



Ekstraksi Informasi Semantik dan Spatiotemporal pada Artikel Online Terkait Bencana di Indonesia

Ashr Hafiizh Tantri^{a,*}, Nur Aini Rakhmawati^b

^a Institut Teknologi Sepuluh Nopember

^b Institut Teknologi Sepuluh Nopember

Naskah Diterima : 5 Februari 2020; Diterima Publikasi : 10 Juli 2020

DOI: 10.21456/vol10iss1pp114-121

Abstract

Indonesia is one country that has a high risk of natural disasters. Ranked in 36 out of 172 countries in disaster index and having 2,372 disaster incidents in 2017, actions need to be taken to minimize the impact of natural disasters. One of it is to do a hazard map modeling. In making hazard maps, several approaches can be used, one of which is the semantic approach to extract disaster information. Therefore, this study aims to develop a system that can be used to extract spatiotemporal and semantic information related to natural disasters in Indonesia. This study uses the NLP method in conducting the information extraction process and carried out using the GATE (General Architecture for Text Engineering) application. In processing Indonesian language articles, it is necessary to develop the plugin because the Indonesian information structure is different from the default information structure in GATE application. The plugin development process is done by using ontology as the basis for determining semantic information. Literature study was carried out related to government regulations that further explained the need for semantic and spatiotemporal information about disaster events. system performance developed produces a precision value of 38% and a recall value of 32%. this is because the system experiences some difficulties in carrying out the information inference process. The reason for low precision rate is because the rules used in the inference process to pair the three types of information still cannot accommodate the variation of information positions in different sentences.

Keywords : Indonesia; Information Extraction; Natural Disaster; Online News Article; Semantic

Abstrak

Indonesia merupakan salah satu negara yang memiliki resiko tinggi dalam bencana alam. Berada di peringkat 36 dari 172 negara berdasarkan indeks risiko terhadap bencana serta terdapat 2.372 kejadian bencana di Indonesia pada tahun 2017, perlu dilakukan tindakan untuk meminimalisir dampak dari kejadian bencana alam. Salah satu cara yang bisa dilakukan adalah melakukan pemodelan peta bahaya. Dalam pembuatan peta bahaya, ada beberapa pendekatan yang bisa digunakan, salah satunya adalah dengan pendekatan semantik untuk mengekstraksi informasi bencana. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem yang bisa digunakan untuk melakukan ekstraksi informasi spatiotemporal dan semantik terkait bencana alam di Indonesia. Penelitian ini menggunakan metode NLP dalam melakukan proses ekstraksi informasi dengan menggunakan aplikasi GATE (General Architecture for Text Engineering). Dalam memproses artikel berbahasa Indonesia, perlu dilakukan pengembangan plugin karena struktur informasi bahasa Indonesia berbeda dengan struktur informasi bawaan pada aplikasi GATE. Proses pengembangan plugin dilakukan dengan memanfaatkan ontologi sebagai dasar dari penentuan informasi semantik. Selain itu juga dilakukan studi literatur terhadap aturan pemerintah yang menjelaskan lebih lanjut mengenai kebutuhan informasi semantik dan spatiotemporal tentang kejadian bencana. performa sistem yang dikembangkan menghasilkan nilai precision sebesar 38% dan nilai recall sebesar 32%. hal ini dikarenakan sistem mengalami beberapa kesulitan dalam melakukan proses inferensi informasi. Penyebab rendahnya precision rate sistem dikarenakan rules yang digunakan dalam proses inferensi untuk memasangkan tiga jenis informasi masih kurang bisa mengakomodasi variasi posisi informasi di kalimat yang berbeda beda.

Keywords : Artikel Berita Online; Bencana Alam; Ekstraksi Informasi; Indonesia; Semantik

1. Pendahuluan

Indonesia merupakan negara tropis yang terletak di antara tiga pertemuan lempeng besar, yaitu Lempeng Eurasia, Lempeng Indo-Australia, dan Lempeng Pasifik. Menurut teori lempeng tektonik,

tempat-tempat yang memiliki kondisi tektonik yang aktif, yang menyebabkan gempa bumi, gunung berapi dan pembentukan dataran tinggi (Balai Besar Wilayah III Denpasar and BMKG, 2013). Menurut World Risk Report 2018, Indonesia masuk dalam kategori resiko tinggi dengan peringkat ke 36 dari

*) Penulis korespondensi: ashrht@yahoo.co.id

172 negara berdasarkan indeks risiko terhadap bencana (Heintze *et al.*, 2018). Pada tahun 2017 terdapat 2.372 kejadian bencana di Indonesia, dimana jumlahnya meningkat 2 kali lipat jika dibandingkan kejadian bencana pada 10 tahun sebelumnya (Pusat Data Informasi dan Humas, 2019).

Berdasarkan data World Health Organization, bencana alam mengakibatkan 90.000 orang kehilangan nyawa dan berdampak kepada 160 juta orang setiap tahunnya (World Health Organization, 2008). Menurut IMF, bencana alam juga memberikan dampak negatif dan signifikan terhadap pertumbuhan ekonomi serta dapat menyebabkan terjadinya pemburukan keseimbangan perdagangan maupun fiskal (Lee *et al.*, 2018). Selain itu, daerah yang mengalami bencana parah akan memiliki migrasi penduduk keluar yang besar, harga rumah di daerah terdampak turun drastis, serta meningkatnya tingkat kemiskinan pada daerah terdampak bencana (Boustan *et al.*, 2017).

Adanya penerapan teknologi bisa membantu mengurangi biaya yang dikeluarkan untuk menangani atau bahkan dampak yang diakibatkan oleh bencana alam (Kirkpatrick, 2018). Salah satu penerapan teknologi dalam *disaster risk reduction* adalah pemodelan peta bahaya/*hazard map*. Penggunaan peta bahaya bermanfaat untuk perencana dan pembuat keputusan dalam memahami dan mengelola daerah terdampak dengan cara: membantu melakukan identifikasi resiko bencana sebelum terjadinya bencana alam tersebut, melakukan mitigasi kerusakan sosial ekonomi yang disebabkan oleh bencana tersebut, dan menetapkan prioritas terhadap lokasi terdampak dengan tingkat bahaya tinggi (Edirisooriya *et al.*, 2018).

Umumnya ada beberapa tahapan yang dilakukan dalam pembuatan peta bahaya, yaitu : pengumpulan data, pengolahan data, pembuatan peta. Dalam proses pengumpulan data maupun pengolahan yang umumnya dilakukan di Indonesia, dibutuhkan waktu yang tidak sebentar karena proses yang dilakukan cukup rumit. Salah satu pendekatan yang bisa digunakan dalam pembuatan peta bahaya adalah pendekatan ekstraksi informasi semantik dengan menggunakan bantuan ontologi (Traverso *et al.*, 2014; Wang and Stewart, 2015, 2014).

GATE adalah salah satu *tools* yang bisa digunakan dalam proses ekstraksi informasi. Sebagaimana yang disinggung pada penelitian Wei Wang, perlu dilakukan pengembangan *plugin* agar GATE bisa digunakan untuk melakukan ekstraksi informasi di negara lain.

Penelitian ini bertujuan untuk mengembangkan sistem yang bisa melakukan ekstraksi informasi spatiotemporal dan semantik terkait bencana alam di Indonesia.

Pengembangan sistem dilakukan dengan cara memanfaatkan GATE sebagai *core* dari sistem. Penyesuaian dilakukan dengan cara mengembangkan

beberapa *plugin* yang dibutuhkan untuk memproses informasi kejadian bencana di Indonesia. Proses pengembangan dilakukan dalam dua tahapan utama, yaitu: pengembangan berdasarkan literatur dan pengembangan berdasarkan hasil *training*. Setelah proses pengembangan selesai, dilakukan pengujian *precision rate* dan *recall rate* yang membandingkan hasil performa sistem dengan responden awam.

2. Kerangka Teori

2.1. Bencana Alam

Bencana adalah peristiwa atau rangkaian peristiwa yang mengancam dan mengganggu kehidupan dan penghidupan masyarakat yang disebabkan, baik oleh faktor alam dan/atau faktor nonalam maupun faktor manusia sehingga mengakibatkan timbulnya korban jiwa manusia, kerusakan lingkungan, kerugian harta benda, dan dampak psikologis (UU No. 24 Tahun 2007). Sementara itu, bencana alam adalah bencana yang diakibatkan oleh peristiwa atau serangkaian peristiwa yang disebabkan oleh alam.

2.2. Artikel Online

Artikel berarti sebuah karya tulis lengkap, misalnya laporan berita atau esai dalam majalah, surat kabar, dan sebagainya (KBBI, 2016). Sementara itu, Online bisa diartikan sebagai sesuatu yang ada di internet. Sehingga, bisa diartikan bahwa artikel online adalah berita yang bisa ditemukan di internet, misalnya berita dari detik.com atau berita dari merdeka.com.

Artikel online memiliki beberapa karakteristik. Salah satunya adalah immediacy (kesegaran), artinya berita bersifat cepat dan langsung. Selain itu juga ada karakteristik Storage and retrieval, yang merujuk pada kemampuan berita dapat disimpan dan diakses lagi kapan pun (Romli, 2012)

2.3. Ekstraksi Informasi

Ekstraksi Informasi adalah salah satu tipe pemrosesan dokumen yang menangkap dan menghasilkan informasi faktual yang tersedia di dalam dokumen tersebut (Okurowski, 1993). Tantangan dalam IE adalah informasi yang diolah umumnya tidak terstruktur. Salah satu metode yang digunakan dalam menangani hal tersebut adalah menggunakan Named-entity recognition.

Named-entity recognition merupakan turunan dari ekstraksi informasi, yaitu proses ekstraksi dan identifikasi jenis kata / entitas kata, terutama entitas bernama / named entity (misal : orang, lokasi, organisasi, nomor, tanggal), dari setiap kata yang ada di dalam sebuah teks yang umumnya tidak terstruktur (misal : artikel koran / berita) (Fauzi Lailil; Setiyaji, Agung, 2017; Nadeau and Sekine, 2007). Named-entity recognition memungkinkan sistem ekstraksi informasi untuk mengenali dan melakukan

klasifikasi unit informasi dari teks yang tidak terstruktur ke dalam kategori tertentu yang sudah dijabarkan sebelumnya (Roy, 2017).

2.4. Informasi Semantik

Secara sederhana, informasi semantik berarti sebuah informasi yang bermakna. Dalam penelitian ini, informasi bermakna yang dibutuhkan adalah informasi terkait kejadian bencana.

Mengikuti aturan pemerintah dalam menetapkan tingkat keparahan bencana (UU No. 24 Tahun 2007), ada beberapa informasi semantik terkait dampak bencana yang digunakan: jumlah korban, kerugian harta benda non bangunan, kerusakan sarana dan prasarana, serta dampak sosial ekonomi yang ditimbulkan.

Selain itu informasi terkait jenis kejadian bencana juga masuk ke dalam informasi semantik yang diproses dalam penelitian ini.

2.5. Informasi Spatiotemporal

Spatiotemporal merupakan sebuah kata yang merupakan hasil dari penggabungan kata *spatial* dan *Temporal*.

Berdasarkan definisi dari KBBI, temporal berarti berhubungan dengan waktu atau berkenaan dengan waktu-waktu tertentu. Informasi waktu / *temporal information* adalah Informasi yang memiliki atribut waktu / *temporal duration*.

Spatial merupakan derivasi dari kata *space* dalam bahasa Inggris. Kata *Spatial* mengacu pada fitur dan fenomena yang terdistribusikan dalam ruang yang terdefinisi secara geografis, sehingga memiliki ukuran dimensi secara fisik (NCERT, 2012). Data spasial, seringkali disebut sebagai data geospasial ataupun informasi geografis, adalah data-data yang memiliki referensi secara langsung ataupun tidak langsung terhadap lokasi spesifik atau data geografis (Information Commissioner's Office, 2014).

Agar lebih mudah memahami informasi *spatiotemporal*, maka asumsikan ada sebuah kalimat berikut: "Gunung Ibu di Halmahera Barat, Maluku Utara, meletus, pada Sabtu (12/1) pukul 17.12 WIT". Pada kalimat tersebut, "Gunung Ibu" "di Halmahera Barat", dan "Maluku Utara" merupakan sebuah informasi spasial. Sementara itu "pada Sabtu (12/1)" dan "Pukul 17.12 WIT" merupakan informasi temporal.

2.6. Ontologi

Ontologi adalah deskripsi formal dari entitas beserta sifat, hubungan, batasan, maupun perilaku dari entitas tersebut (Gruninger and Fox, 1995).

Sebuah ontologi semestinya memerlukan atau mewujudkan semacam pandangan mendunia sehubungan dengan domain yang dibahas. Pandangan mendunia itu seringkali dipahami sebagai sekumpulan konsep (misal: entitas, atribut, poses), dengan definisi dan antar-hubungan mereka, dan

seringkali disebut sebagai konseptualisasi (Gruninger and Uschold, 1996). Konseptualisasi yang ada bisa secara implisit, hanya ada di dalam pikiran seseorang, ataupun terpasang di salah satu bagian perangkat lunak.

2.7. GATE

General Architecture for Text Engineering (GATE) adalah sebuah tools berbasis Java Suite yang dikembangkan oleh Universitas Sheffield pada tahun 1995. Tools ini digunakan oleh ilmuwan, perusahaan, guru, dan siswa untuk macam-macam task terkait pemrosesan bahasa alami/natural language, termasuk ekstraksi informasi dalam banyak bahasa.

Ada beberapa tahapan yang dilakukan oleh tools GATE untuk melakukan ekstraksi informasi dari teks mentah. Secara garis besar, cara kerja aplikasi GATE dalam memproses informasi adalah memasukan sumber artikel ke dalam Language Resource yang selanjutnya di proses dengan menggunakan Processing Resource. Artikel yang sudah diproses selanjutnya bisa dibuka dan akan muncul anotasi – anotasi yang menunjukkan adanya informasi penting pada teks yang nantinya bisa di export menjadi luaran dalam bentuk file XML (Extensible Markup Language).

2.8. Precision dan Recall

Precision dan *Recall* merupakan parameter yang digunakan dalam pengujian hasil klasifikasi. Proses klasifikasi selalu memiliki hasil klasifikasi yang benar dan hasil klasifikasi yang salah. Lebih lanjut lagi, model juga seringkali membutuhkan mengidentifikasi sebuah item sebagai bukan termasuk dalam sebuah klasifikasi. Untuk lebih jelasnya bisa dilihat di confusion matrix pada Gambar 1.

| | | Actual Values | |
|------------------|--------------|---------------|--------------|
| | | Positive (1) | Negative (0) |
| Predicted Values | Positive (1) | TP | FP |
| | Negative (0) | FN | TN |

Gambar 1. Confusion Matrix

Dari *confusion matrix* pada Gambar 1, yang dimaksud dengan TP, TN, FP, dan FN adalah:

1. True Positives (TP), data yang diklasifikasikan oleh model sebagai positif dan label sesungguhnya adalah memang positif (prediksi benar).
2. True Negatives (TN), data yang diklasifikasikan oleh model sebagai negatif dan label

sesungguhnya adalah memang negatif (prediksi benar).

3. False Positives (FP), data yang diklasifikasikan oleh model sebagai positif dan label sesungguhnya adalah negatif (prediksi salah).
4. False Negative (FN), data yang diklasifikasikan oleh model sebagai negatif dan label sesungguhnya adalah positif (prediksi salah).

Precision adalah proporsi kasus prediksi positif yang benar-benar positif. Sementara itu, *recall* adalah proporsi kasus *real positive* yang diprediksi dengan positif dengan benar. Secara umum, rumus dari *precision* bisa dilihat pada persamaan (1) sementara itu rumus dari *recall* bisa dilihat pada persamaan (2).

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

2.9. Kaidah Penulisan Bahasa Indonesia

Dengan melakukan identifikasi kaidah penulisan bahasa Indonesia, maka proses identifikasi informasi spasial, temporal, maupun semantik terkait bencana bisa berjalan lebih baik. Setidaknya ada 2 hal dalam kaidah penulisan bahasa Indonesia yang bisa digunakan untuk membantu proses identifikasi informasi, yaitu preposisi dan penulisan huruf kapital.

Preposisi, atau yang lebih dikenal sebagai kata depan, adalah kata yang terletak di bagian awal frasa dan unsur yang mengikutinya dapat berupa nomina, adjektiva, atau verba (Kusdiyana, 2002). Setiap preposisi memiliki aspek nama masing – masing, sehingga bisa digunakan dalam membantu mengidentifikasi jenis informasi yang ada setelah preposisi.

Dalam kaidah penulisan bahasa Indonesia, ada beberapa frasa tertentu yang harus ditulis dengan menggunakan huruf kapital (Lanin, 2016). Dengan melakukan identifikasi huruf kapital, bisa dilakukan pengelompokan jenis informasi yang ada.

2.9. Clueword

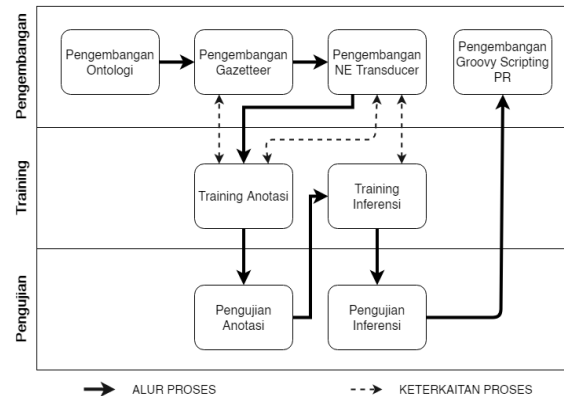
Sebagaimana dalam data wilayah administrasi pemerintahan yang diterbitkan oleh permendagri (Kementerian Dalam Negeri Republik Indonesia, 2017), setiap provinsi memiliki lampiran yang tebalnya mencapai ratusan lembar terkait informasi nama kabupaten, kota, dan kecamatan. Sehingga kurang direkomendasikan untuk memasukkan semua data lokasi mentah Indonesia hingga tingkat kelurahan, terutama jika ada perubahan ataupun pemekaran wilayah. Salah satu solusi yang bisa digunakan untuk mengurangi beban sistem adalah dengan menggunakan *clueword*.

Clueword / *clue* adalah salah satu fitur yang bisa digunakan dalam proses deteksi entitas bernama / *Named-entity recognition*. Penggunaan *gazetteer* list

dan *clue* bisa meningkatkan tingkat akurasi dari proses deteksi entitas bernama (Anggareska and Purwarianti, 2014). Ada beberapa macam *clue* yang bisa digunakan, misalnya: pronomina, nomina, *prefix*, dan *suffix*.

3. Metode

Ada beberapa tahapan yang dilakukan dalam penelitian ini yang ditunjukkan pada Gambar 2.



Gambar 2. Tahapan penelitian

3.1. Pengembangan Ontologi

Pada proses ini dikembangkan ontologi dengan mengikuti salah satu referensi pembuatan ontologi yang sering dirujuk (Noy and McGuinness, 2000). Ontologi dikembangkan dengan melibatkan aturan pemerintahan (UU No. 24 Tahun 2007), ontologi yang sudah ada (Fauzan and Basuki, 2016; Fauziati and Watanabe, 2010), ataupun literatur lainnya (Amri et al., 2006; Data Informasi Bencana Indonesia, 2017).

3.2. Pengembangan Gazetteer

Secara garis besar, pengembangan *gazetteer* dilakukan dengan dua cara, yaitu: penerjemahan *gazetteer* yang sudah ada di GATE ke dalam bahasa Indonesia dan penambahan *clue-word* untuk *gazetteer*.

proses penerjemahan *gazetteer* tidak dilakukan secara langsung dengan cara menerjemahkan setiap baris yang ada di *gazetteer* tersebut, melainkan dilakukan penyesuaian mengikuti kondisi di Indonesia (misal: untuk hari besar tidak di terjemahkan secara mentah namun mengikuti hari besar yang ada di Indonesia).

Sementara itu, untuk proses penambahan *clue-word*, dilakukan penambahan frasa berdasarkan beberapa referensi / penelitian terkait *clue* spasial, temporal, maupun semantik.

3.3. Pengembangan NE Transducer

Proses pengembangan NE Transducer dilakukan dengan cara mengimplementasikan beberapa kaidah

penulisan Bahasa Indonesia ke dalam rules dalam file .jape.

Implementasi dilakukan dengan cara mengubah aturan tata bahasa Indonesia menjadi *linguistic code* dalam format Java Annotation Patterns Engine (JAPE).

3.4. Training Anotasi

Meskipun ada beberapa aturan yang bisa digunakan dalam mendeteksi informasi berharga, seperti Pedoman Umum Ejaan Bahasa Indonesia atau Kaidah Penulisan Bahasa Indonesia. Namun banyak orang yang tidak mengindahkan aturan tersebut dan memilih menulis artikel sesuai dengan gaya bahasanya masing - masing. Oleh karena itu perlu dilakukan training anotasi agar model yang dikembangkan bisa mendeteksi beberapa “penyimpangan” dalam penulisan informasi berharga.

Training anotasi dilakukan dengan cara melakukan analisa struktur informasi yang ada pada beberapa sumber berita. Hasil analisa selanjutnya digunakan untuk melakukan penyesuaian pada NE Transducer maupun Gazetteer sehingga sistem berhasil mengidentifikasi informasi – informasi pada sumber berita tersebut.

Pada tahapan ini, ada 15 artikel berita online dari bnpb.go.id, merdeka.com, dan detik.com yang digunakan.

3.5. Pengujian Anotasi

Pengujian anotasi model dilakukan dengan cara, membandingkan beberapa hasil anotasi manual, hasil anotasi sistem, dan hasil anotasi responden awam. Proses penghitungan precision and recall dilakukan per kata berdasarkan jenis informasi yang ada.

Pada pengujian ini, ada 7 artikel berita online dari merdeka.com dan detik.com yang digunakan. 7 Responden awam dilibatkan sebagai pembanding untuk kinerja dari sistem.

3.6. Training Inferensi

Perlunya dilakukan Training inferensi adalah karena struktur kalimat dalam bahasa Indonesia mempunyai banyak variasi bahkan untuk tipe kalimat formal sekalipun.

Training inferensi dilakukan dengan cara melakukan analisa kombinasi informasi semantik, spasial, dan temporal yang ada pada beberapa sumber berita. Hasil analisa selanjutnya digunakan untuk melakukan penyesuaian pada NE Transducer sehingga sistem berhasil melakukan inferensi pada sumber berita tersebut.

Pada tahapan ini, 15 artikel berita online yang dilibatkan pada training anotasi akan digunakan ulang dalam proses training inferensi. Selain itu juga 7 artikel berita online yang digunakan untuk menguji anotasi juga digunakan dalam proses training inferensi.

3.7. Pengujian Inferensi

Pengujian model dilakukan dengan cara, membandingkan beberapa hasil inferensi manual, hasil inferensi sistem, dan hasil inferensi responden awam. Proses penghitungan precision and recall dilakukan per baris / kalimat yang dibuat oleh sistem ataupun responden awam yang berisikan informasi semantik + informasi spasial + informasi temporal.

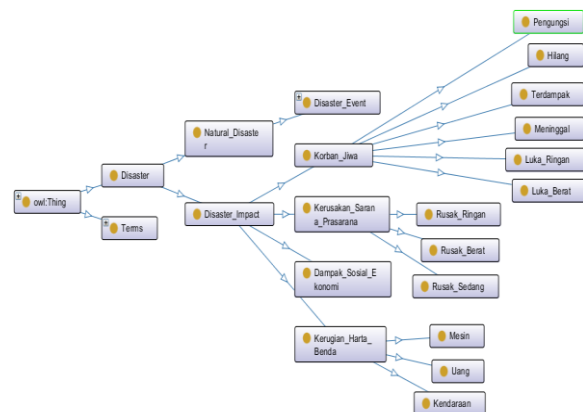
Pada pengujian ini, ada 5 artikel berita online dari detik.com yang digunakan. 6 Responden awam dilibatkan sebagai pembanding untuk kinerja dari sistem.

3.8. Pengembangan Groovy Scripting PR

Pada bagian ini dilakukan penulisan baris kode yang nantinya digunakan untuk melakukan *export* luaran proses inferensi menjadi file dengan ekstensi .csv. Sehingga nanti bisa dimanfaatkan lebih mudah dalam pengolahan informasi kebencanaan.

4. Hasil dan Pembahasan

Ontologi dalam penelitian ini dikembangkan menggunakan aplikasi Protégé 4.3. Hasil akhir dari ontology yang dikembangkan bisa dilihat pada Gambar 3.



Gambar 3. Ontologi bencana

Pada penelitian ini, dihasilkan 32 gazetter dengan rincian 17 gazetter merupakan gazetter yang ditambahkan berdasar referensi terkait, dan 15 gazetter merupakan hasil terjemahan dari gazetter bawaan dari GATE.

Pengembangan NE Transducer menghasilkan 7 file .jape, termasuk file .jape yang digunakan untuk membersihkan anotasi sementara. Selain itu, proses training juga menambahkan 2 file .jape baru dan beberapa pembaruan terhadap file .jape yang ada. sehingga NE Transducer yang dihasilkan pada penelitian ini berjumlah total 9 buah.

Setiap file JAPE yang ada memiliki jumlah rules yang bervariasi, tergantung aturan penulisan bahasa Indonesia dan pengembangan yang dilakukan.

Gambar 4 merupakan contoh salah satu potongan NE Transducer yang dikembangkan untuk menyaring

informasi ekstra dengan bantuan clueword. Gambar 4 berisi rule yang digunakan untuk mendeteksi informasi temporal berdasarkan ada tidaknya clueword yang mengikuti kata lain dalam sebuah frasa.

```
Rule: ExtraClueTemporalSuffix
(
  (
    {Token.kind == word}
    {SpaceToken}
    {Lookup.majorType == "clueword", Lookup.minorType == "suffix_waktu"}
  )
):capital
-->
gate.AnnotationSet capital = (gate.AnnotationSet)bindings.get("capital");
gate.Annotation capitalAnn = (gate.Annotation)capital.iterator().next();
gate.FeatureMap features = Factory.newFeatureMap();
features.put("ExtraInformasi", "Temporal");
features.put("rule", "suffix");
outputAS.add(capital.firstChild(), capital.lastNode(), "ExtraClueT", features); }
```

Gambar 3. Salah satu potongan NE Transducer

Dalam penelitian ini, proses pengujian sistem berjalan dua kali, yaitu: pengujian hasil anotasi dan pengujian hasil inferensi.

Hasil dari pengujian anotasi bisa dilihat pada Tabel 1. Sementara hasil dari pengujian inferensi bisa dilihat pada Tabel 2.

Tabel 1. Komparasi hasil anotasi sistem dengan responden awam

| Jenis Informasi | Precision | | Recall | |
|-----------------|-------------|--------|-------------|--------|
| | Rerata Awam | Sistem | Rerata Awam | Sistem |
| Spasial | 76% | 69% | 63% | 52% |
| Temporal | 86% | 82% | 79% | 47% |
| Semantik | 28% | 29% | 24% | 26% |
| Keseluruhan | 50% | 49% | 44% | 36% |

Sebagaimana yang bisa dilihat pada Tabel 1, performa sistem dalam melakukan anotasi informasi spasial cukup baik dengan nilai precision rate 69% dan recall rate 52%. Penyebab dari recall rate yang sedikit di atas 50% dikarenakan clueword yang bisa digunakan untuk mendeteksi informasi tersebut belum ada di dalam gazetteer spasial.

Performa sistem dalam melakukan anotasi informasi temporal cukup baik dengan nilai precision rate 82% dan recall rate 47%. Penyebab dari cukup rendahnya recall rate disebabkan oleh struktur informasi temporal pada beberapa artikel pengujian berbeda dengan yang dijumpai pada artikel training model.

Performa sistem dalam melakukan anotasi informasi semantik kurang bagus, dengan nilai precision rate 29% dan recall rate 26%. Rendahnya nilai precision dan recall dari sistem disebabkan oleh ketidaksanggupan sistem untuk mengidentifikasi kerugian harta benda, dampak sosial bencana alam, serta informasi korban jiwa. Kesulitan sistem dalam mengidentifikasi kerugian harta benda dan informasi korban jiwa disebabkan oleh dua hal, yaitu : struktur informasi yang sedikit berbeda dengan dari hasil training dan gazetteer yang masih perlu isian clueword. Sementara itu kesulitan sistem dalam

mengidentifikasi dampak sosial bencana alam dikarenakan definisi dari dampak sosial masih ambigu dan terlalu luas.

Tabel 2. Komparasi hasil inferensi sistem dengan responden awam

| Rerata Awam | Precision | | Recall | |
|-------------|-----------|-------------|--------|-------------|
| | Sistem | Rerata Awam | Sistem | Rerata Awam |
| 75% | 38% | 60% | 32% | |

Sebagaimana yang bisa dilihat pada Tabel 2, kemampuan sistem untuk melakukan inferensi kesimpulan lebih buruk dibandingkan kemampuan responden awam.

Kegagalan sistem dalam melakukan identifikasi informasi semantik berpengaruh terhadap kurang baiknya performa sistem dalam melakukan inferensi. Hal ini bisa dilihat dari recall rate sistem yang mencapai nilai 32% dalam proses inferensi.

Proses inferensi yang dilakukan oleh sistem perlu diperbaiki lagi agar bisa menyusun kalimat utuh berisi informasi semantik, informasi spasial, dan informasi temporal dengan lebih baik. Penyebab precision rate sistem yang mencapai nilai 38% dikarenakan rules yang digunakan dalam proses inferensi untuk memasang tiga jenis informasi masih kurang bisa mengakomodasi variasi posisi informasi di kalimat yang berbeda beda.

Hasil inferensi yang didapatkan selanjutnya diproses dengan menggunakan Groovy Scripting PR menjadi file .csv. pada Gambar 4 ditunjukkan tampilan dari contoh luaran file .csv.

| File | Semantik | Temporal | Spasial |
|--|----------|-------------------|---------------------------------------|
| angin-kencang-di-madun-5 Angin Kencang | | | 16-Jan-20 Jawa Timur |
| angin-kencang-di-madun-5 angin kencang | | | 16-Jan-20 di Kabupaten Madun |
| angin-kencang-di-madun-5 angin kencang | | pukul 14.00 WIB | (di) Kabupaten Madun |
| angin-kencang-di-madun-5 Angin Kencang | | (16/1/2020) | Desa Kajang |
| angin-kencang-di-madun-5 hujan deras | | (16/1/2020) | Desa Kajang |
| angin-puting-bellang-terjan Puting Bellang | | | 16-Jan-20 Jawa Barat |
| angin-puting-bellang-terjan Kencana rumah warga | | | 16-Jan-20 di Kabupaten Garut |
| angin-puting-bellang-terjan angin puting bellang | | | 16-Jan-20 di Kabupaten Garut |
| angin-puting-bellang-terjan angin puting bellang | | | 16-Jan-20 di Kabupaten Garut |
| angin-puting-bellang-terjan angin puting bellang | | Kamis (16/1/2020) | kawasan Cobatu |
| angin-puting-bellang-terjan angin puting bellang | | Kamis (16/1/2020) | di kawasan Desa Mertajaya |
| angin-puting-bellang-terjan angin puting bellang | | Kamis (16/1/2020) | di kawasan Desa Mertajaya, Tli Tengah |

Gambar 4. Contoh luaran file Csv

5. Kesimpulan

Secara garis besar, sistem yang dikembangkan memiliki performa anotasi yang kurang memuaskan. Penyebab utama dari performa yang tidak memuaskan adalah karena sistem belum berhasil mengenali beberapa struktur kata terkait informasi yang ada. Salah satu cara yang bisa dilakukan untuk mengatasi hal ini adalah melibatkan lebih banyak variasi struktur informasi dalam proses *training model*.

Selain itu sistem juga memiliki hasil inferensi yang buruk. *recall rate* yang buruk ini disebabkan oleh alur inferensi yang mengutamakan informasi

semantik sebagai fokus utama. Sebagaimana yang bisa dilihat dari performa anotasi, sistem masih kurang bisa mengidentifikasi informasi semantik. Sementara *precision rate* yang buruk disebabkan oleh sistem yang belum berhasil melakukan inferensi untuk berupa kombinasi urutan informasi.

Sehingga agar proses inferensi berjalan lebih baik, selain bergantung pada keberhasilan proses anotasi informasi semantik, juga harus dilakukan *training model* untuk bermacam – macam kombinasi urutan informasi.

Daftar Pustaka

- Amri, M.R., Yulianti, G., Yunus, R., Wiguna, S., Adi, A.W., Ichwana, A.N., Randongkir, R.E., Septian, R.T., 2006. Risiko Bencana Indonesia 218 pp.
- Anggareska, D., Purwarianti, A., 2014. Information Extraction of Public Complaints on Twitter Text For Bandung Government. 2014 Int. Conf. Data Softw. Eng. 1–6.
- Balai Besar Wilayah III Denpasar, BMKG, 2013. Gempabumi [WWW Document]. Badan Meteorol. Klimatologi dan Geofis. Wil. 3 Denpasar. URL <http://balai3.denpasar.bmkg.go.id/tentang-gempa> (accessed 3.27.19).
- Boustan, L.P., Kahn, M.E., Rhode, P.W., Yanguas, M.L., 2017. The Effect of Natural Disasters on Economic Activity in US Counties: A Century of Data. Natl. Bur. Econ. Res. Work. Pap. Ser. No. 23410.
- Data Informasi Bencana Indonesia, 2017. Bencana Menurut Jenisnya di Indonesia Tahun 1815 s/d 2019 [WWW Document]. Badan Nas. Penanggulangan Bencana. URL <http://bnpb.cloud/dibi/tabel1> (accessed 3.27.19).
- Edirisooriya, K. V.D., Vitanage, N.S., Uluwaduge, P., Senevirathna, E.M.T.K., 2018. Understanding Disaster Risk and Mitigation Special Reference to Ratnapura District. *Procedia Eng.* 212, 1034–1039.
- Fauzan, R., Basuki, S., 2016. Purwarupa Website Manajemen Bencana. *Semin. Nas. Ris. Terap.* 5662, 9–10.
- Fauzi Lailil; Setiyoadji, Agung, M.A.M., 2017. Named Entity Recognition Menggunakan Hidden Markov Model dan Algoritma Viterbi pada Teks Tanaman Obat. *J. Pengemb. Teknol. Inf. dan Ilmu Komput.* 1858–1864.
- Fauziati, S., Watanabe, K., 2010. Hazard Ontology Indonesia 2010 Ontology of Volcano System and Volcanic Hazards Assessment.pdf. *Int. J. Geoinformatics* 6, 49–61.
- Gruninger, M., Fox, M.S., 1995. Methodology for the Design and Evaluation of Ontologies 1 Introduction 2 Motivating Scenarios. *Ijcai* 95, 1–10.
- Gruninger, M., Uschold, M., 1996. *Ontologies: Principles, Methods, and Applications*. *Knowl. Eng. Rev.* 11.
- Heintze, H.J., Kirch, L., Kuppers, B., Mann, H., Mischo, F., Mucke, P., Pazdriern, T., Prutz, R., Radtke, K., Strube, F., Weller, D., 2018. World Risk Report 2018 Focus: Child Protection and Children’s Rights.
- Information Commissioner’s Office, 2014. What is spatial information? [WWW Document]. ico. URL <https://ico.org.uk/for-organisations/inspire-regulations/what-is-spatial-information/> (accessed 5.20.19).
- KBBI, 2016. Artikel [WWW Document]. Kamus Besar Bhs. Indones. URL <https://kbbi.web.id/artikel> (accessed 2.4.20).
- Kementerian Dalam Negeri Republik Indonesia, 2017. Permendagri No.137 Tahun 2017 Tentang Kode Dan Data Wilayah Administrasi Pemerintahan [WWW Document]. URL <https://www.kemendagri.go.id/pages/detail/108-permendagri-no137-tahun-2017> <https://www.kemendagri.go.id/pages/detail/108-permendagri-no137-tahun-2017> (accessed 5.20.19).
- Kirkpatrick, H., 2018. Can technology reduce the impact of natural disasters? - IoT Agenda [WWW Document]. Internet Things news, blogs Anal. URL <https://internetofthingsagenda.techtarget.com/blog/IoT-Agenda/Can-technology-reduce-the-impact-of-natural-disasters> (accessed 3.27.19).
- Kusdiyana, E., 2002. Kontrastif Antara Bahasa Jepang Dengan Bahasa Indonesia Ditinjau Dari Segi Preposisi.
- Lanin, I., 2016. Huruf Kapital [WWW Document]. PUEBI Daring. URL <https://puebi.readthedocs.io/en/latest/huruf/huruf-kapital/> (accessed 12.11.19).
- Lee, D., Zhang, H., Nguyen, C., 2018. The Economic Impact of Natural Disasters in Pacific Island Countries: Adaptation and Preparedness. *IMF Work. Pap.* 18, 1.
- Nadeau, D., Sekine, S., 2007. A survey of named entity recognition and classification. *Lingvisticae Investig.* 30, 3–26.
- NCERT, 2012. PRACTICAL WORK in Geography PART II. New Delhi.
- Noy, N.F., McGuinness, D.L., 2000. *Ontology Development 101: A Guide to Creating Your First Ontology* [WWW Document]. Stanford Univ. URL <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html> (accessed 4.1.19).
- Okurowski, M.E., 1993. Information extraction overview. In: *Proceedings of a Workshop on Held at Fredericksburg*. Association for Computational Linguistics, Virginia, pp. 117–121.

- Pusat Data Informasi dan Humas, 2019. Trend Bencana Indonesia Tahun 2003 - 2017 [WWW Document]. BNPB - Badan Nas. Penanggulangan Bencana. URL <https://bnpb.go.id/publikasi/infografis/trend-bencana-2013-2017.html> (accessed 3.27.19).
- Undang-Undang Republik Indonesia Nomor 24 Tahun 2007 Tentang Penanggulangan Bencana, Lembaran Negara RI Tahun 2007. Jakarta.
- Romli, Asep Syamsul M. (2012). *Jurnalistik Online*. Bandung: Nuansa Cendikia Rosdakarya.
- Roy, S., 2017. Named Entity Recognition. *AKGEC Int. J. Technol.* 8, 38–41.
- Traverso, S., Cerutti, V., Stock, K., Jackson, M., 2014. EDIT: A Methodology for the Treatment of Non-authoritative Data in the Reconstruction of Disaster Scenarios 32–45.
- Wang, W., Stewart, K., 2014. Creating Spatiotemporal Semantic Maps from Web Text Documents. *Space-Time Integr. Geogr. GIScience* 157–174.
- Wang, W., Stewart, K., 2015. Spatiotemporal and semantic information extraction from Web news reports about natural hazards. *Comput. Environ. Urban Syst.* 50, 30–40.
- World Health Organization, 2008. Natural events [WWW Document]. WHO | World Heal. Organ. URL https://www.who.int/environmental_health_emergencies/natural_events/en/ (accessed 3.27.19).