

Article

Semantic Fusion with Deep Learning and Formal Ontologies for Evaluation of Policies and Initiatives in the Smart City Domain

Nil Kilicay-Ergin ^{*,†} and Adrian S. Barb [†]

Engineering Department, Great Valley Campus, Penn State University, 30 E Swedesford Rd, Malvern, PA 19335, USA; adrian@psu.edu

* Correspondence: nhe2@psu.edu; Tel.: +1-610-648-3200

† These authors contributed equally to this work.

Abstract: Decision makers and policy analysts at different administrative levels often lack a holistic view of the problem as there are semantic variations in policy documents due to domain-specific content. For example, smart city initiatives are derived from national and international initiatives which may influence the incentives for local participants, but local initiatives reflect the local contextual elements of the city. Balanced assessment of smart city initiatives should include a systemic evaluation of the initiatives at multiple levels including the city, the country in which the city resides as well as at international level. In this paper, a knowledge elicitation methodology is presented for multi-granularity evaluation of policies and initiatives. The methodology is demonstrated on the evaluation of smart city initiatives generated at different administrative levels. Semantic networks are constructed using formal ontologies and deep learning methods for automatic semantic evaluation of initiatives to abstract knowledge found in text. Three smart city initiatives published by different administrative levels including international, national, and city level are evaluated in terms of relevance, coherence, and alignment of multi-level smart city initiatives. Experiments and analysis ultimately provide a holistic view of the problem which is necessary for decision makers and policy analysts of smart cities.

Keywords: multi-level initiatives; policy context; knowledge elicitation; natural language processing; semantic fusion; deep learning; smart city



Citation: Kilicay-Ergin, N.; Barb, A.S. Semantic Fusion with Deep Learning and Formal Ontologies for Evaluation of Policies and Initiatives in the Smart City Domain. *Appl. Sci.* **2021**, *11*, 10037. <https://doi.org/10.3390/app112110037>

Academic Editor: Luis Javier Garcia Villalba

Received: 16 September 2021
Accepted: 12 October 2021
Published: 26 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Policies and initiatives are generated by administrations at different levels in text format and capture various perspectives of the problem context. While policies are designed with the aim to align with upper and lower administrative levels, evaluating multi-level and multi-stakeholder generated policies is challenging because there is ambiguity and variability in the language of free text. In addition, different administrative levels reflect their own domain knowledge and local context. Terms used in one domain might have varying meanings in other contexts. As a result, initiatives, and policies across different levels of administration may not align well leading to a fragmented view of the problem and with possibility of gaps across policies and initiatives. While this a challenge for any multi-domain and multi-level text-based policy, one application area where a balanced assessment of policies and initiatives may be beneficial is smart cities.

Smart city initiatives are proliferating across the world, transforming these initiatives into practice is not straight forward. Studies indicate slow diffusion of smart city initiatives due to several obstacles [1]. Cities are system of systems comprised of transportation network, communication network, infrastructure networks, interconnected citizens, and organizations. Linking initiatives to specific systems within a city or socio-economic issue is difficult due to the complex interrelations among these systems within a city. Balanced assessment of smart city initiatives necessitates an evaluation of the initiatives at multiple levels including solution providers involved with the initiative, organizations

that indirectly benefit from the initiative, the city as a whole, other cities, the country as well as at international level. Thus, complexity stems from the multi-stakeholder nature of cities where multi-level of agencies, perspectives, private and public sectors are involved. This multiplicity of perspectives is evident in the wide range of smart city initiatives across the world which cover a wide range of contexts and situations. For example, smart city initiatives in Barcelona aim to transform the city using advanced digital technologies with a focus on creating a sustainable city and innovative commerce ecosystem [2]. Initiatives in Amsterdam focus on addressing climate change using smart infrastructure technologies. Other smart city initiatives prioritize design of technologies to improve participation of its citizens [3].

Assessment of initiatives is challenging since stakeholders have varying and, in some cases, conflicting objectives. Besides some of the initiatives are derived from national and international levels which may influence the incentives for local participants but do not reflect the local contextual elements of the initiative. Both local and central governments lack a consistent and complete plan for balanced assessment of initiatives and policies for smart cities. Decision makers need methods and tools to assess the broad impact of smart city initiatives and policies. They also need to understand the differences and synergies among multi-domain, multi-level contextual perspectives among initiatives so that gaps and strengths of the initiatives can be identified, and initiatives can be aligned with national and international policy objectives.

The open and complex nature of the problem necessitates knowledge elicitation methodologies that promote holistic view for evaluating multi-level initiatives and policies [4,5]. Formalized semantic networks provides means to structurally represent text [6]. These formalized semantic networks can form a holistic view of knowledge which decision makers and analysts at different levels of administrations can utilize to analyze information. However, generating a formalized semantic network is challenging due to the complexity, ambiguity, and variability of text. Ontologies provide conceptual models of a domain and can be used to reduce complexity and ambiguity in text [6]. Ontologies along with semantic networks can be utilized to construct a formalized semantic network which can be used in automatic semantic evaluation of text. Formal ontologies are also useful in information comprehension especially where the models generated are too large for the analyst to manage or understand. It can reduce text complexity in the area of planning [7] by simplifying and abstracting the text. This abstraction can be applied to complex and adaptive systems in which human knowledge and technologies are interacting [8], such as the area of smart cities. This can be accomplished by revealing the main concepts that are used in a text that experts can use to determine and understand the underlying issues existent in the text and hint to possible solutions to address them.

The usage of formal ontologies requires identification of semantic similarity between in the target text and concepts in the ontology. Traditional methods such as lexical approaches [9] have limited capabilities in this area. Recently, deep learning approaches have obtained high quality results in evaluating natural language similarity. They accomplish this by training models on large corpora of text to generate a high-dimensional vector space in which semantically similar words have closer vector representations.

In this paper, a knowledge elicitation methodology is presented for automated evaluation of text-based policies/initiatives generated by multiple-stakeholders at different administrative levels. Semantic networks are constructed on formal ontologies using semantic fusion methods based on the Word2Vec [10] deep learning techniques for automatic evaluation of knowledge at different abstraction levels. The methodology is demonstrated on smart city initiatives. Three smart city initiatives published by different administrative levels including international, national, and city level are evaluated in terms of relevance, coherence, and alignment of multi-level smart city initiatives. Experiments and analysis ultimately provide a holistic view of the problem to support for decision makers and policy analysts of smart cities in their quest for effective planning of smart city initiatives and policies. The rest of the paper is organized so that Section 2 reviews related work

on smart city initiative studies, text-based analysis methods for smart cities, and natural language processing techniques. Section 3 describes, in detail, the knowledge elicitation methodology for the automatic evaluation of text. An experimental evaluation and analysis of multi-level smart city initiatives are provided in Section 4. Section 5 concludes the paper and provides future directions for the research.

2. Related Work

Earlier studies of smart city initiatives focused on developing a framework to systematically understand smart city initiatives based on review of smart city literature. Smart city initiatives framework in [11] identified a set of factors including technology, management and organization, policy, governance, people and communities, economy, built infrastructure, and natural environment as the fundamental elements of smart city initiatives. In a follow up empirical study [12], this smart city initiatives framework is used to assess smart city initiatives of different cities including Philadelphia, Seattle, Quebec City, and Mexico City. Recent study on smart city initiatives [13] focus on country specific factors and study how characteristics unique to a country may impact the development of smart city initiatives. A comparative analysis between US and China smart city initiatives revealed the impact of government system structure on smart city initiatives. A Smart City Initiative Design (SCID) Framework is constructed in [14] as a framework to support design of smart city initiatives. SCID framework provided common objectives for smart city initiatives, strategies and enabling factors for successful initiative design, and core challenges to be addressed. European Union report [15] analyzed smart city initiatives across European Union to understand the potential impact of the initiatives on European Union 2020 objectives as well as alignment of the initiatives to European Union 2020 objectives. Study emphasized the need for designing smart city initiatives that are based on specific, measurable, achievable, realistic, and time dependent (SMART) objectives and that are aligned with city development plans and European Union 2020 objectives. Another study of smart city initiative [1] constructed a regression analysis to understand the relationship between smart city contextual variables and coverage index which considers number of application domains a city has launched initiatives. Results of the study revealed the importance of understanding the local context as there is no global definition of smart city. Other studies also emphasized the need to understand local contextual factors and integrate disparate and ambiguous information [16,17].

Smart city initiatives are not designed in isolation from the states, regions, governments in which they reside. Higher level administrative levels may impose incentives or constraints on the local cities, so cities need to understand how smart city initiatives align with higher level administrations. Multi-level analysis of initiatives necessitates evaluation of contextual characteristics of knowledge representation [18] at different granularity levels. This requires establishing consensus among decision makers which in a community is knowledge based [19]. Formalization [20] and abstraction [21] of knowledge are some of the key methods for establishing consensus in a community. Ontologies can be used for formalization and abstraction of text [22] to represent, exchange, and interpret knowledge as well as compare knowledge structures from different data sources and formats [6].

There are several classification schemes for linguistic content including Levina's classification of English verbs [23], WordNet semantic network classification of nouns, verbs, adjectives, and adverbs [24]. There are several formal ontologies including DOLCE [25], BFO [26], and SUMO [27]. Among these formal ontologies, The Suggested Upper Merged Ontology (SUMO) provides greater variety and completeness of definitions and has been mapped to WordNet which provides the capability to convert English words to formal terms. Therefore, SUMO ontology can be utilized for reducing ambiguity in text documents using formal terms [28].

Knowledge extraction in smart city domain is challenging due to variations among semantics of data. Several studies tackled knowledge extraction of text from smart city documents to design smart city applications and services. A knowledge-based model

for smart city services mapped static and dynamic data to smart city ontology in [29]. Automatic knowledge extraction of non-taxonomic relations from smart city documents is demonstrated in [30] by combining semantic graph-based method with context information. Formal concept lattices are constructed for smart city problems [31] to extract knowledge that can be utilized for smart city services.

To summarize, previous studies on smart city initiatives focused on understanding the common elements of smart city initiatives rather than evaluating smart city initiatives from a multi-level granularity perspective. Smart city initiatives are embedded in higher level initiatives and policies including country level and international initiatives. It is important to analyze the initiatives at different levels of abstraction to improve alignment of initiatives to higher level objectives.

The objectives of this paper are (1) to provide a knowledge elicitation methodology for multi-granular analysis of text-based policies and initiatives generated by different administrative levels. Such an automated method would support decision makers and analysts in balanced assessment of policies pointing to gaps, and more effective alignment of policies at higher levels, (2) to demonstrate the methodology in evaluation of smart city initiatives generated from multi-level administrative levels to understand alignment and variations among the initiatives. Main methodology utilized in previous research on smart city initiatives is based on use of surveys, interviews, and comparative study of case studies of various smart city initiatives across the world. A recent study of smart city initiative analysis [13] utilized content analysis and text mining to analyze smart city initiatives but the methodology did not focus on use of formalized methods. While there are studies of text document analysis for smart cities utilizing formalized methods, these studies aim to extract knowledge for smart city applications and services rather than use the methodology for analysis of smart city initiatives. This study provides a formalized methodology to evaluate smart city initiatives and differs from other studies in terms of constructing a formalized semantic network structure of smart city initiatives to establish a common framework for analysis of initiatives written at different administrative levels including international, national, and regional.

3. Methodology

Knowledge-based text evaluation is difficult in domain specific areas due to the fact that high accuracy results can be achieved using domain-specific dictionaries that require intensive data engineering methodologies. Even if such domain-specific dictionaries are produced, resulting system have limited success due to the fact that they lack scalability. Scalability is an important characteristic for systems that are comprised of multi domains such as smart city. In the smart city domain, there is a high degree of interaction between providers and citizens leading to complex methods of information delivering [32]. The documents of smart city such as initiatives and policies must address transparency and accountability for all the stakeholders as well as support engagement of its citizens in the process.

In this article we propose a methodology that uses semantic fusion to reduce the uncertainty existent in text by combining general usage and domain specific dictionaries. The methodology is a quantitative method for formal evaluation of text by using the principles of ontological formalization and abstraction. By applying these principles, we can evaluate the similarity and differences among initiatives by associating different concepts to specific areas of the smart city topics. For example, a sentence like “US will invest in new grants to build a research infrastructure for smart cities.” should be more relevant to the areas of infrastructure than to the areas of environment. The methodology is summarized in Figure 1. Our algorithm includes two stages: preprocessing and knowledge extraction. In the preprocessing stage, we generate two dictionaries from two large corpora of data: a general English and a domain specific corpus. In the knowledge extraction stage, we build a semantic map for each document using the previously generated dictionaries as

well as the SUMO ontology and other linguistic resources. The algorithm will be explained in detail in the next sub-sections.

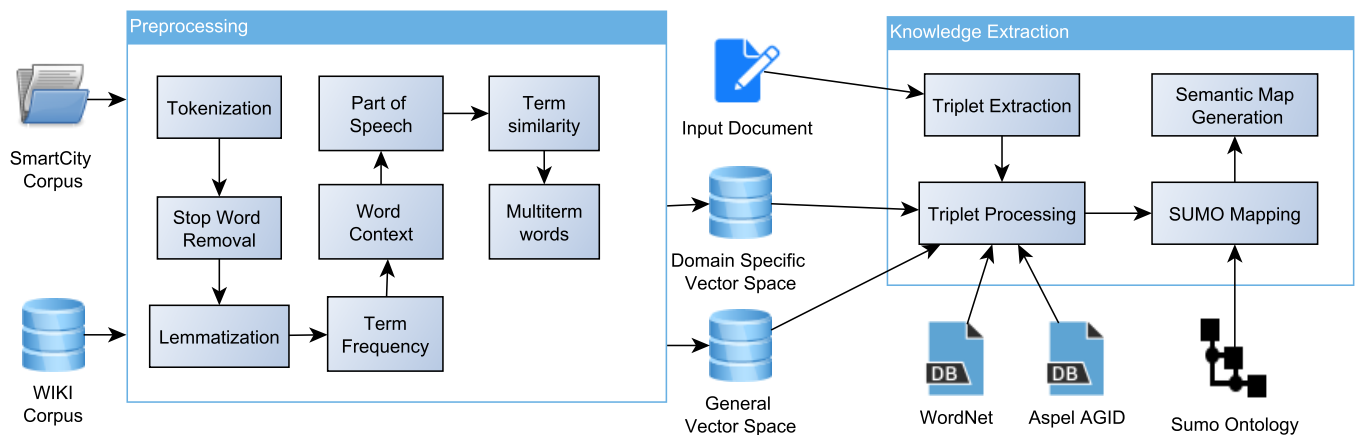


Figure 1. Flow of Knowledge Extraction from Text. Before we can extract the knowledge from the text, we need to generate some domain-specific text statistics. These statistics are used the generation of triplets and of the knowledge maps.

3.1. Preprocessing

The first step in our preprocessing procedure is to generate two dictionaries: one for general English language usage and one for domain-specific usage. For the general language dictionary, we have used the Wikipedia Corpus [33]. For the domain specific dictionary, we used twelve textbooks related to the following areas of smart city: governance, sustainability, information technology, urban development, principles, applications, and solutions. For both indexes, we have processed the document using natural language techniques to match the word form from the documents to be analyzed. Among these operations, we have used tokenization, stop word removal, lemmatization, and part of speech extraction. These operations will be further explained in Section 3.2.

The generated words were stored into a word embedding index [10] which was shown to better identify the semantic similarity between words from semantic relatedness [34]. Word embedding approaches make the assumption that semantic similarity of two words is directly proportional to the probability that the words appear in the same context [35]. In such indexes, similarity between two words is computed by counting the co-occurring words using a moving window of fixed size and using a weighted measure based on the position of the word in the window. The similarity is calculated using the cosine method [36].

Further, for text simplifications we have converted adjectives and adverbs to their noun forms. For example, in a text “new grants” we convert the adjective “new” to its noun form “novelty” due to the fact that the desired triplet form is “grant has property novelty”. For this we used the Automatically Generated inflection Database [37], to remove any word inflections and WordNet [24] to determine the noun form of the adjective.

3.2. Knowledge Extraction

Once we have all the necessary systems in place, we process each target document as shown in Figure 1. First, we extract the relevant triplets from the text. Triplet extraction is performed using the Stanford Core NLP package [38]. Next, we process each triplet by identifying instances and compound words and converting triplet parts to their noun/verb forms. We further map each noun in a triplet into the SUMO ontology [27] using deep learning-based similarity and finally, generate a semantic map for each document.

3.2.1. Triplet Extraction

For each document, we extract the relevant triplets using the open information extraction algorithm [39] that is included in the Stanford NLP package [38]. Before triplets are extracted, we identify and replace named entities using a named entity recognition algorithm (NER) [40]. Example of entities that we replace are countries, cities, organizations, currency, and custom entities that are specific to the smart city subject. We also identify compound words, adjectives, adverbs, phrase modifiers existent in each sentence, and other triplets. For example, the sentence *“The City of Philadelphia established SmartCityPHL in 2017 to better understand and implement smart technology solutions that would improve city service delivery”* contains some instances that have to be replaced before triplets are extracted. In this sentence we find a compound word *“City of Philadelphia”* that will be replaced with *“city”*. Also, *“SmartCityPHL”* is a customized instance and will be replaced with *“initiative”*, while *“2017”* will be replaced with *“year”*. The resulting sentence is *“The City established initiative in year to better understand and implement smart technology solutions that would improve city service delivery”*. This sentence will be used as input for triplet extraction.

Table 1 shows the resulting triplets that were extracted by the Stanford CoreNLP package from the sentence above. As seen from this table, there are six triplets that were extracted but the information in these triplets is repetitive. For example, there are only two distinct actions across the six triplets: *“understand”* and *“establish”*. Considering this, triplets 3 and 4 are very similar and should be reduced. Second issue is the fact that some parts of the sentence did not generate triplets. For example, the *“that would improve city service delivery”* is not represented in the triplet list. Also, some parts of the generated triplets contain multiple words. The next phase of triplet extraction will focus on generating triplets that contain single words for each part.

Table 1. Example triplets generated from a sentence using the Open IE algorithm in the Stanford CoreNLP package.

Cnt	Subject	Action	Object
1	City	established initiative in	year
2	City	understand	technology solutions
3	City	better understand	smart technology solutions
4	City	established	initiative
5	City	understand	smart technology solutions
6	City	better understands	technology solutions

For each generated triplet that has multiple words in one of its parts we look into the dependency tree. For example, for the triplet 3, *“City better understand smart technology solutions”* the dependency tree is shown in Figure 2. By navigating each part of the triplet, we can reduce each triplet part to an atomic term. For example, the action *“better understand”*, can be reduced to *“understand”*, if we add another triplet with the subject *“understand”*, the action *“has property”*, and the object *“better”*. Similarly, the object can be reduced to *“solutions”* while adding two more triplets: *“solution have property smart”* and *“solution relate to technology”*.

Further we look at conjunctions that we find in the text. For example, the terms *“understand”* and *“implement”* are connected by the conjunction *“and”*. For each of the terms connected by a conjunction we will add a new triplet. One example of such new addition is triplet 2.3 (*“city implement solution”*) in Table 2.

The last step in triplet generation is to evaluate parts of the sentence that were not used in the triplets generated by Open IE algorithm. In this sentence, the part *“that would improve city service delivery”* was not included in any triplet generated by the Open IE algorithm. In such cases, we look for subject-object dependencies to generate new triplets. For example Figure 2 shows such a dependency for the verb *“improve”* with the subject *“solution”* and the object *“delivery”*. This triplet will be added to the list of triplets and it is

shown in Table 2 at position 7. The final list of generated triplets is shown in Table 2 and includes the simplification of the object for triplet 7.

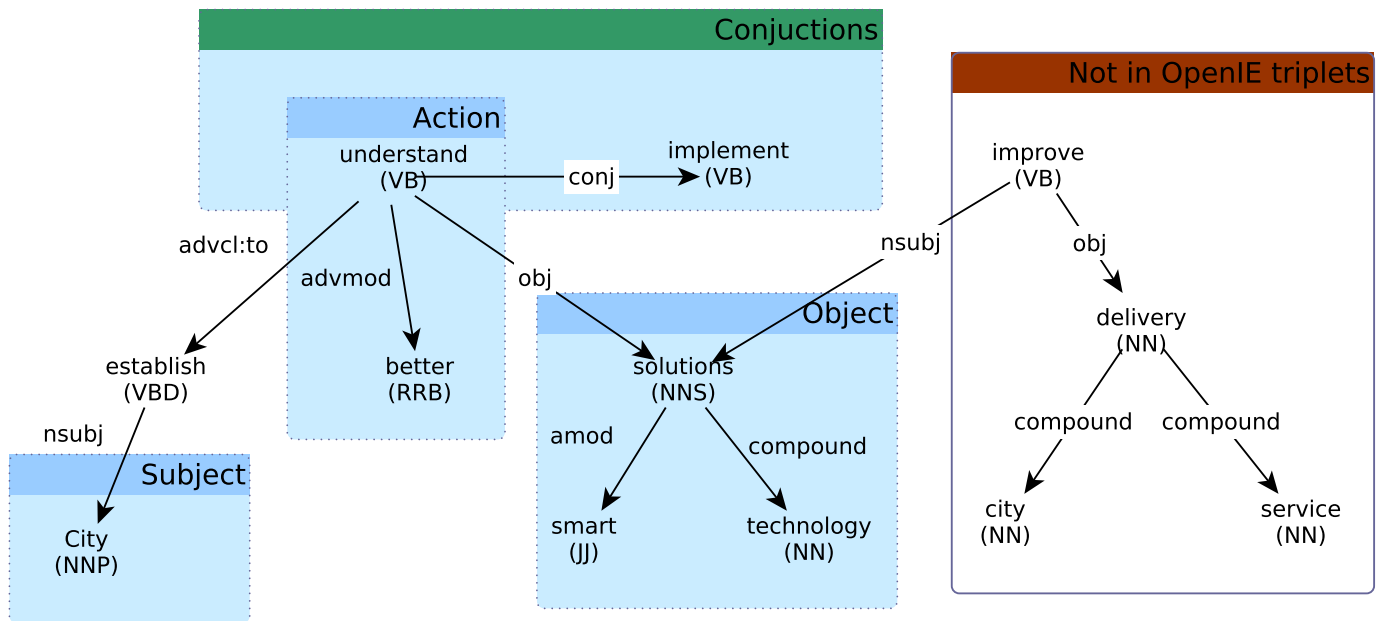


Figure 2. Dependency tree for terms in the triplet “City better understand smart technology solutions”.

Table 2. Example triplets generated after processing the dependency tree using the Stanford CoreNLP package. Duplicates were merged.

Cnt	Subject	Action	Object
1.1	city	establish	initiative
1.2	initiative	relate to	year
2.1	city	understand	solution
2.2	solutions	relate to	technology
2.3	city	implement	solution
3.2	understand	has property	better
3.3	solution	has property	smart
3.2	implements	has property	better
4	city	establish	initiative
7	solution	improve	delivery
8	delivery	relate to	city
9	delivery	relate to	service

3.2.2. Triplet Preprocessing

The next stage is to apply more preprocessing to all the components of a triplet as shown in Figure 1. Our preprocessing consists of three steps: (1) abstraction of triplet action, (2) reducing subjects and objects to their dictionary base, and (3) determining the noun form of subjects and objects.

To represent the information in each document, we create a knowledge map in which nouns are nodes and verbs are edges. However, it is necessary to simplify the map by abstracting the relationship edges formed from verbs. The abstraction of a triplet action is performed using the WordNet lexical classes [24]. WordNet organizes concepts into forty-five lexicographer files based on syntactic category and logical groupings of which 15 apply to verbs. For example, the action in the triplet “City establish initiative” describes a creation activity and hence action type has a “creation” class. On the other hand, the action in the “city understand solution” triplet describes a cognition relationship.

Next, we transform each subject and object to its dictionary noun form. For this we use the Aspell Automatically Generated Inflection Database (AGID) [37]. AGID is part of the Spell Checker Oriented Word Lists (SCOWL) project that is intended to be suitable for use in spell checkers [37]. Using this database, we can remove all the inflections from subjects and objects. For example, by removing the inflection from the subject “understands” will result in “understand”. We further notice that this subject is a verb and not a noun so we will further convert it to its noun form “understanding”. The result of this preprocessing is shown in Table 3.

Table 3. Example triplets generated after applying the preprocessing techniques explained in Section 3.2.2.

Cnt	Type	Term	SUMO Concept
1.1	Subject	city	City
	Action	creation	
	Object	initiative	Content Development
1.1	Subject	city	City
	Action	creation	
	Object	initiative	Content Development
1.2	Subject	initiative	Content Development
	Action	state	
	Object	year	Year Duration
2.1	Subject	city	City
	Action	cognition	
	Object	solution	Procedure
2.2	Subject	solution	Procedure
	Action	state	
	Object	technology	Engineering
2.3	Subject	city	City
	Action	change	
	Object	solution	Procedure
3.2	Subject	understanding	Interpreting
	Action	state	
	Object	quality	Subjective Assessment Attribute
3.3	Subject	solution	Procedure
	Action	state	
	Object	smart	Subjective Strong Positive Attribute
3.2	Subject	implementation	Intentional Process
	Action	state	
	Object	quality	Subjective Assessment Attribute
4	Subject	city	City
	Action	cognition	
	Object	initiative	Content Development
7	Subject	solution	Procedure
	Action	change	
	Object	delivery	Giving
8	Subject	delivery	Giving
	Action	state	
	Object	city	City
9	Subject	delivery	Giving
	Action	state	
	Object	service	Service Process

3.2.3. Concept Mapping into the SUMO Ontology

Finally, mapping concepts into the SUMO ontology is performed using the WordNet to SUMO database provided by [27] which is part of the SUMO project. The results are shown in Table 3 columns three and six. To map the WordNet terms into SUMO, we use the following relation types: equivalence, instance, and subsumption. For example, the WordNet term “service” is equivalent to the SUMO concept “Service Process” while the term “year” is equivalent to the SUMO concept “Year Duration”. At the same time, “year” is subsumed by the SUMO concept “Time Duration”. In such cases, equivalence takes priority over subsumption.

There are some cases where a WordNet term is not mapped into a SUMO concept. In such cases we perform query expansion for the term to identify synonyms and entailments of the word provided in the original text. Research show that query expansion improves knowledge discovery in text [41]. We perform query expansion using both knowledge-based and deep learning methods [10]. For the knowledge-based methods, we use the synonym and entailment facilities of Wordnet while for the deep leaning methods we word similarity in the domain-specific Word2Vec high-dimensional space. Each expanded term added to the concept map will have a relevance probability proportional with its degree of similarity to the original word. We further evaluate the relevance of each expanded term using a Couquet Integral method to identify the terms that are relevant in both general and domain specific indexes. For more information on query expansion, the reader is directed to read [42]. Finally, we map the expanded term set into the SUMO ontology.

3.2.4. Semantic Map Generation, Aggregation, and Reduction

Having each triplet extracted and preprocessed will enable us to generate a semantic map for text. Figure 3 shows the semantic network that was generated from the sentence in our example. We use the SUMO concepts generated for subjects and objects as nodes in the network and the WordNet lexical categories as edges. For example, the node “interpreting” was generated by mapping the word “understand” into the SUMO ontology. This node is connected to the node “Subjective assessment Attribute” that was generated from the word “better”. Processing this sentence returned an unconnected semantic map. The three nodes on the left are not connected to the rest of the map. In such cases we can use the SUMO ontology to connect the network. In this case, we can notice that “Subjective strong positive attribute” is an instance of “Subjective Assignment Attribute”. This relation is represented in Figure 3 with dotted line.

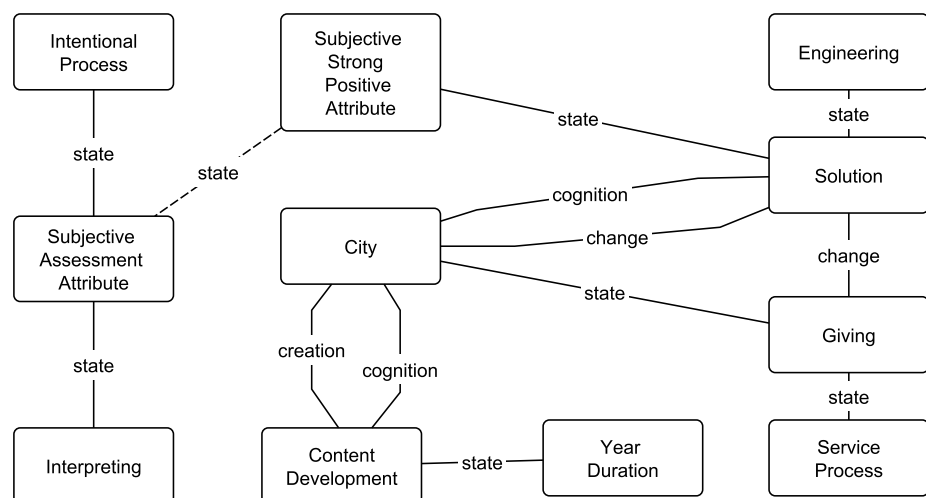


Figure 3. Semantic network generated for the sentence shown in Section 3.2.1. Each noun in the sentence was processed and mapped into a SUMO concept while verbs were used to connect SUMO concepts.

Figure 4 shows an example of mapping WordNet words into the hierarchical structure of the SUMO ontology by navigating the subclass/superclass relationships. For example, the word *city* is mapped into the *City* SUMO concept and can be further abstracted to *Geopolitical Area* or *Region*, for example. Due to the fact that terms in a sentence are abstracted to SUMO concepts, it is possible that, in a sentence, same triplet appear twice. In this case, the triplet relevance is normalized to the [0, 1] interval and the duplication is removed. However, because each document contains several sentences, triplets may appear several times in different sentences of a document. In this case, the relevance of a triplet to the document is aggregated using a weighted average function as shown in the equation below.

$$R_{triplet} = \sum_{triplet_i} R_{triplet_i} / 2^i | R_{triplet_i} > R_{triplet_{i+1}} \tag{1}$$

In this equation, we first sort the instances $triplet_i$ of the same triplet $triplet$ by its relevance $R_{triplet_i}$ in a descending order. Each of these instances will be weighted proportionally to its index in the sequence.

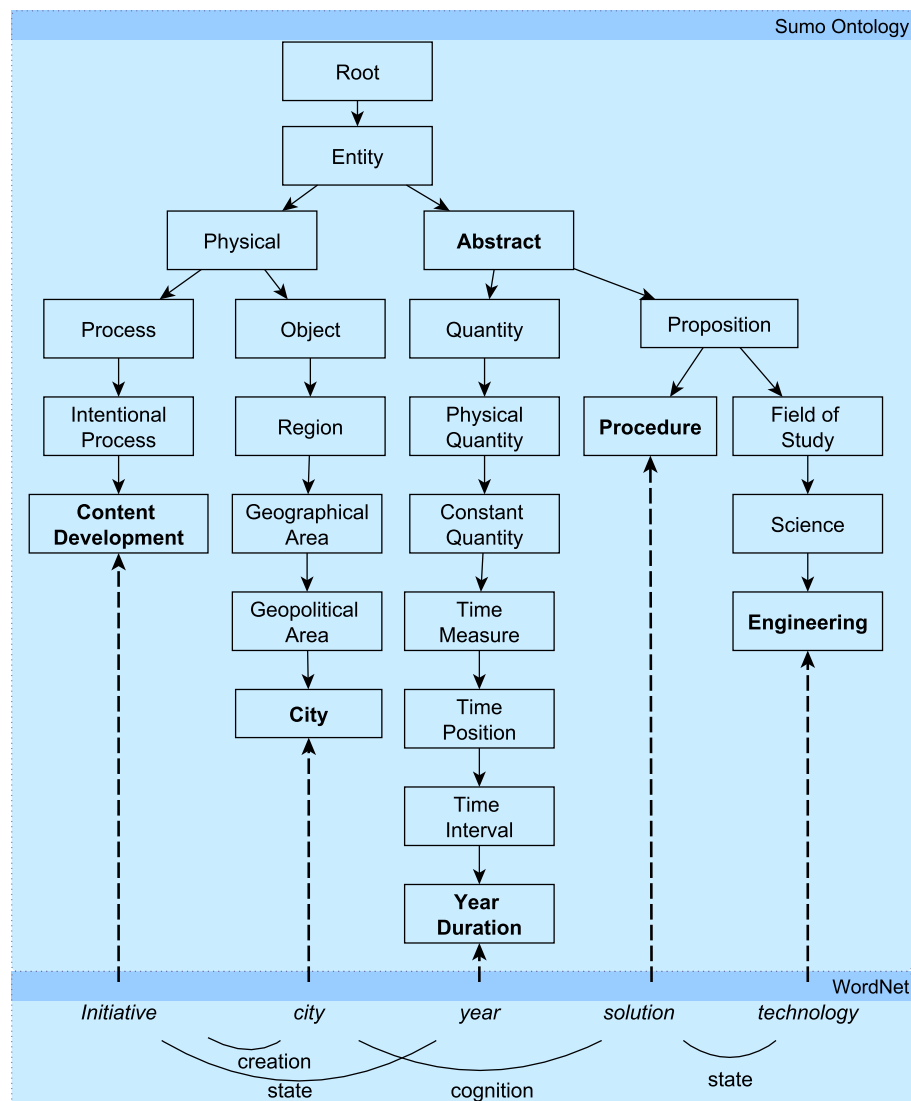


Figure 4. Example mapping of the semantic network generated for the sentence shown in Section 3.2.1 into the hierarchical structure of the SUMO ontology. Each noun in the sentence can be abstracted by navigating the structure of the SUMO ontology.

Over large documents, it is very likely that the semantic maps generated will be very complex and difficult to read. To further simplify these maps for visual inspection, we will apply the PathFinder networks algorithm [43]. Network scaling represents another method based on graph theory. Pathfinder networks are derived from proximities for pairs of entities. The PathFinder network uses the proximity of concepts to generate simplified structures revealing the underlying organization of the semantic network by retaining only the most relevant connections in a semantic network while removing cycles.

4. Experiments

To demonstrate our approach, we used three smart city initiative documents that were generated by different administrative levels: at the international level we evaluated the United Smart Cities (USC) from United Nation [44] which was created to address areas such as urban mobility, sustainable housing, clean energy, and waste management. At the national level we evaluated the United States Smart City fact sheet [45] which was created to address issues such as changing climate and improving city services. At the regional level we evaluated the Philadelphia Smart City Road Map [46] which was created to address issues related to city service delivery.

4.1. Experimental Setup

Using the procedure above, we extracted triplets from each of the three documents. The statistics related to this procedure are shown in Table 4. For example, from the original United Nations, United States, and Philadelphia documents we extracted a number of 183, 425, and 496 unique subjects and objects. These words are connected by 30, 820, and 893 actions to form the triplets. When we further processed the generated triplets, the number of unique concepts typically declined due to abstraction. An exception to this trend was when we matched words into the SUMO ontology. This is because a single word can be mapped into multiple SUMO concepts. For example, in the triplet *organization reduce cost*, the subject *organization* is mapped into two SUMO concepts: *Organization* and *Managing* with different degrees of confidence. At the same time, the object *cost* is also mapped into two SUMO concepts: *Process* and *Attribute*, also with different degrees of confidence. As a result of this step in our process, we added two concepts and three relationships.

Also, data in Table 4 shows the process of abstraction of textual information in the three documents. For example, in the original text, words in the United Smart Cities text from United Nation were only 16.6% of the total number of words in the three documents and only 2.7% of these words were found in the other two documents. As we processed these documents, the percentage of concept overlap over documents increased to 37% after mapping words into the SUMO ontology, 49.5% after abstracting the SUMO concepts to level 5, and 55.2% after mapping the SUMO concepts into the SUMO subject files. The number of relationships/triplets also decreased but at a lower rate since relationships are not mapped in the SUMO ontology but rather into the WordNet lexical classes. This reduction in semantic graph elements will enable us to assess the differences and similarities between the three documents.

Table 4. Statistics for the triplets extracted from the three documents. For each of the processed document, we calculated (1) the unique count of words/concepts and (2) the unique count of relationships between concepts.

Stage	United Nations		United States		Philadelphia	
	Concepts	Triples	Concepts	Triples	Concepts	Triples
Original	183	301	425	820	496	893
Inflected	161	285	339	802	406	862
Replace Action	161	269	339	767	406	819
SUMO	167	967	297	2177	327	2501
SUMO Level 5	88	997	140	1820	150	2066
SUMO Files	16	63	18	137	21	174

4.2. Results

With a complete set of triplets which resulted from the process above, we evaluated the level of correlation among the three documents. The results are shown in Table 5. The correlation of concepts in the semantic map is shown in bold fonts above the main diagonal while the correlation for the relationships among concepts is shown below the main diagonal in italic fonts. These results show that the concepts are more correlated than the relationships. For example, the correlation of concepts between the US Smart City fact sheet and Philadelphia Smart City Road Map documents is very high at 0.67 which is expected since Philadelphia is a city in the US. At the same time, the correlation between how these concepts are related stands at 0.42 which hints to the fact that there may be differences in implementation of the topic of study. Although the concepts used in the USC from UN and Philadelphia Smart City Road Map document seem to have a good correlation, the way they are related is the lowest in our analysis and we attribute this fact to the difference in reach of those two organizations.

Table 5. Correlation of relevance for the three documents. The correlation for relevance of concepts is shown in bold font above the main diagonal, while the correlation for relevance of relationships is shown in italic font below the main diagonal.

	United Nations	United States	Philadelphia
United Nations		0.40	0.46
United States	<i>0.43</i>		0.67
Philadelphia	<i>0.32</i>	<i>0.42</i>	

Further, we generated a conceptual map for each of the three documents. The generated maps varied in size and detail level. For example, the map for the USC from UN document contains 167 SUMO concepts and 967 relations among them, as shown in Table 4. At the same time, the US Smart City fact sheet and Philadelphia Smart City Roadmap semantic maps are much larger, at the same level, with 297 and 327 concepts respectively. Even reducing these maps at the Level 5 in the SUMO ontology will result in rather large semantic maps which are difficult to present here. For that reason, we show the semantic networks at the SUMO subject file level for each document.

The generated networks are shown in Figures 5–7. We also applied the PathFinder networks algorithm [43] to simplify these networks. The PathFinder networks algorithm simplify the semantic network by removing the most irrelevant relationships and maintaining an acyclical graph. In these figures, the relevant connections are represented in solid line having the width proportional with its relevance while the reduced/irrelevant relationships are shown in dotted line. For example, in Figure 5 we notice that there is a relevant connection between *Housing* and *Travel* but this would be reduced when we applied the PathFinder Networks algorithm because each of those concepts have stronger relationships to the central node.

On these figures we can also observe a different approach to categories that relate to processes, geography, and society. The importance of these categories increases with an increase in the locality of the policy. At the most global level, the UN policy only talks about these issues in general terms without too much emphasis. We attribute this to the fact that these issues need to be addressed at a local level. The USA policy emphasizes more on these categories by stressing the role of related categories such as Government and Engineering. At the most local level, the Philadelphia policy has the strongest focus on these policies and adds information on specific topics that are used in the implementation of these policies such as technology, media and transportation. It is also important to mention that, as it will be shown later in this section, the subtopics addressed at each level vary with the locality degree. For example, climate change, which is typically included in these categories, is addressed differently by the three policies.

We also observe that the number of smart city concept areas vary from one document to the other with fewer concept areas addressed by higher level entities that cover more

geographical area. The Philadelphia Smart City Roadmap document covers a total of 20 concept areas of the smart city with the most relevant ones being *Computing, Travel, and Processes*, in the order of their relevance. The US Smart City fact sheet document does not address concept areas such as *Industry, Devices, or Society*. We attribute this fact to the decentralized nature of the US administration, according to which these issues are decided at the local level. For the US Smart City fact sheet document, the most relevant areas are *Computing, Economy, and Travel*, in the order of their relevance. Lastly, the USC from UN document does not address concept areas such as *Industry, Devices, Society, Engineering, or Communications*. We attribute this to scope of UN which needs to address policies across many countries regardless of their technical implementations.

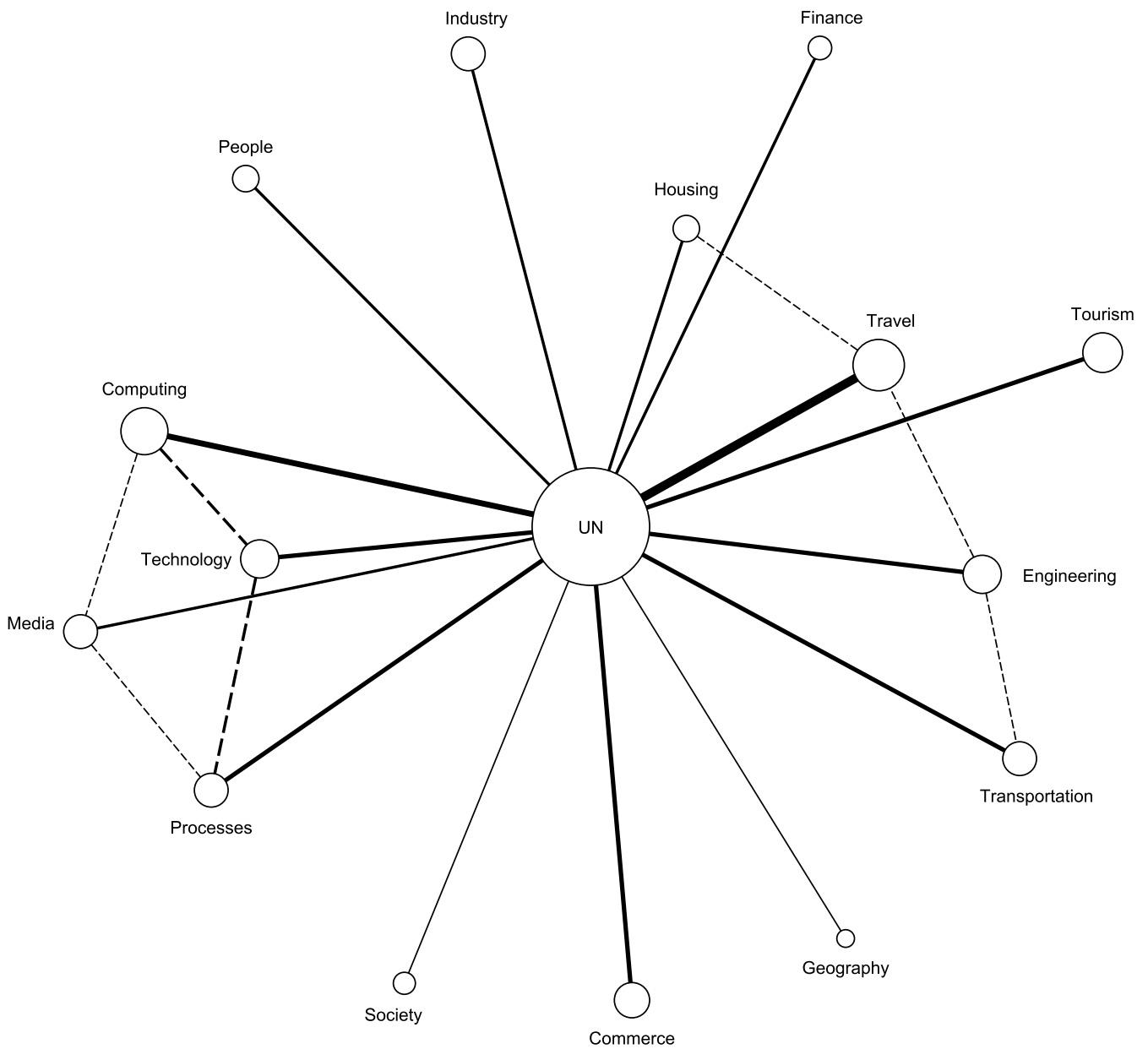


Figure 5. Semantic network generated for the United Nations Smart City document.

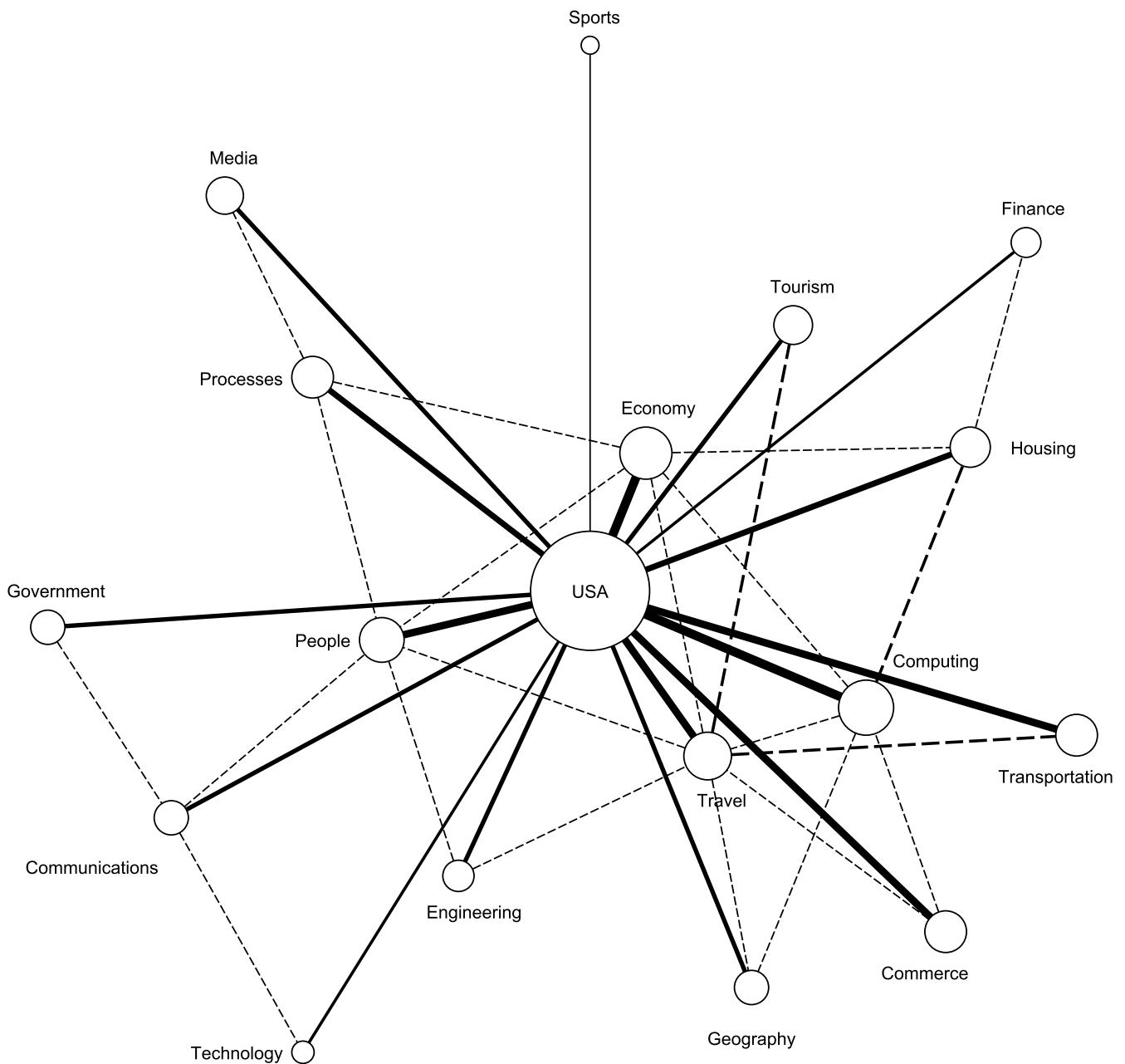


Figure 6. Semantic network generated for the United States Smart City document.

We further analyzed the strength of the relationships between concepts. For example, the USC from UN document elaborates the most on the current state of *Travel* and emphasizes on *communication* in this area. The *social* and *social* aspects of *Travel* are also addressed but to a much lesser degree. The US Smart City fact sheet document elaborates on the current state of the *Economy* and emphasizes on *change* and *change* in this area. The *consumption* aspect is also addressed but to a much lesser degree. Finally, the Philadelphia Smart City Roadmap document elaborates on the current state of the *Computing* and emphasizes on *possession*, *perception*, and *change* in this area. Several other aspects of this relationship are also addressed but to a much lesser degree.

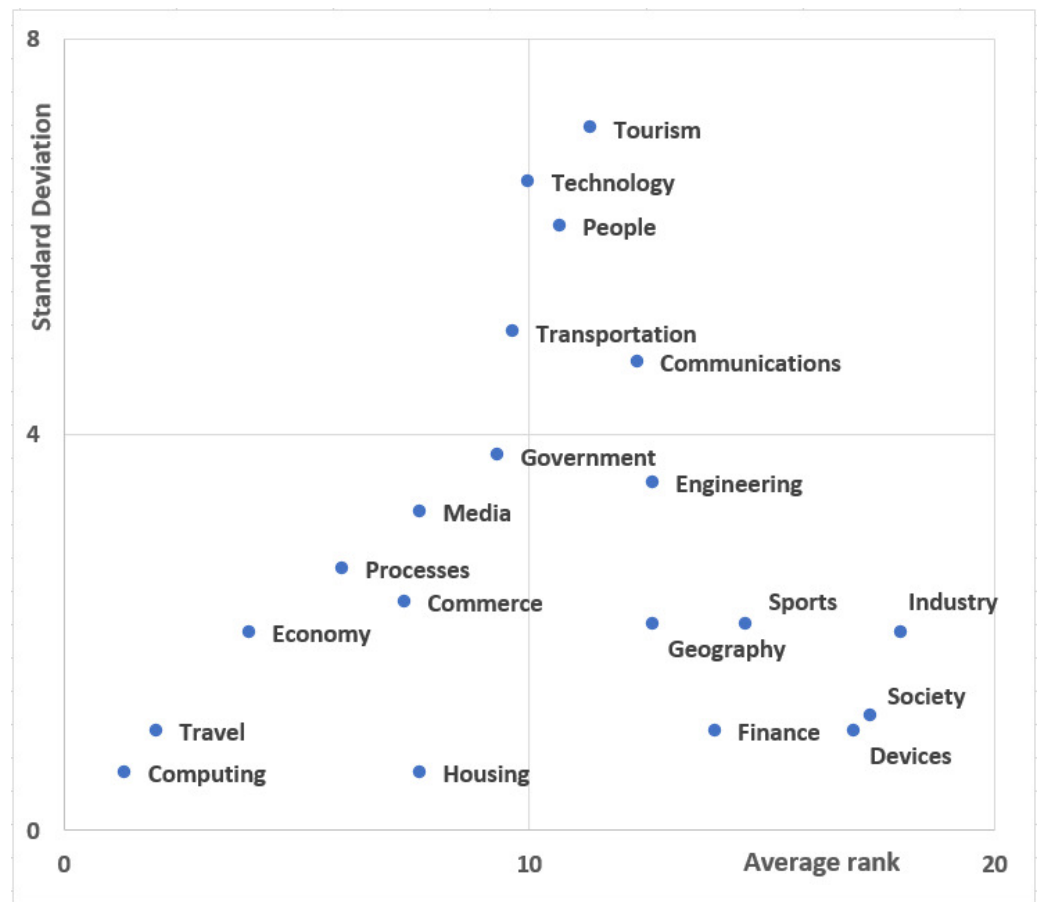


Figure 8. Quadrant of relevance variation for average ranking of each smart city area.

4.3. Analysis and Discussion

To evaluate the effectiveness of our methodology, we analyzed each of the documents and manually extracted a list of smart city topics in those documents. A complete list of these goals is shown in Table 6. For example, “climate change” is explicitly mentioned in the USC from United Nations and US Smart City fact sheet documents but not in the Philadelphia Smart City Roadmap document which mentions “climate resiliency” only. For each of the documents we ranked each topic by computing the similarity of each subject and object in the generated triplets to the description of the topic.

Table 6. Comparison of Smart City objectives for each document analyzed and the predicted rank of the topic in the document.

Goal	U.N.		U.S.A.		PHL.	
	Act	Pred	Act	Pred	Act	Pred
City Services	x	3	x	1	x	1
Clean Energy	x	7		11		11
Climate Change	x	4	x	6		4
Demographic Changes	x	2		9		9
Economic Growth		10	x	3		8
Energy Consumption		9		4	x	3
Public–Private Partnership	x	8		7	x	5
Public Safety		11	x	10		10
Quality of Life	x	5	x	2	x	2
Traffic management		6	x	5		6
Urban Migration	x	1		8		7

As seen in Table 6 the topic *City Services* is ranked consistently high as a topic for all the documents, while the *Urban Migration* is correctly ranked high for the USC from UN document and low for the other two documents. However, *Public Safety* is ranked consistently low although it should be ranked higher for the US Smart City fact sheet document. Overall, the average mean precision was 94% for the USC from UN document, 78% for the US Smart City fact sheet document and 91% for the Philadelphia Smart City Roadmap document. Also, the recall rate at the identified number of topic rate is 71% for the USC from UN document, 83% for the US Smart City fact sheet document, and 75% for the Philadelphia Smart City Roadmap document. For example, the UN document identifies seven relevant smart city areas. If we retrieved the top seven topic as ranked by our algorithm, we would recall 71% of the relevant topics. These results reveal that our methodology is promising in terms of automatically retrieving relevant smart city areas for the US and Philadelphia documents. Further refinement of the methodology is necessary to retrieve smart city areas more effectively for the UN smart city initiative document.

To summarize, our formalized semantic networks of UN, US, and Philadelphia smart city initiative documents provide a common framework which decision makers and policy analysts can utilize to analyze initiatives generated at different administrative levels. Since ambiguity and variations among text is reduced by analyzing semantic networks at the same abstraction or granularity level, comparative analysis can be constructed to understand how the initiatives align or differ from others. Our analysis revealed several smart city areas that are shared among all three initiatives including *Computing* and *Travel*. These smart city areas emerge as common concepts valued across international, national, and city level systems. Using the semantic networks in Figures 5–7, we also observe variations among initiatives that capture localized or system level smart city areas. For example, smart city area *Economy* is important for US Smart City initiative, but this area does not explicitly show up in the Philadelphia smart city semantic network. If decision makers want to scale the Philadelphia smart city initiative to other US cities or across the USA, then they need to address *Economy* area explicitly. A similar comparison can be done for high level administrators. Philadelphia smart city initiative emphasizes smart city areas that are important at local level such as *Housing*, *Industry*, *Devices*, and *Society*. If US government level decision makers could understand the contextual areas that are important for local administrations, then they can design incentives to promote these areas which ultimately would support *Economy* area that is a priority at the country level. The analysis also provides means to evaluate how US initiatives differ from international initiatives such as the USC from UN initiative. Since UN initiative spans across countries, several smart city areas such as *Government* and *Economy* do not show up in the semantic network as these concepts reflect country level areas. In addition to comparison of formalized semantic networks, decision makers can utilize Figure 8 to quantitatively evaluate the relevance variation among smart city initiative documents to understand the strengths or weaknesses of the documents. For example, *Tourism* is a smart city concept that is shared among all three initiatives, but it has high relevance for UN initiative rather than US or Philadelphia initiatives. Analysts could use this information to check if there is a conceptual gap in their initiative and if further analysis is necessary to address this smart city area and align their initiative with higher level administrative initiatives.

5. Conclusions

In this article, we developed a methodology to represent domain-specific knowledge using semantic maps. These maps consist of subject-action-object triplets extracted from text using semantic fusion of knowledge-based and deep learning methods. These semantic maps are further formalized and abstracted to a formal ontology. The methodology can be utilized to establish a common knowledge framework for decision makers and policy analysts who work with multi-level and multi-domain documents that are written by different administrative levels with differing objectives and terminology. The comparison and analysis of formalized semantic networks alleviates fragmented view of knowledge

across different administrative levels leading towards a holistic view of knowledge for decision makers and analysts.

We applied this methodology for analysis of three smart city initiative documents generated at three levels of organizational authority: international, country, and city. Our evaluation showed how each administrative level addressed the subject differently by using different level of abstractions and focused on different aspects of the smart city topic. For example, *Computing* and *Travel* are common smart city topics for all administrative levels. *Economy* is important at the country level whereas *Housing*, *Industry*, and *Services* are important at the local level. Our analysis also provided a way to align initiatives by identifying the variations in relevance of smart city topics. For example, *Tourism* topic showed highest variation in relevance among topics which points to an area that decision makers and analysts might investigate further for alignment of initiatives.

Our methodology provides a promising way to automatically extract knowledge from multi-domain and multi-level text and formally analyze knowledge for decision makers and policy analysts. We demonstrated the methodology on publicly available three multi-level documents. Our future work will scale up the number of policy documents analyzed where the benefit of automatic knowledge extraction will be apparent to decision makers and analysts. We will also evaluate the relevance of these policies to areas of interest for communities such as climate change, impact of natural hazards, and supply. Furthermore, we will extend this work and focus on in-depth evaluation of smart city policy development procedures using mereotopology principles. We will evaluate the smart city information topology at different mereological levels to identify gaps or overlaps in policies. The knowledge gained from this process can be used to address policy areas that need more focus for a consistent allocation of financial and human effort.

Author Contributions: Conceptualization, N.K.-E. and A.S.B.; methodology, A.S.B.; software, A.S.B.; validation, N.K.-E. and A.S.B.; formal analysis, A.S.B.; resources, N.K.-E.; data curation, A.S.B.; writing—original draft preparation, N.K.-E. and A.S.B.; writing—review and editing, N.K.-E.; visualization, A.S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used for the research is publicly available on the web.

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

References

1. Neirotti, P.; De Marco, A.; Cagliano, A.C.; Mangano, G.; Scorrano, F. Current Trends in Smart City Initiatives: Some Stylised Facts. *Cities* **2014**, *38*, 25–36. [[CrossRef](#)]
2. Bakici, T.; Almirall, E.; Wareham, J. A Smart City Initiative: The Case of Barcelona. *J. Knowl. Econ.* **2012**, *4*, 135–148. [[CrossRef](#)]
3. Trindade, E.P.; Hinnig, M.P.F.; Da Costa, E.M.; Marques, J.S.; Bastos, R.C.; Yigitcanlar, T. Sustainable Development of Smart Cities: A Systematic Review of the Literature. *J. Open Innov. Technol. Mark. Complex.* **2017**, *3*, 11. [[CrossRef](#)]
4. Magro, E.; Wilson, J.R. Complex Innovation Policy Systems: Towards an Evaluation Mix. *Res. Policy* **2013**, *42*, 1647–1656. [[CrossRef](#)]
5. Tompson, T. Understanding the Contextual Development of Smart City Initiatives: A Pragmatist Methodology. *J. Des. Econ. Innov.* **2017**, *3*, 210–228. [[CrossRef](#)]
6. Gruber, T.R. A Translation Approach to Portable Ontology Specifications. *Knowl. Acquis.* **1993**, *5*, 199–220. [[CrossRef](#)]
7. Giunchiglia, F.; Villafiorita, A.; Walsh, T. Theories of Abstraction. *AI Commun.* **1997**, *10*, 167–176. [[CrossRef](#)]
8. Hertz, T.; Mancilla Garcia, M.; Schlüter, M. From Nouns to Verbs: How process ontologies Enhance our Understanding of Social-Ecological Systems Understood as Complex Adaptive Systems. *People Nat.* **2020**, *2*, 328–338. [[CrossRef](#)]
9. Resnik, P. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. *J. Artif. Intell. Res.* **1999**, *11*, 95–130. [[CrossRef](#)]
10. Mikolov, T.; Chen, K.; Corrado, G.S.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.

11. Chourabi, H.; Nam, T.; Walker, S.; Gil-Garcia, J.R.; Mellouli, S.; Nahon, K.P.T.A.; Scholl, H.J. Understanding Smart City Initiatives: An Integrative and Comprehensive Theoretical Framework. In Proceedings of the 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2012.
12. Alawadhi, S.; Aldama-Nalda, A.; Chourabi, H.; Gil-Garcia, J.R.; Leung, S.; Mellouli, S.; Nam, T.; Pardo, T.; Scholl, H.J.; Walker, S. Building Understanding of Smart City Initiatives. In *International Conference on Electronic Government*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7443, pp. 40–53.
13. Hu, Q.; Zheng, Y. Smart City Initiatives: A Comparative Study of American and Chinese Cities. *J. Urban Aff.* **2020**, *43*, 504–525. [[CrossRef](#)]
14. Albino, V.; Berardi, U.; Dangelico, R.M. Smart Cities: Definitions, Dimensions, Performance, and Initiatives. *J. Urban Technol.* **2015**, *22*, 3–11. [[CrossRef](#)]
15. Manville, C.; Cochrane, G.; Cave, J.; Millard, J.; Pederson, J.K.; Thaarup, R.K.; Liebe, A.; Wik, A.L.; Wik, M.W.; Massink, R.; et al. *Mapping Smart Cities in the EU*; Publication Office of the European Union: Brussels, Belgium, 2014.
16. Angelidou, M. Smart City Policies: A Spatial Approach. *Cities* **2014**, *41*, S3–S11. [[CrossRef](#)]
17. Mayangsari, L.; Novani, S. Multi-Stakeholder Co-creation Analysis in Smart City Management: An Experience from Bandung, Indonesia. *Procedia Manuf.* **2015**, *4*, 315–321. [[CrossRef](#)]
18. Rodick, D.W. Process Re-engineering and Formal Ontology: A Deweyan Perspective. *Philos. Soc. Crit.* **2015**, *41*, 557–576. [[CrossRef](#)]
19. Miller, B. When is Consensus Knowledge Based? Distinguishing Shared Knowledge from mere Agreement. *Synthese* **2013**, *190*, 1293–1316. [[CrossRef](#)]
20. Woelert, P. Governing Knowledge: The Formalization Dilemma in the Governance of the Public Sciences. *Minerva* **2015**, *53*, 1–19. [[CrossRef](#)]
21. Reed, S.K. A Taxonomic Analysis of Abstraction. *Perspect. Psychol. Sci.* **2016**, *11*, 817–837. [[CrossRef](#)]
22. Valaski, J.; Malucelli, A.; Reinehr, S. Ontologies Application in Organizational Learning: A Literature Review. *Expert Syst. Appl.* **2012**, *39*, 7555–7561. [[CrossRef](#)]
23. Levin, B. *English Verb Classes and Alternations: A Preliminary Investigation*; University of Chicago Press: Chicago, IL, USA, 1993.
24. Fellbaum, C. WordNet. In *Theory and Applications of Ontology: Computer Applications*; Springer: Dordrecht, The Netherlands, 2010; pp. 231–243.
25. Gangemi, A.; Guarino, N.; Masolo, C.; Oltramari, A.; Schneider, L. Sweetening Ontologies with DOLCE. In *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*; Gómez-Pérez, A.; Benjamins, V.R., Eds.; Springer: Berlin/Heidelberg, Germany, 2002; pp. 166–181.
26. Smith, B.; Spear, A.D. *Building Ontologies with Basic Formal Ontology*; The MIT Press: Cambridge, MA, USA, 2015.
27. Pease, A.; Niles, I. IEEE Standard Upper Ontology: A Progress Report. *Spec. Issue Ontol. Agents* **2002**, *17*, 65–70. [[CrossRef](#)]
28. Pease, A. *Ontology*; Articulate Software Press: Angwin, CA, USA, 2011.
29. Bellini, P.; Benigni, M.; Billero, R.; Nesi, P.; Rauch, N. Km4City Ontology Building vs. Data Harvesting and Cleaning for Smart-City Services. *J. Vis. Lang. Comput.* **2014**, *25*, 827–839. [[CrossRef](#)]
30. Qiu, J.; Chai, Y.; Liu, Y.; Gu, Z.; Li, S.; Tian, Z. Automatic Non-Taxonomic Relation Extraction from Big Data in Smart City. *IEEE Access* **2018**, *6*, 74854–74864. [[CrossRef](#)]
31. Miguel-Rodríguez, J.D.; Galán-Páez, J.; Aranda-Corral, G.A.; Borrego-Díaz, J. Urban Knowledge Extraction, Representation and Reasoning as a Bridge from Data City towards Smart City. In Proceedings of the International IEEE Conferences on Ubiquitous Intelligence Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCCom/IoP/SmartWorld), Toulouse, France, 18–21 July 2016; pp. 968–974.
32. Jenkins, H.; Ford, S.; Green, J. *Spreadable Media: Creating Value and Meaning in a Networked Culture*; NYU Press: New York, NY, USA, 2018; Volume 15.
33. Davies, M. The Corpus of Contemporary American English (COCA): 600 Million Words, 1990-Present. Available online: <https://www.english-corpora.org/coca/> (accessed on 5 October 2020).
34. Levy, O.; Goldberg, Y.; Dagan, I. Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Trans. Assoc. Comput. Linguist.* **2015**, *3*, 211–225. [[CrossRef](#)]
35. Boleda, G.; Herbelot, A. Formal Distributional Semantics: Introduction to the Special Issue. *Comput. Linguist.* **2016**, *42*, 619–635. [[CrossRef](#)]
36. Li, M.; Chen, X.; Li, X.; Ma, B.; Vitanyi, P. The similarity metric. *IEEE Trans. Inf. Theory* **2004**, *50*, 3250–3264. [[CrossRef](#)]
37. Atkinson, K. Aspell 0.60.8 Released. 2019. Available online: <https://aspell.net/> (accessed on 5 October 2020).
38. Manning, C.D.; Surdeanu, M.; Bauer, J.; Finkel, J.; Inc, P.; Bethard, S.J.; Mcclosky, D. The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, ML, USA, 23–24 June 2014; pp. 55–60.
39. Angeli, G.; Johnson Premkumar, M.J.; Manning, C.D. Leveraging Linguistic Structure For Open Domain Information Extraction. In Proceedings of the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; pp. 344–354.

40. Finkel, J.R.; Grenager, T.; Manning, C. Incorporating Non-Local Information into Information Extraction Systems by Gibbs Sampling. In *ACL 2005, Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Ann Arbor, MI, USA, 25–30 June 2005*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2005; pp. 363–370.
41. de Boer, M.; Schutte, K.; Kraaij, W. Knowledge based query expansion in complex multimedia event detection. *Multimed. Tools Appl.* **2016**, *75*, 9025–9043. [[CrossRef](#)]
42. Kilicay-Ergin, N.; Barb, A.S. Smart City Document Evaluation to Support Policy Analysis. In Proceedings of the IEEE International System Conference SYSCON2020, Montreal, QC, Canada, 24 August–20 September 2020.
43. Schvaneveldt, R.W. *Pathfinder Associative Networks*; Ablex Series in Computational Sciences; Ablex Publishing Corporation: New York, NY, USA, 1990.
44. United Nations. United Smart Cities (USC)—United Nations Partnership for SDGs Platform. 2015. Available online: <https://sustainabledevelopment.un.org/partnership/?p=10009> (accessed on 5 November 2020).
45. The White House. FACT SHEET: Administration Announces New Smart Cities Initiative to Help Communities Tackle Local Challenges and Improve City Services. 2015. Available online: <https://obamawhitehouse.archives.gov/the-press-office/2015/09/14/fact-sheet-administration-announces-new-smart-cities-initiative-help> (accessed on 5 November 2020).
46. City of Philadelphia. SmartCityPHL Roadmap. 2019. Available online: <https://www.phila.gov/media/20190204121858/SmartCityPHL-Roadmap.pdf> (accessed on 5 November 2020).