BLM3010: Sistem Programlama

Proje 1: Kelime Sayma Programı

Bahar Dönemi

Teslim Tarihi: 18.04.2022 23:59

Proje Kazanımları

- C programlarını etkili bir şekilde yazma

- Komut satırı argümanlarıyla çalışma

- Verilenleri tamamlamak için dosyalardan ve stdin'den okuma

- Yapılandırılmış verileri işleyen yazma işlevleri

Değerlendirme Tablosu (toplam 100 puan)

(10) Programınız için eksiksiz manual sayfası. Tam kredi için uygun yazım ve dil bilgisi.

(10) Tüm komut satırı ve dosya bulunamadı hataları için uygun hata raporlaması.

(20) stdin, tek bir dosya veya birden çok dosyanın doğru komut satırı ayrıştırması ve işlenmesi

(20) fgetc() kullanarak satır sayısı (fscanf() kullanımı olursa 5 puan)

(20) fgetc() kullanarak kelime sayısı (fscanf() kullanımı olursa 5 puan)

(20) fgetc() kullanılarak karakter sayısı (fscanf() kullanımı olursa 5 puan)

fgetc() ile tam kredi için, her dosyada yalnızca tek bir sefer okuma yapmalı.

Gönderim

- Tek bir zip dosyası ogrencino.zip

- Kaynak kod: ogrencino\_ks.c

- Programınız için Manuel Sayfa girişi: man.txt

**Description**

Bu projede, "kelime sayımı"nın kısaltması olan komut satırı wc yardımcı programına benzer şekilde çalışan bir program yazacaksınız.

ks yardımcı programı ayrıca bir dosyadaki karakter ve satır sayısını da sayabilir..

Example output:

$ ./ks c\_dickens.txt

c\_dickens.txt 19199 161007 917028

$ ./ks B.txt

B.txt 1 3960 201961

$ ./ks random.txt

random.txt 69765 295983 6795742

Not: yukarıdaki örnekte random txt dosya içeriği verilmemiştir. Örnek çıktıdaki sayılar gmre bir txt dosyası rasgele oluşturmanız gerekmektedir.

İlk sayı satır sayısı, ikinci sayı kelime sayısını ve üçüncü sayı karakter sayısıdır. Birden fazla dosya belirtilirse, her bir kategorideki toplamlar, ayrı dosya sonuçlarından sonra da rapor edilmesi gerekmektedir.

$ ./ks c\_dickens.txt B.txt

﻿c\_dickens.txt 19199 161007 917028

B.txt 1 3960 201961

toplam 19200 164967 1118989

**Giriş:**

ks [-c -l -w] [file ...]

Hiçbir dosya belirtilmemişse, ks stdin'den okumalıdır:

$ cat c\_dickens.txt | ./ks

﻿-stdin- 19199 161007 917028

Bu durumda dosya adı için "-stdin-" belirtmelisiniz.

Bir kullanıcı ayrıca aşağıdaki gibi bir dosya adı olarak + sembolünü kullanarak stdin'den okumak istediğini belirtebilir:

$ cat c\_dickens.txt | ./ks A.txt + A.txt

﻿ks.txt 1 3960 201961

-stdin- 19199 161007 917028

ks.txt 1 3960 201961

toplam 19201 168927 1320950

Seçenekler

Dosya listesinden önce, çıktıyı yalnızca satır sayısını, kelime sayısını veya karakter sayısını (veya bazı kombinasyonları) bildirmekle sınırlamak için isteğe bağlı komut satırı bağımsız değişkenleri sağlanabilir:

-l satır sayısını yazdır

-w kelime sayısını yazdır

-c karakter sayısını yazdır

Örnekler:

$ cat c\_dickens.txt | ./ks -l B.txt + B.txt

﻿B.txt 1

-stdin- 19199

B.txt 1

toplam 19201

$ cat c\_dickens.txt | ./ks -w B.txt + B.txt

﻿B.txt 3960

-stdin- 161007

B.txt 3960

toplam 168927

$ cat c\_dickens.txt | ./wc -c B.txt + B.txt

﻿B.txt 201961

-stdin- 917028

B.txt 201961

toplam 1320950

Komut satırı argümanları da birleştirilebilir:

$ cat c\_dickens.txt | ./ks -c -l B.txt + B.txt

﻿B.txt 1 201961

-stdin- 19199 917028

B.txt 1 201961

toplam 19201 1320950

Çıktı her zaman komut satırı argümanlarının sırasına bakılmaksızın her zaman aynı sırada - satırlar, kelimeler, karakterler - rapor edilmelidir.

*Tüm komut satırı seçenekleri, tüm dosya adlarından önce gelmelidir.*

Komut satırı seçenekleri ayrılmalıdır: -wlc yerine -w -l -c kullanılmalıdır.

**Nasıl sayılır**

**Karakterler:** Yazdırılabilen ve yazdırılamayan tüm karakterleri içerir. Örneğin, fgetc() her yeni bir karakter aldığında artırın. Çıktınız, builtin fonksiyonu wc -c çıktısıyla eşleşmelidir. Programınız yalnızca yazdırılabilir ASCII karakterleri (boşluklar ve sekmeler dahil) ve LF ('\n') karakterleri içeren metin dosyaları kullanılarak test edilmesi gerekir.

**Satırlar**: Her '\n' ile karşılaşıldığında satır sayısını artırın. Çıktınız, wc -l'de olduğu gibi, yazdığınız ks -l komutu kullanılarak bildirilenlerle eşleşmelidir.

**Kelimeler**: Kelimeleri saymayı nasıl seçebileceğiniz konusunda iki yöntem var. Ancak, tam puan için fgetc() kullanmanız gerekir. fscanf() kullanırsanız şu şekilde olmalıdır;

Satır, sözcük ve karakter sayılarını belirlemek için her geçişte farklı biçim karakterleri olan bir dosyayı birden çok kez okumak için fscanf()'ı kullanabilirsiniz. Sözcükler için, sözcük sınırlarını tanıyan "%s" biçimini kullanabilirsiniz, ancak sonuçtaki sözcüğü depolamak için yeterince büyük bir arabellek belirlemeniz gerekir, bu büyük sözcükler için başarısız olabilir. Daha sonra karakterleri saymak ve yalnızca yeni satır ('\n') sembollerini algılamak için "%c" biçimini kullanarak satırları ve karakterleri okuyabilirsiniz.

Daha verimli yol, belirtilen dosyadan her seferinde bir karakter okuyan fgetc() işlevini kullanmaktır. Bu yöntemle ilgili zorluk, o zaman kelime sınırlarını tespit etmenin bir yolunu bulmanızdır. Bunu yapmak için ctype.h kitaplığını ve isspace() işlevini kullanabilirsiniz. Bu metodu kullanarak, bir dosyadan tek bir geçiş yapabilmeli ve tüm sayımlarınızı gerçekleştirebilmelisiniz.

**isspace() ile kelimeler nasıl sayılır**

Karakterleri ve satırları saymak oldukça basittir. Kelimeleri sayarken, programınız hazır wc programına benzer şekilde çalışmalıdır. Örneğin:

- Taranan geçerli karakter yazdırılabilir olduğunda (boşluk değil) ve önceki karakter bir boşluk karakteri olduğunda bir sözcük başlar (isspace() referansına bakın).

- Bir kelime, taranmakta olan mevcut karakter bir boşluk karakteri olduğunda ve önceki karakter yazdırılabilir olduğunda (boşluk değil) biter.

- Kelime sayısı, geçerli ve önceki karakterlerin her ikisi de boşluk olduğunda veya mevcut ve önceki karakterlerin her ikisi de boşluk olmadığında değişmemelidir.

**Komut Satırı Seçeneklerini Ayrıştırma**

Bu projeye tam bir çözüm, komut satırlarını işleyebilmelidir. Komut satırı seçenekleri için ayrıştırma gereksinimleri aşağıdaki gibidir:

• Komut satırı seçenekleri dosya listesinden önce gelmelidir. Dosya listesinde bir komut satırı seçeneği belirirse, başka bir dosya adı gibi işlem görür.

• Komut satırı seçenekleri bir - (zincir/kısa çizgi) ile başlamalıdır.

• Bilinmeyen seçeneklerle ilgili bir hata bildirmelisiniz (aşağıdaki örneğe bakın).

• İlk komut satırı argümanına - olmadan ulaştığınızda, dosya listesinin başladığını varsayabilir ve kalan her şeyi dosya adı olarak işleyebilirsiniz.

**Hata Durumları**

Hatalı kullanıcı girişiyle ilgili hataları bildirin. İki ana hata koşulu vardır:

(1) Bilinmeyen dosya adı: Hatayı yalnızca stderr'e bildirmelisiniz, ancak kalan dosyaları işlemeye devam etmelisiniz. Çıkmayın. Açılamayan gerçek dosya adı da dahil olmak üzere aşağıdaki biçimi kullanın:

HATA: <dosyaadı> dosyası açılamıyor

(2) Bilinmeyen komut satırı değişkeni: Hatayı stderr'e bildirmeli ve ardından 2 çıkış durum koduyla çıkmalısınız. Kullanıcının yazdığı gerçek geçersiz seçeneği de dahil olmak üzere aşağıdaki biçimi kullanın:

HATA: bilinmeyen seçenek <kullanıcı tarafından yazılan seçenek>

Tüm hata raporlamaları stderr'e yapılmalıdır. Program normal şekilde tamamlanırsa, 0 çıkış durum kodunu kullanmalısınız.

1 numaralı hatalı durumlar için örnekler;

$ ./ks doesnotexist.txt

ERROR(HATA): 'doesnotexist.txt' dosyası açılamıyor

$ ./ks doesnotexist.txt 2> /dev/null

$ ./ks c\_dickens.txt doesnotexist.txt c\_dickens.txt

﻿C\_dickens.txt 19199 161007 917028

ERROR(HATA): 'doesnotexist.txt' dosyası açılamıyor

c\_dickens.txt 19199 161007 917028

toplam 38398 322014 1834056

$ ./ks dickens.txt doesnotexist.txt dickens.txt 2>/dev/null ﻿

c\_dickens.txt 19199 161007 917028

c\_dickens.txt 19199 161007 917028

toplam 38398 322014 1834056

2 numaralı hatalı durumlar için örnekler;

$ ./ks -p c\_dickens.txt

HATA: bilinmeyen seçenek '-p'

$ ./ks -p -l c\_dickens.txt

HATA: bilinmeyen seçenek '-p'

$ ./ks -l -p c\_dickens.txt

HATA: bilinmeyen seçenek '-p'

**KISITLAMALAR**

Kopya: -600

Javadoc yorum satır eksikliği: -110

Gönderim yapıldıktan sonar demo sunumu yapılmaması: 0

Lab içinde 19.04.2022 tarihinde gönderimi yapılan doysa hoca gözetiminde download edilip sanal makinenizde koşturulmalıdır.