Spatial similarity and GIS: the grouping of spatial kinds

Alec Holt

Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
Phone: +64 3 479-8301 Fax: +64 3 479-8311
Email: aholt@infoscience.otago.ac.nz

Presented at SIRC 99 – The 11th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
December 13-15th 1999

ABSTRACT

This paper outlines the general concept of similarity. The background on the research undertaken on similarity is detailed and its application in GIS is discussed. This is an introductory paper to spatial similarity and more detailed concepts such as context and similarity measures can be found in other articles by the author. The example given in this paper is called Huntervale and is a property application. The example aims to highlight the influence of spatial attributes on similarity.

Keywords and phrases**:** similarity, information retrieval, case-based reasoning, spatial similarity.

1.0 INTRODUCTION

Similarity is fundamental for learning, knowledge and thought, for only our sense of similarity allows us to order things into kinds so that these can function as stimulus meanings…………………reasonable expectation depends on the similarity of circumstances and on our tendency to expect that similar causes will have similar effects (Quine 1969:114)

Similarity is important for people to make sense of the objects, structures and actions existent in reality. The process of classifying objects is a fundamental feature of most human pursuits, and the idea that people classify together those things that people find similar is both intuitive and popular across a wide range of disciplines. Quine (1969) does not say that the similarity relation is fundamental to science (which he regards as the most advanced and successful species of the search for knowledge). Scientific theories can not be reduced to logical constructions of simple statements about the similarity between sense data (as many logical positivists such as Carnap (1967) thought), not even similarity between physical objects outside the observer. Quine would rather say that our sense of similarity is basic insofar as it is the starting point for the individual's development of language skills and for the development of theories in a virgin field of study. Our naïve perceptions of similarity are likely to be refined and in some cases contradicted by the scientific theories that eventually evolve. Maybe Quine's quote also stands true for a new language and knowledge storing ability made possible via contemporary information systems.

This paper aims to emphasise, whilst operating in an information science domain, that similarity can be utilised as the basis of a retrieval technique *to sort spatial things into groups*.

2.0 SIMILARITY

This section outlines studies in various disciplines on similarity assessment including philosophy, psychology and geographic information science. There has been a large variety of research on the subject of similarity. Quine (1969) states that some of the research pre-dates the Socrates era. Section 2.1 outlines the research by philosophers and psychologists, who detail the cognitive aspects of similarity. Definitions and understandings of similarity vary from researcher to researcher. Some researchers suggest similarity is an innate quality of humans to categorise data in order to make decisions. Other researchers suggest similarity is more of a leaning experience, a maturation process, which through methods like repetition people can learn to categorise things. Some researchers suggest that similarity is a naïve technique for organising data and suggest that similarity is not used in mature science. This paper is at odds with that statement and takes the view that similarity is useful in naïve and mature fields of science. It is useful in that similarity is the basis of a retrieval technique that allows data to be retrieved and compared.

2.1 Philosophical and psychological background

This section outlines our sense of similarity as a property that is both innate and accumulative. This property is crucial to a person's ability to form expectations and make predictions. The ability to make similarity judgements is considered to be a valuable tool in the study of human perception and cognition and play a central role in theories of human knowledge representation, behaviour and problem solving. Tversky (1977:327) describes the similarity concept as *an organizing principle by which individuals classify objects, form concepts, and make generalization.*

Quine (1969) defines a primitive form of similarity as *our sense of comparative similarity fits in with regularities of nature, so as to afford us reasonable success in our primitive induction's and expectation, it is presumably an evolutionary product of natural selection.* Quine acknowledges, however, that through development people form a more objective sense of a similarity away from the immediate, subjective and animal sense of similarity. Quine suggests that our sense of similarity, our grouping of kinds, is both innate and accumulative. Innate in that our sense of similarity is our foundation blokes of reasoning and induction, our internal check system. The concept of similarity according to Quine is embedded in our innate senses. It is accumulative in that it develops and changes and even turns multiple as one matures, making perhaps for increasingly dependable prediction. Quine presumes that our sense of similarity has evolved through natural selection and it seems to be vital for human success and survival. Our sense of similarity is accumulative in that people learn through maturation and development. Interestingly, our senses compliment new and novel grouping of kinds, they are not superseded. Quine (1969:129) suggests that, o*ur experiences from earliest infancy are bound to have overlaid our innate spacing of qualities by modifying and supplementing our grouping habits little by little, inclining us more and more to an appreciation of theoretical kinds and similarities, long before we reach the point of studying science systematically as such. However, we retain different similarity standards, different systems of kinds, for use in different contexts. We all still say that a marsupial mouse is more like an ordinary mouse than a kangaroo, except when we are concerned with genetic matters. Something like our innate quality space continues to function alongside the more sophisticated regroupings that have been found by scientific experience to facilitate induction.*

Quine (1969:136) encapsulates our historical drive to understand the concept of similarity when he states that, *philosophical or broadly scientific motives can impel us to seek still a basic and absolute concept of similarity, along with such fragmentary similarity concepts as suit special branches of science. This drive for a cosmic similarity concept is perhaps identifiable with the age-old drive to reduce things to their elements. It epitomizes the scientific spirit, though dating back to the pre-Socrates: to Empedocles with his theory of four elements, and above all to Democritus with his atoms.*

Throughout history people have discerned different natural kinds, or species, among animals and plants. This was simply based on how similar or dissimilar different living creatures seem to a person who has some level of personal experience of animals or plants. Gradually these natural kinds have evolved into today's much more objective and theoretically well-founded "taxonomy of life" in biology.

Some milestones in this development of the "taxonomy of life" in biology have been;

- the binomial nomenclature of Linnæus,
- the systematic use of the wider groupings of genera, families, orders and classes,
- the coarse rule that those individuals form a species which in principle could produce fertile offspring,
- the insights into how a species may split into two different species from evolutionary biology,
- the explanation from molecular genetics of why individuals of the same species are very similar to each other but also show small variations,
- the quantitative measurement of the degree of relatedness between individuals, living or fossilized, by means of a chemical analysis of DNA samples.

Hume (Mossner, 1969) was a pioneer of the philosophical study of the concept of similarity. His work *A Treatise of Human Nature* written in 1740 foresaw that if objects are similar in appearance then they will be attended with similar effects. Thus, from causes which appear similar, people expect similar effects. Hume considered surface similarity to be the only form of existing similarity. He believed that when assessing similarity, it is sufficient to consider only simple sensory attributes of objects and he did not consider different perceptions of these attributes by different subjects or by the same subjects but in different contexts. This approach is called *common attribute view of similarity*. Hume's views are limited because his approach equates surface similarity with psychological similarity and thus neglects perceptual capacities of the organisms and assumes common environmental properties. He stated that the degree of similarity of two composite ideas depends on the number of simple ideas they have in common. However, Hume assumed that the similarity between simple ideas are the immediate commonalties. He also observed that arguments from experience are founded on the similarity which humans discover among natural objects (Mossner, 1969).

Wittgenstein (1958:31-32) used commonalties to indicate similarity, in saying that, *something runs through the whole thread - namely the continuous overlapping of those fibres.* He argued that the attributes that situations and objects have in common should be called family resemblances. Family resemblances are *a complicated network of overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail. (Wittgenstein 1958:355).* According to Wittgenstein, the knowledge required to possess a concept or use a linguistic item is an implicit knowledge of the 'family resemblances' between situations and objects.

Popper (1972:422) identifies the significance of point of view to similarity *if similarity and repetition presuppose the adoption of a point of view, or an interest, or an expectation, it is logically necessary that points of view, or interests, or expectations, are logically prior, as well as temporally (or causally or psychologically) prior, to repetition.* What Popper stresses is that similarity between two things is always relative to a certain respect in which they are compared, a certain perspective or interest. They may be similar in one respect but dissimilar in another. For Popper the repetition of similar events is not the basis for empirical theories, not even in the weak psychological sense that expectations fulfilled induce a belief in a general theory. The repeated observation of white swans, for example, is what makes people believe in the general theory that all swans are white. According to Popper (illustrated in figure 1), *two things which are similar are always similar in certain respect and generally, similarity, and with it repetition, always presupposes the adoption of a point of view (Popper, 1972:420-421).*

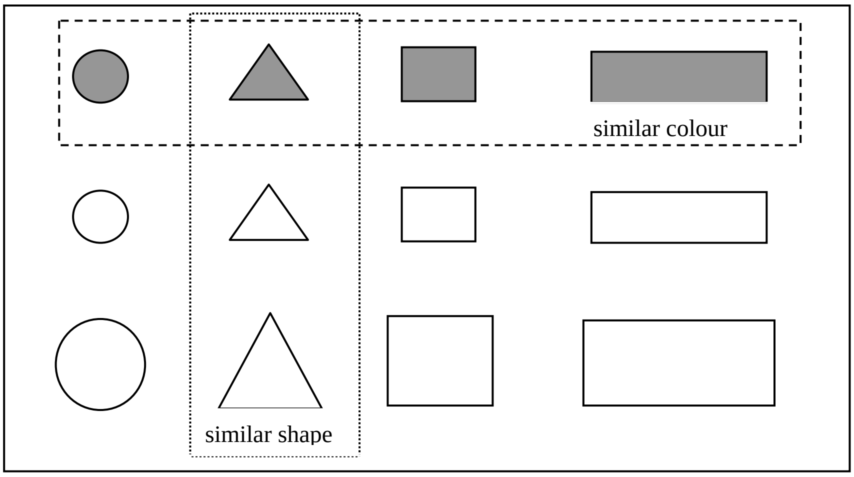


Figure 1 An illustration of Poppers different respects (Adapted from Popper, 1972:421)

Figure 1 illustrates how certain respects can effect similarity. Some of the objects are similar with respect to colour, while others are similar with respect to either shape or area. Therefore, suppose all the objects in figure 1 represent the entire database. Then according to the user point of view each query will give different answers (colour, shape and area). Each answer is correct for the context of the query. Context has a major influence on the information retrieved based on the similarity concept. With context, scale, techniques for retrieval and similarity measures have important roles in spatial similarity. This paper will not detail those roles, instead this paper is an introduction to spatial similarity in GI systems.

3.0 SIMILARITY IN GI SYSTEMS

Holt and Benwell (1997:279) define spatial similarity as *those regions which at a particular granularity (scale) and context (thematic properties) are considered similar*. Spatial similarity (see figure 2) is broadly defined as spatial matching and ranking according to a specific context and scale. More specifically, similarity is governed by context (function, use, reason, goal, users frame-of mind), scale (coarse or fine level), repository (the application, local domain, site and data specifics), techniques (the available technology for searching, retrieving and recognising data) and measure and ranking systems. Similarity is also determined by temporal aspects, that is, the instance the phenomenon occurred could affect the similarity of other phenomena in relation to it. It is important to realise that context in this definition is determined by the user, not automatically by the system.

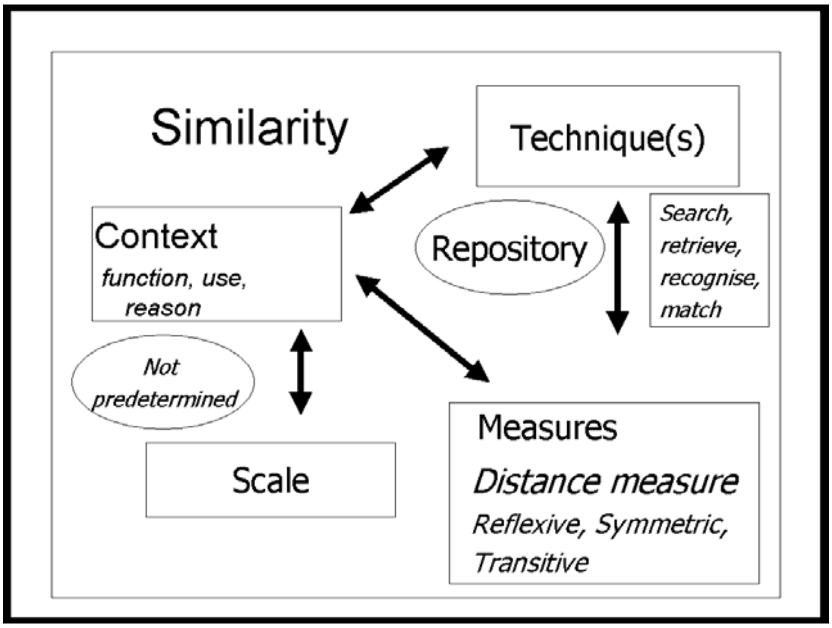


Figure 2. The variables of spatial similarity

There are some distinctive groups currently researching similarity in the milieu of GI systems. These groups use a variety of techniques ranging from *deviation from equivalence*, *feature matching*, *fuzzy membership functions* and *case-based reasoning*. Uses of similarity range from inter-operability (Goodchild *et al.* 1998), conflation (Cobb *et al.* 1998), data retrieval (Holt & Benwell 1999; Flewelling 1997; Bruns & Egenhofer 1996; Papadias and Egenhofer 1997), problem solving (Shi and Yeh, 1999; Holt 1996; Holt *et al.* 1997; Jones & Roydhouse 1994) and exploratory/interpretation (Holt & Benwell 1999).

Cobb *et al.* (1998) present a novel approach to combining maps and associated knowledge (conflation). To determine points which are identical between different maps they describe feature matching and de-confliction and favour the use of inexact reasoning concepts. They implement a system where each feature is considered as a set of attribute-value pairs. From this representation, a degree of matching similarity is determined. For numeric domains a membership matching function is used, while a similarity table is used for linguistic domains. By using a combination of the table and a fuzzy logic membership matching function a composite matching score is computed from the combination of an expert system weight and similarity table values.

Goodchild *et al.* (1998), suggested similarity is relevant to inter-operability. In that it allows a measure of the degree to which *two data sets, software systems, disciplines or agencies use the same vocabulary, follow the same conventions, and thus find it easy to interoperate.* Goodchild *et al.* (1998) suggest that it is only possible to inter-operate over a very narrow domain. Therefore, when considering similarity in the context of inter-operability Goodchild *et al.* (1998) say *the effort to achieve interoperability is thus an effort to extend domains, or to raise the threshold of similarity below which interoperability is possible.* The authors assume the above could also be true for intra-operability.

Configuration similarity developed more recently as a form of content-based retrieval. Bruns and Egenhofer (1996) and Papadias and Egenhofer (1997) initially described spatial structures and configurations (in spatial databases). Once they realised the spatial shape or structure, and given a new instance, they equated similarity by counting the number of transformations it took to morph from an unknown state to a known (structure or configuration). Bruns and Egenhofer (1996) defined similarity as *the assessment of deviation from equivalence*. This raises several questions, for example how is *assessment of deviation* represented and measured ? and how is *equivalence* defined ? Bruns and Egenhofer (1996) used similarity for data retrieval and feature matching.

Egenhofer directs two current research projects with a focus on *similarity.* These include;

1. *Similarity* assessments based on spatial relations and attributes, funded by the National Imagery and Mapping Agency and
2. Heterogeneous geographic databases: spatial *similarity* for the Advanced Research and Development Committee of the Community Management Staff.

These projects include research on numerous database issues including *spatial similarity* retrieval. Researchers include Egenhofer, Flewelling, Goyal, Paiva, Rodríguez and Beard (University of Maine), Bertolotto (Universita di Genova, Italy), Freitas (INPE, Brazil), Sharma (Oracle) and Ubeda (INSA de Lyon, France). In project number 1 spatial similarity measures are developed to overcome the shortcomings of traditional methods (precise spatial concepts, discrete data structures and boolean operators). Egenhofer's team proposed similarity measures based on spatial relations and attributes. Spatial relations are used to capture the distribution of spatial objects through a multi-scale model, allowing analysis of topological, directional and metrical relations. Attribute similarity is measured through a semantic network of feature classes. project number 2 investigates the changes detected whilst analysing multi-scale geographic databases among the different representations for the same geographic area, or different geographic locations. Spatial similarity can be derived using the concepts of the four-intersection and its component invariants. They intend to extend this model to account for qualitative metric properties of spatial relations, and will develop formal models for assessing spatial changes. Egenhofer's team also aim to test their concept for two-dimensional and three-dimensional models.

Papadias and Delis (1997) defined measures for modelling similarity of configurations. They suggested that configuration similarity has developed as a complementary form of content based retrieval. Papadias and Delis (1997) suggested that most approaches follow the same methodology, which includes:

- describing the set of spatial relations allowed in the expression of queries,
- defining measures of similarity between images based on the resemblance between spatial relations (and not on visual characteristics) and
- (in some cases) proposing algorithms for similarity retrieval.

Flewelling (1997) suggested that recent similarity queries have been researched in the object-based spatial (Flewelling 1997; Bruns & Egenhofer 1996) and image database community (Flickner *et al.* 1995; Gudivada 1995; Gudivada & Raghavan 1995). There has been little research on the properties that similarity operators must fulfil and on the differences between field and object models. Flewelling (1997) proposed a solution to the differences between field and object models. He suggested that in order to measure the similarity of one field to another, the similarity of the four field characteristics must be measured. He identified these four fields as theme, extent, time and value (samples) and said that these can be used to derive a four dimensional distance representing the similarity of the two fields. A set of these field similarities could be generated against a user defined scenario (query) or a known state. Flewelling (1997) suggested that this makes it possible to retrieve fields from a database that are highly similar, (but not equivalent to the users query) and to quantify that similarity.

Spatial relations may be classified into directional and topological relations. The frequently used directional relations are the strict directional relations: north, south, east, and west. Some researchers add the mixed directional relations: northeast, northwest, southeast, and southwest. Others use the positional directional relations: left, right, above, and below. Egenhofer and Franzasa [1991] and Egenhofer and Franzasa [1995] point out that there are eight fundamental topological relations that can hold between two planar regions. These relations are disjoint, contains, inside, meet, equal, covers, covered-by, and overlap (Figure 2). This model is called the four-intersection model. A refinement to this model was proposed in Egenhofer and Franzasa [1995] to distinguish between topologically distinct configurations whose empty/nonempty-intersection values are the same. However, directional relations are not sufficient for characterizing spatial similarity because they only consider the spatial orientation of an object while ignoring its spatial extent. In some cases, directional relations do not exist, while in other cases directional relationships may be identical in two images in spite of the fact that the images are not spatially identical. In addition, directional relations are not rotation invariant.

Rodriguez, M. A. aims to define a computational model for the semantic similarity assessment among spatial entity classes. Her main motivation is to enhance the information retrieval and integration mechanism of GI systems. The most recent work *Putting Similarity Assessments into Context: Matching Functions with the User's Intended Operations* is authored by Rodríguez and Egenhofer. The work is based on a new model for the assessment of semantic similarity among entity classes that satisfies cognitive properties of similarity and integrates contextual information. The semantic similarity model represents entity classes by their semantic relations (is-a and part-whole) and their distinguishing features (parts, functions, and attributes). Context describes the domain of an application that is determined by the user's intended operations. Contextual information is specified by a set of tuples over operations associated with their respective entity-class arguments. Based on the contextual information, a partial word-sense disambiguation can be achieved and the relevance of distinguishing features for the similarity assessment is calculated in terms of the features' contribution to the characterization of the application domain.

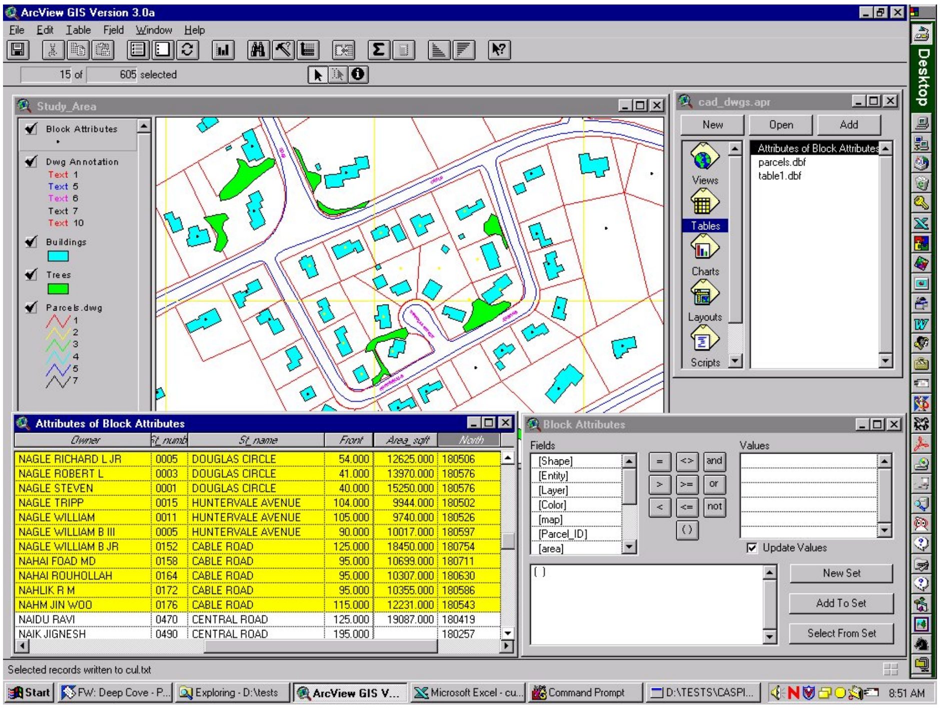
Recently Professor Anthony Gar-On Yeh, Xingwen Wang and Xun Shi from the GIS Research Centre and Centre of Urban Planning and Environmental Management, University of Hong Kong published an article on spatial similarity and GIS. Shi and Yeh (1999) use techniques and methods in-line with Holt and Benwell (1996). The techniques are GI systems combined with CBR find similarities in datasets. The nearest neighbour algorithm is used by the CBR system to find the similarity between a new problem and an old case. Shi and Yeh (1999) use GI systems and CBR to retrieve previous cases to interpret or solve a new problem that assists planners in the Planning Department of Hong Kong in handling planning applications in development control.

Holt & Benwell (1996) identified the usefulness of CBR and similarity in GI systems. Holt (1996) proposed a spatial similarity system (SSS) which allows GI systems to recognise, retrieve, re-use, revise and retain from the past for the present and future. This concept is useful for spatial problem solving, data retrieval, classification and exploratory/interpretation (Holt *et al.* 1997; Holt & Benwell 1999). Geocomputational techniques are increasingly necessary for certain applications for data analysis, data mining and exploratory analysis (Holt 1997; Openshaw & Abrahart 1996).

It is suggested that spatial similarity could be utilised both as a descriptive and exploratory concept in an attempt to satiate the geocomputational need. The SSS is a spatial-artificial intelligence-hybrid and is under continuous research and development. It has arisen from the belief that current GI systems are limited in their reasoning ability and CBR can be integrated to support this deficiency. The primary use of such a system will be to develop reasoning techniques for discovering knowledge about areas that are considered *spatially similar*. The degree of match to a set of criteria (parameters) and circumstances (application) also influence the degree of similarity. The user also governs similarity as they select a set of criteria, defines circumstances and biases the appropriate criteria to achieve the desired result. As a result, based on a set of criteria selected by the user similar instances can be found (Holt 1996). It is not just the attributes that determines similarity. Dubitzky *et al.* (1993) add to this by suggesting that *The relation rather that the objects alone determines to a large degree the similarity between two situations*. This paper builds on this concept by including relations to spatial data. It is the spatial relationships between situations that determine if they are spatially similar or not. Proximity analysis, available in GI systems, allows a relation to be formed between spatial data. This can be used as a similarity measure. The degree of match is the score between a source and a target. In spatial matching a source and a target could be a pixel, region or coverage. The principles that govern spatial similarity are not just the attributes but also the relationships between two phenomena. This is one reason why CBR coupled with a GIS is fortuitous. A GIS is used symbiotically to extract spatial variables that can be used by CBR to determine similar spatial relations between phenomena. These spatial relations are used to assess the similarity between two phenomena (for example proximity and neighbourhood analysis). This paper will illustrate some experiments and results from testing the spatial similarity concept.

4.0 HUNTERVALE EXAMPLE

This example aims to indicate the influence of spatial attributes (from GI systems) on determining spatial similarity. The dataset using ArcView GIS version 3.0a was taken from the AVtutor CAD section using the Cad drawings of data 1996. From the "study area" theme a selection was made this includes Cable road, Douglas circle and Huntervale Avenue (figure 3). A spatial extent was selected and the attributes/entities from the "attributes of bloke attributes" table. These attributes were exported as comma delimited data, which were opened in *Microsoft Excel* and saved as text for further ordering in the next stage. The next stage involved a language used for Case-based Reasoning called CASL. The contents of a case-base are described in a file known as a case file, using the language CASL. The program *Similarity* uses this case file to create a case-base in the computer's memory, which can then be accessed and modified in order to solve problems using Case-Based Reasoning techniques. When new cases are added to the case-base in the computer's memory, they are also appended to the end of the case file. The process of Case-Based Reasoning (CBR) is to find a case that is similar to the current situation, modify the solution to fit the current situation and then to store the case in the case-base. These processes can be carried out using the program *Similarity*. The case base written in CASL was created using *programmers file editor* and then load it into *Similarity*. *Similarity* checks that it is a legal CASL program as it loads the case base.



Attributes of Block Attributes

Owner	St_numb	St_name	Front	Area_sqft	North
NAGLE RICHARD L JR	0005	DOUGLAS CIRCLE	54.000	12625.000	180506
NAGLE ROBERT L	0003	DOUGLAS CIRCLE	41.000	13970.000	180576
NAGLE STEVEN	0001	DOUGLAS CIRCLE	40.000	15250.000	180576
NAGLE TRIPP	0015	HUNTERVALE AVENUE	104.000	9944.000	180502
NAGLE WILLIAM	0011	HUNTERVALE AVENUE	105.000	9740.000	180526
NAGLE WILLIAM B III	0005	HUNTERVALE AVENUE	90.000	10017.000	180597
NAGLE WILLIAM B JR	0152	CABLE ROAD	125.000	18450.000	180754
NAHAI FOAD MD	0158	CABLE ROAD	95.000	10699.000	180711
NAHAI ROUHOLLAH	0164	CABLE ROAD	95.000	10307.000	180630
NAHLIK R M	0172	CABLE ROAD	95.000	10355.000	180586
NAHM JIN WOO	0176	CABLE ROAD	115.000	12231.000	180543
NAIDU RAVI	0470	CENTRAL ROAD	125.000	19087.000	180419
NAIK JIGNESH	0490	CENTRAL ROAD	195.000		180257

Selected records written to cul.txt

Figure 3. The Huntervale example, showing the spatial extent and the database selection.

The case base is called *Hv-case*. The hv_case was run/executed using the case-based reasoning software called *Similarity*. *Similarity* was adapted and compiled by Alec Holt in 1999 from the University of Otago. It was adapted from *Caspian*, which was originally compiled by Ian Pegler in 1997 from the University of Aberystwyth. Caspian is freeware software. *Hv-case* was altered to execute several tests. Tests described below generally show the differences between the *hv-cases* 1-17. The differences are to do with the case definitions which is the case structure (figure 4) and representations of individual case instance (figure5) as well as the indexing and weighting structures. Results of the tests will also be shown.

The global aim of the Huntervale example is to indicate the concepts of spatial similarity. Huntervale is a data set which represents a subdivision (figure 6). The fields in the .dbf table (table 1) are Parcel_id, Map_sheet, Lot_number, Owner, St_number, St_name, Front, Area_acres, Area_sqft, North and East.

```
case definition is
field entity type is (insert, line, polygon);
field st_name type is (huntervale_avenue, douglas_circle, cable_road);
field front type is number weight is 2; ~m
field area type is number weight is 10; ~sqft
field north type is number weight is 15; ~m
field east type is number weight is 15; ~m
end;

index definition is
index on st_name;
```

Figure 4 An example of the case structure

```
case instance parcel_id_81101 is
entity = insert;
st_name = huntervale_avenue;
front = 90;
area = 8570;
north = 180500;
east = 1223907;
solution is
owner = nagendra_setty_atty;
end;
```

Figure 5. An example of a case

Parcel_ID	Area	North	East
81101	8570	180500	1223907
81102	9000	180409	1223959
81103	7517	180349	1223994
81104	9135	180368	1224045
81105	12625	180506	1223975
81106	13970	180576	1224064
81107	15250	180576	1224156
81108	9944	180502	1224226
81109	9740	180526	1224301
81110	10017	180597	1224248
81111	18450	180754	1224215
81112	10699	180711	1224113
81113	10307	180630	1224032
81114	10355	180586	1223945
81115	12231	180543	1223843

Table 1 A summary of the attributes of the case base

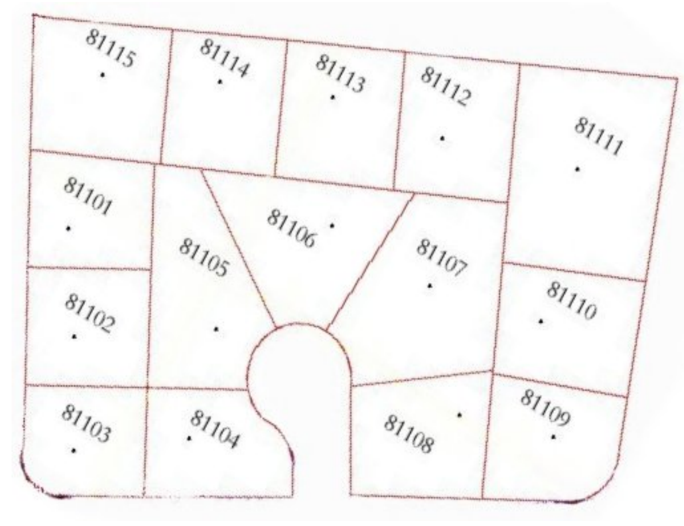


Figure 6 The parcel ID's of huntervale

Test1 was performed to find parcel(s) which have a similar area to the input case. Area was chosen purely to give a starting point to determine the similarity of the other cases and to indicate to influence on spatial properties to the similar area answer. It is acknowledged that area is not a measure of spatial similarity. A test case was chosen for example, parcel number **81111**. The answers to the most similar cases (to test case **81111**) in order, as can be expected were, **81107**, 81106, 81105, 81115, and 81112.

However in test two the same scenario as test one was used but extra criteria were added. Other criteria like giving high preference to parcels nearest to *North* 180349 and *East* 1223994. Then the retrieved cases become, **81105**, 81106, 81107, 81115, and 81112

As we continue to add more criteria, for example, street frontage area, and if we continue to change the weights of the entities (importance weights) allowing for different a context, function and use, the results we alter again. This similarity-matching concept is complex and could aid traditional GIS query and SQL techniques for retrieving information. Understanding the changes in the results (the spatial similarities) according to differences in the queries (spatial and non-spatial attributes) as well as how we could measure spatial similarity is the essence of this paper.

5.0 DISCUSSION

The author considers similarity assessment a useful concept for retrieving and analysing spatial information as it may help researchers describe and explore certain phenomena, its immediate environment and its relationships to other phenomena. This paper will identified that phenomena are similar to each other depending on the type and number of commonalties they share. Similarity is impacted by the spatial properties of entities.

Spatial similarity queries may be utilised in future applications to answer questions such as *Are there spatial phenomena similar to the searched example? Which spatial phenomena have the certain criteria?*

- Agriculture: retrieve the farm layout designs in an archive that are spatially similar to a given farm design,
- Flood diagnostics: retrieve all river coverages and resulting flooding studies that have an area in the same location and or same topology as the one in the query coverage,
- Marketing (spatial product, customer base and supply routes),
- Botany (the classifications of bogland and the modelling of indigenous bush),
- The spatial organisations and groupings at a molecular and chromosome level.

For a list of researchers interested in this topic see http://divcom.otago.ac.nz/sirc/similarity/

ACKNOWLEDGEMENTS

I would like to thank Professor Benwell for the probing questions, thoughtful discussions, and the encouragement to publish this paper. I also express my thanks to Professor Musgrave for his philosophical thoughts, especially his ideas on Karl Popper.

REFERENCES

Bruns, T. & M. Egenhofer (1996) Similarity of Spatial Scenes, in: M. -J. Kraak & M. Molenaar (eds.), *Seventh International Symposium on Spatial Data Handling,* Delft, The Netherlands Taylor & Francis, pp. 173-184.

Carnap, R. (1967). The logical structure of the world. Pseudoproblems in philosophy. Routledge and Kegan Paul, London, UK.

Cobb, M. A., Chung, M. J., Foley III, H., Petry, F. E., Shaw, K. B. & H. V. Miller, (1998), A Rule-based Approach for the Conflation of Attributed vector Data. *GeoInformatica: An International Journal on Advances of Computer Science for Geographical Information Systems*, Vol. 2, Number 1, 7-37.

Dubitzky W. Carville F. & J. Hughes (1993) Case-level Knowledge Modelling in CBR, *Irish Journal of Psychology,* 14:3: 478-479.

Flewelling, D. M. (1997) Comparing Subsets from Digital Spatial Archives: Point Set Similarity. *Ph.D thesis*, University of Maine, Orono, Maine.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D. & P. Yanker, (1995) Query by Image and Video Content: The QBIC System. *IEEE Computer* 28(9): 23-32.

Goodchild, M. F, Egenhofer, M. J. & R. Fegeas., (1998), Interoperating GISs. Report of a Specialist Meeting held under the auspices of the Varenius Project. <http://www.ncgia.ucsb.edu/conf/ interop97/interop_toc.html>

Gudivada, V. (1995) On Spatial Similarity Measures for Multimedia Applications, in: *SPIE,* pp. 363-372.

Gudivada, V. and V. Raghavan (1995) Design and Evaluation of Algorithms for Image Retrieval by Spatial Similarity, *ACM Transactions on Information Systems,* 13 (2): 115-144.

Holt, A. (1996) Incorporating A New Computational Reasoning Approach to Spatial Modelling. *In the proceedings of The 1st International Conference on GeoComputation,* University of Leeds, Leeds, England. 1: 427-442.

Holt, A. (1997) GeoComputation-neologism, or gambit towards progressive research? *Environmental Perspectives, A Quarterly Newsletter published by the Environmental Policy & Management Research Centre,* Issue 15:5-6.

Holt, A. & G. L. Benwell, (1996) Case-based Reasoning and Spatial Analysis. *Journal of the Urban and Regional Information Systems Association,* 8(1) 27-36.

Holt, A. & G. L. Benwell (1997) Using Spatial Similarity for Exploratory Spatial Data Analysis: Some Directions. *The 2nd International Conference on GeoComputation,* University of Otago, Dunedin, New Zealand, pp. 279-288.

Holt, A. & G. L. Benwell (1999) Applying Case-based Reasoning to Spatial Phenomena, *The International Journal of Geographical Information Science 13(1): 9-25.*

Holt, A., Higham, E. C. & G.W. Kearsley, (1997) Predicting International Tourist Flows: Using a Spatial Reasoning System. *The Pacific Tourism Review: An Interdisciplinary Journal.* 1:4 30 pp.

Hume, D. (1739) A treatise of human nature. ed. E. C. Mossner, Harmondsworth Penguin Books, 1969.

Kolodner J. (1993) *Case-Based Reasoning.* San Mateo, Morgan Kaufmann, Publishers.

Lilburne, L. (1998) Scale issues in environmental data modelling. I*nternal Report, Landcare Research NZ Ltd.,* Lincoln, New Zealand.

Medin, D. and Ortony, A. (1989). Comments on Part I: Psychological essentialism. In Similarity and Analogical Reasoning, pages 179--195, New York, NY. Cambridge University Press.

Openshaw, S. & R. J. Abrahart (1996) GeoComputation. *In the proceedings of The 1st International Conference on GeoComputation,* University of Leeds, Leeds, England. 1: 665-666.

Papadias, D. & M. J. Egenhofer, (1997), Algorithms for Hierarchical Spatial Reasoning. *GeoInformatica: An International Journal on Advances of Computer Science for Geographical Information Systems*, Vol. 1, Number 3, 251-274.

Papadias, D. & Delis, B. (1997) Relation-based Similarity. *Proceedings of the 5th ACM Workshop on GIS*, Las Vegas, ACM Press.

Pegler, I. 1996. CASL description document v2.1, University of Wales, Aberystwyth, September 1996 Creating a Case-Base using CASL. 16 pages.

Popper K. R. 1972, The Logic of Scientific Discovery, London: Hutchinson 480 p.

Quine W. V. 1969, Ontological Relativity and Other Essays. New York, Columbia University Press. 165p

Shi, X and Yeh, A. G. O. (1999) The Integration of Case-Based System (CBS) and GIS in Development Control *Environment and Planning B: Planning and Design* vol 26, 345-364.

Tversky, A. (1977) Features of Similarity. *Psychological Review* 84(4): 327–352.

Tversky, A. & D. H. Krantz, (1970) The dimensional representation and the metric structure of similarity data. *Journal of Mathematical Psychology,* Vol. 7. p572-597.

Wallach, M. A. (1958) On Psychological similarity. *Psychological Review*, 65(2):103-116.

Wittgenstein L. (1968) Philosophical Investigations, Basil Blackwell, Oxford, 250pages.