**Προεπεξεργασία**

Για την καλύτερη απόδοση του Μοντέλου Διανυσματικού Χώρου προχωρήσαμε στην προεπεξεργασία των κειμένων της συλλογής και των ερωτημάτων. Οι παρεμβάσεις μας ήταν οι εξής:

* Αφαίρεση των stopwords
* Stemming

Για την αφαίρεση των stopwords χρησιμοποιήσαμε την λίστα που παρέχει η βιβλιοθήκη ntlk. Διατρέχοντας κάθε κείμενο και κάθε ερώτημα, διατηρήσαμε μόνο τις λέξεις που δεν περιέχονταν σε αυτή τη λίστα. Για την εφαρμογή του stemming χρησιμοποιήθηκε ο Snowball Stemmer από τη βιβλιοθήκη nltk. Ο συγκεκριμένος αλγόριθμος επιλέχθηκε γιατί πρόκειται για μια βελτιωμένη έκδοση του Porter Stemmer, ο οποίος χρησιμοποιείται πολύ συχνά στην ανάκτηση πληροφορίας και είναι υπολογιστικά φθηνός και αλγοριθμικά απλός.