AI Phase 2 of the assignment.

Dataset Description: In the phase 1 of the assignment, a set of matches was provided. Now the dataset for this assignment is the commentary section of those set of matches. Note that the commentary is not entirely free text, it has some structure and on the basis of the questions asked, you need to see what kind of data would you need to save or recover from that semi-structured commentary. [For some of the questions, temporal features like which ball, which over etc might be useful.]

The questions that have been asked admittedly don't have a specific grammar but follow a pattern - they would talk about match number etc. You should try and extract this information from the questions using some trivial techniques like a set of regular expressions, which will make the answers more relevant and accurate.

Basically there are two techniques to solve these questions, you should try with (i) and if the recall is too low, you can go ahead and try to answer the question using technique (ii). Here's a brief description of the two techniques; - these are more of intuitive notions and heuristics rather than proper techniques.

(i). You'll try to find sentences that match the question's unigrams the most; this can be extended to bigram and trigram, but note that you'd have to understand the trade off between precision and recall. Another thing that can be done is to look at the bold and italicized keywords and try to rank the relevant sentences according to that and return those sentences or relevant information from them.
(ii). Second technique is what is known as query expansion. If you cannot find the keyword anywhere in your dataset, you should try to expand the query - what it means is you should try to look for words that have same or similar meaning as the keyword and try to search for those words. Now, how do you find the related words - try to use **synsets** from nltk's ***Wordnet*** interface which is a large concept graph.

1. a) When was Ryder ***dismissed*** in match 1?
   b)How was Ryder ***dismissed*** in the first match ?
2. What was the ***best boundary*** of the match 2 ? [look for keywords like 'classy','fine' etc ...]
3. During what overs was Virat Kohli ***active*** the most in the match 4 ?
4. Which over was the most ***interesting over*** of the match 3 ?
5. a) When were ***powerplays*** signalled by umpire in match 4 ?
   b) Who were the bowlers ***immediately after powerplay*** started in match 4?
6. How is the ***weather*** on the day of the match 4 ? [look for weather keywords as well as some notion that conveys the idea that weather is being discussed of present day only.]