Community-Connect: COVID-19 Small Business Marketplace with Automated Regulation Matching and Company Lead Retrieval

Erik Coltey[1], Shahin Vassigh[2], Shu-Ching Chen[1]
[1]*Knight Foundation School of Computing and Information Sciences*
Florida International University
Miami, FL, United States
[2]*College of Communication, Architecture and the Arts*
Florida International University
Miami, FL, United States
ecolt003@cs.fiu.edu, svassigh@fiu.edu, chens@cs.fiu.edu

***Abstract*—Periods of unique economic distress such as the COVID-19 pandemic can be quite difficult for small businesses. Challenges acquiring the supplies necessary to adhere to safety regulations created in the wake of such events can introduce stress on these businesses. This is further exacerbated when supply chains have slowed down, leading to global shortages from most large suppliers. This paper proposes a platform to aid such businesses in procuring COVID-19 related supplies such as Personal Protective Equipment (PPE) from one another, leveraging advanced data acquisition, integration, and Natural Language Processing (NLP) methods. With the pandemic end in sight, the platform described in this paper can be reused for other emergencies such as hurricanes, floods, among others. The proposed platform supports business transactions within a Buyer's Club (BC), keyword-based sourcing of new businesses to join the platform, and matching products to relevant regulations using greater-than-word length encoding, helping businesses comply with the ever-changing regulatory landscape.**

***Index Terms*—COVID-19, Disaster, Natural Language Processing, Data Acquisition, Data Retrieval, User Interfaces**

I. Introduction

COVID-19 has had devastating effects across all economic sectors, with small businesses being hit exceptionally hard. Between February and April 2020, the number of active business owners contracted from 15.0 million to 11.7 million, which was the largest reduction of owners in recent history. This surpassed even the contraction from the Great Depression [1]. A survey was also conducted on 5,800 small businesses at the beginning of the pandemic, indicating they had grappled with huge layoffs, and only a few had enough cash on hand to weather the first few months of the pandemic [2]. This is also juxtaposed with shortages of reliable Personal Protective Equipment (PPE) across healthcare centers, with many organizations having to refer to using additive technologies such as 3D printing to help alleviate the PPE shortage [3].

The pandemic took everyone by surprise as many businesses were without a plan to deal with the issues of procurement during an emergency. Although the availability of vaccines has brought the pandemic almost to an end, it has had a lasting impact, compelling businesses to plan for the future. Challenges include the slowdown of supply chains globally, the loss of existing supplier networks due to slowdowns and suppliers closing, and keeping updated on the status of regulations. Building a platform to host small business products and transactions is necessary to help both inform and reconnect businesses that may have been split off from existing networks of trusted suppliers. When it comes to finding small local business leads to rebuild these networks, the most common methods are to use specialized software. However, such software can be cumbersome or expensive to use. Many business-facing marketplaces only provide basic guidance on what products to purchase during the pandemic, which can lead to confusion with products such as face masks, of which there are many variants with potentially different sets of regulatory requirements.

In this paper, we propose a platform for Business to Business (B2B) transactions among small local businesses to help alleviate some of the challenges brought on by disasters such as COVID-19. This platform focuses on allowing businesses to procure items in bulk either from approved suppliers or from other approved businesses that may have a surplus. Data acquisition strategies are proposed to allow administrators to source companies with minimal inputs. Business sourcing can help rebuild broken B2B networks, with the administrator acting as a middle man to connect suppliers with buyers. Usage of the Universal Sentence Encoder (USE) grater-than-word length encoder architecture allows for information on relevant regulations and guidance to be displayed without human intervention, saving both buyers and suppliers time in researching the regulations and guidelines which apply to their products.

II. Related Work

Business to business (B2B) commerce focuses on the communications wherein which businesses are both the consumer and suppliers of products. Sales between businesses can reach

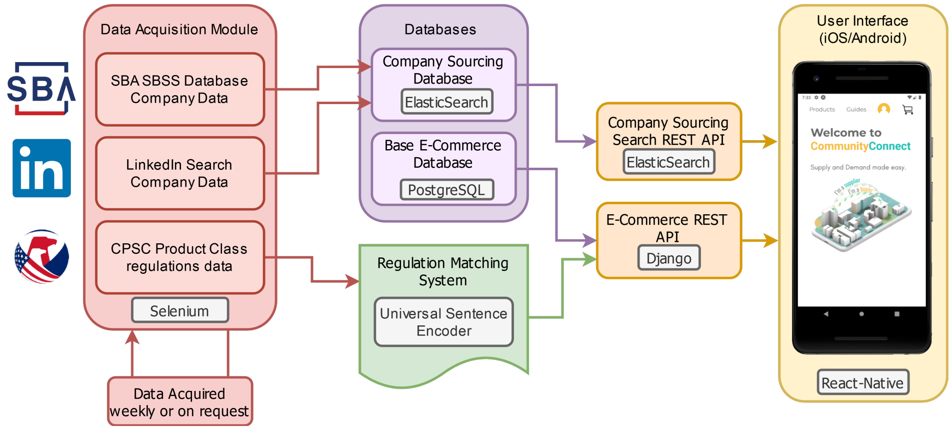


Fig. 1: Block Diagram of the application architecture, denoting data sources and the integrated regulation matching system.

the many billions of dollars per year globally, which represents a significant market share, especially as businesses are likely to buy in bulk to meet the needs of their entire workforce. However, B2B businesses have been hit hard by COVID-19, with many of the markets that B2B firms supply to have vanished in a matter of months [4]. This is extremely challenging where local or global supply chains become increasingly constrained. Shortages can lead to businesses not being able to procure supplies from existing suppliers, making it significantly more difficult to operate throughout such challenging conditions.

Advancements in pre-trained NLP models and data acquisition have allowed for the creation of increasingly sophisticated text-based systems. Models such as CoVe, ELMo, GPT-3, and BERT have proven their usefulness in creating meaningful word embeddings based on extensive corpora of text [5]. These models can be used for downstream tasks such as summarization and semantic similarity, which are incredibly useful when dealing with regulatory data. Attempts have been made to pair advanced NLP methods with such data, especially in the realm of building code compliance, using both traditional rule-based and deep learning based methods to create automated compliance checking systems [6], [7].

Data acquisition has been a driving force for business growth in recent years, with big data being a necessary tool for businesses to function as efficiently as possible. Such business intelligence tools have gained increased interest over the years, being considered a revelation to most companies [8]. Many challenges are present when considering the multi-source data acquisition, transmission, and fusion needed to develop novel and useful business intelligence systems [9]. Useful problem-specific data can at times be difficult to find when using traditional web searching methods [10], and using such data will necessitate the development of custom data acquisition software.

III. B2B For COVID-19

To begin providing value for local businesses during emergency situations such as the COVID-19 pandemic, it is first necessary to build the platform where they could conduct marketplace transactions, further serving the purpose of integrating novel algorithms and generating new datasets for a relatively unexplored domain.

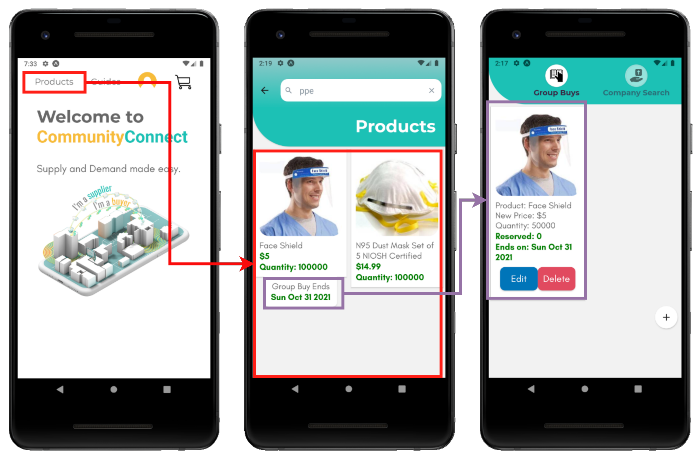


Fig. 2: Usage diagram demonstrating the group buy functionality of the B2B platform.

The proposed application was developed for mobile devices, with the intent of allowing local businesses to participate in a BC. BCs are used primarily to take advantage of the collective buying power of their members. With this system, administrators can negotiate with suppliers in the BC to sell supplies in bulk at a discount, whilst they focus on collecting the buyers necessary to fill such bulk orders. In this case, such bulk buys are demonstrated using the concept of a group buy as shown in Figure 2. The price of an item is reduced during a specific time window, and businesses can order that product at the reduced price until the group buy quantity limit is reached, or allotted time runs out.

Figure 1 demonstrates a block diagram of the end-to-end architecture of the marketplace platform. The front-end built using the React-Native to achieve parity on both Android and iOS devices with only a single code base. The back-end server architecture was built using Django, ElasticSearch, and PostgreSQL. Using Django allows for the simplified development and deployment of algorithms and data acquisition techniques with Python, using existing libraries such as Selenium and Tensorflow [11]. The back-end consists of different modules. One is a data acquisition module for collecting information to train and deploy different multimedia algorithms. There is also a database module exposed to traditional REST APIs to interact with the website front-end, and the regulation matching algorithm to aid businesses in knowing what to adhere to when purchasing products.

IV. Sourcing Local Businesses

As the pandemic continues to shutter businesses and slow down supply chains, traditional supply networks become strained. Many products that otherwise would be significantly easier to buy now become near impossible to procure. Sourcing local businesses as both buyers and suppliers is necessary to help alleviate procurement issues during such disasters. Integrating sourcing capabilities within the platform itself rather than as an external service can then allow for easy access to companies for administrators without needing to learn any external software. Integrating the sourcing system into the platform itself also allows it to be easily extended in the future for automated business lead acquisition based on trends from both within the application such as order volume or external multimedia data.

A. Data Acquisition

To automatically source for local businesses based on simple search queries, a portion of the data acquisition module is dedicated towards searching for companies using the small business association (SBA) SBSS Database and LinkedIn. These services were chosen as their search results are comprised of a large portion of small businesses, and can include valuable lead information such as phone numbers, location, company category, and description. These companies can be acquired by platform administrators using a simple keyword search as well as a defined location. Selenium is thereafter used to collect relevant companies based on the search query. Once these companies have been collected, they can then be automatically inserted as JSON files into a NoSQL database such as ElasticSearch to allow for search and retrieval.

B. Administrator Interface

Figure 3 demonstrates the home page for this searching, where an administrator can define keywords for the businesses they wish to acquire, as well as a location to search within. The administrator can also search already found companies in search of potential leads for new platform users.

V. Regulation and Guidance Matching

One point which affects B2B transactions is the adherence to changing federal regulation. As regulations regarding products during disasters such as COVID-19 can be quite stringent, it is necessary for businesses, especially recently created ones to understand the legal landscape, which includes the

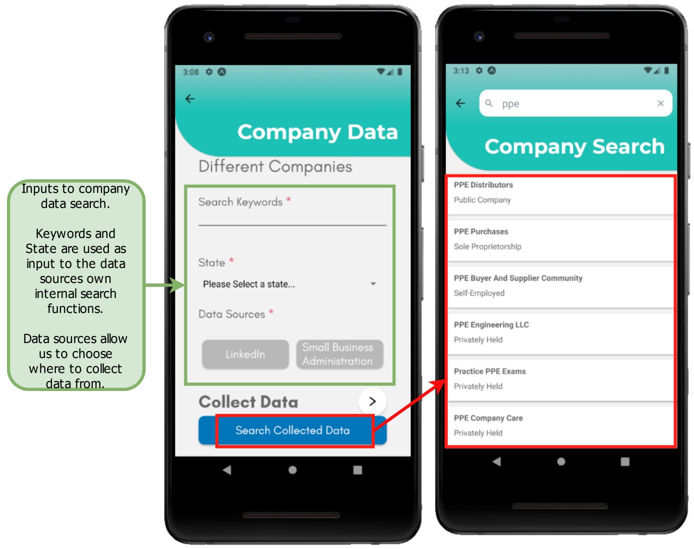


Fig. 3: Usage diagram for company data acquisition.

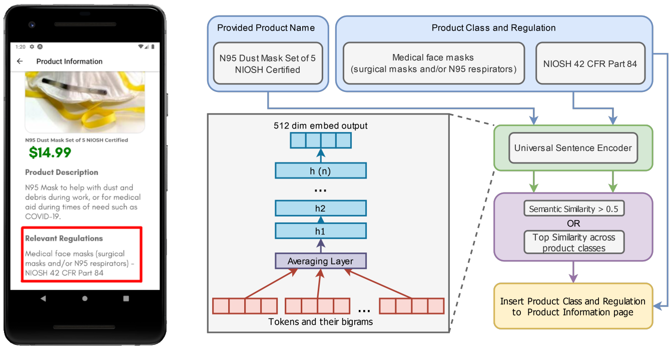
regulations to follow as they procure products for workplace use. This can lead to reduced compliance costs, which have been said to disproportionately affect smaller firms in the United States [12].

A. Regulation and Guideline Dataset

In the United States, most consumer products tend to be regulated by the Consumer Product Safety Commission (CPSC), which creates safety guidelines for many classes of such products. To power the regulation and guidance matching system a dataset was created using data acquired from the CPSC website, as well as hand-generated augmentations using the CPSC web page relating to product regulations for common PPE items during COVID-19. This dataset consists of 266 product class names, as well as the Electronic Code of Federal Regulations (eCFR) references and related guideline links for each class of products. With this in mind, matching regulations to product listings could be considered a semantic similarity problem, where an embedding vector v_0 is that of the name of the product, and vector v_1 being that of the product class name in the CPSC dataset.

B. Universal Sentence Encoder

Since typical product names and the product classes in the CPSC dataset contain multiple words, a greater-than-word length encoder is necessary to extract meaningful embeddings for semantic similarity. The Universal Sentence Encoder (USE) is a good fit for this task, as it specializes in creating embeddings for phrases, sentences, and paragraphs, without the need for pre-processing to remove stop-words and other characters that may interfere with similarity results. The USE variant used for this application is that which is formulated as a Deep Averaging Network (DAN), taking advantage of its improved resource usage characteristics for a deployed application when compared to the more accurate, yet slower



(a) *Product page with regulations.*

(b) *Block diagram of Regulation and Guideline matching system.*

Fig. 4: Architecture and output of regulation matching.

transformer model. The model used was acquired from Tensorflow Hub, which is pre-trained using a variety of sources such as Wikipedia, news, question and answer pages, discussion forums, and augmented with data such as the SNLI corpus [13].

C. Matching Process

Once 512-dimensional embedding vectors v_0 and v_1 have been generated using USE, the semantic similarity S can be determined to find out whether the regulations are related to that product or not.

$$S = \frac{v_0 \cdot v_1}{||v_0|| \times ||v_1||}$$

A threshold of 0.5 is used for the semantic similarity result, wherein product classes where $S > 0.5$ will then be displayed on the product information page, as demonstrated in Figure 4. It is possible that multiple regulations can be matched to a single product. If no product class has semantic similarity S greater than 0.5 to the product name, a fallback is implemented where the largest S among all product classes is displayed instead. This can aid imbalanced text similarity, wherein a long product name may not have a score higher than 0.5 across any given product class. A disclaimer can be included to ensure that businesses still conduct their own research, but the provided regulations and guidelines can help inform them on the first steps to avoid compliance violations.

VI. Conclusion and Future Work

In this paper, a novel platform with the use case of aiding local businesses during the COVID-19 pandemic was proposed. The platform employs a Buyer's club system to allow for collective buying power during times of need. This application is equipped with data acquisition and regulation matching modules to enhance the user experience while complying with a very dynamic regulatory environment, as seen during emergency situations. Data acquisition is used to automatically source new businesses to invite to the platform, as well as creating datasets on different product regulations. An NLP method for matching regulations to products based on product names is also proposed. This is designed to help inform small businesses about the regulations their products must adhere to for compliance with current laws. Future work in this domain could include the use of trend analysis to acquire company data in the background without an administrator in the loop. Automating the collection of regulatory and guidance documents from more agencies such as the FDA, CDC, NIOSH, OSHA could also aid in improving the regulation matching system. Such data could be preprocessed using NLP methods to ensure consistency and help improve matching across a wider array of product classes. The USE model could also be improved by training on product names and descriptions across e-commerce sites to help improve embedding results.

VII. Acknowledgments

This research is partially supported by NSF OIA-1937019, NSF OIA-2029557, and NSF CNS-1952089.

References

[1] R. Fairlie, "The impact of covid-19 on small business owners: Evidence from the first three months after widespread social-distancing restrictions," *Journal of economics & management strategy*, vol. 29, no. 4, pp. 727–740, 2020.

[2] A. W. Bartik, M. Bertrand, Z. B. Cullen, E. L. Glaeser, M. Luca, and C. T. Stanton, "How are small businesses adjusting to covid-19? early evidence from a survey," National Bureau of Economic Research, Tech. Rep., 2020.

[3] N. Vordos, D. A. Gkika, G. Maliaris, K. E. Tilkeridis, A. Antoniou, D. V. Bandekas, and A. Ch. Mitropoulos, "How 3d printing and social media tackles the ppe shortage during covid – 19 pandemic," *Safety Science*, vol. 130, p. 104870, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753520302678

[4] P. Cankurtaran and M. B. Beverland, "Using design thinking to respond to crises: B2b lessons from the 2020 covid-19 pandemic," *Industrial Marketing Management*, vol. 88, pp. 255–260, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019850120304144

[5] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained models for natural language processing: A survey," *Science China Technological Sciences*, pp. 1–26, 2020.

[6] J. Song, J. Kim, and J.-K. Lee, "Nlp and deep learning-based analysis of building regulations to support automated rule checking system," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, vol. 35. IAARC Publications, 2018, pp. 1–7.

[7] J. Zhang and N. M. El-Gohary, "Semantic nlp-based information extraction from construction regulatory documents for automated compliance checking," *Journal of Computing in Civil Engineering*, vol. 30, no. 2, p. 04015014, 2016.

[8] A. Ahlemeyer-Stubbe and S. Coleman, *A practical guide to data mining for business and industry*. John Wiley & Sons, 2014.

[9] S. Pouyanfar, Y. Yang, S.-C. Chen, M.-L. Shyu, and S. S. Iyengar, "Multimedia big data analytics: A survey," *ACM Computing Surveys*, vol. 51, no. 1, pp. 10:1–10:34, 2018.

[10] D. Brickley, M. Burgess, and N. Noy, "Google dataset search: Building a search engine for datasets in an open web ecosystem," in *The World Wide Web Conference*, 2019, pp. 1365–1375.

[11] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265–283.

[12] F. Chittenden, S. Kauser, and P. Poutziouris, *Regulatory burdens of small business: a literature review*. SBS Research Directorate, 2002.

[13] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Céspedes, S. Yuan, C. Tar *et al.*, "Universal sentence encoder," *arXiv preprint arXiv:1803.11175*, 2018.