Ontology-Based Approach for Neighborhood and Real Estate Recommendations*

Wissame Laddada
wissame.laddada@liris.cnrs.fr
Université de Lyon, LIRIS UMR 5205
Lyon, France

Fabien Duchateau
fabien.duchateau@univ-lyon1.fr
Université de Lyon, LIRIS UMR 5205
Lyon, France

Franck Favetta
franck.favetta@liris.cnrs.fr
Université de Lyon, LIRIS UMR 5205
Lyon, France

Ludovic Moncla
ludovic.moncla@insa-lyon.fr
INSA Lyon, LIRIS UMR 5205
Lyon, France

ABSTRACT

Suggesting services or products to people is a task that should be handled by recommendation systems due to the important increase of information and the multitude of user criteria. In fact, when expressing wishes for a product, a user is influenced by his/her tastes or priorities. These influential characteristics tend to be challenging regarding their integration into recommendation systems, because interaction between the products/services and the user has to be captured through its preferences. Recommendation systems for neighborhood and real estate search are no exception, and to achieve reliable recommendation, we developed an ontology NAREO (Neighborhood And Real Estate Ontology) where environment characteristics related to user preferences are modeled with other geo-semantic descriptions. This ontology can be enriched by SWRL (Semantic Web Rule Language) rules that enhance the semantics of our knowledge base and allow reasoning process through built-ins. To illustrate a use case, we provide a basic set of predefined rules for the recommendation context. User preferences are managed through SPARQL queries taking into account the result of inferences.

CCS CONCEPTS

• **Information systems** → **Location based services**; **Spatial-temporal systems**.

KEYWORDS

Ontology, SWRL reasoning, Recommendation Systems, Spatial modeling

*This work has been partially funded by LABEX IMU (ANR-10-LABX-0088) from Université de Lyon, in the context of the program "Investissements d'Avenir" (ANR-11-IDEX-0007) from the French Research Agency (ANR), during the HiL project†.

LocalRec'20, November 3, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-8160-4/20/11…$15.00
https://doi.org/10.1145/3423334.3431452

ACM Reference Format:
Wissame Laddada, Fabien Duchateau, Franck Favetta, and Ludovic Moncla. 2020. Ontology-Based Approach for Neighborhood and Real Estate Recommendations. In *4th ACM SIGSPATIAL Workshop on Location-Based Recommendations, Geosocial Networks, and Geoadvertising (LocalRec'20), November 3, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3423334.3431452

1 INTRODUCTION

With the tremendous quantity of available objects, users become familiar with recommendation systems. These systems are even essential on the web to discover new movies or related e-commerce products [5]. More recently, location-based social networks have been exploited, mainly focused on recommending places such as restaurants and attractive areas [12, 17]. Despite this trend, searching and finding an ideal accommodation, either for purchase or for rental, is still a very tiresome task [7, 20]. And this is worse when users do not have prior knowledge about their future city of residence (e.g., case of job transfers). There exist many generic or specialized websites that offer description of available real estates, including pictures and even virtual visits. And customers usually have a precise idea about their ideal accommodation (e.g., number of rooms, maximum price, type of neighborhood, essential services and transportation means): one may look for a cosy apartment in a vibrant neighborhood with many pubs while another may prefer a house with garden in a quiet residential area close to schools and parks. Thus, the question is how recommendation systems need to be adapted for neighborhood and real estate search.

When buying or renting a real estate, one takes into account multiple criteria about the accommodation and its neighborhood. Existing solutions, typically web applications, are not able to perform complex queries using a combination of preferences and possibly restrictions. Besides, only a few information are structured (e.g., type of accommodation, number of rooms, address), and the remaining ones (e.g., specific rooms, building level, neighborhood ambiance, surrounding amenities) are usually - when available - scattered into textual descriptions which cannot be easily exploited in an automated manner. A second issue deals with the vagueness of the neighborhood concept. As explained in Delmelle's study [8], the definition, borders and perception of a neighborhood tend to be subjective, and they suffer from drastic changes over time.

Several works were designed to tackle this recommendation challenge. A first one aims at detecting similar neighborhoods between cities [11], which is an acceptable solution in case of job transfer, but limited to a few large cities. Similarly, the VizLIRIS tool shows how to detect an ideal arrival area similar to a given starting neighborhood [4]. It works for a whole country (France), but it assumes that the user perception about his starting neighborhood is consistent with the quantitative numbers that describe it. HoodSquare recommends neighborhoods by exploiting Foursquare check-ins, advising for instance areas for education, arts, food or parks [19]. Spatial Pattern Matching is another proposition to take into account constraints in order to recommend places [13]. However, these constraints are limited to distance between places. Finally, South Korea researchers directly recommend, for a few cities in their country, the most relevant neighborhood and accommodation based on similar user profiles [18]. The modeling of an accommodation (using the SEED layout –design representation for spatial layouts–) is accurate but difficult to obtain automatically from a textual accommodation description. Authors also have defined a partial ontology about location, however it is not available and its components (i.e., classes, predicates, rules) are not described. In another study [6], an ontology gathering some relevant classes (apartment, house, price, home, etc) to describe attributes for a real estate is presented. It is also enriched by object and data properties (*hasPrice*, *hasArea*, etc.) that allow the establishment of semantic relationships between classes. This ontology is suggested to study the extraction of fuzzy information from real estate offer advertisements and not for recommendation. However, a fragment of its taxonomy can be embellished with our concepts for the recommendation process. An ontology-based multi-criteria spatial decision support system was proposed by Malczeweski [14] for the house selection. This ontology encompasses several and relevant entities and properties such as house, building characteristics, neighborhood quality, accessibility (to public transportation, commerce, education, etc.). SWRL rules are also adopted to i) assign the user preferences (weights) to all the attribute concepts (e.g., accessibility to education is an attribute) and ii) to determine a score for a house according to the weights values. However, expressing user preferences in terms of weights through rules is a heavy task and difficult to extend. Moreover, the spatial reasoning process cannot be managed by the ontology and the spatial information cannot be directly reached. In fact, the spatial information/functions are handled through the GIS engine (ArcGIS) and are not modeled through the ontology. Finally, their ontology is not publicly available, thus limiting its use by other researchers.

To the best of our knowledge, there is no complete and functional ontology about neighborhood and real estate. Yet, this is a crucial component for solving such recommendation due to the complexity of the preferences and constraints that can be expressed about the accommodation and its neighborhood. In this paper, we therefore propose an ontology named NAREO (Neighborhood And Real Estate Ontology), dedicated to neighborhood recommendation firstly and real estate search for future work. It has been designed with the support of researchers in social sciences, mainly to support the description of a neighborhood. It enables reasoning based on SWRL rules, that can be enriched if needed. It is freely available, and we demonstrate its use in a use case for Lyon, France by detailing both the data integration aspect (using Open Street Map data and other data resources related to neighborhoods and the transportation system) and the recommendation part (reasoning ans SPARQL queries).

In the rest of this paper, Section 2 describes the knowledge representation developed for NAREO ontology from the conceptualization to the modeling step. We determine also in this section the concepts (classes) taxonomy and how they are related to each other through object properties and data properties. Then, we present Section 3, in which the semantics of NAREO is improved thanks to SWRL rules. These rules are included to infer some relevant data and facts related to criterion. The ontology is enriched with individuals and relations assertions, drawing on different resources in order to present a use case in Section 4. It shows how SWRL rules enhance the semantic of NAREO and describes some queries using SPARQL language to express user preferences. The last section is about future work.

2 ONTOLOGY-BASED KNOWLEDGE REPRESENTATION

A shared ontology consolidates the philosophical idea that several partial ontologies can be modeled then shared to form a global ontology where a possible enrichment is carried out which is our aim. Generally, in order to build an ontology three steps are to be followed: conceptualization, formalization and modeling process. This section covers two steps: conceptualization and modelling aspect of NAREO. The formalization step is not considered in the current paper because we did not formalize the concepts for this first version of NAREO.

2.1 Conceptualization

People looking for a new estate have several preferences to choose the right accommodation. These preferences represent characteristics for the real estate, amenities around it, neighborhood, etc. Since the neighborhood is an important feature for most users, we assume that they will search a real estate *within* the neighborhood that satisfies their characteristics. These latter may gather amenities (school, hospital, transportation, etc.) or guarantee a proximity to some services that allow activities like leisure, shopping, etc. It could also be related to the neighborhood ambiance or type of landscape. Describing characteristics and conceptualizing knowledge about neighborhood recommendation lead us to identify concepts and relationships between them. Thus, in order to design our ontology we need to answer questions about the characteristics:

- What are all the general characteristics that influence the neighborhood recommendation?
- What can be the abstraction (concepts) to describe all the characteristics?
- Which relationships (roles) allow us to highlight the semantic knowledge about the preferences?

For this purpose, we analyzed preferences expressed by people that affect the abstraction according to the different individual needs or objectives. Parents for example, may prefer to be near schools or kindergartens. Others require a proximity to some general food store or supermarket. Also people without a personal vehicle reach

Table 1: Atomic concepts

Atomic concept	Interpretation
Spatial Entity	Encapsulates all features from the environment that contribute to express people preferences
Geometry	Each spatial entity is related to a geometry that ensures the possibility of spatial reasoning between features from the real world
Semantic Entity	Knowledge that completes the definition of other concepts
Neighborhood	Entity that describes an area where a real estate may be located
Public transportation	Semantics about stations that are features (sub concept of spatial entity). It is also completed by information such as the different lines of the transportation network (bus station, subway station, etc.)
Amenity	Spatial entity coupled with one or more activities (e.g., shopping, education, health care)
Person	Individuals who are about to make decision and select their future neighborhood
Real estate	Accommodations (apartment, house, etc.) available for rental or purchase, that may be a potential housing for a person
Workplace	Place where a person works, which is a relevant information for filtering a recommendation based on the distance criterion

their workplace using public transportation or by foot. Hence, the neighborhood must be close to public transportation (subway station, bus station, bike station, etc.) or to the workplace. Other users may provide more general ideas about their ideal neighborhood: type of landscape, quiet or vibrant, more or less close to a city center, etc. They may also express priorities between these preferences. For instance, if most people tend to favour a secure area, this criterion may be essential for some users. From this preferences study, terms (concepts) and relevant spatial aspect and roles (e.g., proximity) emerge. In addition, the majority of characteristics appears to be related to some features from a city (environment, infrastructure). Note that the full list of concepts has been discussed with social science researchers. Taking into account preferences to choose the right neighborhood leads somehow to establish the spatial relationships between these features, which are mainly composed of spatial entities. Formally, this emphasizes the need for a spatial entity concept that encapsulates all the required features and from where spatial reasoning may occur. In geographic information science, this reasoning around spatial relationships includes topological models that describe them such as RCC8 [16] or 9-intersections [9]. Consequently, a spatial entity requires other concepts to clearly describe spatial data like knowledge about geometries. Besides, to enhance the flexibility and the scalability of our ontology, semantic entity concept has to be considered. Such entity is useful to capture other complementary concepts that help to collect a specific piece of information about the spatial entity and reinforce our policy for searching data. For instance, a real estate will have a category such as apartment, house, loft, etc. These categories will be considered as a semantic entity, since it is not a spatial information. Plus, each category may have characteristics (e.g., garden, swimming pool, etc.) for a house which enhance the semantic when we get to filter data through SPARQL queries.

At first sight, the main atomic concepts to design the ontology are summarized in Table 1. This first stage of conceptualization highlights the domain and scope of the NAREO ontology by defining the general atomic concepts. In the next step, we design the ontology by establishing the taxonomy between entities (classes/concepts) and determining properties. These latter, may also be depicted through taxonomies. Our ontology for neighborhood recommendation is presented in the next section with more explanations.

2.2 Ontology modeling

Broadly speaking, an ontology is based on logical theories. More precisely, First-Order Logic (FOL) and Description Logic (DL) [2]. In this specific context, the classes that we must define for NAREO ontology are considered as concepts (DL definition) or unary predicate (FOL definition). Likewise, the properties that depict the relationships are roles (DL definition) or binary predicates (FOL definition). The semantic web technologies provide a set of vocabularies and languages to describe the component of an ontology with different level of semantic expressiveness such as RDF (Resource Description Framework), RDFS (Resource Description Framework Schema) or OWL (Web Ontology Language). Based on DL, OWL is the most complete language that allows a high level of expressiveness. In fact, with OWL an ontology can be enriched with axioms and complex definitions of concepts (classes). This complexity combines several characteristics like adding constraint cardinalities, universality and existentiality on properties within the definition of concepts. In our knowledge representation of neighborhood recommendation, the definitions of concepts and properties are not based on heavy formal axioms and the ontology is not really heavyweight. However, for future work and more relevant ontology formalization we adopt OWL for the description. Also, technical prowess motivated us to rather opt for an OWL description than another representation. The ontology is illustrated here and built with Protégé[1] by importing an existing description of the spatial dimension.

The conceptualization presented above, motivates us to consider the ontology classification. Indeed, using spatial dimension (general classes/properties) and features characterizing a city within a knowledge base is not a new task and some semantic descriptions do already exist [1, 15]. Consequently, we have an *upper-level* ontology and a *domain/application* ontology. An upper-level ontology is an ontology where general concepts and properties are defined

[1]https://protege.stanford.edu

with a high degree of abstraction and from which other ontologies (domain/application ontology) may be designed or extended. If an ontology targets a specific problem we call it a domain ontology (or an application ontology if the vocabulary is very specific to a particular application or system).

2.2.1 Upper-ontology Level. A spatial dimension is used in several applications or services (routing, tourism, wayfinding, etc.). Thus, an ontology describing this dimension is independent of domain of interest or application. That is why we consider it as an upper ontology. In our description, we used the OGC GeoSPARQL[2] standard that represents geospatial knowledge on the semantic web. As illustrated in Figure 1, this ontology mainly gathers three classes and the super class is *Spatial Object*. This class clusters the *Feature* and *Geometry* classes that provide the possibility to associate a geometry (*Point, LineString*, etc.) to features from the real world. The association between the two classes is handled by the property *has Geometry*. Each geometry is serialized by geographic coordinates through *asWKT* or *asGML* data properties. In order to processes spatial information and ensure spatial reasoning, the GeoSPARQL ontology also presents spatial relations (functions) from RCC8/9-intersections theories. This ontology is extended with a domain/application ontology which is specific to neighborhood recommendations.

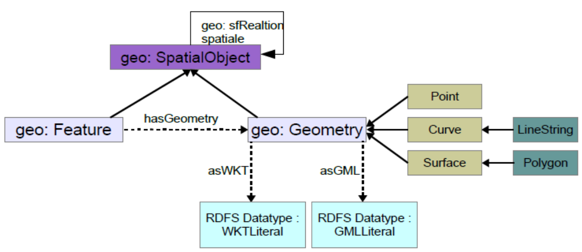

Figure 1: GeoSPARQL standard ontology (geo: prefix for the namespace: http://www.opengis.net/ont/geosparql)

2.2.2 Domain/application ontology. As explained, people preferences are mainly related to daily life activities and consequently to the infrastructure and characteristics of cities (e.g., location of amenities, services, shopping). In addition, social science researchers have defined six environment variables to summarize the characteristics of a neighborhood [3]:

- *Type of building* represents the most common buildings in the neighborhood (from large housing complexes to individual houses);
- *Usage* describes local activities (residential, services, etc.);
- *Landscape* defines the quantity of surrounding natural elements (e.g., fields, forest, urban);
- *Social class* denotes the degree of wealth using five levels of value;
- *Morphological position* indicates the distance level of a neighborhood from the city center (rural up to central);
- *Geographical position* stands for the direction towards the city center of the closest city (eight cardinal values plus central).

In order to capture the domain knowledge of those characteristics, we first combine the GeoSPARQL ontology with another taxonomy as depicted in the partial overview of our concepts organization (Figure 2). The taxonomy is inspired from the tags classification of OpenStreetMap (OSM). They are collected from wiki pages[3] where they are organized and maintained. Three principal keys and some of their values are used in our ontology; namely, *Shop*, *Amenity*, and *Leisure*. Each key is considered as a class where their values are their sub classes. The taxonomy of the sub classes is organized as suggested for the values classification in OSM. Since each component of this taxonomy is related to a location from the environment –that influences the recommendation– we add an *Environment* top class above the taxonomy.

Besides, as explained in the conceptualization, the **Neighborhood** and the **Real estate** concepts have to be integrated into the ontology. Therefore, we include a *Housing* class that congregates both of them. The Neighborhood is *equivalent to* a class *IRIS* (a French administrative acronym for "*Ilots Regroupés pour l'Information Statistique*") which refers to the census areas of a territory, or division of municipalities/districts "units". Besides, in order to handle the spatial aspect, the class Feature from the upper-level ontology (GeoSPARQL ontology) is extended by the concept *Spatial entity* which is introduced in our conceptualization. Broadly speaking, in order to process spatial data, the spatial entity gathers all the top classes with their taxonomy listed so far, in addition to the *Workplace* class.

The *Semantic Entity* concept is also introduced as a class into our ontology. It deals with some metadata to complete semantics from other classes as succinctly explained in the conceptualization. For example, the *Transport stations* class which is a sub class of *Amenity* – and a value for Amenity key from OSM – can be linked to a *Public Transportation* (subClass of the semantic entity). It provides the possibility to associate each station by the line (bus, subway, etc.) which serves it. Furthermore, a class for *Itinerary* description is defined as a sub class of the semantic entity, in order to represent alternative paths or combination of trajectories (with different means of transportation) to join for instance the workplace. When addressing a recommendation problem, we have to take into account the person preferences to whom we suggest relevant real estates. In this study, the preferences are considered into the SPARQL query. However, to reach a reliable recommendation for a real estate in future work, the ontology will evolve and the semantic user profile will be enriched. Consequently, a class named *Person* is included to ease the mapping with a user profile.

2.2.3 Properties definition. With the presented knowledge representation, we consider also properties to add the necessary metadata and perform the recommendation process. In addition to the relationships from the GeoSPARQL ontology (*has serialization, equals, intersects*, etc.), we developed other object and data properties to

[2]http://www.opengeospatial.org/standards/geosparql

[3]https://wiki.openstreetmap.org/wiki/Map_Features

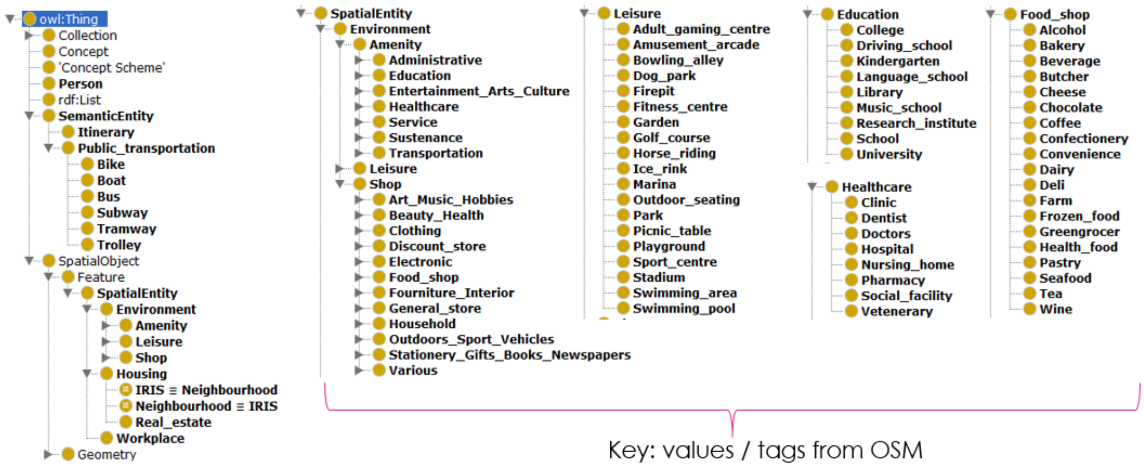

Figure 2: Partial overview of classes from NAREO ontology

Figure 3: Fragment of properties from NAREO ontology (blue: object properties, green: data properties)

carry out the relationships between entities from NAREO ontology. Figure 3 shows an overview of the properties included in the ontology.

When expressing preferences, the distance criteria often comes up and tends to be very important. Thus, the object property *hasDistance* is determined between the *Spatial entity* and *Itinerary* classes. An itinerary is also related to an IRIS (EquivalentTo:Neighborhood) from where the distance should be calculated. This leads us to include the property *itineraryFromIRIS*.

Formally, suppose we have the following assertions:

- *IRIS(s)* : *"s"* is an instance of IRIS in the knowledge base
- *Workplace(k)*,
- *Itinerary(i)*,
- *hasDistance(k,i)*,
- *itineraryFromIRIS(i,s)*

The assumption to make through those facts is that an itinerary *"i"* exists to reach the workplace *"k"* from the IRIS *"s"*. Furthermore, information about the distance are asserted by means of data properties. In fact, the itinerary can refer to a combination of trajectories with different public transportation. Consequently, each itinerary has a combination of different distances related to different transportation. Data properties such as $subway\_distance$, $bus\_distance$, $tram\_distance$ are subsumed by the $distance\_value$ data property. The distance value will be communicated as a time value (trajectory duration). To clearly understand these properties and keeping in view the facts below, we can assert additional facts about distance related to the itinerary *"i"*, for instance:

- $subway\_distance(i, "15")$,
- $tram\_distance(i, "10")$

These facts means that it would take 25 minutes (15+10) to reach the workplace *"k"* from the IRIS *"s"* while considering the itinerary*"i"*. The latter is a combination of two trajectories: by subway and tram. The transportation system is also related to stations. As explained through the domain ontology, each station has one or more lines (e.g., bus or subway) from the transportation network. The top-level object property that allows to formally consider this fact is *hasStation*, from where several properties are derived such as *hasBusStation*, *hasTramwayStation*, *hasSubwayStation*, etc.

To fulfill more semantics and metadata, we add other data properties such as *salary*, *age*, *diploma*, *gender*, etc. for a person description, and *atmosphere*, $code\_IRIS$, nom_COM, nom_IRIS, etc. for the neighborhood description. The atmosphere data property provides the ability to formalize general information about the district (mostly, the environment defined by social researchers, as explained in Section 2.2.2). Some of the properties such as *nearby* or $satisfy\_criterion$ are used for the reasoning process through SWRL rules. More explanation about this reasoning mechanism is presented in the next section. Finally, NAREO[4] merges the two main ontologies –spatial and domain / application– encompassing a total of 306 classes, 59 object properties and 35 data properties,

[4]The NAREO ontology is publicly available: https://doi.org/10.5281/zenodo.3904419

as shown in Table 2. In addition to existing generic concepts, we defined 228 new classes, 16 object properties and 42 new relations (object and data properties).

Table 2: Ontology metrics for NAREO

	Total	Domain/application
#Class	306	228
#Object Property	59	16
#Data Property	35	26

3 SWRL RULES DEFINITION FOR NAREO

In addition to the model, our NAREO ontology is provided with a set of rules that enable reasoning and enhance the semantics of our knowledge base. New rules can also be added if needed.

In this context, this section lays out some rules that infer relevant information related to preferences such as distance criterion. In fact, in an ontology knowledge base described with OWL, we can not infer new values for individuals while using axiomatic reasoning (complex concepts definition). Inferences are only about new facts between existing classes and individuals from the ontology or to assert their types considering concepts formalization. Sometimes, when we attempt to infer a specific fact, we need first to restrict and calculate values about individuals and since OWL description does not allow this mechanism, we exploit SWRL rules to infer for instance values related to distances.

SWRL rules are defined as a conjunction of predicates forming a head (the results of inference) and a body (the conditions to have an inference). In our case, we design *DL-safe rules* which means that every argument (individual) in the head has to be in the body in order to maintain both consistency and decidability. We use SWRL with an extension of the *Pellet* reasoner and a set of APIs (e.g., OWL) to customize and define our own n-ary predicates named built-ins. They can be considered as methods where the parameters are the n-arguments associated to the predicates. The reasoning engine (in our case Pellet) infers new values related to data properties through these built-ins. The other predicates forming a rule are in fact classes or object/data properties. We define here four major DL-safe rules including built-ins to calculate distances and infer values for a data property.

3.1 Inference for proximity criteria

The first rule, shown in SWRL Rule 1, is about proximity between a neighborhood and a food shop. In this rule, the body is built from an *IRIS(?i)* predicate with a geometry ?g1 defined by the predicate *hasGeometry(?i,?g1)*. This geometry is described by a WKT serialization (geographic coordinates) through the data property *asWKT(?g1,?w1)*. The *Food_shop* predicate is also used in this rule but it may be replaced by other sub classes (kindergarten, hospital, Swimming_pool, etc.). This feature is related to geographic coordinates described by *?w2* in the rule. The inference asserts that an IRIS (?i) is not far from (nearby) a feature Food_shop(?f), if all the conditions in the body are fulfilled. Technically, the predicate *distance_criterion_food_shop* which is a built-in, represents a method with three parameters, including both location coordinates and parameter ?d which is a boolean value. This value is fulfilled (with the value *true*) if the calculated distance between the two entities (?w1,?w2) does not exceed for example 300 m. This new fact can be used as a filter in a SPARQL query for neighborhood recommendation.

SWRL Rule 1

```
IRIS(?i) ^ hasGeometry(?i,?g1) ^ asWKT(?g1,?w1) ^
Food_shop(?f) ^ hasGeometry(?f,?g2) ^ asWKT(?g2,?w2) ^
distance_criterion_food_shop(?d,?w1,?w2) ^
=> nearby_food_shop(?i, ?f)
```

3.2 Inference for neighborhood atmosphere criteria

The second rule depicted by SWRL Rule 2 deals with the atmosphere and environment of a neighborhood. As in the previous definition, this rule is about distance criteria. The objective is to use the inference result for the next rule that determines whether a neighborhood is animated or not. As for the previous one, the SWRL Rule 2 is formulated with different predicates, mainly IRIS and Sustenance (super class of: Pub, Bar, Restaurant, etc.). These predicates are related to their geographic coordinates, respectively, ?w1 and ?w2. The built-in *distance_criterion_sustenance* is verified through a method as for the built-in *distance_criterion_food_shop* and the assumption made here for the object property *nearby_sustenance(?i, ?s)* is fulfilled if all the predicates are satisfied.

SWRL Rule 2

```
IRIS(?i) ^ hasGeometry(?i,?g1) ^ asWKT(?g1,?w1) ^
Sustenance(?s) ^ hasGeometry(?s,?g2) ^ asWKT(?g2,?w2) ^
distance_criterion_sustenance(?d,?w1,?w2) ^
=> nearby_sustenance(?i, ?s)
```

Next, we define SWRL Rule 3 that handles an OWL class description (*nearby_sustenance* >= 5). This rule infers the fact that a neighborhood ?i is animated (atmosphere(?i, "true")) if there are more than five relations (minimal cardinality restriction) that assert a proximity of the IRIS ?i from a Sustenance (pub, bar, restaurant, etc.).

SWRL Rule 3

```
IRIS(?i) ^ (nearby_sustenance >= 5)(?i)
=> atmosphere(?i, "true")
```

3.3 Inference for distance from workplace

The home-work distance is considered as a very important criteria when someone is looking for a housing. Generally, this distance is expressed with a temporal value. To deal with this criteria, we design SWRL Rule 4 below that infers the different trajectories forming an itinerary. The body of the rule stipulates the fact that when we have an IRIS (?i), its geographic coordinates ?w1 are captured (the same information are collected for the workplace) from the knowledge base. Different itineraries may be proposed to reach the workplace. As we know, the reasoning process allows only inferences about individuals in the knowledge base. Which leads us to create first the instances that should describe the different itineraries. We set five possible itineraries between each IRIS and the workplace. These possibilities are determined in the rule by ?iT1,...,iT5. Furthermore, to capture the association between, for instance the workplace ?k, itinerary ?iT1, IRIS ?i, we use the relations hasDistance(?k, ?iT1) and itineraryFromIRIS(?iT1,?i). Since SWRL does not adopt the unique name assumption, we assert that all individuals are different into the knowledge base and we exploit this assertion using *differentFrom* properties in the rule. Without this assertion, the rule may collect one itinerary from the knowledge base to fulfil the different variables representing itineraries in the rule.

SWRL Rule 4

```
IRIS(?i) ^ hasGeometry(?i,?g1) ^ asWKT(?g1,?w1) ^
Workplace(?k) ^ hasGeometry(?k,?g2) ^ asWKT(?g2,?w2) ^
hasDistance(?k, ?iT1) ^...^ hasDistance(?k,iT5) ^
differentFrom(?iT1,?iT2) ^ differentFrom(?iT1,?iT3) ^
...^ differentFrom(?iT4,?iT5) ^
itineraryFromIRIS(?iT1,?i) ^...^
itineraryFromIRIS(?iT5,?i) ^
work_distance_criterion (?w1,?w2,?dv1,...,
?dv5,?v1,?v2,?v3,?v4...)
=> subway_distance(?iT1,?v1 ) ^...^
tram_distance(?iT5,?v35) ^distance_value(?iT1,?dv1)^...^
distance_value(?iT5,?dv5)
```

The *work_distance_criterion* predicate is a built-in customized in order to generate five optimal itineraries and it fills in the variables ?v1, ?v2,...?v5. For each itinerary we have seven distance values that are described by seven relations. These latter are clustered by the distance_value data property (Figure 3). Hence, each variable (?v1, ?v2,...?v35) represents a temporal value for one trajectory constituting an itinerary. When all the predicates are satisfied, the rule infers the different temporal values for each trajectory. Plus, for each itinerary, the built-in returns "0" as a temporal value, if the transportation related to the data property is not adopted into the itinerary. For example, the rule can return the assertion *subway_distance(?iT2, '0')* if there is no trajectory related to a subway for the second itinerary (?iT2).

Finally, the built-in returns also values (?dv1...?dv5) representing the temporal distance value of an itinerary (e.g., $?dv1 = ?v1 + ... ?v5$). This operation simplifies querying data when users express preferences based on distance to work (see Section 4).

The semantic enhancement presented in this section can evolve by adding other rules. For example, it is possible to infer the social class of a neighborhood if we analyze the influential values using several built-ins. In the next section, we present a use case to understand how rules are used and how the querying process manages inferences.

4 USE CASE

To understand how using this model can be relevant and effective, we explain here its exploitation. First, we introduce a data enrichment process using OSM. Thereafter, we point out how using SWRL rules is important to enhance the semantic through reasoning mechanism and built-ins. At the end, the model suggests recommendations by means of SPARQL queries.

4.1 Data enrichment through NAREO

The main contribution is the NAREO ontology, but it does not contain any instances. To be useful, it needs to be populated with instances from the targeted area, for instance a country, a region or a city. These instances mainly consist of spatial information that can be easily extracted from cartographic providers such as OpenStreetMap or Bing Maps.

In this use case, we show how to use NAREO for recommending neighbourhoods in Lyon. Thus, we instantiate all the environment sub classes with *Lyon city* information using Overpass[5] API (turbo) for OSM. More precisely, we collected data from OSM on a 10km radius around Lyon with values for Shop, Amenity and Leisure keys. Figure 4 shows an example of the individual "crêche Masséna" which is an instance of the *Kindergarten* class. This individual has *Point5775* as a *Geometry* with the serialization (asWKT) "*POINT (4.8552157 45.7645912)*". The same applies to all the spatial data enrichment. However, we noticed that information about transportation are not completely available in OSM. Consequently, we collected datasets about the city transportation system (considering Lyon city) from a French open data website [6]. In addition to this enrichment, information related to IRIS are also stored into the knowledge base from the same website. At the end, the integration process gathered about 86,990 instances.

Note that the ontology is enriched entirely in Java with the JENA API[7]. The integration script can be easily adapted for populating other areas, and is publicly available at https://gitlab.liris.cnrs.fr/fduchate/nareo.

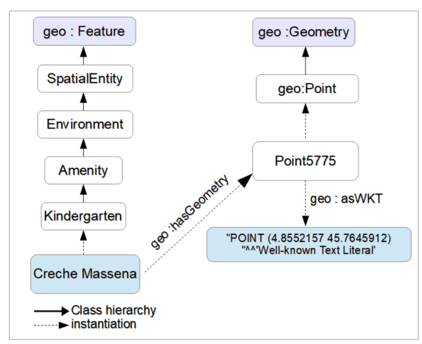

Figure 4: Example of spatial data enrichment

4.2 Reasoning process through SWRL

To clearly understand how information are inferred, we describe an example of inferences through the rules defined in this paper (see Section 3) considering some data and fact assertions. These are presented in Table 3. For instance, *Franprix is a Supermarket and has Geometry Point51975 with POINT (4.867083 45.7737778) as a serialization*, is an assertion from the second row of the table. The inferences are mainly related to recommendation criterion.

[5]https://overpass-turbo.eu
[6]https://www.data.gouv.fr
[7]https://jena.apache.org

Table 3: Example of data assertion

Individual label	Geometry	asWKT	Class
Casino	Point51975	POINT (4.8655005 45.7711646)	Supermarket
Franprix	Point51581	POINT (4.867083 45.7737778)	Supermarket
Boulangerie Régis Grand	Point55750	POINT (4.8618856 45.7705183)	Bakery
Picard	Point62958	POINT (4.8618621 45.77022)	Frozen_food
Provifruits	Point54006	POINT (4.8660036 45.7709304)	Greengrocer
Martins Boucher	Point54389	POINT (4.8669555 45.7733211)	Butcher
Leader Price Express	Point56688	POINT (4.8628103 45.7724536)	Convenience
The Brew Brothers	Point50562	POINT (4.856848 45.7697452)	Pub
Le Waldeck Sweet Bar	Point50623	POINT (4.8583287 45.7698096)	Pub
Le Select	Point57491	POINT (4.8667944 45.7731651)	Bar
Le Charpenne	Point57512	POINT (4.8664263 45.7730669)	Bar
Okawali	Point61447	POINT (4.863124 45.769918)	Restaurant
McDonald's	Point52337	POINT (4.8633017 45.7700648)	Fast_food
Sapori di casa	Point57474	POINT (4.8612948 45.7708478)	Restaurant
Le Béranger	Point57529	POINT (4.8609165 45.7699466)	Restaurant
Le Bistrot du Potager-Stalingrad	Point70582	POINT (4.8586965 45.7706483)	Restaurant
Le Hoggar	Point72993	POINT (4.867677 45.771059)	Restaurant
692660301 (Tonkin-Sud)	MultiPolygon129	MULTIPOLYGON (((4.86594583533633 45.77...	IRIS

4.2.1 Proximity criteria. In the previous section, we detailed two rules (SWRL Rule 1 and SWRL Rule 2) to infer proximity criteria between an IRIS and other spatial entities (Food_shop and Sustenance). Two built-ins were respectively defined according to the entities in question: distance_criterion_food_shop and distance_criterion_sustenance. When the reasoning process is applied to the set of assertions presented in Table 3, SWRL Rule 1 will infer the semantic triples summarized in Table 4. Indeed, the whole body of the rule is fulfilled and every predicate is satisfied. Each triple means that the IRIS named Tonkin-Sud is nearby some *Food_shop*. The same applies for the SWRL Rule 2. We summarize the inference results for the latter rule in Table 5. The inferences about sustenance proximity help us to verify the neighborhood atmosphere (animated or not).

Table 4: Inferences for proximity to Food shops

IRIS	Inference	Food_shop
692660301 (Tonkin-Sud)	nearby_food_shop	Casino
692660301 (Tonkin-Sud)	nearby_food_shop	Franprix
692660301 (Tonkin-Sud)	nearby_food_shop	Boulangerie Régis Grand
692660301 (Tonkin-Sud)	nearby_food_shop	Picard
692660301 (Tonkin-Sud)	nearby_food_shop	Provifruits
692660301 (Tonkin-Sud)	nearby_food_shop	Martins Boucher
692660301 (Tonkin-Sud)	nearby_food_shop	Leader Price Express

4.2.2 Neighborhood atmosphere criteria. Once we have the inference about proximity to sustenance, we can verify whether an IRIS is animated or not, using the SWRL Rule 3 defined in Section 3. The rule is satisfied based on the number of sustenance places nearby the IRIS Tonkin-Sud. From the triples presented in Table 5 we can affirm that there are more that five sustenance locations nearby the neighborhood Tonkin-Sud. Therefore, the rule deduces the fact

Table 5: Inferences for proximity to Sustenance

IRIS	Inference	Sustenance
692660301 (Tonkin-Sud)	nearby_sustenance	The Brew Brothers
692660301 (Tonkin-Sud)	nearby_sustenance	Le Waldeck Sweet Bar
692660301 (Tonkin-Sud)	nearby_sustenance	Le Select
692660301 (Tonkin-Sud)	nearby_sustenance	Le Charpenne
692660301 (Tonkin-Sud)	nearby_sustenance	Okawali
692660301 (Tonkin-Sud)	nearby_sustenance	McDonald's
692660301 (Tonkin-Sud)	nearby_sustenance	Sapori di casa
692660301 (Tonkin-Sud)	nearby_sustenance	Le Béranger
692660301 (Tonkin-Sud)	nearby_sustenance	Le Hoggar

that this neighborhood is animated, adding the following triple: `"692660301" "atmosphere" "true"`.

4.2.3 Distance from workplace criteria. In the last rule (SWRL Rule 4), we define an inference of temporal distance for each trajectory forming one itinerary. Suppose we add these data assertions[8]:

- Itinerary(T1), Itinerary(T2), Itinerary(T3)
- Workplace(Nautibus), hasGeometry(Nautibus, Point9600)
- asWKT(Point9600, POINT(4.865868 45.782342))
- hasDistance(Nautibus, T1), hasDistance(Nautibus, T2)
- hasDistance(Nautibus, T3)
- itineraryFromIRIS(T1, 692660301)
- itineraryFromIRIS(T2, 692660301)
- itineraryFromIRIS(T3, 692660301)

SWRL Rule 4 will infer the following temporal distance values for each itinerary (T1, T2, T3) by assigning the different trajectories through several data properties:

- walking_distance(T1, 11), bus_distance(T1, 2)
- distance_value(T1, 13), train_value(T1, 0)

[8]Note that *Nautibus* is the building name of our LIRIS laboratory.

- subway_distance(T1, 0), car_distance(T1, 0)
- tram_distance(T1, 0), Trolley_distance(T1, 0)
- walking_distance(T2, 5), tram_distance(T2, 5)
- distance_value(T2, 10), train_value(T2, 0)
- subway_distance(T2, 0), car_distance(T2, 0)
- bus_distance(T2, 0), Trolley_distance(T2, 0)
- walking_distance(T3, 19),distance_value(T3, 19)
- tram_distance(T3, 0), train_value(T3,0)
- subway_distance(T3, 0), car_distance(T3, 0)
- bus_distance(T3, 0), Trolley_distance(T3, 0)

All the inferred data are stored as an RDF model to allow SPARQL queries. These queries provide the possibility to express some preferences for neighborhood recommendation.

4.3 SPARQL queries for recommendation

Users preferences are the main components that lead to a relevant recommendation. This section describes how the SPARQL query language is used after the reasoning process to obtain recommendations according to distance, proximity and neighborhood atmosphere[9]. The latter can be related to some semantic knowledge that enhance the recommendation.

A. Itinerary preferences:

SPARQL Query 1 is about detecting neighborhoods (IRIS) for an accommodation by taking into account the distance to a specific workplace.

SPARQL Query 1

```
select ?name_IRIS ?itinerary ?temporal_distance
  where{?x a base:IRIS;
         base:nom_IRIS ?name_IRIS.
         ?w a base:Workplace;
         rdfs:label \"Nautibus\";
         base:hasDistance ?itinerary.
         ?itinerary base:ItineraryFromIRIS ?x;
         base:distance_value ?temporal_distance.
         FILTER (?temporal_distance > 0 &&
                 ?temporal_distance <= 20)
          }
```

In this example, we suppose that a user works in a place named "Nautibus". From this place, we capture itineraries –generated through SWRL Rule 4 – to all IRIS. Then, for each itinerary, we collect the temporal distance value (inferred previously) on which a refinement is applied. This refinement is specified by the FILTER function that allows restriction on arithmetic expressions. The restriction in this query claims that the temporal distance value should not exceed 20 minutes. The minimum value restriction is considered to avoid getting other itineraries that are not initialized through SWRL when there are less than five possibilities to join the workplace. Formally, this query recommends all the neighborhoods from where we reach the Nautibus workplace in less than 20 minutes. The semantic of this query can be enhanced by adding preferences on transportation systems. For example, the triple, `[?itinerary base:tramway_distance ?tmprl_distance]` and the restriction `[FILTER (?tmprl_distance > 0)]` may be added to the query to specify that a user prefers to take the tramway as means of transportation.

[9]For all SPARQL queries, we specify the term "base" for the prefix of NAREO IRI (Internationalized Resource Identifier).

B. Preferences for the proximity:

Shopping is an activity that people cannot avoid. Hence, having conveniences not far from their future housing can be very satisfying and is a frequent query from real estate buyers. SPARQL Query 2 is proposed to compute the number of food shops classified per category nearby in the neighborhood.

SPARQL Query 2

```
select ?name_IRIS (count(distinct ?shop) as ?count)
  where {?x a base:IRIS;
         base:nom_IRIS ?name_IRIS;
         base:nearby_food_shop ?shop.
         } GROUP BY ?name_IRIS
```

In this query, we use aggregate functions allowed by SPARQL (count with GROUP BY). Information about proximity (nearby_food_shop) have to be already inferred through SWRL Rule 1.

C. Preferences for neighborhood atmosphere:

The neighborhood atmosphere is probably one of the most important criterion. Indeed, a user may prefer a lively neighborhood when others such as families would prefer a quiet place.

SPARQL Query 3

```
select ?name_IRIS
   where{?x a base:IRIS;
         base:nom_IRIS ?name_IRIS;
         base:atmosphere true.
         }
```

In this context, SPARQL Query 3 can be used to gather all the neighborhoods that are animated or quiet, expressed respectively by "true"/"false" values. To obtain more diversity in terms of atmosphere or about the neighbourhood environment, new predicates or built-ins need to be written.

D. Sustenance Typology:

One may want to know what kind of sustenance are nearby each neighborhood. Thus, we propose SPARQL Query 4 which may help to improve the recommendation process if we join it with neighborhood atmosphere query to refine the result by adding restriction on sustenance typology.

SPARQL Query 4

```
select ?name_IRIS ?name_sustenance ?class
  where {?x a base:IRIS;
         base:nom_IRIS ?name_IRIS;
         base:nearby_sustenance ?tag.
         ?tag rdfs:label ?name_sustenance;
         a ?class."
         FILTER( STRSTARTS(STR(?class),str(base:)) )
         }
```

For example, SPARQL Query 5 suggests a refining recommendation considering an animated neighborhood and add a restriction on the number of bars nearby.

SPARQL Query 5

```
Select ?name_IRIS (count(distinct ?tag) as ?count)
  where {?x a base:IRIS;
          base:nom_IRIS ?name_IRIS;
          base:atmosphere true;
          base:nearby_sustenance ?tag.
          ?tag rdfs:label ?name_sustenance;
          a base:Bar.
          } GROUP BY ?name_IRIS
          HAVING (?count < 4)
```

The ability to express preferences through SPARQL evolves based on the set of knowledge representation. For instance, recommendation should consider security rating of the neighborhood. Hence, it has to be structured into the knowledge base in order to guarantee its inclusion through SPARQL queries. More enhancement about recommendation through our NAREO ontology are presented in the next section.

5 CONCLUSION AND FUTURE WORK

An ontological representation for neighborhood recommendation appears to be a helpful contribution to suggest relevant neighborhoods based on user preferences. The representation has to be generic and independent of the city where a user is searching for a new housing. Moreover, it must support the scalability in order to update the knowledge base when needed. In this paper, we have presented the NAREO ontological model which enables the representation and population of data related to neighborhood recommendation. This representation considers spatial data about the environment forming a neighborhood (amenity, services, leisure, shops). We reuse the GeoSPARQL standard ontology that eases the spatial representation in NAREO. The taxonomy presented gathers the main concepts related to a neighborhood and relationships to semantically link stored data. SWRL rules with customized built-ins are suggested in order to manage inferences on different criteria such as distance to work, proximity to convenience or sustenance that leads us to propose a rule to infer the atmosphere of the neighborhoods. Characteristics about neighborhood may also be presented through other criteria like natural elements in the surrounding area or distance level of a neighborhood from the city center. We did not consider these characteristics in the present study. Adding semantics describing them would be useful in order to integrate related preferences. In this proposal, preferences are managed by means of SPARQL queries using data inferences through SWRL rules. Functions provided by the query language are also applied to filter values and get a relevant result for recommendation.

Our future work focuses on the evolution of NAREO by adding concepts related to characteristics for real estates (e.g., type of accommodation, number of rooms, presence of a fireplace or swimming pool). This extension is also related to spatial data, especially if we try to represent the layout of the housing (e.g., kitchen oriented south, bedrooms at the second floor) and the SEED layout could be an interesting foundation [10], in addition to spatial relations from the Geosparql ontology. For example, the relation "within" will capture real estates available into the recommended neighborhoods. Then, we will add preferences about real estate (number of rooms, orientation, etc.) considering the semantic representation in the future version of NAREO. The latter will also take into account users profile to propose more accurate recommendation. For example, a neighborhood close to a school will be privileged if the user has children. This recommendation will be handled first by defining a semantic rules to infer the spatial relation between neighborhood and schools. These inferred facts will be used to filter triples in SPARQL queries.

REFERENCES

[1] Ghislain Atemezing and Raphaël Troncy. 2012. Comparing Vocabularies for Representing Geographical Features and Their Geometry. *Terra Cognita 2012 Workshop* (01 2012), 3.

[2] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider. 2003. *The Description Logic Handbook: Theory, Implementation and Applications.* Cambridge University Press.

[3] Nelly Barret, Fabien Duchateau, Franck Favetta, and Loïc Bonneval. 2020. Predicting the Environment of a Neighborhood: a Use Case for France. In *International Conference on Data Management Technologies and Applications (DATA).* SciTePress, 294–301.

[4] Nelly Barret, Fabien Duchateau, Franck Favetta, Maryvonne Miquel, Aurélien Gentil, and Loïc Bonneval. 2019. À la recherche du quartier idéal. In *EGC*. 429–432.

[5] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. 2013. Recommender systems survey. *Knowledge-Based Systems* 46 (2013), 109 – 132.

[6] J.R. Campaña C.D. Barranco and J.M. Medina. July 2006. A Framework for Fuzzy Data Extraction from Non-Structured Sources using Ontologies. In *11TH International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2006)*, Vol. 2. 1826–1833.

[7] Cynthia Chen and Haiyun Lin. 2012. How Far Do People Search for Housing? Analyzing the Roles of Housing Supply, Intra-household Dynamics, and the Use of Information Channels. *Housing Studies* 27, 7 (2012), 898–914.

[8] Elizabeth C. Delmelle. 2015. Five decades of neighborhood classifications and their transitions: A comparison of four US cities, 1970–2010. *Applied Geography* 57 (2015), 1 – 11.

[9] M. Egenhofer and J. Herring. 1991. *Categorizing Binary Topological Relationships Between Regions, Lines, and Points in Geographic Databases.* Technical Report. Department of Surveying Engineering, University of Maine.

[10] Makoto Inoue and Hideyuki Takagi. 2008. Layout algorithm for an EC-based room layout planning support system. In *IEEE Conference on Soft Computing in Industrial Applications.* IEEE, 165–170.

[11] Géraud Le Falher, Aristides Gionis, and Michael Mathioudakis. 2015. Where Is the Soho of Rome? Measures and Algorithms for Finding Similar Neighborhoods in Cities. *ICWSM* 2 (2015), 3–2.

[12] Huayu Li, Yong Ge, Richang Hong, and Hengshu Zhu. 2016. Point-of-interest recommendations: Learning potential check-ins from friends. In *SIGKDD*. ACM, 975–984.

[13] Yun Li, Yixiang Fang, Reynold Cheng, and Wenjie Zhang. 2019. Spatial pattern matching: a new direction for finding spatial objects. *SIGSPATIAL Special* 11, 1 (2019), 3–12.

[14] Jacek Malczewski and Mohammadreza Jelokhani-Niaraki. 2012. An ontology-based multicriteria spatial decision support system: a case study of house selection. *Geo-spatial Information Science* 15, 3 (2012), 177–185.

[15] Ikrom Nishanbaev, Erik Champion, and David Mcmeekin. 2019. A Survey of Geospatial Semantic Web for Cultural Heritage. *Heritage* 2 (05 2019), 1471–1498.

[16] David A Randell, Zhan Cui, and Anthony G Cohn. 1992. A spatial logic based on regions and connection.. In *3rd International Conference on Knowledge Representation and Reasoning*. 165–176.

[17] Hao Wang, Manolis Terrovitis, and Nikos Mamoulis. 2013. Location recommendation in location-based social networks using user check-in data. In *SIGSPATIAL*. ACM, 374–383.

[18] Xiaofang Yuan, Ji-Hyun Lee, Sun-Joong Kim, and Yoon-Hyun Kim. 2013. Toward a user-oriented recommendation system for real estate websites. *Information Systems* 38, 2 (2013), 231–243.

[19] Amy X Zhang, Anastasios Noulas, Salvatore Scellato, and Cecilia Mascolo. 2013. Hoodsquare: Modeling and recommending neighborhoods in location-based social networks. In *Social Computing*. IEEE, 69–74.

[20] Leonard V Zumpano, Ken H Johnson, and Randy I Anderson. 2003. Internet use and real estate brokerage market intermediation. *Journal of Housing Economics* 12, 2 (2003), 134–150.