

Information Retrieval Cheatsheet

Edoardo Riggio

October 25, 2021

Operating Systems - SA. 2021
Computer Science
Università della Svizzera Italiana, Lugano

Contents

1	Introduction	2
1.1	Text Information Systems	2
1.2	Relevance	2

1 Introduction

1.1 Text Information Systems

Text Information Systems involve three main capabilities:

- Text Retrieval

Information Retrieval is a field concerned with the structure, analysis, organization, storage searching, and retrieval of information.

- Text Analysis

Analyze large amounts of text data in order to discover interesting patterns buried in text.

- Text Organization

Annotate a collection of text documents with meaningful topical structures so that scattered information can be connected and navigated.

While text retrieval is part of information retrieval, text analysis and text organization are part of text mining.

Differently from queries done on DBMS, queries in search engines make use of natural language. It is much harder to compare the text query to the document text and determining what is a good match and what is not a good match. This is the core issue of information retrieval. There are many different ways of writing the same thing, thus an identical matching of words is not enough.

1.2 Relevance

A document is said to be relevant when it contains the information that a person was looking for when he/she submitted the query to the search engine.

In order to understand what the user is asking for in the query, we use something that is known as **NLP** (Natural Language Processing). NLP is concerned with developing techniques for enabling computers to understand the meaning of natural language text.

2 Text Access