Geodesy and Cartography
ISSN 2029-6991 / eISSN 2029-7009
2023 Volume 49 Issue 4: 216–221
https://doi.org/10.3846/gac.2023.18111

UDC 528.935

APPLICATION OF NATURAL LANGUAGE PARSING FOR IDENTIFYING NON-SURVEYED BOUNDARIES TOWARDS ENHANCED SYSTEMATIC LAND TITLING: RESULTS FROM PRELIMINARY EXPERIMENT

Joseph O. ODUMOSU [1,2*], Victor C. NNAM [3], Olurotimi A. KEMIKI [4], Abdulkadir ABUBARKAR [5], Micheal A. OYEBANJI [1], Sunday O. BABALOLA [2]

[1]Department of Surveying and Geoinformatics, Federal University of Technology, Minna, Nigeria
[2]Department of Surveying and Geoinformatics, Federal University, Oye Ekiti, Nigeria
[3]Department of Surveying and Geoinformatics, Enugu State University of Technology, Nssuka, Nigeria
[4]Department of Estate Management and Valuation, Federal University of Technology, Minna, Nigeria
[5]Department of Computer Science, Federal University of Technology, Minna, Nigeria

Received 19 November 2022; accepted 22 November 2023

Abstract. The need for the adoption of systematic land titling (SLT) in Nigeria cannot be overemphasised. Nonetheless, the problems of speed and cost of geospatial data acquisition, as well as identification of non-surveyed boundaries, remain unresolved, impeding the effectiveness of SLT for non-surveyed boundaries. The integration of language into Artificial Intelligence (AI) has allowed Natural Language Parsing (NLP) to effectively serve as a tool for communication between humans and computer systems. This study presents preliminary results of testing a prototype application that utilises NLP to convert textual descriptions into graphic sketches as a tool towards the production of a-priori sketches that can aid SLT in non-surveyed boundaries. The study determines that NLP alone cannot be used to achieve the required accuracy in geospatial data for SLT; however, the study concludes that NLP can be integrated alongside other ancillary information to enhance SLT in peri-urban regions.

Keywords: systematic land titling, non-surveyed boundaries, Natural Language Parsing, artificial intelligence, cadastral mapping.

Introduction

All human activities, be it religious, economic, cultural, etc., are performed on land, thus making land considerably the most valuable natural resource available to man. Unfortunately, land is limited in supply; hence the need for the continuous modification and adoption of various land tenure, policies and reform systems for its effective administration (Banire, 2006; Atilola, 2013; Otubu, 2018). In Nigeria, the need for land reforms is particularly obvious with over 95% of the entire country (about 900,000 km^2) not properly titled. Efforts to improve the land titling practice in the country, thus ameliorating the consequent challenges posed by the ineffectiveness of the initial land titling system, have brought the idea of systematic land titling (SLT) to the fore (Presidential Technical Committee on Land Reform, 2013; Oluwadare & Kufoniyi, 2017).

SLT is a system of registration whereby a specific geographical location is steadily worked through so that all adjacent parcels of land within the area are adjudicated upon and or surveyed, issued titles to and registered. It is a system that is usually initiated by the government or its appropriate agency. One of the key hindrances to this system in many developing nations has been attributed to the high costs of acquiring the needed geospatial information required for initiating and eventually completing the land titling process (Atilola, 2013). In a bid to address this problem, a fit-for-purpose (FFP) survey has been proposed as a means of obtaining reliable spatial data that could be integrated into the land titling procedure, especially in non-urban areas (Manyoky et al., 2012; Jazayeri et al., 2014). The FFP survey approaches are generally associated with the collection of accurate survey information about the environment using high-resolution images. Semi-automatic extraction of visible boundaries from

*Corresponding author. E-mails: *odumossu4life@yahoo.com; joseph.odumosu@fuoye.edu.ng*

high-resolution imageries obtained via unmanned aerial vehicles (UAV) is often incorporated with local knowledge from human operators and this practice has brought about significant improvement in cadastral mapping (cadastration) in terms of cost, time and accuracy (Crommelinck et al., 2016).

Nevertheless, the implementation of the FFP technique in land titling is not without challenges. Some challenges associated with FFP surveys include identifying boundaries of individual landholders (especially when such land parcels do not have any physically well-defined boundary forms), the cost of acquiring high-resolution imageries to be used as a base map, deployment of skilled manpower, etc. Attempts to improve boundary identification from FFP surveys have led to the application of several machine learning algorithms. Some of the algorithms used include agglomerative segmentation (García-Pedrero et al., 2017), multistage combinatorial grouping (MCG) (Pont-Tuset et al., 2017), random forest (Crommelinck et al., 2019a), a combination of fully convolutional networks and combinatorial grouping (Persello et al., 2019) and incorporation of additional functionalities such as connect around selection/optimal paths into the MCG technique (Crommelinck et al., 2019b).

Whereas mapping visible boundaries from high-resolution satellite imageries has improved significantly due to developments in supervised machine learning algorithms, identifying non-visible boundaries from imageries still remains an emerging area of research interest that is yet to be fully developed. Therefore, this technique is mostly inadequate for automated mapping in peri-urban areas (where some of the land parcels do not have physically recognised boundaries).

Although some peri-urban lands have neither visible boundaries that can be recognised via imageries nor cadastral survey plans/sketches, the individual landholders of such land often have very correct descriptive knowledge of their property boundaries by referring to specific land features. Therefore, this study presents a novel approach wherein language parsing technique is used to convert textual descriptions into spatial representations (sketches) as a-priori spatial information. These sketches can later be further integrated with other spatial information available within the vicinity of such land parcels as a basis for SLT in peri-urban areas. This means that the proposed method is expected to complement existing deep learning techniques of boundary extraction earlier mentioned towards achieving improved reliability in the application of artificial intelligence (AI) in the cadastral mapping process.

1. Natural language parsing for boundary identification

The goal of natural language processing/parsing (NLP) is to make computers understand unstructured text and retrieve meaningful information from it. NLP is a subfield of AI that basically involves fostering interactions between computers and humans, such that the computers can implement unstructured human language/instructions (Kibble, 2013). The essence of NLP is to make the computer system understand and execute regular human interactions without the human having to structure commands into the systems' language. Language parsing for geospatial positioning and analysis is similar to the principle of geocoding. Essentially, geocoding involves the translation of text-based information about location (address, zip code, names of localities, etc.) into numerical geographic coordinates such as longitude and latitude (Owusu et al., 2018). Similar to the principle of geocoding, language parsing uses explicit reference datasets such as digital road networks or identified land features to identify the location that best matches the input address (Owusu et al., 2018).

Improvements in computational linguistics have led to the use of algorithms and AI-based methods that use language for achieving specific tasks. The integration of language into the field of AI in the form of NLP is usually implemented using either named entity recognition (NER), sentiment analysis (SA), text summarisation (TS), aspect mining (AM) or topic modelling (TM). Sohail (2020) identified that NLP can be used for boundary disambiguation, mapping crime reports and entity extraction, although the issue of accuracy of AI for geospatial analysis is still a subject of concern.

Therefore, this study proposes that a-priori estimates of the position of unsurveyed land boundaries can be obtained via NLP. If the a-priori location is identified using the NER, for instance, other AI tools such as pattern and shape recognition can later be used to carry out a hierarchical or sequential repositioning of the boundary based on the road geometry from FFP, survey plan of adjoining properties (where available) and other ancillary information within the specific area. Consequently, the use of NLP for identifying non-surveyed boundaries can be explored for SLT not as a stand-alone solution but to serve as an initial estimate for unidentified boundaries, which would later be integrated with other spatial information. Also, this method is suggested only for SLT where the expanse of land to be regularised is small.

2. Material and methods

This section presents the materials and methods used for designing and implementing the NLP application in this study. For the purpose of the study, the study area chosen was part of the Federal University of Technology, Minna, Nigeria. The study area was chosen to serve as a typical example of a peri-urban area where some buildings and land parcels are well defined and can easily be seen from imagery, while other areas do not have any physically distinguishable trait that defines their limits. Let us assume that SLT is to be carried out within this said portion of land within the Federal University of Technology, Minna, Nigeria, as shown in Figure 1.

Figure 1. The study area

2.1. Materials

Specific materials used in the design and implementation of the application are:

(1) Natural Language Parsing Toolkit (NLTK): a Python programming library – used for parsing user-supplied land addresses.

(2) Django: a Python framework for web development – used for building the application backend, thus making the application web-based.

(3) Hypertext Markup Language (HML) and Cascading Style Sheet/Bootstrap (CSS) – used for the application front-end design and styling the application front end, respectively.

(4) JQuery: a JavaScript framework for interaction on the front end and plotting on the Google Map.

(5) Google Earth image: to serve as a base map and reference map for guiding the users' description. Eventually, it is auto-linked to the application via web service. In a well-developed system, this can be replaced with an orthophoto.

2.2. Method

The conceptual design of the NLP system for boundary identification is shown in Figure 2. The application architecture is designed such that the user provides specifications by typing the textual description of his land parcel into the system. The description must include a primary reference point, from where the system automatically takes its spatial reference. At the backend, the language parser converts the texts to coordinates and returns the coordinates into an online interface. At the online interface, a

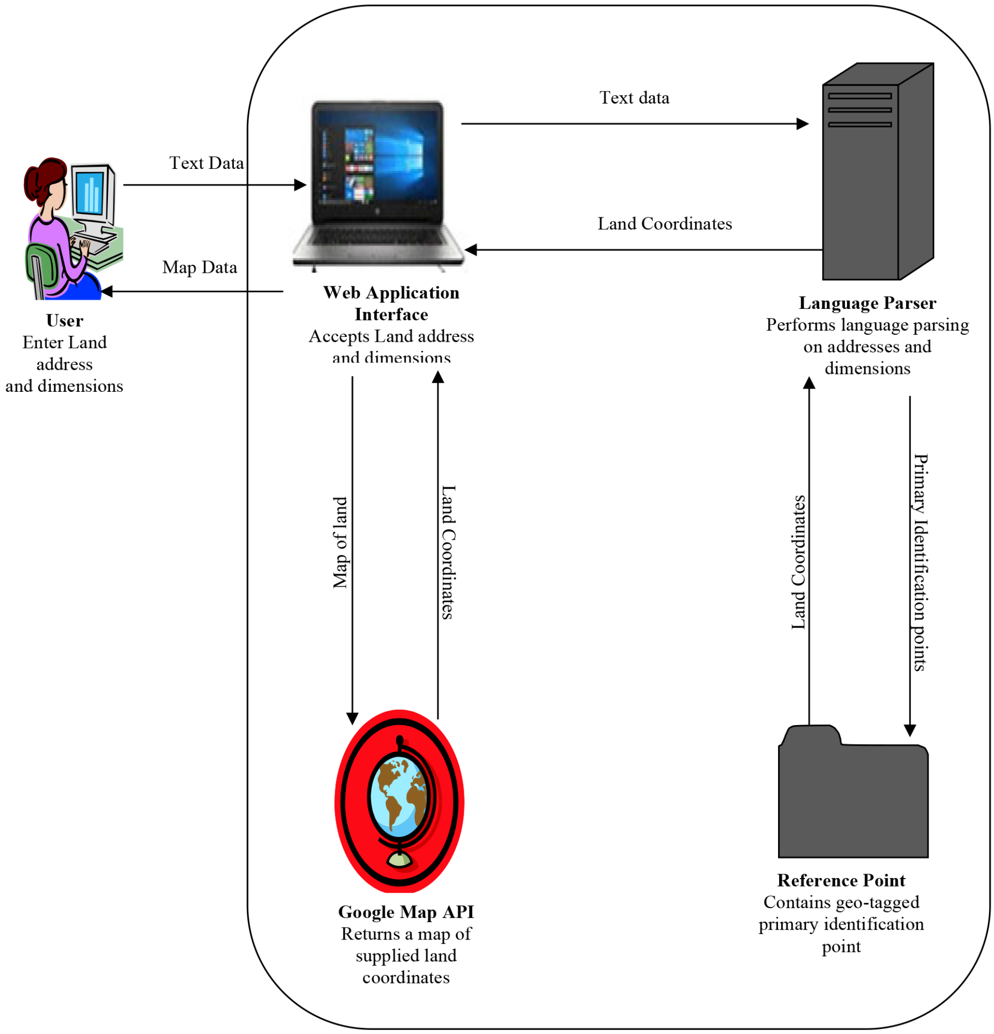


Figure 2. Application design

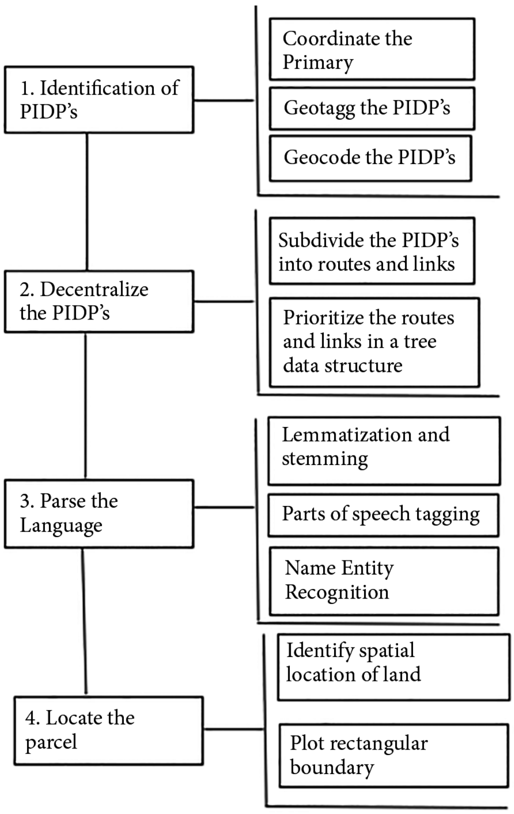


Figure 3. Design implementation workflow

Google application programming interface (Google API) is linked which in turn returns a plot of the spatial location and dimension of the land parcel to the user.

The design was implemented by a sequential implementation of the steps presented in the workflow diagram shown in Figure 3. The implementation begins by identifying primary identification points (PIDPs) or key landmark features which are likely to attract the attention of any person who desires to locate or describe a parcel within the area. Some of the common PIDP used in the study include bus stops, gates, etc. A major rule guiding the choice of the PIDP is that such points must be conspicuous and near permanent landmark features which can easily be identified and whose coordinates can easily be determined on the adopted base map (in this case, Google Earth). For this study, a total of 32 PIDPs were identified and coordinated. These points are then geotagged and geocoded so as to allow the system to identify them whenever a user describes a land parcel using such PIDPs.

The geocoded PIDPs are then subdivided and linked into routes. For instance, all PIDPs along a specific route are aggregated to form a link, with the system taking note of their order of arrangement. All possible links and sublinks are identified accordingly in a tree-like data structure, such that each PIDP is a parent while the subsequent PIDPs following it in that link and other connected

a)

b)

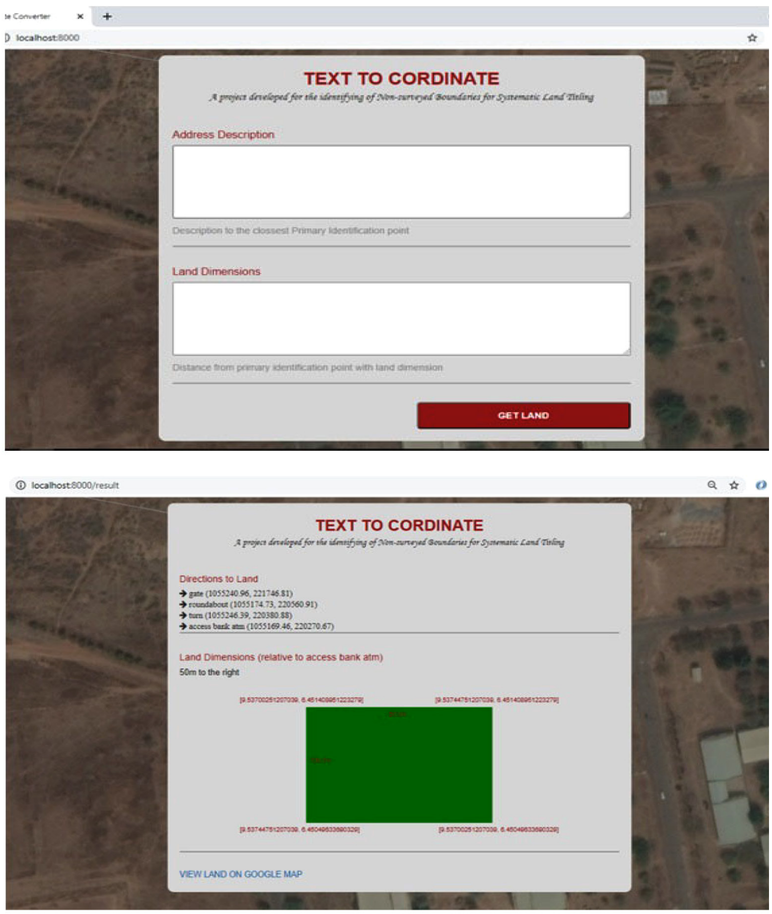


Figure 4. Application front-end: a – textual description entry form (description page);
b – description has been geocoded into coordinates

sub-links are children. For this reason, the geotagged PIDPs were formatted into the Javascript Object Notation (JSON) format. A screenshot of the application front end for description entry and result display is shown in Figures 4a and 4b.

Two major language parsing techniques (i.e. the Part of Speech [PoS] tagging and the NER) were used to parse the textual descriptions into system language. PoS tagging is a branch of AM. AM identifies the different aspects in the text. In this case, it is used in conjunction with SA so as to extract complete information from the text. NER is then used to extract the entities in the text. The extracted entities are then identified and transformed using the geocoding information obtained earlier. Once the parcel of land is located, the dimensions of the parcel as specified are plotted on the Google Map.

3. Model testing and results

The application was tested using a textual description for a parcel within the study area. This was done in order to test the application and ascertain the correctness of the textual descriptions in relation to actual survey measurements. The selected parcel of land's dimensions were 25 m by 50 m and is located off the main road from the University roundabout. The textual description of the parcel used for the application testing is as given below:

"Enter through the *University main gate* and at the *University roundabout*, turn *right* towards *School of Environmental Technology*; then move *50 m along the road*. The parcel is located on your *left-hand side, 50 m long by 25 m wide*".

The above textual description is transformed via NLP such that the key words written in italics are identified and mapped on the Google Map environment by linking with the Google API. Based on the transformation, the graphical description of the identified parcel is shown in Figure 5a, while an overlay of the NLP-derived boundary with the precise survey is shown in Figure 5b.

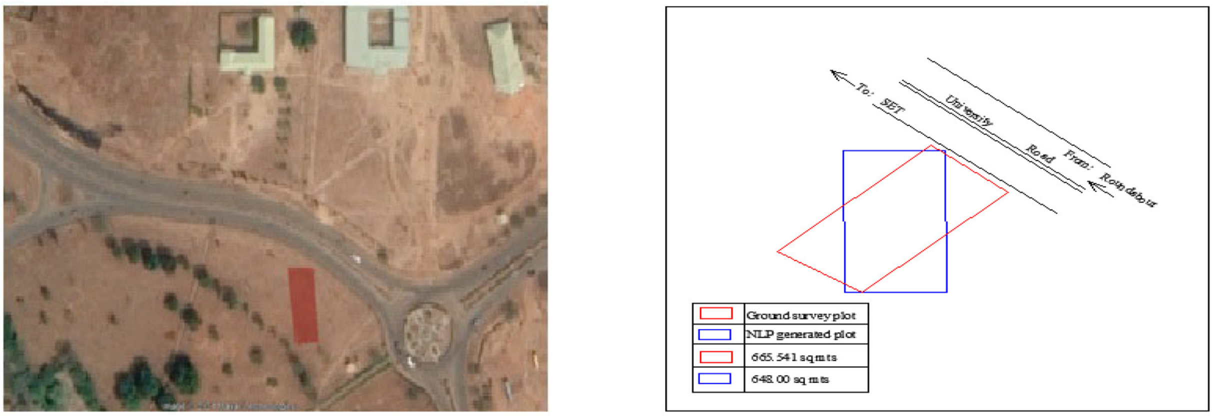
3.1. Discussion of results

Analysis of the overlap between the two boundaries as identified shows that the automatically identified boundary tallies with the precisely surveyed boundary by about 68%, although it does not align with the true orientation of the actual property. The discrepancy observed between them is a consequence of the fact that the automatically identified boundary does not consider parcel orientation (angles and alignments) in plotting the boundary while the actual survey is properly referenced to consider the true orientation of the parcel. Careful observation of the behaviour of the system indicates that in the worst case, the NLP-generated plot would lose about 50% of the actual land area and this is a major limitation of the proposed method. This shows that the adjacency accuracy for the automatic delineation is poor.

Furthermore, it is observed that the length and breadth of the parcel obtained from the automatic delineation is very close to that of the precise survey. This is because the correct dimensions were specified during the textual description. Invariably, it can be inferred that the length accuracy depends on the correctness of the description.

Again, the automatically delineated boundary takes into cognisance the dimensions of the parcel as a perfect rectangle, with the measurements taken anticlockwise with respect to the specified starting point. The inability of the NLP to identify orientations is a major limitation of the system and this makes it unable to properly represent even the dimensions of parcels that are not rectangular. For instance, when the NLP was tested for a five-sided parcel, it simply returned a pentagon as against the actual shape of the parcel (not shown in this paper).

This striking limitation of the NLP, however, can be overcome by incorporating a pattern recognition algorithm into the system. Using pattern recognition, the system automatically aligns the width (breadth) of the rectangle with the nearest road (lineament) feature around it. And in the case of non-rectangular parcels, the pattern recognition would be structured to fit the parcel into the nearest unoccupied parcel around the area. For this reason, it is



a) b)

Figure 5. Plot of the described parcel: a – the NLP-generated plot using the Google API (shown in blue); b – overlay of the NLP-generated plot with plot of actual ground survey (shown in red)

recommended that the intelligence of the system should be guided by limiting the application of this method to peri-urban areas where most parcels are already properly defined and there are only a few parcels that are yet to be surveyed in an SLT exercise. Also, restricting the application of this proposed method to SLT in a well-surveyed peri-urban area would allow the proposed application to learn shapes and patterns from the several other available survey plans and sketches that would have been loaded on the system. The application would then be left with only a few unidentified boundaries to resolve, for which incorporation of NLP and other deep learning methods could be used.

Conclusions

The use of NLP for identifying non-surveyed boundaries has been examined in this study and its major limitations are identified. It has been shown that NLP can be used to provide a sketch of the approximate location of parcels. It has also been identified that the provided sketch from NLP alone is not suitable for a title document. Nevertheless, as in the case of SLT where all land parcels within a specified area are registered simultaneously, the NLP can be further complemented with shape, pattern and geometry recognition algorithms to match adjacent roads and existing survey plans to provide a reasonable sketch for non-surveyed boundaries towards achieving an efficient SLT system. Notwithstanding the limitations noted above, this study has provided practical evidence for the capability of using NLP to provide graphical sketches of land parcels in a bid to achieve an efficient and low-cost SLT scheme.

Declarations

The authors declare no conflict of interest.

References

Atilola, O. (2013, May 6–10). Systematic land titling in Nigeria: Geoinformation challenges. In *FIG Working Week 2013* (pp. 1–13), Abuja, Nigeria. https://doi.org/10.1177/1365480212474731

Banire, M. A. (2006). *Land management in Nigeria: Towards a new legal framework*. Ecowatch Publication Lagos.

Crommelinck, S., Bennett, R., Gerke, M., Nex, F., Yang, M. Y., & Vosselman, G. (2016). Review of automatic feature extraction from high-resolution optical sensor data for UAV-based cadastral mapping. *Remote Sensing*, *8*(8), 689. https://doi.org/10.3390/rs8080689

Crommelinck, S., Koeva, M., Yang, M. Y., & Vosselman, G. (2019a). *Robust object extraction from remote sensing data*. arXiv:1904.12586. https://arxiv.org/abs/1904.12586

Crommelinck, S., Koeva, M., Yang, M. Y., & Vosselman, G. (2019b). Application of deep learning for delineation of visible cadastral boundaries from remote sensing imagery. *Remote Sensing 11*(21), 2505. https://doi.org/10.3390/rs11212505

García-Pedrero, A., Gonzalo-Martín, C., & Lillo-Saavedra, M. (2017). A machine learning approach for agricultural parcel delineation through agglomerative segmentation. *International Journal of Remote Sensing*, *38*(7), 1809–1819. https://doi.org/10.1080/01431161.2016.1278312

Jazayeri, I., Rajabifard, A., & Kalantari, M. (2014). Geometric and semantic evaluation of 3D data sourcing methods for land and property information. *Land Use Policy*, *36*, 219–230. https://doi.org/10.1016/j.landusepol.2013.08.004

Kibble, R. (2013). *Introduction to Natural Language processing: Undergraduate study in computing and related programmes*. University of London, Department of Computing, Goldsmiths.

Manyoky, M., Theiler, P., Steudler, D., & Eisenbeiss, H. (2012). Unmanned aerial vehicles in cadastral applications. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *38*, 57–62. https://doi.org/10.5194/isprsarchives-XXXVIII-1-C22-57-2011

Oluwadare, C. O., & Kufoniyi, O. (2017, March). Comparative study of sporadic and systematic methods of land titling and registration in Ondo State, Nigeria. In *Environmental Design and Management International Conference (EDMIC): Responsive Built Environment* (pp. 9–13). The Faculty of Environmental Design and Management, Obafemi Awolowo University, Ile-Ife.

Otubu, A. (2018). The land use act and land administration in 21st century Nigeria: Need for reforms. *Journal of Sustainable Development, Law and Policy*, *9*(1), 80–108. https://doi.org/10.4314/jsdlp.v9i1.5

Owusu, C., Lan, Y., Zheng, M., Tang, W., & Delmelle, E. (2018). Geocoding fundamentals and associated challenges. In H. A. Karimi & B. Karimi. (Eds.), *Geospatial data science techniques and applications*. CRC Press. https://doi.org/10.1201/b22052

Persello, C., Tolpekin, V. A., Bergado, J. R., & de By, R. A. (2019). Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sensing of Environment*, *231*, 111253. https://doi.org/10.1016/j.rse.2019.111253

Pont-Tuset, J., Arbeláez, P., Barron, J. T., Marques, F., & Malik, J. (2017). Multiscale combinatorial grouping for image segmentation and object proposal generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(1), 128–140. https://doi.org/10.1109/TPAMI.2016.2537320

Presidential Technical Committee on Land Reform. (2013). *Draft regulations on land use act, 2013.* Abuja, FCT Nigeria.

Sohail, S. (2020). *Geospatial natural language processing*. https://medium.com/geoai/geospatial-natural-language-processing-ee3fc6ea6939