**Administrative**

|  |  |
| --- | --- |
| **S/No** | **Items** |
| 1 | Flask Attendance |
| 2 | Start recording |
| 3 | Self introduction   * History linked in * Work in the area of NLP / Dialog systems |
| 4 | Some expectations   * Lesson organisation * (15 + 5) format mins talks, followed by activities or 2 or 3 minutes to switch out * Hope as much as possible, we learn from each other |

**Concepts**

|  |  |
| --- | --- |
| S/No | Items |
| 1 | Phrase grammar structure - Phrase structure grammars model the internal structure of a sentence in terms of a hierarchically organized representation. The sentence *Every boy has a bike*, for instance, is taken to consist of a noun phrase (*every boy*) and a verb phrase (*has a bike*), where the former consists of a determiner (*every*) and a noun (*boy*), and the latter consists of a verb (*has*) and a noun phrase (*a bike*), which in turn consists of a determiner (*a*) and a noun (*bike*). |
| 2 | Markov model – originally used to as [**stochastic mode**l](https://en.wikipedia.org/wiki/Stochastic_model) used to [model](https://en.wikipedia.org/wiki/Mathematical_model) randomly changing systems.  <https://en.wikipedia.org/wiki/Markov_model#:~:text=In%20probability%20theory%2C%20a%20Markov,it%20assumes%20the%20Markov%20property).>  A **stochastic model** is a tool for estimating probability distributions of potential outcomes by allowing for random variation in one or more inputs over time. |
| 3 | Difference between **Information retrieval and Information extraction**   * Information retrieval is ***based on a query*** - you specify what information you need and it is returned in human understandable form. * **Information extraction** is about structuring unstructured information - given some sources ***all of the (relevant) information*** is structured in a form that will be easy for processing. This will not necessary be in human understandable form - it can be only for use of computer programs. |
| 4 | **Unicode** is an [information technology](https://en.wikipedia.org/wiki/Information_technology) (IT) [standard](https://en.wikipedia.org/wiki/Technical_standard) for the consistent [encoding](https://en.wikipedia.org/wiki/Character_encoding), representation, and handling of [text](https://en.wikipedia.org/wiki/Character_(computing)) expressed in most of the world's [writing systems](https://en.wikipedia.org/wiki/Writing_system). The standard is maintained by the [Unicode Consortium](https://en.wikipedia.org/wiki/Unicode_Consortium), and as of March 2020, there is a repertoire of 143,859 characters, with Unicode 13.0 (these [characters](https://en.wikipedia.org/wiki/Character_(computing)) consist of 143,696 graphic characters and 163 format characters) covering 154 modern and historic [scripts](https://en.wikipedia.org/wiki/Script_(Unicode)), as well as multiple symbol sets and [emoji](https://en.wikipedia.org/wiki/Emoji). |

**Concepts / Further Explanation**

|  |  |
| --- | --- |
| S/No | Items |
|  |  |

**Supplementary Resources**

|  |  |
| --- | --- |
| S/No | Resource details |
|  |  |