**Course: Advance Bio Informatics**

**Module Title: Biomedical annotated corpora**

**Module No: 121**

Biomedical researchers are interested

(i) Understanding data

(ii) Context information

(iii) Background knowledge

(iv) Curated databases &

(v) Literature extensive

**What is Annotated Corpora?**

Dataset for extraction of disease/treatment entities relations

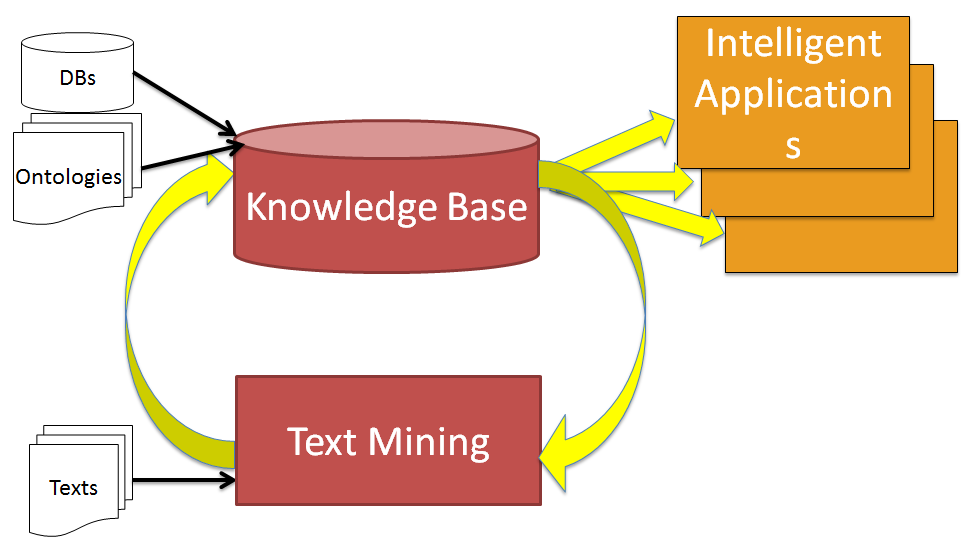
Corpora are usually constructed for training or evaluation purposes during the development of particular system

**Annotation Consumers**

The linguistic community typically uses annotation as training data or for specific tasks. An abundance of tools that can *produce* annotations in the specific format of those resources.

Biomedical annotation typically used for gene set enrichment analysis.





**Information deluge**

Bio-databases, controlled vocabularies and bio-ontologies encode small fraction of information.

**Linking text to dbs and ontologies**

Curators struggling to process scientific literature

Discovery of facts & events crucial for gaining insights in biosciences: need for text mining