1.what is document retrieval

Document retrieval is sometimes referred to as, or as a branch of, **text retrieval**. Text retrieval is a branch of [information retrieval](https://en.wikipedia.org/wiki/Information_retrieval) where the information is stored primarily in the form of [text](https://en.wikipedia.org/wiki/Natural_language). Text databases became decentralized thanks to the [personal computer](https://en.wikipedia.org/wiki/Personal_computer) and the [CD-ROM](https://en.wikipedia.org/wiki/CD-ROM). Text retrieval is a critical area of study today, since it is the fundamental basis of all [internet](https://en.wikipedia.org/wiki/Internet) [search engines](https://en.wikipedia.org/wiki/Search_engine).

2.controlled indexing vs. natural language indexing

(1) controlled indexing:

In [library and information science](https://en.wikipedia.org/wiki/Library_and_information_science) controlled vocabulary is a carefully selected list of [words](https://en.wikipedia.org/wiki/Word_(linguistics)) and [phrases](https://en.wikipedia.org/wiki/Phrase), which are used to [tag](https://en.wikipedia.org/wiki/Tag_(metadata)) units of information (document or work) so that they may be more easily retrieved by a search.[[3]](https://en.wikipedia.org/wiki/Controlled_vocabulary#cite_note-3)[[4]](https://en.wikipedia.org/wiki/Controlled_vocabulary#cite_note-4)Controlled vocabularies solve the problems of [homographs](https://en.wikipedia.org/wiki/Homographs), [synonyms](https://en.wikipedia.org/wiki/Synonyms) and [polysemes](https://en.wikipedia.org/wiki/Polyseme) by a [bijection](https://en.wikipedia.org/wiki/Bijection) between concepts and authorized terms. In short, controlled vocabularies reduce ambiguity inherent in normal human languages where the same concept can be given different names and ensure consistency.