

PROGRAMLAMA II DERSİ

2. PROJE

Bu projede karakter tabanlı LZ77, DEFLATE sıkıştırma algoritmaları kullanılarak dosya sıkıştırma uygulaması yapmanız istenmektedir. Proje için C / C++ programlama dili kullanılacaktır.

Projede istenenler sırasıyla aşağıda verilmektedir.

- Girdi olarak kullanılacak sıkıştırılmamış dosya, metin dosyası (metin.txt) şeklinde olacaktır.
- Sıkıştırılacak dosyadaki her bir karakter okunacaktır.
- En son her bir sıkıştırma algoritması için çıktı dosyası (metin dosyasının sıkıştırılmış hali) oluşturulacaktır.
- Kullandığınız sıkıştırma algoritmalarının performansları birbiri ile karşılaştırılacaktır.

LZ77 Sıkıştırma Algoritması

Genellikle hazırlanan metin belgelerinde, kendini tekrar eden kısımlar bulunmaktadır. Özellikle;

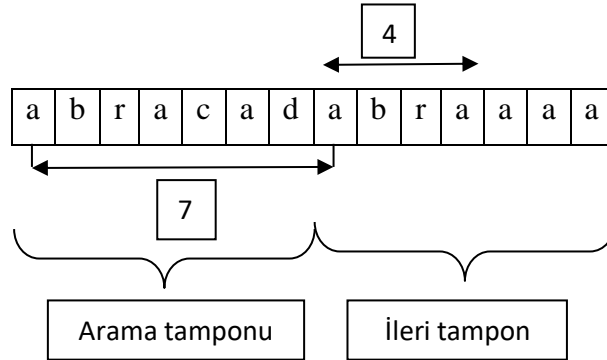
- bağlaç ve zamirler,
- diğer kelimelerin başına ve sonuna gelen ekler,

metinlerin içinde sık sık tekrar edilmektedirler.

LZ77 algoritmasının amacı, metin belgelerindeki tekrar eden bölümleri ortadan kaldırmak ve buna yönelik metin belgesini sıkıştırmaktır.

LZ77 yaklaşımında sözlük, daha önce kodlanmış serinin bir parçasıdır. Algoritmadaki arama tamponunun büyüklüğü, daha önce kodlanmış serinin ne büyüklükte bir parçasında arama yapılacağını belirler. Arama tamponu büyütüldükçe, sıkıştırma oranı artar, fakat aynı zamanda sıkıştırma zamanı da artar.

Örnek: “abracadabra” kelimesini LZ77 algoritması ile sıkıştırma.



İleri tamponun ilk karakteri olan a, arama tamponunda sondan başa doğru aranır. İkinci karşılaştırmada benzerlik bulunur, fakat bu karakterden sonra b karakteri değil de d karakteri yer aldığı için benzerlik uzunluğu sadece 1’ dir. Arama devam ettirilir. İki karakter sonra bir a daha bulunur, sonrasında c yer aldığı için bunun da benzerlik uzunluğu 1’ dir. Aramaya devam edilir. Arama tamponunun başında, yani ileri tamponda aranan karakterden 7 uzaklıkta (offset=7) bir a daha bulunur. Bu defa benzerlik uzunluğu 4’ tür (abra).

İleri tamponda “abra” serisinden sonra yer alan a karakteri ile birlikte [7,4,C(a)] şeklinde üçlü olarak kodlanır. İleri tamponun en sonundaki a karakteri ise [0,0,C(a)] şeklinde kodlanır.

DEFLATE Sıkıştırma Algoritması

Deflate Sıkıştırma Algoritması kayıpsız veri sıkıştırma formatıdır. Huffman Kodlaması ve LZSS algoritmasının bir bileşiminden oluşmaktadır.

Sıkıştırılacak olan veri birbirini takip eden bloklar kümesi olarak düşünülmektedir ve her blok LZSS algoritması ve Huffman kodlamasının birlikte kullanılması ile sıkıştırılmaktadır. Her blok için oluşturulan Huffman ağacı bir önceki ve bir sonraki bloktan bağımsızdır. Sıkıştırılabilen blokların büyüklüğü değişken olabilmektedir.

Deflate algoritmasında eğer Huffman ağacının kodlama yapamayacak kadar büyümesi durumu oluşursa, yeni bir Huffman ağacı oluşturmak için o bloğu sonlandırarak yeni bir blok başlatmaktadır.

Her blok, sıkıştırılmış verilerin bulunduğu kısım ve bu verileri temsil eden Huffman kod ağaçlarının saklandığı kısım olmak üzere iki kısımdan oluşur. Kullanılan Huffman ağaçları her

bloğun başında sıkıştırılmış veriden önce yer alır, ve bu ağaçlar da Huffman kodlaması kullanılarak sıkıştırılır.

Son Teslim Tarihi : 18 Mayıs 2020

Proje ile ilgili Arş. Gör Burcu KIR SAVAŞ ve Arş. Gör. Dilara GÖRMEZ ile irtibata geçebilirsiniz.

Proje ile ilgili sorularınız sadece e-destek sisteminde bulunan forumdan alınacaktır.

Proje teslimleriniz e-destek sistemi üzerinden yapılacaktır.

Değerlendirme sisteme yüklenilen rapor ve proje kodu ile yapılacaktır.