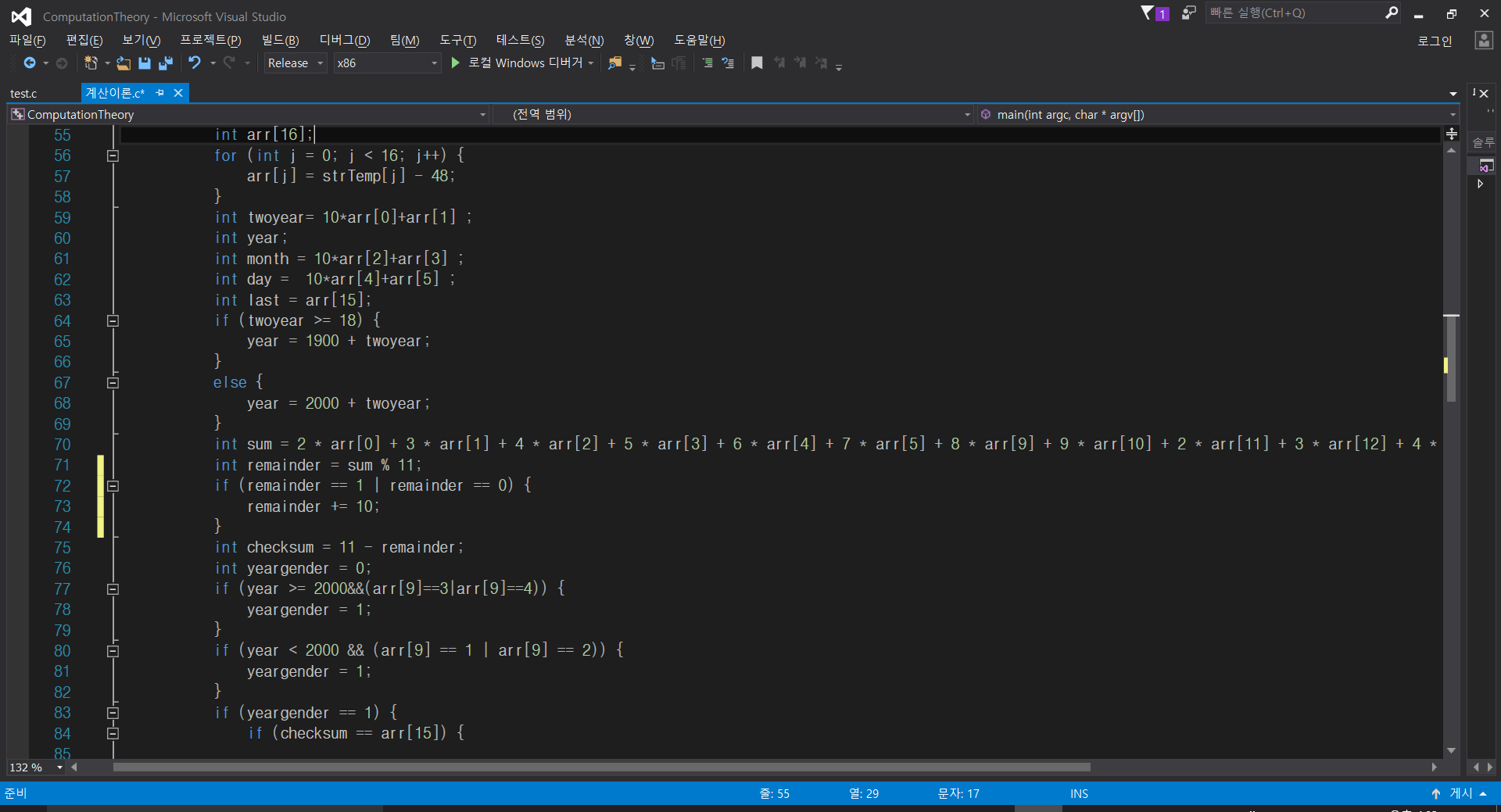
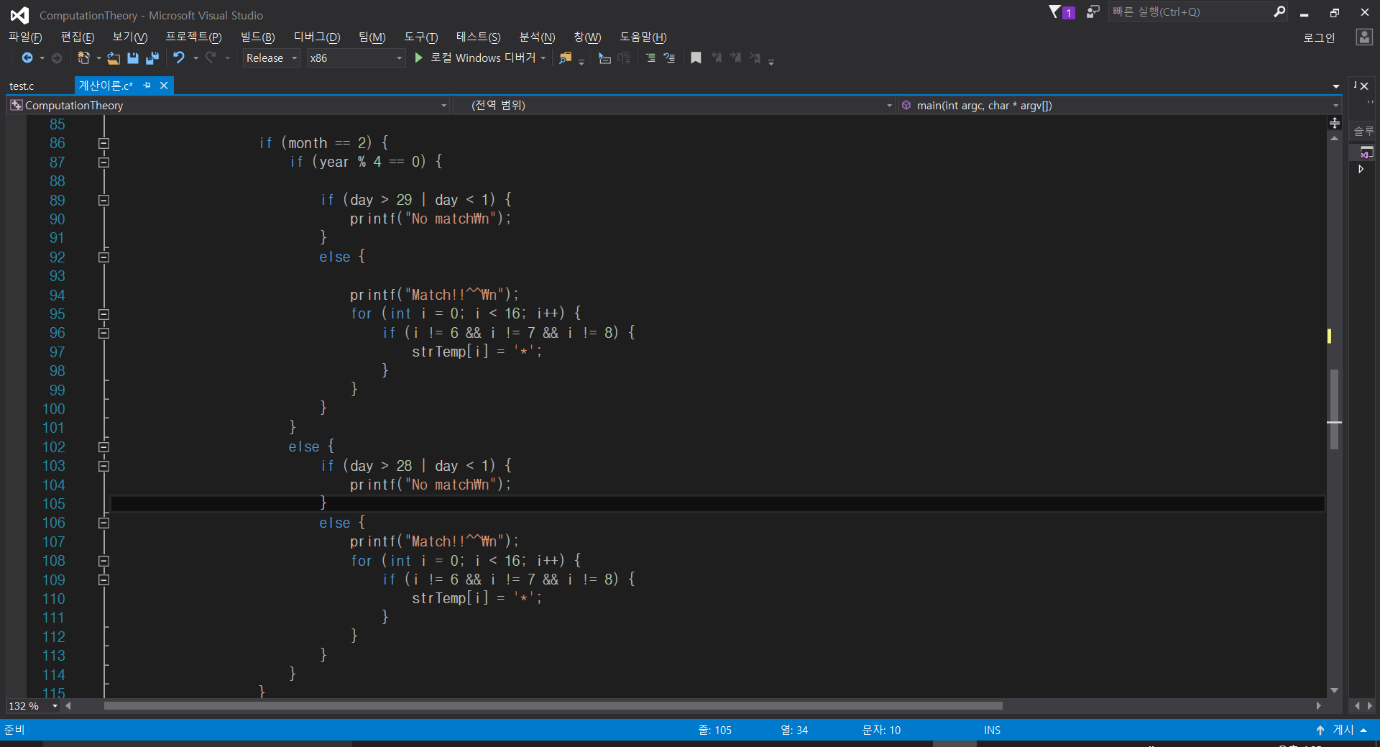
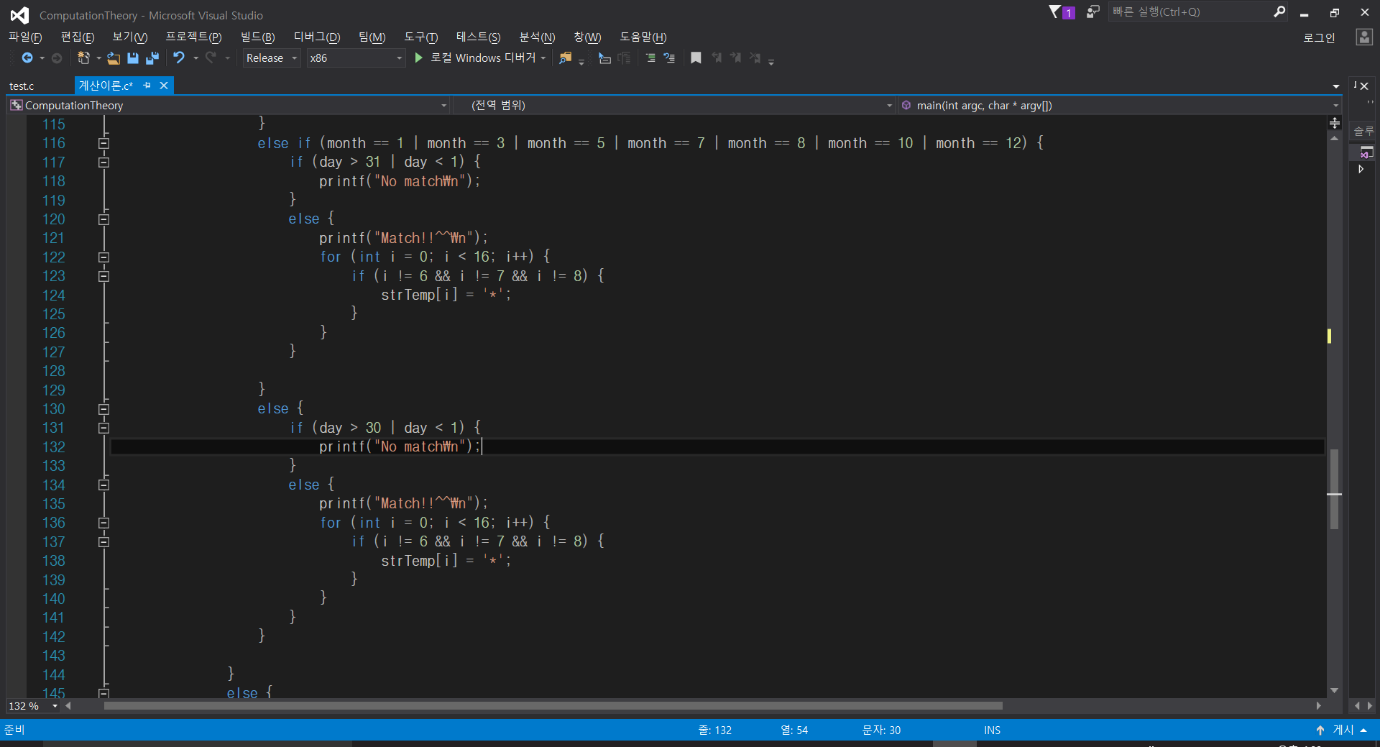
Theory of Computation Report

2016320152 전지연

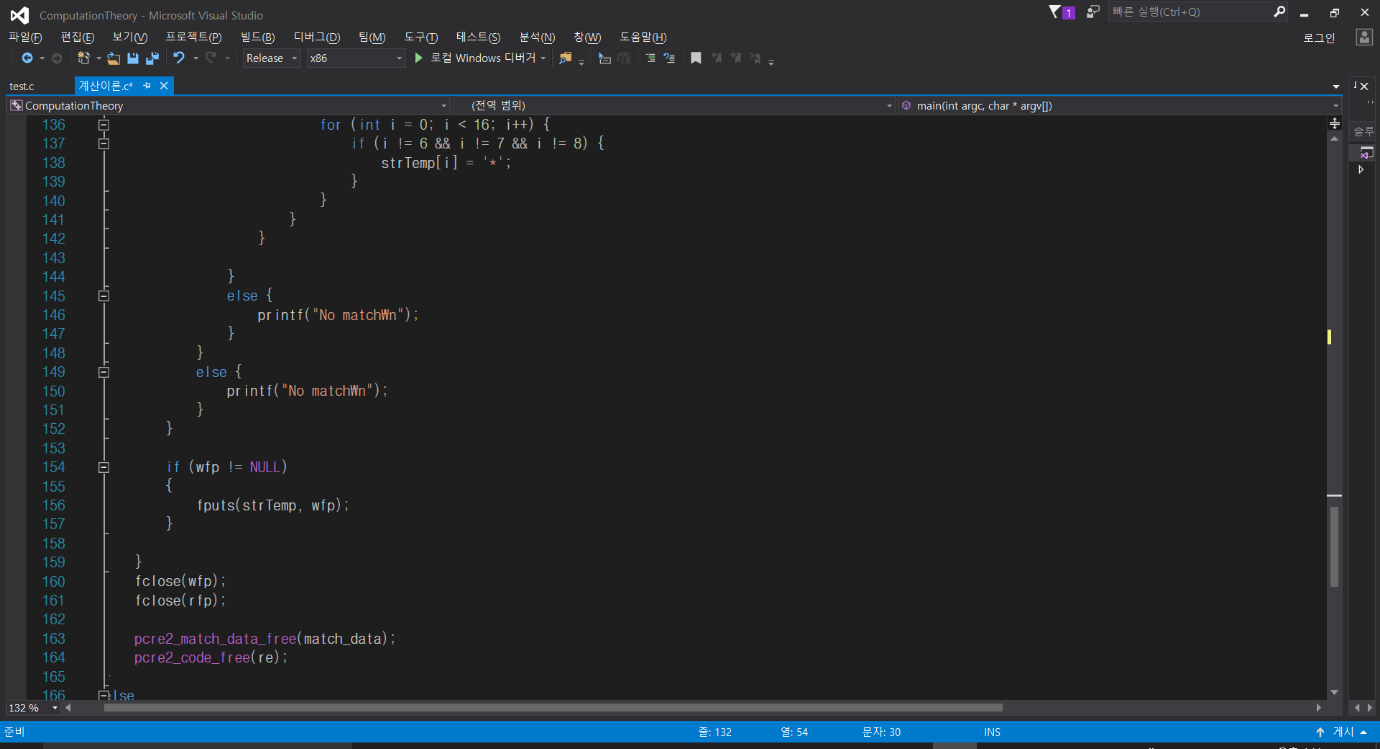
My program receives input.txt file via command line, and checks if the input has a valid social security number through regular expression and additional routine, and then change that valid social security number to ‘\*’ string. First, through regular expression([0-9]{2}(0[1-9]|1[0-2])(0[1-9]|[1,2][0-9]|3[0,1])\\s-\\s[1-4][0-9]{6}), it checks input has valid form of security number. For example, if input is “hello world!”, it is not valid form of social security number, so my program prints “No match!”. The part of regular expression “[0-9]{2}” says year and the part “(0[1-9]|1[0-2])” says month from January to December and the part “(0[1-9]|[1,2][0-9]|3[0,1])”says Day from first to 31th and the part “\\s-\\s” says space and ‘-‘ and space, and the part “[1-4][0-9]{6}”says a number from 1 to 4 and six numbers from 0 to 9. So if input is “971389 – 9123456”(invalid month, invalid day, invalid first number after ‘-‘) or “971230-1234567”(there is not space before and after ‘-‘) is not valid. However, there are some matters. First, there is difference of number of days for months.(달마다 일의 개수가 다르다) For example, January has 31 days but June has 30 days. Moreover, February has different days whether the year is leap or common, if the year is leap year, there are 29days else there are 28 days.(게다가, 2월은 윤년인가 평년인가에 따라서 가지는 날의 수가 다르다, 윤년이면 29일까지 있고 평년이면 28일까지 있다.) Lastly, we must check the last number of security number(checksum value) and the number meaning gender(if year is larger than 2000, the gender number must be 3 or 4, otherwise, the number must be 1 or 2)(주민등록번호 뒤 7자리의 첫 숫자인 성별을 나타내는 숫자는 2000년 이후 태어난 사람이면 3혹은 4이고 1900년 이후 태어난 사람이면 1혹은 2이다). So I added some codes for check these matters. (그래서 나는 앞에서 말한 이런 문제들을 체크하기 위한 코드를 추가했다.) If the program checks input through regular expression, then it checks input’s validity through additional codes.



At line 55-58, I changed character type input array to integer type by deducting 48. (문자형의 인풋 배열을 각각의 요소에 48을 뺌으로써 정수형의 배열로 바꾸어 주었다.) And then at line 59,61,62, I changed each integer to value by calculating(ex)10\*arr[0]+arr[1]) and substituted that value to each variance(twoyear,month,day).(각각의 정수들을 두개씩 모아서 10진수 계산으로 연도 값, 달 값, 일 값으로 바꿔서 각각의 변수들에 대입했다.) The range of year is from 1918 to 2017, so if the value of variance “twoyear” is larger than 18, at line 64-69, I substituted the value that twoyear plus 1900 to variance “year” otherwise, I substituted the value twoyear plus 2000 to variance “year”. (twoyear 에 있는 값이 18 이상이면 year 라는 변수에 twoyear에 1900을 더한 값을 대입했고 그렇지 않으면 year 변수에 twoyear에 2000을 더한 값을 대입했다. ) And then, at line 70-75, I checked checksum by calculating sum and remainder. Line 71-74 is for the case that remainder is 0 or 1 so checksum is larger than10.(71-74줄은 remainder가 0또는 1이라서 checksum이 10이상인 경우를 위해 있다. remainder가 0이나 1이면 remainder에 더하기 10을 해주어 checksum에서 remainder를 뺐을 때 10이상인 경우가 없게 한다.) At line 76-82, I declared variance “yeargender” that is for checking the number of gender(first value is 0)(성별을 나타내는 주민등록번호의 숫자를 확인하기 위해 yeargender 라는 변수를 0으로 초기화해서 선언했다). When the year is larger than 2000, if the value of arr[9] is 3 or 4, the value of “yeargender” will be 1.(년도가 2000년 이상일때, arr[9]값, 즉 성별을 나타내는 수가 3또는 4이면 yeargender의 값은 1이된다.) When the year is not large than 2000, (1918-1999) if the value of arr[9] is 1 or 2, the value of “yeargender” will be 1.(년도가 2000보다 작을 때, 성별을 의미하는 숫자가 1 또는 2이면 yeargender의 값은 1이된다.) If the value of “yeargender” is 1, then input has valid gender number, so the program implements next codes. (yeargender값이 1이면 입력은 유효한 성별을 나타내는 숫자를 가지고 있으므로 다음 코드를 실행한다.) At line 84, the validity of checksum is checked. (84번 줄에서 checksum의 유효성이 확인된다.)

After checking the validity of checksum, it checks if the month is 2 or not(line 86). If the month is 2, February, it checks wheter the year is leap year or not.(line 87). If the year is leap year, condition for day is 0<days<30 (윤년이면 일은 1부터 29까지이다.) and checks whether it has valid day or not. If input is valid, the program prints “Match!!^^\n” and change only the numbers to ‘\*’(line92-97).(입력이 유효한 주민등록번호이면 “Match!!^^”를 출력하고 공백이나 -를 제외한 숫자만 \*으로 바꾼다.)

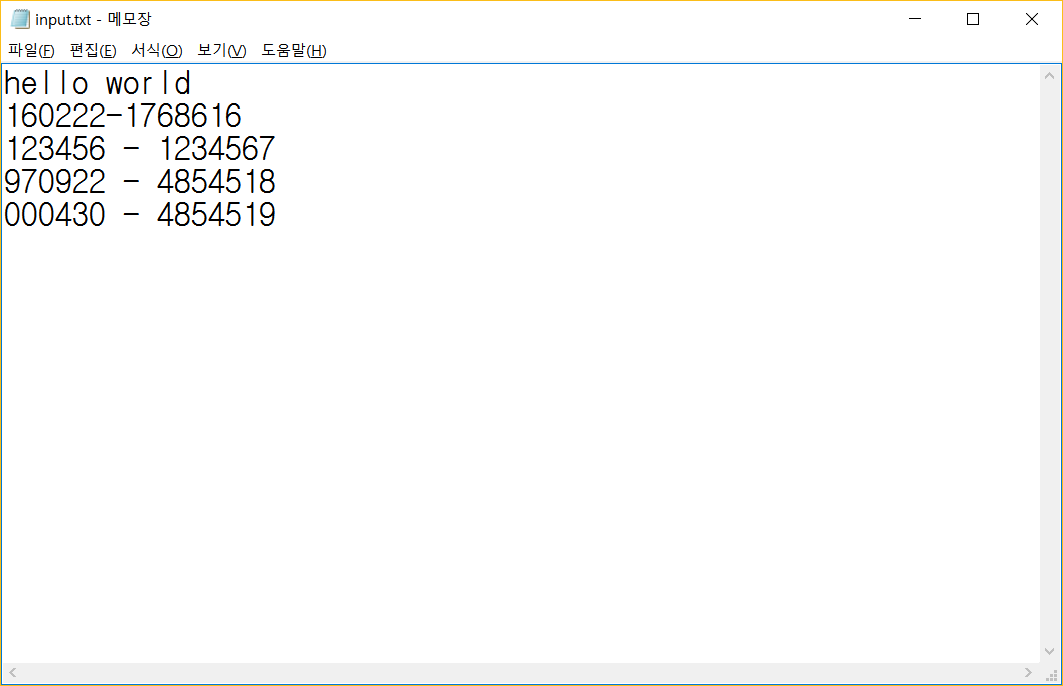
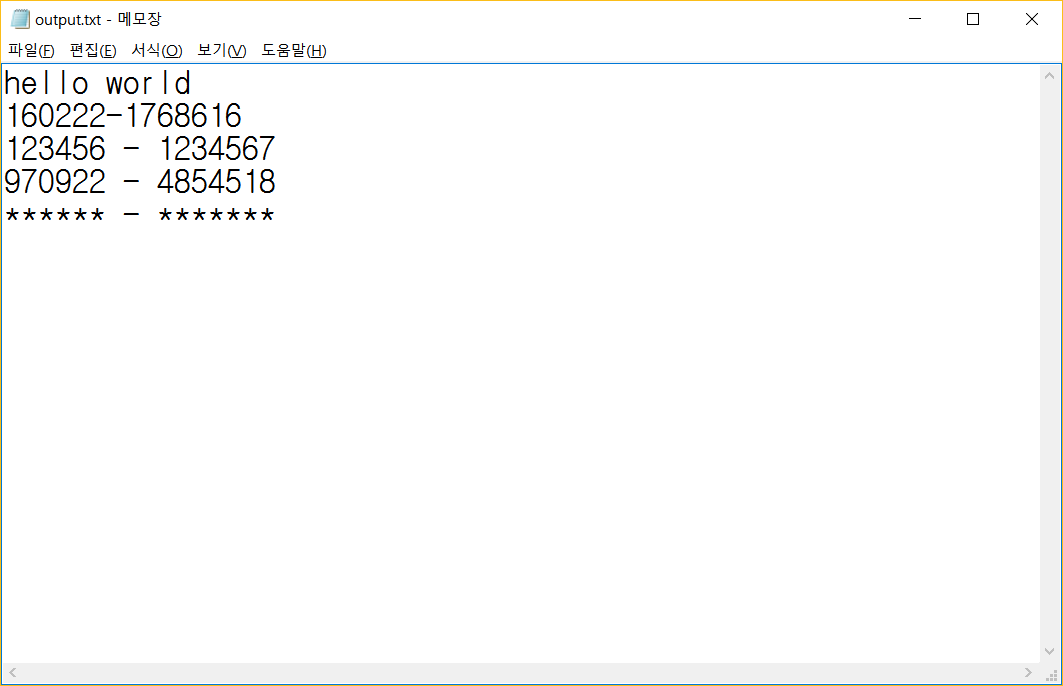
If the month is January or March or May or July or August or October or December, it has 31 days so the condition of day is 0<day<32 and checks whether it has valid day or not(line 116-117). The next step is the same with upper code. (다음단계는 위에서 말한 코드와 같다. 유효하면 match 출력하고 숫자 \*로 바꾸고 유효하지 않으면 No match만 출력.(line 118-124)

For the other months, the condition of day is 0<day<31 because these months have 30 days. Also, next step is the same with upper code. (line 130-138)

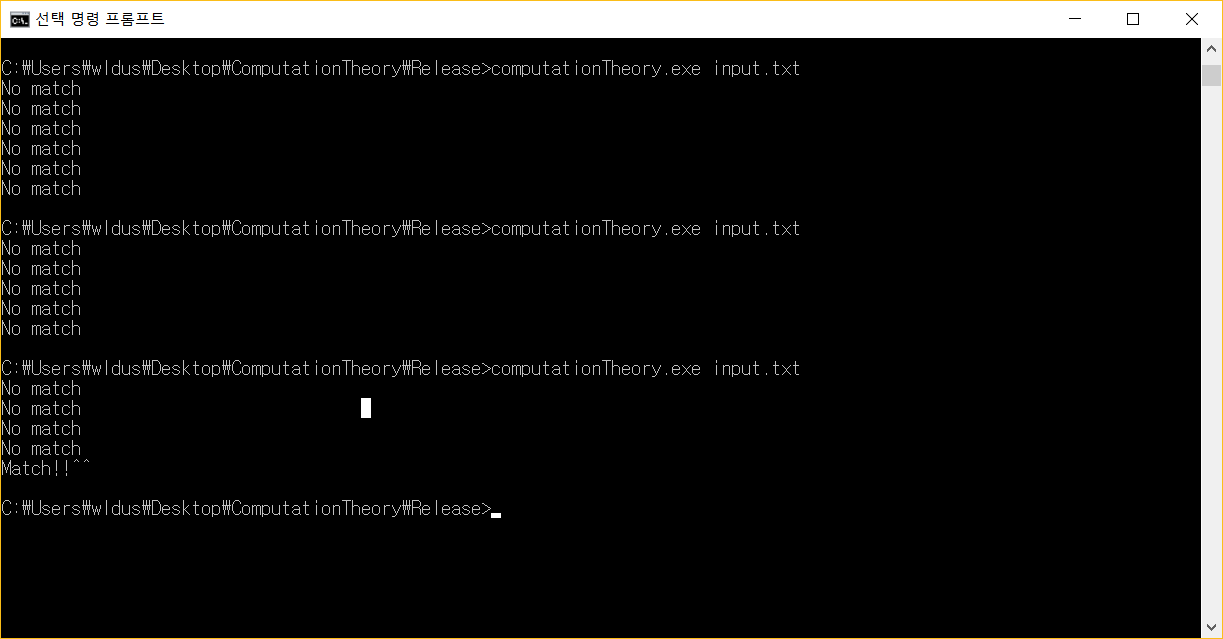
At line 146 and 150, it prints “No match” for the input having invalid checksum and having invalid gender number.(146번 줄에서는 유효하지 않은 checksum을 가진 인풋에 대해 no match를 출력, 150번 줄에서는 유효하지 않은 성별을 나타내는 숫자를 가진 인풋에 대해 no match를 출력한다.

The progress of this program is making integer array, checking gender number, checking checksum, checking February and whether the year is leap year, checking day and checking month having 31days ,checking day and checking the other month, checking day. For invalid input, it prints out “No match” and output.txt file is not changes. For valid input, it prints out “Match!!^^” and output.txt file is changed, only numbers of input is changed to ‘\*’.(프로그램의 진행과정은 인풋으로 정수형 배열을 만들고 그로부터 checksum을 만들고 유효한 성별숫자인지 확인하고 유효하면 checksum을 확인하고 유효하면 2월인지 확인후 2월이면 윤년을 확인해서 그에 맞는 날짜를 확인하고 31일을 가지는 달을 확인해서 가지는 달에 대해 유효한 날짜를 가지는지 확인하고 나머지 달에 대해 (30일가지는 달) 유효한 날짜를 가지는지 확인하고 날짜까지 모두 확인한 유효한 주민등록번호는 숫자를 \*로 바꾸는 것 이다. 그리고 유효하지 않은 인풋에 대해서는 No match를 출력하고 output 파일의 그 인풋에 대해서는 변화가 없는데 유효한 인풋에 대해서는 Match를 출력하고 output파일의 그 인풋의 숫자만 \*로 바뀌는 프로그램이다.)

[Input.txt] [Output]

[cmd]



I think the application of regualr expression on real world is very efficient because if in my program, I did not use regular expression, I had to add condition of year and month then the code will be longer and more complicated. So, by using regular expression, I could check invalid input form (ex. “Hello world” or “123456 – 1234567”) On real world, (현실세계에서도) if the regular expression is used, program will be more efficient.