

Research Seminar Homework 2

Futamata Takumi

Dataset

Datasets come from 10 different articles about Tokyo 2020 Summer Olympics.

All these articles were written after the closing ceremony by global mass media including NY Times, CNN, BBC, Al Jazeera, France24, Tokyo Shimbun.

A Memorable Olympics, but for the Right Reasons?

A feat of organizational planning and execution in a pandemic left even gold medalists hurrying home from an atmosphere drained of pomp and full of anxiety.

TOKYO — The kaleidoscopic stands of Tokyo's Olympic Stadium — a disordered mosaic of white and green and reddish-brown seats — were designed to evoke the sun-dappled floor of a Japanese forest.

Over the past three weeks, though, they have had another, useful effect: Squint your eyes, blur them just so, and it almost appears as if spectators are arrayed around the sweeping structure, as if the 68,000-seat stadium is not essentially empty, as if these are a normal Olympic Games.

Of course, they were not. On Sunday night in Tokyo, a stripped-down closing ceremony in Japan's sprawling national stadium will bring these extraordinary Games to an end, concluding an Olympics that, in some sense, felt like an illusion — at times convincing and fully welcome, at others jarringly off-key.

Example: NY Times

Dealing with dataset

First of all, we should clean text in datasets with punctuations.

As a next step, we apply MapReduce approach to count words and get top 10 popular words.



Photo by Tokyo 2020 / Uta MUKUO

Top 10 popular words

We can first see Olympic-specific words such as "Olympic", "games" and "athletes".

Also we can find words related to the host country, Japan, such as "Tokyo" and "Japan" among the top 10 words.

More distinctive is the word "pandemic" which symbolizes the Tokyo 2020 Olympics as an event held during the pandemic of Covid-19.

	word	count
games	118	
tokyo	81	
olympic	80	
olympics	72	
athletes	56	
pandemic	45	
world	31	
chinese	30	
japan	30	
people	29	

Additional analysis: Word Cloud



Word Cloud is a diagram in which frequently occurring words are arranged like a cloud with a size proportional to their frequency.

It provides a more visual understanding of the frequency of words appearing in articles about the Tokyo Olympics.



How we can use the analysis results?

The data covered in this analysis were articles written after the Tokyo Olympics had come to a close.

Therefore, it is possible to intuitively understand the overall impression of the Games from the frequently occurring vocabulary.

This analysis can also be applied to articles on other topics, allowing for a quick and intuitive understanding of the content of the article through frequently-used vocabulary.



A complex network graph is displayed against a dark, textured background. The graph consists of numerous small, glowing blue and orange nodes connected by thin, translucent blue lines forming a hexagonal lattice. A bright, glowing white starburst or lens flare effect is positioned in the upper left quadrant, casting light rays across the network. The overall aesthetic is futuristic and represents concepts like connectivity, data flow, or artificial intelligence.

Thank you!