Analysis of Science and Technology Trend Based on Word Usage in Digitized Books JINHYUK YUN, Department of Physics, KAIST, South Korea, PAN-JUN KIM, Asia Pacific Center for Theoretical Physics, South Korea, HAWOONG JEONG, Department of Physics & KI for BioCentury, KAIST, South Korea — Throughout mankind's history, forecasting and predicting future has been a long-lasting interest to our society. Many fortune-tellers have tried to forecast the future by "divine" items. Sci-fi writers have also imagined what the future would look like. However most of them have been illogical and unscientific. Meanwhile, scientists have also attempted to discover future trend of science. Many researchers have used quantitative models to study how new ideas are used and spread. Besides the modeling works, in the early 21st century, the rise of data science has provided another prospect of forecasting future. However many studies have focused on very limited set of period or age, due to the limitations of dataset. Hence, many questions still remained unanswered. Fortunately, Google released a new dataset named "Google N-Gram Dataset." This dataset provides us with 5 million words worth of literature dating from 1520 to 2008, and this is nearly 4% of publications ever printed. With this new time-varying dataset, we studied the spread and development of technologies by searching "Science and Technology" related words from 1800 to 2000. By statistical analysis, some general scaling laws were discovered. And finally, we determined factors that strongly affect the lifecycle of a word.

Jinhyuk Yun
Department of Physics, KAIST, South Korea

Electronic form version 1.4