LAPORAN

**Pengembangan Produk Inovasi Fakultas Sains dan Teknologi 2024 pada Kategori *Islam and Digital Transformation***

Pengembangan Dataset Pedoman Pengenalan Entitas Bernama pada Terjemahan Al-Quran berbahasa Indonesia (E-IndQNER) dengan Metode

***Self-Training* menggunakan *Large Language Models*: Menuju Islamicpedia**



# Oleh:

**Ria Hari Gusmita, M.Kom**

NIP: 198208172009122002

# Khodijah Hulliyah, Ph.D.

NIDN: 2128018202

# Kamal Fikry Musa, Lc., M.A

NIDN: 2128018202

# Program Studi Teknik Informatika dan Ilmu Al-Qur’an dan Tafsir

**UIN Syarif Hidayatullah Jakarta Februari 2024**

## LATAR BELAKANG

Perkembangan teknologi saat ini telah menghadirkan transformasi yang men- dalam pada kehidupan manusia. Kecerdasan buatan (*Artificial Intelligence/AI*) me- rupakan salah satu bidang ilmu yang memberikan dampak besar pada teknologi. ChatGPT, yang diperkenalkan di bulan November 2022, merupakan terobosan tek- nologi kecerdasan buatan terkini yang telah mengubah cara kita berinteraksi dengan komputer dan internet. ChatGPT diklaim sebagai teknologi yang memiliki kemam- puan lebih canggih dibandingkan pendahulunya, seperti *Google*. Hal ini ditandai dengan salah satunya adalah fakta menarik seperti jumlah anggota (*member*) seba- nyak lebih dari 1 juta orang di 5 hari pertama kemunculannya.[1](#_bookmark0) ChatGPT mam- pu memahami dan menghasilkan teks dengan tingkat kemiripan interaksi manusia yang semakin tinggi, membuka peluang baru dalam berbagai sektor di masyarakat.[2](#_bookmark0) Kemampuan *ChatGPT* tersebut terlihat di Gambar [1.1](#_bookmark1) ketika menjawab pertanya- an dari seseorang yang meninggalkan sholat Subuh karena terlambat bangun dari tidurnya. Selain memberikan jawaban yang benar, paragraf terakhir menunjukk- an *ChatGPT* terlihat memotivasi pengguna untuk memperbaiki ibadahnya karena Allah Subhanahuwwata’ala Maha Pengampun dan Maha Penyayang. Kenyaman- an yang didapatkan dari *ChatGPT* seperti ini membuat jumlah akses terhadapnya meningkat tajam, di mana angkanya mencapai 1.7 milyar di bulan Oktober 2023.[3](#_bookmark0)

Fakta yang telah dijelaskan di paragraf sebelumnya tidak serta merta mem-

buat *ChatGPT* dapat digunakan untuk memperoleh informasi tentang semua hal yang terkait dengan keislaman. Gambar [1.2](#_bookmark2) menunjukkan bagaimana *ChatGPT* menyediakan jawaban dari pertanyaan terkait salah satu isu terkini di Indonesia, yakni, *Apakah LGBT haram menurut Al-Qur’an*. Bagian jawaban yang diberi ko- tak berwarna merah menunjukkan informasi yang kontradiktif, di mana hal ini dapat berpotensi mengelirukan pengguna, khususnya umat Islam. Hal yang ditunjukkan di Gambar [1.2](#_bookmark2) merupakan salah satu faktor yang menyebabkan munculnya keputus- an bahwa penggunaan teknologi AI sebagai pedoman keagamaan Islam dinyatakan haram. Ketetapan ini dihasilkan pada musyawarah nasional alim ulama dan Konfe- rensi Besar Nahdlatul Ulama pada bulan September 2023.[4](#_bookmark0) Adanya jawaban yang dapat mengelirukan dari *ChatGPT* disebabkan karena *ChatGPT* menggunakan in- formasi yang berasal dari internet, seperti, halaman web, buku elektronik, *Wikipe- dia*, artikel penelitian, dan respon penggunanya terhadap jawaban yang diberikan.[5](#_bookmark0) Seluruh sumber informasi tersebut memiliki satu kesamaan, yakni, dihasilkan oleh manusia. Perbedaan koneksi fisik yang terjadi di otak manusia menjadi hal uta-

1[https://explodingtopics*.*com/blog/chatgpt-users](https://explodingtopics.com/blog/chatgpt-users)

2[https://timesofindia*.*indiatimes*.*com/readersblog/marketing-savvy/chatgpt-](https://timesofindia.indiatimes.com/readersblog/marketing-savvy/chatgpt-and-its-impact-on-society-50445/)

[and-its-impact-on-society-50445/](https://timesofindia.indiatimes.com/readersblog/marketing-savvy/chatgpt-and-its-impact-on-society-50445/)

3[https://explodingtopics*.*com/blog/chatgpt-users#region](https://explodingtopics.com/blog/chatgpt-users#region)

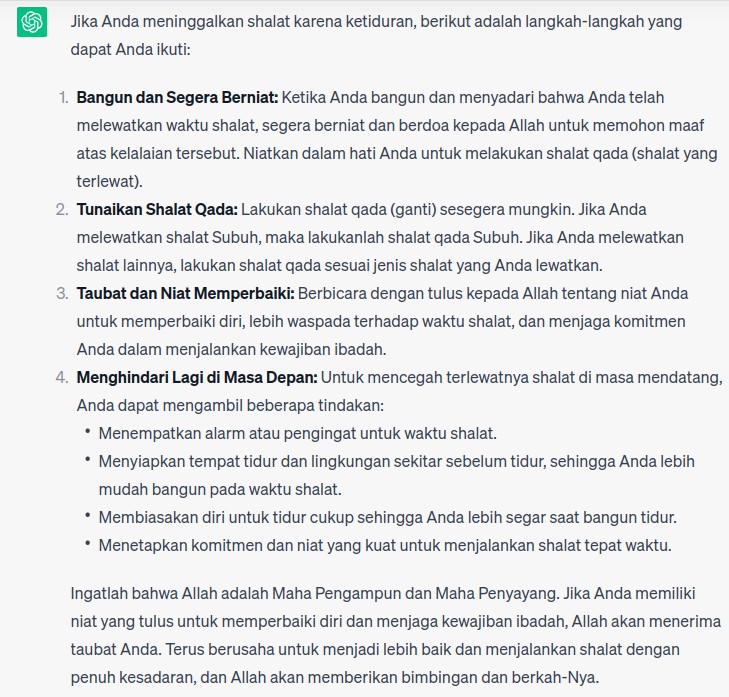
4[https://www*.*antaranews*.*com/berita/3733755/munas-nu-sebut-haram-jawaban-ai-](https://www.antaranews.com/berita/3733755/munas-nu-sebut-haram-jawaban-ai-dijadikan-panduan-fatwa)

[dijadikan-panduan-fatwa](https://www.antaranews.com/berita/3733755/munas-nu-sebut-haram-jawaban-ai-dijadikan-panduan-fatwa)

5[https://analyticsindiamag*.*com/behind-chatgpts-wisdom-300-bn-words-570-gb-](https://analyticsindiamag.com/behind-chatgpts-wisdom-300-bn-words-570-gb-data/)

[data/](https://analyticsindiamag.com/behind-chatgpts-wisdom-300-bn-words-570-gb-data/)

ma yang menjadikan manusia memiliki cara berfikir yang berbeda satu sama lain.[6](#_bookmark0) Perbedaan inilah yang berkontribusi pada adanya informasi kontradiktif yang diha- silkan oleh *ChatGPT*. Untuk membantu umat muslim (dan manusia pada umumnya) terhindar dari informasi yang mengelirukan terkait dengan keislaman seperti dije- laskan di atas, maka teknologi seperti *ChatGPT* perlu dilengkapi dengan sumber informasi yang akurat, dalam hal ini yang berasal dari Al-Qur’an dan hadis.



**Gambar 1.1** Tampilan Layar Penggunaan *ChatGPT*

6[https://www*.*sciencedaily*.*com/releases/2013/02/130206131048*.*htm#](https://www.sciencedaily.com/releases/2013/02/130206131048.htm)



**Gambar 1.2** Tampilan Informasi dari *ChatGPT* yang Dapat Mengelirukan

Dataset pedoman merupakan kumpulan data yang telah diberi label oleh pa- kar di bidang tertentu melalui metodologi yang telah disepakati. Keterlibatan pa- kar pada pembuatan dataset pedoman menjadikan dataset pedoman sebagai data- set standar untuk penelitian bidang terkait. Pada bidang pengolahan bahasa ma- nusia (*natural language processing*), dataset pedoman memiliki peranan penting untuk mengevaluasi kinerja suatu sistem melalui pengujian memanfaatkan dataset pedoman tersebut. Selanjutnya, berdasarkan hasil evaluasi, kita dapat mengeta- hui sejauhmana kinerja dari sistem tersebut dengan memanfaatkan ukuran kinerja menggunakan *f-measure*, *accuracy*, *precision*, *recall*, atau ukuran kinerja lainnya. Mekanisme ini membuat penelitian pada pembuatan atau pengembangan sistem ter- tentu dapat menjadi berkelanjutan dan penemuan yang dihasilkan pada akhirnya dapat bersifat relatif mapan (*mature*). Pentingnya ketersediaan dataset pedoman melatarbelakangi banyaknya kegiatan untuk membuat dataset tersebut. Kegiatan yang dimaksud berbentuk tugas bersama (*shared-task*) pada konferensi internasio- nal, workshop, kerjasama antara institusi penelitian, dan penelitian individu.

Salah satu contoh pemanfaatan dataset pedoman untuk keperluan peneliti-

an adalah penggunaan dataset pedoman dalam bidang penelitian pengenalan entitas bernama (PEB). PEB merupakan bidang penelitian yang mengidentifikasi entitas bernama (EB) pada teks, dan selanjutnya mengelompokkannya ke kelas yang se- suai. Kelas EB yang umum digunakan adalah nama orang (*person*), nama loka- si/tempat (*location*), dan nama organisasi (*organization*). Lebih lanjut, PEB me- rupakan proses fundamental pada pembangunan teknologi/sistem yang mengolah bahasa manusia, seperti *ChatGPT*. Untuk sistem PEB berbahasa Indonesia, ter- dapat 4 (empat) dataset pedoman pertama yang dipublikasikan secara bersamaan dengan *transformer-based pre-trained language model*, IndoBERT [[1,](#_bookmark4) [2].](#_bookmark5) Dataset tersebut adalah NERGrit[7](#_bookmark0), NERProsa[8](#_bookmark0), NER UI[9](#_bookmark0), dan NER UGM.[10](#_bookmark0) Seluruh da- taset ini hanya diperuntukkan bagi sistem PEB dengan domain informasi bersifat umum. Selanjutnya, IndQNER diperkenalkan pada bulan Juni 2023 sebagai data- set pedoman dari domain khusus, yakni terjemahan Al-Qur’an berbahasa Indonesia [[3].](#_bookmark6) Proses pelabelan entitas bernama pada IndQNER dilakukan secara manual serta melibatkan pakar di bidang ilmu Al-Qur’an dan tafsir. Karena keterbatasan sum- ber daya, IndQNER hanya mencakup entitas bernama dari 8 (delapan) terjemahan surat di Al-Qur’an. Surat tersebut antara lain, Al-Baqarah, Ali-’Imran, An-Nisa, Al-Maidah, Al-An’am, Al-A’raf, Yunus, dan An-Nahl.

Sebagai dataset yang dibangun dari terjemahan Al-Qur’an berbahasa Indo-

nesia, IndQNER dapat berperan penting bagi solusi penyediaan sumber informa- si akurat bagi teknologi AI yang akan dimanfaatkan umat muslim khususnya un- tuk mendapatkan informasi tentang keislaman. Oleh karena itu, melalui proposal ini penulis mengusulkan rencana penelitian untuk mengembangkan IndQNER se- hingga dataset pedoman tersebut dapat mencakup seluruh surat pada terjemahan Al-Qur’an berbahasa Indonesia. Dengan demikian, IndQNER dapat meningkatkan nilai manfaatnya secara signifikan bagi pengembangan teknologi AI yang menyedi- akan informasi berdasarkan Al-Qur’an. Dataset hasil pengembangan IndQNER se- lanjutnya disebut dengan E-IndQNER (*Extended* IndQNER). Karena pembangun- an E-IndQNER dilakukan pada 106 surat pada terjemahan Al-Qur’an berbahasa Indonesia, maka pelabelan akan dilakukan secara otomatis dengan pendekatan *self- training* menggunakan *large language models/LLM*. Hasil pelabelan ini selanjutnya akan diverifikasi oleh pakar di bidang Al-Qur’an. Karena itulah, peneliti berkolabo- rasi dengan Lajnah Pentashihan Mushaf Al-Qur’an (LPMQ), Kementerian Agama Republik Indonesia. LPMQ merupakan lembaga yang berperan pada proses penter- jemahan Al-Qur’an berbahasa Indonesia. Selanjutnya, E-IndQNER akan menjadi sumber daya penting dalam pembanguan *Islamicpedia*, yakni basis pengetahuan (*knowledge base*) tentang keislaman yang sumber datanya berasal dari Al-Qur’an, tafsir dan hadis. *Islamicpedia* ini akan hadir untuk memfasilitasi pembangunan tek- nologi, seperti chatGPT, yang dapat memberikan informasi akurat bagi umat Islam khususnya.

7[https://github*.*com/indobenchmark/indonlu/tree/master/dataset/nergrit ner-](https://github.com/indobenchmark/indonlu/tree/master/dataset/nergrit_ner-grit) [grit](https://github.com/indobenchmark/indonlu/tree/master/dataset/nergrit_ner-grit)

8[https://github*.*com/indobenchmark/indonlu/tree/master/dataset/nerp ner-prosa](https://github.com/indobenchmark/indonlu/tree/master/dataset/nerp_ner-prosa)

9[https://github*.*com/indolem/indolem/tree/main/ner/data/nerui](https://github.com/indolem/indolem/tree/main/ner/data/nerui) 10[https://github*.*com/indolem/indolem/tree/main/ner/data/nerugm](https://github.com/indolem/indolem/tree/main/ner/data/nerugm)

## PERUMUSAN MASALAH

IndQNER dikembangkan dengan memanfaatkan daftar entitas bernama serta kelas relevan terkait dari ontologi konsep di Al-Qur’an yang dibangun oleh peneliti Leeds University, selanjutnya disebut OLU.[11](#_bookmark0) Selama proses pelabelan manual di- lakukan, ditemukan kandidat entitas bernama dan kelas entitas bernama. Selain itu, ditemukan pula bentuk kemunculan entitas bernama di teks terjemahan Al-Qur’an berbahasa Indonesia (selanjutnya disebut teks) yang berbeda dengan teks terjemah- an dalam Bahasa Inggris, sumber dokumen pembangunan OLU. Kedua fakta ter- sebut berkontribusi dalam memperbesar jumlah entitas bernama dan kelas entitas bernama di IndQNER. Pembangunan E-IndQNER akan dilaksanakan dengan me- tode *self-training* dengan *LLMs*, di mana pelabelan akan dilakukan secara otomatis pada 106 surat pada teks. Selanjutnya, 106 surat pada terjemahan Al-Qur’an ber- bahasa Indonesia disebut sebagai teks. Berdasarkan fakta-fakta tersebut, penelitian ini akan fokus pada masalah berikut ini.

1. Bagaimana mengidentifikasi pola kemunculan entitas bernama di teks?
2. Bagaimana mengidentifikasi kandidat entitas bernama dan kelas entitas ber- nama di teks?
3. LLM mana yang terbaik untuk melakukan prediksi pada teks di proses *self- training*, di mana distribusi entitas bernama dari beberapa kelas cenderung tidak seimbang?

## BATASAN MASALAH

Pembangunan E-IndQNER akan dilakukan pada ruang lingkup sebagai ber-

ikut.

1. Dari setidaknya 3 (tiga) versi terjemahan Al-Qur’an berbahasa Indonesia yang tersedia, penelitian ini hanya akan menggunakan versi yang dihasilkan oleh Kementerian Agama Republik Indonesia tahun 2019.
2. Pemilihan LLM yang akan digunakan di proses *self-training* dilakukan mela- lui survey pada artikel penelitian yang menggunakan LLM pada teks Bahasa Indonesia. Tiga metode yang paling banyak digunakan atau memberi kinerja baik selanjutnya akan dipilih sebagai LLMs di penelitian ini.
3. Seperti halnya IndQNER, E-IndQNER akan dievaluasi dengan 2 (dua) ske- nario, yakni, berbasis *supervised learning* dan *transfer learning*.

11[https://corpus*.*quran*.*com/](https://corpus.quran.com/)

## TUJUAN PENELITIAN

Melalui penelitian ini, dataset pedoman untuk proses pengenalan EB pada terjemahan Al-Qur’an berbahasa Indonesia yang lengkap (dari seluruh surat di Al- Qur’an) akan dihasilkan. Dataset pedoman yang dimaksud diharapkan dapat men- jadi pelopor standardisasi penelitian di bidang integrasi Teknologi dan Komunika- si (TIK) dan Islam secara umum, dan penelitian PEB pada terjemahan Al-Qur’an berbahasa Indonesia secara khusus. Standardisasi ini pula yang akan membawa penelitian-penelitian terkait menjadi berkesinambungan menuju tahapan di mana perkembangan penelitian berada di level matang (*mature*), sehingga siap digunakan untuk membangun *Islamicpedia* serta teknologi AI di bidang keislaman. Pencapa- ian visi besar ini diperkuat dengan dilakukannya kerjasama personel yang berasal dari bidang teknik informatika dan ilmu Al-Qur’an dan tafsir di UIN Syarif Hidaya- tullah Jakarta. Lebih lanjut, kerjasama personel di penelitian ini diharapkan dapat membentuk komunitas penelitian di bidang integrasi keilmuan yang menjadi aktua- lisasi visi UIN Jakarta, yakni keunggulan integrasi ilmu keislaman, keindonesiaan, dan sains.

## PENELITIAN SEJENIS

Pada tahun 2017, [[4]](#_bookmark7) membuat dataset sistem PEB pada dokumen berbahasa Indonesia pertama dengan memanfaatkan DBpedia berbahasa Indonesia (*DBpedia Entities Expansian/DEE*). Penelitian ini diawali dengan analisis pada EB yang te- lah dihasilkan di pengembangan sistem PEB sebelumnya [[5,](#_bookmark8) [6].](#_bookmark9) Berdasarkan hasil analisis tersebut, kriteria dari masing-masing kelas EB yang terdiri dari *Person, Lo- cation, Organization*, didefinisikan. Terdapat 11 cara untuk *Person*, 6 cara untuk *Location*, dan 5 cara untuk *Organization*. Setiap EB selanjutnya diproses dalam 4 tahap, yakni pembersihan nama (*name cleansing*), normalisasi nama (*name norma- lization*), ekspansi nama (*name expansion*), dan validasi nama (*name validation*). Seluruh nama di tahap terakhir divalidasi menggunakan Kamus Besar Bahasa Indo- nesia (KBBI)[12](#_bookmark0) dan korpus NLTK untuk Bahasa Inggris.[13](#_bookmark0) Penelitian ini mengha- silkan lebih banyak EB untuk kelas *Location* (lebih dari 2 kali lipat) dan *Person* (2 kali lipat). DEE dievaluasi melalui pengukuran kinerja sistem PEB yang dibangun dengan menggunakan perangkat lunak *Stanford Named Entity Recognition (NER)*, di mana sistem menghasilkan kenaikan sebesar 3.41 poin nilai *precision* dan nilai *recall* yang dua kali lebih tinggi dibandingkan nilai yang diperoleh di [[5]](#_bookmark8) dan [[6].](#_bookmark9)

Dataset DEE yang dihasilkan oleh [[4]](#_bookmark7) ternyata memiliki kelemahan di mana terdapat kesalahan pada pendefinisian dan pelabelan kelas EB *Person*. Oleh karena itu, [[7]](#_bookmark10) memperbaiki dataset tersebut dan menciptakan dataset baru yang disebut dengan *Modified DBpedia Entities Expansion* atau M-DEE. Penulis mendefinisikan kriteria baru bagi EB *Person* yang ada dan telah dihilangkan di dataset, dan meng- ubah kriteria untuk EB yang ada di dataset. Selain itu, terdapat penambahan 1 kri-

12[https://kbbi*.*kemdikbud*.*go*.*id/Beranda](https://kbbi.kemdikbud.go.id/Beranda) 13[https://www*.*nltk*.*org/howto/corpus*.*html](https://www.nltk.org/howto/corpus.html)

teria baru untuk kelas *Organization*. Kriteria baru ini memperbesar ukuran M-DEE di mana terdapat EB baru sebanyak 2036 dari kelas *Location*, lebih dari 6521 dari kelas *Person*, dan 352 dari kelas *Organization*. M-DEE dievaluasi dengan teknik yang sama dengan DEE, di mana sebuah sistem PEB dibuat menggunakan *Stanford NER* untuk diuji menggunakan bagian pengujian dataset. Evaluasi ini menunjukk- an bahwa jumlah kriteria yang banyak pada kelas *Person* tidak memberi pengaruh positif pada sistem PEB, karena nilai *precision* hanya meningkat sebanyak 3 poin dan nilai *recall* relatif sama dibandingkan hasil pada [[5]](#_bookmark8) dan [[6]](#_bookmark9) untuk kelas *Person*. Sementara itu, penggunaan *Gazetteers* untuk menambah EB dari kelas *Location* ter- bukti efektif, karena nilai *precision* dan *recall* meningkat sebanyak masing-masing

5.75 dan 24.46 poin.

Usaha untuk melakukan standardisasi dataset pedoman PEB pada domain informasi umum pertama kali dilakukan oleh [[8].](#_bookmark11) Penelitian ini dilakukan dengan cara melabeli kembali dataset PEB yang telah tersedia secara publik dari [[9]](#_bookmark12) (se- lanjutnya disebut dataset S&N). Dataset ini memiliki daftar EB dari 5 kelas, yakni *Person Location*, *Organization*, waktu *Time*, dan kuantitas *Quantity*. Penulis meng- awali pelabelan dengan cara menganalisis kualitas pelabelan pada dataset S&N. Hasil analisis menunjukkan terdapat banyak kesalahan pelabelan pada EB, teruta- ma dari kelas *Organization*, lalu diikuti dengan EB dari kelas *Location* dan *Person*. Pelabelan kembali pada dataset S&N dilakukan secara manual oleh 3 orang Indo- nesia. Dari 5 kelas EB, para pelabel fokus melabeli EB dari *Person*, *Location*, dan *Organization*. Hal ini dikarenakan, penulis fokus pada usaha untuk meningkatkan kinerja model PEB dalam menangani ambiguitas pada kata benda. Evaluasi dataset pedoman dilakukan dengan mengukur kinerja sistem PEB yang dibangun dengan teknik *Bi-directional Long Short-Term Memory* (BiLSTM) dan *Conditional Ran- dom Field* (CRF). Hasil evaluasi menunjukkan bahwa pelabelan kembali dataset S&N ini meningkatkan kinerja sistem PEB secara signifikan, di mana nilai F1 score bertambah dari 76.11% menjadi 90.85%.

Dataset pedoman pertama untuk PEB dari domain khusus dirilis pada bulan Juni 2023. Dataset ini dinamakan IndQNER, Named Entity Recognition benchmark dataset from the Indonesian translation of the Quran. IndQNER berisi 3117 kalimat yang berasal dari 8 (delapan) terjemahan surat AL-Qur’an, dan 2475 entitas berna- ma yang merepresentasikan 18 kelas entitas bernama. Proses pelabelan dilakukan secara manual oleh 8 orang pelabel, di mana setiap 2 orang pelabel melabeli teks yang sama. Hal ini dilakukan untuk memenuhi unsur representatif dan menghindari bias pada hasil pelabelan. Proses pembangunan IndQNER diawali dengan pendefi- nisian daftar entitas bernama dan kelas entitas bernama, di mana proses ini meman- faatkan ontologi konsep di Al-Qur’an yang dibangun di Leeds University. Setelah itu, pedoman pelabelan didefinisikan berdasarkan hasil analisis teks terjemahan Al- Qur’an berbahasa Indonesia. Pelabelan manual dilakukan mengikuti pedoman yang dihasilkan, di mana pelabel juga mengidentifikasi kandidat entitas bernama dan kelas entitas bernama. Kandidat tersebut selanjutnya diverifikasi oleh pakar yang berasal dari ilmu Al-Qur’an dan tafsir. Salah satu kontribusi dari penelitian ini ada- lah tersedianya pedoman pelabelan yang didesain secara komprehensif dan dapat

diakses secara publik. Pengukuran dengan *Inter-Annotator Agreement* dilakukan untuk mengetahui tingkat representatif hasil pelabelan dari 2 pelabel. Kualitas ha- sil pelabelan IndQNER dievaluasi melalui 2 skenario yang berbasiskan *supervised learning* dan *transfer learning*. Diketahui bahwa hasil pengujian pada skenario per- tama, yakni yang menggunakan BiLSTM dan CRF, lebih baik dibandingkan dengan skenario kedua yang menggunakan pendekatan *fine-tuning* IndoBERT (*pre-trained language model*) untuk Bahasa Indonesia. Nilai F1 pada skenario pertama adalah 98%, sedangkan skenario kedua hanya mampu menghasilkan nilai F1 sebesar 71%.

## KAJIAN TEORI

Berikut ini diberikan kajian teori yang melandasi penelitian ini meliputi te- ori tentang dataset pedoman, pengenalan EB, pendekatan *self-training*, dan *large language models/LLMs*.

### Dataset Pedoman

Dataset pedoman menyediakan acuan standar yang dapat digunakan untuk membandingkan suatu sistem dengan sistem lainnya. Dengan kata lain, dataset pedoman berperan untuk mengevaluasi kinerja sebuah sistem terhadap kinerja sis- tem lainnya. Dataset pedoman yang umumnya digunakan di bidang pembelajaran mesin dan pengolahan bahasa manusia terdiri dari satu atau beberapa dataset, satu atau beberapa metrik evaluasi terkait, dan mekanisme untuk meningatkan kinerja sebuah sistem. Lebih lanjut, sebuah dataset pedoman menetapkan sebuah standar untuk menilai sistem berbeda dari bidang penelitian yang sama.[14](#_bookmark0) Pembuatan da- taset pedoman untuk PEB dimulai dari konferensi *Message Understanding* keenam yang dilaksanakan tahun 1995 [[10].](#_bookmark13) Konferensi ini menghasilkan dataset pedoman PEB pertama dengan nama MUC-6. Dataset MUC-6 diperbaharui menjadi MUC-7 melalui konferensi yang sama dua tahun berikutnya [[11].](#_bookmark14) CoNLL merupakan data- set pedoman selanjutnya yang dihasilkan melalui konferensi, yakni *Conference on Computational Natural Language Learning* [[12].](#_bookmark15) Dataset ini berisi EB dari kelas *Person, Location, organization,* dan *miscellaneous*. Kelas terakhir diperuntukkan bagi EB yang bukan berasal dari tiga kelas pertama. Di tahun 2005, *The Lingu- istic Data Consortium* membuat dataset pedoman PEB dengan nama dataset ACE (*Automatic Content Extraction*) yang bersifat multilingual dari dokumen berbahasa Inggris, Arab, dan Cina.[15](#_bookmark0) Selanjutnya, dataset *OntoNotes* dihasilkan dari kolabo- rasi sekelompok institusi di Amerika Serikat yang berisi berbagai jenis data seperti percakapan telepon, siaran langsung, berita, dan *talk show*. Dataset ini banyak di- gunakan pada pengembangan sistem PEB berbahasa Inggris hingga saat ini. Di sisi lain, terdapat dataset pedoman yang dihasilkan dari sebuah workshop, yakni *Semantic Evaluation*.[16](#_bookmark0) Dataset ini dikenal dengan nama *SemEval Corpus*. Sela-

14[https://ruder*.*io/nlp-benchmarking/index*.*html#what-is-a-benchmark](https://ruder.io/nlp-benchmarking/index.html#what-is-a-benchmark) 15https://catalog.ldc.upenn.edu/LDC2006T06

16https://semeval.github.io/

in dataset yang telah dijelaskan sebelumnya, terdapat pula dataset pedoman dari domain kedokteran seperti *MEDLINE*, *PubMed*, dan *PubMed Central*.[17](#_bookmark0)

### Pengenalan Entitas Bersama

Pengenalan EB (PEB) merupakan proses fundamental di area studi Pengo- lahan Bahasa Manusia (PBM). Berdasarkan [[13],](#_bookmark16) PBM merupakan bidang ilmu yang mempelajari metode untuk membuat sebuah komputer (mesin) dapat mem- baca, menganalisis, dan pada akhirnya mengetahui makna yang terkandung dalam bahasa manusia. Lebih lanjut, PBM memungkinkan sebuah mesin untuk melakuk- an perkerjaan yang melibatkan bahasa manusia, seperti mengakomodasi komunika- si manusia dan mesin, manusia dan manusia, serta memproses teks atau suara [[14].](#_bookmark17) Berikut ini diberikan uraian tentang konsep EB dan PEB secara lebih detil.

### Entitas Bernama

Entitas bernama merupakan sebuah kata atau frase pada teks yang secara langsung mengacu/merepresentasikan obyek unik di dunia nyata [[15].](#_bookmark18) Contoh EB antara lain nama orang tertentu (B.J. Habibie), organisasi tertentu (Palang Merah Indonesia), dan lokasi tertentu (Gedung Dewan Pers). Selanjutnya, setiap EB me- miliki kelas yang sesuai yakni nama orang, nama organisasi, atau nama lokasi. Se- lain mengacu pada obyek fisik yang ada di dunia nyata, EB juga digunakan untuk mengacu pada obyek non-fisik unik seperti waktu (jam 15.00 WIB), merk produk (Wardah), nama kegiatan (Pekan Olahraga Nasional 2020), dan sebagainya. PEB merupakan sistem yang mengidentifikasi EB pada sebuah teks untuk selanjutnya mengelompokkan EB tersebut ke dalam kelas yang sesuai. Sebagai contoh, pada kalimat *Nil Maizar, pelatih Semen Padang menanti aksi konkrit pemerintah yang sudah membekukan PSSI demi perbaikan organisasi sepakbola tanah air itu.*, frase *Nil Maizar* dan *PSSI* mengacu pada obyek unik orang dan organisasi tertentu. Oleh karena itu, sebuah PEB akan mengidentifikasi keduanya sebagai EB dengan kelas masing-masing yakni nama orang dan nama organisasi.

### Evaluasi Pengenalan Entitas Bernama

Secara umum, evaluasi sistem PEB dilakukan dengan membandingkan ha- sil identifikasi EB dan kelasnya dengan data di dataset pedoman. Terdapat 3 jenis teknik evaluasi sistem PEB meliputi teknik yang dipublikasikan pada konferensi *Message Understanding* (dikenal dengan MUC), *exact match*, dan teknik dari da- taset *Automatic Content Extraction* (dikenal dengan ACE). Berikut ini penjelasan masing-masing teknik tersebut.

* + - 1. Evaluasi MUC

Evaluasi MUC didasarkan pada dua komponen, yakni kemampuan sistem

17[https://www.nlm.nih.gov/databases/download/pubmed*medline.html*](http://www.nlm.nih.gov/databases/download/pubmedmedline.html)

mengidentifikasi EB dengan kelas yang benar (CT) dan kemampuan mene- mukan teks yang tepat (ET). CT terpenuhi ketika sistem mengidentifikasi EB dengan kelas yang benar tanpa tergantung pada cakupan EB tersebut. Dengan kata lain, jika EB berbentuk frase dan sistem mampu mengidentifikasi seba- gian dari frase tersebut, maka CT telah terpenuhi. Di sisi lain, ET fokus pada apakah sistem dapat menentukan kelas suatu EB dengan benar tanpa meli- hat apakah EB tersebut secara keseluruhan diidentifikasi oleh sistem. Dari CT dan ET, diperoleh 3 komponen penilaian, yakni jumlah jawaban benar (NCA), jumlah jawaban yang dikembalikan oleh sistem (NAS), dan jawaban pada dataset pedoman (NPE). Nilai akhir dari evaluasi MUC disebut *Micro Averaged F-Measure* (MAF) diperoleh dari rata-rata nilai *precision* dan *reca-*

*ll*. *Precision* berasal dari NCA/NAS, dan *recall* diperoleh dari NCA/NPE.

* + - 1. Exact Match

Terdapat 2 sub teknik yang digunakan pada teknik *Exact Match* (EM), yakni teknik dari IREX (*Information Retrieval and Extraction Exercise*) dan Co- NLL (*Conference of Natural Language Learning*). Teknik EM menggunakan metrik yang sama dengan MUC yaitu MAF, namun dengan perolehan *preci- sion* dan *recall* yang berbeda. *Precision* merupakan persentase EB benar yang ditemukan oleh sistem, sedangkan *recall* mengacu pada persentase EB benar yang ditemukan oleh sistem serta muncul di dataset pedoman. Dalam hal ini, sebuah EB dinyatakan benar jika berasal dari hasil benar pada perbandingan secara tepat terhadap EB di dataset pedoman.

* + - 1. Evaluasi ACE

ACE merupakan teknik evaluasi yang kompleks dibandingkan dua teknik se- belumnya. Teknik ini dilengkapi dengan komponen seperti hasil perbanding- an sebagian (*partial match*) dan EB benar dengan kelas yang salah (*wrong type*). Lebih lanjut, setiap kelas EB memiliki bobot yang ditentukan di awal yang membentuk nilai akhir secara keseluruhan. Misalnya, jika kelas *Person* diberi bobot 1 dan *Location* memiliki skor 0.5, maka pada penghitungan nilai akhir, 1 EB dari kelas *Person* harus didampingi oleh 2 EB dari kelas *Loca- tion*. ACE juga memungkinkan pemberian skor sangat rendah (penalti) bagi suatu EB untuk mengakomodir EB yang tidak terdeteksi (*missed entity*) dan kesalahan pada kelas EB. Nilai akhir pada ACE disebut dengan EDR (*Entity Detection and Recognition*) yang diperoleh dengan mengurangi skor 100% dengan nilai penalti.

* 1. **Pemanfaatan Metode *Semi-supervised Learning* untuk Pengenalan Entitas Bernama**

Dalam pembelajaran mesin (machine learning) terdapat empat metode ben- tuk pembelajaran yakni 1) *Unsupervised learning*, suatu metode model pembela- jaran dalam pembelajaran mesin yang menggunakan data-data yang memiliki label. Tujuan dari pembelajaran ini adalah dengan melakukan identifikasi terhadap suatu

data berdasarkan pola dan tren data 2) *Supervised learning*, dimana metode ini me- mungkinkan model dapat mempelajari dari data-data yang sudah memiliki label se- hingga model tersebut mampu melakukan prediksi terhadap data baru berdasarkan informasi yang didapat sebelumnya [[16].](#_bookmark19) 3) *Semi-supervised learning*, metode ini memungkinkan mesin dapat mempelajari pola suatu data yang melibatkan sekelom- pok data yang sudah memiliki label dan juga dari sebagian besar data yang belum memiliki label [18](#_bookmark0). Dan yang terakhir 4) *Reinforcement learning*, merupakan pola pembelajaran yang relatif baru dalam pembelajaran mesin yang mengoptimalkan pembelajaran berdasarkan *reward* jika model dapat memberikan informasi yang di- butuhkan, dan *punishment* jika model tersebut memberikan informasi yang keliru. Dengan begitu model dapat belajar dari hasil interaksi antara agent (model) dan enviroment (yang memberikan feedback) berdasarkan respon yang diberikan. Da- lam penelitian ini, kami menggunakan pendekatan pembelajaran *semi-supervised* yang saat ini sedang populer. Salah satu metode dari *semi-supervised learning* yang akan digunakan untuk pengenalan entitas bernama dalam penelitian ini adalah *Self- Training (ST)* [[17],](#_bookmark20) merupakan teknik yang menggunakan seperti *teacher model* yang sudah dilatih pada data yang memiliki label untuk melakukan prediksi pada data yang belum memiliki label dan kemudian melatih model baru (kita sebut seba- gai *student model*) dengan dataset dari hasil prediksi yang dilakukan oleh *teacher model*. Maksud dari teacher model tersebut adalah suatu model yang sudah dila- tih dengan pendekatan *supervised-learning*. Pada dasarnya, terdapat tiga langkah untuk menerapkan *semi-supervised learning* dengan menerapkan teknik *ST* seperti yang terlihat pada Gambar. [1.3.](#_bookmark3)



**Semi-supervised self-training method**

1

IndQNER

Teacher model

2

Lots of unlabeled data

Teacher model

Pseudo-labels

IndQNER

3

New dataset

Student model

Predictions

Most confident pseudo-labels

**Gambar 1.3** Teknik semi-supervised learning dengan menerapkan metode *ST*

18[https://machinelearningmastery*.*com/what-is-semi-supervised-learning/](https://machinelearningmastery.com/what-is-semi-supervised-learning/)

### Langkah pertama

Sebuah *teacher model* dilatih dengan menggunakan *supervised-learning* pa- da dataset yang memiliki label untuk domain spesifik IndQNER. Model yang di- kembangkan dapat memanfaatkan *neural networks* atau *language models* dengan bertujuan untuk dapat memprediksi *named entity* pada kata atau frasa yang berhu- bungan dengan terjermahan teks Al-Quran dalam bahasa Indonesia.

### Langkah kedua

Selanjutnya dengan menggunakan *teacher model* dilakukan proses prediksi *named-entity* pada dataset yang belum memiliki label. Hasil prediksi dataset yang belum memiliki label tersebut disebut dengan *pseudo-labels*.

### Langkah ketiga

Dikembangkan model baru yang disebut dengan *student model* yang dilatih dengan menggunakan gabungan dataset IndQNER dan dataset yang sudah memi- liki *pseudo-labels* dengan tingkat *confident* tertentu. Model *student* ini berfungsi sebagai *teacher model* baru yang digunakan untuk memprediksi *named-entity* pada kata atau frasa.

* 1. ***Large Language Models***

Model bahasa berukuran besar (*Large Language Models/LLMs*) merupakan algoritma pembelajaran mesin yang dapat mengenali, merangkum, menerjemahkan, memprediksi, dan menghasilkan bahasa manusia berdasarkan kumpulan data teks yang sangat besar. LLMs memungkinan untuk menyediakan imitasi bahasa manu- sia yang dihasilkan komputer paling meyakinkan hingga saat ini. Dalam beberapa tahun terakhir, penggunaan model bahasa skala besar (*Large Language Model*) se- perti BERT [[18],](#_bookmark21) ROBERTA [[19],](#_bookmark22) T5 [[20],](#_bookmark23) GPT-4 [[21],](#_bookmark24) dan LLAMA[19](#_bookmark0), banyak

digunakan dan berdampak sangat signifikan untuk berbagai macam aplikasi seperti sistem tanya jawab (*Question-Answering*) [[22]](#_bookmark25) dan peringkasan dokumen (*summa- rization*) [[23].](#_bookmark26) Model-model ini mampu mencapai performa yang sangat baik dalam mengerjakan tugas-tugas terkait bidang pemrosesan bahasa alamai (NLP) dan *artifi- cial intelligence (AI)*. Salah satu contoh yang sangat popular saat ini adalah aplikasi ChatGPT[20](#_bookmark0) yang mampu memanfaatkan LLM untuk mengerjakan berbagai tugas sehingga para pengguna dapat menjadikan aplikasi tersebut sebagai layaknya *per- sonal assisstant* virtual. Terdapat tiga jenis kategori terkait penggunaan LLM untuk menyelesaikan tugas-tugas dalam bidang NLP dan AI, yakni pertama sebagai *En- coder LLM*; kedua, *encoder-decoder LLM*, dan ketiga, *decoder LLM*.

19https://ai.meta.com/llama/ 20https://chat.openai.com/

## METODOLOGI PENELITIAN

Penelitian yang diusulkan ini akan dilakukan melalui beberapa tahapan yang dijelaskan pada bagian berikut.

1. Studi literatur

Tahapan ini bertujuan untuk mempelajari arsitektur pembangunan dataset pe- doman yang dilakukan secara otomatis dengan pendekatan *self-training*. Se- lain itu, survei juga akan dilakukan pada artikel penelitian yang menggunakan LLMs pada teks Bahasa Indonesia. Survei ini ditujukan untuk mengetahui 3 LLMs yang paling banyak digunakan dan memiliki kinerja baik, untuk se- lanjutnya ketiga LLMs tersebut akan digunakan untuk melakukan pelabelan otomatis pada teks.

1. Pembuatan pedoman verifikasi oleh pakar

Pedoman untuk melakukan verifikasi hasil pelabelan otomatis oleh model *self-training* dirancang agar seluruh pakar dapat memberikan keluaran yang standar. Selain itu, mekanisme identifikasi kandidat entitas bernama dan en- titas bernama juga akan didefinisikan mengingat terjemahan Al-Quran berba- hasa Indonesia dibuat dengan prosedur tertentu yang berbeda dibandingkan dengan terjemahan di bahasa lainnya.

1. Pelabelan otomatis dengan *self-training* menggunakan LLMs

Pelabelan otomatis dengan *self-training* akan dilakukan dalam 3 siklus bagi 3 LLMs yang akan digunakan. Pengukuran kinerja setiap LLM akan dilakukan untuk mengetahui LLM mana yang relevan bagi kompleksitas teks terjemah- an Al-Qur’an berbahasa Indonesia. Pengukuran kinerja ini akan dilakukan melalui metrik evaluasi *F1* berdasarkan nilai *precision* dan *recall*

1. Analisis hasil verifikasi

Setelah kegiatan pelabelan otomatis selesai, pakar akan memverifikasi hasil pelabelan secara manual. Terdapat 6 orang pakar yang berasal dari Lajnah Pentashihan Mushaf Al-Qur’an (LPMQ), Kementerian Agama Republik In- donesia, di mana 3 orang pakar akan memverifikasi hasil pelabelan dokumen yang sama. Hasil verifikasi ini akan dievaluasi menggunakan metrik *Interra- tor agreement*. Jika ketiga atau mayoritas pakar menyetujui hasil pelabelan tertentu, maka hasil tersebut dapat diterima. Sebaliknya, ketidaksesuaian *di- sagreement* hasil verifikasi pakar akan diselesaikan melalui diskusi langsung oleh seluruh pakar untuk menentukan hasil akhir verifikasi.

1. Evaluasi hasil pelabelan

Dataset E-IndQNER yang dihasilkan akan dievaluasi dengan prosedur yang sama dengan IndQNER. Dalam hal ini, 2 sistem PEB yang masing-masing dibangun dengan BiLSTM-CRF dan *fine-tuning* IndoBERT akan diuji meng- gunakan E-IndQNER. Lalu, pengukuran kinerja sistem akan dilakukan de- ngan metrik evaluasi *F1* berdasarkan nilai *precision* dan *recall*. Selain itu,

kualitas E-IndQNER juga akan dievaluasi untuk mengetahui seberapa besar dampaknya terhadap IndoBERT dalam mendukung sistem PEB pada domain umum.

## RENCANA PEMBAHASAN

Laporan dari penelitian yang akan dilaksanakan ini berisi komponen sebagai berikut:

1. BAB 1. Pendahuluan

Bab ini membahas mengenai latar belakang penelitian pembuatan dataset pe- doman E-IndQNER dari terjemahan Al-Qur’an berbahasa Indonesia. Selan- jutnya, bagian ini menjabarkan permasalahan yang akan diselesaikan pada penelitian. Hal yang akan dicapai pada penelitian dibahas di bagian tujuan penelitian di BAB 1 ini. Sebelum membahas bagian tujuan penelitian, ter- lebih dahulu akan dibahas batasan penelitian yang membahas ruang lingkup penelitian yang akan dilakukan, asumsi-asumsi penelitian jika diperlukan dan kendala-kendala penelitian yang dihadapi. Bagian terkahir dari penelitian ini adalah pembahasan manfaat penelitian.

1. BAB 2. Kajian Teori

Bab ini akan memaparkan kajian teori yang melandasi kegiatan penelitian ini meliputi metodologi pembuatan data pedoman pengenalan EB disertai de- ngan teori-teori yang melandasinya, termasuk pendekatan *self-training* serta LLMs terkait. Selain itu, pembahasan tentang PEB juga diberikan secara komprehensif.

1. BAB 3. Metodologi Penelitian

Bab ini akan menguraikan tahapan penelitian yang telah dijelaskan di sek- si/bagian 7. Tahapan tersebut meliputi studi literatur, pembuatan pedoman verifikasi oleh pakar, pelabelan otomatis dengan *self-training* menggunakan LLMs, analisis hasil verifikasi, dan evaluasi hasil pelabelan.

1. BAB 4. Hasil dan Pembahasan

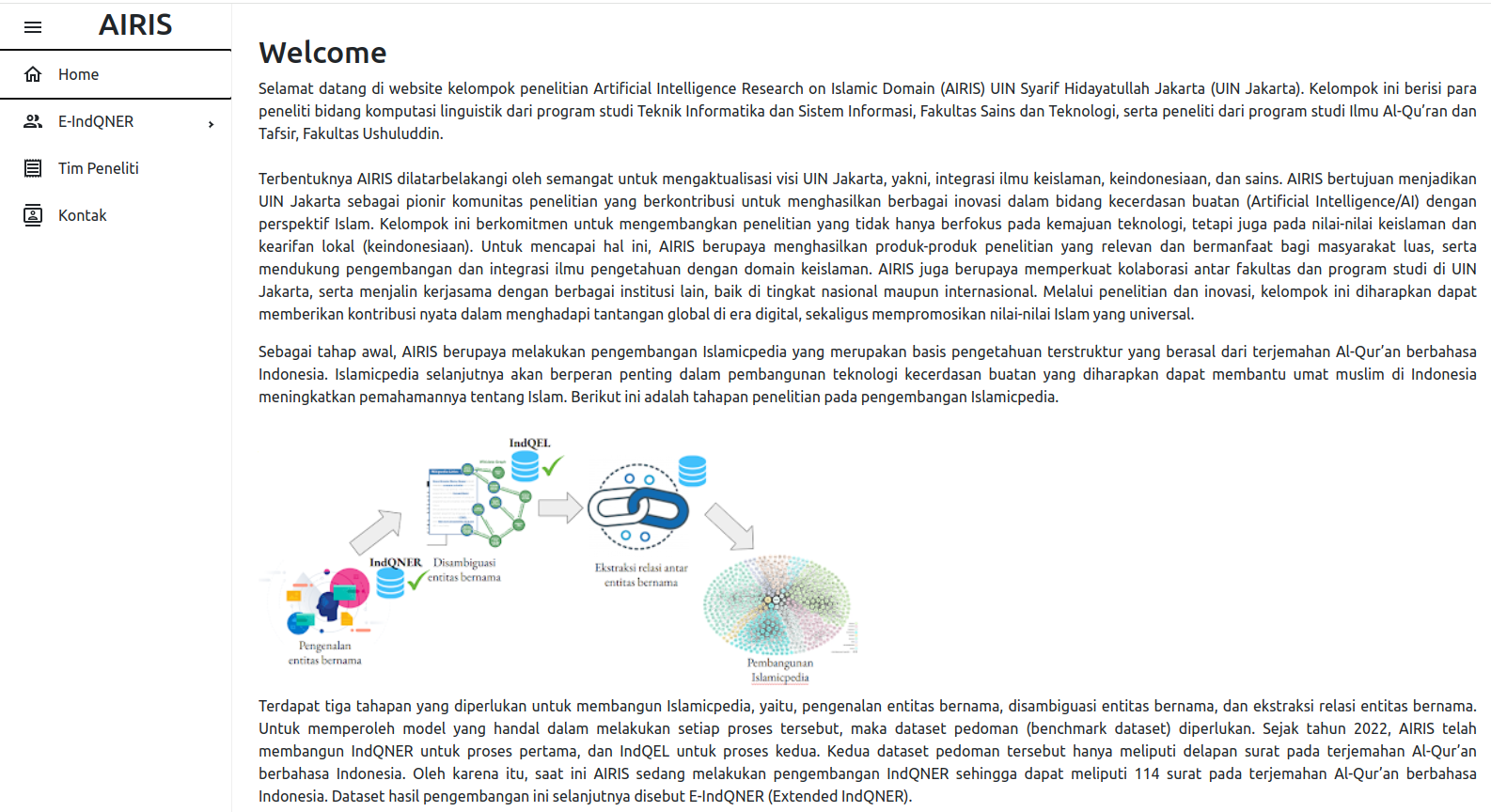
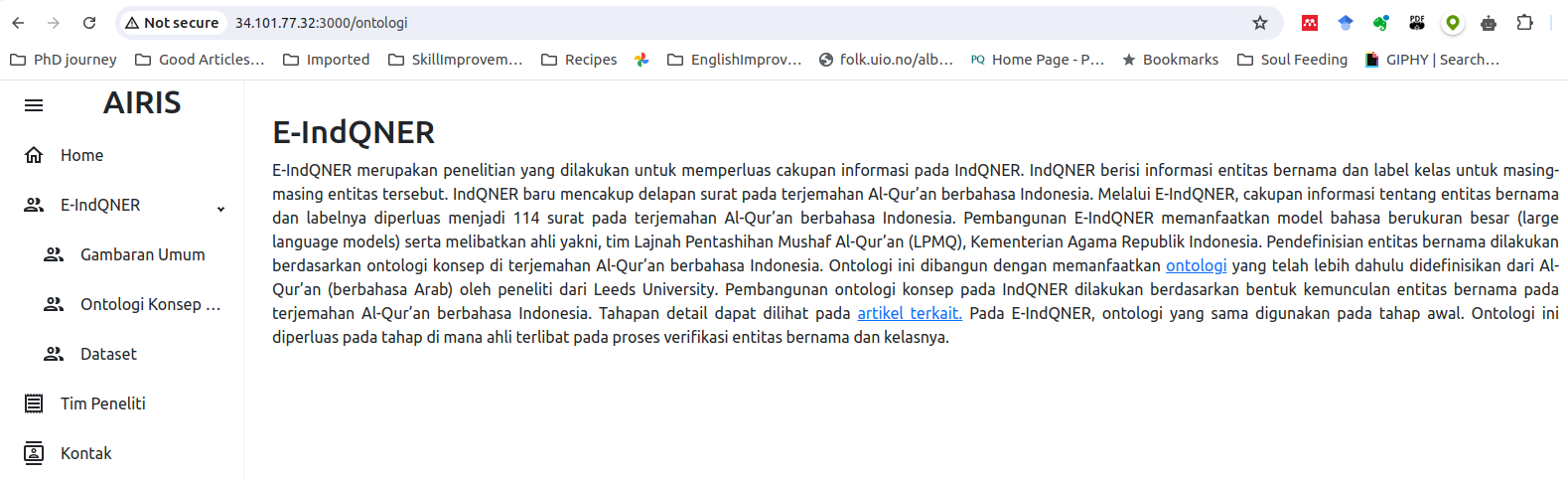
Bab 4 ini akan melaporkan apa yang dilakukan di setiap tahapan-tahapan pe- nelitian beserta hasil yang dicapai pada setiap tahapannya.

1. BAB 5. Penutup

Bab ini akan menyajikan kesimpulan dari penelitian yang telah dilakukan dan langkah yang akan dilakukan selanjutnya berdasarkan hasil evaluasi dari ke- giatan penelitian ini.

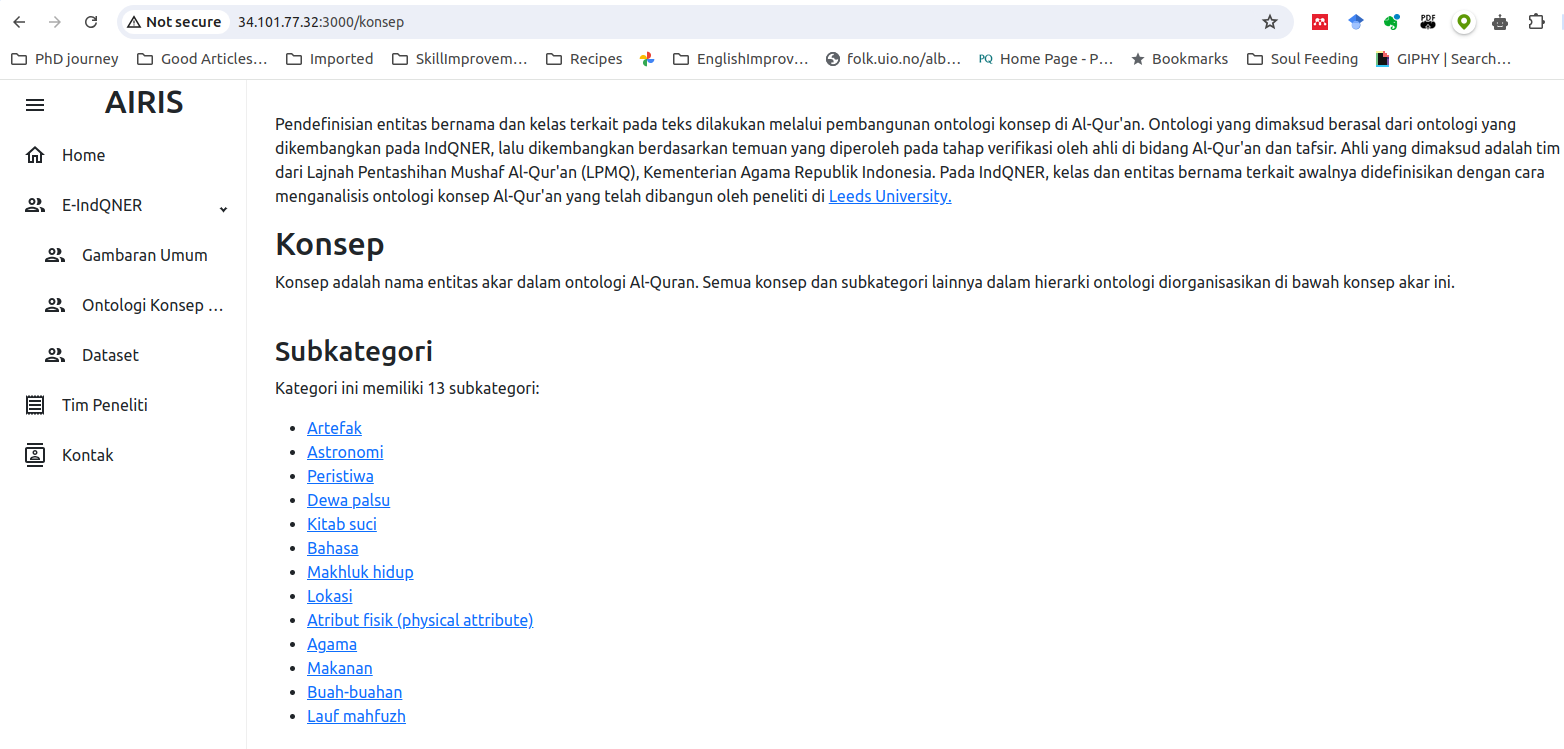
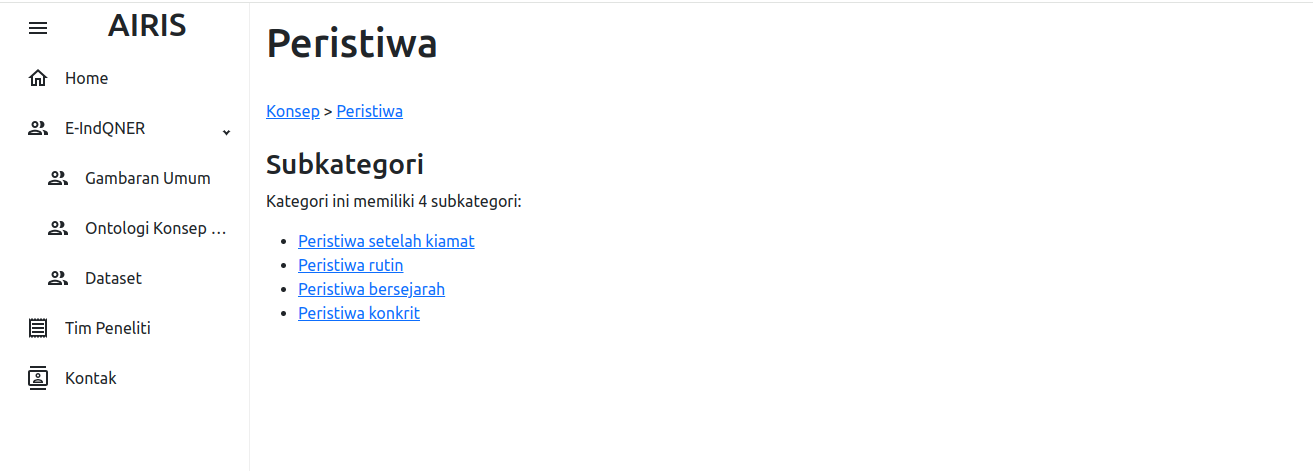
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **NO** | **KEGIATAN/SUB KEGIATAN/JENIS BELANJA** | **URAIAN VOLUME** | | | | | | | **VOL** | **SATUAN** | **HARGA SATUAN** | **SUB TOTAL** | **TOTAL** |
|  |
| **A** | **PRA PENELITIAN** |  |  |  |  |  |  |  |  |  |  |  | **2000000** |  |
| **A.1** | **Rapat Pembahasan Desain Operasional Penelitian** | | | | | | | | | | | | |  |
| 1 | Uang Saku Rapat Tim Penelitian | 10 | org | x | 2 | kali | x | 1 | bln | 20 | OK | 100000 | 2000000 |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| A.2 | **Pelaksanaan Konsultansi dengan Konsultan Expert** | | | | | | | | | | | | **41000000** |  |
| 1 | Honor Konsultan |  |  |  |  |  |  |  |  | 1 | Paket | 41000000 | 41000000 |  |
| 2 | Uang Saku FGD |  |  |  | 10 | kali | x | 2 | kali | 20 | OK | 100000 | 2000000 |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **TOTAL** | | | | | | | | | | | | | **45000000** |  |

# DOKUMENTASI



**Gambar 1.5 Output1\_bagian2**

**Gambar 1.4 Output1\_bagian1**

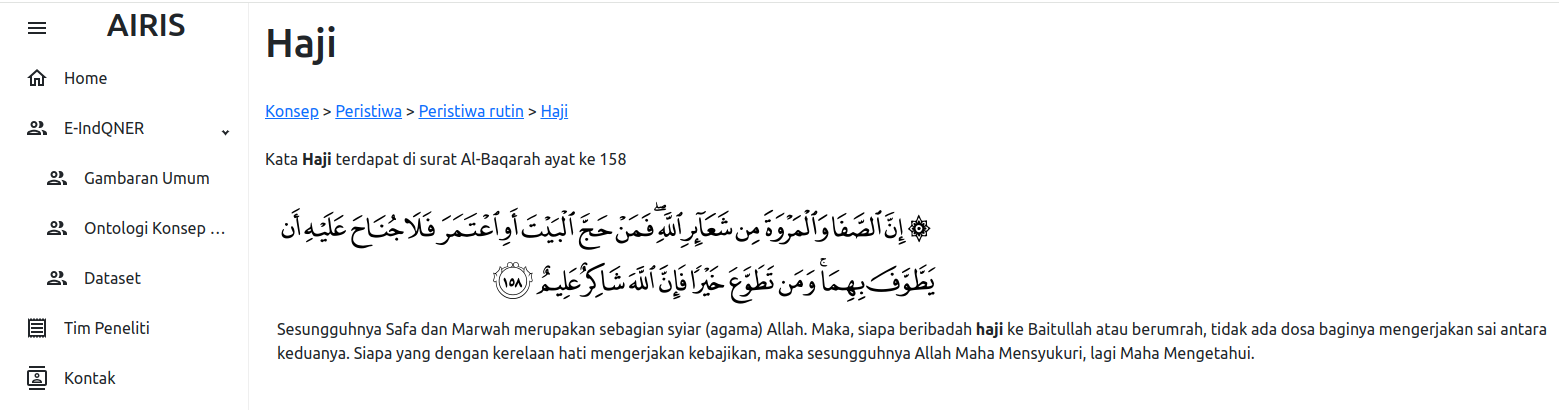


**Gambar 1.7 Output1\_bagian4**

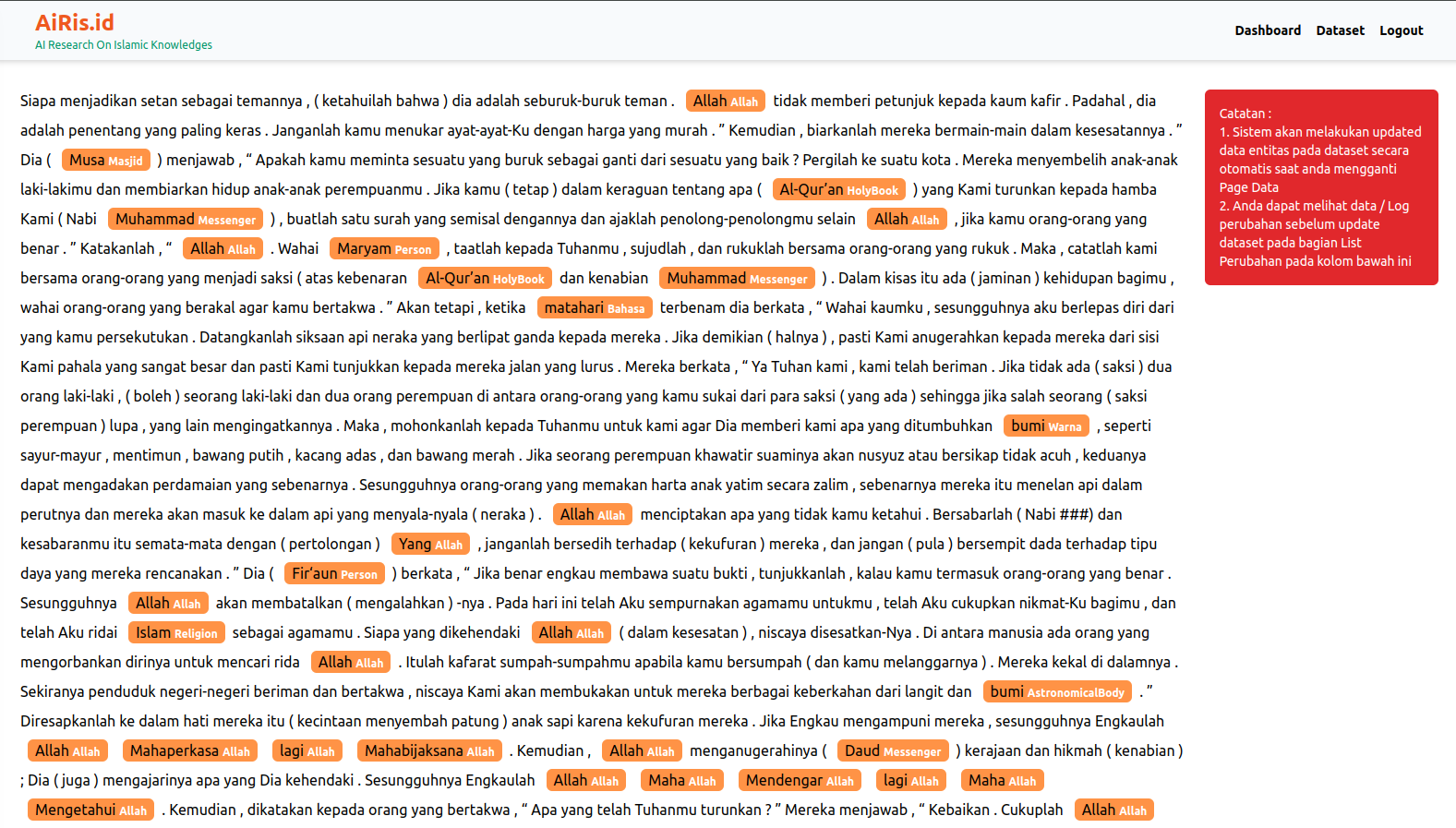
**Gambar 1.6 Output1\_bagian3**



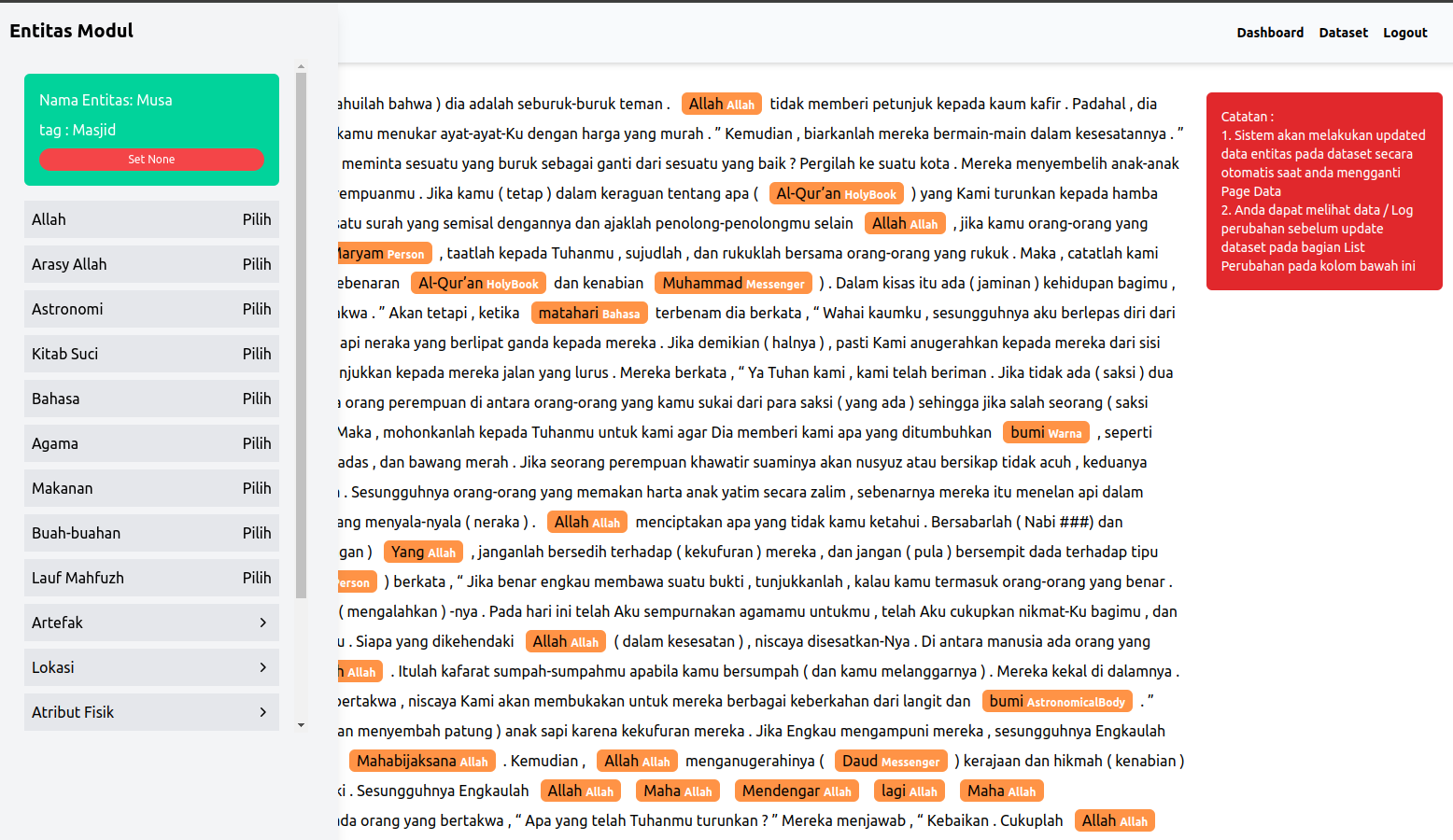
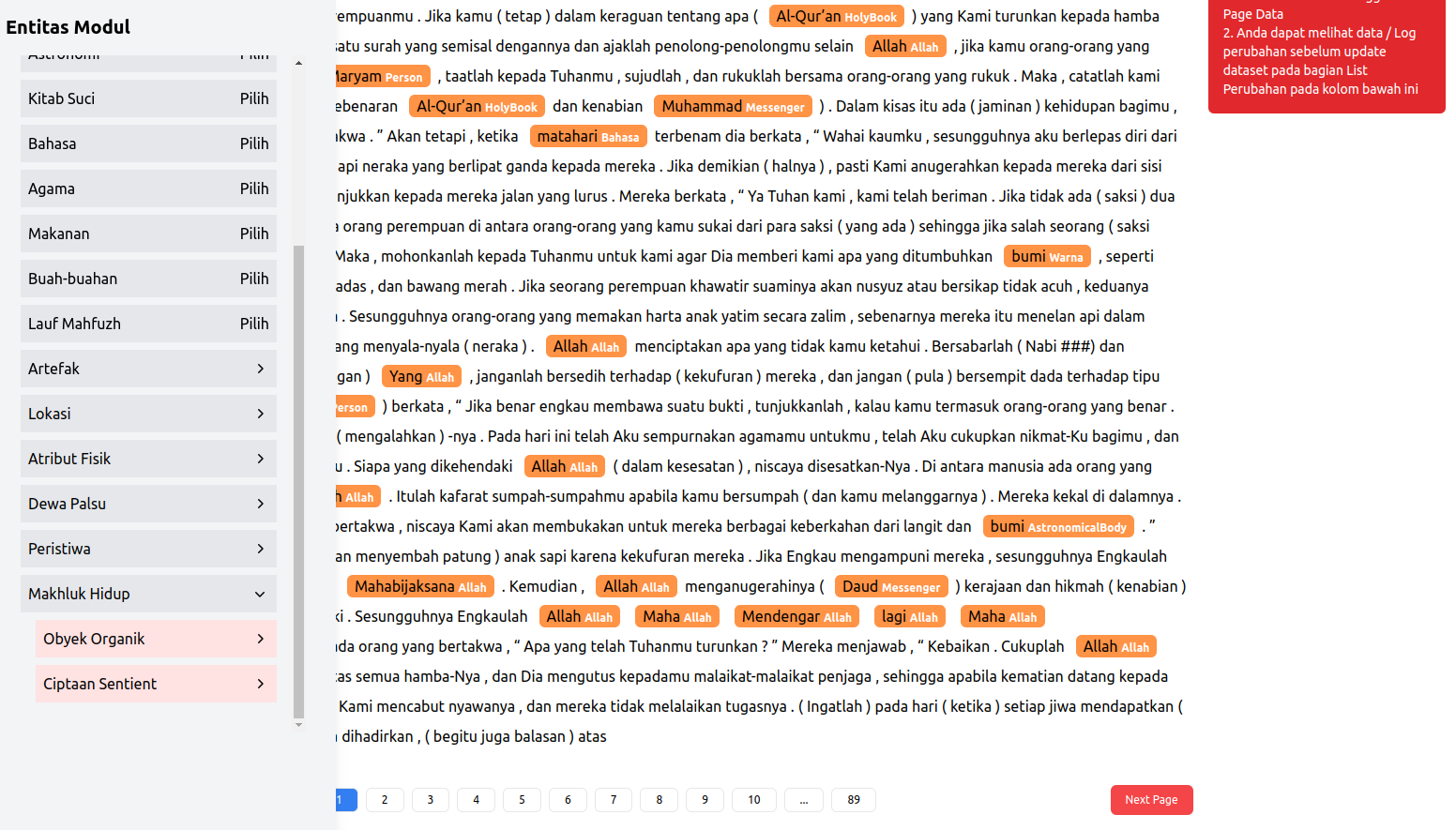
**Gambar 1.8 Output1\_bagian5**



**Gambar 1.9 Output1\_bagian6**



**Gambar 1.10 Output2\_bagian1**

****

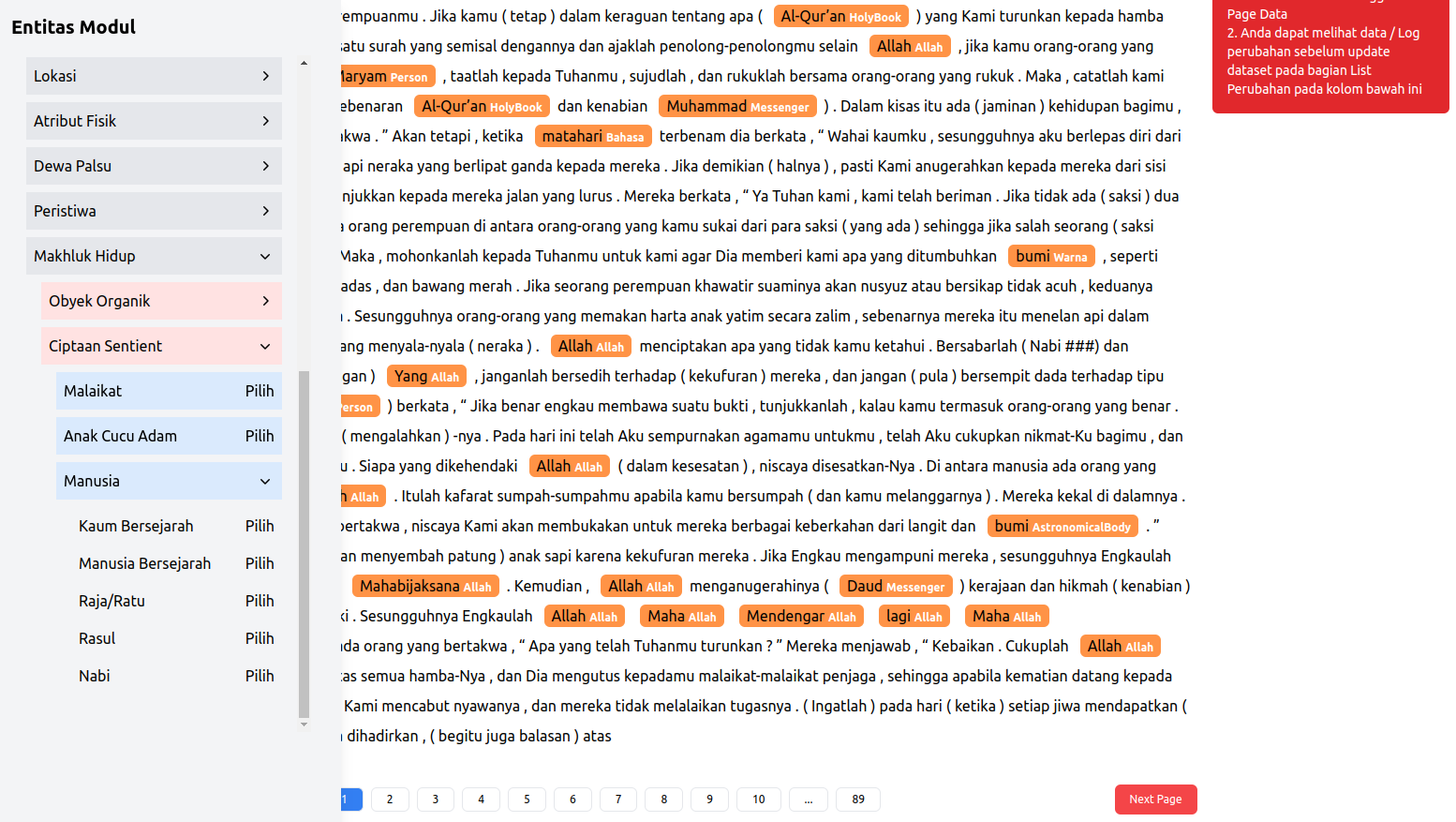
**Gambar 1.12 Output2\_bagian3**

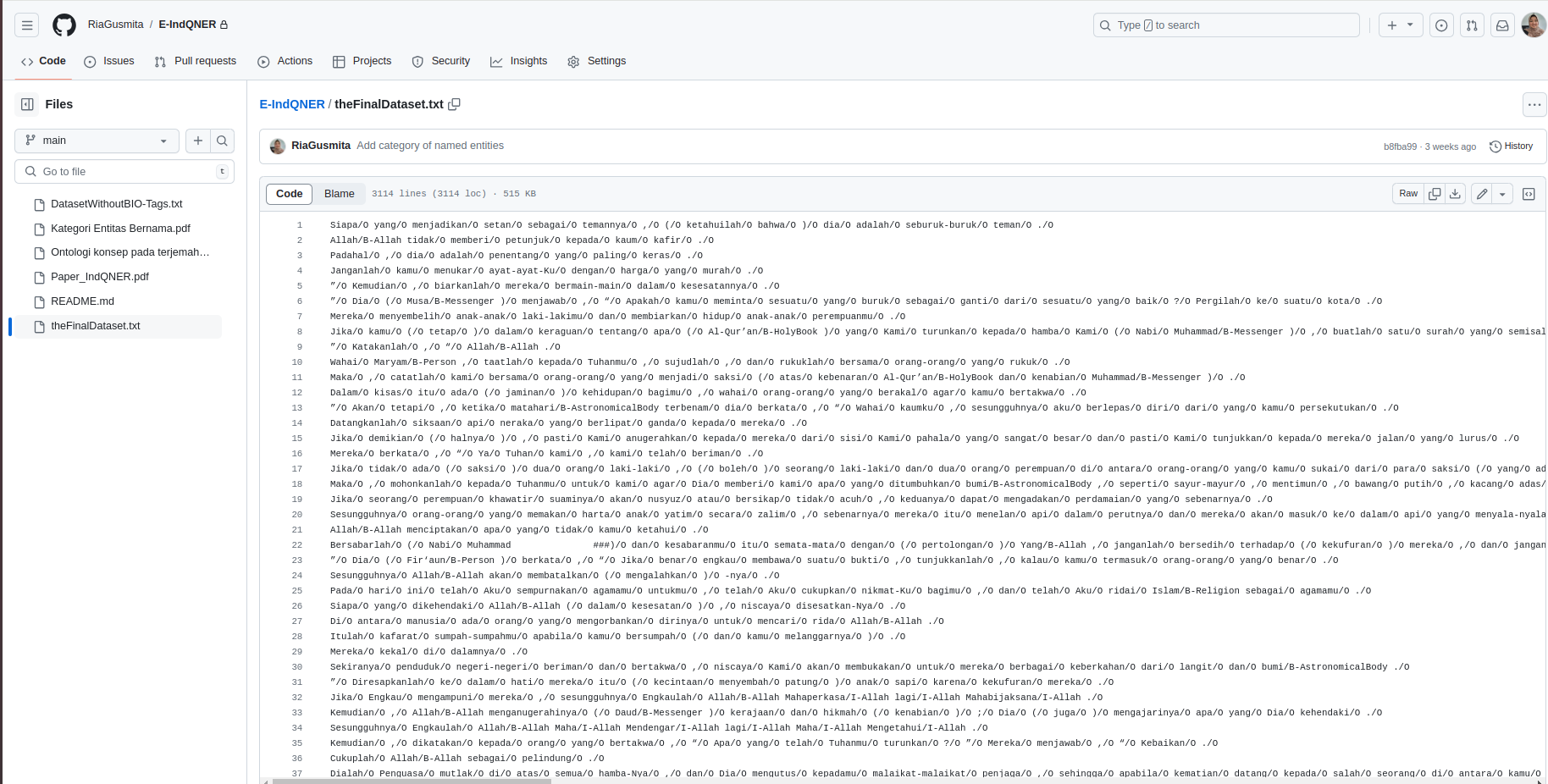
**Gambar 1.11 Output2\_bagian2**

****

**Gambar 1.14 Output2\_bagian5**

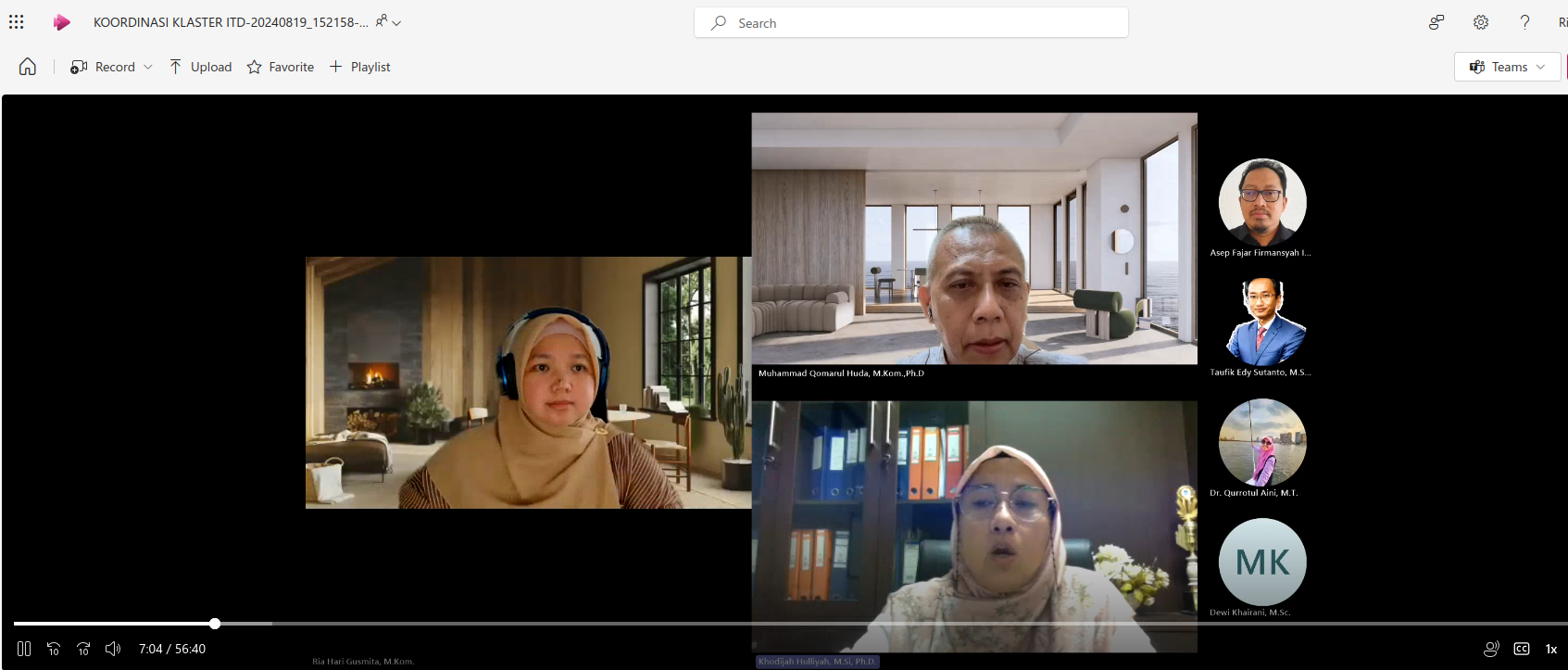
**Gambar 1.13 Output2\_bagian4**

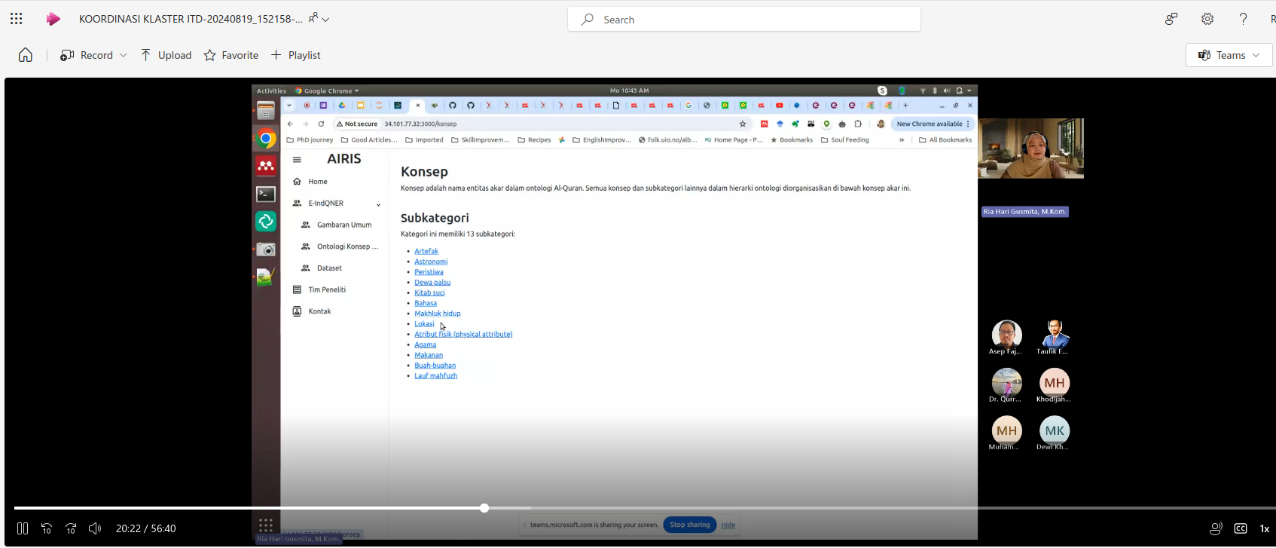
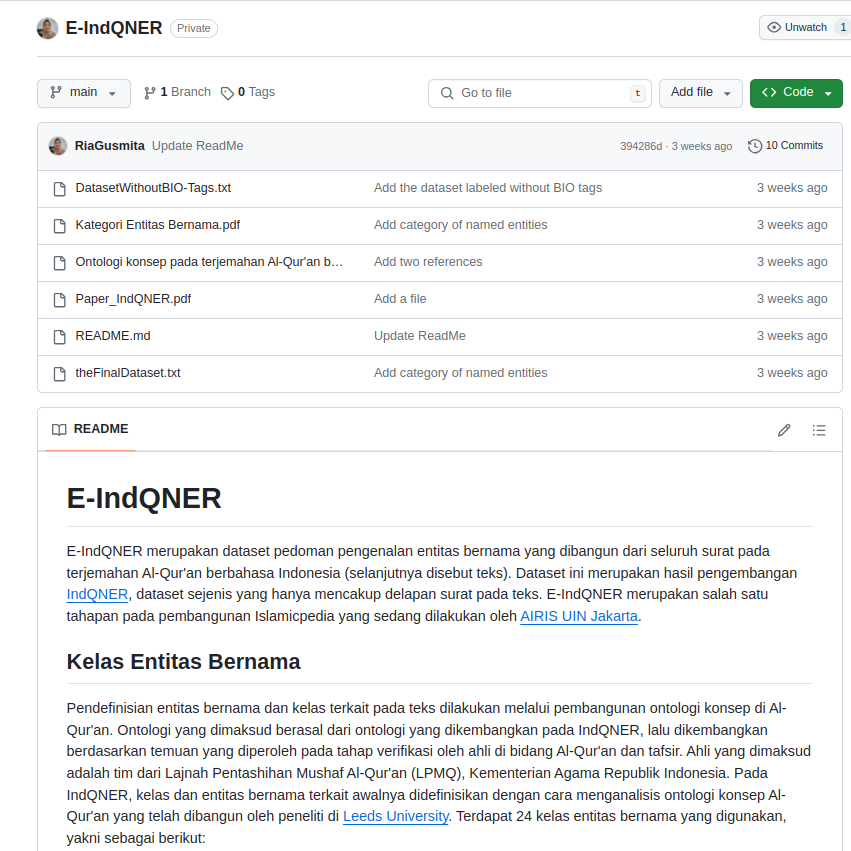
****

****

**Gambar 1.16 RapatKoordinasi\_bagian1**

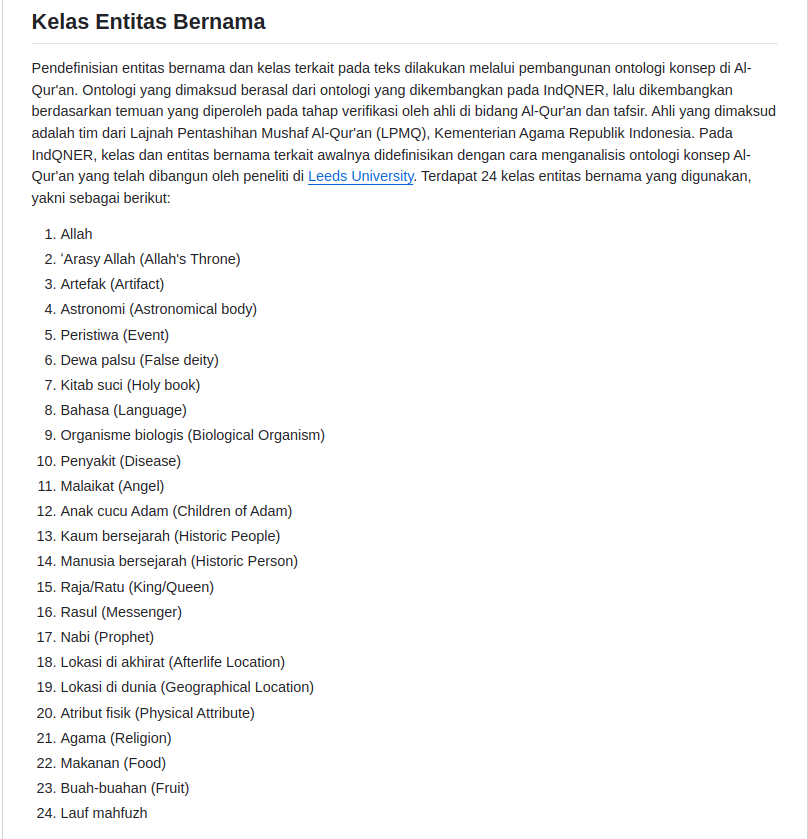
**Gambar 1.15 Dataset\_E-IndQNER**

****

****

**Gambar 1.18 Repositori\_bagian1**

**Gambar 1.17 RapatKoordinasi\_bagian2**

****

**Gambar 1.19 Repositori\_bagian2**

# DAFTAR PUSTAKA

1. B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. So- leman, R. Mahendra, P. Fung, S. Bahar, and A. Purwarianti, “IndoNLU: Ben- chmark and Resources for Evaluating Indonesian Natural Language Unders- tanding,” 9 2020.
2. F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “*{*I*}*ndo*{*LEM*}* and

*{*I*}*ndo*{*BERT*}*: A Benchmark Dataset and Pre-trained Language Model for

*{*I*}*ndonesian *{*NLP*}*,” in *Proceedings of the 28th International Conference on Computational Linguistics*, (Barcelona, Spain (Online)), pp. 757–770, In- ternational Committee on Computational Linguistics, 2020.

1. R. H. Gusmita, A. F. Firmansyah, D. Moussallem, and A.-C. Ngonga Ngomo, *IndQNER: Named Entity Recognition Benchmark Dataset from the Indonesian Translation of the Quran*, vol. 2. Springer Nature Switzerland, 2023.
2. I. Alfina, R. Manurung, and M. I. Fanany, “DBpedia entities expansion in automatically building dataset for Indonesian NER,” *2016 International Con- ference on Advanced Computer Science and Information Systems, ICACSIS 2016*, pp. 335–340, 2017.
3. A. Luthfi, B. Distiawan, and R. Manurung, “Building an Indonesian named entity recognizer using Wikipedia and DBPedia,” *Proceedings of the Interna- tional Conference on Asian Language Processing 2014, IALP 2014*, pp. 19– 22, 2014.
4. R. A. Leonandya, B. Distiawan, and N. H. Praptono, “A Semi-supervised Al- gorithm for Indonesian Named Entity Recognition,” *Proceedings - 2015 3rd International Symposium on Computational and Business Intelligence, ISCBI 2015*, no. March, pp. 45–50, 2016.
5. I. Alfina, S. Savitri, and M. I. Fanany, “Modified DBpedia entities expan- sion for tagging automatically NER dataset,” *2017 International Conferen- ce on Advanced Computer Science and Information Systems, ICACSIS 2017*, vol. 2018-Janua, pp. 216–221, 2018.
6. S. O. Khairunnisa, A. Imankulova, and M. Komachi, “Towards a Standardi- zed Dataset on *{*I*}*ndonesian Named Entity Recognition,” in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Compu- tational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop*, pp. 64–71, Association for Computational Linguistics, 2020.

15

1. Y. Syaifudin and A. Nurwidyantoro, “Quotations identification from Indone- sian online news using rule-based method,” *Proceeding - 2016 Internatio- nal Seminar on Intelligent Technology and Its Application, ISITIA 2016: Re- cent Trends in Intelligent Computational Technologies for Sustainable Energy*,

pp. 187–194, 2017.

1. N. Chinchor and B. Sundheim, “*{*MUC*}*-5 Evaluation Metrics,” in *Fifth Mes- sage Understanding Conference ({MUC}-5): Proceedings of a Conference Held in Baltimore, {M}aryland, August 25-27, 1993*, 1993.
2. N. Chinchor and P. Robinson, “MUC-7 Named Entity Task Definition,” *Pro- ceedings of the Sixth Message Understanding Conference MUC6*, no. Septem- ber, p. 21, 1997.
3. E. F. Tjong Kim Sang, “Introduction to the *{*C*}*o*{*NLL*}*-2002 Shared Task: Language-Independent Named Entity Recognition,” in *{COLING}-02: The 6th Conference on Natural Language Learning 2002 ({C}o{NLL}-2002)*, 2002.
4. T. Al-Moslmi, M. G. Oca*{\*˜n*}*a, A. L. Opdahl, and C. Veres, “Named Enti- ty Extraction for Knowledge Graphs: A Literature Overview,” *IEEE Access*, vol. 8, pp. 32862–32881, 2020.
5. J. Jurafsky, Daniel; Martin, “An introduction to natural language processing through prolog,” *Speech and Language Processing: An Introduction to Na- tural Language Processing, Computational Linguistics, and Speech Recogni- tion*, vol. 2, no. May, pp. 1–306, 2008.
6. J. Li, A. Sun, J. Han, and C. Li, “A Survey on Deep Learning for Named Entity Recognition,” *IEEE Transactions on Knowledge and Data Engineering*, pp. 1– 1, 2020.
7. Q. Liu and Y. Wu, *Supervised Learning*, pp. 3243–3245. Boston, MA: Spri- nger US, 2012.
8. Z. Shi, F. Tonolini, N. Aletras, E. Yilmaz, G. Kazai, and Y. Jiao, “Rethinking semi-supervised learning with language models,” in *Findings of the Associa- tion for Computational Linguistics: ACL 2023* (A. Rogers, J. Boyd-Graber, and N. Okazaki, eds.), (Toronto, Canada), pp. 5614–5634, Association for Computational Linguistics, July 2023.
9. J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” *CoRR*, vol. abs/1810.04805, 2018.
10. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zett- lemoyer, and V. Stoyanov, “Roberta: A robustly optimized BERT pretraining approach,” *CoRR*, vol. abs/1907.11692, 2019.
11. C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou,

W. Li, and P. J. Liu, “Exploring the limits of transfer learning with a unified text-to-text transformer,” *J. Mach. Learn. Res.*, vol. 21, jan 2020.

1. OpenAI, “GPT-4 technical report,” *CoRR*, vol. abs/2303.08774, 2023.
2. M. Zaib, D. H. Tran, S. Sagar, A. Mahmood, W. E. Zhang, and Q. Z. Sheng, “Bert-coqac: Bert-based conversational question answering in context,” in *Pa- rallel Architectures, Algorithms and Programming - 11th International Sym- posium, PAAP 2020, Shenzhen, China, December 28-30, 2020, Proceedings* (L. Ning, V. Chau, and F. C. M. Lau, eds.), vol. 1362 of *Communications in Computer and Information Science*, pp. 47–57, Springer, 2020.
3. T. S. Barros, C. E. S. Pires, and D. C. do Nascimento, “Leveraging BERT for extractive text summarization on federal police documents,” *Knowl. Inf. Syst.*, vol. 65, no. 11, pp. 4873–4903, 2023.