saya ingin membuat sistem koreksi ejaan pada dokumen text jurnal menggunakan react dan express. tapi saya punya kendala pada saat proses preprocessing text. dalam jurnal biasa terdapat angka, simbol-simbol, link, email, script, alamat email ilmifaian@gmail.com rumus prsamaan, dan lain sebagainya. nah saat pengeekan ejaan saya menggunakan dictionary lokkup untuk menentukan kata yang termasukk dalam kata dengan ejaan yang salah.. tetapi angka simbol-simbol dan lain sebagainya yang saya sebukan tadi juga dianggap sebagai kata yang salah karena tidak terdappat dalam basis data kata. bagaimana caranya agar sistem dapat memfilter bahwa pada saat proses preprocessing text itu hanya memfilter kata

dengan link sebagai berikut: http://ppid.uho.ac.id

1. Kamu;
2. Dia;
3. Apakah;
4. Dan lain sebagainya.

ChatGPT

Untuk menangani masalah tersebut, Anda dapat melakukan beberapa langkah dalam proses preprocessing teks sebelum melakukan koreksi ejaan. Berikut adalah beberapa saran yang mungkin berguna[3].

teks sebelum melakukan koreksi ejaan.