Towards Interoperability of Social Media: Venue Matching by Categories

Thiago H Silva[1,2∞], **Mark Fox**[2∞]

[1]Universidade Tecnológica Federal do Paraná
Departamento Acadêmico de Informática
Curitiba, Brasil

[2]University of Toronto
Dept. of Industrial Engineering and Computer Science
School of Cities
Toronto, Canada

∞ Contributed equally to this work

thiagoh@utfpr.edu.br, msf@eil.utoronto.ca

Abstract. *Social media has become a significant source of data, important for different research, including those interested in using venue categories to understand different city dynamics and urban social behavior. Those research often uses data from a single platform, however, integrating data from multiple platforms affords many benefits; for instance, we can have a richer vision of the venues in the city. Nevertheless, the integration of data from multiple social media platforms poses several issues. One issue is the identification and integration of activities, which is implicitly represented by categories of places. This is because of the lack of standard venue categories across social media. Thus, this paper focuses on the venue integration problem through categories of venues. We propose an ontology to support the integration of venues across different data sources in this context. In addition, we propose a novel approach for this problem. Experimental results show that our approach is a promising alternative for this issue.*

1. Introduction

Over the last decade, social media has become a major source of data, fueling research and insights in the fields of sociology, urban geography, economics, etc. Some of this research is based on events captured by social media. An Event is composed of: (i) information about an activity that is taking place; (ii) information about people and groups of people that participate in the activity, and (iii) information about the venue or place where the activity takes place [Silva et al. 2019].

Several urban analyses use venue categories to understand different city dynamics and urban social behavior [Silva et al. 2017, Tsutsumi et al. 2019]. For example, Silva et al. (2017) introduces a novel approach to identify cultural boundaries between urban societies, considering users' preferences for food preferences measured by the category of places visit in Foursquare. Tsutsumi et al. (2019) presents a model that captures significant virtual relationships among businesses that are generated by users in the virtual world, which by exploring the categories of business enables the identification of business that represents non-obvious relations that might deserve particular attention of business owners, for instance, for new partnerships. Although events are captured by multiple social media platforms, research often uses data from a single platform [Senefonte et al. 2020, Ferreira et al. 2020, Kobellarz et al. 2019]. Yet, integrating data from multiple platforms affords a number of benefits, namely: **Complementary** use cases, for example, information about a venue, such as fax numbers, email, etc., may not be available in just one system but spread across multiple systems; **Additive** use cases, where some venues may be missing in one system, so, by integrating two systems, we might have a more complete vision of the venues in the city; and **Confirmatory** use cases, where, for example, information about a venue, such as telephone number, may appear in multiple systems and be used to confirm correctness.

Integration of data from multiple social media platforms poses several issues. One issue is the identification and integration of venues, i.e., to accurately match the same venue across social media platforms, e.g., the same venue in Yelp and Foursquare, where basic information, such as venue names and addresses, could be insufficient (i.e., incomplete, erroneous or too general). A second issue is the integration of venues by activities, where it is not required to match only the same venue across systems. In Yelp, for instance, the explicit activity is the review, but the implied activity is what takes place at the venue, such as eating, social interaction, exercise, etc. These activities are often implied by the category of the venue; thus, venue categories are a proxy for activities. The problem is exacerbated by a lack of standard venue categories across social media. For example, Yelp and Foursquare do not use the same venue categories. Even where there are overlaps in category names, their interpretation may differ.

The focus of this paper is on the venue integration by activities (or category) problem. Given a category of venues in a one system, e.g., Foursquare, we want to find the corresponding match in another system, e.g., Yelp. As the number of categories is large on social media sites like those exemplified, manual identification is not practical. This type of matching enables, for example, richer information to improve studies such as [Silva et al. 2017, Tsutsumi et al. 2019]. It also helps identify the same venue across social media because it provides more evidence to assist this task.

The main contributions of our study can be summarized in:

- An ontology to support the integration of data across different data sources, based on their similarity;
- A new approach to perform matching in different social media sources by categories of places. This step is key to enable interoperability. Experimental results shows that our approach is a promising alternative for this problem.

The remainder of this paper is organized as follows. Section 2 presents the use cases and competency questions regarding the integration of venues in multiple social media. Section 3 presents the ontology to support the integration of data from multiple social media. Section 4 discusses the problem of category matching for venue integration, also presenting a new strategy and its evaluation in this context. Finally, Section 5 concludes the study and presents future work.

2. Use Case and Competency Questions

The use case focuses on the task of integrating venues found in multiple social media datasets regardless of whether it is performed manually, semi-automatically or automatically. Each dataset contains a unique identifier for a venue, along with a set of attributes associated with the venue. These attributes include venue name, address, phone number, email address, hours of operation, venue categories and ratings. We assume a venue's attributes' values are correct unless evidence is found to cast doubt on their validity. Two types of commonly found evidence (Rahm & Do, 2000) are:

- Single-Source problems, such as misspellings (e.g., Joe's Dinnner instead of Joe's Diner), incomplete information (e.g., an address without a street number), or contradictions (e.g., postal code not matching the city); or
- Multi-Source Problems, where data from multiple sources may confirm or contradict each other. For example, differences in address, phone, and venue categories for the same venue in Yelp and Foursquare datasets.

The agent performing venue integration first performs single-source problem analysis. Each attribute for a venue is analysed and assigned a degree of validity and the methods used to determine it. An overall assessment of the validity of a venue's information may be performed, based upon the validity of its attributes.

In the next stage of venue integration, a pair of venues to be integrated, each from a different social medium, are analysed from a multi-source problem perspective. For each attribute, the similarity of values for the two venues is generated. In the case of the venue category attribute, the degree of similarity between the categories assigned to each venue can be based on the method defined in Section 4.2. To perform the integration task, the agent needs to know what venues are being integrated, the types of multi-source problems exist, the methods used to determine them, and the degree of certainty that the two venues are the same.

Based on the use case, the ontology pattern needs to support the agent by answering the following competency questions. For a single venue:

1. What is the source of an attribute's value, e.g., the social medium?
2. When was the value sourced?
3. What is the validity of the value, i.e., the degree of certainty that the value is correct?
4. How was the validity of the value determined?

For a pair of venues to be integrated:

1. What are the two venues being integrated?
2. What is the degree of similarity of the two venues?
3. How was the degree of similarity determined?
4. What is the degree of similarity of the values of the same attribute in the two venues?
5. What method was used to determine similarity of an attribute across the two venues?

3. Similarity Ontology

The purpose of the Similarity Ontology is to support the integration of data from multiple social media; venue matching, and venue category matching are two of many matching tasks – see Section 4 for one exemplification. In particular, the Similarity Ontology supports the measuring of similarity of two properties, and of two classes based on the similarity of its properties. It does not specify a particular method of determining similarity, for which many methods exist [Gan et al., 2013], but allows for the identification of the methods used.

The method for ontology design and evaluation is a simplified version of the Ontology Engineering process defined in Grüninger & Fox (1995). The process begins by defining a set of usage scenarios. Based on the scenarios, we identify a set of competency questions that the ontology must answer. These are the requirements for what is to be represented and the deductions to be performed. Next, the terminology (i.e., concepts and properties) required to answer the competency questions are selected and refined. The semantics of the terminology are defined by constructing a set of axioms that define and/or constrain their interpretation. The ontology pattern is defined using Description Logic (DL) and published using the Web Ontology Language (OWL). Finally, the ontology is evaluated based on the extent to which it is able to answer the competency questions.

The competency questions defined in Section 2 identify three types of meta-information that need to be captured by the venue pattern in order to support matching:

1. The validity of a property's value. In other words, how certain the value for a property is correct.
2. The provenance of a property's value. This includes both the dataset(s) from which it was sourced, and any transformations performed.
3. The degree of similarity between the values of two properties, or classes.

This information is meta to the values of a property or to an individual of a class. Consequently, we first introduce the Proposition class which addresses the first two items.

3.1 Proposition Class

Section 2 identifies the following single-source problem competency questions:

1. What is the source of an attribute's value, e.g., the social medium?
2. When was the value sourced?

3. What is the validity of the value, i.e., the degree of certainty that the value is correct?
4. How was the validity of the value determined?

The focus of these questions is the value of a property. Consider a version of the class Venue defined to have a data property hasCategory whose range is xsd:string and an object property hasSource whose range is SocialMedium:

$$\text{Venue} \sqsubseteq \leq 1\ \text{hasCategory.xsd:string} \sqcap \text{hasSource.SocialMedium}$$

Description Logic does not allow for the attachment of meta-information (annotations) directly to a data property value. Therefore, we have to use intermediation to capture the meta-information. The data property hasCategory becomes an object property and its value is restricted to an individual of a class, in this case the Category class. We can then attach the meta information as properties of the Category class.

$$\text{Venue} \sqsubseteq \leq 1\ \text{hasCategory.Category} \sqcap \text{hasSource.SocialMedium}$$

Our approach to representing the meta-information required to address the single-source problem competency questions is based on Fox & Huang (2005), where a Proposition "is a class that provides the basic properties for representing both the validity of an individual, and what its validity is dependent upon." We define the Proposition class as having a measure of validity (hasValidity) that ranges between zero and one, representing the degree of certainty that the proposition is true. The method used to derive the validity value, e.g., average, min-max, is specified by the hasValidityMethod property. The method used to generate the proposition is defined by re-using the PROV-O ontology [Lebo et al., 2013] property prov:wasGeneratedBy with a range of prov:Activity. An Activity can be a simple copy from a social media dataset, or in the case of an address, retrieved from an authoritative source. Finally, the time when the proposition was generated is specified using prov:generatedAtTime.

$$\begin{aligned}
&\text{Proposition} \sqsubseteq \text{prov:Entity} \\
&\quad \sqcap =1\ \text{hasValidity.xsd:float}[\geq 0][\leq 1] \\
&\quad \sqcap =1\ \text{hasValidityMethod.Method} \\
&\quad \sqcap =1\ \text{prov:wasGeneratedBy.prov:Activity} \\
&\quad \sqcap =1\ \text{prov:generatedAtTime.xsd:dateTime}
\end{aligned}$$

With the definition of Proposition, we can now define Category as a subclass of Proposition. The inheritance of Proposition properties provides for the representation of both validity and provenance.

$$\text{Category} \sqsubseteq \text{Proposition} \sqcap \leq 1\ \text{hasValue.xsd:string}$$

3.2 Similarity Class

The degree of similarity between two properties, e.g., hasCategory, or two classes, e.g., Venue is addressed by defining a general Similarity class. The Similarity class addresses the following multi-source problem competency questions:

1. What are the two venues being integrated?
2. What is the degree of similarity of the two venues?
3. How was the degree of similarity determined?
4. What is the degree of similarity of the values of the same attribute in the two venues?
5. What method was used to determine similarity of an attribute across the two venues?

The Similarity class is used for representing the similarity between two classes, and the similarity between two properties. It has properties for similarity of the two entities being compared, the method used to compute similarity, e.g., a distance metric, the activity used to generate the similarity metric and time the similarity was generated.

$$
\begin{aligned}
\text{Similarity} \sqsubseteq\ & {=}1\ \text{hasSimilarity.xsd:float}[\geq 0][\leq 1] \\
& \sqcap {=}1\ \text{hasSimilarityMethod.Method} \\
& \sqcap {=}1\ \text{prov:wasGeneratedBy.prov:Activity} \\
& \sqcap {=}1\ \text{prov:generatedAtTime.xsd:dateTime}
\end{aligned}
$$

The VenueSimilarity class represents the similarity of two venues. It is a subclass of Similarity with the additional properties that identify the two venues and individual similarity between each property of the venues.

$$
\begin{aligned}
\text{VenueSimilarity} \sqsubseteq\ & \text{Similarity} \\
& \sqcap {=}2\ \text{hasVenue.Venue} \\
& \sqcap \forall\ \text{hasPropertySimilarity.PropertySimilarity}
\end{aligned}
$$

The PropertySimilarity class specifies the similarity between two properties. In addition to inheriting the Similarity class properties, it identifies the two property values, which are individuals of type Proposition, whose similarity is being measured.

$$
\text{PropertySimilarity} \sqsubseteq \text{Similarity} \sqcap {=}2\ \text{hasProperty.Proposition}
$$

As an example, we can represent the similarity between the categories for two different venues using the sFOX method (defined in section 4) as follows:

Venue(v1) hasSource(v1,yelp) Category(c1) hasCategory(v1,c1)
Venue(v2) hasSource(v1,4squ) Category(c2) hasCategory(v2,c2)

VenueSimilarity(vs1) hasVenue(vs1,v1)hasVenue(vs1, v2)
PropertySimilarity(ps1) hasProperty(ps1,c1) hasProperty(ps1, c2)
hasPropertySimilarity(vs1, ps1) hasSimilarity(ps1, 0.8)
hasSimilarityMethod(sFOX)

With this ontology, we can compute the overall similarity of two classes, e.g., venues, as a function of the similarities of the class's properties, e.g., hasCategory. Extensions can be made to include such things as weights for properties if their importance in determine overall class similarity differs.

4. Matching Categories

In the previous section we introduced an ontology for representing the results of the venue matching process. In this section we describe a new approach to venue matching in different social media sources by categories of places. The method described herein is an example of a value (I.e., sFox) for the hasSimilarityMethod property inherited from the Similarity class.

4.1 Related Work

In the field of category similarity problem, a class of studies explore structure-based metrics [Zheng et al. 2010, Zhu et al. 2017, Deng et al. 2019]. For instance, Deng et al. (2019) calculated the category similarity in two steps. First, they used a manual step to match all first-level category labels in two distinct categories hierarchy — for example, Food (in system 1) and Catering Service (in system 2). Next, the authors traced the category tree for the second and third-level categories and assigned its parent category. They explore this trace to compute distances/similarities of categories in two systems. While this approach could be interesting to match categories of a few venues, such as deciding if two venues are the same, it is not practical when making a broad matching as proposed in this paper. Our approach proposed in this paper does not demand manual matching and works with category hierarchies with varied sizes, which could have any number of levels.

Another class of study to the problem of category similarity explores content-based metrics, which rely on semantic information of categories of entities [Chen et al. 2018, Čerba et al. 2016, Ballatore et al. 2015]. Chen et al. (2018) presented an approach that combines a generic lexical database with a professional controlled vocabulary to compute the relatedness of any two terms in the thesaurus. For instance, "river" and "stream" are semantically similar, while "river" and "boat" are dissimilar but semantically related, so relatedness refers to this latter case. While this approach points to a direction that potentially helps the category similarity problem, it faces some practical challenges — being one of the most critical the demand for controlled vocabulary for different contexts of interest, which are typically expensive to obtain, thus, hampering generalization.

4.2 Our Approach

Integration by categories. We propose a solution called sFOX that matches categories in two different systems based on words definitions and Sentence-BERT (sBERT), which is a modification of the BERT network using siamese and triplet networks that can derive semantically meaningful sentence embeddings, i.e., semantically similar sentences close in vector space [Reimers & Gurevych, 2019].

More specifically, for each category name *c* we get its definition *def* from WordNet [Miller 1995]; we use Python library PyDictionary for that. Next, we compute sentence embeddings for *c+def*, using sBert based on two pretrained sentence-transformer models (stsb-roberta-large and stsb-roberta-base) provided by the creators of sBert [Reimers & Gurevych, 2019]. This step provides two sentence embeddings for all categories of both systems; one is provided by exploring stsb-roberta-large and the other by stsb-roberta-base. In possession of sentence embeddings, for each category in System-1, we find the ten most similar embeddings on System-2 -- five provided by stsb-roberta-large and five by stsb-roberta-base. After that, we sort the ten candidates by the cosine similarity value.

4.3 Evaluated Approaches

We compared our proposal with three different ones. The problem tackled by all of them is the same: for a given category on system1, find the ten most similar categories on system2 in descending order. The first approach, called Levenshtein, considers the Levenshtein similarity, a string metric for measuring the similarity between two sequences [Levenshtein 1966], to compute this measure for a given category *c* in System-1 to all categories in System-2 to select the ten most similar ones to c.

The second approach, called Levenshtein+Structure, first executes the Levenshtein approach. In possession of the ten most similar categories to *c* given by the previous step, it computes the category similarity of *c* to this set using a structure-based approach, as Deng et al. (2019) performed, and add this similarity value to the Levenshtein similarity. Following the algorithm for the structure-based approach, we must establish a connection between first-level categories on the two systems studied, Foursquare and Yelp, in this study. For instance, the 'Nightlife' category in Yelp was manually matched to 'Nightlife Spot' on Foursquare. Categories for other levels in the hierarchy do not need to be manually matched. The structure-based similarity category S_{struct} is calculated according to:

$$S_{struct} = e^{(-D/2\alpha)},$$

where $D = p_1 + p_2$, with $p_1 + p_2$ representing the distance from their shared root parent node to the node representing a certain category for system1 and system2, respectively, α is the maximum distance, which could be derived either from system1 or system2, in this study α=4. If there is not shared root node, the similarity is 0.

The third approach computes, for a particular category, word embeddings using the sBert, computing next the cosine similarity using the embeddings. For a given category in system1, it finds the ten most similar ones on system2 (five using the pre-trained model stsb-roberta-large and five stsb-roberta-base).

4.4 Experimental Results

For each approach, our evaluation considers three different experiments. The first one regards the match with the best candidate suggested by the approach under assessment, i.e., with the highest similarity score. This experiment refers to the case of an automatic approach, where the matching follows what is recommended. The following two experiments consider the case where there is human intervention in the matching process. In experiment 2, the approach selects the three best candidate categories so that the user can choose the most appropriate one. Experiment 3 is similar to 2 but offers ten possibilities to users. Experiment 3 is more expensive because it provides more cases to be analyzed; however, it could be worth the cost depending on the situation.

To evaluate the approaches described in Section 3, we prepared a test set containing 300 Foursquare random categories. We manually tried to match these categories on Yelp category system. Categories with no match found were disregarded, 25 categories in total. Thus, 275 words were considered in the evaluation; this is our ground truth.

According to Table 1, the method sFOX is better in all evaluations. It is worth noting that it is considerably superior in the match with the first candidate, the most challenging one because it considers automatic evaluation. This means that sFOX is a promising approach to perform an automatic evaluation without human intervention in the selection process.

Table 1. Results for the evaluation of all considered approaches for category matching in three different experiments.

	Levenshtein	**Levenshtein + Structure**	**sBert**	**sFOX**
Experiment 1: **Match with the first candidate**	0.71	0.72	0.71	**0.82**
Experiment 2: **Match with any of the top 3**	0.79	0.77	0.83	**0.88**
Experiment 3: **Match with any of the top 10**	0.82	0.79	0.88	**0.89**

The structure-based approach alone does not produce satisfactory results in the investigated problem, below 0.1 in all three experiments (omitted in the analysis). This is because the range of possibilities to match is big, and several possibilities end up having the same similarity but are not related. The problem is partly related to the fact that in some cases, more than one first-level category in Yelp had to be matched in one in Foursquare. For instance, 'Shop & Service' in Foursquare were matched with 'Shopping' and 'Local Services' on Yelp because they refer to similar types of places. For this reason, instead of considering the results of this approach by itself, we applied it after selecting a set of candidates. This is a common practice; for example, Deng et al. (2019) also used a similar strategy. We see that this strategy helps the Levenshtein approach on experiment 1, being better then sBert; however, it does not help on experiments 2 and 3. With more candidates (as in experiments 2 and 3), the chances of adding noise due to the problems of structure-based similarity increase leading to errors, especially in a gray area of the decision space.

5. Conclusion

The task of integration of data from multiple social media platforms poses several issues. To contribute to this issue, first, we proposed the Similarity Ontology to support the integration of data from multiple social media. This ontology paves the way to support the interoperability of social media. This is a bigger problem than addressed in this paper; however, our contributions are essential steps towards this goal.

Second, we propose a venue matching technique across different social media, which is helpful to feed our proposed ontology. In the context of social media that offers data that can be associated with venues, such as Yelp and Foursquare, one fundamental issue is venue matching across different social media. This problem has different facets; for instance, (i) one can be interested in identifying the same venue in different systems, or (ii) identify venues of the same type, i.e., has the same category, across social media. This paper focused on the latter case, thus, contributing to venue integration by categories problem. Given a category of venues in a particular system, e.g., Yelp, our proposed solution, sFOX, finds the corresponding match in another system, e.g., Foursquare. Automatic approaches like ours are essential as the number of categories is high on social media sites like those illustrated. Our solution helps to provide richer information that can improve existing and new studies in different contexts and help identify the same venue across different social media because it allows for more robust evidence for this problem.

As future work, it is important to address some issues that can emerge when integrating venues. For instance, some chains in dense areas can be spatially located close to each other. In this case, they will have several attributes in common, including the categories, making the venue matching problem harder. This and other issues should be better investigated to improve integrating the same venue in different social media. In addition, we envision new approaches to provide a more comprehensive data interoperability across social media. For instance, we envision providing new strategies and ontologies to integrate users (and groups of users) across social media.

6. Acknowledgements

This research was supported, in part, by the Natural Science and Engineering Research Council of Canada and the University of Toronto School of Cities Urban Challenge Grant. This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, by the project GoodWeb (Grant 2018/23011-1 from São Paulo Research Foundation - FAPESP), and by the Brazilian National Council for Scientific and Technological Development - CNPq (grant 310998/2020-4).

7. References

Ballatore, A., Bertolotto, M., & Wilson, D. C. (2015). A structural-lexical measure of semantic similarity for geo-knowledge graphs. ISPRS International Journal of Geo-Information, 4(2), 471-492.

Čerba, O., & Jedlička, K. (2016). Linked Forests: Semantic similarity of geographical concepts "forest". Open geosciences, 8(1), 556-566.

Chen, Z., Song, J., & Yang, Y. (2018). An approach to measuring semantic relatedness of geographic terminologies using a thesaurus and lexical database sources. ISPRS International Journal of Geo-Information, 7(3), 98.

Deng, Y., Luo, A., Liu, J., & Wang, Y. (2019). Point of Interest Matching between Different Geospatial Datasets. ISPRS International Journal of Geo-Information, 8(10), 435.

Ferreira, A. P., Silva, T. H., & Loureiro, A. A. (2020). Uncovering spatiotemporal and semantic aspects of tourists mobility using social sensing. Computer Communications, 160, 240-252.

Gan, M., Dou, X., & Jiang, R. (2013). From ontology to semantic similarity: calculation of ontology-based semantic similarity. The Scientific World Journal, 2013.

Grüninger, M., and Fox, M.S. (1995), "Methodology for the Design and Evaluation of Ontologies", Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal.

Holton, M. and Alexander, S. (1995) "Soft Cellular Modeling: A Technique for the Simulation of Non-rigid Materials", Computer Graphics: Developments in Virtual Environments, R. A. Earnshaw and J. A. Vince, England, Academic Press Ltd., p. 449-460.

Kobellarz, J. K., Graeml, A. R., Reddy, M., & Silva, T. H. (2019, October). Parrot talk: retweeting among Twitter users during the 2018 Brazilian presidential election. In Proceedings of the 25th Brazilian Symposium on Multimedia and the Web (pp. 221-228). Rio de Janeiro, Brazil.

Lebo, T., Sahoo, S., and McGuinness, D. (2013), "PROV-O: The PROV Ontology", W3C Recommendation 30 April 2013. https://www.w3.org/TR/prov-o/

Levenshtein, V. I. (1966, February). Binary codes capable of correcting deletions, insertions, and reversals. In Soviet physics doklady (Vol. 10, No. 8, pp. 707-710).

Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM, 38(11), 39-41.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. IEEE Data Eng. Bull., 23(4), 3-13.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

Senefonte, H., Frizzo, G., Delgado, M., Lüders, R., Silver, D., & Silva, T. (2020, October). Regional Influences on Tourists Mobility Through the Lens of Social Sensing. In International Conference on Social Informatics (pp. 312-319). Pisa, Italy.

Silva, T. H., de Melo, P. O. V., Almeida, J. M., Musolesi, M., & Loureiro, A. A. (2017). A large-scale study of cultural differences using urban data about eating and drinking preferences. Information Systems, 72, 95-116.

Silva, T. H., Viana, A. C., Benevenuto, F., Villas, L., Salles, J., Loureiro, A., & Quercia, D. (2019). Urban computing leveraging location-based social network data: a survey. ACM Computing Surveys (CSUR), 52(1), 1-39.

Tsutsumi, D. P., Fenerich, A. T., & Silva, T. H. (2019). Towards business partnership recommendation using user opinion on Facebook. Journal of Internet Services and Applications, 10(1), 1-23.

Zheng, Y., Fen, X., Xie, X., Peng, S., & Fu, J. (2010, November). Detecting nearly duplicated records in location datasets. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 137-143). San Jose, CA, USA.

Zhu, Y., Zhu, A. X., Song, J., Yang, J., Feng, M., Sun, K., ... & Zhao, H. (2017). Multidimensional and quantitative interlinking approach for Linked Geospatial Data. International Journal of Digital Earth, 10(9), 923-943.