Focus of this course

The most popular application of Information Retrieval is searching information on the World Wide Web, by means of **search engines** as Naver, Google, Yahoo etc.

There are other applications, like **document databases** (e.g., as found in libraries) and **text searches**, always relying on **complex mathematical models** (e.g., based on probability theory) to assess the *relevance* of the retrieved information or the *efficiency* of the retrieving algorithms.

In this course, however, we shall focus exclusively on text search, because it is the simplest way to start the study of this wide topics.

Focus of this course (cont)

Text search, in its simplest form, refers to the search of the occurrences of a word in a text, i.e. determine whether a word occurs in a given text and, if so, where.

All text editors include such a feature, as well as scripting languages as Perl and awk.

There is a famous utility called `grep` under UNIX which allows to search a set of strings in a text very efficiently.

Focus of this course (cont)

A **text** is nothing else than a string of characters and a word is a special case of a **pattern**.

In general, a pattern can describe a set of words, not only a single word. For example, the pattern `abra.*bra` defines the set of words starting by `abra` and ending by `bra`.

The efficiency of a search algorithm is usually measured by **the number of letter comparisons**: the lower, the better.