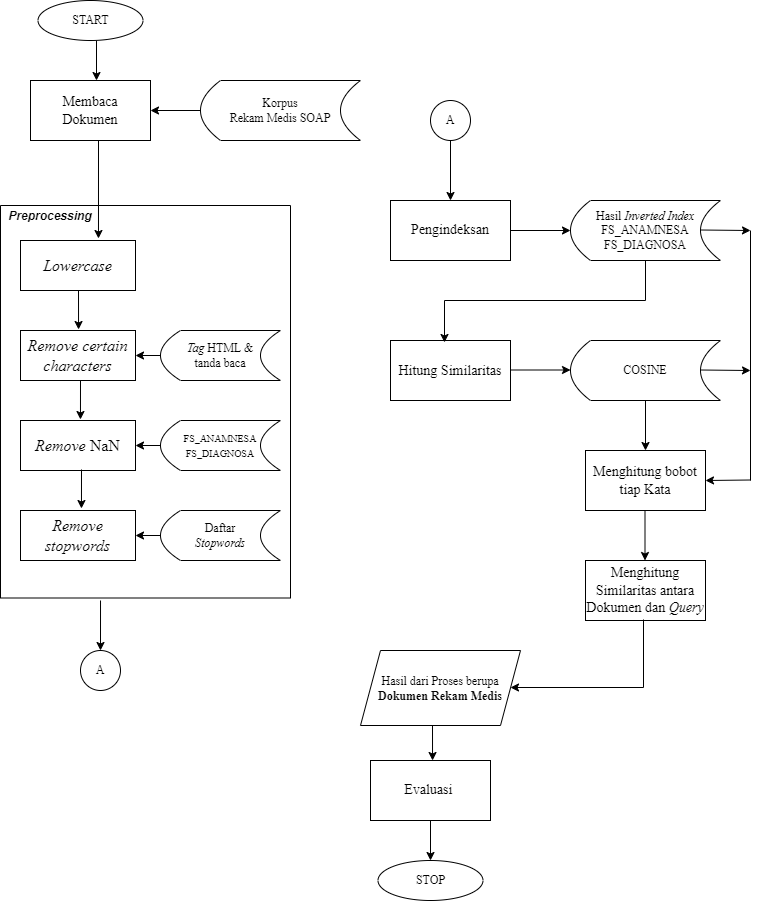
Dokumen *Design* STKI

**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)

***Design flowchart:***



# **Metodologi**

Berikut adalah proses implementasi sebagai analisis *design* pada STKI:

1. *Preprocessing*
   1. *Lowercase,* membuat semua huruf menjadi huruf kecil.
   2. *Remove certain characters,* seperti *tag* HTML, tanda baca, dan karakter lainnya.
2. *Tag* HTML yang dimaksudkan seperti <p> < /p>, <br />, <strong> ,

<p style="text-align: left;">kontrol</p> .

1. HTML *characters* *entities,* seperti &nbsp, &gt, &amp, &lt.
2. Simbol-simbol seperti (\_)!?@/=;:#$\*^~{}\.
   1. *Remove* NaN berdasarkan tiap kolom FS\_ANAMNESA dan FS\_DIAGNOSA.
   2. *Remove stopwords,* seperti kata sambung, kata depan, atau kata-kata lain yang tidak mengandung makna signifikan. Beberapa *stopwords* yang kami temukan adalah seperti bisa, dlm, dalam, daan, dan, dapat, dari, denagn, dengan, di.
3. Implementasi *Inverted Index,* menggunakan data dari kolom FS\_ANAMNESA dan FS\_DIAGNOSA. Selain itu, kami juga membuat 1 *database* yang berisi semua kolom tanpa melakukan *preprocessing.*
4. Implementasi penghitungan nilai *similarity* antar *terms* menggunakan *Vector Space Model (VSM).* Formulanya didasarkan pada *cosine similarity.*
5. Implementasi penghitungan bobot TF-IDF untuk tiap *term* dari tiap dokumen.
6. Implementasi penentuan dokumen yang mirip dengan *input query* dengan membandingkan nilai *vector space.*
7. *Input* *query* berupa *word* atau *sentences* dan *output* berupa dokumen rekam medis.
8. Evaluasi
   1. *Computation time* (proses *indexing* dan *querying*),
   2. *Precision,* dan
   3. *Recall.*