19.1.3 A third attempt: DataIndexedStringSet

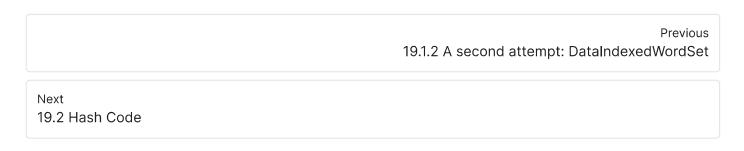[Hashing, Video 3] - ASCII and Data Indexed String Sets

There is a character format called **ASCII**, which has an integer per character. Here, we see that the largest value (i.e., the base/multiplier we need to use) is 126. Let's just do that. The same thing as `DataIndexedEnglishWordSet`, but just with base `126`.

```java
public static int asciiToInt(String s) {
    int intRep = 0;
    for (int i = 0; i < s.length(); i += 1) {
        intRep = intRep * 126;
        intRep = intRep + s.charAt(i);
    }
    return intRep;
}
```

What about adding support for Chinese? The largest possible representation is 40959, so we need to use that as the base.

So... to store a 3-character Chinese word, we need an array of size larger than **39 trillion** (with a T)!. This is getting out of hand... so let's explore what we can do to improve this, namely, using hashCode.

Last updated 1 year ago