A Fuzzy Ontology Generation Framework for Handling Uncertainties and Non-uniformity in Domain Knowledge Description

Muhammad Abulaish
Member, IEEE & IEEE-CS
Department of Mathematics
Jamia Millia Islamia (A Central Univ.)
Jamia Nagar, New Delhi – 25, India
abulaish@ieee.org

Lipika Dey
Member, IEEE & IEEE-CS
Department of Mathematics
Indian Institute of Technology
Hauz Khas, New Delhi – 16, India
lipika@maths.iitd.ernet.in

Abstract

Since web documents are not fully structured sources of information and in Internet almost everything, especially in the realm of search, is approximate in nature, it is not possible to utilize the benefits of a domain ontology straight away to extract information from such a document. One way of overcoming this problem is the postulation of a "fuzzy ontology" by adding a value for degree of membership to each term that is imprecise in nature. In this paper, we propose a fuzzy ontology generation framework in which a concept descriptor is represented as a fuzzy relation which encodes the degree of a property value using a fuzzy membership function. The fuzzy ontology framework provides appropriate support for application integration by identifying the most likely location of a particular term in the ontology. The applicability of the fuzzy ontology structure in retrieving and curating information from text documents to answer imprecise queries has been thoroughly experimented.

Keywords: Semantic Web, Fuzzy ontology structure, Ontology enhancement.

1. Introduction

As envisaged by Berners-Lee, the Semantic Web (SW) [7] promises to make the Web a meaningful experience, and ontology [1] is increasingly being accepted as the knowledge-management structure that can eventually realize it. Ontologies represent domain knowledge in a structured and machine-interpretable form. Ontology represents a method of formally expressing a shared understanding of information. Ontologies are being increasingly used to support a great variety of tasks and as a structure to represent meaning of data ontologies are emerging as the main area of interest for the success of the SW paradigm. Though ontology plays a key role by defining concepts and relationships in an unambiguous way, one of the chief bottlenecks that crops up while designing domain-specific applications is that the ontology itself is a pre-defined structure with crisp concept descriptions and inter-concept relations. It is unlikely that all application developers will strictly adhere to the concept descriptors used in the ontology. Moreover, if the application involves text processing, it is absolutely unreal to expect that concept descriptions appearing within the text will be easily identifiable as ontology concepts. Besides, for most of the domains, other than the strictly technical ones like the medical domain, it is found that knowledge-modeling experts differ in their conceptualization of a domain. There may be ambiguity both in the set of concepts used to define a domain and also in the use of inter-concept relations.

One way of overcoming this problem is the postulation of a "fuzzy ontology", in which inter-concept relations can be represented as fuzzy relations rather than as crisp associations. Thus every inter-concept relation is viewed as a fuzzy association whose strength can be ascertained from application parameters. Thus rather than looking for exact concept descriptions, a fuzzy ontology based application identifies the most likely concept from the ontology. The fuzzy ontology also serves as an ideal tool for handling multiple ontology descriptions created for the same domain. Incorporating imprecision into the ontology structure itself helps in resolving ambiguities arising due to differences in user requirement specification and concept descriptions embedded in text documents.

In this paper we have proposed the model for representing and generating a *fuzzy ontology*, which can be used for designing text-processing tools. The proposed framework exploits fuzzy logic and reasoning mechanisms to incorporate fuzzy membership functions into the rigid ontology structure. In the proposed design, a concept descriptor is represented as a fuzzy relation, which encodes the degree of a property value using a fuzzy membership function. The fuzzy concept descriptors could be either manually generated, or as we have shown, can be extracted through text mining. Other than concept descriptors, generic semantic relations extracted from text documents are also associated strengths and represented as a fuzzy relation.

2. Related work on fuzzy ontology structure

Though ontology is meant to represent knowledge in an unambiguous structured format, it is practically impossible to assume that all application developers will agree to any such unique structure. Enhancement of crisp ontology structures to a fuzzy ontology structure is viewed as a potential solution to this problem and has received attention from a number of research groups.

Widyantoro and Yen [2] have shown how fuzzy membership values associated to ontology concepts, along with a concept hierarchy, can be used for intelligent text information retrieval. Starting with a set of manually tagged abstracts of papers from several IEEE Transactions, a fuzzy ontology is built on the collection of keywords. The abstracts are tagged based on their title, authors, publication date, abstract body, and author supplied keywords. The hierarchical arrangement of the terms in the newly generated ontology is dependent on their co-occurrence measures. The drawback of this system is its dependence on user judgment about the relevance of articles to user queries which is provided manually.

Wallace and Avrithis [5] have extended the idea of ontology-based knowledge representation to include fuzzy degrees of membership for a set of inter-concept relations defined in an ontology. The membership of these relations are used to judge the context of a set of entities, the context of a user and the context of the query for the purpose of intelligent information retrieval. A fixed set of commonly encountered semantic relations have been identified and their combinations are used to generate fuzzy, quasi-taxonomic relations.

Quan *et al.* [8] have proposed an automatic fuzzy ontology generation framework – FOGA. They have incorporated fuzzy logic into formal concept analysis to handle uncertainty information for conceptual clustering and concept hierarchy generation. However, the quality of clustering is dependent on assignment of meaningful labels to initial class names, attributes and relations. This is done manually and requires domain expertise. This system is also not designed to extract fuzzy relational concepts from unstructured or semi-structured text documents.

3. Mathematical model of an ontology

An ontology represents a model of a domain that defines the concepts existing in that domain, their properties and the relationships between them and is typically represented as a knowledge base. For example, a plant ontology can specify relationship among various categories of plants like *algae*, *legumes*, *ferns* and so on, and also specify structural organization of plants in terms of parts and sub-parts like *stems*, *leaves*, *cells* etc.

Definition (*Ontology*) – An Ontology Θ is defined as a triplet of the form $\Theta = (C, P, \Re)$, where

- C is a set of concepts defined for the domain.
- P is a set of concept properties. A property $p \in P$ is defined as a ternary relation of the form p(c, v, f), where $c \in C$ is an ontology concept, 'v' is a property value associated with 'c' and 'f' defines restriction facets on v. Some of the restriction facets are – *type* (f_t), *cardinality* (f_c), and *range* (f_r). The *type* facet f_t may be any one of the standard data types that are supported by ontology editors. Thus $f_t \in$ {*Boolean*, *integer*, *float*, *string*, *symbol*, *instance*, *class*, …}. The *cardinality* facet f_c defines the upper and lower limits on the number of values for the property. The *range* facet f_r specifies a range of values that can be assigned to the property.
- $\Re \subseteq C \times C$. $\Re$ is a set of binary semantic relations which can be either one-to-one, one-to-many, or many-to-many. $\Re$ is recursively defined as follows:

a. A set of atomic relations is defined as $\Re_a = \{\approx, \uparrow, \downarrow, \nabla, \Delta\}$ which have the following interpretations:

For any two ontological concepts $C_i, C_j \in C$

- $\approx$ denotes the equivalence relation. $C_i \approx C_j \Rightarrow$ *C_i is equivalent to C_j.* The *synonym* relation of natural language is modeled in an ontology using the equivalence relation. For example, through WordNet [3] we obtain that the word "*inn*" is synonymous to "*hotel*". If two concepts C_i and C_j are declared equivalent in an ontology then instances of concept C_i can also be inferred as instances of C_j and vice-versa.
- $\uparrow$ denotes the generalization relation. $C_i \uparrow C_j \Rightarrow$ *C_i is a generalization* of *C_j.* When an ontology specifies that C_i is a generalization of C_j, then C_j inherits all property descriptors associated with C_i, and these need not be repeated for C_j while specifying the ontology. $\downarrow$ is the inverse of $\uparrow$. Hence, $C_i \uparrow C_j \Rightarrow C_j \downarrow C_i$, i.e., C_i is a generalization of C_j implies that C_j is a specialization of C_i. The relations $\uparrow$ and $\downarrow$ correspond to the semantic relations "h*ypernym*" and "*hyponym*" respectively of WordNet. These relations are usually denoted by an arrow super

scribed with "is-a" or "kind-of", where the arrow is directed from the specialized class to the generalized class. Ontologies can also accommodate multiple inheritances, whereby a concept can acquire properties through multiple paths of specialization.

* $C_i \nabla C_j \Rightarrow$ *C_i has part C_j.* Δ is inverse of ∇. Hence $C_i \Delta C_j \Rightarrow$ *C_i is a part of C_j.* In an ontology, a concept which is defined as aggregation of other concepts is expressed using the relation ∇. The "has-part" relation is equivalent to the "*holonym*" relation of WordNet.

b. If $\Re_1, \Re_2 \in \Re$ be any two relations defined between concept-pairs in Θ and o denotes a composition operation, then $\Re_1$ o $\Re_2$ is a valid relation.

4. The fuzzy ontology model

We now explain how the ontology structure is extended to accommodate fuzzy descriptions and relations. Traditionally concepts are described in an ontology using a <*property, value, constraints*> framework. The fuzzy ontology structure is created as an extension to the standard ontology structure. In this structure, property descriptors are accompanied by qualifiers along with values for defining a concept. The proposed fuzzy ontology structure stores concept descriptions in a <*property, value, qualifier, constraints*> framework, where the value and the qualifier are both defined as a fuzzy set. This framework allows defining the property-value of a concept with differing degrees of fuzziness, without actually changing the concept description paradigm. Such concept descriptions can be termed as imprecise concept descriptions.

Mathematically, a fuzzy ontology (Θ_F) can be defined as follows.

Definition (*Fuzzy Ontology*) – A Fuzzy Ontology, Θ_F, is a quadruple of the form

$\Theta_F = (C, P_F, \Re_F, M)$, *where:*

- C has same interpretation as mentioned in section 3.
- P_F is a set of fuzzy concept properties. A property $p_f \in P_F$ is defined as a quadruple of the form p_f (c, v_f, q_f, f), where c $\in$ C is an ontology concept, 'v_f' represents fuzzy attribute values and could be either *fuzzy numbers* or *fuzzy quantifiers*, 'q_f' models linguistic qualifiers and are *hedges*, which can control or alter the strength of an attribute value and *f* is the restriction facet on v_f.
- $\Re_F$ is a set of fuzzy inter-concept relations between concepts. Like fuzzy concept properties, $\Re_F$ is defined as a quadruple of the form $\Re_F$ (c_i, c_j, t, q_f), where c_i, $c_j \in$ C are ontology concepts, 't' represents relation type, and 'q_f' models relation strengths and are linguistic variables, which can represent the strength of association between concept-pairs <c_i, c_j>.
- The choice of *fuzzy numbers* or *fuzzy quantifiers* for values is dictated by the nature of the underlying attribute and also its restriction facets. The complete range of values over which an attribute can take values defines the *universe of discourse* M. The universe of discourse is decomposed into a collection of fuzzy sets. Each fuzzy set is defined over a *domain* that overlays part of the universe of discourse.

Since the essence of fuzzy sets is to be able to control the degree of imprecision rather than bind a single membership function to a definition, we propose the use of application-specific fuzzy-membership functions for fuzzy quantifiers and qualifiers. Though the membership functions themselves change depending on the nature of the domains, their role in modifying fuzzy attribute values remains unchanged across applications. For appropriate fuzzyfication of concept descriptions each attribute is also associated with a qualifier set which is a collection of hedges. Since the qualifiers associated to different properties are usually different, hence the hedge sets are also different though may be overlapping. To maintain uniformity of using concept descriptions, every value is always assumed to be accompanied by a qualifier. Hence to model values without a qualifier, we have used the qualifier "*null*". For every qualifier set, we have included the value "*null*" to indicate the absence of any qualifier.

An interesting aspect of modeling attributes as fuzzy sets is that with an underlying set of numeric values, one can associate different fuzzy quantifier sets to represent different aspects of the same attribute. For example, a single price value can be interpreted as being "*close to*" or "*far away*" from another value of price, and at the same time can also be interpreted as "*cheap*" or "*expensive*." Moreover, hedges can also be applied to create new fuzzy sets with different meanings. Thus modeling an attribute as a fuzzy set allows a single attribute to contribute to different types of imprecision in concept description.

Fuzzy qualifiers are used in fuzzy models to dynamically create new fuzzy sets and change the meaning of linguistic variables. This enables the modification of existing fuzzy sets temporarily to provide different meaning to the underlying linguistic variable. Most of the applications consider linguistic qualifiers as those elements that modify the value of a fuzzy number. However, modeling qualifiers become more complex when the fuzzy quantifier set is itself graded. For example, the weather domain uses three values *hot*, *cold*, and *cool* to model the weather condition in terms of temperature. In this case, fuzzy modeling of the

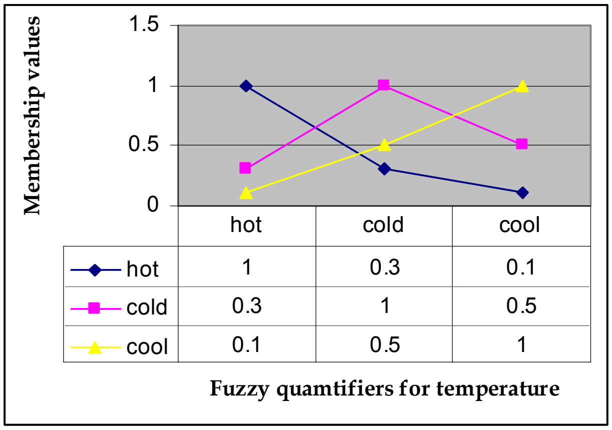


Figure 1. Fuzzy membership functions for temperature values

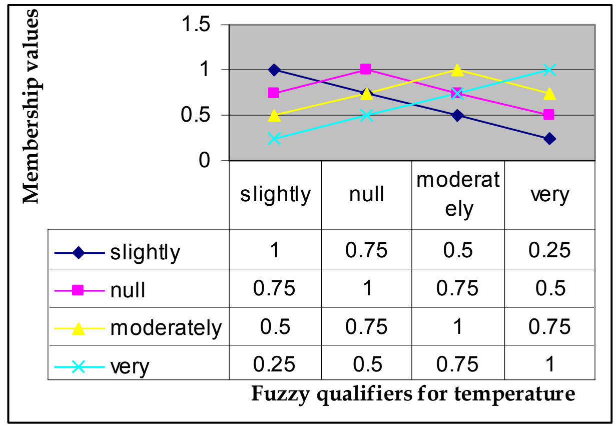


Figure 2. Fuzzy membership functions for weather qualifier set

temperature can be achieved by the membership table shown in Figure 1. As we can see, the weather value "*cool*" can be interpreted to be as "*cold*" to some extent, and vice versa, where the extent is defined by the fuzzy membership values. An interesting thing to observe is that since "*cool*" and "*cold*" are basically intensity variations of the same temperature, where "*cool*" is an intensified version of "*cold*", thus the weather which is "*very cold*" can be considered to be "*cool*" with a higher membership value than the weather which is simply "*cold*". Thus in this case we want that rather than working as an intensifier, which hardens or reduces the membership value for "*cold*" to "*cool*", the intensifier "*very*" should increase the membership value of "*cold*" to "*cool*". Obviously, this is a special situation occurring due to the gradation among the fuzzy quantifiers themselves.

To take care of all such situations, we have adopted a generalized approach to model fuzzy quantifier and qualifier sets. In this scheme both fuzzy quantifiers and fuzzy qualifiers can be modeled as graded sets, with the similarity between two variables defined as a function of their relative positions in the set. This allows us to control and combine the effects of qualifiers over quantifiers in a more context dependent way. The next section presents detailed description of the modeling scheme with specific references to domains indicating the types of values for which a particular modeling is suitable.

4.1 Encoding domain knowledge using concept descriptors

Since the proposed fuzzy ontology structure has been integrated with text information retrieval applications, and text abounds in vague descriptions, the qualifier sets for a domain has been extracted through text-mining. Thus, each domain yields a set of qualifiers, which are then modeled as hedges for that domain. These qualifiers are later on used for further information retrieval. As we have discussed earlier, the role of a modifier for a domain does not remain static. Rather it is defined as a function of both the qualifier and the value it is trying to modify. In case of matching a pair of <*value*, *qualifier*> tuples, the overall effect is to be determined as a function of the distance between the qualifiers, and the value pairs. When values match, but qualifiers do not match the overall aim is to always decrease the precision of an associated value.

Qualifier sets are modeled as graded sets. A graded set is an ordered collection of elements. The similarity between two objects in the graded set is defined as a function of their relative positions within the set. The position of "null" is selected depending on the nature of qualifiers used. For most of the qualifier sets, "null" occupies a central position, with dilution hedges occurring towards its left and intensification hedges occurring towards its right. However, if a domain includes only intensification hedges then "null" is located as the first element in an ascending ordered set. Similarly, for a set of only dilution hedges, "null" occupies the extreme right position in an ascending ordered set. Figure 2 shows the modeling of a sample qualifier set for the domain of values from weather.

We now show how the fuzzy memberships are computed for qualified variables. Fuzzy memberships for qualified variables are computed using composition of the fuzzy membership values for the variables and the qualifiers. The similarity between two qualified variables <q_i, v_i> and <q_j, v_j> is expressed as a fuzzy membership function denoted by $\mu_{(q_i,v_i)}(q_j,v_j)$.

Since qualifiers are modeled as graded sets, fuzzy membership functions for these sets can be designed using their relative positions within the set. The distance between two qualifiers in the collection reflects their degree of dissimilarity. The distance between the qualifier

q_i at position i and the qualifier q_j at position j within a set is defined by using equation 1.

$$d(q_i, q_j) = | i - j | \text{.................................}(1)$$

The fuzzy membership function for the qualifier set is then defined as given in equation 2.

$$S = f(q_i, q_j) = 1 - \frac{d(q_i, q_j)}{MAX + 1} \text{.....................}(2)$$

where, $MAX = \max \{d(q_i, q_j), \forall q_i, q_j \in Q$, where Q is the qualifier set$\}$. f is commutative in nature. Figure 2 shows the fuzzy membership functions derived for the qualifier sets for *temperature* property of weather.

In order to compute the fuzzy membership of compositions, we have taken the dilution or intensification aspects of both the qualifiers and values into account. An element t_i is a dilution with respect to another element t_j in the graded set if $i<j$ in the ordered set $\{t_i, t_j\}$. Conversely t_j is an intensifier with respect to t_i. This information is encoded in terms of a function as given in equation 3.

$$Sgn(t_i, t_j) = \begin{cases} +1, & if\ i < j \\ -1, & if\ i > j \\ 0, & if\ i = j \end{cases} \text{.......................}(3)$$

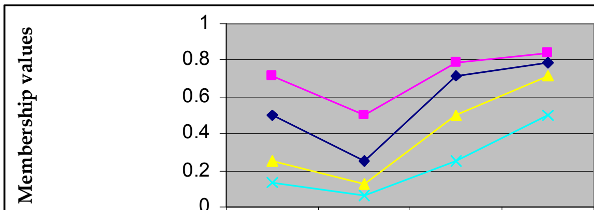
The elements t_i and t_j can represent a pair of qualifiers q_i and q_j or a pair of values v_i and v_j. The composite fuzzy membership function is defined as shown in equation 4.

$$\mu_{(q_i, v_i)}(q_j, v_j) = \begin{cases} \left[\mu_{v_i}(v_j)\right]^{\frac{1}{d(q_i, q_j)+1}} & if\ Sgn(q_i, q_j) \times Sgn(v_i, v_j) = -1 \\ \left[\mu_{v_i}(v_j)\right]^{d(q_i, q_j)+1} & if\ Sgn(q_i, q_j) \times Sgn(v_i, v_j) = +1 \\ \left[\mu_{v_i}(v_j)\right] \times f(q_i, q_j) & if\ Sgn(q_i, q_j) \times Sgn(v_i, v_j) = 0 \end{cases} \text{......}(4)$$

Figure 3 shows the composition of fuzzy quantifiers and qualifiers for the *temperature* property of weather concept.

We now present the fuzzy modeling mechanism to handle numeric attributes. For example, in the *weather* domain the temperature property may be expressed at multiple levels of granularity. While at the lowest level they may comprise of numeric values, for describing long term weather conditions usually linguistic variables like *hot, cold, cool* etc. are used and each numeric value can be mapped into these linguistic variables by using fuzzy membership functions. Figure 4 shows the modeling of temperature values by using these fuzzy sets.

Moreover, numeric attributes can also be expressed as *fuzzy numbers,* which simply represent fuzzy numeric intervals over the domain of particular variable. Fuzzy numbers are generally represented using bell-shape, triangular or trapezoidal membership function along with a fuzzy quantifier defined over the numeric domain with appropriate fuzzy functions. A subset of hedges known in the domain of fuzzy set theory like, *few*, *somewhat, small*, *average*, *more or less*, *many*, *very*, *high* etc. can also be used directly on crisp numbers to convert them into fuzzy sets through the process called approximation [4].



	null cold	slightly cold	moderately cold	very cold
null cool	0.5	0.25	0.71	0.79
slightly cool	0.71	0.5	0.79	0.84
moderately cool	0.25	0.125	0.5	0.71
very cool	0.13	0.06	0.25	0.5

Qualified temperature values

Figure 3. Composition of fuzzy quantifiers and qualifiers for temperature

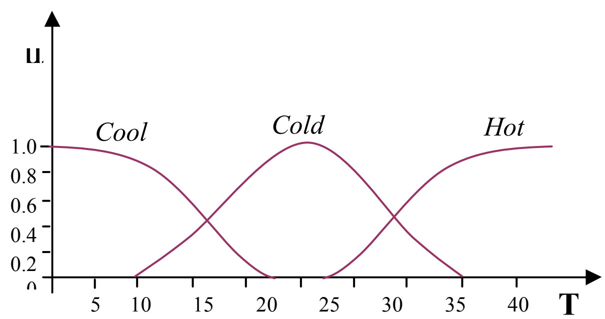


Figure 4. Fuzzy membership functions to represent temperature values

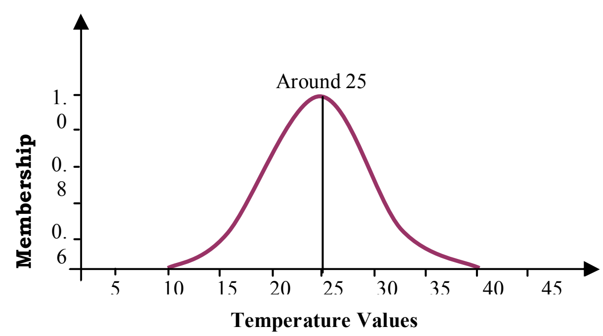


Figure 5. A fuzzy number "around 25" to represent the value of temperature

A bell-shape membership function to represent a fuzzy number "around 25" to represent the temperature value is shown in figure 5. Fuzzy numbers are generally represented through two principal attributes – a central value and a degree of spread around the value. The degree of spread is also called the expectancy (e) of the fuzzy number. When e=0, the fuzzy number is a single point and correspond to a normal scalar value. As the expectancy increases, the number becomes fuzzier.

5. Application of fuzzy ontologies

The proposed fuzzy ontology structure has been integrated with a text mining and query processing application. Starting with a seed ontology in which some domain concepts are defined, more concepts and concept descriptors are extracted from free-form text documents of the domain. Concept descriptions not encountered earlier are fuzzified through manual intervention, where the knowledge-engineer chooses the appropriate set of hedges, membership functions etc. and the requisite ontology is thereafter enhanced automatically. The knowledge extracted about domain instances are also curated into a database, which is then used for query processing.

Figure 6 shows a partial snapshot of a weather database about different countries of the world that has been created using the proposed fuzzy-ontology based text mining with web documents. This database is used to answer queries like "*Which country has weather similar to Andorra*", and the answers would be Australia and Albania, in that order.

The fuzzy ontology structure is also ideally suited for handling non-uniformity in domain descriptions. We illustrate this through an example. The concept *publication* defined in WordNet states that a publication *is a kind of* book, while [6] defines a publication as *synonymous* to book. The associated properties are also different and often non-overlapping. While none of them can be assumed to be wrong, our proposal is to convert each ontology into a fuzzy ontology where the relations are weighted to indicate the uniformity of the relations across ontologies. The weighting scheme is based on a simple weight-propagation mechanism, similar to that defined in [6]. Thus the relationship between *book* and *publication* is associated with a weight of 0.8 rather than 1, to indicate that it is not a consistent definition across ontologies.

Weather : Table

CName	Qclimate	Vclimate	Qsummer	Vsumme	Qwinter
Afghanistan	Null	semiarid	Null	hot	semiarid
Albania	mild	temperate	Null	dry	cloudy
Algeria	Null	semiarid	Null	hot	mild
sirocco	Null	Null	Null	Null	Null
American Samoa	Null	tropical	Null	Null	Null
Andorra	Null	temperate	Null	Null	snowy
Angola	Null	semiarid	Null	dry	Null
Anguilla	Null	tropical	Null	Null	Null
Antarctica	most	moderate	Null	Null	Null
Antiqua	Null	tropical	Null	Null	Null
Arctic Ocean	Null	Null	Null	Null	Null
Argentina	mostly	temperate	Null	Null	Null
Armenia highland contin	Null	Null	Null	hot	Null
Aruba	Null	tropical	Null	Null	Null

Figure 6. A curated Weather database (partial) from text documents

6. Conclusion

In this paper we have presented a fuzzy ontology generation framework in which concept descriptors and inter-concept relations are represented as fuzzy relations. This work has been integrated with a text-mining system such that, starting with a seed ontology a domain ontology can be extended with new knowledge extracted from text documents. The proposed ontology structure is also suitable to resolve the inconsistencies in concept descriptions and inter-concept relations present across multiple ontologies that define the same domain.

References

[1] D. Fensel, I. Horrocks, F. van Harmelen, D. L. McGuinness, and P. Patel-Schneider, "OIL: Ontology Infrastructure to Enable the Semantic Web", *IEEE Intelligent Systems 16(2)*, 2001, pp. 38-45.

[2] D. H. Widyantoro, and J. Yen, "A Fuzzy Ontology-based Abstract Search Engine and its User Studies", *Proceedings of the 10th IEEE International Conference on Fuzzy Systems*, Melbourne, Australia, 2001, pp. 1291-1294.

[3] G. Miller, "Wordnet: An online lexical database", *International Journal of Lexicography*, 3(4), 1997.

[4] L. A. Zadeh, "A Computational Approach to Fuzzy Quantifiers in Natural Languages", *Computational Mathematics Applications 9*, 1983, pp. 149-184.

[5] M. Wallace, and Y. Avrithis, "Fuzzy Relational Knowledge Representation and Context in the Service of Semantic Information Retrieval", *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Budapest, Hungary, 2004.

[6] S. Castano, A. Ferrara, and S. Montanelli, "H-MATCH: an Algorithm for Dynamically Matching Ontologies in Peer-based Systems", *Proceedings of the 1st International Workshop on Semantic Web and Databases (SWDB'03)*, 2003, pp. 231-250.

[7] T. Berners-Lee, "Semantic Web Road Map", W3C Design Issues, 1998, URL: www.w3.org/DesignIssues/Semantic.html

[8] T. T. Quan, S. C. Hui, and T. H. Cao, "FOGA: A Fuzzy Ontology Generation Framework for Scholarly Semantic Web", *Proceedings of the 2004 Knowledge Discovery and Ontologies Workshop (KDO'04)*, Pisa, Italy, 2004.