Dokumen Implementasi STKI

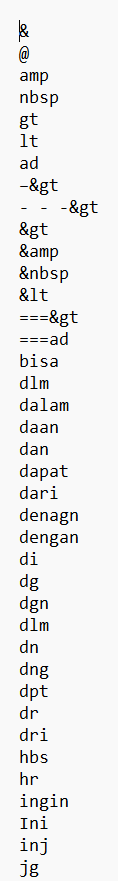
**Kelompok 2 STBI Magister Ilmu Komputer UGM**  
Anggota:

* Salsabila Nurulfarah Mahmudah (21/475591/PPA/06131)
* Dian Novitaningrum (21/483928/PPA/06177)
* Kartika Rizqi Nastiti (21/485636/PPA/06220)

Berikut adalah akses menuju Github : <https://github.com/Salsabila54/medical-record-information-retrieval>

*Progress* implementasi pada STKI, adalah:

1. Melakukan *list stopwords* dan *replacement typo* pada kolom FS\_ANAMNESA & FS\_DIAGNOSA
   1. *List Stopwords*



* 1. *List Typo Words*



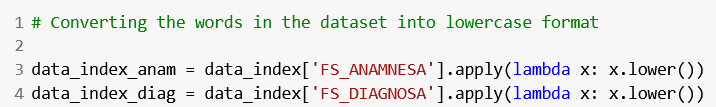
1. Mendapatkan data



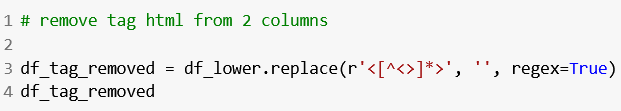
1. *Preprocessing* data rekam medis
   1. *Remove NaN*



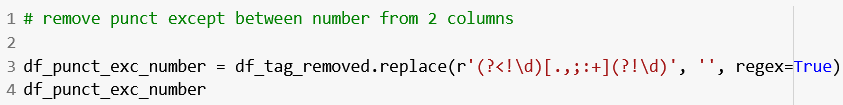
* 1. *Lowercase*



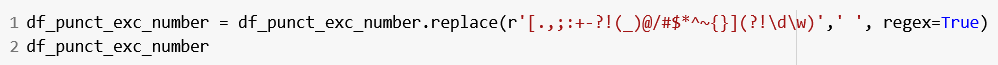
* 1. *Remove Tag HTML*



* 1. *Remove Punctuation*



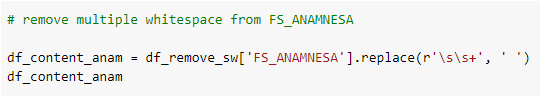




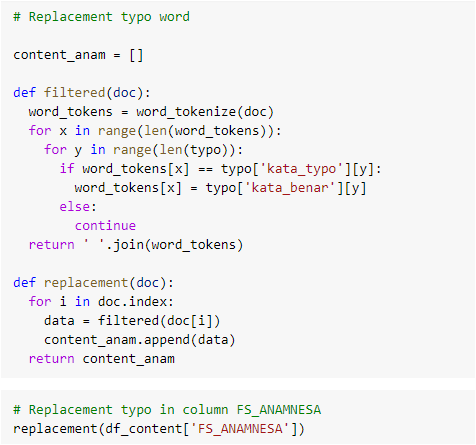
* 1. *Remove Stopwords*



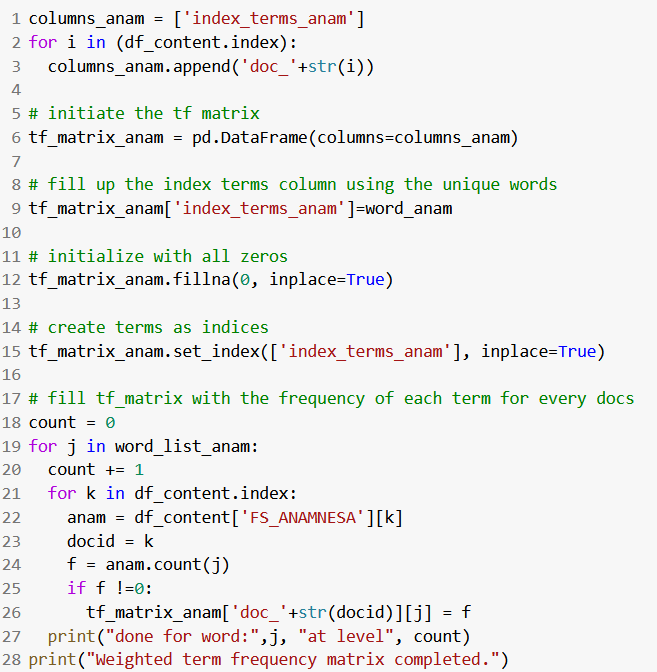
* 1. *Remove Multiple Whitespaces*

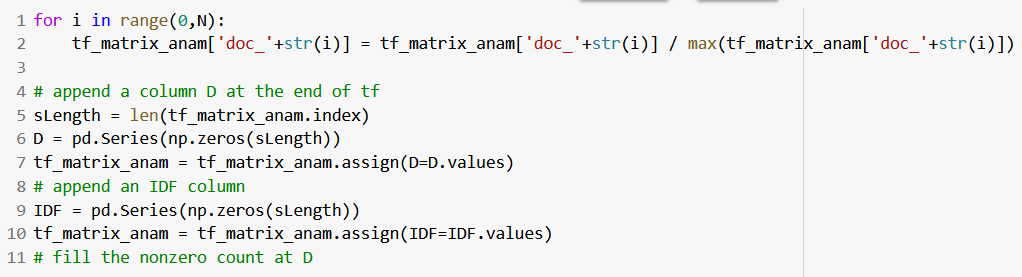
**

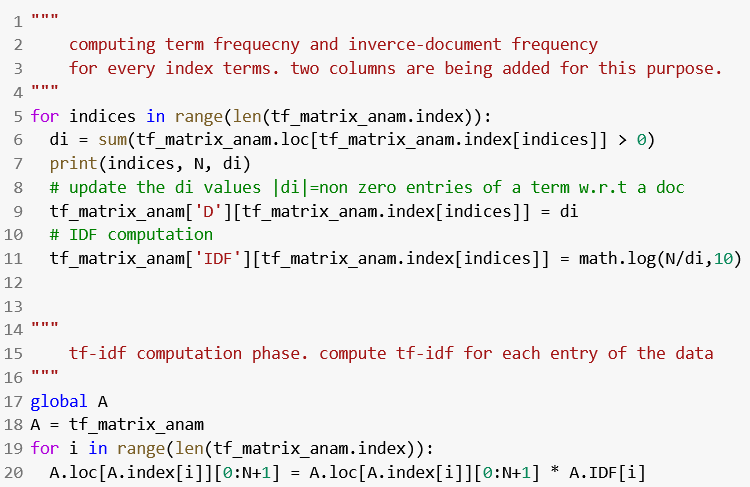
* 1. *Replace Typo*

**

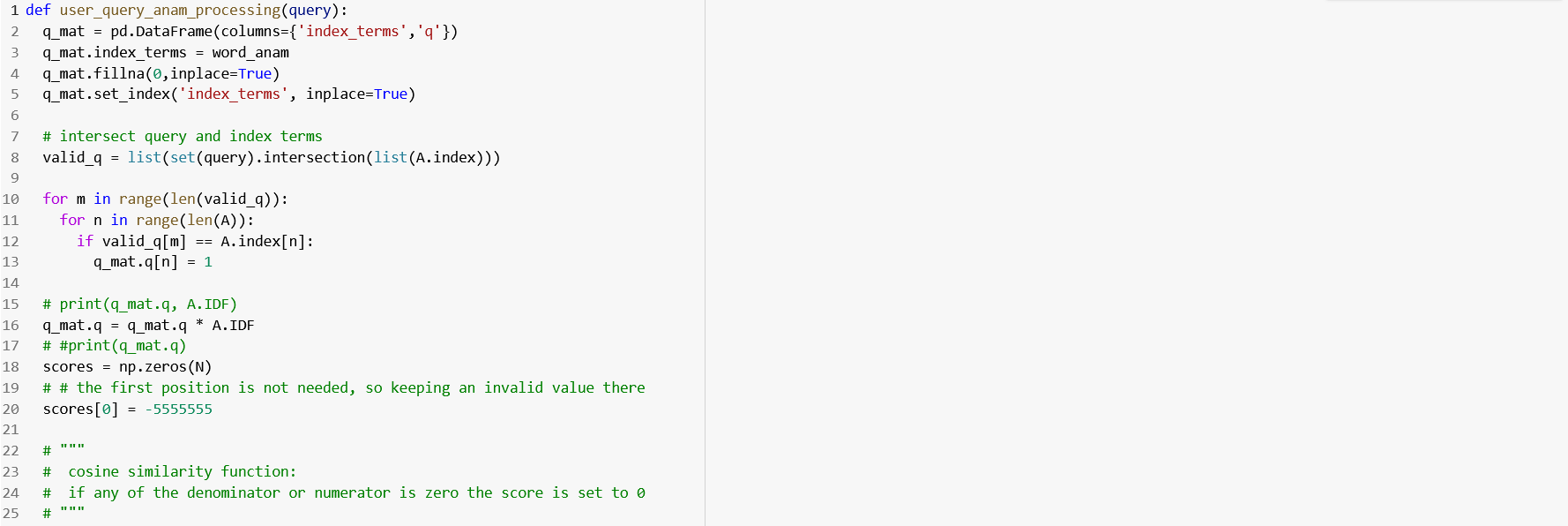
1. Pembuatan Matrix TF-IDF matrix



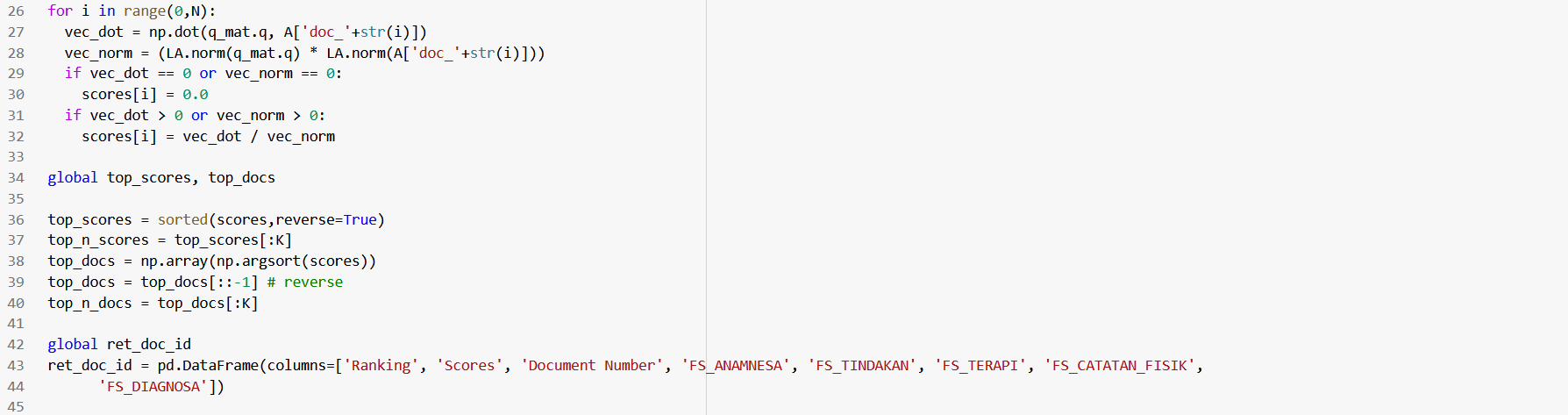


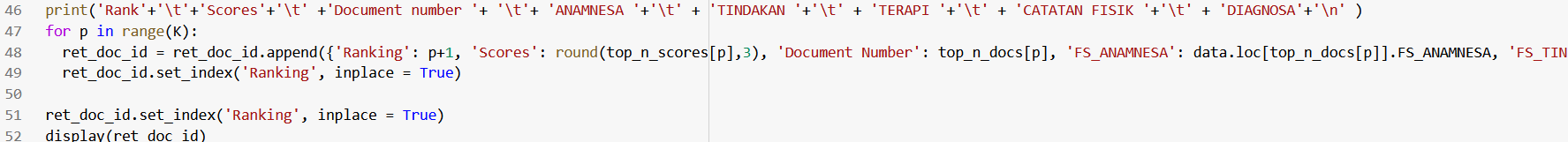


1. Pembuatan *cosine similarity* antara *Document* dan *Query*. Awal *query* yang diinputkan adalah pemilihan FS\_ANAMNESA atau FS\_DIAGNOSA. Jika menginginkan anamnesa menginputkan angka 0, sedangkan diagnosa menginputkan angka 1.

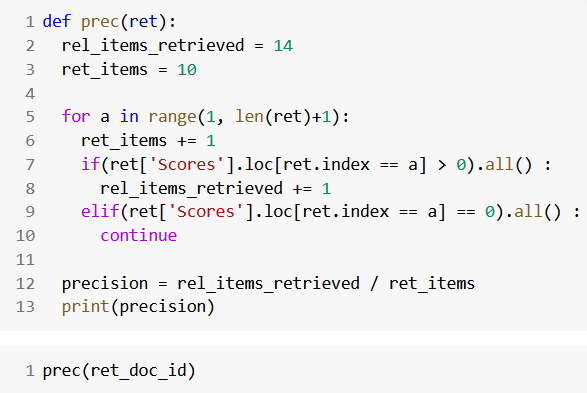


1. Melakukan perangkingan berdasarkan *cosine similarity.*





1. Evaluasi
   1. *Precision*



* 1. *Recall*

