**Short-Answer Questions [25 points]:**

**1. Investigating one search engine (Google or Bing), and figure out what tricks they are using. (15%)**

**I have investigated all the queries on Google search engine.**

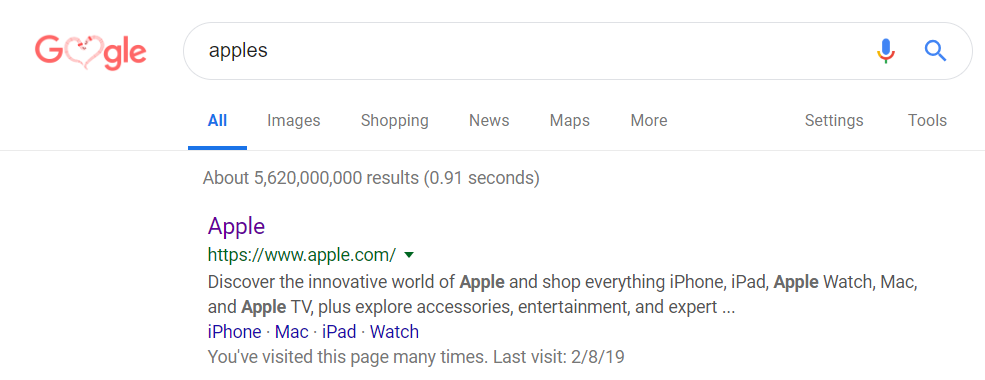
**(a) Does it use stemming? How do you know? (Explain your testing method in all the following sub-questions as well)**

i) Yes it uses stemming. Type for word **Apples** The first thing it returns is “Apple.com” here it stems the word “Apples” to its root word **Apple** and then returns the top results.

Testing mechanism: Entered the above mentioned word on Google Chrome and found the result as mentioned.

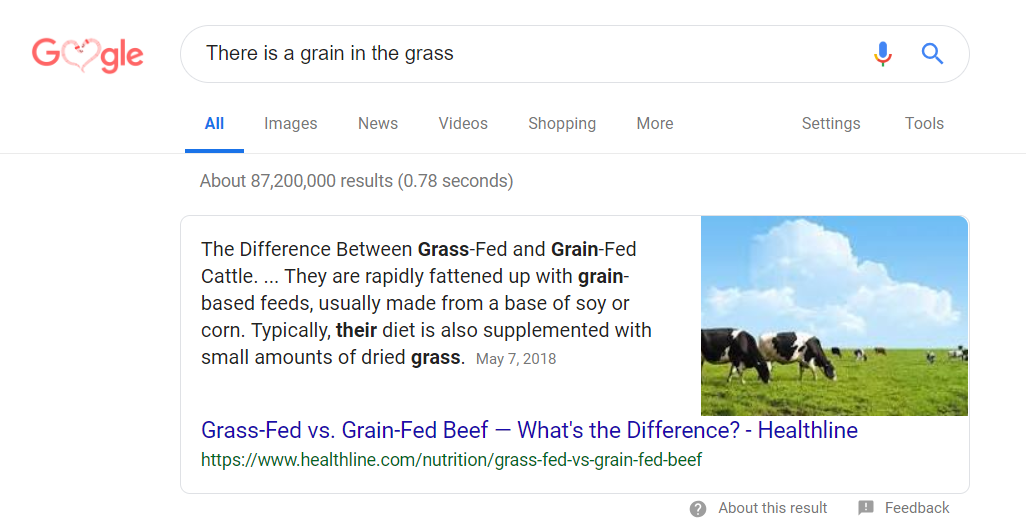
When entered “**Apples**” in double quotes it does not return the website “apple.com” this time.

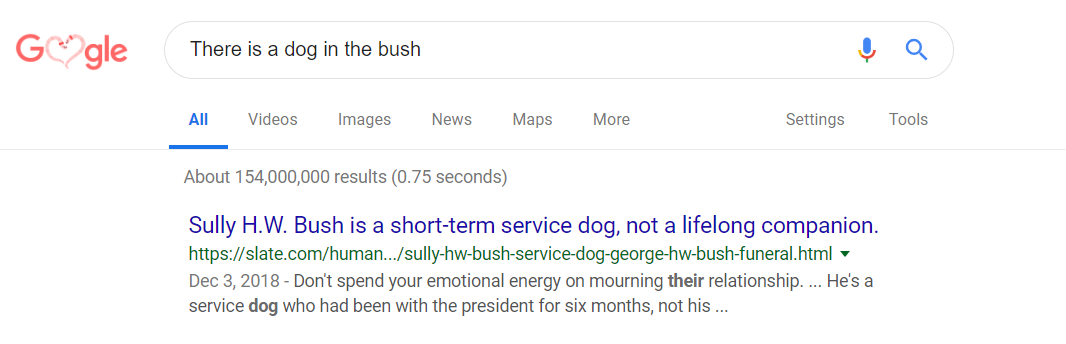
Experimented with the word “**processed”** as well, when used normally it performs stemming and finds documents containing the word “**process”** and not **“processed”** as well, but used double quotes it looks only for files which contain the exact word **processed.**



**(b) Does it filter stop words?**

Yes, it filter stop words. An example can be seen below where stop words like is “a” “the” “are” are filtered and query returned is focused on other words (non-stop words) like grass, grain.





**(c) Does it use spell correction?**

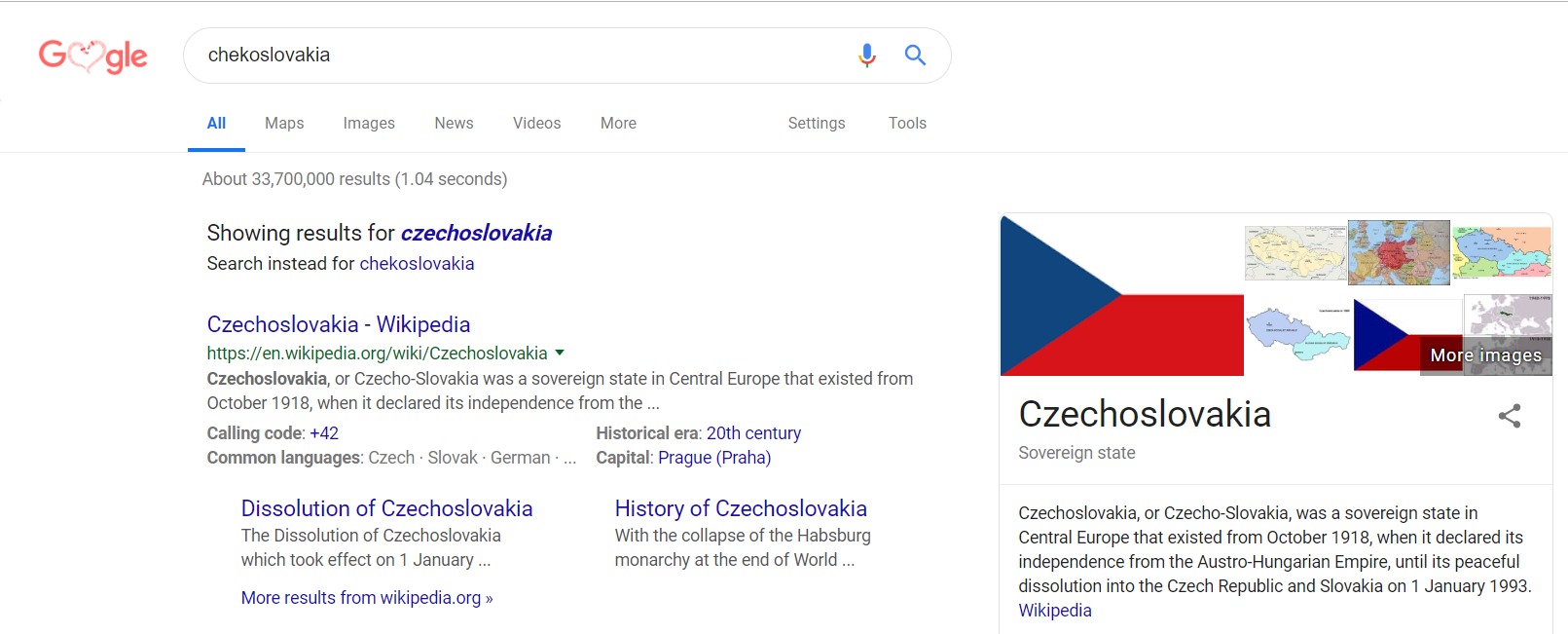
Yes it uses spell correction

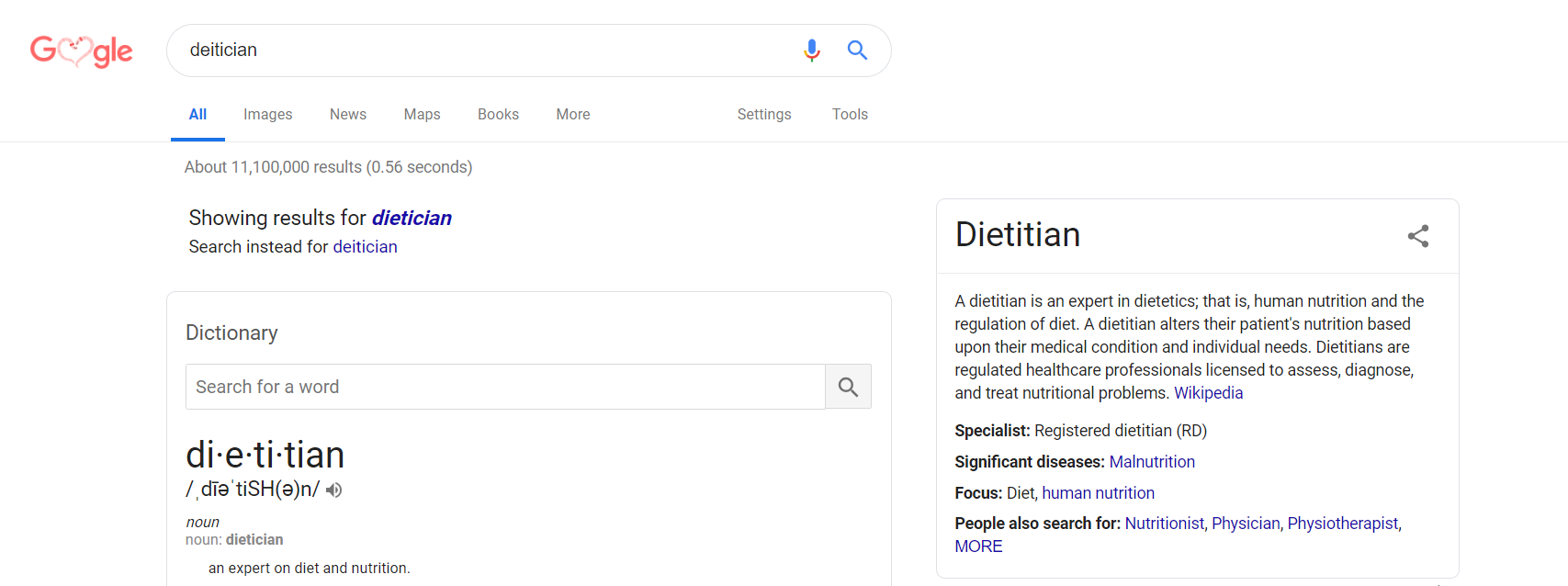
Typed in word: **chekoslovakia**, it returns the right intended word to be searched

Actual: **Czechoslovakia** it did perform spell correction and returned the top results for the word.

Testing mechanism: Entered the above mentioned word on Google Chrome and found the result as mentioned.

When entered “**chekoslovakia”** in double quotes it does not perform any spell check but returns results for the original query word entered.





**2. For a particular search query, your IR system returns 14 relevant documents and 16 irrelevant documents. There are a total of 80 relevant documents in the collection. (10%)**

**(a) What is the precision of the system on this search?**

14/30 =0.466

**(b) What is the recall of the system on this search?**

14/80 = 0.175