Lauh mahfuzAnnotation Guideline of the Creation of Indonesian Named Entity Recognition (NER) Benchmark Dataset from the Indonesian Translation of the Quran

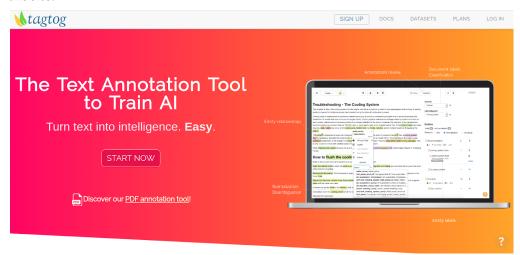
- 1. A Named Entity (NE) is a name on text that refers to a specific object in the world. NEs exist in the form of proper nouns, and they belong to particular classes, such as, *Person*, *Location*, *Organization*, etc. For example:
 - a. Salah satu perguruan tinggi yang menjunjung tinggi nilai keislaman adalah Universitas Islam Negeri Syarif Hidayatullah Jakarta, atau UIN Syarif Hidayatullah Jakarta. (One of universities that implement Islamic values is The State Islamic University Syarif Hidayatullah Jakarta or UIN Syarif Hidayatullah Jakarta.)
 - b. **Roemah Rempah** bersebelahan dengan salah satu restoran terekstrim sampai saat ini, **Katsura**. (**Roemah Rempah** is located adjacent to **Katsura**, one of the most extreme restaurants.)
 - c. **Syahrini** tidak lagi eksis di dunia entertainment yang telah membesarkan namanya. (**Syahrini** is no longer in the entertainment industry, where she gained fame.)

In the above sentences, phrases in bold are NEs since each of them directly refers to a particular object in the world. They are a particular organization (The State Islamic University Syarif Hidayatullah Jakarta and UIN Syarif Hidayatullah Jakarta), a particular place (Roemah Rempah and Katsura), and a particular person (Syahrini).

- 2. Every NE on text is labeled with a corresponding class. For example:
 - a. Salah satu perguruan tinggi yang menjunjung tinggi nilai keislaman adalah Universitas Islam Negeri Syarif Hidayatullah Jakarta<Organization>, atau UIN Syarif Hidayatullah Jakarta<Organization>. (One of universities that implement Islamic values is The State Islamic University Syarif Hidayatullah Jakarta<Organization> or UIN Syarif Hidayatullah Jakarta<Organization>.)
 - b. Roemah Rempah<Location> bersebelahan dengan salah satu restoran terekstrim sampai saat ini, Katsura<Location>. (Roemah Rempah<Location> is located adjacent to Katsura<Location>, one of the most extreme restaurants.)
 - c. Syahrini<Person> tidak lagi eksis di dunia entertainment yang telah membesarkan namanya. (Syahrini<Person> is no longer in the entertainment industry, where she gained fame.)

Organization, Location, and Person, are labels from the corresponding classes, i.e., Organization, Location, and Person, respectively.

- 3. The annotation process is performed using a web-based text annotation tool, Tagtog. These are the steps to perform annotation using Tagtog.
 - a. Visit https://tagtog.net/ on a web browser. This is the display of the home page of the site:



b. Do log in by entering the respective *username* and *password*.



c. Choose the corresponding annotation project. For example, NER-Benchmark-From-AlBaqarah



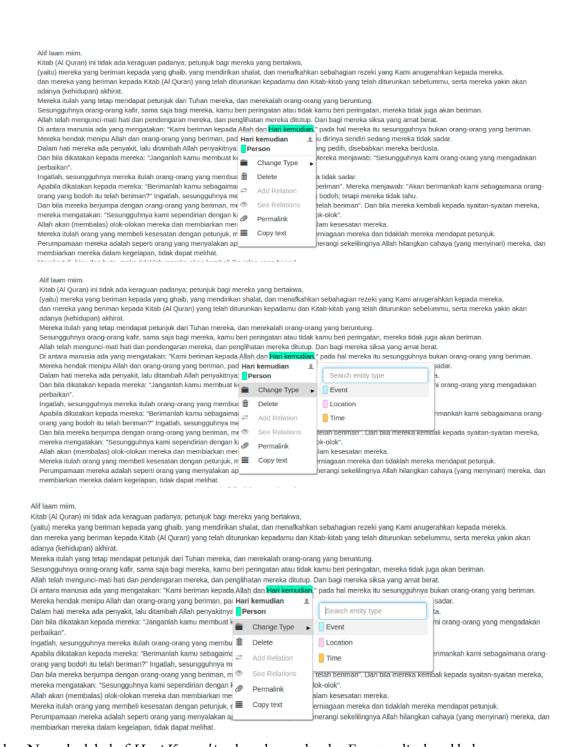
d. Choose a corpus that will be annotated. For example, Al-Baqarah Verses.txt.



- e. The corpus contains a collection of sentences in the Indonesian language whose NEs will be annotated with corresponding labels.
- f. When an NE is found, select the NE. The first selection on Tagtog will be automatically colored with the color of the first label option. For example, the first label option is Person, represented by green color



g. In the example of point f, *Hari Kemudian* (the judgement day) is an NE from *Event* class. Since the selection is the first and the NE is automatically labeled with *Person* (green color), then we need to change the label into *Event*. This is done by clicking the *Change Type* option, followed by selecting the corresponding label (*Event*).



h. Now the label of *Hari Kemudian* has changed to be *Event* as displayed below:



4. There are 20 NE classes options that are used to annotate the Indonesian translation of the Quran. These classes are derived from the Quran concepts ontology that was created at Leeds University. The following are list of all classes, as well as the corresponding labels,

descriptions, and sample of NEs. The descriptions are taken from https://corpus.quran.com/concept.jsp?id=concept.

Table 1. List of all classes and the corresponding labels, descriptions, and sample of NEs $\,$

NE Class	Label	Description	Sample of NEs
Allah	Allah	Allah (God in Islam) and all Allah's names that are known as Asmaul Husna	 Allah Tuhan Yang Maha Esa (The Unique, The Only One) Yang Maha Pengasih (The Most or Entirely Merciful)
Allah's Throne	Throne of Allah	The seat of Allah's power and authority	'Arasy (Allah's Throne)
Artifact	Artifact	Man-made constructions that are mentioned in the Quran	 Ka'bah (Kaaba) Masjidilaqsa (Al-Aqsa mosque) Masjidilharam (Masjid al-Haram) Tabut (the chest)
Astronomi cal body	Astronom ical body	Astronomical objects that are mentioned in the Quran	 Bintang Syi'ra (Sirius) Bumi (the earth) Bulan (the moon) Matahari (the sun)
Event	Event	Temporal events	 Hari kiamat (day of resurrection) Hari akhir (the last day) Subuh (dawn) Lailatulqadar (Night of Decree) Jum'at/hari Jum'at (Friday) Haji (Hajj) Bulan Ramadan/Ramadan (Ramadan) Hari Sabat (Sabbath) Umrah (Umra) Jahiliah (Al-Jahiliyah)
False deity	False deity	The worship of false gods mentioned in the Quran	 Al-'Uzza (Al-Uzza) Al-Lata (Allat) Wadd (Wadd) Suwā' (Suwa) Ya'ūq (Yauq) Nasr (Nasr) Manata (Manat) Ba'l (Baal)

			9. <i>Yagūs</i> (Yaghuth)
Holybook	Holy book	Holy books and other religious texts that are mentioned in the Quran	 Al-Qur'an (Quran) Injil (Injeel) Taurat (Torah) Zabur (Zabur)
Language	Language	The languages mentioned in the Quran	Bahasa Arab (Arabic)
Angel	Angel	The creations of Allah mentioned in the Quran, known as angels	 Malaikat maut (Azrael) Harut (Harut) Marut (Marut) Jibril (Jibreel) Malik (Malik) Mikail (Mikaeel)
Person	Person	Individual human beings or groups of people mentioned in the Quran	 Anak cucu Adam (Children of Adam) Kaum 'Ad/'Ad (Aad) Orang-orang Ansar (Ansar) Orang-orang Arab Badui (Bedouin) Bani Israil (Children of Israel) Penghuni gua (companions of the cave) Pasukan bergajah (companions of the elephant) Orang-orang yang membuat parit (companions of the pit) Penduduk Rass (companions of the Rass) Penduduk Aikah (companions of the wood) Ya'juj dan Ma'juj (Gog and Magog) Penduduk Madyan (Madyan) Penduduk negeri Hijr (people of al-Hijr) Orang-orang Quraisy (Quraysh) Bangsa Romawi (Romans) Samud/(kaum) Samud (Thamud) Kaum Tubba' (Tubba)

			1
			18. Habil (Abel) 19. Azar (Azar) 20. Qabil (Cain) 21. Zulkifli (Dhul Kifl) 22. Zulqarnain (Dhul Qarnayn) 23. Haman (Haman) 24. Imran (Imran) 25. Jalut (Jalut) 26. Luqman (Luqman) 27. Maryam (Maryam) 28. Fir'aun (Pharaoh) 29. Qarun (Qarun) 30. Samiri (Samiri) 31. Talut (Talut) 32. Uzair (Uzair)
Messenger	Messenger	The messengers of Allah mentioned in the Quran	 Daud (David) Ilyas (Elijah) Hud (Hud) Ibrahim (Ibrahim) Ismail (Ishmael) Isa putra Maryam/Isa Almasih putra Maryam (Jesus, messiah,) Lut (Lut) Muhammad (Muhammad) Musa (Musa) Nuh (Nuh) Saleh (Salih) Syuʻaib (Shuayb) Yunus (Yunus)
Prophet	Prophet	The prophets of Allah mentioned in the Quran	1. Adam (Adam) 2. Ahmad (Ahmad) 3. Ayyub (Ayyub) 4. Ilyasa' (Elijah) 5. Harun (Harun) 6. Idris (Idris) 7. Ishaq (Isaaq) 8. Israil (Israel) 9. Sulaiman (Solomon) 10. Yahya (Yahya) 11. Ya'qub (Yaqub) 12. Yusuf (Yusuf) 13. Zakaria (Zechariah)

Beast of the earth	Sentient	The sentient creation mentioned in the Quran	Makhluk bergerak dari bumi (creature from the earth)
Afterlife Location	Afterlife Location	Locations in the afterlife	 Surga Firdaus (Firdous) Surga 'Adn (Garden of eden) Salsabil (Salsabil) Sidratulmuntaha (Lote tree) Saqar (Saqar) Sijjīn (Sijjeen) Pohon zaqum (Zaqqum)
Geographic al location	Geographi cal location	Geographical locations mentioned in the Quran	1. Negeri Babilonia (Babylon) 2. Badar (Badr) 3. Bakkah (Bakkah) 4. Makkah (Makkah) 5. Hunain (Hunayn) 6. Iram (Iram) 7. Madinah (Medinah) 8. Yasrib (Yathrib) 9. Ahqaf (Al-Ahqaf) 10. Mesir (Egypt) 11. Negeri Saba' (Sheba) 12. Gunung Judiy (Al-Judi) 13. Marwah (Al-Marwah) 14. Safa (Al-Safa) 15. Arafah (Mount Arafat) 16. Bukit Sinai (Mount Sinai)
Color	Color	The different colors that are mentioned in the Quran	 Hijau (green) Kuning tua (yellow) Merah padam (dark) Putih (white)
Religion	Religion	The major religions, or other systems of ancient belief, that are mentioned by name in the Quran	 Islam (Islam) Nasrani (Christian) Yahudi (Judaism) Majusi (Magians) Sabiin (Sabians)
Food	Food	The food mentioned in the	 Manna (Manna) Salwa (Quails)

		Quran	
Fruit	Fruit	The fruit mentioned in the Quran	1. Kurma (Palm) 2. Anggur (Grave)
The book of Allah	The book of Allah	The book of Allah mentioned in the Quran	Lauh mahfuz (Clear record)

- 5. The annotation process is performed following the below steps:
 - a. Find NEs in Table 1 on the Indonesian translation of the Quran. Then, label them accordingly.
 - b. NEs from *Person*, Messenger, and Prophet have an additional rule. If their appearance is accompanied by a title, then the title is excluded in the annotation. Example:
 - i. Jika kamu (tetap) dalam keraguan tentang apa (Al-Qur'an) yang Kami turunkan kepada hamba Kami (Nabi Muhammad), buatlah satu surah yang semisal dengannya dan ajaklah penolong-penolongmu selain Allah, jika kamu orang-orang yang benar. (And if you are in doubt about what We have sent down upon Our Servant [Muhammad], then produce a Surah the like thereof and call upon your witnesses other than Allah, if you should be truthful) (2:23)

The verse contains an NE from the *Messenger* class that is written with its title, i.e., *Nabi Muhammad* (Prophet Muhammad). Nabi is the title, and therefore, it is not annotated. In other words, the word *Muhammad* is the only one labeled as a messenger.

- c. NEs can appear in the form of their synonyms that are not in Table 1. A synonym of an NE is an NE, and therefore, it has to be labeled. In the Indonesian translation of the Quran, synonyms are generally written in brackets. Examples:
 - i. Katakanlah (Nabi Muhammad), "Ruhulkudus (Jibril) menurunkannya (Al-Qur'an) dari Tuhanmu dengan hak untuk meneguhkan (hati) orang-orang yang telah beriman dan menjadi petunjuk serta kabar gembira bagi orang-orang muslim (yang berserah diri kepada Allah)." (Say, [O Muhammad], "The Pure Spirit has brought it down from your Lord in truth to make firm those who believe and as guidance and good tidings to the Muslims.") (16:102)

ii. (Ingatlah) ketika Isa putra Maryam berkata, "Wahai Bani Israil, sesungguhnya aku adalah utusan Allah kepadamu untuk membenarkan kitab (yang turun) sebelumku, yaitu Taurat, dan memberi kabar gembira tentang seorang utusan Allah yang akan datang setelahku yang namanya Ahmad (Nabi Muhammad)." Akan tetapi, ketika utusan itu datang kepada mereka dengan membawa bukti-bukti yang nyata, mereka berkata, "Ini adalah sihir yang nyata." (Whatever is in the heavens and whatever is on the earth exalts Allah, and He is the Exalted in Might, the Wise.) (61:1)

In the first example, *Jibril*, an NE, is written in brackets after *Ruhulkudus*. In this case, *Ruhulkudus* is a synonym of *Jibril*. In the second example, an NE, *Nabi Muhammad*, is written in brackets after *Ahmad*. In this case, Ahmad is a synonym of *Muhammad*. Therefore, both *Ruhulkudus* and *Ahmad* are NEs and must be labeled accordingly.

- d. Synonyms of an NE can appear without additional information written in brackets. Example:
 - Dihalalkan bagimu pada malam puasa bercampur dengan istrimu. Mereka adalah pakaian bagimu dan kamu adalah pakaian bagi mereka. Allah mengetahui bahwa kamu tidak dapat menahan dirimu sendiri, tetapi Dia menerima tobatmu dan memaafkanmu. Maka, sekarang campurilah mereka dan carilah apa yang telah ditetapkan Allah bagimu. Makan dan minumlah hingga jelas bagimu (perbedaan) antara benang putih dan benang hitam, yaitu fajar. Kemudian, sempurnakanlah puasa sampai (datang) malam. Akan tetapi, jangan campuri mereka ketika kamu (dalam keadaan) beriktikaf di masjid. Itulah batas-batas (ketentuan) Allah. Maka, janganlah kamu mendekatinya. Demikianlah Allah menerangkan ayat-ayat-Nya kepada manusia agar mereka bertakwa. (It has been made permissible for you the night preceding fasting to go to your wives [for sexual relations]. They are clothing for you and you are clothing for them. Allah knows that you used to deceive yourselves, so He accepted your repentance and forgave you. So now, have relations with them and seek that which Allah has decreed for you. And eat and drink until the white thread of dawn becomes distinct to you from the black thread [of night]. Then complete the fast until the sunset. And do not have relations with them as long as you are staying for worship in the mosques. These are the limits [set by] Allah, so do not approach them. Thus does Allah make clear His ordinances to the people that they may become righteous.) (2:187)

- In the above translation, *fajar* (dawn) is a proper noun as well as a synonym of *subuh* (dawn). Therefore, it must be labeled with the same label of *subuh*, i.e., *Event*.
- e. In the case of NE *Allah*, the synonyms are names that appear in <u>Asmaul Husna</u>. The synonyms appear in particular forms in the Indonesian translation of the Quran, as follows. In the following examples, Allah's synonyms are those in bold.
 - i. One name in Asmaul Husna that is preceded by the word *Yang* (The). Examples:
 - 1. Sesungguhnya bagi orang-orang yang beriman dan beramal saleh, (Allah) **Yang Maha Pengasih** akan menanamkan rasa cinta (dalam hati) mereka. (19:96)
 - 2. Mereka (yang menyekutukan Allah) berkata, "Allah mengangkat anak." Mahasuci Dia. Dialah **Yang Mahakaya**. Milik-Nyalah apa yang ada di langit dan apa yang ada di bumi. Kamu tidak mempunyai alasan kuat tentang ini. Pantaskah kamu mengatakan tentang Allah apa yang tidak kamu ketahui? (109:68)
 - ii. Two names in Asmaul Husna that are preceded with the word *Yang* (The) and connected with the word *Lagi* (also/as well as/and). Those two names together are an NE. Examples:
 - 1. (Suatu tanda juga atas kekuasaan Allah bagi mereka adalah) matahari yang berjalan di tempat peredarannya. Demikianlah ketetapan (Allah) **Yang Mahaperkasa lagi Maha Mengetahui**. (36:38)
 - 2. Dia tidak dapat dijangkau oleh penglihatan mata, sedangkan Dia dapat menjangkau segala penglihatan itu. Dialah **Yang Mahahalus lagi Mahateliti**. (6:103)
 - 3. Tidaklah mereka menyiksa (membakar) orang-orang mukmin itu, kecuali karena mereka beriman kepada Allah Yang Mahaperkasa lagi Maha Terpuji. (85:8)
 - iii. One or two names in Asmaul Husna that is/are preceded with the words *Tuhan Yang*. Examples:
 - 1. Tuhan kamu adalah **Tuhan Yang Maha Esa**. Tidak ada tuhan selain Dia Yang Maha Pengasih Lagi Maha Penyayang. (2:163)
 - 2. Seperti (pengutusan para rasul sebelummu) itulah, Kami (juga) mengutusmu (Nabi Muhammad) kepada suatu umat yang sungguh sebelumnya telah berlalu beberapa umat agar engkau bacakan kepada mereka (Al-Qur'an) yang Kami wahyukan kepadamu, padahal mereka ingkar kepada **Tuhan Yang Maha Pengasih**. Katakanlah, "Dia Tuhanku, tidak ada tuhan selain

- Dia. Hanya kepada-Nya aku bertawakal dan hanya kepada-Nya aku bertobat." (13:30)
- 3. Alif Lām Rā. (Ini adalah) Kitab (Al-Qur'an) yang Kami turunkan kepadamu (Nabi Muhammad) agar engkau mengeluarkan manusia dari berbagai kegelapan pada cahaya (terang-benderang) dengan izin Tuhan mereka, (yaitu) menuju jalan **Tuhan Yang Mahaperkasa lagi Maha Terpuji**. (14:1)
- 4. (Ingatlah) Ayyub ketika dia berdoa kepada Tuhannya, "(Ya Tuhanku,) sesungguhnya aku telah ditimpa penyakit, padahal Engkau **Tuhan Yang Maha Penyayang** dari semua yang penyayang."(21:83)
- f. NEs can appear in brackets, functioning as additional information for corresponding common nouns. In this case, only the NEs must be labeled accordingly. Examples:
 - i. dan demi negeri (Makkah) yang aman ini, (And [by] this secure city [Makkah],) (95:3)
 - ii. Kami pun telah mengangkat gunung (**Sinai**) di atas mereka untuk (menguatkan) perjanjian mereka.182) Kami perintahkan kepada mereka, "Masukilah pintu gerbang (**Baitulmaqdis**) itu sambil bersujud". Kami perintahkan pula kepada mereka, "Janganlah melanggar (peraturan) pada hari Sabat." Kami telah mengambil dari mereka perjanjian yang kukuh. (4:154)
 - iii. Lalu, Malaikat (**Jibril**) memanggilnya ketika dia berdiri melaksanakan salat di mihrab, "Allah menyampaikan kabar gembira kepadamu dengan (kelahiran) Yahya yang membenarkan kalimat dari Allah, (menjadi) panutan, menahan diri (dari hawa nafsu), dan seorang nabi di antara orang-orang saleh." (3:39)