

## گزارش پروژه هافمن

شایان درویش پور و وحید رضائی

در ابتدا به اشکالاتی که در فایل های ارسالی موجود بود اشاره ای می‌کنم. در این فایل ها کاراکتر هایی با کد عددی ۶۵۵۳۳ وجود دارد که خارج از محدوده اُسکی است که در توضیحات پروژه اعلام نموده اید.

کد گذاری:

الگوریتم ما در ابتدا فراوانی هر کاراکتر را در متن پیدا می‌کند و با استفاده از این داده ها درخت هافمن را تولید می‌کند. پس از آن ابتدا تمامی این فراوانی ها را روی فایل ذخیره می‌کند و بین هرکدام از آنها کاما ذخیره می‌کند. بیشینه حجم این بخش برای فایل های ارسالی ۵۰۰ بایت بود که نسبت به حجم فایل بسیار ناچیز است. پس از آن شروع به ذخیره سازی کد های معادل هر کاراکتر در فایل می‌کنیم.

به خاطر وجود مشکلاتی که در الصال رشته های طولانی وجود دارد بنابراین یک حد برای آن قرار دادیم و اجازه ندادیم طول رشته بیشتر از ۱۰۰۰ شود و هرکار طول رشته به ۱۰۰۰ می‌رسید آنرا به بخش های ۸ تایی تقسیم می‌کردیم و هر قسمت ۸ تایی را به مقدار عددی آن تبدیل کرده و پس از آن روی فایل می‌نوشتیم. با این روش سرعت اجرای برنامه تا حد خوبی افزایش پیدا کرد.

دیکد کردن:

در این عملیات ابتدا از روی فایل میزان فراوانی هر حرف که در ابتدای عملیات کدگذاری نوشته شد بارگذاری می‌شود. و به وسیله آن درخت هافمن دوباره تولید می‌شود و پس از آن عملیات دیکد کردن آغاز می‌شود. در این مرحله یک استریم روی `FileInputStream` نوشته شد که هر کاراکتر ورودی از فایل را به یک رشته از ۰ و ۱ باینتری تبدیل می‌کند و در هر مرحله یکی از این مقادیر را خروجی می‌دهد. یک تابع با استفاده از این استریم در هر مرحله یک ورودی می‌خواند و روی درخت حرکت می‌کند. وقتی به ریشه رسید با بررسی اینه آیا این مقدار وارد شده می‌تواند درست باشد کاراکتر را روی فایل خروجی می‌نویسد.

به این ترتیب فایل کد و دیکد می‌شود.

خروجی برای فایل های تست:

Harry Potter 7 - Deathly Hollows.txt

Harry Potter 2 - Chamber of Secrets.txt

0compress ratio is : 1.6970034013605442

saving\_space is : 41.07259895894925

File is 100.0 percent same to decompressed file

Harry Potter 6 - The Half Blood Prince.txt

Harry Potter 3 - The Prisoner Of Azkaban.txt

Harry Potter 4 - The Goblet Of Fire.txt

Harry Potter 1 - Sorcerer's Stone.txt

Harry Potter 5 - Order of the Phoenix.txt

Hapworth.txt

Ocompress ratio is : 1.8081148381426688

saving\_space is : 44.69377835385613

File is 100.0 percent same to decompressed file

Bananafish.txt

Ocompress ratio is : 1.673552138321475

saving\_space is : 40.24685714285714

File is 100.0 percent same to decompressed file

Zooey.txt

Ocompress ratio is : 1.7613982970070534

saving\_space is : 43.22692364928547

File is 100.0 percent same to decompressed file

Franny.txt

Ocompress ratio is : 1.7407362225136478

saving\_space is : 42.5530423813445

File is 100.0 percent same to decompressed file

بخش هایی که خروجی برای آنها چاپ نشده فایل هایی است که شامل کاراکتر خارج از محدوده اسکی است.