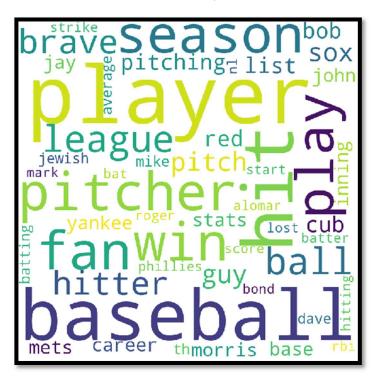
# Report Lab\_05 - Alessio Serra S277756

- # of clusters found: 4
- Clustering used: MiniBatchKMeans with the following parameters:
  (n\_clusters=4, init\_size=1024, batch\_size=2048, random\_state=20)

#### WordCloud created by the clusters:

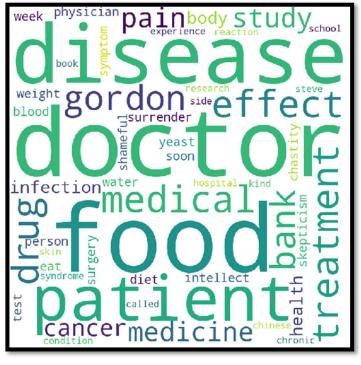


## **Cluster 1: Sport**

Looking at this picture, we can note that all the words belong to the world of sport.

In particular there are many specific terms of baseball.

So, we can conclude that the main topic of the first cluster is **Sport (Baseball in particular)**.



### **Cluster 2: Medicine**

Looking at this picture, we can note that all the words belong to different branches of *medical* science.

So, we can conclude that the main topic of the second cluster is **Medical Science**.



### **Cluster 3: Space**

Looking at this picture, we can note that all the words belong to technology word, in particular to aereospace science, with term such as *shuttle*, *orbit*, *launch*, *moon*, *nasa*, etc.

So, we can conclude that the main topic of the third cluster is the **Space**.



#### **Cluster 4: Crime**

Looking at this picture, we can note that all the words come from the word of crime.

Probably, it is because documents that are in this cluster are crime news.

So, we can conclude that the main topic of the fourth cluster is **Crime**.