**Course: Advance Bio Informatics**

**Module Title: Need of corpora**

**Module No: 126**

**Need of Corpora**

Annotation of corpus, assist in the development and evaluation of an IE system. Annotated corpora are becoming increasingly common within biomedical information extraction research.Discover & extract unstructured knowledge hidden in data.

Aids to construct hypotheses from associations derived from data.Protein-protein interactions associations of genes – phenotypes. Functional relationships among gene.

**Why biomedicine?**

MEDLINE: 16,000,000 references, 40,000 added per month. Dynamic nature of the domain: new terms (genes, proteins, chemical compounds, drugs) constantly created. Impossible to manage such information overload







