# Merging intervals problem

**Interval (definition)** An interval is a time period with a start and an end. The start and end values are included within the intervals, i.e. [4,9] means the interval includes 4,5,6,7,8,9 i.e. 4 and 9 are included here.

**Merge distance (definition)** is a value, when combined with two separate intervals allows them to be merged if they overlap across the merge disance. Examples

* Given two intervals [1,5] and [10,15] and a merge distance of 5, the two intervals overlap across this merge distance allowing them to be merged to a new interval of [1,15].
* Similarly given two intervals [1,5] and [11,15] and a merge distance of 5, you cannot merge these two intervals since they do not overlap across the merge distance.

**Problem** Given a list of intervals [start, end] you have to merge them based on a specified **merge distance**. Your intervals are arriving in a particular order, as they arrive you should merge them according to the specified **merge distance** as each interval is received. Some of these intervals will be removed[[1]](#footnote-1) (in the arrival stream they will be marked as removed) – in that situation you should treat the original interval as if it never existed. Example:

Your **merge distance is 7** – assume in the following example the input is arriving in order

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sequence** | **Start** | **End** | **Action** | **Output** |
| 1 | 1 | 20 | ADDED | [1,20] |
| 2 | 55 | 58 | ADDED | [1,20] [55,58] |
| 3 | 60 | 89 | ADDED | [1,20] [55,89] |
| 4 | 15 | 31 | ADDED | [1,31] [55,89] |
| 5 | 10 | 15 | ADDED | [1,31] [55,89] |
| 6 | 1 | 20 | REMOVED | [10,31] [55,89] |

Write a program in a language of your choice that you can demonstrate on your laptop. The input to your program should be a file and a merge interval. The file should have the following format: Arrival time, Start, End, Action – you can assume the file is sorted by arrival time. The Action column is a string value which is either “ADDED” or “REMOVED” (for “DELETED” see the extra credit question), which tells you that your interval was added or removed.

Removed means that the original interval was removed from the input stream. In the example above, when Removed is called on the action in sequence 6, it is actually removing the interval that showed up in sequence 1.

You should also have a test input file that demonstrates all the edge cases. The program should output the merged intervals on the console as every row is read from the file.

What are some of the challenges with this problem as your input grows – how would you address them?

## Extra credit

You can delete “parts” of an interval. Delete is different than remove – while remove will remove an intervals from the original stream, delete will delete an interval block from the current set of merged intervals. So, you could have:

1. 1,6, added OUTPUT: [1,6]
2. 5,7, added OUTPUT: [1,7]
3. 2,3, deleted OUTPUT: [1,2] [3,7]

Build a modified solution to incorporate this change.

1. Removed is different than deleted. I.e. remove means that the original interval is now removed. Delete means that that interval is deleted from the overall merged interval. [↑](#footnote-ref-1)