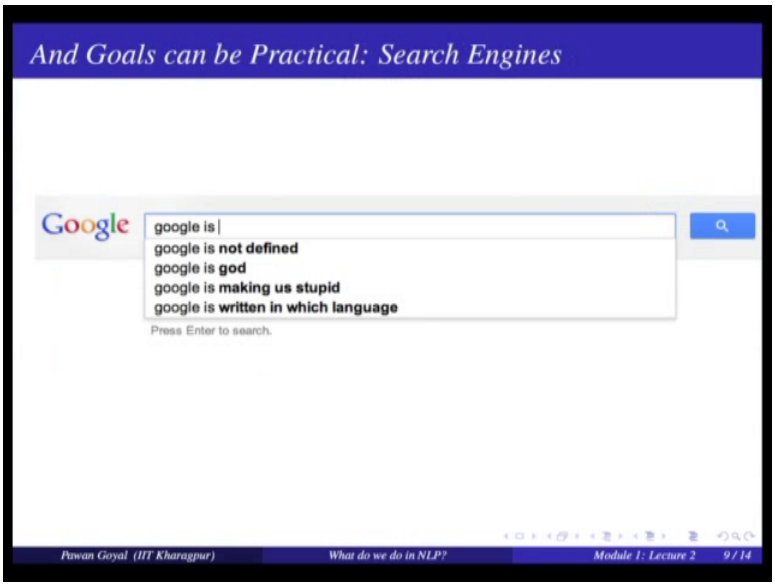
So, we have some goals that are very very ambitious, but they are very other goals many other goals that are also very practical for example, finding out if this, my query is incorrect I am trying to correct it. Suppose in Google you type a query, so world cup 2014 and you missed out on in r. So, Google will give you some sort of reply that Ok, are you looking for world cup 2014. So, instead of w o l d did you mean w o r l d?

(Refer Slide Time: 06:50)



So, this kind of automatic query correction which problem that you can think of solving by using NLP, so in very very systematic manner.

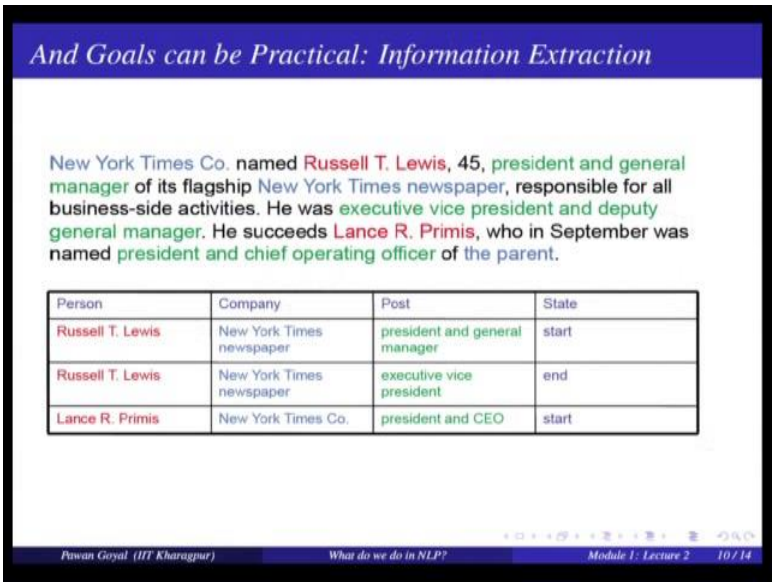
(Refer Slide Time: 07:19)



Search engines and query completion; again in search engines, if you type somewhat like if you type a start typing a query, they also try to predict what is the complete query that that you are planning. So, if you type a query Google is they will try to complete it with what you have queried before or what other users have queried with these types.

So, again the language modelling concept that we will discuss in this course, goes behind all these completion tasks.

(Refer Slide Time: 07:52)



Then there is the very very important application on information extraction; that is you have a lot of unstructured data in the sense of news report and whatever is. So, from where you want to identify what are the entities of interest and what are the various relations between these entities. So, for example, if you look at this text, from here you can identify that Russell is a person who works on the post of president and general manager in the company New York times newspaper and he just started his post at this movement. So, this information is available in the text data, but in a unstructured format.

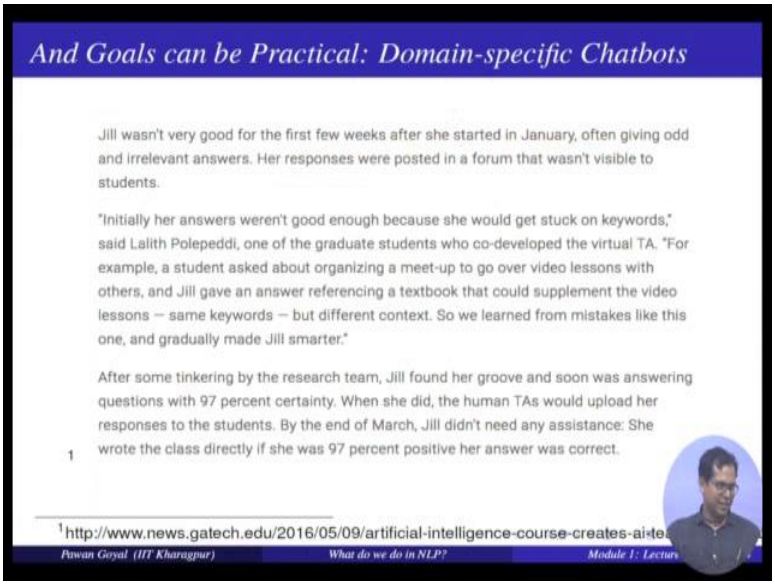
So, now can I have an application or a system that converted information to a very very structured format? So, like here you have seeing it in a data set format. So, you are finding out various persons, to what company, what post they are on and did they tenure start or end. So, given lot of text data can you automatically build up such instruction data sets of information relations between entities? So, this is the task of information extraction, and this is a very very practical goal of NLP. So, in this course we will also deal with this problem for certain lectures that how do I start extracting relations between entities from text data, what are the different algorithms that go behind it.

(Refer Slide Time: 09:22)

Then if you have heard about this course; so this is recent news, in one of the course in (Refer Time: 09:32) professor used a chatbot as a TA for the course. So, what happened in the course, so the student whatever queries the students were having, they also (Refer Time: 09:43) in chatbot that can try to analyze the queries and try to give a some

readymade answers, and it was interesting that after some training, the chatbot was so good that the students failed to notice.

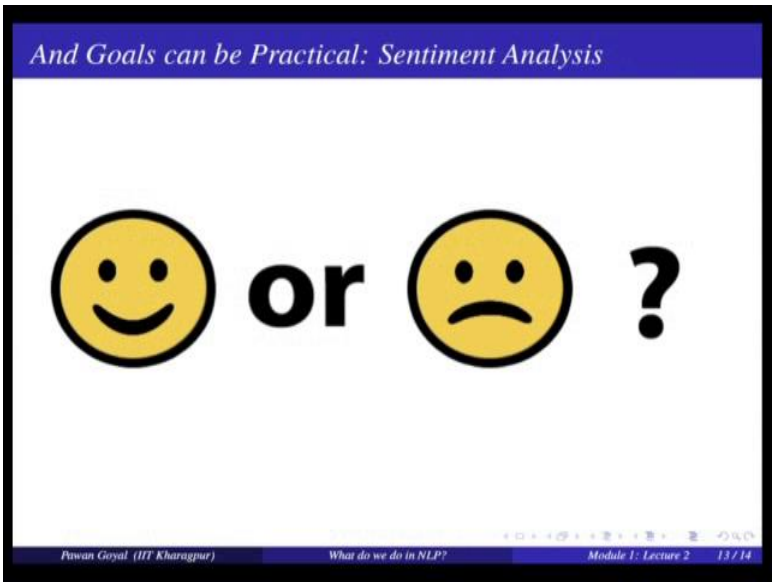
(Refer Slide Time: 09:57)



So, from the article if you see that after sometime when the chatbot learnt from the way issues for querying and the responses that where ideal for the queries, it was giving answers with some roughly like 97 percent certainty and TA is the actual, the human TA were actually check the responses first and then they will upload on the photo.

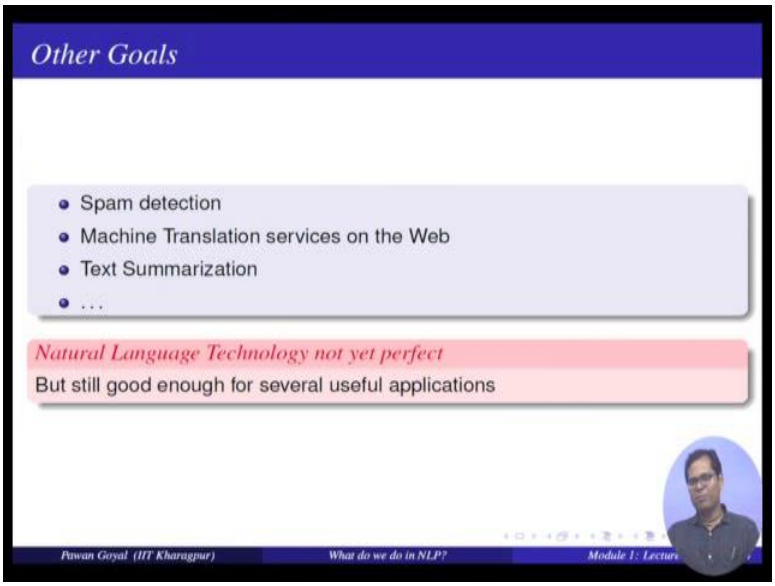
So, this was again very very practical goal. So, you can contrast it with the open domain chatbot that we talked about. So, this was very very domain is specific chatbot. So, they built it only for their own course. So, domain was fixed to their course and the kind of queries you can expect are also limited in number fixed to a certain domain. So, building these domains a specific chartbots or conversation agent is it practical very very practical goal and also coming up to be important application in recent days. So, thinking, starting from the conversation agent that can help you by some product on an e-commerce website, instead of you having to search everything; can you just provide your specifications to the chatbot and it can search a product for you or in the case of any flight booking system or banking system where you can give your queries and the chatbot can come up with the possible reply by looking into the document and so on. And this is the practical application that can be solved using NLP.

(Refer Slide Time: 11:28)



Then there is a problem of sentiment analysis again lot of work has gone in NLP on this and this is again a very very practical code. So, from your all your tweets, all your opinions and comments that you provide in social media, can a tool find out what are various sentiments of users and with that you can also find out are they some transitions and sentiments of the users over the years; a lot of research was done with recent presidential elections and in India the Lok Sabha elections, a lot of research was on finding out what are people's opinions and sentiments about various political parties and leaders; many of them actually came across to even predicting who will be the winner of the elections.

(Refer Slide Time: 12:18)



And there are many many other goals. So, we talked about some interesting goals like building say sentiment analysis, building domains, specific conversational agents, doing query completion or auto correction of the query, but there are many other goals like spam deduction. So, you will see that if your using Gmail or any other the web service many of the emails going to the spam folder directly without even bothering you. So, what it happening at the back end? Once and email comes in, so the system tries to see is it spam or not by doing again text analysis over there and if it just spam it is not even shown to you in your inbox it is directly send to some a spam folder. So, spam deduction is again a very very practical goal in not only in your emails also on with social media even on tweets, YouTube comments, even YouTube videos finding out what are its spams is again very interesting and challenging problem.

Then you have the problem on machine translation services on the web. So, think about opening a web page from some other country. So, like from China or Japan suppose you are going to visit that country and you want to read that page. So, you can use Google translate to load that page in English or in any other language for you and that is really really helpful. So, again this is a very e practical application where NLP is used. And finally text summarization, so given a big news article or scientific article can I summarize that in short; and then there are many many other applications where NLP is actually used.

So, remember one of the some of the previous slides that we saw in this lecture. So, we found that NLP technology is not perfect. So, there are many places where the systems make blunders. So, it is not perfect, but it is still good enough for many many good applications. So, you can know that by, the way you are using it in your daily life. So, you can a still use that in many day to day life applications and that is what is guiding this field that so there are lot of applications and that you can think of. So, can you come up with nice ideas, nice algorithms for solving that problem and actually people would use that and benefit from that? So, lot of applications you can think off where you can help the society by doing text processing and analytics.

So, in this lecture we discussed what are some of the things that we do in NLP. So, next lecture onwards we will start talking about see why is NLP hard, what makes the language processing a difficult ask to handle.

Thank you.