

Peningkatan Urgensi Daerah Rawan Bencana melalui Analisis Geoparsing pada Berita Kebencanaan dengan Text Mining

Muhammad Syaifur Rohman^{1*}, Nurul Anisa Sri Winarsih², Galuh Wilujeng Saraswati³
 Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
 *e-mail Corresponding Author: syaifur@dsn.dinus.ac.id

Abstract

This research aims to enhance the Disaster Vulnerability Map through the utilization of Geoparsing method by Text Mining on disaster news reports. The increase in casualties and damages caused by natural disasters reported by BNPB from 2020 to 2021 necessitates effective disaster management and preparedness for future events. BPBD Jawa Tengah employs disaster news reports as a means to raise public awareness. However, the creation of an accurate Disaster Vulnerability Map requires geospatial data on the frequency of disaster occurrences, which is not available within the reports. Thus, Geoparsing is employed to process the disaster reports data. The findings of this study demonstrate that Geoparsing can enhance the accuracy of the Disaster Vulnerability Map and provide insights into the level of urgency for disaster preparedness in the Preparedness Disaster Management phase.

Keywords: Text Mining; Geoparsing; Disaster Prone Area; Disaster Management

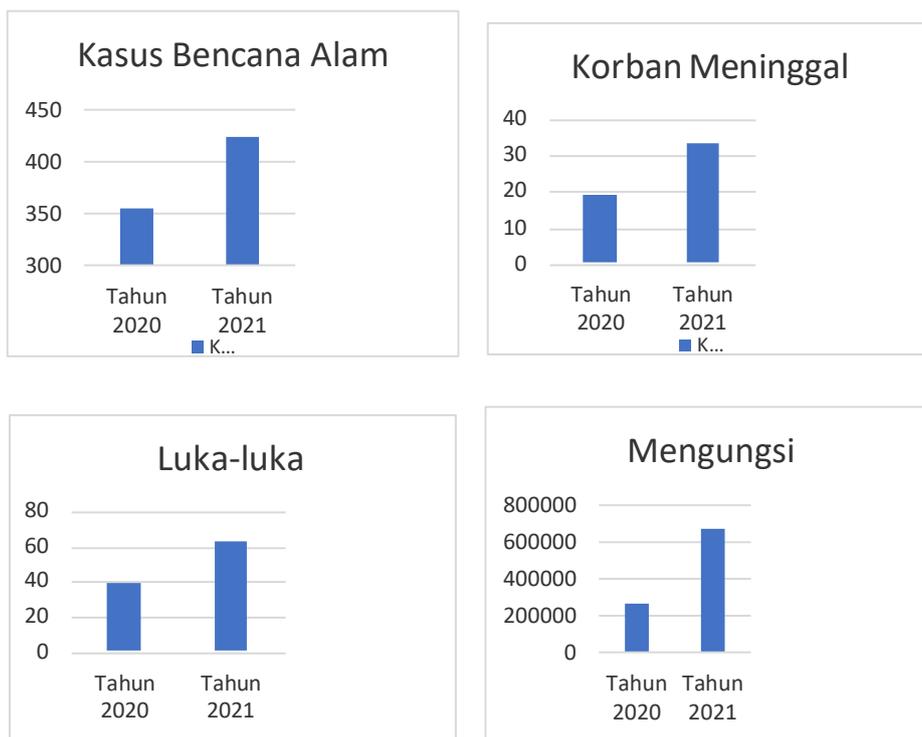
Abstrak

Penelitian ini bertujuan untuk meningkatkan Peta Rawan Bencana melalui penggunaan metode *Geoparsing* yang didapat melalui *Text Mining* pada berita laporan kebencanaan. Dalam kurun waktu tahun 2020 hingga 2021, terjadi peningkatan korban dan kerugian akibat bencana alam yang dilaporkan oleh BNPB. Oleh karena itu, penanganan dan persiapan yang efektif diperlukan untuk mengurangi dampak bencana di masa depan. BPBD Jawa Tengah menggunakan berita laporan kebencanaan sebagai upaya untuk meningkatkan kesadaran masyarakat. Namun, untuk menghasilkan Peta Rawan Bencana yang akurat, diperlukan data geospasial mengenai frekuensi kejadian bencana yang tidak tersedia dalam laporan tersebut. Dalam penelitian ini, dilakukan pengolahan data laporan kebencanaan menggunakan metode *Geoparsing*. Hasil penelitian menunjukkan bahwa *Geoparsing* dapat meningkatkan akurasi Peta Rawan Bencana dan memberikan informasi mengenai tingkat urgensi persiapan terhadap bencana di fase *Preparedness Disaster Management*.

Kata kunci: Text Mining; Geoparsing; Disaster Prone Area; Disaster Management

1. Pendahuluan

Pada tahun 2021 BNPB menyatakan bahwa terjadi kasus 2.841 Bencana alam ditunjukkan pada Gambar 1. Secara kumulatif terhitung ada sebanyak 672.736 orang menderita dan mengungsi. Dibandingkan dengan tahun lalu yaitu bulan November tahun 2020, pada bulan November tahun 2021 terjadi kenaikan kejadian bencana yaitu sebesar 19,4 % yang artinya menaikkan juga menaikkan jumlah korban terkena bencana seperti korban meninggal (73,7 %), luka luka (59 %), dan mengungsi (153 %). Peningkatan pada jumlah korban terkena bencana menunjukkan bahwa ada dua kemungkinan yang terjadi yaitu jumlah bencana yang terjadi lebih banyak daripada tahun sebelumnya atau persiapan akan datangnya bencana di waktu yang akan datang masih kurang sehingga terjadi peningkatan. Dengan adanya kenaikan yang seperti ini maka ditahun berikutnya dalam penanggulangan bencana alam memerlukan tindakan yang tepat sehingga korban terkena bencana bisa berkurang.



Gambar 1. Data Korban Bencana BNPB

Disaster Management adalah rangkaian konsep dan strategi untuk mengurangi dampak yang terjadi oleh bencana melalui berbagai analisis dan management yang baik antara manusia dengan aspek aspek yang lainnya [1]. Karena kerugian yang ditimbulkannya, berbagai penelitian *Disaster Management* [2]–[5] memegang peranan untuk penyelamatan hidup manusia melalui beragam cara hingga berbagai media yang bisa digunakan. Evaluasi pada *Disaster Management* juga dilakukan sehingga dalam pelaksanaannya, solusi yang digunakan selalu sesuai dengan keadaan bagaimana dan kapan bencana terjadi. *Disaster Management* memiliki beberapa fase yang berbeda tergantung melalui kapan bencana terjadi [6]. Solusi untuk mengurangi dampak terjadinya bencana salah satunya adalah tahap *Preparedness*. Dalam *Disaster Management* kita mengenal istilah tahap *preparedness* yang artinya persiapan untuk menanggulangi bencana sebelum terjadinya bencana untuk mengurangi kerugian yang akan berdampak pada manusia [7]. Dikatakan bahwa *preparedness* dengan rencana dan persiapan yang baik bisa mengurangi dampak yang terjadi oleh bencana. Dalam *preparedness*, memberikan laporan adalah salah satu cara untuk membuat manusia lebih siap jika ada bencana [7] maka dari itu BPBD Jawa Tengah melalui Halaman Websitenya selalu memberikan laporan kebencanaan. Masyarakat yang selalu mengunjungi Website BPBD Jawa Tengah dalam Gambar 2, mereka akan mendapatkan informasi terkait bencana sehingga meningkatkan kewaspadaannya. Sehingga harapannya adalah semakin banyak masyarakat mengetahui bahwa ada informasi bencana maka persiapan akan datangnya bencana di waktu yang akan datang lebih siap sehingga bisa mengurangi dampak kerugian yang akan terjadi.

Lalu karena kerugian bencana yang ditimbulkan semakin meningkat dari tahun sebelumnya, *Preparedness* dalam *Disaster Management* meningkatkan urgensitas dalam mencari solusi sebanyak mungkin sehingga kerugian yang di hasilkan oleh bencana dapat berkurang. Dalam mencari solusi data memegang peranan yang sangat penting, karena dengan contoh sebelumnya dengan data kita bisa mengetahui bahwa korban dampak dari bencana meningkat dari tahun sebelumnya. Sehingga akan urgensitasnya sebuah data maka dalam berbagai bentuknya, data sangat dibutuhkan untuk dijadikan informasi yang bermanfaat [8] terutama untuk dijadikan solusi untuk mengurangi dampak bencana. Salah satu data yang dapat dimanfaatkan lebih banyak fungsi lainnya adalah data laporan kebencanaan. Seperti yang diketahui, membaca laporan bencana artinya mengetahui bahwa kejadian bencana tersebut sudah terjadi dan sudah ada kerugian serta korban jiwa. Padahal dengan laporan

laporan yang sudah dibuat sebelumnya, bisa dilakukan pengolahan data yang bisa digunakan untuk Disaster Management preparedness seperti pengambil informasi lokasi bencananya. Namun dalam pemanfaatan data laporan bencana diperlukan proses dan tujuan yang tepat untuk dijadikan solusi untuk mengurangi dampak bencana.



Gambar 2. Halaman Laporan Berita BPBD Jawa Tengah

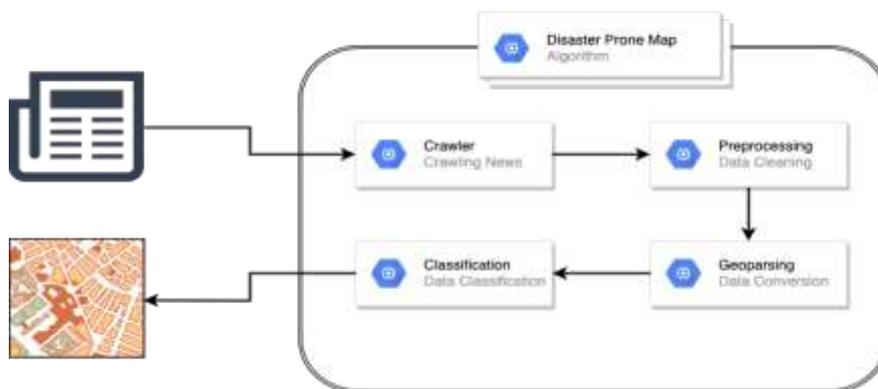
Pemanfaatan IT dapat membantu mengolah data yang dinamis yaitu data yang sulit untuk diproses oleh manusia dan mampu memberikan hasil pengolahan yang lebih baik dengan lebih cepat seperti data laporan bencana BPBD Jawa Tengah [8]. Salah satu pemanfaatan IT untuk melakukan preparedness terhadap bencana alam adalah menerapkan GIS (*Geographical Information System*). Menerapkan GIS sangat membantu untuk mengetahui resiko terjadinya kebencanaan disuatu lokasi [9]. Dalam penelitian [10] menggunakan GIS sangat bermanfaat untuk melakukan pemodelan untuk mengetahui resiko terjadinya bencana banjir di negara Malaysia. Mereka mengestimasi tingkat terjadinya bencana banjir dengan melakukan penelitian Peta Rawan Bencana. [11] Dalam penelitian lain untuk mengetahui tingkat pada Peta Rawan Bencana, mereka menerapkan GIS dengan menggunakan data jumlah banyak terjadinya bencana disuatu titik lokasi terutama yang paling dekat dengan pemukiman. Maka data laporan bencana BPBD Jawa Tengah akan sangat bermanfaat untuk preparedness dalam Disaster Management untuk meningkatkan Peta Rawan Bencana.

GIS pada umumnya adalah terdiri dari data spasial yang digunakan untuk diolah sehingga didapatkan informasi terhadap lokasi kebencanaan [12]. Data spasial harus terdiri dari data koordinat atau bentuk lain seperti polygon, line atau titik. Untuk mengembangkan Peta Rawan bencana dibutuhkan data spasial yang sesuai untuk memberikan informasi kebencanaan berdasarkan lokasi. Namun untuk meningkatkan Peta Rawan bencana dari laporan BPBD Jawa Tengah, laporan tersebut tidak ada informasi spasial yang berupa kordinat melainkan hanya berupa penulisan nama tempat berupa teks. Laporan bencana BPBD Jawa Tengah yang berupa data teks harus diolah terlebih dahulu sehingga bisa dibuat Peta Rawan Bencana dengan menggunakan metode *Geoparsing* [13]. Dengan alasan yang sama dalam penelitian [14] menyebutkan masifnya data di internet mengandung banyak potensi data geospasial yang bisa diolah. Dikatakan pada awalnya data yang ada memang tidak terdapat

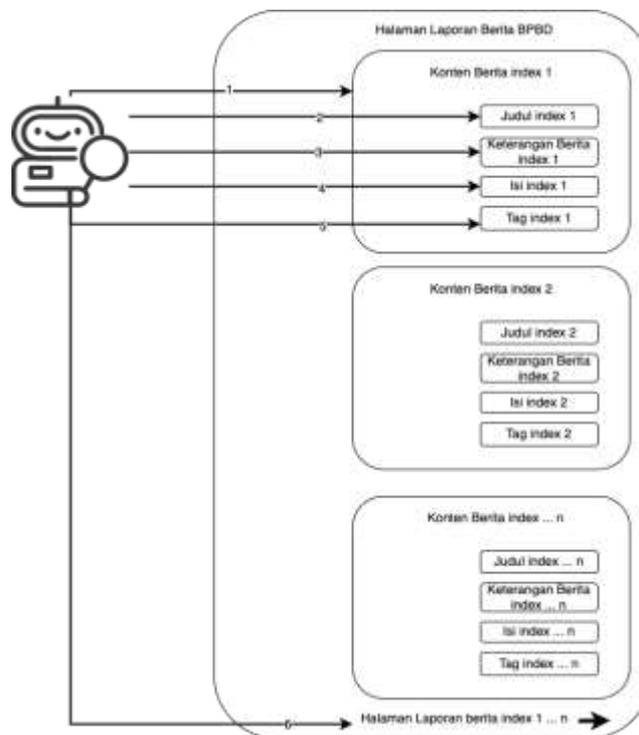
data spasial tetapi dengan menggunakan metode *Geoparsing*, data yang berupa teks dapat dikonversi menjadi data spasial. Hal yang sama juga ada di laporan bencana BPBD Jawa Tengah yaitu potensi data geospasial yang dapat digunakan untuk Peta Rawan Bencana pada Tahap preparedness dalam Disaster Management sehingga jumlah korban dan kerugian terhadap terjadinya bencana bisa berkurang.

2. Metode Penelitian

Dalam peningkatan Peta Rawan Bencana dengan metode *Geoparsing* dilakukan berupa tahapan tahapan untuk membuat sebuah algoritma yang utuh. Ditunjukkan oleh Gambar 3 bahwa untuk mendapatkan berita pada halaman BPBD Jawa Tengah mulanya diperlukan *Crawler* berita lalu hasil *Crawler* tersebut dilakukan *Preprocessing* yaitu Pembersihan Data. Tahapan pembersihan data memegang peranan yang sangat penting karena data harus disesuaikan terlebih dahulu hingga menjadikan data tersebut bisa diolah dan dianalisis lebih lanjut. Setelah selesai dengan pembersihan data dilakukan *Geoparsing* dengan Data Kebencanaan yang sudah dibersihkan sehingga bisa dilakukan proses *Classification* untuk membuat peta Rawan Bencana. Langkah Langkah tersebut selanjutnya akan dijelaskan dengan detail dalam paragraf paragraf setelah ini.



Gambar 3. Metode *Geoparsing Disaster Prone Map*



Gambar 4. Robot *Scrapper* Berita BPBD Jawa Tengah

Tantangan terbesar untuk membuat peta rawan bencana adalah dengan penyediaan data dari berapa banyak terjadinya bencana terjadi dalam suatu daerah. Sumber data yang digunakan dalam penelitian ini adalah bersumber dari berita kebencanaan yang ada di BPBD Jawa Tengah. Berita dalam Website BPBD Jawa Tengah berjumlah hingga ribuan sehingga sangat kesulitan untuk didapatkan maka diperlukan cara untuk mengambil berita tersebut menggunakan robot *scraper* [15]. *Robot scraper* memungkinkan bahwa data berita yang berjumlah ribuan halaman bisa di ambil dalam hitungan menit yang tidak mungkin dilakukan oleh manusia. Ini memungkinkan untuk mengambil judul, keterangan (author dan waktu), isi berita dan tag yang diberikan untuk berita tersebut seperti ditunjukkan pada Gambar 4.

Setelah data diperoleh dari hasil robot *scraper* maka data tersebut haruslah melakukan pembersihan data. Dalam membuat peta rawan bencana setiap satu berita harus mewakili kejadian yang terjadi dalam satu daerah. Maka untuk mendapatkan data tersebut hanya digunakan tag pada komponen berita hasil *scraper* seperti ditunjukkan pada komponen Gambar 5.



Gambar 5. Komponen Berita BPBD Jawa Tengah

Dalam membuat peta rawan bencana diperlukan jumlah terjadinya bencana terhadap suatu daerah, maka dalam tag yang terkumpul dalam berita hanya diambil teks yang mewakili nama daerah. Untuk pengambilan data dengan yang mengandung unsur nama daerah dilakukan beberapa tahapan yaitu tahapan pertama adalah tahap Parsing. Karena tag tersebut memiliki beberapa informasi maka Parsing dilakukan untuk memecah string tersebut menjadi beberapa bagian untuk menjadi set symbol [16]. Parsing hanya memecah sebuah string menjadi beberapa bagian dalam hal ini ditentukan bahwa pemecahan *string* dengan pemisah karakter tanda koma (,).

Parsing tidak menunjukkan arti kata tetapi parsing dapat menunjukkan bagian mana saja yang harus dipisah maka dalam langkah selanjutnya pada tahap *Data Cleaning* adalah *Lexical Parsing*. *Lexical parsing* menempati posisi tahap kedua setelah parsing ditunjukkan oleh Gambar 6 yaitu artinya *Lexical Parsing* melakukan tahapan yang cukup penting peranannya karena harus disesuaikan dengan penelitian yang dilakukan. *Lexical parsing* sangat bermanfaat untuk mengekstraksi geospasial dari data tidak terstruktur. Dalam tahapan ini terjadi proses mengidentifikasi kata mana saja dalam potongan potongan string yang sudah dilakukan oleh parsing. Penelitian ini tidak dapat menggunakan geoname yang tersedia karena penulisan nama daerah pada website BPBD Jawa Tengah tidak memenuhi standar seperti menggunakan kata singkatan. Maka Langkah yang harus kami lakukan adalah melakukan mencocokkan seberapa banyak kata yang serupa muncul.



Gambar 6. Proses *Data Cleaning*

Selanjutnya adalah tahap *Semantic analysis* untuk menentukan mana kata yang mengandung arti dan sesuai dengan data penelitian ini yaitu kata yang mengandung nama

daerah. Proses ini dilakukan secara manual karena walaupun jumlah kata yang di hasilkan oleh parsing menjadi ratusan ribu namun setelah dilakukan proses *lexical parsing* dapat di kelompokkan menjadi beberapa kata saja.

Nama Daerah	Jumlah Laporan
Banjarnegara	41
Banyumas	199
Batang	11
Blora	34
Boyolali	9
Brebes	101
Cilacap	192
Demak	13
Grobogan	71
Jepara	45
Karanganyar	15
Kebumen	49
Kendal	62
Klaten	15
Kudus	79
Magelang Kab	248
Magelang Kota	
Pati	55
Pekalongan Kab	5
Pekalongan Kota	10
Pemalang	24
Purbalingga	64
Purworejo	26
Rembang	15
Salatiga	3
Semarang Kab	18
Semarang Kota	270
Sragen	37
Sukoharjo	8
Surakarta	15
Tegal Kab	24
Tegal Kota	18
Temanggung	64
Wonogiri	28
Wonosobo	38

Gambar 7. Hasil *Semantic Analysis*

Pada akhirnya data pada *lexical parsing* dibagi menjadi dua jenis yaitu kata yang mengandung nama daerah dan kata yang tidak mengandung nama daerah. Pada tahap *Semantic Analysis* ini memberikan hasil sebuah baris kata yang mengandung unsur nama daerah dan frekuensi muncul dalam *corpus* laporan bencana selanjutnya disebut *corpus* laporan bencana seperti di sajikan pada Gambar 7.

Geoparsing adalah metode cara mengekstrak nama tempat dalam bentuk teks dan membantu peneliti untuk menyesuaikan nama tempat tersebut dengan kordinatnya [12]. Sebelumnya peneliti sudah menyiapkan beberapa peta polygon daerah dalam provinsi Jawa Tengah Indonesia yang memili data lengkap berupa kordinat dengan nama daerahnya. Untuk menyesuaikan dengan metode *Geoparsing* dilakukan penyesuain penyebutan nama daerah dengan hasil pembersihan data dari *Semantic Analysis* seperti Gambar 7. Selanjutnya *Geoparsing* dilakukan menyesuaikan nama nama daerah yang konsisten dalam tag tersebut dengan memberikan langkah tambahan yang tidak dilakukan oleh tahap *preprocessing* seperti merubah nama daerah yang sudah di data cleaning menjadi menggunakan huruf kecil semua dan penyesuaian nama daerah yang sama dengan berbeda bentuk wilayah (penyebutan kabupaten dan kota). Lalu yang terakhir pada tahap *Geoparsing* peta rawan bencana adalah memetakan dan mencocokkan semua nama daerah yang dikumpulkan dalam bentuk *corpus* laporan bencana dalam suatu daerah dengan nama daerah yang terdapat pada kumpulan polygon dalam peta Jawa Tengah hingga didapatkan kordinat kordinat polygon yang dibutuhkan yang merepresentatif dengan *corpus* laporan bencana.

Setelah data *corpus* laporan bencana selesai dipetakan dengan metode *Geoparsing* dilakukan klasifikasi terhadap jumlah terjadinya bencana dalam suatu daerah dengan melakukan pewarnaan. Dalam melakukan klasifikasi terhadap data tersebut, penggunaan

bencana dengan metode *Geoparsing* yang dibuat ini masih membutuhkan data pembantu seperti data satelit dan pengklasifikasian jenis bencana sehingga bisa memberikan persiapan yang spesifik pada *Preparedness Disaster Management*.

Daftar Pustaka

- [1] A. Usman, "Integrated disaster risk management in Indian environment: Prediction, prevention and preparedness," *GHTC 2017 - IEEE Glob. Humanit. Technol. Conf. Proc.*, vol. 2017-January, pp. 1–6, 2017.
- [2] J. Widagdo, D. D. Putra, B. Syihabuddin, T. Juhana, E. Mulyana, and A. Munir, "Android-based Disaster Management Application for After-Disaster Rapid Mobile Assessment," *IoTaIS 2020 - Proc. 2020 IEEE Int. Conf. Internet Things Intell. Syst.*, pp. 201–204, 2021.
- [3] T. Pujadi and W. Sardjono, "Evaluation of Knowledge Management System for Disaster Management Using Factor Analysis," *Proc. 2018 Int. Conf. Inf. Manag. Technol. ICIMTech 2018*, no. September, pp. 327–332, 2018.
- [4] V. Nunavath and M. Goodwin, "The Role of Artificial Intelligence in Social Media Big data Analytics for Disaster Management -Initial Results of a Systematic Literature Review," *2018 5th Int. Conf. Inf. Commun. Technol. Disaster Manag. ICT-DM 2018*, no. 1, pp. 1–4, 2019.
- [5] F. A. Bazley, A. Maybhate, C. S. Tan, N. V. Thakor, C. Kerr, and A. H. All, "Enhancement of bilateral cortical somatosensory evoked potentials to intact forelimb stimulation following thoracic contusion spinal cord injury in rats," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 5, pp. 953–964, 2014.
- [6] S. Raffel and S. Stephens, "Extended Abstract: How Hurricane Visualization Tools Affect the Public's Perception of Risk and Preparedness," *IEEE Int. Prof. Commun. Conf.*, vol. 2020-July, pp. 186–187, 2020.
- [7] T. D. Scheneid and L. Collins, *Disaster Management and Preparedness*. CRC Press, 2000.
- [8] National Research Council, *National Research Council. Improving disaster management: the role of IT in mitigation, preparedness, response, and recovery*. 2007.
- [9] L. Gigović, D. Pamučar, Z. Bajić, and S. Drobniak, "Application of GIS-interval rough AHP methodology for flood hazard mapping in Urban areas," *Water (Switzerland)*, vol. 9, no. 6, pp. 1–26, 2017.
- [10] H. Mojaddadi, B. Pradhan, H. Nampak, N. Ahmad, and A. H. bin Ghazali, "Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS," *Geomatics, Nat. Hazards Risk*, 2017.
- [11] I. C. Nicu, "Tracking natural and anthropic risks from historical maps as a tool for cultural heritage assessment: A case study," *Environ. Earth Sci.*, vol. 76, no. 9, p. 330 2017.
- [12] C. J. Rupp, P. Rayson, I. Gregory, A. Hardie, A. Joulain, and D. Hartmann, "Dealing with heterogeneous big data when geoparsing historical corpora," *Proc. - 2014 IEEE Int. Conf. Big Data, IEEE Big Data 2014*, no. September 2013, pp. 80–83, 2015.
- [13] J. de Bruijn, H. de Moel, B. Jongman, J. Wagemaker, and J. Aerts, "TAGGS: Grouping Tweets to Improve Global Geotagging for Disaster Response," *Nat. Hazards Earth Syst. Sci. Discuss.*, pp. 1–22, 2017.
- [14] L. Nizzoli, M. Avvenuti, M. Tesconi, and S. Cresci, "Geo-semantic-parsing: AI-powered geoparsing by traversing semantic knowledge graphs," *Decis. Support Syst.*, vol. 136, no. March, p. 113346, 2020.
- [15] M. S. Rohman, H. A. Santoso, and G. W. Saraswati, "Pemanfaatan Topic-Focused Crawler untuk Pembangunan Corpus Berita Bencana menggunakan Teknik Scrapy CSS Selector," *Semin. Nas. APTIKOM 2019*, pp. 250–258, 2019.
- [16] J. W. Karl, "Mining location information from life- and earth-sciences studies to facilitate knowledge discovery," *J. Librariansh. Inf. Sci.*, vol. 51, no. 4, pp. 1007–1021, 2019.