International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 781-793

http://www.issr-journals.org/ijias/

Hybrid Distance Based Measures for Geospatial Domain

K. Saruladha and E. Thirumagal

Pondicherry Engineering College, Department of Computer Science and Engineering, Puducherry, India

ABSTRACT: Current researchers of search engines focus more on semantic based information retrieval as syntax based retrieval yield less precision. Retrieving relevant information from diverse heterogeneous web resources remains as a challenge. Distance based measures play a major role in the information retrieval systems. This work focuses on retrieving relevant concepts using geospatial datasources to aid geospatial applications in predicting floods, locating underground pipes and cables and testing the quality of water. Geospatial data characterizes geographical features of the real world using spatial extent and location. This paper proposes HDSM algorithm for geospatial information retrieval which adapts the existing distance based measures viz., Manhattan distance, Euclidean distance, Vector cosine and bray Curtis for the geospatial domain to identify related concepts to the geo-spatial query concept. All these four proposed hybrid distance based measures combine the advantages of geometric and network semantic similarity models. The meaning of the geospatial concepts are captured from the expressive knowledge of the geospatial concept properties and geospatial relations. These proposed four Hybrid distance based measures have been tested using Ordnance Survey Master Map data source and ordnance survey ontology for varying semantic similarity thresholds. The experimental results are reported in this paper. The Hybrid Manhattan distance based measure has yielded the precision of 95%.

KEYWORDS: Hybrid model, Geo-spatial relations, Properties, Semantic similarity, Conceptualisation, Distance Measures.

1 INTRODUCTION

The semantic information retrieval system's [1] goal is to predict the relevant geo-spatial concepts that are best suited for the geo-spatial query concept given by the user. The semantic similarity measure compares the geo-spatial concepts for determining the degree of relevance between them. Based on the relevance of the geo-spatial concepts, the semantic information retrieval task compares the synonyms of the geo-spatial query concepts and search for the geo-spatial information which are relevant to the geo-spatial query concept given by the user. The well thought-out representation of semantics is essential for the competent information retrieval. The semantic description must include the properties [2] i.e. the geo-spatial concepts' features and also the spatial relations [3] i.e. the relationship between the features of the geo-spatial concepts for representing the geo-spatial concepts.

Schwering [4] proposed the Hybrid semantic similarity computational model which uses the geo-spatial properties and the geo-spatial relations for computing the semantic similarity among the geo-spatial concepts. The other semantic similarity computation models except the Hybrid semantic similarity computational model either use geo-spatial properties or the geo-spatial spatial relations for representing the geo-spatial concepts. The Hybrid semantic similarity computational model uses the conceptual hulls for describing the concepts. The properties of the geo-spatial concepts are represented as the vector points in the conceptual hull [5]. For measuring the semantic similarity between the geo-spatial concepts, the Hybrid model compares the vector points between the geo-spatial concepts. The Euclidean distance method is applied between the matching vector points between the geo-spatial concepts. The Hybrid semantic similarity computational model uses the Euclidean semantic distance measure for computing the semantic distance between the geo-spatial concepts which also does not yield better precision and recall. If the semantic distance between the vector points are more, it indicates that the semantic similarity is less. If the semantic distance between the vector points are less, it indicates that the semantic similarity is more. Hence the various hybrid distance measures are used to test the information retrieval.

Corresponding Author: E. Thirumagal (tmagaleng@gmail.com)

The main objective of this paper is to propose various hybrid distance based measures such as Hybrid Manhattan distance measure, Hybrid Vector Cosine distance measure and Hybrid Bray Curtis distance measures for the Hybrid model of the geo-spatial domain and to predict the distance measure which yields good relevant information retrieval. The hybrid distance measures are evaluated using the performance metrics such as precision, recall and f-measure. The Hybrid distance measure which yields high precision is said to be best semantic similarity distance measure for the geo-spatial domain. The hybrid distance measures which are said above are not used for geo-spatial information retrieval, hence this paper attempts to use those distance measures for computing semantic similarity and to retrieve the relevant geo-spatial information.

The geo-spatial datasource used for experiments in this paper is Ordnance Survey Master Map [5] which has four layers such as Topography Layer, OSMasterMap Address Layer, Imagery Layer, and Integrated Transport Network (ITN) Layer provides the information about the topography, address of the location, aerial images and the vehicle movement tracking. Further this paper proposes the HDSM (Hybrid Distance Similarity Measure) algorithm which depicts how the above said hybrid distance measures are used in the geo-spatial domain. In section 2, the related works are depicted. The HDSM (Hybrid Distance Similarity Measure) algorithm is described in section 3. In section 4, the semantic similarity computation using various distance measures are described. In section 5, the experimental results and analysis are discussed. In section 6, the conclusion and the future enhancements are depicted.

2 RELATED WORKS

This section discusses the related works. Peter Gardenfors [6] introduced the design of conceptual space which has the geographical features of the geo-spatial concepts. The semantic similarity model that uses the conceptual space as the knowledge representation is called as the Geometric model. Based on the matching geographical features, the semantic similarity is computed called as the Feature model proposed by Rodriguez [7] and Egenhofer. Rada et al [8] proposed the semantic network which is used in Network model [9] considers only the spatial relations. Goldstone [10] introduced the Alignment model which considers the alignable features for computing the semantic similarity. The Transformational model uses the edit distance method which was introduced by Levenshtein [11] transforms one geo-spatial concept to another. Based on the number of transformations, the semantic similarity is computed. The above said semantic similarity models uses either geographical features or the spatial relations for representing the geo-spatial concepts and does not yield better information retrieval. Hence Schwering [4] introduced the Hybrid model whose semantic description has the geographical features and the spatial relations. The Hybrid model proposed by Schwering uses the Euclidean distance measure for computing the semantic distance which does not yield better precision, recall, f-measure and relevant information retrieval for most of the geo-spatial query concept. Hence various hybrid distance measure are tried for the Hybrid model.

The Euclidean distance measure [5], [12], Manhattan distance measure [12], [13] and the vector cosine method [12] are used for the Image Retrieval Application. The Image Retrieval Application is developed using five histograms for which the distance measures are applied for retrieving the relevant images. The five histograms used are RGB histogram [14], an HSV-based histogram [15], Jain and Vailaya’s histogram [16], an HSV-based histogram with soft decision [17] and the histogram used in the QBIC system [18]. For this Image Retrieval Application the Manhattan distance measure yields good precision for all the five histograms. The Bray Cutis distance [19] measure is used in the clustering applications which provides better results. The next section discuss in detail about the HDSM algorithm.

3 HDSM ALGORITHM

The HDSM (Hybrid Distance Similarity Measure) algorithm is given below depicts how various semantic distance measures such as Hybrid Euclidean distance measure, Hybrid Manhattan distance measure, Hybrid Vector Cosine distance measure and Hybrid Bray Curtis distance measure are applied to the Hybrid model of the geo-spatial domain. The geo-spatial query concept is pre-processed. Then the dimensional and the relational semantic distance computation are done using various distance measures. The semantic distance computation is done using the dimensional and the relational semantic distance as the distance measures are applied to the Hybrid semantic similarity computation model. The semantically similar concepts are retrieved for which the documents are retrieved and the performance metrics such are precision, recall and f-measure are calculated.

Here α and β are the ratios which are used for the non-matching and matching dimensions and relations respectively. The α and β values ranges from 0 to 1. The various values of α and β are tested for computing semantic similarity and relevant information retrieval. When α value is 0.8 and β value is 0.2 it yields good results and it also yields good information retrieval. The prediction of the best semantic distance measure for the geo-spatial domain is made with the evaluation of the performance metrics such as precision, recall and f-measure. Precision is the ability of the geo-spatial information retrieval system to retrieve more number of the relevant concepts from the retrieved concepts of the geo-spatial query concept. Recall is the ability of the geo-spatial information system to retrieve the concepts from the datasource. F-measure is the accuracy of the precision and the recall obtained for the geo-spatial query concept.

Algorithm 1: HDSM Algorithm

HDSM algorithm (QC, DS, SV, GO)
// DS is the Datasource, SV is the Shared Vocabulary and GO is Geo-spatial Ontology.
Input: Query Concept QC
Output: Retrieval of the related geo-spatial concepts.

Step 1: Pre-processing

Pre-processing of the geo-spatial Query concept QC.
Based on the dimensions available in GO, the related geo-spatial concepts from DS are retrieved.
The dimensional and relational quantitative data are obtained from SV.

Step 2: Dimensional semantic distance computation using various distance metrics.

For all the extracted relevant geo-spatial concepts from DS
For all the QC dimensions and the DS concept's dimensions
If the dimensions are matching
Compute $\text{Euclideandistance}_{dim} = \sqrt{(qd_1 - dd_1)^2 + ... + (qd_n - dd_n)^2}$ // qd is the quantitative dimensional value of the Query concept. dd is the quantitative dimensional value of the datasource concepts.
Compute $\text{Manhattandistance}_{dim} = |qd_1 - dd_1| + ... + |qd_n - dd_n|$
Compute $\text{VectorCosinedistance}_{dim} = \dfrac{(qd_1 \times dd_1) + ... + (qd_n \times dd_n)}{\sqrt{(qd_1)^2 + ... + (qd_n)^2} \times \sqrt{(dd_1)^2 + ... + (dd_n)^2}}$
Compute $\text{BrayCurtisdistance}_{dim} = \dfrac{|qd_1 - dd_1| + ... + |qd_n - dd_n|}{|qd_1 + dd_1| + ... + |qd_n + dd_n|}$
Else compute standardised distance// for the non-matching dimensions, the concept that have dimensions are aggregated.
End if
$\text{Semdist}_{Eucliddim} = \alpha * \text{standarddistance} + \beta * \text{Euclideandistance}_{dim}$
$\text{Semdist}_{Man\,dim} = \alpha * \text{standarddistance} + \beta * \text{Manhattan distance}_{dim}$
$\text{Semdist}_{Vectordim} = \alpha * \text{standarddistance} + \beta * \text{VectorCosinedistance}_{dim}$
$\text{Semdist}_{Braydim} = \alpha * \text{standarddistance} + \beta * \text{BrayCurtisdistance}_{dim}$
End for (all dimensions).

Step 3: Relational semantic distance computation using various distance metrics.

For all the extracted relevant geo-spatial concepts from DS
For all the QC relations and the DS concept's relations
If the relations are matching
Compute $\text{Euclideandistance}_{rel} = \sqrt{(qr_1 - dr_1)^2 + ... + (qr_n - dr_n)^2}$ // qr is the quantitative relational value of the Query concept. dr is the quantitative relational value of the datasource concepts.
Compute $\text{Manhattandistance}_{rel} |qr_1 - dr_1| + ... + |qr_n - dr_n|$
Compute $\text{VectorCo sinedis tance}_{rel} \dfrac{(qr_1 \times dr_1) + ... + (qr_n \times dr_n)}{\sqrt{(qr_1)^2 + ... + (qr_n)^2} \times \sqrt{(dr_1)^2 + ... + (dr_n)^2}}$
Compute $\text{BrayCurtis distance}_{rel} = \dfrac{|qr_1 - dr_1| + ... + |qr_n - dr_n|}{|qr_1 + dr_1| + ... + |qr_n + dr_n|}$
Else compute standardised distance.
End if
$\text{Semdist}_{Euclidrel} = \alpha * \text{standarddistance} + \beta * \text{Euclideandistance}_{rel}$
$\text{Semdist}_{Manrel} = \alpha * \text{standarddistance} + \beta * \text{Manhattan distance}_{rel}$
$\text{Semdist}_{Vectorrel} = \alpha * \text{standarddistance} + \beta * \text{Vectordistance}_{rel}$
$\text{Semdist}_{Brayrel} = \alpha * \text{standard distance} + \beta * \text{BrayCurtis distance}_{rel}$
End for (all relations).

End for (all extracted relevant concepts).

Step 4: Semantic distance computation

//DSC is Datasource Concept

$$\mathrm{Semdist}_{\mathrm{Euclid}}(QC,DSC) = \sqrt{[\mathrm{Semdist}_{\mathrm{Eucliddim}} - \mathrm{Semdist}_{\mathrm{Euclidrel}}]^2}$$

$$\mathrm{Semdist}_{\mathrm{Man}}(QC,DSC) = |\mathrm{Semdist}_{\mathrm{Mandim}} - \mathrm{Semdist}_{\mathrm{Manrel}}|$$

$$\mathrm{Semdist}_{\mathrm{Vector}}(QC,DSC) = \frac{\mathrm{Semdist}_{\mathrm{Vectordim}} + \mathrm{Semdist}_{\mathrm{Vectorrel}}}{\sqrt{[\mathrm{Semdist}_{\mathrm{Vectordim}}]^2} \times \sqrt{[\mathrm{Semdist}_{\mathrm{Vectorrel}}]^2}}$$

$$\mathrm{Semdist}_{\mathrm{Bray}}(QC,DSC) = \frac{|[\mathrm{Semdist}_{\mathrm{Braydim}}] - [\mathrm{Semdist}_{\mathrm{Brayrel}}]|}{|[\mathrm{Semdist}_{\mathrm{Braydim}}] + [\mathrm{Semdist}_{\mathrm{Brayrel}}]|}$$

//Less the semantic distance, more the semantic similarity. More the semantic distance, less is the semantic similarity.

The relevant semantically similar concepts are retrieved.

Step 5: Retrieval of documents for the semantically similar concepts.

//Rlc is relevant concepts and Rtc is retrieved concepts.

Compute precision, recall and f-measure where

precision=|Rlc ∩ Rtc|/ Rtc,

recall = |Rlc ∩ Rtc|/Rlc and

f-measure = (2*(precision*recall))/(precision + recall).

This section discussed in detail about the HDSM algorithm and also depicts how the various distance measures such as Hybrid Euclidean distance measure, Hybrid Manhattan distance measure, Hybrid Vector Cosine distance measure and Hybrid Bray Curtis distance measure are applied to the Hybrid Model of the geo-spatial information system. The next section discuss about the computation of each distance measure for the sample geo-spatial hydrological ontology and the retrieval of the relevant geo-spatial concepts for the specified geo-spatial query concept.

4 SEMANTIC SIMILARITY COMPUTATION USING VARIOUS DISTANCE MEASURES FOR THE HYBRID MODEL

This section discuss about the Hybrid semantic similarity computational model and the application of various semantic distance measures such as Hybrid Euclidean distance measure, Hybrid Manhattan distance measure, Hybrid Vector Cosine distance measure and Hybrid Bray Curtis distance measure for the Hybrid model of the geo-spatial domain and also depicts the sample calculation of each distance measure.

4.1 EXISTING HYBRID SEMANTIC SIMILARITY COMPUTATIONAL MODEL

Gardenfors introduced the design of the conceptual space which provides the geo-spatial concepts and its corresponding features. Rada et al proposed the semantic networks which uses the relationship between the geo-spatial concepts. The Hybrid model proposed by Schwering combined the work of Gardenfors and Rada et al and this model uses the Euclidean distance measure for computing the semantic distance among the geo-spatial concepts which does not yield better precision. The conceptual space and the semantic network are used for computing the semantic similarity in the Hybrid model. In this Model the dimensions or the properties are considered as the vector points in the conceptual hull.

The Euclidean semantic distance measure is used for computing the semantic distance between the matching vector points of the geo-spatial query concept and the relevant concepts. In the Fig. 1 the query concept hull is watercourse and the related concept hull is channel. q_1, q_2 and q_3 are the matching dimensional vector points between the geo-spatial query concept and the related geo-spatial concept. d_1, d_2 and d_3 are the semantic distances between the dimensions. The other distance measures are not attempted for the geo-spatial domain for computing the semantic distance. Hence this paper attempts various semantic distance measures for the Hybrid model. The next section discuss about the semantic distance measures.

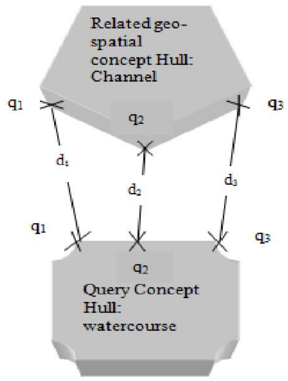

Fig. 1. Hybrid Model Conceptual Hull

4.2 PROPOSED SEMANTIC DISTANCE MEASURES

The proposed semantic distance measures for Hybrid model are

- Hybrid Euclidean distance measure
- Hybrid Manhattan distance measure
- Hybrid Vector Cosine distance measure and
- Hybrid Bray Curtis distance measure

These semantic distance measures are used for computing the semantic distance among the geo-spatial concepts. These semantic distance measures are the distance measures which are used for calculating the distance between two points in the n-dimensional space. The geo-spatial ontology [20] is needed for providing the semantics of the geo-spatial domain. The geo-spatial ontology used for experiments in this paper is Hydrology.owl which is the Ordnance Survey ontology. The datasource used for experiments in this paper is Ordnance Survey Master Map. The Ordnance Survey Master Map has four layers such as Topography Layer, OSMasterMap Address Layer, Imagery Layer, and Integrated Transport Network (ITN) Layer provides the information about the topography, address of the location, aerial images and the vehicle movement tracking. For viewing the Ordnance Survey Master Map, the Geographic Mark-up language viewer is used. The shared vocabulary converts the qualitative dimensional value into quantitative dimensional value. The sample ontology is given by Fig.2, for which the various semantic distance measures are applied and the sample calculation are done and they are tested for the information retrieval and the prediction of the best semantic distance measure for the Hybrid model of the geo-spatial domain is done. Here 'pond' is the geo-spatial query concept considered. Fig.2 shows the related geo-spatial concepts of the query concept pond.

For computing the semantic distance between the geo-spatial query concept and its related concepts, the dimensional values of the geo-spatial concepts are needed which is given by the shared vocabulary. Totally 21 dimensions such as elevation, gradient, gauge level, altitude difference, height relative to the environment, soil moisture, salinity, number of days waterlogged, etc,. Here for the sample calculation three dimensions are considered for which the dimensional values are given by the Table 1.

Table 1. Geo-spatial concepts and its dimensional values.

Geo-spatial Concepts	Dimensions		
	Elevation(meters)	Gradient (%)	Altitude Difference(meters)
River	610	79	80
Pond	600	78	56
Stream	760	84	78
Nant	670	67	89
Sea	800	77	90
watercourse	277	73	87
Lake	345	78	60
Dam	890	75	67

The every semantic distance measure is applied to the sample tree shown in the Fig.2 and for the dimensional values given by the Table 1. The next sub-sections discuss the distance measures and its application in the geo-spatial domain.

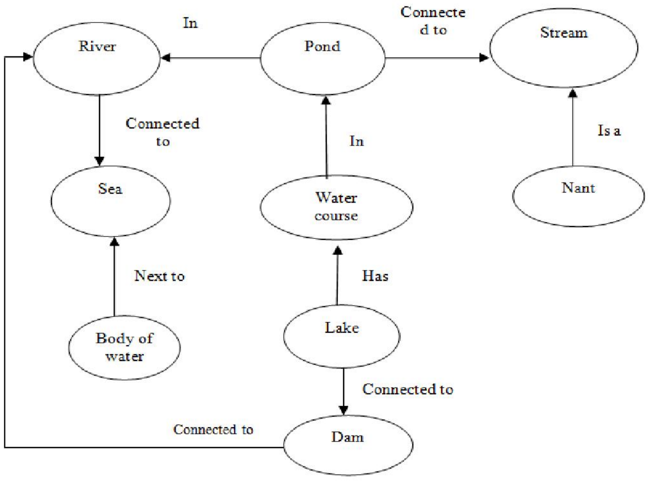

Fig. 2. Sample tree containing the related concepts of the query concept 'Pond'

4.2.1 HYBRID EUCLIDEAN DISTANCE MEASURE

This section discuss about the Hybrid Euclidean distance measure and its application in the geo-spatial domain. The Euclidean distance method is used for computing the distance between two points in n-dimensional space. The general Euclidean distance is given by equation (1).

$$\text{Euclideandistance}(p,q)=\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+....+(p_n-q_n)^2} \quad (1)$$

where $p_1, p_2, ..., p_n$ and $q_1, q_2, ..., q_n$ are the points in the n-dimensional space. This general equation of the Euclidean distance is mapped to the geo-spatial domain which is given by the equations (2), (3) and (4). As said above this distance measure is applied for the Hybrid model which will be considering both the properties or the dimensions and the spatial relations for representing the geo-spatial concepts. For the geo-spatial query concept given, the related geo-spatial concepts are obtained from the datasource based on the semantics in the geo-spatial ontology. The dimensional values of the needed geo-spatial concepts are obtained from the shared vocabulary. For all the matching and the non-matching dimensions of the query concept and the related geo-spatial concept, the dimensional semantic distance is computed which is given by equation (2) and (3).

$$\text{Euclidean distance}_{\text{dim}} = \sqrt{(qd_1 - dd_1)^2 + ... + (qd_n - dd_n)^2} \quad (2)$$

where qd is the quantitative dimensional value of the Query concept. dd is the quantitative dimensional value of the datasource concepts.

$$\text{Semdist}_{\text{Eucliddim}} = \alpha * \text{standarddistance} + \beta * \text{Euclideandistance}_{\text{dim}} \quad (3)$$

where standardised distance is for the non-matching dimensions, the geo-spatial concepts that have dimensions are aggregated. The α and β are the ratios of the matching and the non matching dimensions. The α value is 0.8 multiplied to the non-matching dimensions and β value is 0.2 multiplied to the matching dimensions. Likewise for the matching and the non-matching relations, the relational semantic distance is computed which is given by the equation (4).

$$\text{Semdist}_{\text{Euclidrel}} = \alpha * \text{standarddistance} + \beta * \text{Euclideandistance}_{\text{rel}} \quad (4)$$

The semantic distance is computed by aggregating the dimensional and the relational semantic distance is given by equation (5).

$$\text{Semdist}_{\text{Euclid}}(QC, DSC) = \sqrt{[\text{Semdist}_{\text{Eucliddim}} - \text{Semdist}_{\text{Euclidrel}}]^2} \quad (5)$$

where QC is the Query Concept and DSC is the datasource concept.

Computation using Hybrid Euclidean distance method

Consider the sample tree shown in the Fig.2 and the dimensional values given in the Table 1 for the geo-spatial concepts. Pond is the query concept considered here. The related concept to the query concept pond is watercourse. So the dimensional semantic distance between pond and watercourse is calculated using equation (2) and (3). The α and β values took are 0.8 and 0.2 respectively. Hence,

$$\text{Semdist}_{\text{Eucliddim}} = 0.8 * 0 + 0.2 * \sqrt{(600-277)^2 + (78-73)^2 + (56-87)^2} = 66.55$$

The geo-spatial concept watercourse has its relation with the lake so dimensions of the watercourse and lake are considered.

$$\text{Semdist}_{\text{Euclidrel}} = 0.8 * 0 + 0.2 * \sqrt{(277-345)^2 + (73-78)^2 + (87-60)^2} = 14.66$$

The semantic distance computation is done by aggregating the dimensional and the relational semantic distance using equation (5).

$$\text{Semdist}_{\text{Euclid}}(\text{pond}, \text{watercours e}) = \sqrt{(66.55-14.66)^2} = 5.189$$

Here all the dimensions are matching and so the standardised distance is 0. The semantic distance between pond and watercourse is 5.189 when three dimensions are considered. For the experiments conducted in this paper 21 dimensions are considered. The query concept pond is connected to stream so the semantic distance between the pond and stream is computed.

$$\text{Semdist}_{\text{Eucliddim}} = 0.8 * 0 + 0.2 * \sqrt{(600-760)^2 + (78-84)^2 + (56-78)^2} = 32.32$$

The geo-spatial concept stream has its relation with the Nant so dimensions of the stream and nant are considered.

$$\text{Semdist}_{\text{Euclidrel}} = 0.8 * 0 + 0.2 * \sqrt{(760-670)^2 + (84-67)^2 + (78-89)^2} = 18.44$$

The semantic distance computation is done by aggregating the dimensional and the relational semantic distance using equation (5).

$$\text{Semdist}_{\text{Euclid}}(\text{pond}, \text{stream}) = \sqrt{(32.32-18.44)^2} = 1.388$$

The semantic distance between pond and stream is 1.388. The semantic distance between pond and watercourse is 5.189. Thus the geo-spatial concept stream is more similar to the query concept pond. Likewise the semantic distance is computed for all the related concepts of the geo-spatial query concept. Then the performance metrics such as precision, recall and f-measure are computed for evaluation. The next sub-section discuss about the Manhattan distance measure.

4.2.2 HYBRID MANHATTAN DISTANCE MEASURE

This section discuss about how the Hybrid Manhattan distance measure is used for computing the semantic similarity between the geo-spatial concepts. The general Manhattan distance is given by the equation (6).

$$\text{Manhattan distance }(p,q) = |p_1 - q_1| + |p_2 - q_2| + \ldots\ldots\ldots + |p_n - q_n| \quad (6)$$

where $p_1, p_2, \ldots p_n$ and $q_1, q_2, \ldots\ldots q_n$ are the points in the n-dimensional space. The mapping of the general Manhattan distance to the geo-spatial domain is given by equations (7), (8), (9) and (10). The matching and the non-matching dimensions are considered. For the matching dimensions the Manhattan distance is used for computation which is given by equation (7).

$$\text{Compute } \text{Manhattandistance}_{\text{dim}} = |qd_1 - dd_1| + \ldots\ldots + |qd_n - dd_n| \quad (7)$$

where qd is the quantitative dimensional value of the Query concept. dd is the quantitative dimensional value of the datasource concepts.

If the dimensions are not matching then the standardised distance is considered. The dimensional semantic distance is given by equation (8).

$$\text{Semdist}_{\text{Man dim}} = \alpha * \text{standard distance} + \beta * \text{Manhattan distance}_{\text{dim}} \tag{8}$$

Likewise the relational semantic distance is computed and given by the equation (9).

$$\text{Semdist}_{\text{Manrel}} = \alpha * \text{standard distance} + \beta * \text{Manhattan distance}_{\text{rel}} \tag{9}$$

The Hybrid computational model aggregates the dimensional and the relational distance which is given by equation (10).

$$\text{Semdist}_{\text{Man}}(\text{QC,DSC}) = |\text{Semdist}_{\text{Mandim}} - \text{Semdist}_{\text{Manrel}}| \tag{10}$$

where QC is the Query Concept and DSC is the datasource concept.

Computation using Hybrid Manhattan distance method

For the semantic similarity computation, Fig.2 and Table 1 are considered for the sample tree consisting of the geo-spatial concepts and the dimensional values of the geo-spatial concepts. The query concept pond is considered for which the related geo-spatial concept is watercourse and so the semantic distance between pond and watercourse must be computed. The Manhattan dimensional semantic distance is computed using equations (7) and (8).

$$\text{Semdist}_{\text{Man dim}} = 0.8*0+0.2*[|600 – 277|+|78 – 73|+|56 - 87| = 71.8$$

The geo-spatial concept watercourse has its relation with the lake so dimensions of the watercourse and lake are considered. The relational semantic distance is computed using equation (9).

$$\text{Semdist}_{\text{Manrel}} = 0.8*0+0.2*[|277 - 345|+|73 - 78|+|87 - 60| = 20$$

The dimensional and the relational semantic distance are aggregated in the Hybrid semantic similarity computational model for computing the semantic distance which is given by the equation (10).

$$\text{Semdist}_{\text{Man}} \text{ (Pond, watercourse)} = |71.8-20| = 51.8/10=5.18$$

The query concept pond is connected to stream, so the semantic distance between the pond and stream is computed.

$$\text{Semdist}_{\text{Man dim}} = 0.8*0+0.2*[|600 – 760|+|78 – 84|+|56 - 78| = 37.6$$

The geo-spatial concept stream has its relation with the nant. So dimensions of the stream and nant are considered.

$$\text{Semdist}_{\text{Manrel}} = 0.8*0+0.2*[|760 - 670|+|84 - 67|+|78 - 89| = 23.4$$

$$\text{Semdist}_{\text{Man}} \text{ (Pond, stream)} = |37.6 – 23.4| = 14.2/10 = 1.42$$

The semantic distance between pond and stream is 1.42. The semantic distance between pond and watercourse is 5.18. Thus the geo-spatial concept stream is more similar to the query concept pond. Likewise the semantic distance is computed for all the related concepts of the geo-spatial query concept. Then the precision, recall and f-measure are computed. The next sub-section discuss about the Vector cosine distance measure.

4.2.3 HYBRID VECTOR COSINE DISTANCE MEASURE

This section discusses how the Vector Cosine distance can be applied to the geo-spatial domain. The general vector cosine distance is given by equation (11).

$$\text{Vector Cosine distance}(p, q) = \frac{(p_1 \times q_1) + (p_2 \times q_2) + ... + (p_n \times q_n)}{\sqrt{(p_1)^2 + ... + (p_n)^2} \times \sqrt{(q_1)^2 + ... + (q_n)^2}} \tag{11}$$

where $p_1, p_2, ...p_n$ and $q_1, q_2, ...q_n$ are the points in the n-dimensional space.

This general Vector Cosine distance mapped to the geo-spatial information retrieval system which is given by the below equations (12), (13), (14) and (15).

The matching and the non-matching dimensions are considered. For the matching dimensions the Vector Cosine distance is used for computation which is given by equation (12).

$$\text{VectorCosinedistance}_{\dim} = \frac{(qd_1 \times dd_1) + \ldots\ldots + (qd_n \times dd_n)}{\sqrt{(qd_1)^2 + \ldots. + (qd_n)^2} \times \sqrt{(dd_1)^2 + \ldots. + (dd_n)^2}} \quad (12)$$

where qd is the quantitative dimensional value of the Query concept. dd is the quantitative dimensional value of the datasource concepts.

If the dimensions are not matching then the standardised distance is considered. The dimensional semantic distance is given by equation (13).

$$\text{Semdist}_{\text{Vectordim}} = \alpha * \text{standarddistance} + \beta * \text{VectorCosinedistance}_{\dim} \quad (13)$$

The relational semantic distance is computed and given by the equation (14).

$$\text{Semdist}_{\text{Vectorrel}} = \alpha * \text{standarddistance} + \beta * \text{Vectordistance}_{\text{rel}} \quad (14)$$

The Hybrid Model combines the dimension and the relation which is given by equation (15).

$$Semdist_{Vector}(QC, DSC) = \frac{Semdist_{Vector\dim} \times Semdist_{Vectorrel}}{\sqrt{[Semdist_{Vector\dim}]^2} \times \sqrt{[Semdist_{Vectorrel}]^2}} \quad (15)$$

where QC is the Query Concept and DSC is the datasource concept.

Computation using Hybrid Vector Cosine distance method

For the semantic similarity computation, Fig.2 and Table 1 are considered for the sample tree consisting of the geo-spatial concepts and the dimensional values of the geo-spatial concepts. The query concept pond is considered for which the related geo-spatial concept is watercourse and so the semantic distance between pond and watercourse must be computed. The Vector Cosine dimensional semantic distance is computed using equations (12) and (13).

$$\text{Semdist}_{\text{Vectordim}} = 0.8 * 0 + 0.2 * \left[\frac{(600 \times 277) + (78 \times 73) + (56 \times 87)}{\sqrt{(600)^2 + (78)^2 + (56)^2} \times \sqrt{(277)^2 + (73)^2 + (87)^2}} \right] = 0.194$$

The geo-spatial concept watercourse has its relation with the lake so dimensions of the watercourse and lake are considered. The relational semantic distance is computed using equation (14).

$$\text{Semdist}_{\text{Vectorrel}} = 0.8 * 0 + 0.2 * \left[\frac{(277 \times 345) + (73 \times 78) + (87 \times 60)}{\sqrt{(277)^2 + (73)^2 + (87)^2} \times \sqrt{(345)^2 + (78)^2 + (60)^2}} \right] = 0.99.$$

The semantic distance aggregation is done using the equation (15).

$$\text{Semdist}_{\text{Vector}}(\text{pond}, \text{watercourse}) = \frac{0.194 + 0.99}{\sqrt{(0.194)^2} \times \sqrt{(0.99)^2}} = 6.16$$

The query concept pond is connected to stream, so the semantic distance between the pond and stream is computed.

$$\text{Semdist}_{\text{Vectordim}} = 0.8 * 0 + 0.2 * \left[\frac{(600 \times 760) + (78 \times 84) + (56 \times 78)}{\sqrt{(600)^2 + (78)^2 + (56)^2} \times \sqrt{(760)^2 + (84)^2 + (78)^2}} \right] = 0.199$$

The geo-spatial concept stream has its relation with the nant. So dimensions of the stream and nant are considered.

$$\text{Semdist}_{\text{Vectorrel}} = 0.8 * 0 + 0.2 * \left[\frac{(760 \times 670) + (84 \times 67) + (78 \times 89)}{\sqrt{(760)^2 + (84)^2 + (78)^2} \times \sqrt{(670)^2 + (67)^2 + (89)^2}} \right] = 0.199$$

$$\text{Semdist}_{\text{Vector}}(\text{pond}, \text{stream}) = \frac{0.199 + 0.199}{\sqrt{(0.199)^2} \times \sqrt{(0.199)^2}} = 1.02$$

The semantic distance between pond and watercourse is 6.16 and the semantic distance between pond and stream is 1.02. The stream has the less semantic distance and so stream is more relevant concept to the given geo-spatial query concept pond. Then the precision, recall and f-measure are computed. The next sub-section discuss about the Bray Curtis distance measure.

4.2.4 HYBRID BRAY CURTIS DISTANCE MEASURE

This section discusses how the Bray Curtis distance can be incorporated into the geo-spatial domain. The general Bray Curtis distance is given by equation (16).

$$\text{Bray Curtis distance}(p,q) = \frac{|p_1 - q_1| + |p_2 - q_2| + \ldots\ldots + |p_n - q_n|}{|p_1 + q_1| + |p_2 + q_2| + \ldots\ldots + |p_n + q_n|} \quad (16)$$

where $p_1, p_2, \ldots p_n$ and $q_1, q_2, \ldots\ldots q_n$ are the points in the n-dimensional space. The mapping of the general Bray Curtis distance to the geo-spatial domain is given by the equations (17), (18), (19) and (20). The matching and the non-matching dimensions are considered. For the matching dimensions the Bray Curtis distance is used for computation which is given by equation (17).

$$\text{BrayCurtisdistance}_{dim} = \frac{|qd_1 - dd_1| + \ldots\ldots + |qd_n - dd_n|}{|qd_1 + dd_1| + \ldots + |qd_n + dd_n|} \quad (17)$$

where qd is the quantitative dimensional value of the Query concept. dd is the quantitative dimensional value of the datasource concepts.

If the dimensions are not matching then the standardised distance is considered. The dimensional semantic distance is given by equation (18).

$$\text{Semdist}_{Braydim} = \alpha * \text{standarddistance} + \beta * \text{BrayCurtisdistance}_{dim} \quad (18)$$

The relational semantic distance is computed and given by the equation (19).

$$\text{BrayCurtis distance}_{rel} = \frac{|qr_1 - dr_1| + \ldots\ldots + |qr_n - dr_n|}{|qr_1 + dr_1| + \ldots + |qr_n + dr_n|} \quad (19)$$

The Hybrid computational model aggregates the dimensional and the relational distance which is given by equation (20).

$$\text{Semdist}_{Bray}(QC, DSC) = \frac{|[\text{Semdist}_{Bray\,dim}] - [\text{Semdist}_{Brayrel}]|}{|[\text{Semdist}_{Bray\,dim}] + [\text{Semdist}_{Brayrel}]|} \quad (20)$$

where QC is the Query Concept and DSC is the datasource concept.

Computation using Hybrid Bray Curtis distance method

For the semantic similarity computation, Fig.2 and Table 1 are considered for the sample tree consisting of the geo-spatial concepts and the dimensional values of the geo-spatial concepts. The query concept pond is considered for which the related geo-spatial concept is watercourse and so the semantic distance between pond and watercourse must be computed. The Bray Curtis dimensional semantic distance is computed using equations (17) and (18).

$$\text{Semdist}_{Braydim} = 0.8 * 0 + 0.2 * \left[\frac{|600-277| + |78-73| + |56-87|}{|600+277| + |78+73| + |56+87|} \right] = 0.06.$$

The geo-spatial concept watercourse has its relation with the lake so dimensions of the watercourse and lake are considered. The relational semantic distance is computed using equation (19).

$$\text{Semdist}_{Brayrel} = 0.8 * 0 + 0.2 * \left[\frac{|277-345| + |73-78| + |87-60|}{|277+345| + |78+73| + |60+87|} \right] = 0.0217$$

$\text{Semdist}_{Bray}(\text{pond, watercourse}) = \frac{|0.06 - 0.0217|}{|0.06 + 0.0217|}$ = 0.46*10=4.6. The query concept pond is connected to stream, so the semantic distance between the pond and stream is computed.

$$\text{Semdist}_{Braydim} = 0.8 * 0 + 0.2 * \left[\frac{|600-760| + |78-84| + |56-78|}{|600+760| + |78+84| + |56+78|} \right] = 0.022.$$

The geo-spatial concept stream has its relation with the nant. So dimensions of the stream and nant are considered.

$$Semdist_{Brayrel} = 0.8*0+0.2*\left[\frac{|760-670|+|84-67|+|78-89|}{|760+670|+|84+67|+|78+89|}\right] = 0.0135$$

$$Semdist_{Bray}(pond, stream) = \frac{|0.022-0.0135|}{|0.022+0.0135|} = 0.23*10=2.3$$

The semantic distance between pond and watercourse is 4.6 and the semantic distance between pond and stream is 2.3. The stream has the less semantic distance and so stream is more relevant concept to the given geo-spatial query concept pond. All the distance measure depicts that stream is more relevant than the watercourse but the semantic distance varies for every semantic distance measure. Based on the semantic distance, the precision, recall and the f-measure values are computed. The next section discusses the experimental results and analysis.

5 EXPERIMENTAL RESULTS AND ANALYSIS

This section discuss in detail about the experimental results and analysis. The experiments are conducted using Netbeans 7.2.1 IDE. The geo-spatial datasource and the ontology are used for capturing the knowledge about the domain. Ordnance Survey Master Map is the datasource used for conducting experiments. The Geographic Mark-up Language (GML) viewer is used for viewing the Ordnance Survey Master Map. For conducting experiments 108 geo-spatial concepts, 20 relations and 21 dimensions or properties are considered. The various distance based measures are applied to the Hybrid model. The performance metrics used are precision, recall and f-measure. The performance metric values for 10 sample relevant geo-spatial concepts for the geo-spatial query concept using various semantic distance measures are shown in Table 2. The precision, recall and the f-measure values are computed for various similarity thresholds. Fig.3 depicts that the Hybrid Manhattan distance measure has yielded the maximum precision of 0.95 at the maximum similarity threshold 5. Then the Hybrid Euclidean distance measure has yielded the precision of 0.8 at the maximum similarity threshold likewise the Vector Cosine distance measure and Bray Curtis similarity measure yields the precision of 0.70 and 0.77 at the maximum similarity threshold. The Fig.4 depicts that except the Hybrid Manhattan distance measure, other hybrid distance measures mostly yield the recall of 1 for varying similarity threshold. Hybrid Vector cosine distance measure yields the maximum recall. The f-measure is computed which is maximum for the Hybrid Manhattan distance measure. Hence from the Figure 3, 4 and 5 it is depicted that the Hybrid Manhattan distance measure yields good precision, recall, f-measure and the relevant information retrieval as the Hybrid Manhattan distance measure retrieves all the relevant neighbour geo-spatial concepts.

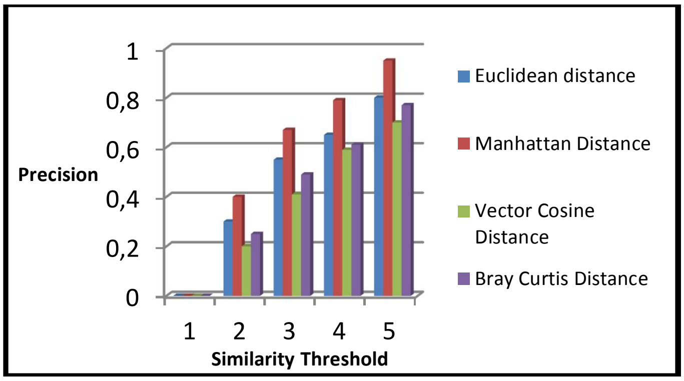

Fig. 3. Average precision yielded by each distance measure

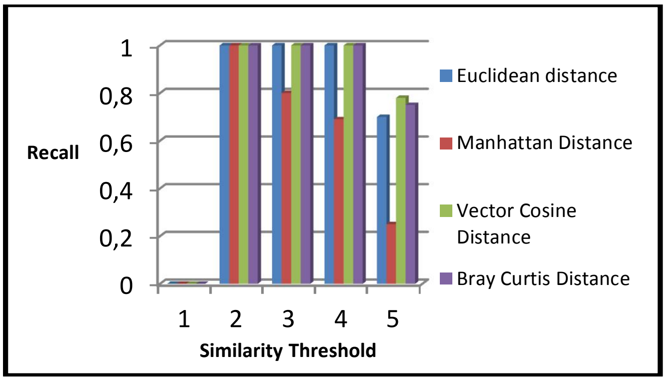

Fig. 4. Average recall yielded by ach distance measure

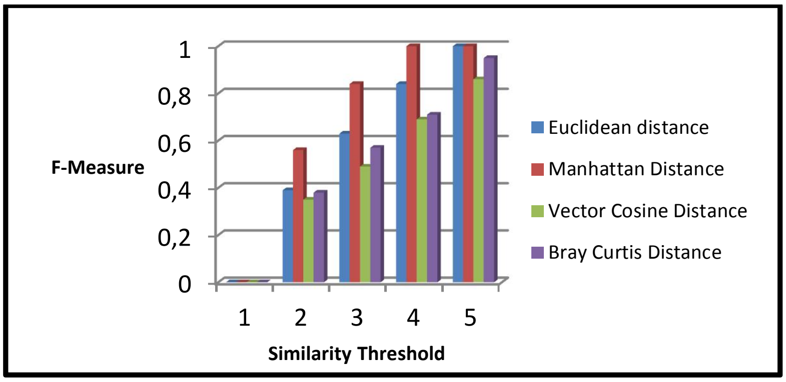

Fig. 5. Average f-measure yielded by each semantic distance measure

Table 2. Semantic distance measures and its performance metrics for the query concept 'Pond'

Semantic Distance Measure	Geo-spatial concepts	Precision				Recall				F-Measure			
		Similarity Thresholds				Similarity Thresholds				Similarity Thresholds			
		2	3	4	5	2	3	4	5	2	3	4	5
Euclidean Measure	Floodplain	0.21	0.39	0.52	0.63	1.00	1.00	1.00	0.77	0.35	0.52	0.69	0.87
	River	0.19	0.46	0.63	0.79	1.00	1.00	1.00	0.69	0.32	0.48	0.64	0.80
	Canal	0.22	0.45	0.55	0.8	1.00	1.00	1.00	0.69	0.36	0.54	0.72	0.90
	Sea	0.26	0.5	0.56	0.81	1.00	1.00	1.00	0.69	0.41	0.62	0.83	1.03
	Lake	0.31	0.48	0.51	0.65	1.00	1.00	1.00	0.76	0.47	0.71	0.95	1.18
	Water course	0.29	0.47	0.49	0.59	1.00	1.00	1.00	0.9	0.45	0.67	0.90	1.12
	Dam	0.23	0.42	0.6	0.75	1.00	1.00	1.00	0.68	0.37	0.56	0.75	0.93
	Channel	0.32	0.47	0.59	0.81	1.00	1.00	1.00	0.69	0.48	0.73	0.97	1.21
	Stream	0.23	0.53	0.62	0.82	1.00	1.00	1.00	0.62	0.37	0.56	0.75	0.93
	Backwater	0.27	0.39	0.51	0.63	1.00	1.00	1.00	0.76	0.43	0.64	0.85	1.06
	Pool	0.22	0.51	0.68	0.83	1.00	1.00	1.00	0.59	0.36	0.54	0.72	0.90
	Reservoir	0.29	0.54	0.65	0.81	1.00	1.00	1.00	0.69	0.45	0.67	0.90	1.12
Manhattan Distance	Floodplain	0.39	0.58	0.75	0.92	1.00	0.81	0.71	0.17	0.56	0.84	1.12	1.40
	River	0.28	0.49	0.6	0.9	1.00	0.79	0.6	0.2	0.44	0.66	0.88	1.09
	Canal	0.3	0.5	0.6	0.91	1.00	0.82	0.72	0.36	0.46	0.69	0.92	1.15
	Sea	0.38	0.55	0.65	0.9	1.00	0.79	0.73	0.2	0.55	0.83	1.10	1.38
	Lake	0.42	0.62	0.8	0.92	1.00	0.8	0.65	0.18	0.59	0.89	1.18	1.48
	Water course	0.41	0.62	0.8	0.92	1.00	0.81	0.73	0.18	0.58	0.87	1.16	1.45
	Dam	0.39	0.6	0.79	0.9	1.00	0.81	0.71	0.2	0.56	0.84	1.12	1.40
	Channel	0.49	0.55	0.65	0.85	1.00	0.79	0.64	0.37	0.66	0.99	1.32	1.64
	Stream	0.39	0.58	0.79	0.9	1.00	0.8	0.73	0.2	0.56	0.84	1.12	1.40
	Backwater	0.43	0.63	0.81	0.91	1.00	0.84	0.71	0.19	0.60	0.90	1.20	1.50
	Pool	0.37	0.53	0.65	0.87	1.00	0.8	0.64	0.36	0.54	0.81	1.08	1.35
	Reservoir	0.39	0.58	0.68	0.88	1.00	0.81	0.75	0.36	0.56	0.84	1.12	1.40
Vector Cosine Distance Measure	Floodplain	0.19	0.3	0.48	0.65	1.00	1.00	1.00	0.8	0.32	0.48	0.64	0.80
	River	0.18	0.4	0.59	0.7	1.00	1.00	1.00	0.71	0.31	0.46	0.61	0.76
	Canal	0.19	0.4	0.5	0.69	1.00	1.00	1.00	0.71	0.32	0.48	0.64	0.80
	Sea	0.2	0.49	0.53	0.7	1.00	1.00	1.00	0.71	0.33	0.50	0.67	0.83
	Lake	0.25	0.45	0.52	0.7	1.00	1.00	1.00	0.8	0.4	0.6	0.8	1.00
	Water course	0.2	0.43	0.41	0.68	1.00	1.00	1.00	1.00	0.33	0.5	0.67	0.83
	Dam	0.19	0.39	0.5	0.72	1.00	1.00	1.00	0.7	0.32	0.48	0.64	0.80
	Channel	0.25	0.44	0.49	0.7	1.00	1.00	1.00	0.71	0.4	0.6	0.8	1
	Stream	0.19	0.5	0.6	0.7	1.00	1.00	1.00	0.8	0.32	0.48	0.64	0.80
	Backwater	0.2	0.35	0.49	0.65	1.00	1.00	1.00	0.82	0.33	0.50	0.67	0.83
	Pool	0.2	0.45	0.6	0.71	1.00	1.00	1.00	0.62	0.33	0.50	0.67	0.83
	Reservoir	0.25	0.5	0.6	0.71	1.00	1.00	1.00	0.75	0.40	0.60	0.80	1.00
Bray Curtis Distance Measure	Floodplain	0.2	0.35	0.5	0.71	1.00	1.00	1.00	0.78	0.33	0.50	0.67	0.83
	River	0.19	0.43	0.6	0.81	1.00	1.00	1.00	0.7	0.32	0.48	0.64	0.80
	Canal	0.2	0.42	0.54	0.7	1.00	1.00	1.00	0.7	0.33	0.50	0.67	0.83
	Sea	0.24	0.49	0.56	0.72	1.00	1.00	1.00	0.7	0.39	0.58	0.77	0.97
	Lake	0.3	0.46	0.52	0.75	1.00	1.00	1.00	0.78	0.46	0.69	0.92	1.15
	Water course	0.25	0.45	0.45	0.7	1.00	1.00	1.00	0.88	0.4	0.6	0.8	1
	Dam	0.22	0.4	0.58	0.75	1.00	1.00	1.00	0.69	0.36	0.54	0.72	0.90
	Channel	0.29	0.46	0.55	0.72	1.00	1.00	1.00	0.71	0.45	0.67	0.90	1.12
	Stream	0.2	0.52	0.61	0.73	1.00	1.00	1.00	0.7	0.33	0.50	0.67	0.83
	Backwater	0.24	0.37	0.5	0.72	1.00	1.00	1.00	0.8	0.39	0.58	0.77	0.97
	Pool	0.21	0.49	0.65	0.79	1.00	1.00	1.00	0.61	0.35	0.52	0.69	0.87
	Reservoir	0.27	0.52	0.64	0.79	1.00	1.00	1.00	0.72	0.43	0.64	0.85	1.06

6 CONCLUSION

Thus this paper has discussed HDSM algorithm which includes various Hybrid semantic distance measures such as Hybrid Euclidean distance measure, Hybrid Manhattan distance measure, Hybrid Vector Cosine distance measure and Hybrid Bray Cutis distance measure for the Hybrid Model (combines conceptual space containing the geographical features of geo-spatial concepts and the semantic network containing the spatial relations between the geo-spatial concepts) of the geo-spatial information system. It is inferred that the Manhattan distance measure for the Hybrid model yields 15% increase of precision, recall and f-measure. It yields better relevant information retrieval as it retrieves all the neighbour geo-spatial concepts. In this paper the semantic similarity measure for computing the relevance of information in done. In future the general approach of semantic similarity computation can be developed for inspecting the semantic similarity between the various application tasks and the similarity judgement of the application tasks can be analysed. The conceptual contexts are not attempted with the semantic distance measures for the geo-spatial domain so the conceptual context can be included with the semantic distance measures.

REFERENCES

[1] Kay Khaing Win, "Measuring Geospatial Semantic Similarity between Geospatial Entity Classes", University of Computer Studies, Yangon, Myanmar, *IEEE,* 2006.

[2] Schwering.A, "Hybrid model for semantic similarity measurement", *4th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE05)*, Agia Napa, Cyprus, Springer, 1449-1465, 2005.

[3] Resnik. P, "Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language", *Journal of Artificial Intelligence Research*, 11: 95-130, 1999.

[4] Schwering, A. and M. Raubal, "Measuring semantic similarity between geospatial conceptual regions", *1st International Conference on GeoSpatial Semantics*, Mexico City, Mexico, Springer, 93-109, 2005.

[5] Schwering, A. and M. Raubal, "Spatial relations for semantic similarity measurement", *2nd International Workshop on Conceptual Modeling for Geographic Information Systems (CoMoGIS2005)*, Klagenfurt, Austria., Springer, 259-269, 2005.

[6] Gärdenfors.P, "How to make the Semantic Web more semantic. Formal Ontology in Information Systems", *Proceedings of the Third International Conference (FOIS2004)*. L. Vieu. Amsterdam, NL, IOS Press. 114: 153-164, 2004.

[7] Rodríguez, A. and M. J. Egenhofer, "Comparing geospatial entity classes: An asymmetric and context-dependent similarity measure", *International Journal of Geographical Information Science*, 18(3): 229-256, 2004.

[8] Rada, R., H. Mili, E. Bicknell and M. Blettner, "Development and application of a metric on semantic nets", *IEEE Transactions on systems, man, and cybernetics*, 19(1):17-30, 1989.

[9] Resnik, P. "Using information content to evaluate semantic similarity in a taxonomy", *14th International Joint Conference on Artificial Intelligence (IJCAI)*, Montreal, Canada, Morgan Kaufmann, 448-453, 1995.

[10] Goldstone, R. L. and J. Son "Similarity - Cambridge Handbook of Thinking and Reasoning" K. Holyoak and R. Morrison. Cambridge, Cambridge University Press: 13-36, 2005.

[11] Levenshtein, I. V, "Binary codes capable of correcting deletions, insertions, and reversals", *Soviet Physics Doklady*, 10(8): 707-710, 1966.

[12] A. Vadivel, A.K. Majumdar, Shamik Sural, "Performance comparison of distance metrics in content-based Image retrieval applications" Computer Science and Engineering, Indian Institute of Technology, Kharagpur, India.

[13] Hosseinabady, M. and Nunez-Yanez, "Run-time stochastic task mapping on a large scale network-on-chip with dynamically reconfigurable tiles", J.L. Computers & Digital Techniques, Vol. 6, pp. 1-11, *IET Journals & Magazines*, 2012.

[14] M.J. Swain and D. H. Ballard, "Color Indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.

[15] S.Sural, G. Qian and S.Pramanik,"Segmentation and Histogram Generation using the HSV Color Space for Content Based Image Retrieval", *IEEE Int. Conf. on Image Processing*, Rochester, NY, 2002.

[16] A. Jain and A. Vailaya, "Image Retrieval using Color and Shape", *Pattern Recognition*, 29(8), pp. 1233–1244, 1996.

[17] S. Sural, "Histogram Generation from the HSV Color Space using Saturation Projection", in Multimedia Systems and Content-based Image Retrieval, S.Deb eds, Idea Group Publishing, Hershey, PA, USA, 2003.

[18] W. Niblack, R. Barber,W. Equitz, M. Flickner, E. Glasman, D. Pektovic, P. Yanker, C. Faloutsos, and G. Taubin,, "The QBIC Project: Querying Images by Content using Color Texture and Shape", *Proc. SPIE Int. Soc. Opt. Eng., in Storage and Retrieval for Image and Video Databases*, vol. 1908, pp. 173-187, 1993.

[19] Michael G. Michie," Use of the Bray-Curtis similarity measure in cluster analysis of foraminiferal data**",** *Journal of the International Association for Mathematical Geology*, Volume 14, Issue 6, pp 661-667,1982.

[20] Ahlqvist, O., "Using Semantic Similarity Metrics to Uncover Category and Land Cover Change", *1st International Conference on GeoSpatial Semantics*, Mexico City, Mexico, Springer, 107-119, 2005.