

Heuristic to Build RCC8 for Event Locations

Majed Ayyad

University of Trento.

Via Sommarive 14 - 38123 Povo

Trento, Italy

Email: ayyad@disi.unitn.it

Abstract—Events that are detected and reported by humans to actionable knowledge bases in multi-tier responding agencies have significant amount of spatial information. Humans have intuitive ability to triage repeated or duplicated events based on their spatio-temporal information. However, this cognitive process is not modeled easily and human ability is limited in situations where large number of events are reported simultaneously. The likelihood of two events to be the same is higher if they occur on the same place and time. In this work, we focus only on calculating location equivalence of events. For this purpose we use RCC8 theory to represent spatial relations between regional locations. The algorithm designed approximates the arbitrary shape of regions into circles and build region connection relations based on the size of the circle. The end result is a region of circular tiles with explicit RCC8 relations that could be used to reason on the relation between the locations of events. Additionally, we outline some experiments to evaluate the precision and recall of the results based on the used corpus. These results indicate that although the task is challenging, automated methods are capable of building spatial regional relations between events.

I. INTRODUCTION

DESPITE advances in technology and ubiquitous computing, a large volume of events are still detected and reported by human beings. Statistics about emergency rooms or command and control rooms still report about receiving hundreds of calls per day from the public. In such situations, operators log each call and forward it to a dispatcher or commander. Typically, dispatchers and operators sit next to each other, so they can rapidly share information among one another in case the system is jammed with calls and there are emergencies to reply to. The efficiency of this system is measured by the number of abandoned calls and the response time of the first responder (the dispatched resource to the event site). To improve both elements, commanders could benefit from having a system that can work on the

triage of events logged by operators and improve the decision making process before taking any action. In such situations, there is a need to triage repeated or duplicated events in real-time. We call this as the event matching problem which is the process of finding similarity between two events and is computed as 3-tuples $\langle e_1, e_2, R \rangle$, where R specifies a similarity relation. Possible forms for R are: equivalence ($=$), sub-event (\subset), and mismatch (\perp).

Davidson[2] and Quine [3] argue that two events are identical if they occupy the same portion of space and time. Therefore, for two events to match or (when $x=y$? ; if x and y are events) is to find the necessary and sufficient conditions for identical events. The matching criteria depends on a set of elements which we can summarize as: time, location, physical objects, cause and effect, existential conditions and properties. Therefore to match two events we need to calculate their time equivalence, location equivalence, causal equivalence and properties equivalence using other arguments such as participants and objects.

Comparing the location of two events is not always a straight forward, especially when events are reported using natural language. Different qualitative spatial relations are used to express the location of an event with other spatial entities. For the orientation aspect, events are described using qualitative terms such as “north of”, “in front of”, “behind”, etc. Many approaches and calculi have been used to express the orientation of one object on reference to another. Most approaches use points as the basic spatial entities and use different versions of JEPD orientation relations. Distance qualitative relations are also used when describing the location of events. For the distance aspect, terms such as “near”, “far”, “close to” are commonly used. As mentioned by [7] combining the orientation and distance aspects is called positional information

In this work, we use the Region Connection Calculus (RCC8) theory to partially solve the event matching problem. A location is defined as an inherently grounded spatial entity, a location includes geospatial entities such as countries, mountains, cities, rivers, etc. It also includes classificatory and ontological spatial terms, such as edge, corner, intersection[4]. The location element covers both locations and places (where a place is considered a

functional category), and is assumed to be associated with a region whenever appropriate[10]

The main objective of this paper is to demonstrate how connectedness relations between geographical spaces could be calculated automatically. The following five topological relations between locations are built: (1) Equal (2) Externally Connected (3) Disconnected (4) Tangential Proper Part, and (5) Non-Tangential Proper Part. In this work, and by using a dataset of a country we build RCC8 relations between cities, towns, villages, suburbs and points of interests.

For this purpose, we use an approximation technique to represent a region as circular shape. Furthermore, we represent a country map from circular tiles. The radius of the circle is calculated based on the type of the region being a city or a hamlet as an example. Other parameters are also considered if available such as the area and population of a region. We show that our heuristic algorithm to build RCC8 relations between country regions and places is likely to achieve acceptable results

This paper is organized as follows. Section 2 includes background information on RCC8. Section 3 describes the estimation problems and their formulation. Experimental results are presented in Section 4 and the paper is concluded in Section 5.

II. THE REGION CONNECTION CALCULUS

A. RCC8 relations

There are different aspects of space related to describing the event location on reference to another object. The location of an event could be expressed using a combination of orientation relations, distance relations and topological relations. While orientation and distance relations are important, in this paper we focus only on topological relations. Topology in mathematics concerned with the most basic properties of space, such as connectedness, continuity and boundary, while in qualitative spatial reasoning, the focus is on mereotopology [5].

In the Region Connection Calculus, regions are the basic spatial entities and relationships between spatial regions are defined in terms of the binary relation $C(x, y)$, meaning spatial entity x connects with spatial entity y , which is true if

and only if the closure of region x is connected to the closure of region y , i.e. if their closures share a common point[7]. Using the relation C , many versions of RCC could be found for instance RCC1, RCC2, RCC3, RCC5, RCC8, RCC15, and RCC23. The most common used and researched version is RCC8, which defines the following eight Jointly Exhaustive and Pairwise Disjoint (JEPD) relations: disconnected (DC), externally connected (EC), partially overlaps (PO), equal (EQ), tangential proper part (TPP), nontangential proper part (NTPP), tangential proper part inverse (TPPi) and nontangential proper part inverse (NTPPi) [8]. The intended meaning of these relations is illustrated in table (1).

B. Reasoning using RCC Relations

Since events are spatio-temporal entities, it is natural to use spatio-temporal reasoning to reason about the location of events. Studying how people report about the location of events, we notice that qualitative knowledge is used to express the event location as could be seen from the following example:

Event 1 : 8 Palestinians are arrested across the West Bank

Event 2: Thursday eight Palestinians arrested from Jerusalem, Jenin and Hebron, according to local and security sources.

In these two events, the event location is expressed using different qualitative representations which are used with different levels of granularity and expressiveness. When performing reasoning about the location of the two events, we may need to know if West Bank contains Jerusalem, Jenin and Hebron. Other aspects of event locations are usually described qualitatively, such as distance, orientation and topology.

Furthermore, There are many places that share the same or similar names ("AL-Tireh" :a neighborhood in Ramallah city;"AL-Tireh" : a Village in Ramallah region and "AL-Tireh": a village north of Jenin city) . Also some places have multiple names(e.g. AL-Manarah square is also called Lions square). Some places are called after the most famous point of interest found near that place.

With RCC we can reason if two events have the same location by using the connection relations as explained in the

TABLE I.
DEFINING RCC8 RELATIONS[10][13]

Name	Symbol	Relation	Meaning	Definition
Equals	EQ	EQ(x,y)	X is identical with y	$X = Y$
Disconnected	DC	DC(x,y)	X is disconnected from	$X \cap Y = \emptyset$
Externally Connected	EC	EC(x,y)	X is externally connected to y	$i(X) \cap i(Y) = \emptyset, X \cap Y \neq \emptyset$
Partially Overlap	PO	PO(x,y)	X partially overlaps y	$i(X) \cap i(Y) \neq \emptyset, X \not\subseteq Y, Y \not\subseteq X$
Tangential Proper Part	TPP	TPP(x,y)	X is tangential proper part of y	$X \subset Y, X \not\subseteq i(Y)$
Non-Tangential Proper Part	NTPP	NTPP(x,y)	X is non-tangential proper part of y	$X \subset i(Y)$

following rules :

Disconnected: Since one event cannot take place into two separate locations , and we have two events with disconnected locations, we can deduce that these are two different events .

$$\frac{(e_1 \text{ in } x) \wedge (e_2 \text{ in } y) \wedge (DC(x,y))}{e_1 \not\cong e_2}$$

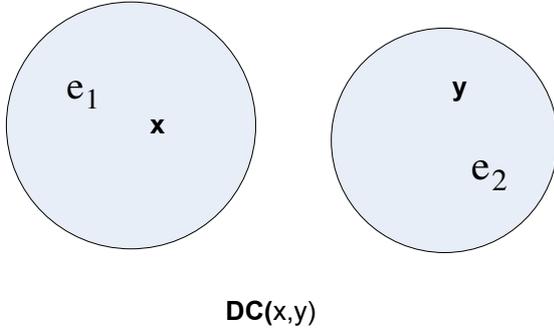


Fig. 1 Disconnected regions

Equal: If the two regions are equal , then at least one condition is met in the matching criteria, therefore it is possible that these two events are matched.

$$\frac{(e_1 \text{ in } x) \wedge (e_2 \text{ in } y) \wedge (EQ(x,y))}{e_1 \cong e_2}$$

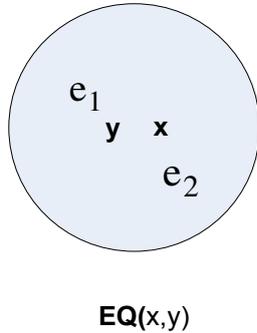


Fig. 2 Equal regions

As in the following two events, if we know that Radio street and Al-Ersal street are the same street from our knowledge base then this condition is met.

Event 3 : On 11 May 13, 11:14 hrs, reportedly, a car accident was reported in Radio street

Event 4 : On 11 May 13, 11:18 hrs, a car accident was reported in Al-Ersal street

Externally Connected: with externally connected regions, there is a possibility that the two events are taking place at the border of these two regions, therefore it possible that these two events have equal location and therefore a possible match.

$$\frac{(e_1 \text{ in } x) \wedge (e_2 \text{ in } y) \wedge (EC(x,y))}{e_1 \cong e_2}$$

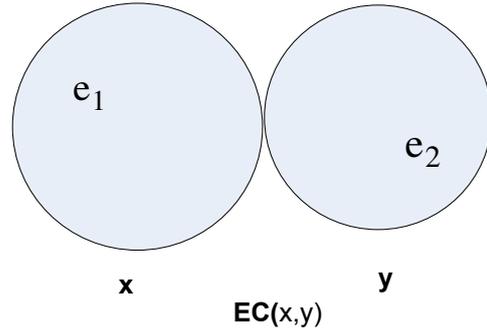


Fig. 3 Externally Connected regions

Event 1 : On 31 Mar 13, 0930 hrs, approximately 40 people demonstrated at DCO Beit-EL, NE Ramallah. It ended peacefully at 1440 hrs.

Event 2 : On 31 Mar 13, between 0945-1200 hrs, families protested near City Inn Hotel, NE Ramallah against prisoners conditions.

Non tangential proper part: The semantic of the non tangential proper part is that region R1 is totally inside region R2 and that they are not equal and do not share any border.

$$\frac{(e_1 \text{ in } x) \wedge (e_2 \text{ in } y) \wedge (NTTP(x,y))}{e_1 \cong e_2}$$

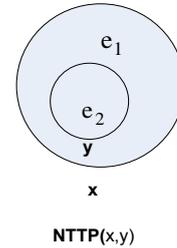


Fig. 4 Equal regions

Event 1 : A house in fire , in Jaffa street , second floor , near store AL-Manara close to AL-Families park

Event 2 : A smoke is seen ,near supermarket AL-Manara in Ain-Munjid area.

In these two events , Al-Families park is located in Ain-Munjid area.

Tangential proper part: in TTP relations, there might be more than two regions involved in the event. If x,y, and z are regions then y and x might be connected through a TPP , also y and z might be connected through a TPP.

$$\frac{(e_1 \text{ in } x) \wedge (e_2 \text{ in } y) \wedge (TTP(x,y))}{e_1 \cong e_2}$$

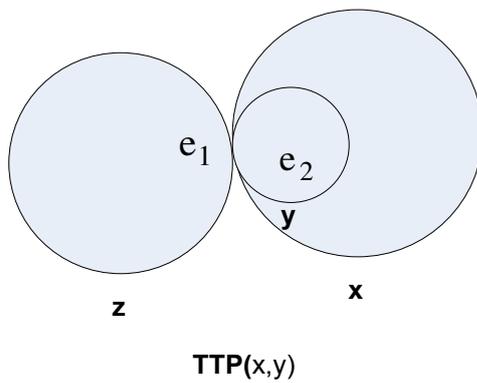


Fig. 5 Different TPP relations

Event 1 : A teen is injured in clashes near Jerusalem

Event 2 : A 17-year-old student was injured Thursday morning during clashes in the town of Abu Dis.

A main advantage for using RCC to reason about the location of events is that as examined by [9] , RCC is structurally similar to the way people reason about space and is a model of people's conceptual knowledge of spatial relationships).

III. CALCULATING RCC RELATIONS

A. Region Ontology

In this paper all the examples are taken for events located in populated places. A populated place is an area of land inhabited by people. Therefore cities, villages , hamlets , towns, townships ,etc. are type of populated places. By definition, what mainly characterize an entity from another is its area. It is common to find the following definitions: a village is small human settlement, or a city is a large settlement and a hamlet is just a few dwellings[12]. Location and regions are more important for our work, however places are sometimes used to describe a region by its functional place like "city center". A city center is a circle on a map to indicate the center of the city and it is only perceived by the human mind.

We have noticed that the three themes of geography (location, place and region) are used to describe where an event occurred or is happening. An observer uses relative location to describe the event when the observer is not familiar with the area. Also absolute locations are used when the observer knows the address of the event. Functional locations such as 'city center' or formal name such as 'name of the city', or vernacular region such as 'at the south area of the city' are all used to describe an event location.

To model our regions, we use a region ontology where the country regions are classified into populated places and administratively declared places as shown in Figure

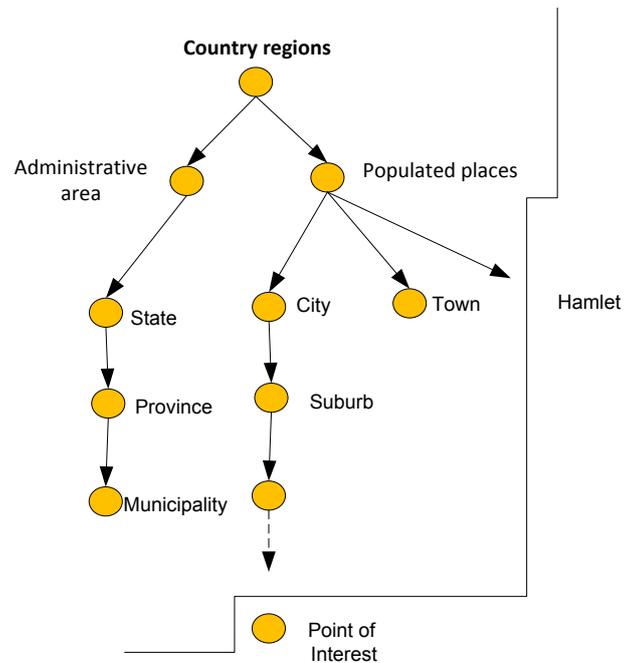


Fig. 6 Classification of a Country Region

Populated places are classified into extended entities such as city and non-extended entities such as a point-of-interest. All populated places are disjoint classes and are continuous and have no holes. Suburbs and neighborhoods are part of a larger entity and is represented in one of the following forms :

NTTP: a suburb(S) has an NTTP relation with a town (T) if a suburb lies in a town and shares no border with it. The relation is denoted by S NTTP T

TPP: a suburb(S) has a TPP relation with a town (T) if a suburb lies in a town and shares borders with it. The relation is denoted by S TPP T

EC and DC , this relation holds between suburb of a larger entity such as a city or town.

B. The Algorithm

The proposed methodology for calculating RCC between geographical regions is to approximate the exact region tiles by circular tiles as shown in Table (2). In the case of a country regions, the frame of reference is the partition of the country into cells which share boundaries but do not overlap. RCC relationships could then be calculated by using the longitude and latitude of the region as the center of the cell and then calculating the distance between cells.

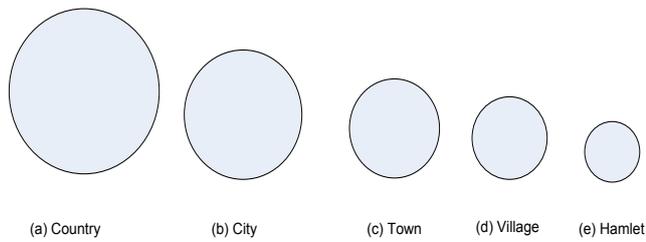


Fig. 7 Region Classification based on approximate area size

The difference between each type is identified by a set of features specially the size of the region. By comparing the distance between center of the cells and the reference distance, we can calculate the following relations:

Disconnected (DC): if two cells, R1 and R2, share no border then the relation between them is denoted by R1 DC R2. This is calculated using the following formula

$$DISTANCE(R1, R2) > (2 * \alpha + c); \alpha \text{ denotes a constant that represents the maximum radius of a town and } c \text{ denotes an error margin constant}$$

Externally connected (EC): if two regions, R1 and R2, share borders then the relation between them is denoted by R1 EC R2.

$$DISTANCE(R1, R2) < (2 * \alpha + c)$$

Equals (EQ): the relation between each town, or any other location type, and itself is denoted by R1 EQ R2.

$$DISTANCE(R1, R2) < c$$

Both DC and EC relations are bidirectional. The algorithm is basically divided into three main parts : (1) calculates relations between town or cities (2) calculates relations

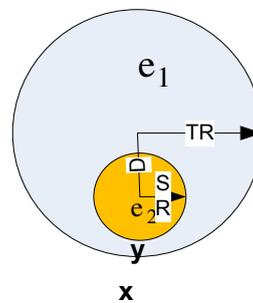
point of interests and suburbs or town with no suburbs. Following pseudo code illustrates how to calculate relations among towns in a country.

```

Pseudocode for RCC8 Relations among towns/cities
Declare region Radius  $\alpha$  // represents the maximum radius in meters
Declare  $c$  // denotes an error margin constant defined in meters

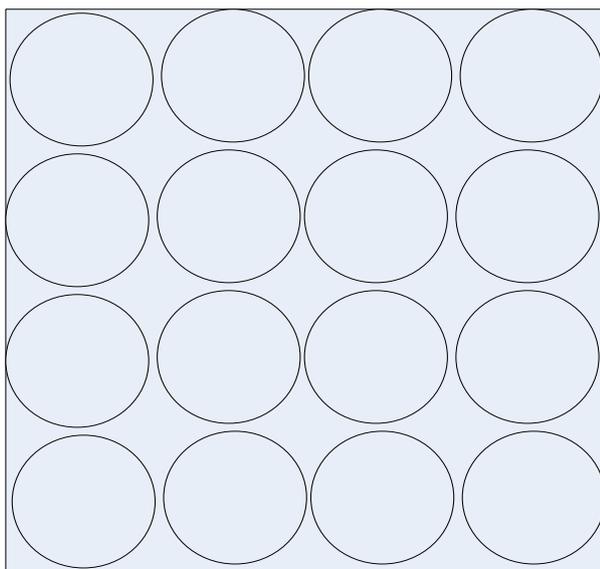
Input region dataset containing longitude, latitude, place name
TOWNS_SET = FIND_ALL_LOCATIONS_BY_TYPE ("TOWN")
POIS_SET = FIND_ALL_LOCATIONS_BY_TYPE ("POIS")
SUBERBS_SET = FIND_ALL_LOCATIONS_BY_TYPE ("SUBERB")
BEGIN
    Build RCC8 Relations among towns
    Build RCC8 Relations among Suburbs and Towns
    Build RCC8 Relations among places of interests, towns and suburbs and towns
END
Output set of Relations between all regions {EQ,DC ,EC} // same type.
    
```

C. Building Town Suburb Relations

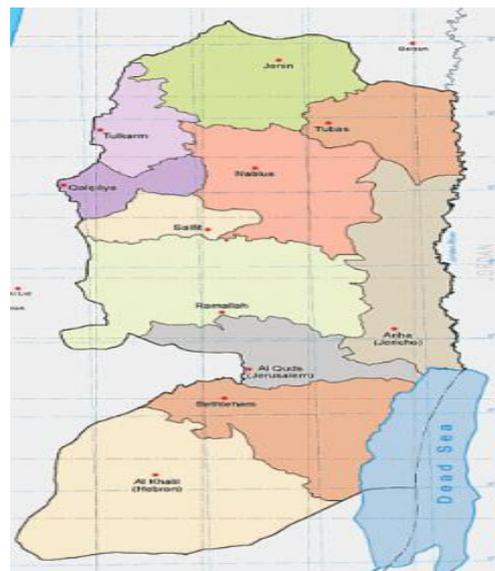


NTTP(Town,Suburb)

TABLE II.
(A) APPROXIMATION USING CIRCULAR TILES (B) EXACT REGION TILES



(a) Approximation using circular tiles



(b) Exact region tiles

between suburbs and towns (3) calculates relations between

Fig. 8 NTTP(Town, Suburb)

The second part of the algorithm is concerned with building the relations between towns and suburbs.

$TownRaduis > Distance + SuburbRaduis + Constant$

This part of the algorithm try to build the town or city suburbs only based on the input data which are are the lat,lon and suburb name.

D. Building Suburb- Suburb Relations

Building the suburb-suburb relations , follows the same approach for towns except we limit the comparison among a city or town suburbs.

Externally connected (EC): if two regions, S1 and S2, share borders then the relation between them is denoted by S1 EC S2.

Equals (EQ): the relation between each suburb, or any other location type, and itself is denoted by S1 EQ S2.

$$DISTANCE(S2, S2) < c$$

E. Building POI Relations

Specifying the relations among points of interests, villages, suburbs and towns: a point of interest can be located either in a village, suburb (town). The set of relations are all the relations listed above considering the semantics and the context of point of interests. At this stage we are mainly considering the NTPP relation between a point and a region (suburb,village and town).

NTPP: a POI(S) has an NTPP relation with a town (T) if a POI lies in a town and shares no border with it. The relation is denoted by S NTPP T

IV. EXPERIMENTAL RESULTS AND VALIDATION

country region. All experiments were conducted on an Intel(R) Core(TM) i7 2.00 GHz running 32-bit Windows 7 Operating System with 4 GB of RAM.

A. An illustrative example

To build the data set for this experiment, we used Palestinian regions . We collected the shape



Figure 6 –

Fig. 9 Map of a region from a shape file

files from different municipalities like the one in Figure (6) and loaded the shape files into **PostGIS/PostgreSQL** database using the right coordination system for the selected region. The total spatial entities for this experiment is 5957 entity classified as shown in table ().

TABLE II.

SPATIAL ENTITIES PER TYPE

Type	Count
locality	144
hamlet	23
village	323
pois	5337
suburb	39
region	7
town	81
Border Crossing	1
city	10

The challenging question at this point is how to select the best radius for each region type. Obviously the algorithm will produce wrong results if the radius is chosen too small or too large. In order to select the best radius, we created a visual map that can help the user to select the best radius. As shown in figure (7) , choosing a radius of 800 meter will create more relations than 400 meters. Also we enhanced the algorithm by considering the area of the region. If the area of the region is found, then we can calculate the radius using the formula $Area = \sqrt{(Area)/3.14}$ and thus we can get more reliable address

B. Validation of results

To develop our ground truth database for region relations, we had to build up the relations manually from existing maps. The ground truth data might include attribute data about the area size or population size of the region. However, not all towns or cities have these attributes filled. At this moment, we manually built the EC relationship between all towns, cities and villages. Also suburbs relations were built for two cities. Point of interests relations with their suburbs are built for nine suburbs.

The results are validated by computing precision , recall as shown in table

TABLE III.
EC PRECISION AND RECALL PER REGION

Reg ion Id	Total EC	EC relations	EC relations	EC relations	EC relations
	Relations (GT-Expert)	Built By System	Built (True)	missed	Built (false)
			<true positive>		<false positive>
94	8	6	5	3	1

91	6	2	2	4	0
236	8	9	7	1	1
196	9	8	8	1	0
177	6	8	6	1	2
.....					

For each region , we calculate the EC relations manually as shown in column (2). Column-3 represents the total relations built by the algorithm for each region ; column-4 shows how much of the calculated relations are true; column-5 shows how much relations are missed and column-6 represents how much relations are false.

- A – Number of relevant relations not retrieved
- B – Number of relevant relations retrieved
- C – Irrelevant relations retrieved

$$\text{Precision} = \frac{|B|}{|B| + |C|};$$

$$\text{Recall} = \frac{|B|}{|A| + |B|};$$

Precision = 0.82926829

Recall = 0.90265487

C. Discussion of results

Since the approach relies on approximating the area using a circle region. Selecting the radius (R) might produce wrong results as shown in the following cases. When the radius R is much smaller than region radius (RR) (R << RR), the algorithm creates no relations between the two regions. This is equivalent to region A is disconnected from region B

. This could be improved by using the area of the region to calculate the radius and overriding the estimated one.

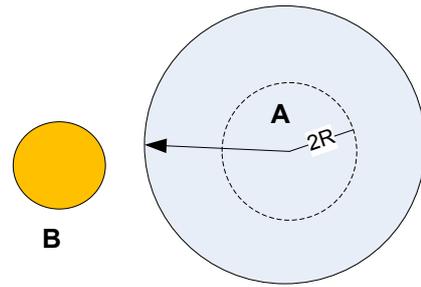


Fig. 10 City radius larger than double of selected radius

A second case occurs when the selected radius R is much larger than region radius (R >> RR)

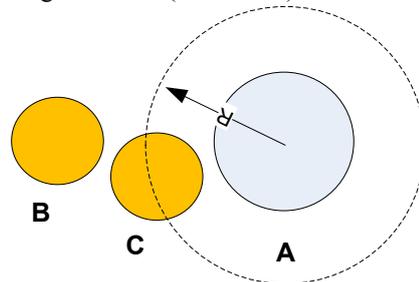
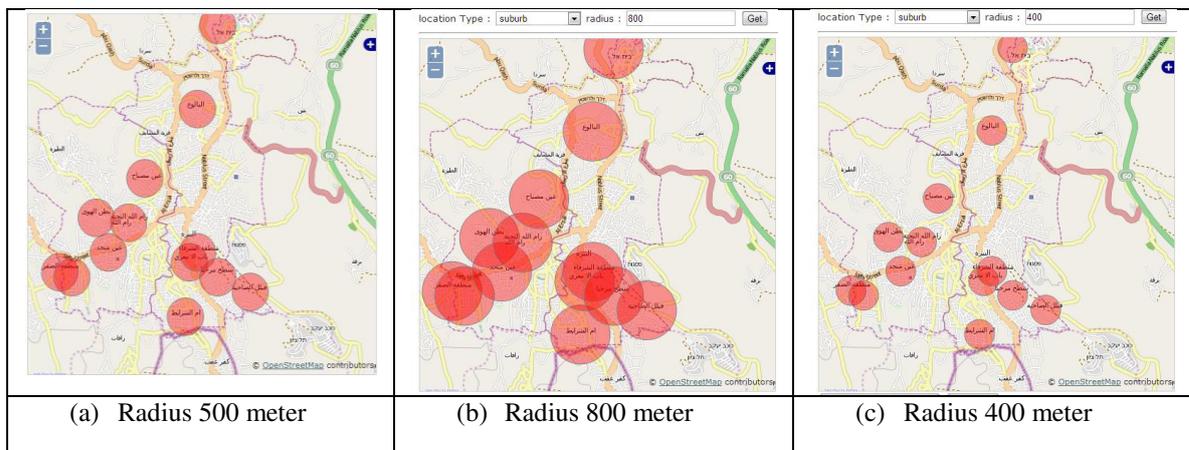


Fig. 11 City radius less than double of selected radius

When region radius is much lesser than selected radius , it is possible to make an EC relation with a region although there is another region in between. This is equivalent to having the following relations : a) A Externally connect to B b) A Externally connect to C c) C Externally connect to B

TABLE IIV.
DIFFERENT REGION RADIUS BASED ON VISUAL MAP



II. CONCLUSION

Using an automated method to build RCC relations between geographic regions is challenging especially if data has only attributes related to longitude and latitude. Although locating events is best done by its address, which is the more accurate among other methods like post address or boundary, the boundary approach in many rural areas is the only option available. However, from our experiments we found encouraging results. With such results it is now possible to use the new data set to find automatically the matching relationships between a pair of events such as in the two events presented earlier:

Event 1 : “8 Palestinians are arrested across the West Bank “

Event 2 : “Thursday eight Palestinians arrested from Jerusalem, Jenin and Hebron, according to local and security sources”. Since Jerusalem, Jenin and Hebron has NTPP relationships with West Bank, we can infer that the location of these two events is the same.

However, we did not consider the issue of integrating topological relation with other binary relations such as distance constraints and directional constraints. In future research, the similarity measure between location A and location B is computed using the three types of constraints.

REFERENCES

- [1] J. M. Zacks, B. Tversky, "Event structure in perception and conception". *Psychological Bulletin*, 2001, 3-21
- [2] D. Davidson, "The individuation of events", in Davidson, 1985, p 179 doi: <http://dx.doi.org/10.1037//0033-2909.127.1.3>
- [3] W. V. Quine, "Events and reification", 1985
- [4] J. Pustejovsky, "ISO-Space: The Annotation of Spatial Information in Language", Proceedings of the Sixth Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation isa-6, 2011
- [5] A. G. Cohn and J. Renz, "Qualitative Spatial Representation and Reasoning. Handbook of Knowledge Representation", pages 551-596, 2008.
- [6] B. L. Clarke. "A calculus of individuals based on 'connection". *Notre Dame J. Formal Logic*, (22), 1981. doi:10.1305/ndjfl/1093883455
- [7] J. Renz and B. Nebel. "Qualitative spatial reasoning using constraint calculi". In *Handbook of Spatial Logics*, pages 161-215. Springer, 2007. ISBN 978-1-4020-5586-7. doi:10.1007/978-1-4020-5587-4_4
- [8] D. A. Randell, Z. Cui, A. G. Cohn, "A Spatial Logic based on Regions and Connections". In B. Nebel, C. Rich, and W. Swartout (Eds.), *Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann, San Mateo, CA, 1992, 165-176 DOI: 10.1016/j.artint.2008.10.009
- [9] M. Knau, R. Rauh, J. Renz, "A cognitive assessment of topological spatial relations: Results from an empirical investigation". In *Proceedings of the 3rd International Conference on Spatial Information Theory (COSIT'97)*, volume 1329 of *Lecture Notes in Computer Science*, pages 193-206, 1997 doi: 10.1007/3-540-63623-4_51
- [10] Z. C. David, A. Randell, A. G. Cohn, "A spatial logic based on regions and connection" in *3rd International Conference on knowledge representation and reasoning*, vol. 1, 1992, pp. 165-176.
- [11] T. Bittner, J.G. Stell, "Approximate qualitative spatial reasoning" Department of Computer Science, Northwestern University doi: 10.1023/A:1015598320584
- [12] <http://vocab.org/places/schema.html>
- [13] J. Renz, "A canonical model of the Region Connection Calculus", in: *Proc. 6th International Conference on Principles of Knowledge Representation and Reasoning (KR-98)*, Trento, Italy, 1998. DOI: 10.3166/jancl.12.469-494