Database search engines

Shatakshi Kulkarni

Entrez

- https://www.ncbi.nlm.nih.gov/Web/Search/entrezfs.html
- Search NCBI databases - NLM (nih.gov)
- Global Query Cross Database search system
- It is an integrated search and retrieval system and allows to search discrete health sciences databases
- Allows access to large database only by a single query string and Efficiently retrieves sequences, structures and references
- It allows users to search for and retrieve information from multiple NCBI databases through a single interface.
- Provides view of gene, gene sequences and chromosome maps
- Ability to integrate information by cross-referencing from NCBI based on pre-existing & logical relationships between individual entries
- Databases linked to Entrez retrieval system:

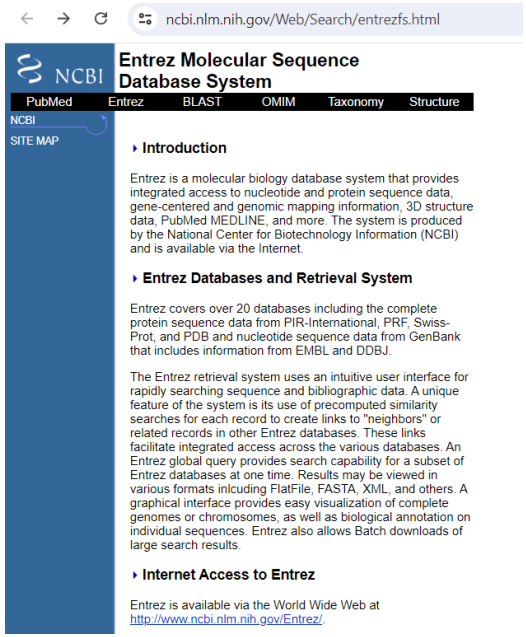
a) PubMed
b) Books (Online books)
c) Nucleotide sequence database
d) Protein databases
e) Genome databases
f) Structure databases

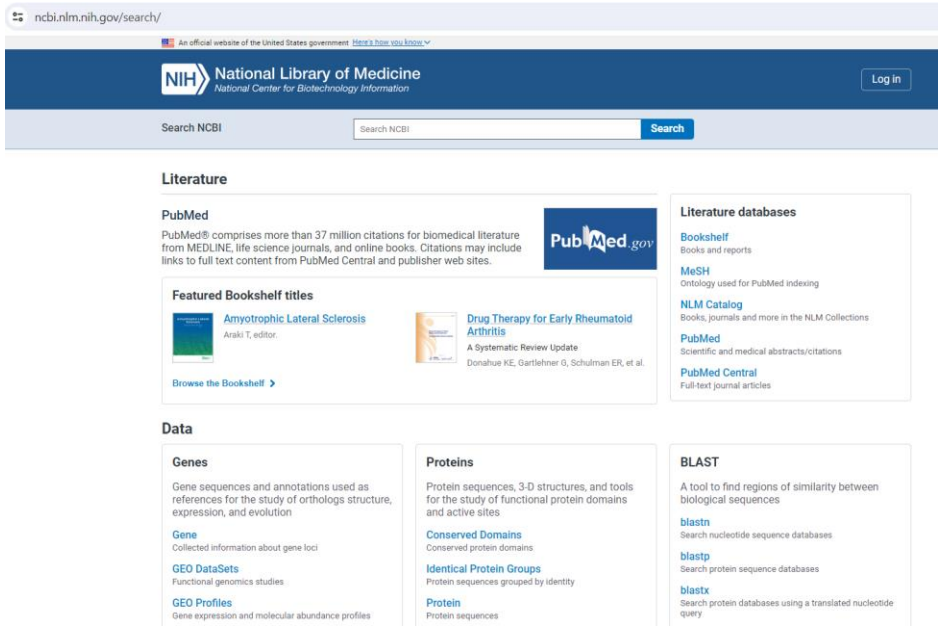
Entrez

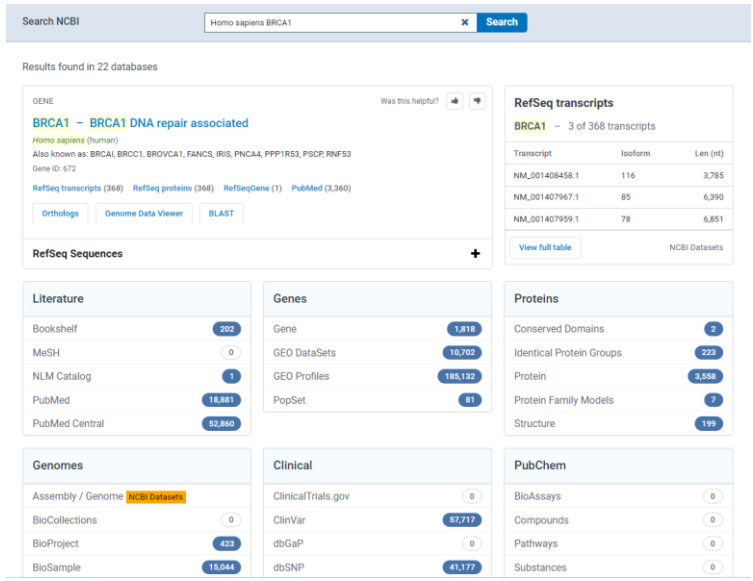
❑ Entrez search fields:

- Keyword allows to search a set of indexed terms
- Accession allows to search accession numbers
- Author Name
- Affiliations of authors
- Journal Title
- E.C. Numbers
- Feature Key searches for particular DNA feature
- SeqId is string identifier
- Title Words
- Text Words
- Organism
- PubMed ID
- Publication and modification date
- Protein Name

Entrez







Homepage

After searching for a query

SRS/Sequence Retrieval System

- **Retrieval:**

→ It is an information indexing and retrieval system designed for libraries with a flat file format such as the EMBL nucleotide sequence databank, the SwissProt protein sequence databank or the Prosite library of protein subsequence consensus patterns.

→ Retrieval system maintained by EBI

→ The SRS search engine allows researchers to perform precise searches, retrieve specific sequences, and conduct comprehensive analyses.

→ Not integrated as Entrez but allows to query multiple databases

→ Offers direct access to sequence analysis applications

→ Queries launched using Quick text search

- **Input/Output Flexibility:**

→ Supports various formats, with particular adaptation to GCG/Genetics Computer Group programs, making it a versatile tool for bioinformatics research.

- SRS (Sequence Retrieval System) is a system designed to support the data structure of various biological databases (libraries) by creating special indices.
- **These indices help manage and navigate complex data formats, including:**

→ **Lists of sub-entities:** For example, feature tables in a genomic database that list genes, exons, or other annotations.

→ **Hierarchically structured data-fields:** For instance, taxonomic classification that organizes species into a hierarchy of kingdoms, phyla, classes, etc.