Dr. Dapeng Dong

Maynooth University
National University of Ireland Maynooth

CS211FZ Data Structures & Algorithms (II)

Lab 3 – Merge Sort and Applications

Objectives

- Understand how merge sort work in principle
- Exercise with a standard implementation of merge sort and be able to customize it for a given problem
- Reflect the knowledge learned in the class

NOTE:

- **Do NOT use "package" in your source code**
- **You must submit the source code files, i.e., the ".java" files.**
- **You are allowed to use course reference books or class notes during the lab.**
- **Sharing your work with others is NOT allowed.**

Task 1: Understanding Merge Sort

An optimized implementation of merge sort is given on the Moodle course page (***SortingRecords.java***). The implementation can only sort integer values. The source code is provided without any comment. Your first task is to understand the code and then write a comment to explain how the method "***merge( )***" works.

Task 2: Applying Merge Sort to Applications

You are given a dataset that provides patient reviews on specific drugs. The data was gathered from online pharmaceutical review sites. The dataset contains 161,297 records. Each record consists of six attributes. A tiny fragment of the dataset is shown in Figure 1.

一行6个数据f

[[1,2,3,4,5,6],[…],[…]…]

3 5

	code	drugName	condition	rating	date	usefulCount
2	206461	Valsartan	Left Ventricular Dysfunction	9	May 20, 2012	27
3	95260	Guanfacine	ADHD	8	April 27, 2010	192
4	92703	Lybrel	Birth Control	5	December 14, 2009	17
5	138000	Ortho Evra	Birth Control	8	November 3, 2015	10
6	35696	Buprenorphine / naloxone	Opiate Dependence	9	November 27, 2016	37
7	155963	Cialis	Benign Prostatic Hyperplasia	2	November 28, 2015	43
8	165907	Levonorgestrel	Emergency Contraception	1	March 7, 2017	5
9	102654	Aripiprazole	Bipolar Disorde	10	March 14, 2015	32
10	74811	Keppra	Epilepsy	1	August 9, 2016	11
11	48928	Ethinyl estradiol / levonorgestrel	Birth Control	8	December 8, 2016	1
12	29607	Topiramate	Migraine Prevention	9	January 1, 2015	19
13	75612	L-methylfolate	Depression	10	March 9, 2017	54
14	191290	Pentasa	Crohn's Disease	4	July 6, 2013	8
15	221320	Dextromethorphan	Cough	4	September 7, 2017	1
16	98494	Nexplanon	Birth Control	3	August 7, 2014	10
17	81890	Liraglutide	Obesity	9	January 19, 2017	20
18	48188	Trimethoprim	Urinary Tract Infection	9	September 22, 2017	0
19	219869	Amitriptyline	ibromyalgia	9	March 15, 2017	39
20	212077	Lamotrigine	Bipolar Disorde	10	November 9, 2014	18
21	119705	Nilotinib	Chronic Myelogenous Leukemia	10	September 1, 2015	11
22	12372	Atripla	HIV Infection	8	July 9, 2010	11
23	231466	Trazodone	Insomnia	10	April 3, 2016	43

Figure 1

The first line is the header of the file, which consists of six attributes, as listed in Table 1.

code:	A unique code assigned to the drug
drugName	Name of the drug
condition	Name of condition
rating	10 start patient rating
date	The date of the review created
usefulCount	Number of users who found the review was useful

Table 1

Within each record, attributes are separated using "***Tab***" characters, i.e., the file is in a Tab-Separated Values format.

Your task is to modify the provided merge sort so that the modified merge sort can sort the records in an ascending order based on the '***rating***', however, if records have the same rating scores, then they should be sorted based on the '***usefulCount'***. A sample of sorted records is shown in Figure 2. When reading the file to your program, you should consider removing the first line.

	A	B	C	D	E	F
1	code	drugName	condition	rating	date	usefulCount
2	230728	Camphor / m	Pain	1	August 25, 2017	0
3	107325	Implanon	Birth Control	1	December 13, 2011	0
1366	103586	Synthroid	Underactive	1	December 6, 2017	1
1367	207071	Emollients	Dry Skin	1	October 30, 2017	1
161290	126311	Viibryd	Depression	10	September 7, 2011	693
161291	169147	Vilazodone	Depression	10	September 7, 2011	693
161292	107655	Implanon	Birth Control	10	July 19, 2010	730
161293	52305	Adipex-P	Weight Loss	10	October 19, 2008	796
161294	139141	Phentermine	Weight Loss	10	October 19, 2008	796
161295	182560	Mirena	Birth Control	10	April 1, 2009	1247
161296	131116	Levonorgestr	Birth Control	10	April 1, 2009	1247
161297	96616	Sertraline	Depression	10	July 31, 2008	1291
161298	119152	Zoloft	Depression	10	July 31, 2008	1291
161299						

Figure 2

Task 3: Saving results

After sorting the records, you need to store the sorted results (including the header) to a file "***Lab3_DrugsCom_Results.tsv***".