Article

Generating Spatial Knowledge Graphs with 2D Indoor Floorplan Data: A Case Study on the Jeonju Express Bus Terminal

Hanme Jang [1] (ID)**, Kiyun Yu** [1,2] **and Jiyoung Kim** [3,*] (ID)

1   Department of Civil and Environmental Engineering, Seoul National University,
    Seoul 08826, Republic of Korea; janghanie1@snu.ac.kr (H.J.); kiyun@snu.ac.kr (K.Y.)
2   Institute of Engineering Research, Seoul National University, Seoul 08826, Republic of Korea
3   Social Eco-Tech Institute (SETI), KonKuk University, Seoul 05029, Republic of Korea
*   Correspondence: elliekim@konkuk.ac.kr

Abstract: With the boom in online information, knowledge graphs like Freebase, Wikidata, and YAGO have emerged, thanks to the introduction of the RDF (Resource Description Framework). As RDF data grew, more and more spatial data was incorporated into it. While we have a lot of 2D data for outdoor spaces, mapping indoor spaces in 3D is challenging because it is expensive and time-consuming. In our research, we turned 2D blueprints into detailed 3D maps and then translated this into RDF format. We used the Jeonju Express Bus Terminal in South Korea as our test case. We made an automated tool that can turn 2D spatial data into 3D data that fits the IndoorGML standard. We also introduced terms like 'loc', 'indoorgml-lite', and 'bloc' to describe indoor spaces in the RDF format. Once we put our data into a GraphDB database, we could easily search for specific details and routes inside buildings. This work fills a significant gap in knowledge graphs concerning indoor spaces. However, the production of large-scale data across varied areas remains a challenge, pointing towards future research directions for more comprehensive indoor spatial information systems.

Keywords: RDF; ontology; IndoorGML; navigation

1. Introduction

With the exponential growth of information on the Internet and the proposal of the RDF (Resource Description Framework), a semantic web standard model for representing resources on the web like Freebase [1], Wikidata [2], and YAGO [3] have emerged. Searching through hundreds of GB of data stored in the knowledge graph for users has become a paradigm of information retrieval known as KBQA [4–7] and it is now well-recognized that, as the size of these massive Knowledge Bases increases, the performance to integrate and accumulate the various type of data could be enhanced [8]. Additionally, with the evolution of web semantic technology, various data types, including spatial data formats like shapefile, CityGML, KML, and GeoJSON [9], have been converted to RDF. For example, Auer et al. combined Open Street Map (OSM) data with DBpedia and Geonames to represent 2D spaces in RDF [10]. Punjani et al. have offered RDF datasets mixing DBpedia General Administrative Divisions (GADM) and OSM [11]. Dsouza et al. attempted to interlink existing Knowledge Bases by combining Wikidata, OSM, and DBpedia [12]. Moreover, some studies introduced how to generate an ontology for CityGML [13] as a representative of 3D spatial data. In the case of [14–16], there have been attempts to provide services such as flood management, photovoltaic power generation analysis, and Point of Interest (POI) search, which were traditionally conducted using Geographic Information System (GIS) data, using Resource Description Framework (RDF) data instead. However, technical limitations existed, as it was necessary to build data and develop spatial functions such as intersection and union, which are not supported in the existing Knowledge Base Question Answering (KBQA) ecosystem.

The prevalence of 3D indoor spatial data is limited compared to the widespread and consistently constructed 2D spatial information used for cadastral systems in many countries, which often cover entire territories. This is primarily due to the high costs and time-consuming nature of its construction [17]. There have been a few attempts to convert indoor space features to RDF triple data. For instance, Chen et al. modeled indoor sensors, doors, furniture, devices, etc., in RDF format and suggested some activities that can be performed in indoor environment [18]. Fensel et al. built a system after performing RDF modeling on devices like thermometers, hygrometers, and air quality measuring instruments that monitor indoor environments [19]. However, no attempt was made to systematically produce indoor spatial data.

When producing indoor space data, 2D spatial information from sources like CAD drawings or 3D models are typically used. While 2D spatial information offers advantages such as ease of acquisition, abundance, and cost-effectiveness, it struggles to represent 3D indoor spaces. Conversely, using 3D models guarantees data accuracy but comes with the disadvantages of high cost and time [20–22]. Given the necessity of 3D representation for various applications like firefighting, emergency situations, and navigation [23,24], relying solely on 2D spatial information for indoor space data construction remains a limitation [25–27].

Currently, for example, Jamsil Station in Seoul can be identified on Freebase as m.0281d_p, while the nearby Lotte Tower is recognizable on Wikidata as Q12616233. However, the 'natuur' store in Figure 1 located inside the building on the third floor cannot be found on Freebase, Wikidata, or WorldKG. Consequently, querying data in RDF format for this store is not possible. Additionally, navigating a path efficiently between stores inside a building present a challenge. If systematic indoor spatial data were to be established, it would be feasible to use RDF-based indoor spatial information to query interior objects, facilitate pathfinding, and link with external data.

Figure 1. Jamsil station, Nauur store inside the building.

Therefore, this study aims to store indoor space information using 2D indoor spatial data, construct 3D indoor space data representing indoor routes, and design a schema to produce RDF data. We also aimed to produce RDF data according to the schema, input it into the database, and demonstrate the usability of the generated data. To realize these objectives, we employed Safe Software's FME for automating the production of 3D data and developed a parser to convert the produced 3D spatial information into RDF data. We also utilized graphDB, a widely known GraphDatabase, for data operations.

This study's contributions are:

- Designing a process to create 3D spatial information compliant with standard formats for representing building interiors and indoor routes using 2D spatial information so that an efficiency in data production can be expected.

- Enabling the use of 3D indoor spatial information as RDF data by developing a parser to generate data from the created 3D spatial information.

The remainder of this paper is structured as follows: Section 2 reviews the relevant research on spatial information's RDF, indoor space information data model. Section 3 describes the data and methods to generate 3D spatial data and RDF triples. Section 4 shows the converted data and presents case studies utilizing the data. Section 5 and 6 give a discussion and the conclusion of the paper, respectively.

2. Theoretical Background

2.1. RDF

RDF has been introduced as a standard for representing web data and was first proposed by Tim Berners-Lee in 2006 [28]. It is composed of a triple model consisting of a subject, predicate, and object, as depicted in Figure 2. In this structure, the subject represents the data of interest, the predicate describes the subject or represents the relationship between the subject and the object, and the object conveys the content or value related to the predicate.
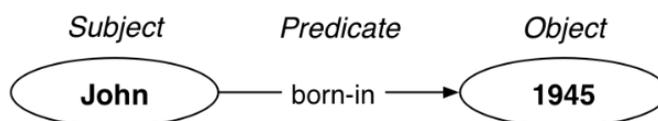

Figure 2. RDF triple.

In RDF, each element is associated with a distinct namespace, which encompasses a collection of identifiers to denote resources. These namespaces are crucial to prevent potential clashes between identifiers bearing the same name but with varied interpretations. Every namespace is paired with a unique Uniform Resource Identifier (URI) to set it apart from other namespaces. Thanks to these namespaces, it becomes feasible to merge and utilize data from various RDF databases, including Linked Open Data (LoD), FreeBase, and Wikidata. In recent times, the transformation of spatial information into RDF format has emerged as one of the significant topics in the research domain aiming for the integration of geographic information and semantic web technologies. Regarding spatial information, multiple RDF namespaces have been instituted by various organizations to aid in the seamless exchange and amalgamation of spatial data on the Internet. Table 1 shows some of the notable RDF namespaces dedicated to spatial information.

Table 1. Namespaces used in spatial information area.

Name	URL	Description	Access date
wd	http://www.wikidata.org/entity/	Entities in Wikidata	
osm	https://www.openstreetmap.org/	Entities in OpenStreetMap	
lgdo	http://linkedgeodata.org/ontology/	Entities in LinkedGeodata [10]	10 October 2023
wkgs	http://www.worldkg.org/schema/	Entities in worldKG [12]	
geosparql	http://www.opengis.net/ont/geosparql/	Geosparql ontology [29]	

2.2. Indoor Spatial Data to RDF

Following the initial discussion, there have been numerous attempts to express indoor environments. However, there are a few cases that model indoor spaces from an RDF perspective, especially those that implement indoor objects and paths. For instance, Pauwels et al. enhanced data interoperability by analyzing the semantics of IFC data, a CAD exchange format representing building structures and functionalities, transitioning it to RDF format [30]. Nevertheless, this model primarily focused on IFC objects, overlooking the interconnections between objects and the construction of indoor pathways. Several studies have highlighted navigation within architectural floorplans, emphasizing the identification

of indoor routes connecting spaces or entities. For example, Matuszka et al. stored POIs within buildings in RDF format and navigated between them using the relational term 'hasAvailable' [31]. However, the predicates were made only among a limited number of bases. Fleiner et al. developed an Indoor Navigation Linked Data at Obuda University by extracting spaces from two-dimensional architectural blueprints of the university buildings [32]. The semantics used to represent these spaces utilized the LOC namespaces, which are extensions from the iLOC namespaces [33]. This approach further differentiated spaces and classified corridors and rooms into "Room" and "Hallway", respectively. The research also introduced SPARQL syntax for identifying routes transitioning between floors. Chalhoub et al. developed a hospital navigation system employing hLOC, a spatial representation ontology that shares namespaces with both iLOC and iACC [34]. However, these two studies predominantly depicted movement paths between POIs, making it challenging to achieve geometrically precise routes. Furthermore, the lack of standardized data production methods complicates the consistent creation of RDF data for detailing indoor spatial contexts.

2.3. Indoor Spatial Data Format

As the range of spatial information expanded to include interiors, various formats have been developed to represent indoor spaces. Traditionally, previous studies constructed data by directly inputting objects from data source into databases manually. However, this method was not standardized and relied on the experience of the worker, making it unsuitable for handling large-scale data. To make the RDF data more accurate and universal, an automated process is necessary for data production and conversion. This study aimed to utilize the standard format for indoor space data. The widely recognized standards for indoor data are as follows [35–37]:

- The Indoor Mapping Data Format (IMDF) [38] was developed by Apple as an indoor mapping standard and is used in the Apple Map, ESRI, and other industrial fields. IMDF is primarily used for developing commercial navigation software within indoor spaces, making it advantageous for depicting attributes of large structures such as airports and shopping malls. The geometrical and attribute information is in a JSON-based format and was adopted as a standard by the Open Geospatial Consortium (OSM) in 2021.
- OSM indoor [39] can be seen more as a tagged schema in OSM rather than an independent data model, designed to represent indoor objects. It can depict individual spaces, doors, and hallways, and can be converted into the XML and GeoJSON formats. Through these details, users can navigate inside buildings and access indoor-specific information. The data, being open source, can be utilized by developers to create indoor navigation apps or enhance existing mapping solutions.
- CityGML [40], an open data model based on GML, is a data exchange standard for city and city features currently defined by ISO TC211. As a specialized application called Application Domain Extensions (ADEs), CityGML offers additional frameworks for urban modeling, excavating, water management, etc. CityGML possesses a concept of differentiating urban models based on accuracy, from LoD1-4, with LoD4 being capable of representing indoor objects. However, it does not explicitly represent indoor paths.
- IndoorGML [41] was developed by the OGC as a standard for representing indoor spatial models and functions as an exchange format. IndoorGML can represent attributes of individual indoor spaces, connections between spaces, and indoor Points of Interest (POI). It is characterized by its compatibility with BIM, CityGML, and various spatial information data.

2.4. IndoorGML Schema

As detailed in Section 2.3, numerous data formats are available for portraying the indoor environment of buildings. Among them, IndoorGML stands out as an open spatial

information representation standard to describe indoor environment and routes. It has the advantage of being compatible with various formats, such as CityGML and IFC data, that depict urban spaces [42]. Moreover, rather than focusing on representing the specific shapes of indoor objects, IndoorGML primarily emphasizes representing the volumes occupied by spaces within the building and the connectivity between these spaces and adjacent other spaces, as depicted in Figure 3. In detail, it classifies spaces, such as rooms, corridors, and staircases, into cellular spaces, with each space being allocated a unique identifier known as PrimalSpaceFeature. Additionally, to achieve a more accurate representation of indoor navigation, larger areas such as corridors and staircases can be subdivided into multiple subspaces. This approach allows for a finer granularity in depicting realistic indoor pathways. Within this model, the components named CellSpace and CellSpaceBoundary signify the volume and its entrances of these spaces, respectively. Conversely, each CellSpace can be transformed into a State, allowing for the representation of movement pathways connecting rooms through doors in a State–Transition structure, much like the Figure below. The MultiLayeredGraph visually portrays the relationships between spaces. The elements named State and Transition mirror route nodes and segments respectively. This model further incorporates efficiency attributes, such as linking the CellSpaceBoundary, to diminish file size, and binding attributes that depict spatial relationships. The two primary features, PrimalSpaceFeature and MultiLayeredGraph, are interconnected via a duality relationship. This relationship distinguishes the corresponding features from one another.
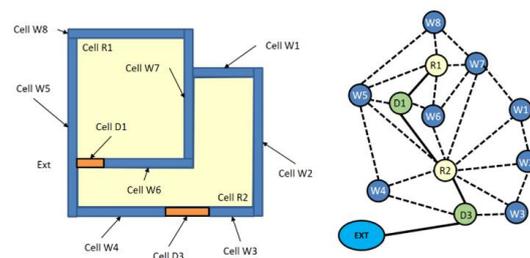

Figure 3. IndoorGML PrimalSpaceFeature and MultiLayeredGraph [24].

3. Methodology

Our methodology is organized into two critical subsections: 3D data construction and RDF data generation. Th first, 3D data construction, involves transitioning from 2D shapefile-managed spatial data to 3D indoor spatial representations. To achieve this, we describe the properties of the employed 2D shapefile data and outline a methodology for generating 3D indoor spatial data. This generation process includes matching 2D entities with the attributes of 3D spatial information, executing 3D spatial operations, and subsequently producing 3D objects. The RDF data generation focuses on changing the made 3D indoor space data into RDF format. To help with this change, an RDF structure (specifically RDF OWL) is set up. The method includes using namespaces that fit the 3D indoor space data's layout to design the RDF data schema, and then developing a parser to make RDF data that matches the designed schema.

3.1. Data

The research focuses on the Jeonju Express Bus Terminal, a building in South Korea that was newly constructed in 2016. Buses in South Korea have the second-highest transportation share after private cars. As a regional hub facility, the building frequently accommodates passengers, offering various types of points of interest (POIs), such as a bus ticket office, shops, restrooms, and boarding areas. However, the Jeonju Express Bus Terminal is currently managed as 2D indoor spatial data created using building floorplans and assigned a single address in national-level spatial databases. The 2D indoor spatial data—represented in shapefile format and derived from building blueprints, as illustrated in Figure 4a—identifies distinct spaces such as rooms, open squares, corridors, and en-

trances. This dataset facilitates mapping of movement pathways through corridors and includes attributes such as District Code, City Name, Room Type, and Room Name. Each room is categorized into spaces commonly found in bus terminals, such as ticket offices, restrooms, etc. The specific types of rooms are summarized in Table 2. Furthermore, entrance information for each space is stored in point form, as shown in Figure 4b. A shapefile has been produced for each floor; there is no Z-coordinate.
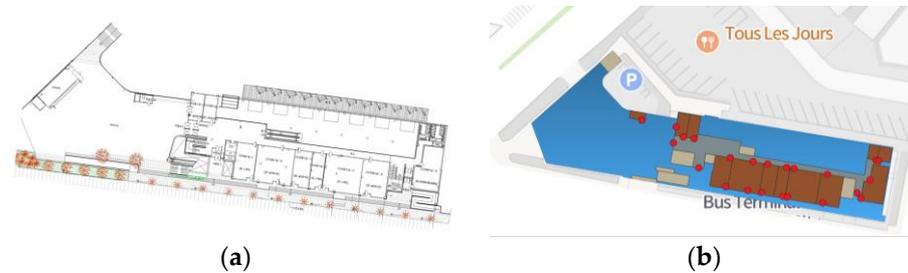

(a) (b)

Figure 4. The 2D floorplan image (**a**) room polygons and door points generated from the image (**b**).

Table 2. Type and number of rooms.

No.	Type	Number of Features	No.	Type	Number of Features
1	Accomodation	2	7	InfoCenter	1
2	AirChamber	1	8	NursingRoom	1
3	Atrium	1	9	Ticket	1
4	BookStore	1	10	Toilet	8
5	Deck	1	11	Office	6
6	Hallway	3	12	Residence	16

3.2. 3D Feature Generation

This section outlines the process of extracting information from 2D shapefiles for the production of 3D spatial data and its conversion into IndoorGML objects. Given that each 2D shapefile lacks a Z-coordinate, our first step was to assign 3D coordinates to every 2D shapefile. When creating a floorplan, the height of each floor entered is 3 m, below the typical floor height of a bus terminal or public facility. This measurement was accepted due to the difficulty in obtaining exact values. Door heights were set at 2.1 m, in compliance with Korean construction regulations. Utilizing the IndoorGML data structure analysis from Section 2.4, a correspondence table maps 2D objects to 3D IndoorGML features. Table 3 primarily categorizes the spatial processes into the extraction of PrimalSpaceFeature and MultiLayeredGraph. Figure 5 fully illustrates this process.

- PrimalSpaceFeature entities generation

The shapefile depicts rooms as closed polygons and the entrances as points. To represent door features as rectangular solids, a buffer operation was applied to the point objects, resulting in line segments from their intersection with the room polygons. It is crucial to depict actual indoor spaces with wall thickness; thus, according to the IndoorGML thick model, spaces adjacent to each other should be separated. To accomplish this, a negative buffer operation was applied to each indoor space to account for wall thickness. Line segments representing doors, positioned in the middle of separated room spaces, underwent a vertical positive buffer operation to bridge the gaps between rooms, facilitating passage. Door thickness was made slightly greater than that of walls. Rooms and doors represented by 2D polygons underwent a 3D extrusion process, being swept to heights of 3m and 2.1m, respectively. Subsequently, a 3D intersection operation between rooms and doors was performed to define the spaces for doors, rooms, and their boundary surfaces. These boundary surfaces, delineating elements on either side of rooms and doors, were

designated as the partialboundedBy property for each CellSpace. All geometric data was stored in gml:Solid and gml:Surface formats. Vertical transportation areas were identified, 3D extruded, and then the ceiling face was isolated. A 3D intersection with the upper floor's objects ensured connectivity for each vertical transportation space.

Table 3. The 3D feature generation method.

| Construction Element		IndoorGML Feature	Generation Method
Primal-SpaceFeature	Wall (Polygon)	CellSpace	3D polygon extrusion
	Door (point)	CellSpace	3D polygon extrusion, intersection with solids from Wall polygon
		CellSpaceBoundary	Intersection between wall solid and door solid
Multi-LayeredGraph	Points (line) connecting center of wall (point)	State	Center point of the 3D solid
		Transition	Extraction of the connection relationship between centers.
	Path (line)	State and Transition	Extraction of the movement path's point
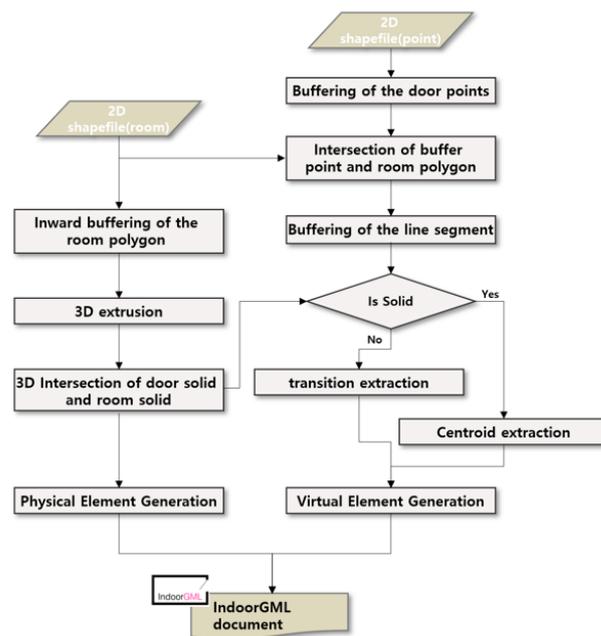
Figure 5. Workflow to generate IndoorGML features.

- MultiLayeredGraph entities generation

Entities that represent indoor navigation routes are composed of States and Transitions within the IndoorGML MultiLayeredGraph framework. Each State is positioned at the centroid of the corresponding CellSpace, which includes rooms and doors. Given that the shape of each CellSpace resembles a prism, the coordinates for each State are determined as follows:

$$C_x = \frac{1}{6A}\sum_{i=0}^{N-1}(x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$
$$C_y = \frac{1}{6A}\sum_{i=0}^{N-1}(y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \qquad (1)$$
$$C_z = \frac{1}{2}(z_h - z_l)$$

where $A = \frac{1}{2}\sum_{i=0}^{N-1}(x_i y_{i+1} - x_{i+1} y_i)$

Furthermore, by utilizing the CellSpaceBoundary objects stored within the partial-boundedBy property, adjacent CellSpaces were identified and extracted. Subsequently, the

State associated with each CellSpace was connected to generate both States and Transitions. In the final step, a duality relationship was established, linking each CellSpace with its corresponding State, as well as linking each CellSpaceBoundary with the Transition derived from it.

3.3. IndoorGML to RDF

IndoorGML is an XML document format that consists of tags and values presented in text form. Various established methods exist for transforming XML documents into RDF, such as GeoTriples [43], which publishes geospatial data as RDF. Methods for converting CityGML, a standard 3D city model format, into RDF have been documented [12]. To date, there are no known examples of converting IndoorGML into RDF. Additionally, there is a recognized need for developing a new data model that includes various POIs specifically designed for bus terminals. This section explores the development of an OWL ontology based on IndoorGML and the subsequent generation of RDF data.

3.3.1. Schema Mapping

This section outlines the ontology design process for indoor spatial data. The UML diagram for the IndoorGML core module shows the relationships between classes and their semantic information. This diagram highlights features like duality, partialboundedBy attribute, and geometry. These features are part of the IndoorGML core module, including indoorgml:CellSpace and indoorgml:CellSpaceBoundary. Indoorgml:CellSpace represents individual spaces separated by walls, such as offices, restrooms, and cafes. The geometry of indoorgml:CellSpace typically uses the gml:Solid method. Neighboring CellSpaces with shared surfaces are depicted as indoorgml:CellSpaceBoundary. Here, partialboundedBy serves as an attribute to link to the respective indoorgml:CellSpaceBoundary. Conversely, the indoorgml:State category denotes nodes within a structure, such as doors, corridors, and spaces. Elements from the indoorgml:State category appear as geometric points in IndoorGML. The indoorgml:Transition class covers edges, showing close relationships between two indoorgml:State entities, like doors, staircases, or elevators. Entities from the indoorgml:Transition class are shown as gml:Curve within the OGC Geography Markup Language framework. To create RDF data in detail, existing namespaces were used to show the structure of IndoorGML's indoor features. Additionally, a new semantic ontology was developed to convey various room features' meanings. The loc ontology, introduced by Obuda [32], served as the foundational structure. This ontology details different indoor components such as places, buildings, rooms, and entrances. The loc ontology was selected for its detailed room descriptions to depict the main spaces in indoor features, and we added some room types and attributes by creating a custom ontology for the bus express terminal. The connection relationships between 'State' and 'Transition' characterize the indoor movement network in IndoorGML. We defined the State–Transition relationship using the Indoorgml-lite namespace [44] in this study. Significantly, Indoorgml-lite defines both the State and Transition nodes and the 'connects' predicate, indicating the link between these nodes. Additionally, the 'duality', 'partialboundedby' attributes, and the geometric information of each object were integrated, referencing the namespaces of OWL [45] and geosparql [29] namespace.

3.3.2. RDF Data Generation

Figure 6 illustrates the detailed process of converting an IndoorGML model into RDF, starting from the initial reading of the IndoorGML file to the final storage of data in an RDF triple-store within a GraphDB database utilized by the system. This conversion includes selecting specific indoor attributes, transforming IndoorGML into RDF, and verifying the file schema afterward. Historically, the CityJSON format, a specialized JSON variant incorporating both attributes and geometric details, was developed for converting CityGML documents into binary data [46]. In a similar vein, Ledoux introduced IndoorJSON, designed specifically for handling IndoorGML files and their principal labels [47]. Inspired by

this transformation approach, we converted the original IndoorGML file into a key-value format by utilizing the IndoorGML XML Schema Definition (XSD) to dissect the XML file's structure. Throughout this process, critical elements that mirror IndoorGML's inherent graph structure were maintained. This key-value representation of IndoorGML was subsequently converted into an RDF Turtle text file, adhering to a pre-established schema and namespaces.
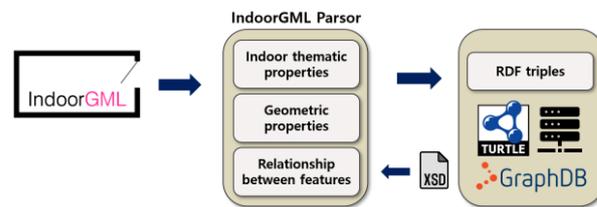

Figure 6. Parsing IndoorGML to RDF turtle.

4. Data Generation and Result

4.1. 3D Object Generation from 2D Shapefile

With regard to 3D spaces, this section presents the outcomes of overlay operations performed between doors (with point geometric information) and rooms (with polygon geometric information), as well as the results of buffer applications and 3D extrusions for each object. In the IndoorGML 3D thick model, as the thickness of the walls existing between spaces is considered in the modeling, the adjacent rooms from the original data shown in Figure 4b were shaped into a rectangle using a negative buffer operation, whereas a positive buffer operation was applied to the door to fill the gap between the walls suitable for the IndoorGML thick model, as shown in Figure 7a. Following this, a 3D extrusion process was carried out, enabling the visualization of the derived 3D object in Figure 7b.
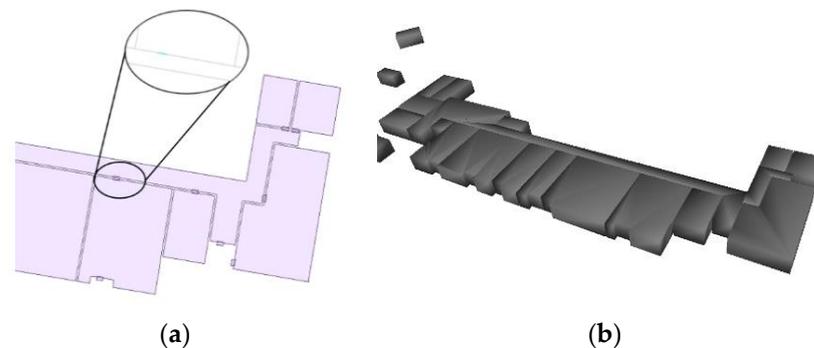

(**a**) (**b**)

Figure 7. Room and door geometry with buffer (**a**) and 3D extrusion (**b**).

For openings, due to the overlapping of the door and room within the 3D space, an intersection operation was performed to distinguish between the CellSpace and CellSpaceboundary. Subsequently, each CellSpaceBoundary surface was stored using the xlink:href attribute associated with its neighboring CellSpace. As shown in Figure 8, conducting the intersection operation results in two CellSpaceBoundaries for each door. This suggests that the space inside the building is navigable for people and other moving objects.

For the indoor routing paths, each space is categorized into one of four types: corridor, room, door, or vertical movement area. Primarily, doors act as connectors between rooms and corridors, or between two rooms, facilitating vertical movement. Within this framework, State features formed from adjacent spaces are interconnected through Transition features. Especially for vertical movement passages such as staircases and elevator shafts, they were first separated from the 2D shapefile and then the vertical entities were extruded in 3D, after which the ceiling surface was separated, and a 3D intersection was performed with objects on the upper floor. A cubic vertical passage was created at the location of the

intersection surface, and the center of gravity was extracted as a State to create a vertical movement Transition. In IndoorGML, vertical movement passages do not conceptually have a special meaning, so they were made as small objects to represent spaces that can be moved through. Generated States and Transitions are shown in Figure 9.
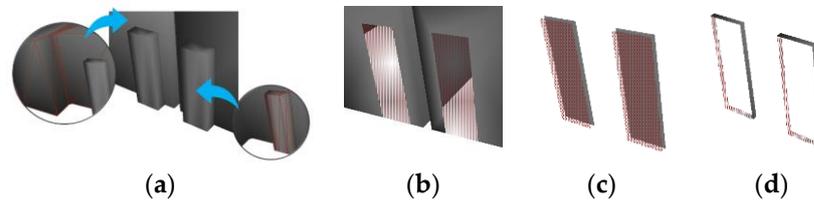

Figure 8. Intersection (**a**), room solid with opening (**b**), boundary surface (**c**), door solid with opening (**d**).
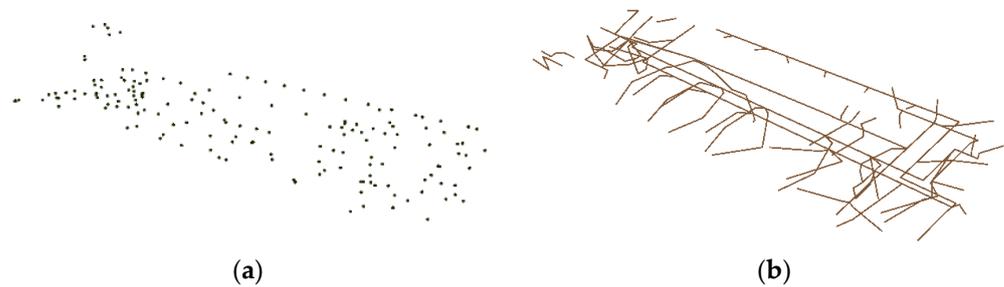

Figure 9. States (**a**) and Transitions (**b**).

4.2. RDF Data Generation

As described in Section 3.3.1, the structure of the RDF data generated in this study is illustrated in Figure 10 below. Each node is originated from LOC and indoorgml-lite namespaces and additional entities such as Accommodation, BookStore, and Ticket, which are not defined in LOC, were introduced to represent points of interest in the express bus terminal with custom namespace bLOC. Indoor routes utilize the 'connects' predicate from indoorgml-lite to define the relationship between each space and its entrance. Additionally, a 'path' predicate was incorporated between States to facilitate efficient pathfinding.
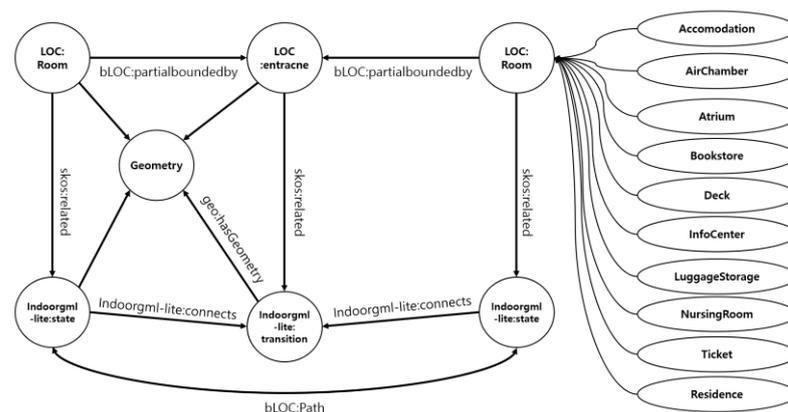

Figure 10. RDF schema.

Figure 11 is an example of a room classified as 'office' and a Transition. A room has attributes such as name, partialboundedBy, and room number. The hasGeometry predicate can be used to retrieve its geometric information. Furthermore, a Transition includes attributes like connects, gmlId, and skos:related, which indicates the corresponding boundary feature in IndoorGML.

```
bloc:262 a <http://lod.nik.uni-obuda.hu/oloudindoor#Office>;
<http://lod.nui-obuda.hu/oloudindoor#isPartOf> bloc:8000;
<https://www.opengis.net/ont/geosparql#hasGeometry>bloc:263,bloc:264;
bloc:Name "office4";
bloc:PARTIALBOUNDEDBY bloc:7907;
bloc:SI "45113";
bloc:gmlId "CellSpace_12";
bloc:roomId "4511331485331000008";
bloc:roomNumber "310";      bloc:7794 a indoorgml-lite:Transition;
                              indoorgml-lite:connects "S_18", "S_tr_57";
                            bloc:gmlId "T_tr_57T_18";
                            skos:related "Door_CellSpace_tr_57CellSpace_18"
```

Figure 11. Office room and Transition in RDF Turtle format.

Commonly, prior research addressing the conversion of data formats has employed a comparison table of matched object pairs to assess data generation accuracy, as referenced in [48,49]. In alignment with this evaluation methodology, this study evaluates the conversion accuracy in Table 4 by comparing the quantity of objects represented in the transformed RDF triples with the number of features present in the IndoorGML documents.

Table 4. Type and number of entities from 2D shapefile to RDF triples.

2D Shapefile Features	Count	RDF Triples		Count	Accuracy (%)
Room	53		CellSpace	61	>100
Door	76		CellSpaceBoundary	76	100
-		Subject	State	204	100
-			Transition	203	100
-			connects	406	100
-			partialboundedBy	152	100
-		Predicate	path	203	100
-			related	274	100

Firstly, a total of 53 rooms extracted from the 2D shapefile were converted into CellSpace, and 8 objects were additionally created for vertical movement. The number of doors is 76, with each opening being converted into CellSpaceBoundary. The partial-boundedBy predicate was generated exactly twice the number of CellSpaceBoundaries. Since States were additionally created to represent corridor spaces, it was observed that the number of States and Transitions does not exactly match the number of CellSpaces and CellSpaceBoundaries. This is because additional States and Transitions were created to represent corridors, in addition to single States and Transitions representing individual spaces. These objects help provide a more detailed representation of the path when moving through a long corridor space. The number of connects predicates was twice that of Transitions because each Transition is connected to two States. Finally, the number of related predicates was twice the combined number of CellSpaces and CellSpaceBoundaries so that CellSpace is associated with State, and CellSpaceBoundary is associated with Transition to refer to each other. Upon verifying the number of generated RDF objects, it was confirmed that the data conversion was correctly performed.

4.3. Visualization and Case Study

The 3D indoor spatial information generated from the 2D shapefile was converted to the RDF turtle format, in accordance with the defined schema above. The data can represent rooms, corridors, indoor routes, and their respective geometric information. Rooms were defined not only to depict semantics like toilets, offices, and hallways, which are already defined in the LOC namespace, but also entities like accommodations and nursing rooms. The IndoorGML 3D model output and indoor paths are illustrated in Figures 12 and 13. Visualization was carried out using the Inviewer developed by Busan National University, confirming the generation of features such as geometric information, entrances, and States for each space.
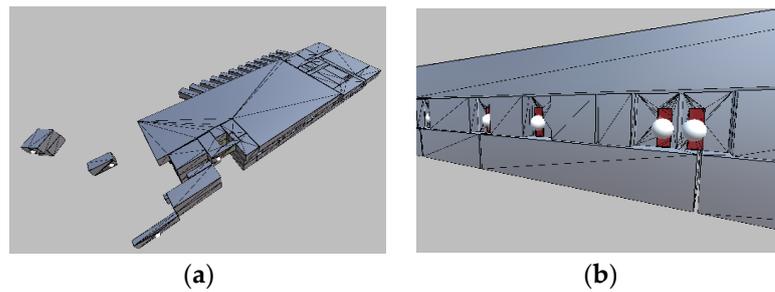
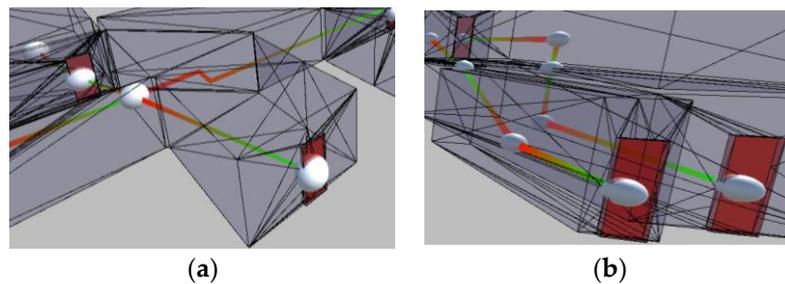**Figure 12.** Overall view (**a**) entrances and walls (**b**).

Figure 13. Horizontal path (**a**) vertical path (**b**).

To evaluate the practicality and compatibility of the data created in this research, two scenarios are presented below. The first is to query information about the geometry and opening of LuggageStorage. Starting from the node with LuggageStorage type, and using the geo:hasGeometry predicate, the geometry of the room can be accessed. By following the bloc:partialboundedBy predicate, the geometry of the connected opening can be found. In Figure 14 below, the type of each node and additional predicates between nodes are presented.
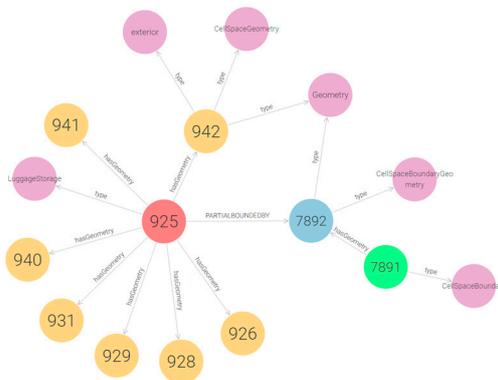

Figure 14. LuggageStorage node at the center (red), geometry nodes (orange, blue), door node (green) and node type (purple).

For the second scenario, we focused on the outdoor area around the express bus terminal. Here, we utilized data from OSMTTL [50], recognized as the most accessible knowledge graph in South Korea for spatial RDF information. This data was seamlessly merged with the RDF dataset generated through our approach. Following this integration, we applied a pathfinding technique to further demonstrate the utility of our combined data set. No direct link existed between our study's data and the neighboring road, represented as osm:way in the RDF data from OSM, so we manually connected the nearest entrance of the building. For constructing the SPARQL query to navigate from a nearby motel to a bookstore on the third floor, we utilized the osmkey:name:en and skos:related predicates.

These were used to identify the node representing the hotel and the State representing the bookstore, respectively.

The specific query is provided in Figure 15, and the results, supplemented with spatial data outside, are illustrated in the accompanying Figure 16. The OSM node representing a motel can be connected to a node on the road nearby, and then by using the bloc:path predicate that represents the State–State connection inside the building, it is possible to search the shortest path to the node corresponding to a bookstore. By verifying the type of cellspace node that has a skos:related predicate at the end of the path, it can be confirmed that the pathfinding has been accurately performed.

```
SELECT ?pathIndex ?edgeIndex ?edge
WHERE {
    ?src osmkey:tourism "motel".
    ?src osmkey:name:en "Sillauidalbam Mo-tel".
    ?dstCellspace a bloc:BookStore.
    ?dst owl:equivalentClass ?dstCellspace

    SERVICE path:search {
        [] path:findPath path:shortestPath ;
           path:sourceNode ?src ;
           path:destinationNode ?dst ;
           path:pathIndex ?pathIndex ;
           path:resultBindingIndex ?edgeIn-dex ;
           path:resultBinding ?edge ;
           path:maxPathLength 30 ;
        .
        SERVICE <urn:path> {
            ?start bloc:Path ?end
        }
    }
}
```

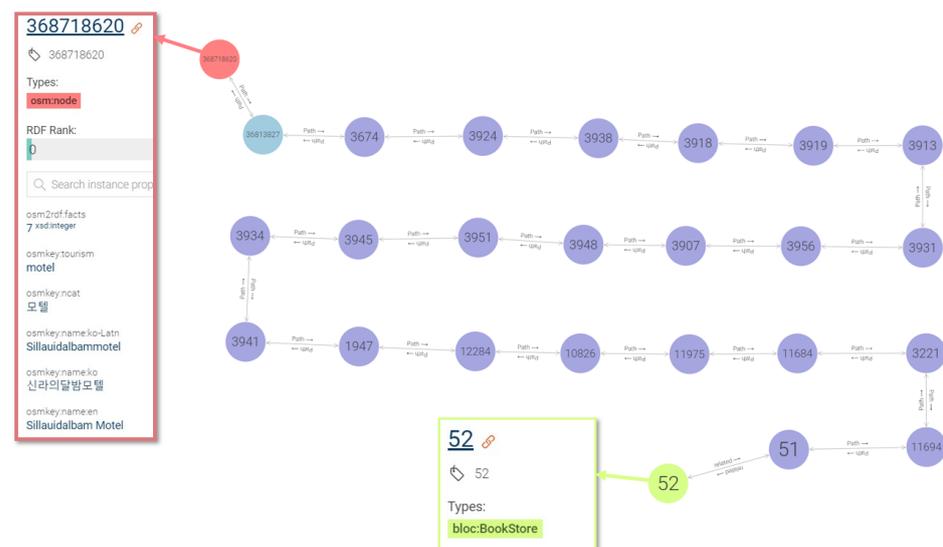**Figure 15.** SPARQL query for pathfinding outdoor from indoor space.

Figure 16. Pathfinding result visualization with node properties.

5. Discussion

This study's approach to transforming 2D spatial information into 3D spatial data and its subsequent conversion into RDF format marks a significant advancement in the field of semantic web and knowledge graph data. By employing IndoorGML as the foundation for 3D spatial data generation and RDF transformation, our methodology has bridged a critical gap in representing indoor spaces within RDF databases. Traditionally, RDF data models have predominantly focused on outdoor spatial information, as seen in platforms like Wikidata and WorldKG. The lack of comprehensive methods to incorporate indoor spatial data posed a substantial limitation, particularly for applications requiring detailed indoor navigation and space utilization information. Our findings demonstrate a successful conversion process, enabling complex indoor spatial details to be queried and utilized effectively.

Compared to existing methods for RDF data representation of spatial information, our study introduces several innovative elements. Previous approaches have lacked the integration of detailed 3D spatial data into RDF models. By introducing a new parser that systematically converts IndoorGML to RDF, this research addressed these limitations. The utilization of namespaces has allowed for a more detailed representation of indoor features, State–Transition relationships, and additional semantics not previously captured. Comparing the number of objects obtained from the original data with the number of RDF triples, beyond simply converting 2D polygons to 3D solids, various classes and predicates that can express the initially introduced semantics were created. This detailed representation is not only a significant technical achievement but also a practical one, as it facilitates more accurate and useful queries within indoor environments. Finally, through the two scenarios presented, practical usability of the generated data in the real world was demonstrated. Furthermore, it was confirmed that the data produced in our research can be linked with outdoor data structured in RDF format.

While our method represents a substantial advancement in RDF data representation of indoor spatial information, it is not without limitations. One of the primary challenges faced in this study was the scale of data generation. Our focus on a single three-story express bus terminal in Jeonju City, South Korea, provides a valuable proof of concept but also highlights the need for scalability in future implementations. Additionally, automatically integrating this indoor spatial data with existing outdoor RDF datasets remains a challenge that needs to be addressed. Looking ahead, we aim to expand our dataset to encompass larger and more diverse indoor environments. This expansion would not only enhance the robustness of our method but also facilitate the integration of indoor and outdoor spatial data, thereby creating a more comprehensive spatial knowledge graph on a large scale. Furthermore, the development of a more sophisticated query system that seamlessly navigates between indoor and outdoor spaces is an essential goal for future research, which would significantly enhance the practical applications of RDF in real-world scenarios.

6. Conclusions

In recent years, there has been a marked interest in establishing RDF data spatial information and its generation, aiming to enrich spatial data for knowledge base system. While large-scale spatial data for outdoor spaces can be produced on a national scale, the generation of data for indoor spaces presents challenges due to the substantial time, manpower, and cost involved. The lack of indoor spatial information in constructing knowledge graphs, which aid in addressing user queries, is a significant gap, especially when considering spatial information RDFs such as WorldKG, Wikidata, and OSMTTL. Consequently, this study sought to augment the richness of knowledge graph data by converting 2D spatial information, typically formulated during a building's construction phase, into 3D spatial information and then transforming this into RDF. Our data pertains to a three-story express bus terminal building in Jeonju City, South Korea. We invented an automated method to transform 2D spatial details of rooms and doors into 3D geometric information, adopting the widely recognized IndoorGML indoor spatial information standard format. Attributes from 2D spatial information, including space names, room numbers, and intended uses, were stored separately, and used during RDF data creation. A newly invented parser was engineered for RDF data generation, probing suitable namespaces for data encapsulated in the IndoorGML format and adopting them for data modification. We introduced the 'loc' namespace for the semantic description of indoor spaces and 'indoorgml-lite' to articulate the State–Transition framework for indoor routing. Furthermore, the 'bloc' namespace was proposed to convey additional semantics. Upon inputting the resulting RDF data into a GraphDB equipped to process such data, it became evident that one could query attributes of each space, access geometric details, and trace indoor pathways, which was impossible before constructing the RDF data from indoor spatial data. The importance of the indoor RDF data became clearer when we combined it with existing outdoor RDF data. In this paper, our main focus was on data design, its automated transformation, and

its application within buildings. We acknowledge that our data is not yet on a large scale. Moving forward, we plan to develop more extensive data sets and merge them with outdoor spatial information. This approach will better address questions about indoor spaces, providing details on POI and specific distances. Our future goals include expanding the dataset and developing a system tailored to respond to queries about spatial information with knowledge graph.

Author Contributions: Conceptualization, Hanme Jang; methodology, Hanme Jang; software, Hanme Jang; validation, Hanme Jang; formal analysis, Hanme Jang and Jiyoung Kim; investigation, Hanme Jang; resources, Jiyoung Kim; data curation, Hanme Jang and Jiyoung Kim; writing—original draft preparation, Hanme Jang; writing—review and editing, Hanme Jang and Jiyoung Kim; visualization, Hanme Jang; supervision, Jiyoung Kim; project administration, Kiyun Yu; funding acquisition, Kiyun Yu and Jiyoung Kim All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; Taylor, J. Freebase: A collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, Vancouver, BC, Canada, 9–12 June 2008; pp. 1247–1250.
2. Vrandečić, D.; Krötzsch, M. Wikidata: A free collaborative knowledgebase. *Commun. ACM* **2014**, *57*, 78–85. [CrossRef]
3. Suchanek, F.M.; Kasneci, G.; Weikum, G. Yago: A core of semantic knowledge. In Proceedings of the 16th International Conference on World Wide Web, Banff, AB, Canada, 8–12 May 2007; pp. 697–706.
4. Shu, Y.; Yu, Z.; Li, Y.; Karlsson, B.F.; Ma, T.; Qu, Y.; Lin, C.Y. Tiara: Multi-grained retrieval for robust question answering over large knowledge bases. *arXiv* **2022**, arXiv:2210.12925.
5. Ye, X.; Yavuz, S.; Hashimoto, K.; Zhou, Y.; Xiong, C. Rng-kbqa: Generation augmented iterative ranking for knowledge base question answering. *arXiv* **2021**, arXiv:2109.08678.
6. Gu, Y.; Su, Y. Arcaneqa: Dynamic program induction and contextualized encoding for knowledge base question answering. *arXiv* **2022**, arXiv:2204.08109.
7. Chen, S.; Liu, Q.; Yu, Z.; Lin, C.Y.; Lou, J.G.; Jiang, F. ReTraCk: A flexible and efficient framework for knowledge base question answering. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations, Online, 1–6 August 2021; pp. 325–336.
8. Zhou, J.; Xu, W. End-to-end learning of semantic role labeling using recurrent neural networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; pp. 1127–1137.
9. Apeh, M.M.; Abdul Rahman, A. A review of 3D spatial data models for surface and subsurface integration of spatial objects. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *48*, 15–23. [CrossRef]
10. Auer, S.; Lehmann, J.; Hellmann, S. Linkedgeodata: Adding a spatial dimension to the web of data. In *The Semantic Web-ISWC 2009*; Springer Berlin Heidelberg: Chantilly, VA, USA, 2009; pp. 731–746.
11. Punjani, D.; Singh, K.; Both, A.; Koubarakis, M.; Angelidis, I.; Bereta, K.; Stamoulis, G. Template-based question answering over linked geospatial data. In Proceedings of the 12th Workshop on Geographic Information Retrieval, Seattle, WA, USA, 6 November 2018; pp. 1–10.
12. Dsouza, A.; Tempelmeier, N.; Yu, R.; Gottschalk, S.; Demidova, E. Worldkg: A world-scale geographic knowledge graph. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Virtual, 1–5 November 2021; pp. 4475–4484.
13. Vinasco-Alvarez, D.; Samuel, J.S.; Servigne, S.; Gesquière, G. From Citygml to Owl. Ph.D. Thesis, École Centrale de Lyon, Ecully, France, 2020.
14. Hamzei, E.; Tomko, M.; Winter, S. Translating place-related questions to GeoSPARQL queries. In Proceedings of the ACM Web Conference, Online, 25–29 April 2022; pp. 902–911.

15. Homburg, T.; Staab, S.; Janke, D. Geosparql+: Syntax, semantics and system for integrated querying of graph, raster and vector data. In *The Semantic Web–ISWC 2020*; Springer International Publishing: Athens, Greece, 2020; pp. 258–275.

16. Akroyd, J.; Mosbach, S.; Bhave, A.; Kraft, M. Universal Digital Twin-A Dynamic Knowledge Graph. *Data-Centric Eng.* **2021**, *2*, e14. [CrossRef]

17. Basiri, A.; Lohan, E.S.; Moore, T.; Winstanley, A.; Peltola, P.; Hill, C.; Amirian, P.; Silva, P.F.E. Indoor location based services challenges, requirements and usability of current solutions. *Comput. Sci. Rev.* **2017**, *24*, 1–12. [CrossRef]

18. Chen, L.; Nugent, C.D.; Wang, H. A knowledge-driven approach to activity recognition in smart homes. *IEEE Trans. Knowl. Data Eng.* **2011**, *24*, 961–974. [CrossRef]

19. Fensel, A.; Tomic, S.; Kumar, V.; Stefanovic, M.; Aleshin, S.V.; Novikov, D.O. Sesame-s: Semantic smart home system for energy efficiency. *Inform. Spektrum* **2013**, *36*, 46–57. [CrossRef]

20. Jang, H.; Yu, K.; Yang, J. Indoor reconstruction from floorplan images with a deep learning approach. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 65. [CrossRef]

21. Teo, T.A.; Cho, K.H. BIM-oriented indoor network model for indoor and outdoor combined route planning. *Adv. Eng. Inform.* **2016**, *30*, 268–282. [CrossRef]

22. Gunduz, M.; Isikdag, U.; Basaraner, M. A review of recent research in indoor modelling & mapping. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *41*, 289–294.

23. Kim, Y.J.; Kang, H.Y.; Lee, J. Development of indoor spatial data model using CityGML ADE. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *40*, 41–45. [CrossRef]

24. Kang, H.K.; Li, K.J. A standard indoor spatial data model—OGC IndoorGML and implementation approaches. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 116. [CrossRef]

25. Srivastava, S.; Maheshwari, N.; Rajan, K.S. Towards generating semantically-rich indoorgml data from architectural plans. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, 591–595. [CrossRef]

26. Mirvahabi, S.S.; Abbaspour, R.A. Automatic extraction of IndoorGML core model from OpenStreetMap. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 459–462. [CrossRef]

27. Tessema, L.S.; Jäger, R.E.; Stilla, U. Extraction of IndoorGML Model from an Occupancy Grid Map Constructed Using 2D LiDAR. In Proceedings of the German Society for Photogrammetry, Remote Sensing and Geoinformation, Vienna, Austria, 20–22 February 2019.

28. Berners-Lee, T. Linked Data-Design Issues. Available online: http://www.w3.org/DesignIssues/LinkedData.html (accessed on 12 October 2023).

29. Battle, R.; Kolas, D. Geosparql: Enabling a geospatial semantic web. *Semant. Web J.* **2011**, *3*, 355–370. [CrossRef]

30. Pauwels, P.; Van Deursen, D. IFC-to-RDF: Adaptation, aggregation and enrichment. In Proceedings of the First International Workshop on Linked Data in Architecture and Construction, Ghent, Belgium, 28–29 March 2012; pp. 1–3.

31. Matuszka, T.; Gombos, G.; Kiss, A. A new approach for indoor navigation using semantic web technologies and augmented reality. In Proceedings of the Virtual Augmented and Mixed Reality Designing and Developing Augmented and Virtual Environments: 5th International Conference, VAMR 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, 21–26 July 2013; Proceedings, Part I 5. Springer: Berlin/Heidelberg, Germany, 2013; pp. 202–210.

32. Fleiner, R.; Szász, B.; Piros, P. Indoor navigation Linked Data at Obuda University. In Proceedings of the 2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 12–14 May 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 25–30.

33. Simon-Nagy, G.; Fleiner, R. Ontology Extension for Personalized Accessible Indoor Navigation. In Proceedings of the 16th International Conference on Global Research and Education Inter-Academia, Iași, Romania, 25–28 September 2017; Springer: Cham, Switzerland, 2018; pp. 281–288.

34. Chalhoub, N.; Simon-Nagy, G. Indoor Navigation Based on Linked Data at Honvéd Hospital, Budapest. In Proceedings of the 2018 IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 17–19 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5.

35. Li, K.J.; Zlatanova, S.; Torres-Sospedra, J.; Pérez-Navarro, A.; Laoudias, C.; Moreira, A. Survey on indoor map standards and formats. In Proceedings of the 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Pisa, Italy, 30 September–3 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–8.

36. Chen, J.; Clarke, K.C. Modeling standards and file formats for indoor mapping. In Proceedings of the International Conference on Geographical Information Systems Theory, Applications and Management, Online, 23–25 April 2017; SciTePress: Setúbal, Portugal, 2017; Volume 2, pp. 268–275.

37. Yang, C.T.; Chen, S.T.; Den, W.; Wang, Y.T.; Kristiani, E. Implementation of an Intelligent Indoor Environmental Monitoring and Management System in Cloud. *Future Gener. Comput. Syst.* **2019**, *96*, 731–749. [CrossRef]

38. IMDF. Available online: https://docs.ogc.org/cs/20-094/index.html (accessed on 15 October 2023).

39. Goetz, M. Using crowdsourced indoor geodata for the creation of a three-dimensional indoor routing web application. *Future Internet* **2012**, *4*, 575–591. [CrossRef]

40. CityGML. Available online: https://www.ogc.org/standard/citygml/ (accessed on 15 October 2023).

41. IndoorGML. Available online: https://www.indoorgml.net/ (accessed on 15 October 2023).

42. Deng, Y.; Ai, H.; Deng, Z.; Gao, W.; Shang, J. An Overview of Indoor Positioning and Mapping Technology Standards. *Standards* **2022**, *2*, 157–183. [CrossRef]

43. Kyzirakos, K.; Savva, D.; Vlachopoulos, I.; Vasileiou, A.; Karalis, N.; Koubarakis, M.; Manegold, S. GeoTriples: Transforming Geospatial Data into RDF Graphs Using R2RML and RML Mappings. *J. Web Semant.* **2018**, *52*, 16–32. [CrossRef]

44. Huang, C.Y.; Chiang, Y.H.; Tsai, F. An Ontology Integrating the Open Standards of City Models and Internet of Things for Smart-City Applications. *IEEE Internet Things J.* **2022**, *9*, 20444–20457. [CrossRef]

45. Available online: https://www.w3.org/OWL/ (accessed on 17 October 2023).

46. Ledoux, H.; Arroyo Ohori, K.; Kumar, K.; Dukai, B.; Labetski, A.; Vitalis, S. CityJSON: A Compact and Easy-to-Use Encoding of the CityGML Data Model. *Open Geospat. Data Softw. Stand.* **2019**, *4*, 1–12. [CrossRef]

47. Available online: https://github.com/tudelft3d/indoorjson (accessed on 17 October 2023).

48. Teo, T.A.; Yu, S.C. The extraction of indoor building information from BIM to OGC IndoorGML. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 167–170. [CrossRef]

49. Adouane, K.; Stouffs, R.; Janssen, P.; Domer, B. A model-based approach to convert a building BIM-IFC data set model into CityGML. *J. Spat. Sci.* **2020**, *65*, 257–280. [CrossRef]

50. Bast, H.; Brosi, P.; Kalmbach, J.; Lehmann, A. An Efficient RDF Converter and SPARQL Endpoint for the Complete OpenStreetMap Data. In Proceedings of the 29th International Conference on Advances in Geographic Information Systems, Beijing, China, 2–5 November 2021; pp. 536–539.