

Bil 214 Mini Proje#1

Girdi olarak kullanılacak olan csv dosya formatı aşağıda verilmiştir:

Cites : negatif olmayan tam sayı (boş geçilemez)

Authors : Virgül ile ayrılmış isim listesi. İsimlerde büyük harf küçük harf aynıdır. (Ornek: T Ozyer, Y Kocak, B Karatay)

Title : String formatında başlık bilgisidir. (Ornek: Sistem Programlama)

Year : Tamsayıdır(Örnek: 2006). Boş olabilir.

Source : String (Boş olabilir)

Publisher : String (Boş olabilir)

ArticleURL : String (Boş olabilir)

CitesURL : String (Boş olabilir)

GSRank : Tamsayı (≥ 0)

QueryDate: Tarih alanıdır. YYYY-MM-DD HH:mm Örnek: 2021-08-09 20:07

Type: String, boş geçilebilir. (Ornek: PDF, CITATION, HTML, ...)

DOI: String, boş geçilebilir.

ISSN: String (Boş olabilir)

CitationURL : String (Boş olabilir)

Volume :Tam Sayı (Boş olabilir)

Issue : Tam Sayı (Boş olabilir)

StartPage : Negatif Olmayan Tam Sayı (Boş olabilir)

EndPage : Negatif Olmayan Tam Sayı (Boş olabilir)

-ECC : Boş alandır. Estimated Citation Count bilgisidir. Tam Sayı (≥ 0), Cites degeri atanacaktır.

-CitesPerYear : Boş Alandır. Decimal olarak belirlenmiştir. $Cites * 365 / Age$ olarak hesaplanır.

-CitesPerAuthor : Cites / AuthorCount değeridir. Girdi olarak verilen inputta boştur.

-AuthorCount : Boş alandır. Tamsayı(≥ 0) olacaktır. Authors alanındaki yazar sayısı girilir.

-Age : Boş alandır. Query Date-Year alanı(Year, 1 Ocak ve verilen Year alan değerinden oluşur) olarak hesaplanır. Gün cinsinden sonuç döndürülür.

Abstract : String (Boş geçilebilir)

FullTextURL : String (Boş geçilebilir)

RelatedURL: String (Boş geçilebilir)

Örnek csv girdisi:

Cites,Authors,Title,Year,Source,Publisher,ArticleURL,CitesURL,GSRank,QueryDate,Type,DOI,ISSN,CitationURL,Volume,Issue,StartPage,EndPage,ECC,CitesPerYear,CitesPerAuthor,AuthorCount, Age,Abstract,FullTextURL,RelatedURL

36,"T Ozyer, Y Kocak, B Karatay","System Programming",2021,"Turkish Journal of ABC","DergiPark Online

Library","https://onlinelibrary.dergipark.com.tr/doi/abs/20.1020/abcjc.2322","https://scholar.google.com/scholar?

cites=3267544328290472108&as_sdt=2002&scioldt=2008&hl=en",12,"2021-08-09

18:09:53","", "20.1020/abcjc.2322", "", "", 36,12,3,220,,,,,"This paper is about blah blah lorem

ipsum", "", "https://scholar.google.com/scholar?q=related:KRtY-2jHB3NP:scholar.google.com/&scioq=author:%33tansel+ozyer%33&hl=en&as_sdt=2007&as_ylo=2018&as_yhi=2021"

36,"T Ozyer, Y Kocak, B Karatay","System Programming",2021,"Turkish Journal of ABC","DergiPark Online

Library","https://onlinelibrary.dergipark.com.tr/doi/abs/20.1020/abcjc.2322","https://scholar.google.com/scholar?cites=3267544328290472108&as_sdt=2002&sciodt=2008&hl=en",12,"2021-08-09 18:09:53","","20.1020/abcjc.2322","","",36,12,3,220,,,,,"This paper is about blah blah lorem ipsum","","https://scholar.google.com/scholar?q=related:KRtY-2jHB3NP:scholar.google.com/&scioq=author:%33tansel+ozyer%33&hl=en&as_sdt=2007&as_ylo=2018&as_yhi=2021"

1- process.sh adlı bash scriptini yazınız. Bu script n(>0) tane csv dosyasını girdi olarak kabul edecektir. Buna göre aşağıda örnek komut verilmiştir,

Örnek : ./process.sh -s "Year Title" -i "a.csv b.csv" -o c.csv

Burada, ilgili komut -s ile sıralama alanlarını (alan sayısı bağımsızdır, sadece örnektir. 1,2,3,... adet sıralama alanı alabilir) dikkate alarak n tane csv dosyasının içindeki kayıtları verilen alanlara göre artan sırada c.csv dosyasında sıralayacaktır.

Başında '-' karakteri olan alanlar(alan ismine dahil değildir sadece boş alan olduğunu belirtmek için konmuştur) tamamen boş geldiğinden hesaplayarak dolduracaktır.

SortBy alanları - ile verilen alanlardan(-) biri olamaz. Hata mesajı verilmelidir. Komut satırında input csv girilmemesi veya option kullanılıp sıralama için referans alan verilmemesi durumunda hata mesajı verilmelidir.

Ayrıca, c.csv dosyasında ilave Total Pages adlı sütun eklenecektir (End Page-Start Page +1). İki alandan biri eğer boşsa, boş bırakılacaktır.

2- extract.sh adlı bash scriptini yazınız. Bu script n(>0) tane csv dosyasını girdi olarak kabul edecektir. Buna göre aşağıda örnek komut verilmiştir,

Örnek : ./extract.sh -i "a.csv b.csv" -t e.csv -o d.csv -v f.csv

Burada, t flagi a ve b içerisindeki tüm yazarları dikkate alarak e.csv içerisine virgül ile yazarları ayrıştırıp yazarları kodlayacaktır. Tüm satırlarda tekrarlanan yazarlar yalnız bir kere yer alacaktır(unique). Authors alanının boş olması durumunda o satırdaki author bilgisi dikkate alınmayacaktır.

Örnek e.csv :

"T Ozyer",1

"Y Kocak",2

"B Karatay",3

...

"Huseyin Kagit",4

Burada, o flagi a ve b içerisindeki tüm yazarları dikkate alarak d.csv içerisine virgül ile yazarları ayrıştırıp yazarları kodlayacaktır. Tüm satırlarda tekrarlanan yazarlar yalnız bir kere yer alacaktır(unique). Authors alanının boş olması durumunda o satırdaki author bilgisi dikkate alınmayacaktır.

Örnek d.csv :

Title, Authors.....

"System Programming",1,2,3 #sadece anlatim amaclı commenttir 1: T Ozyer, 2: Y Kocak 3:B Karatay

"Intro to C",2,3,4

Burada, v flagi f.csv dosyası içerisindeki yazar kodlamasını dikkate alarak ilk satır ve ilk sütundaki yazar noları (e.csvdeki en son sütun)na göre kaç satırda birlikte yer aldıklarını hesaplayacaklardır. Matrisin içindeki değerler birlikte yer alma durumunu içerecektir. Dolayısıyla simetrik matristir.

Örnek d.csv :

,1,2,3,4

1,22,12,9,0

2,

3,

4,

Üstte matris içi sayılar rastgele verilmiştir. Ekstra 4 kodlu yazar da eklenmiştir. Sadece örnek amaçlıdır.