

毕业设计（论文）

题目名称：跨媒体信息检索系统的设计与实现

——图片检索

院系名称：软件学院

班 级：RB软工卓越171

学 号：201719044105

学生姓名：于锦江

指导教师：郭丽

2021 年 6 月

**摘要**

“跨媒体信息检索系统的设计与实现”是一个面向互联网用户的信息检索系统。与传统信息检索不同的是，跨媒体信息检索的对象不仅是单一类型的文本对象，而是包括图像、声音以及视频等类型的数据。

…

关键词：跨媒体；信息检索；语义分析；图片识别；词向量计算；

**Abstract**

# 第一章 项目简介

本章的主要内容是“跨媒体搜索引擎”的项目背景、现实意义与实现目标的介绍，此部分内容主要通过分析目前市场现存的产品，进而来阐述本系统在当前环境下能解决的问题。

### 1.1 项目背景

对于信息检索来说，图像对信息的表达更为直观且包含更多更丰富的意义，是人们获取信息不可或缺的重要来源，然而图片在计算机中的识别过程，还远未达到人脑在图片识别过程中能够直接理解其中语义信息的能力，基于语义的图像检索是研究的高级目标。

图像检索，又称图像检索 ，系统是一个电脑浏览的系统，从一个大型的数字图像数据库去检索和检索图像。大多传统和一般图像检索的方式是利用一些增加元数据(metadata)的方法，例如：字幕、关键词或是图像的说明，如此一来就可以透过注解词完成检索。人工的图像注解是费时、费力并且昂贵；为了解决这个问题，已经有大量的研究在做自动图像注解方面上。此外，越来越多的社会网络应用和语义网已经产生了数个以网络为基底发展的图像注解工具。

第一个以微电脑为基底开发的图像数据库检索系统，是由80年代麻省理工学院里的Banireddy Prasad、Amar Gupta、Hoo-min Toong、and Stuart Madnick所共开发出来的。这是记载于1987年2月发行的IEEE Transactions on Industrial Electronics。另一种方法的图像检索是以内容为基底的图像检索（Content-based Image Retrieval，简称CBIR）技术，其目的是为了避免使用文字上的描述而是以视觉相似性为基底透过用户提供查询的图像或是用户指定的图像特征来检索图像。

从20世纪70年代开始，有关图像检索的研究就已开始，当时主要是基于文本的图像检索技术（Text-based Image Retrieval，简称TBIR），利用文本描述的方式描述图像的特征，如绘画作品的作者、年代、流派、尺寸等。到90年代以后，出现了对图像的内容语义，如图像的颜色、纹理、布局等进行分析和检索的图像检索技术，即基于内容的图像检索CBIR技术。CBIR属于基于内容检索（Content-based Retrieval，简称CBR）的一种，CBR中还包括对动态视频、音频等其它形式多媒体信息的检索技术。

考虑基于文本的检索和基于内容的检索的优势和劣势，尽可能对图片本身及其环绕文本所蕴藏信息的利用，以及对分布式并行计算云平台并行图片检索的应用，达到准确快速对图片进行检索的目的。

在检索原理上，无论是基于文本的图像检索还是基于内容的图像检索，主要包括三方面：一方面对用户需求的分析和转化，形成可以检索索引数据库的提问；另一方面，收集和加工图像资源，提取特征，分析并进行标引，建立图像的索引数据库；最后一方面是根据相似度算法，计算用户提问与索引数据库中记录的相似度大小，提取出满足阈值的记录作为结果，按照相似度降序的方式输出。

为了进一步提高检索的准确性，许多系统结合相关反馈技术来收集用户对检索结果的反馈信息，这在CBIR中显得更为突出，因为CBIR实现的是逐步求精的图像检索过程，在同一次检索过程中需要不断地与用户进行交互。

图像检索的发展是一个从简单到复杂、从低级到高级的过程，从最初的文本信息查询发展到基于内容的图像检索。同时随着人们对图像理解、图像识别研究的不断深入，提出了基于图像语义的检索,充分利用了图像的语义信息, 提高了图像检索系统的能力。另外，为了解决语义鸿沟的问题，人们提出了基于反馈的信息检索技术，利用人机交互行为，改进系统的能力，提高检索结果的准确性。最后，随着人工智能和信息技术的发展，一种智能的基于知识的信息检索系统成为信息检索领域的发展方向。基于知识的信息检索技术将基于视觉特征和基于文本语义的技术结合在一起，通过建立知识库，实现自动提取语义和图像特征的功能，并且充分考虑到用户特征对检索系统的影响, 这是建立高效、实用、快速的图像检索系统的必然的发展方向。并且图像检索领域的关键问题是对人类视觉机制的进一步了解,即探求人是如何去感知图像内容的, 这个问题的解决能够进一步优化数据特征索引技术, 解决对大规模数据库检索速度的瓶颈问题。随着多媒体数据压缩技术和互联网的迅速发展，信息的形式多种多样,视觉信息数据不仅包括单幅的图像数据还包括视频数据，针对视频数据的特点，进行高速、可靠的检索也是一个需要研究的课题。将信息检索技术推向实用化, 也是信息技术发展的主要目标。

### 1.2 研究状况调研情况

到目前为止，已经推出的较为成熟的跨媒体搜索的产品有“百度识图”、“Google 识图”、“微信搜歌”等，它们都是将多媒体作为自己的搜索输入条件，来搜索出用户想 要的搜索结果，而且准确率已经达到了十分可观的程度。由于跨媒体搜索的语义解析大致可以分为“文字语义解析”、“图片语义解析”、 “音频语义解析”、“视频语义解析”四个方面，现在分别介绍：

（1）文字语义解析：文字语义的解析经过多年的发展与运用，已经十分成熟，很多文字解析的算法已经非常成熟。

（2）图片语义解析：现在图片的语义解析的算法研究已经十分成熟，相关的识图软件诸如“百度识图”、“Google 识图”也已经推出。

（3）音频语义解析：现在音频语义的解析相关的引擎已经较为成熟，相关的识别软件诸如“微信搜歌”也已经推出。

（4）视频语义解析：现在视频语义的解析算法研究还不是十分成熟，要推出相应的产品可能还需要经过一段时间。

总体而言，现在单个文件格式的解析算法的研究已经趋于成熟，但如何将各种算法集合到一起，实现真正的跨媒体搜索，到目前为止还没有一个成熟的方案。因此，本文提出了将文字解析、图片解析、音频解析、视频解析集于一体，并且考虑到各种算法的差异性与多样性，提出了一个可扩展的、可以自由配置解析算法的跨媒体搜索平台。

### 1.3 项目目标

本系统的目标是实现对音频，视频，图片，文本这三类文件资源，基于语义内容进行分析以实现对跨媒体资源的检索。主要包括一下三点：

一、对上传的音视频文件基于音频内容的语音识别并存储分析数据到Mysql，定时同步到搭建好的ElasticSearch搜索引擎框架上。

二、对通过爬虫爬去的文本类型的网页内容并存储在Mysql，定时到同步到ElasticSearch搜索引擎。

三、使用前端搜索页面进行搜索数据是不仅对匹配的关键词进行返回，也会返回一些根据搜索关键词进行词向量计算，返回高度相似的搜索关键词的匹配结果。第2章 系统需求

# 第3章 系统分析

# 第4章 系统设计

系统设计的优良程度关乎决定系统运行时的稳定性、以及发布后的安全性、用户使用时的便利性、后期维护时的成本性。本章将从架构设计、工作流程设计，类设计，以及数据库设计这四个方面来阐述本系统的设计思路。

## 4.1 架构设计

跨媒体信息检索系统的设计与实现，根据需求分析主要分为三个模块的任务：

音视频存储与检索方案，图片存储与检索方案，搜索引擎构建。这三个模块对应的分别是音频、视频、图片、以及文本类型媒体数据的文件分析。

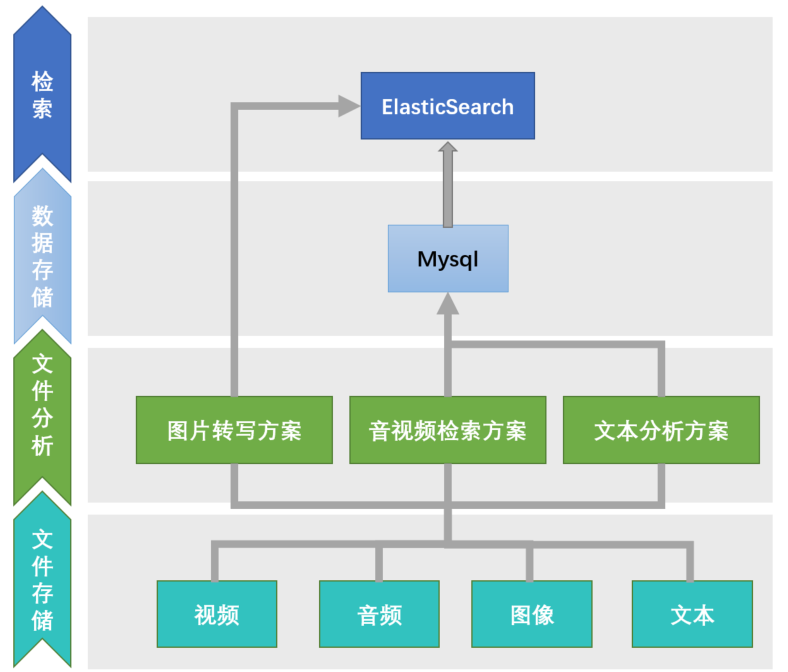


图4-1 系统架构设计图

## 4.2 工作流程设计

本章主要描述图片检索的工作流程，首先使用VGG16对数据集进行特征向量提取，并将特征向量和图片路径作为键值对存储在h5文件中（HDFS5），通过 group 有效的将多种 dataset 进行管理和划分。检索时通过提取被检索图片的特征向量，与h5文件中的向量集进行夹角余弦距离计算特征向量之间的距离。获得距离最小的数据集。

　　图像检索过程简单说来就是对图片数据库的每张图片抽取特征(一般形式为特征向量)，存储于数据库中，对于待检索图片，抽取同样的特征向量，然后并对该向量和数据库中向量的距离（相似度计算），找出最接近的一些特征向量，其对应的图片即为检索结果。

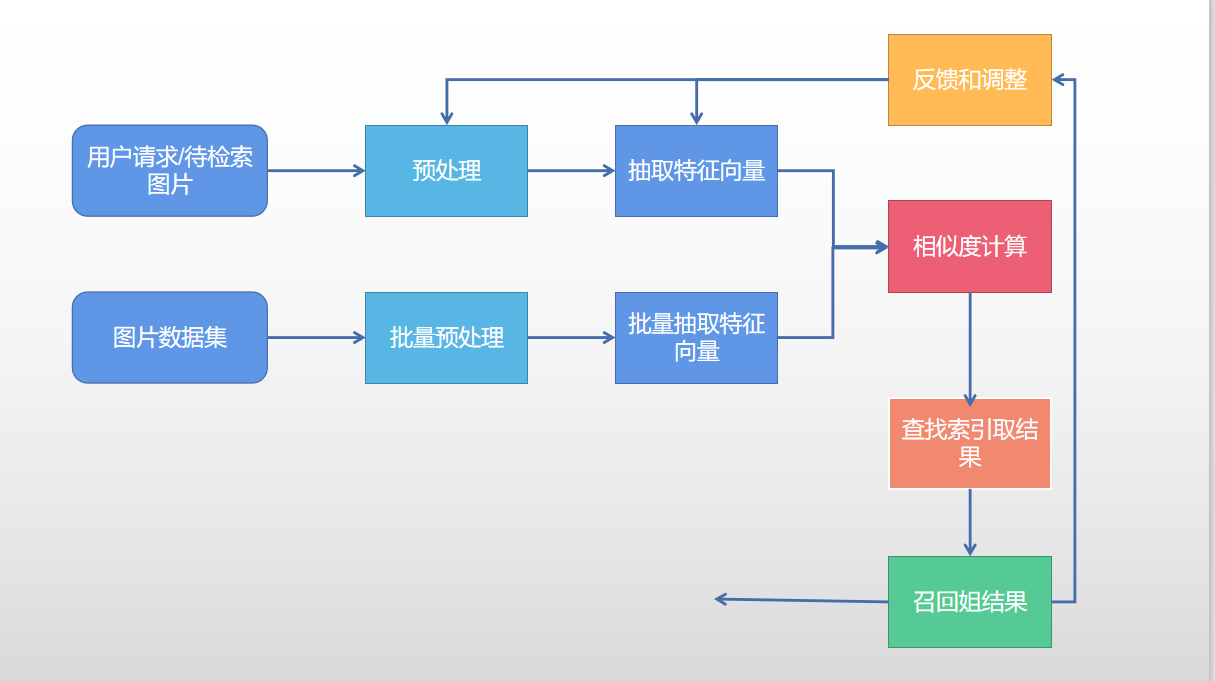


图4-2图片检索流程图

## 4.2 设计类

## 4.3 数据库设计

### 4.3.1 概念模型

### 4.3.1 关系模型

### 4.3.1 物理模型

# 第5章 系统实现

# 第6章 系统测试

# 第7章 结束语

# 致谢

# 参考文献