Mandatory Exercise 1 INF240:

Crack this ciphertext that has been encrypted with the vigenere cipher:

ocwyikoooniwugpmxwktzdwgtssayjzwyemdlbnqaaavsuwdvbrflauplo
oubfgqhgcscmgzlatoedcsdeidpbhtmuovpiekifpimfnoamvlpqfxejsm
xmpgkccaykwfzpyuavtelwhrhmwkbbvgtguvtefjlodfefkvpxsgrsorvg
tajbsauhzrzalkwuowhgedefnswmrciwcpaaavogpdnfpktdbalsisurln
psjyeatcuceesohhdarkhwotikbroqrdfmzghgucebvgwcdqxgpbgqwlpb
daylooqdmuhbdqgmyweuikmvswrvnqlszdmgaoqsakmlupsqforvtwvdfc
jzvgsoaoqsacjkbrsevbel

The vigenere cipher:

The vigenere cipher is a variation of the shift cipher.
Encryption is done by having a chosen key k with integers from 0 to 25, these are indices in the English alphabet. Indices corresponding to a word is often used, for example the word "monitor" becomes {12,13,14,8,19,13,17}.
We then use these indices to shift a letter ki spaces to the right, where 0<=i<keylength, we perform this shift on every letter in the plaintext, starting over from ki=0 when we're at the end of the keylength.
Because we perform this shift ciphered letters can have (keylength) variations, and this makes frequency analysis charts hard to apply directly on the ciphertext.

Breaking the vigenere cipher:

The question is now, how do we break it?
In vigenere cipher the keyword and keylength are secret, we need to find a way to obtain them.

Obtaining the keylength:

We start by trying to obtain the keylength, since from that it is easier to conjecture the keyword. A way of doing this is putting two copies of the ciphertext against eachother, but one of the texts which we will call the shifttext is shifted i times to the right.
Then we compare the characters at each position to see if they are the same. We count the number of characters in the same position that are equal, then we shift the shifttext 1 more space to the right and repeat the process.
The key length is the number of the shifts where we have the highest equal characters, if we now execute the method on the exercise's ciphertext we get:

# of shifts   : # of equal chars
Displaced 1: 13
Displcaed 2: 14
Displcaed 3: 13
Displaced 4: 9
Displaced 5: 17
Displaced 6: 24
Displaced 7: 12
Displaced 8: 9
Displaced 9: 15
Displaced 10:11

Key length: 6

6 shifts has the highest number of equal chars and we guess that this is the keylength.

Obtaining the key:

Next is obtaining the key. Since we have a probable keylength we can use it to compute each key in the keylenght.
We do this by computing the dot products of W*Aj, where W is a vector containing the frequency probability of letters at positions i + keylength in the ciphertext where 0<=i<keylength, Aj is the vector containing the frequency analysis for the English alphabet shifted by j spaces to the right and 0<=j<=25.

After we've obtained all the dot products we compare and see which dot product is the largest, the j with the highest dot product is most likely the key. We then increase i and repeat the process to find the next key.
If we apply this procedure on the ciphertext with the keylength we found earlier we get the decryption key vector: {19,12,15,14,22,8}
With this, we can find the last secret, keyword. Let every key in the decryption key vector be negative and reduce by modulo 26.
For example first key will be -19 mod 26 = 7 = h
Repeat and we get the keyword «holmes»

Deciphering the text:
When we've found all the keys we try and apply the decryption key vector on the ciphertext and see if the result makes sense, if it does we have succesfully cracked the ciphertext.
This is the plaintext we get after deciphering the ciphertext (given asis without formatting for readability):

keylength: 6
decryption key vector: {19,12,15,14,22,8}
keyword: «holmes»

```
«holmeshadbeenseatedforsomehoursinsilencewithhislongthinbackcurvedoverachemicalvesselin
whichhewasbrewingaparticularlymalodorousproducthisheadwassunkuponhisbreastandhelookedfr
ommypointofviewlikeastrangelankbirdwithdullgreyplumageandablacktopknotsowatsonsaidhesud
denlyyoudonotproposetoinvestinsouthafricansecuritiesfhhkndgcagvlfspcmatwbzqxlcucndmikrb
kclkuowtafgwkcwqfomontz»
```

It seems it is correct, because almost all of the text is sensible except for a strip at the end. This might be because the ciphertext given in the book and online are different, we have used the ciphertext in the book. There is a possibility of human error here since it was transcribed from the book by hand.

End notes:
Attached is the java program created to crack the given ciphertext, it uses java 1.6 and contains some more detail on how it works in the comments and javadoc.