atitit 知识压缩的总结 学习方法总结v2 t55.docx

目录

[1.1. 压缩目的 1](#_Toc23353)

[1.2. 压缩分类 无损压缩 有损压缩 1](#_Toc1664)

[2. 功能指标 2](#_Toc8572)

[2.1. 压缩比：有损压缩一般10:1=====30:1 2](#_Toc10882)

[2.2. 压缩速度：主要与压缩算法复杂度有关 2](#_Toc12332)

[3. 常见知识压缩方式 2](#_Toc28906)

[3.1. 抽象化提升一层 概念化 2](#_Toc23982)

[3.2. 分类打包 专题化 类比压缩，举一反三 2](#_Toc24024)

[3.3. 骨架 ，目录化 大纲化 归纳整理 2](#_Toc20393)

[3.4. 记住搜索关键词 关键点 类似于视频压缩的关键帧 2](#_Toc4665)

[3.5. ，摘要 ，丢弃细节部分 包含了重要信息，远少于原文本的一半 2](#_Toc30129)

[3.6. 知识头（what 暂忽略how） 3](#_Toc12355)

[3.7. 打包压缩 体系化 知识树 3](#_Toc24138)

[3.8. 结构化，增加分类tag属性等 3](#_Toc21742)

[4. 识别重点和细节 3](#_Toc24845)

[4.1. 常用的频繁的为重点 3](#_Toc22011)

[4.2. 面试高频问题为重点 3](#_Toc12026)

[4.3. 概念复杂的 或理解复杂的点一般根据自身情况适当调整 3](#_Toc12693)

[5. 有效提升压缩质量的综合方法 3](#_Toc8752)

[5.1. VBR技术的应用。](#_Toc17290)*[variablebitrate](#_Toc17290)*[的缩写，意思是可变比率 3](#_Toc17290)

[6. 压缩的还原 4](#_Toc12460)

[6.1. 填充 4](#_Toc13736)

[7. 主流的文本摘要方式抽取式（extractive），另一种是生成式（abstractive）。 4](#_Toc16830)

[7.1. 抽取式（extractive）简单 4](#_Toc3024)

[7.2. 生成式（abstractive）意思的基础上重新生成概要。 4](#_Toc25677)

## 压缩目的

减少体积占用。。。

## 压缩分类 无损压缩 有损压缩

无损压缩 主要是文本文件 zip rar gz

有损压缩主要 视频mp4 音频mp3

Wav非压缩

# 功能指标

## 压缩比：有损压缩一般10:1=====30:1

## 压缩速度：主要与压缩算法复杂度有关

# 常见知识压缩方式

## 抽象化提升一层 概念化

Mybatis hibenate jpa 》》orm

## 分类打包 专题化 类比压缩，举一反三

## 骨架 ，目录化 大纲化 归纳整理

如何目录化，按照大纲来即可

## 记住搜索关键词 关键点 类似于视频压缩的关键帧

通过关键词还远内容

## ，摘要 ，丢弃细节部分 包含了重要信息，远少于原文本的一半

包含了原文本中的重要信息，其长度不超过或远少于原文本的一半”

归纳与演绎\_百度百科

标题摘要 vs 关键词摘要

## 知识头（what 暂忽略how）

## 打包压缩 体系化 知识树

大量小文件占有大量体积，打包会压缩空余，缩小体积

## 结构化，增加分类tag属性等

# 识别重点和细节

## 常用的频繁的为重点

## 面试高频问题为重点

## 概念复杂的 或理解复杂的点一般根据自身情况适当调整

# 有效提升压缩质量的综合方法

## VBR技术的应用。*variablebitrate*的缩写，意思是可变比率

PS：VBR是*variablebitrate*的缩写，意思是可变比率，就是MP3文件压制的时候[声音](https://baike.baidu.com/item/%E5%A3%B0%E9%9F%B3" \t "https://baike.baidu.com/item/%E6%9C%89%E6%8D%9F%E5%8E%8B%E7%BC%A9/_blank)元素较多，比率较高时，将自动减低压缩[比特率](https://baike.baidu.com/item/%E6%AF%94%E7%89%B9%E7%8E%87" \t "https://baike.baidu.com/item/%E6%9C%89%E6%8D%9F%E5%8E%8B%E7%BC%A9/_blank)，在比特率需求比较低时自动升高比特率，这样做的目的是在保证音质基本不被损害的情况下增加文件在线播放时的速度，和减少在本机播放时所占的[系统资源](https://baike.baidu.com/item/%E7%B3%BB%E7%BB%9F%E8%B5%84%E6%BA%90" \t "https://baike.baidu.com/item/%E6%9C%89%E6%8D%9F%E5%8E%8B%E7%BC%A9/_blank)……这是Xing发展的算法，他们将一首歌的复杂部分用高Bitrate编码，简单部分用低Bitrate编码。主意虽然不错，可惜Xing编码器的VBR算法很差，音质与CBR相去甚远。幸运的是，Lame完美地优化了VBR算法，使之成为MP3的最佳编码模式。这是以质量为前提兼顾文件大小的方式，推荐编码模式。

# 压缩的还原

## 填充

# **主流的文本摘要方式**抽取式（extractive），另一种是生成式（abstractive）。

## 抽取式（extractive）简单

## 生成式（abstractive）意思的基础上重新生成概要。

目前主流的文本摘要自动生成有两种方式，一种是抽取式（extractive），另一种是生成式（abstractive）。

抽取式顾名思义，就是按照一定权重，从原文中寻找跟中心思想最接近的一条或几条句子。而生成式是计算机通读原文，在理解整篇文章意思的基础上，重新生成概要。

抽取式摘要目前已经相对成熟，但抽取质量及内容流畅度均差强人意。伴随着深度学习的研究，生成式摘要对质量和流畅度都有很大的提升，但目前也涉及到原文本长度过长、抽取内容不佳等问题的限制。

**文本摘要的发展概况**

抽取式摘要是一种比较成熟的方案，其中Text rank排序算法以其简洁、高效的特点被工业界广泛运用。大体思想就是先去除文章中的一些停用词，之后对句子的相似度进行度量，计算每一句相对另一句的相似度得分，迭代传播，直到误差小于0.0001，再对上述方法得到的关键语句进行排序，即可获得摘要。抽取式摘要主要考虑单词词频，并没有过多的语义信息，像“猪八戒”、“孙悟空”这样的词汇都会被独立对待，无法建立文本段落中完整的语义信息。

入门\_ 文本摘要自动生成技术的前世今生 - 简书.html