Which book should I pick? - Text Visualization based on readability and genre

- Readability Visualization

- Readability is the ease in which text can be read and understood and it directly affects difficulty of a book.

- The longer words, sentences and paragraphs have less readability. For the expression, we converted each character to a visible point of which the brightness decreases gradually during reading process.

- If the sentence has commas or a punctuation, the brightness of the points next to them is slightly increased.

- There is a clear difference between two books.
'My sister's keeper' has clear and concise sentence and paragraph style.

- But 'The Critique of Pure Reason' is extremely hard to understand even for well educated people not only because of the difficulty of contents but also because of the long and complex sentences and paragraphs.

- Next, Genre Visualization

- General Word Frequency Dictionary

- We collected 4,000 digital books to make the 'General Word Frequency Dictionary'. It contains frequency and rank of each word in the 4,000 books.

- More than half million words are in this dictionary, but the 20% of high ranked words covers more than 98% of whole words in use.

- Genre Word Frequency Dictionary

- As same way of making 'General Word Frequency Dictionary' we made multiple 'Genre Word Frequency Dictionaries'.
Some books represent 4 genres - Fantasy, Philosophy, Custom, Science Fiction- were selected.

- Then we make 'Genre Word Frequency Dictionary' with these 15 books.

- The 'Genre Word Frequency Dictionary' has to be refined to represent a genre identity. So We designed an algorithm to compare 'Genre Word Frequency Dictionary' with the 'General Word Frequency Dictionary'.

- Then measured a rank difference of a same word between two dictionaries.
If a word has higher rank in the 'Genre Word Frequency Dictionary', We recorded the word into a database with a rank distance.

- We call these words as 'Genre Identity Words' and the database as 'Genre Identity Dictionary'. Each genre has a special set of extraordinary words.
- Now we can calculate the 'Genre Closeness' with comparing to each 'Genre Identity Dictionary'. For the experiment, we picked 'Harry Potter book 3 - Harry Potter and the Prisoner of Azkaban- the representative fantasy novel and analyzed its 'Words Dictionary'.

- Then we made the algorithm to find the same words and to connect the two same words with a line. The line thickness and transparency are proportional to distance of each line. Therefore, the closer genre has more bright visualization image.

- The most complex and bright image implies the closest genre of the subject book, so users are able to instantly describe the genre closeness.

- The book 'Solaris' has high closeness figure not only in the science fiction but also philosophy. It means this genre closeness visualization can the multi-genre of the books.

- Combine visualization of readability and genre

a) Define genre-hue: Fantasy, Philosophy, Science Fiction, Custom genre.

b) Set the genre-hue as base hue in the readability visualization with hue rotation.

c) Make a lengthy paragraph colored with de-saturated tone due to the saturation setup function in the algorithm.