

Increasing amounts of multimedia content are being produced and stored on a daily basis. In order to make this data useful, computer applications are required that facilitate search, browsing, and navigation through these large data collections.

The first part of this thesis describes our approach to carry out multimedia search and indexing by connecting the textual information and visual content. The experiments were carried out considering the TRECVID Video Hyperlinking task.

Different combinations of monomodal queries are experimentally evaluated, and the impact of both parameters and single features are discussed to identify their contributions. The Automatic Feature Selection (AFS) algorithm gain the best-performing approach at the TRECVID 2017 video hyperlinking challenge. The proposed algorithm includes three different monomodal queries based on enriched feature sets.

The second part of this thesis is related to textual information analysis for discovering of research collaborations among multiple authors on single or multiple topics. Identifying the most relevant scientific publications on a given topic is a well-known research problem. The Author-Topic Model (ATM) is a generative model that represents the relationships between research topics and publication authors. It allows us to identify the most important authors on a particular topic. Specifically, we exploited an exploratory data mining technique, i.e., Weighted Association Rule (WAR) mining, to analyze publication data and to discover correlations between ATM topics and combinations of authors.

The applicability of the proposed approach was validated on real data acquired from the Online Mendelian Inheritance in Man catalog of genetic disorders and from the PubMed digital library. The results confirm the effectiveness of the proposed strategy.