Text Analysis Tools

Word Frequency Table

Task
To build a tool to produce a table of word frequencies in a given range (i.e. the entire text <WORK>, or by act <ACT>, scene <SCENE>, or character <SP>). This tool should work on **both modern spelling texts and diplomatic transcriptions**.

Usage Examples
User wants to know the most/least frequent words used in the play.
User wants to know most frequent words used by a particular character.

Optional Additional Task
To build in the option to show results in context. For example, if I click on the word "love", I am shown a list of all of the instances of the word in context (e.g. 5 words before and after "love").

Letter Frequency Table

Task
To build a tool to produce a table of letter (and letter-form) frequencies in the entire text or in a given range (i.e. page range <PAGE> or signature range <SIG>). This tool should be designed to work on the **diplomatic transcriptions only**.

Usage Examples
User wants to know which letters are used between signatures A3r and C2v.
User wants to know how many times the long-s-s-i ligature and the swash capital R are used (tagged {{s}{s}i} and {*R} or {R*} respectively) throughout the entire text.

Optional Additional Task
To build in the option to show results in context. For example, if I click on "æ" (a digraph, tagged as {ae}), I am shown a list of all of the instances of the letter-form in context (e.g. 5 words before and after the occurrence of the letter-form "æ").

Word and Letter Clouds

Task
To build a tool to produce word and letter cloud visualisations of the frequency data generated by the tools described above. The size of the word/letter should be proportionate to its relative frequency (i.e. more frequent words are larger, less frequent words smaller, etc). Think Wordle.net, but feel free to be creative.

Usage Examples
A user wants to visualise the relative frequency of words in an entire play, or used by a particular character.

Optional Additional Task
To build in the option of being able to mouse-over the word/letter to reveal its frequency count.

Doubling Chart

Task
To produce a chart illustrating a possible distribution of parts (characters). The chart should be arranged by act/scene on the *x*-axis and by part/character on the *y*-axis. X-Y values should give the number of lines spoken by a character at a particular point in the play (with a dash to indicate that they're not on stage, and a 0 value to indicate that they are on stage but do not speak). This tool should work with the **modern-spelling text only**, because there will not be any ambiguities about speakers and stage directions.

Here is an example of part of a doubling chart for Shakespeare's *Hamlet*:

Table 61
Hamlet (F)

Principal Parts — Men

	1					2		3				4							5		Totals
	1	2	3	4	5	1	2	1	2	3	4	1	2	3	4	5	6	7	1	2	
#1 Hamlet	-	103	-	44	97	-	149	76	227	22	148	-	20	21	-	-	-	-	130	203	1,240
#2 Claudius-King	-	95	-	-	-	-	37	39	7	50	-	32	-	45	-	67	-	113	14	30	529
#3 Polonius	-	2	68	-	-	76	149	23	13	9	8	-	-	-	-	-	-	-	-	-	348
Clowne	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	71	-	71
#4 Horatio	83	51	-	22	14	-	-	-	9	-	-	-	-	-	-	12	25	-	11	50	277
#5 Laertes	-	7	52	-	-	-	-	-	-	-	-	-	-	-	-	49	-	42	18	38	206
1 Player-King	-	-	-	-	-	-	48	-	48	-	-	-	-	-	-	-	-	-	-	-	96
#6 Rosincrane	-	-	-	-	-	-	46	12	11	14	-	0	8	5	-	-	-	-	-	-	96
Osricke	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	36	36
#7 Ghost	0	-	-	0	87	-	-	-	-	-	6	-	-	-	-	-	-	-	-	-	93
Fortinbras	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8	-	-	-	-	21	29
#8 Marcellus	45	2	-	8	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	60
Fellow-Lucianus	-	-	-	-	-	-	0	-	7	-	-	-	-	-	-	-	-	-	-	-	7
#9 Barnardo	28	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	28
Guildensterne	-	-	-	-	-	-	26	5	21	4	-	0	1	0	-	-	-	-	-	-	57

Boys

	1	2	3	4	5	1	2	1	2	3	4	1	2	3	4	5	6	7	1	2	Totals
#1 Gertrude-Queene	-	10	-	-	-	-	22	8	3	-	46	11	-	-	-	18	-	21	7	9	155
#2 Ophelia	-	0	20	-	-	37	-	33	18	-	-	-	-	-	-	29	-	-	-	-	137
#3 Boy Player-Queen	-	-	-	-	-	-	0	-	26	-	-	-	-	-	-	-	-	-	-	-	26

Total Lines Principal Parts 3,491

Things to note:
1. There were no female actors in Shakespeare's day, so boy actors performed the female characters. This is why the chart above splits the roles across boy actors and men. Don't bother with this for the tool – character genders have not been tagged in the texts, and users can figure it out for themselves.
2. The above chart has made a suggestion about the doubling of parts (e.g. that actor #7 plays the Ghost and Fortinbras). Don't worry about doing this – just producing a chart that gives the lines for each character by scene will be enough.

Usage Examples
A user wants to see which characters are on and off stage at a given point, and try to work out which parts might be doubled (i.e. played by the same actor).

(Very) Optional Additional Task
To build in the option to define charted roles as played by boy actors and mature actors, and to generate a series of possible doublings. This is a task in and of itself, so it's very much optional!

Prop Chart

Task
To build a tool to produce a chart listing all stage props in a play (tagged <PROP>) along the *x*-axis and their use in a given act/scene on the *y*-axis (with a dash – to indicate that it is not being used, and an "x" to indicate its use). This tool should work with the **modern-spelling text only.**

Usage Examples
A user wants to see which props are on and off stage at a given point.

Optional Additional Task
None.