As our data set consists of the works of Shakespeare, we have decided that the primary users would be students or researchers studying English or Drama. This could allow them to search through his works to find certain quotes, characters, stage directions and titles of the play. It will also help the students/researchers analyse the main feelings and moods of the play.

Possible Queries:

* A user could search for the phrase “to be or not to be”. Our system will return the title of the play this occurs in (“Hamlet”) and the scene(s) and line(s) on which it occurs.
* A user could search for “love” or “hate” to see how many times the words appear in a specific play, in order to get the leading emotion of that play.
* A user could search for a character such as “Romeo” to see in which play(s) and scene(s) they appear.

Our system will be able to support the different Boolean operators (\*, +, -) and the actual Boolean words (AND, OR, NOT) when searching for results.

All fields (title, full text, stage directions etc.) can be searched and we will allow the user to filter which fields of the files they are searching.

Standard stemming practices will be followed i.e. removing ‘ing’, ‘ed’ etc. from the end of words. Lucene’s built-in implementation of the Porter stemming algorithm will be used to achieve this.

Because of the nature of the system and the style of Shakespearean English, our system will not make use of stopwords. This decision was made after realising that with the standard stopwords, the famous quote “to be or not to be” would not appear in the search results.

The result of the query will be displayed as a list of results on the left-hand side of the screen, showing the title of the play; the act/scene it occurs in and the line of searched text. If the user clicks on an item in the list, the entire text document is displayed on the right hand side at the relevant point, with the search query highlighted.

For a task of this nature, relevance feedback is not necessary.

Character names and play titles will be presented as autocomplete suggestions as the user types their query.

Additional Features:

* History – allow users to select previous queries to speed up search.