基于信号指纹的电磁异常自动寻找与分类

# 0 引言

地震地磁异常检测是研究地震预警预报以及其他空间电磁异常现象的基础。很多研究都表明地震发生前后会有相应的短时电磁异常。在研究异常的过程中，各种人为和重复出现的干扰影响了我们对于电磁异常的判断，然而各种干扰成因复杂，很难通过单一的人工挑选的方式一一排除。对于这些干扰和异常的判断一种可靠的方式就是将所有干扰一一找出，并进行分类。那么在后续的研究中可通过查表的方式去索引干扰信息，对真正地震相关的电磁异常的寻找起到关键性的推动作用。

本文结合了短时傅里叶变换、小波变换和文本相似性哈希计算，将一定时间长度内的波形数据转换为整型数据(指纹)，利用指纹进行索引分类，使得每个整形数组所对应的波形相似。短时傅里叶变换的作用在于将时序数据转换为二维傅里叶谱数据，这个过程相当于将时序数据转换为与时间无关的数据，二维小波变换中采用了声音识别中比较成熟的希尔小波变换，进一步提取特征。之后将所得图形二值化进行文本相似性哈希计算，其原理在于相似的文本所得到的哈希值相同。最终利用所得数值进行分类统计，并建立索引，用以之后的快速查找与分类。

本文将上述波形指纹提取过程应用于GM-4的单分量数据，尝试加入人为异常并进行分析，以期为后续工作提供参考服务。

# 1.指纹提取过程

本质来讲本文所叙述的过程是一个无监督优化的聚类问题，对于无监督优化过程需要指定数据相似度标准，通常来讲地学中常用的相似度标准是互相关(COR)：

(1.1)

通常来说，对于短时异常秒采样数据n通常都大于300，而对于一天的数据而言每隔2秒进行向量提取的话通常要提取4万多个向量，显然在利用互相关系数作为“距离”对所有向量进行聚类的过程中需要计算海量的计算，同时还会面对噪声的问题。所以本文利用了文本相似性哈希的原理将信号特征进行压缩。

首先对于时序数据进行短时傅里叶变换(STFT)

(1.2)

其中w(t)是窗函数，本文中取汉宁窗：

(1.3)

短时傅里叶变换的意义与小波变换的意义类似均是将信号利用基函数进行展开，最终使得时序数据的特征与时序不相关，这个过程有利于进行特征的提取，使得其相比于互相关过程的优势在于弱化了窗函数延时长短的影响。

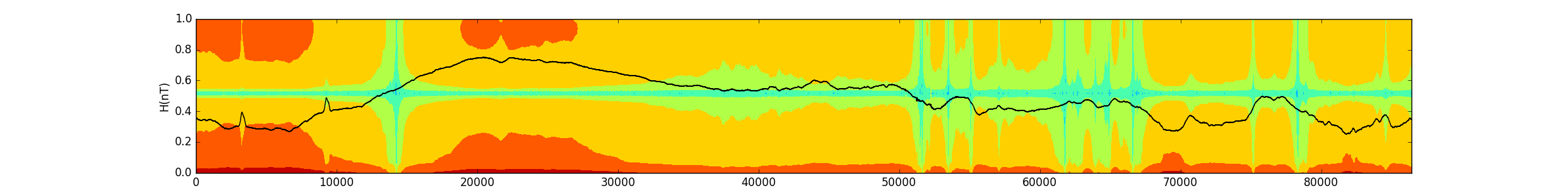


图1 短时傅里叶变换的结果(黑线为波形数据)

可以看到不同特征信号所对应的频谱图不尽相同。

之后对短时傅里叶变换所得结果进行取窗，进行二维小波变换进一步提取特征，这里选取小波基函数为希尔小波变换的子波，其母基函数为：

(1.4)

希尔小波变换常应用于信号特征的提取过程，计算过程比较简单，处理速度快。

截取图1的一定长度(600s)的二维图形进行小二维小波变换，得到结果：

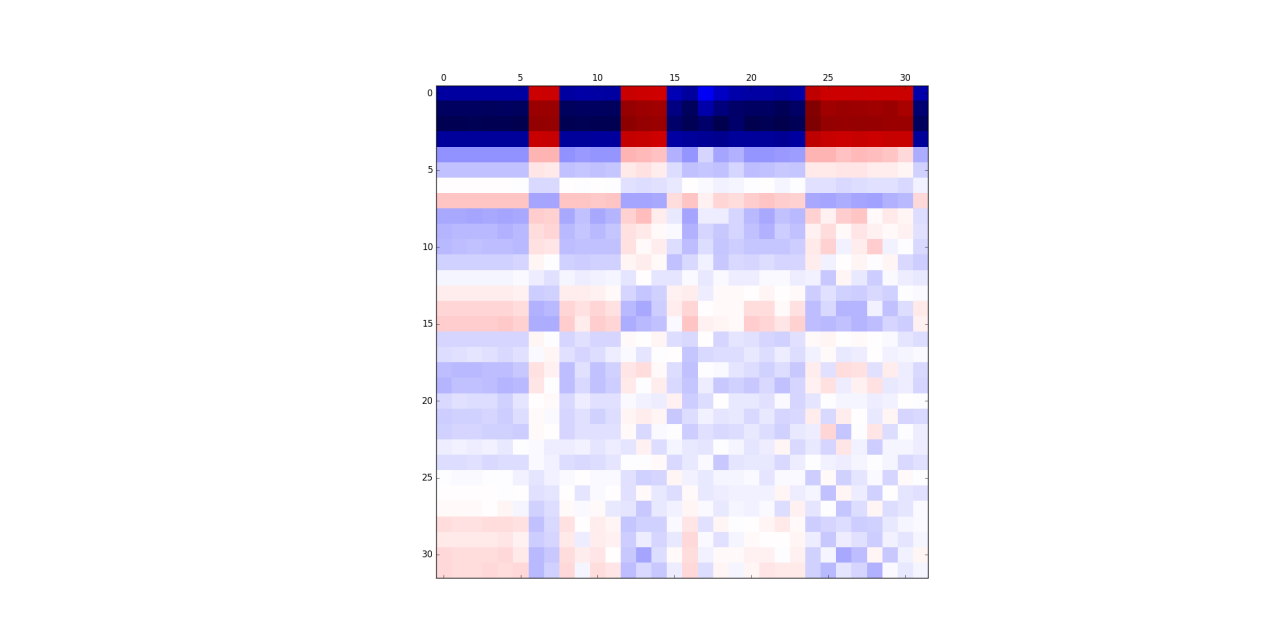


图2对局部图像作二维小波变换

对于短时傅里叶变换，选取的参数如下表所示。

表1 指纹提取过程参数

|  |  |
| --- | --- |
| **参数** | **值** |
| 原始数据采样率 | 1 |
| 短时变换时间窗 | 60(点) |
| 短时变换时间窗延迟 | 10(点) |
| 短时傅里叶变换重采样点数 | 32 |
| 二维小波变换时间窗 | 60(点) |
| 二维小波变换时间窗延迟 | 6(点) |
| 二维小波变换重采样点数 | 32 |

至此傅里叶变换和小波变的过程已经完成，在电磁异常信号的提取与分类的过程中需要解决的问题有两个，第一就是噪声信号的处理，第二就是与其他正常数据作对比，这个过程再人工选取异常的时候通常的做法就是人为的提取与分析信号的特征，这里为了自动化的处理这个过程提出了如下的公式过程。

首先将二维小波变换后的数据展开成一维向量其长度为N=，对于一天的数据而言总共有M=86400/10/6=1440个向量。将一天的M个向量写成矩阵的形式

显然对于某些点的数据其波动比较大，这里进行的处理为对除以其方差，使得波动较大的部分的权值更小：

(1.5)

其中：

经过以上处理过程解决了两个问题，第一就是如何去对比相对值大小的问题，这里在公式中表现为减平均值的过程，第二个问题就是如何应对误差的问题，这里选取的方式为对纵向对比数据加入了除以方差的过程，以减少波动较大的数据的影响。最后选取数据向量 较大的300个数据(占比300/1024=29.3%)进行文本相似性哈希的计算，对于向量中前300大的数据大于0的取1小于0的取-1。其他较小的数据则取0。最终结果如下：

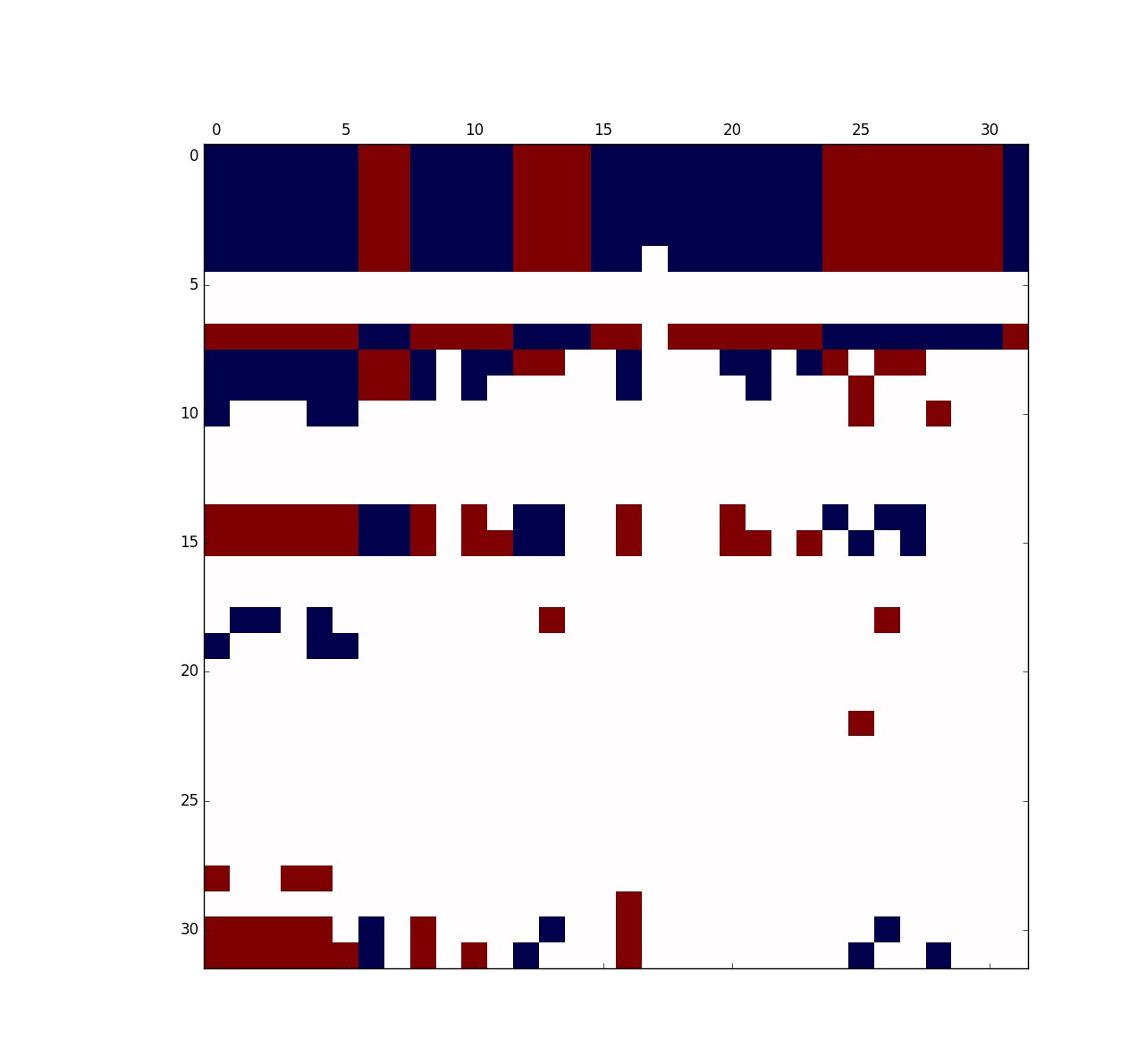


图3 处理后的二值化数据

将所得的01向量进行文本转换为哈希值。

# 2.计算结果

对于平稳过程，将所得哈希值绘制在波形图上如下：

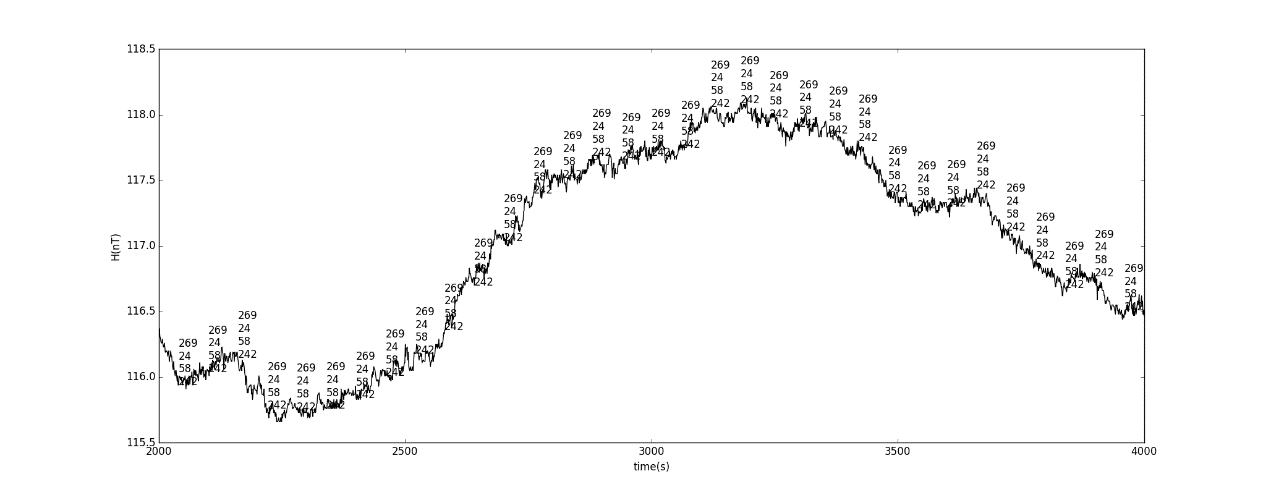
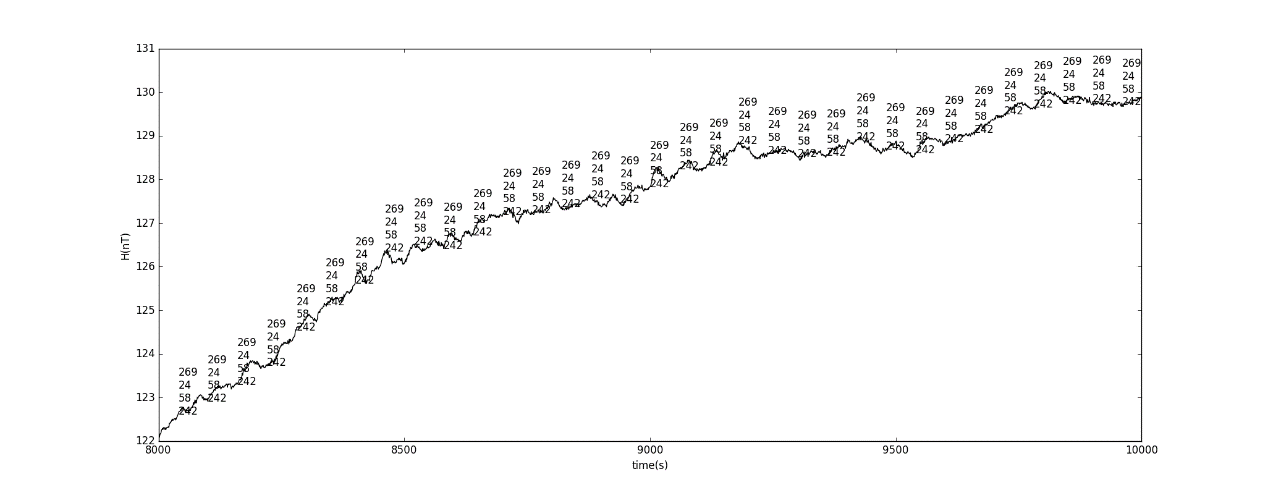


图4 未加入人为干扰数据时波形与哈希值

可见对于平稳随机过程所得的哈希值基本相同，并无明显区别。这说明计算随机过程对于整体的哈希值影响很小。

下面加入人为干扰：

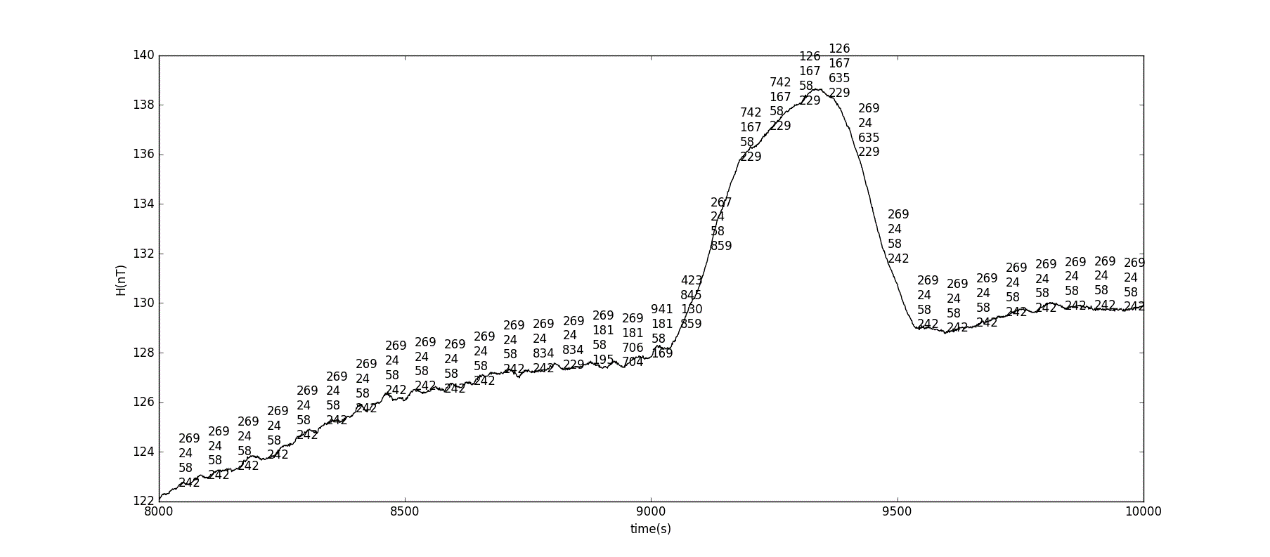
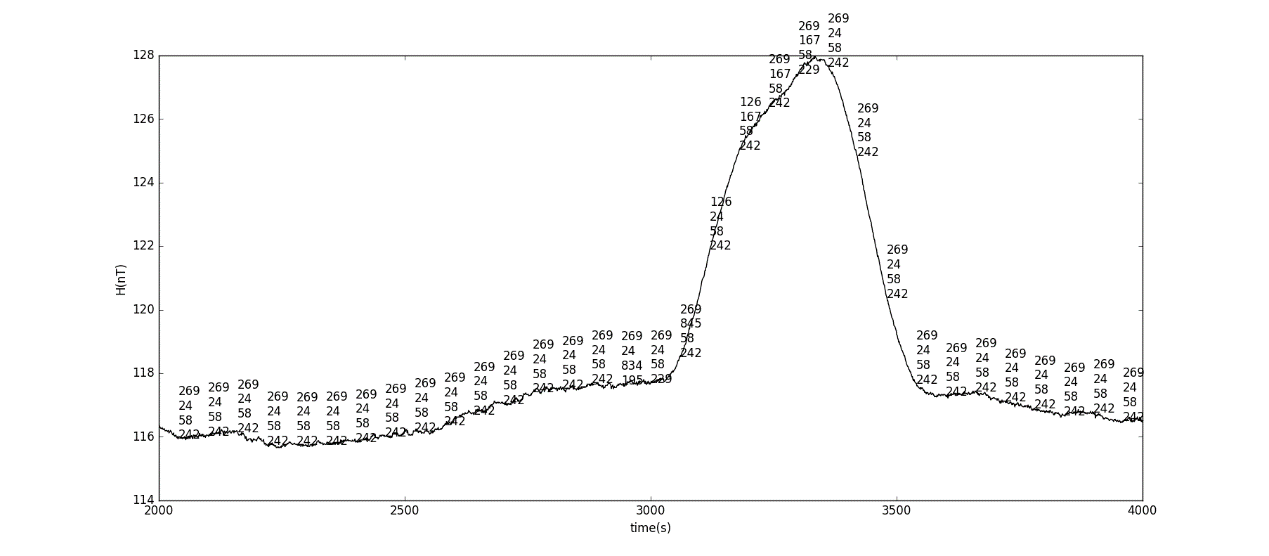


图5加入人为干扰后所得的哈希值

在相同的两地方加入相同的干扰之后所得的哈希值出现了波动，并且因为加入的干扰波形一致，所以出现了相同的哈希值：{269,167,58,229}，

# 3结论与讨论

通过本文研究可以看到波形指纹的方法对于异常的寻找与分类是本文的算法在寻找异常的过程中是行之有效的，并且在一定程度上可以对相同的异常进行分类与识别。计算过程中有效的去除了随机噪声的影响，使得噪声数据并未影响哈希值的产生，同时使得对于相同哈希值的波形具有了一定的相似特性。在计算过程中未对原始数据进行任何人为加工，相信如果加入滤波过程可能会对异常的识别更加准确。

本文不同于传统的异常识别算法，而是结合了傅里叶变换小波变换以及计算机科学中的文本相似性哈希的计算过程，而发展出了新的对于地磁异常识别与分类的思路。这使得对于大规模自动化的地磁数据处理与分类提出了十分有价值的参考。并帮助人们完善与分析地磁特征。

**参考文献**

Withers M, Aster R, Young C, et al. A comparison of select trigger algorithms for automated global seismic phase and event detection[J]. Bulletin of the Seismological Society of America, 1998, 88(1):95-106.

Geller R J, Mueller C S. Four similar earthquakes in central California[J]. Geophysical Research Letters, 1980, 7(10):821–824.

Gibbons S J, Ringdal F. The detection of low magnitude seismic events using array-based waveform correlation[J]. Geophysical Journal International, 2006, 165(1):149-166.

Bobrov D, Kitov I, Zerbo L. Perspectives of Cross-Correlation in Seismic Monitoring at the International Data Centre[J]. Pure and Applied Geophysics, 2014, 171(3):439-468.

Manber U. Finding Similar Files in a Large File System[C]// Usenix Winter Technical Conference. 1994:1--10.

Henzinger M. Finding near-duplicate web pages: a large-scale evaluation of algorithms[C]// International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2006:284-291.

Potthast M, Stein B. New Issues in Near-duplicate Detection[M]// Data Analysis, Machine Learning and Applications. Springer Berlin Heidelberg, 2008:601-609.

Pamulaparthy L, Rao D C V G, Rao D M S. A Near-Duplicate Detection Algorithm to Facilitate Document Clustering[J]. International Journal of Data Mining & Knowledge Management Process, 2014, 4(6):39-49.

Suresh S. An Efficient Internet forum crawling Technique[J]. Ijmca, 2014, 2(6).

Haitsma J, Kalker T. A Highly Robust Audio Fingerprinting System[C]// Ismir 2002, International Conference on Music Information Retrieval, Paris, France, October 13-17, 2002, Proceedings. DBLP, 2002:107--115.

Jaap Haitsma, Ton Kalker. A Highly Robust Audio Fingerprinting System With an Efficient Search Strategy[J]. Journal of New Music Research, 2003, 32(2):211-221.