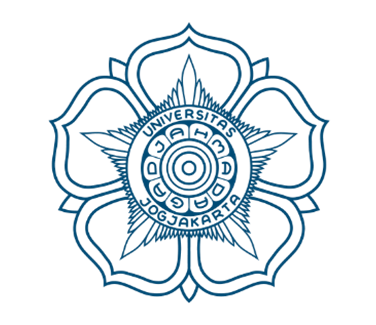
**SISTEM TEMU KEMBALI INFORMASI**

**TUGAS 5 METODE DAN HASIL PEMBUATAN INDEKS**

**DATA ANGGOTA ORGANISASI MAHASISWA (ORMAWA) ATAU UNIT KEGIATAN MAHASISWA (UKM)**

****

Disusun oleh :

Syuhra Putri Ganiswari 22/507327/NPA/19811

Ifta Mu’izana 22/507333/NPA/19813

**PROGRAM STUDI S1 ILMU KOMPUTER**

**DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS GADJAH MADA**

**YOGYAKARTA**

**2022**

**Sistem Temu Kembali Informasi Data Anggota Organisasi Mahasiswa (Ormawa) atau Unit Kegiatan Mahasiswa (UKM)**

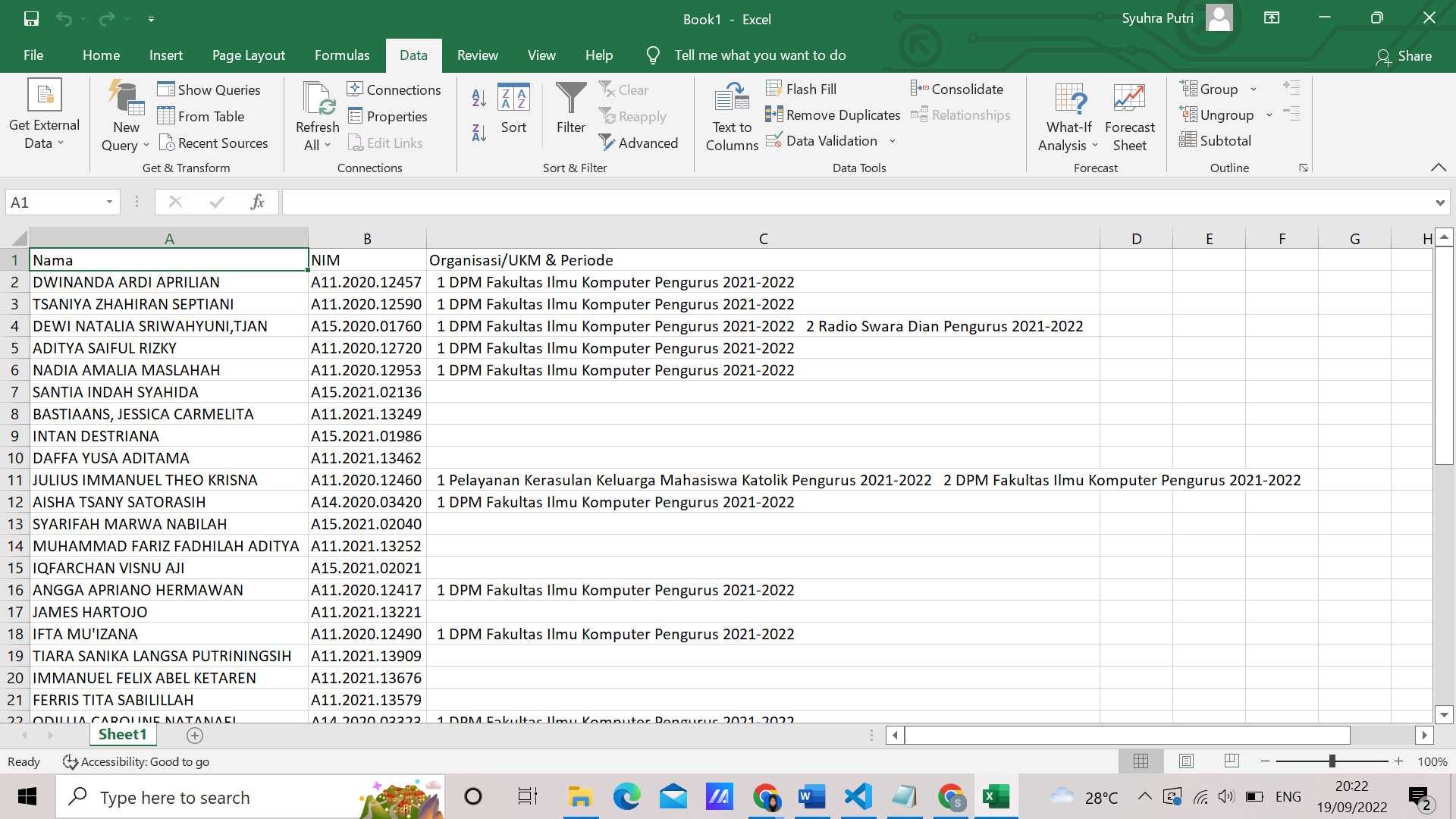
Anggota Kelompok :

1. Syuhra Putri Ganiswari      (22/507327/NPA/19811)
2. Ifta Mu’izana                      (22/507333/NPA/19813)

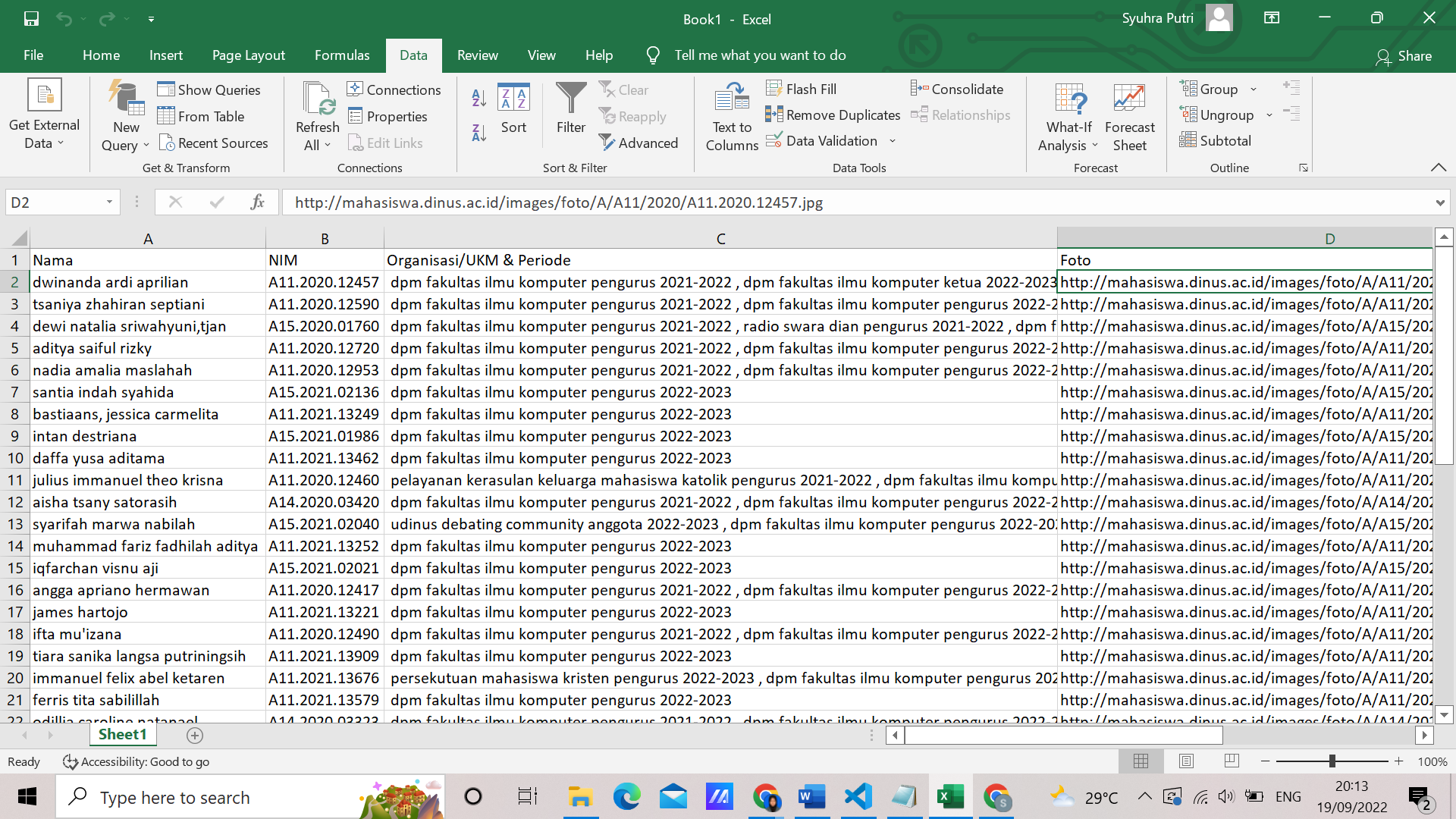
Tugas 5 – Index Contruction (Proses Pengindeksan)

1. Metode dan Tahapan Inverted Index yang Digunakan

Dalam pembuatan suatu sistem *Information Retrieval*, diperlukan adanya *indexing* supaya dokumen siap untuk diproses. Pada tahap Indexing sistem temu kembali yang kami buat, kami menerapkan metode *inverted index* dimana ada beberapa tahapan atau langkah - langkah yang harus dilakukan sebelum melakukan *inverted index*. Pada *source code* yang sebelumnya, telah kami perbarui supaya didapatkan data nama dan organisasi/ukm dengan huruf kecil pada setiap kata. Selain itu, kami juga menghilangkan spasi beserta angka yang tidak digunakan. Kami juga menambahkan kolom data berupa url foto mahasiswa.



*gambar 1 dataset sebelum cleaning data*



*gambar 2 dataset setelah cleaning data*

Setelah data telah bersih dan telah diubah menjadi huruf kecil di setiap katanya, langkah selanjutnya adalah tokenisasi tiap kalimat yaitu dilakukan pemecahan tiap kolomnya menjadi potongan kata dan menjadi kumpulan term. Kemudian term tersebut diberi ID dokumen.

**doc ID Document Tokenized term**

0 [dwinanda adri aprilian] [‘dwinanda’,’adri’,’aprilian’]

1 [tsaniya zhahiran septiani] [‘tsaniya’,‘zhahiran’,’septiani’]

2 [dewi natalia sriwahyuni] [‘dewi’,’natalia’,’sriwahyuni’]

3 [aditya saiful rizky] [‘aditya’,’saiful’,’rizky’]

dst.

Setelah tokenisasi, data baru dapat dilakukan *inverted index* seperti berikut :

**Term DocFreq PostingLists**

dwinanda 1 0

adri 1 0

aprilian 1 0

tsaniya 1 1

zhahiran 1 1

septiani 1 1

dewi 1 2

natalia 1 2

sriwahyuni 1 2

aditya 1 3

saiful 1 3

rizky 1 3

…. …. ….

…. …. ….

1. Referensi :
2. <https://www.markijar.com/2015/07/information-retrieval.html>
3. <https://stackoverflow.com/questions/73553369/i-use-the-word-tokenize-function-on-my-dataframe-by-writing-word-dict-but-afte>
4. <https://www.youtube.com/watch?v=aUgrfDhPc4Y&t=282s>
5. <https://iq.opengenus.org/idea-of-indexing-in-nlp/>