Using Python and available libraries:

1. Write a script that extracts available a) **personal names** and b) **dates** from column K “newsdescp”. Hint: check out the Flair and Spacy libraries.

**Example:**

|  |  |  |
| --- | --- | --- |
| **Text** | **Names** | **Dates** |
| Cast member Mel Gibson (R) and Oksana Grigorieva attend the premiere of the film ''Edge of Darkness'' in Los Angeles January 26, 2010. Earlier this week, the agency's Patrick Whitesell informed the actors' representatives that he would no longer be represented by the agency. | Mel Gibson, Oksana Grigorieva, Patrick Whitesell | January 26, 2010 |

1. Apply text analytics to identify 5 other news articles that are similar to the following article. Consider methods such as cosine / pairwise similar to compute similarity or clustering algorithms such as k-means.
   1. “Human Rights Watch says government-controlled health services in Egypt have been pressured into playing down the number of casualties during anti-government protests. The group has documented the deaths of 297 people, but says the final toll is likely to be significantly higher. Human Rights Watch says the vast majority of the deaths in Cairo, Alexandria and Suez were on January 28 and 29 as a result of live gunfire as riot police fought running battles with protesters. A significant proportion came as a result of rubber bullets fired at too close a range and from teargas canisters fired into the crowds at very close range. Human Rights Watch says the actual number of deaths is likely to be an underestimate because the organisation had only included those deaths it had verified itself at key hospitals in the three major cities.”

You may share your scripts and other relevant information in any file format that is convenient. Good luck!